SIAM J. OPTIM.
Vol. 18, No. 1, pp. 223–241

A COPOSITIVE PROGRAMMING APPROACH TO GRAPH PARTITIONING*

JANEZ POVH† AND FRANZ RENDL‡

Abstract. We consider 3-partitioning the vertices of a graph into sets S_1, S_2, and S_3 of specified cardinalities, such that the total weight of all edges joining S_1 and S_2 is minimized. This problem is closely related to several NP-hard problems like determining the bandwidth or finding a vertex separator in a graph. We show that this problem can be formulated as a linear program over the cone of completely positive matrices, leading in a natural way to semidefinite relaxations of the problem. We show in particular that the spectral relaxation introduced by Helmberg et al. (1995) can equivalently be formulated as a semidefinite program. Finally we propose a tightened version of this semidefinite program and show on some small instances that this new bound is a significant improvement over the spectral bound.

Key words. semidefinite programming, copositive programming, graph partitioning problem, bandwidth problem, vertex separator problem

AMS subject classifications. 90C22, 90C27

DOI. 10.1137/050637467

1. Introduction.

We consider the following partition problem on graphs, and we denote it as the MIN-CUT problem (MCP). Let $G = (V, E)$ be an undirected graph on n vertices, given by its (weighted) adjacency matrix $A \geq 0$, so $a_{ij} > 0$ implies the edge $(ij) \in E(G)$ with weight a_{ij}. For given integers m_1, m_2, and m_3 summing to n, we are interested in the following NP-complete problem: find subsets S_1, S_2, and S_3 of $V(G)$ with cardinalities m_1, m_2, and m_3, respectively, such that the total weight of edges between S_1 and S_2 is minimal. More formally, let (S_1, S_2, S_3) be a partition of V with $|S_i| = m_i$ for $i = 1, 2, 3$. The total weight of edges between sets S_1 and S_2 will be denoted as $\mathrm{cut}(S_1, S_2)$. Hence

$$\mathrm{cut}(S_1, S_2) = \sum_{i \in S_1,\, j \in S_2} a_{ij}.$$

We define the MCP as the following optimization problem:

$$(\mathrm{MCP}) \qquad \begin{array}{rl} \min & \mathrm{cut}(S_1, S_2) \\ \text{such that} & (S_1, S_2, S_3) \text{ partitions } V(G) \\ & \text{and } |S_i| = m_i,\ i = 1, 2, 3. \end{array}$$

The optimal value of this problem will be denoted as OPT_{MC}.

REMARK 1. *If $m_1 = 0$ or $m_2 = 0$, then the MCP is trivial: $OPT_{MC} = 0$. Therefore, we assume from now on that $1 \leq m_1 \leq m_2$. If $m_3 = 0$, $m_1 = \lfloor \frac{n}{2} \rfloor$, and $m_2 = \lceil \frac{n}{2} \rceil$, we get the NP-complete bisection problem as a special case (see [7]).*

*Received by the editors August 3, 2005; accepted for publication (in revised form) September 15, 2006; published electronically March 19, 2007. Partial support by the Marie Curie Research Training Network MCRTN-CT-2003-504438 (ADONET) is gratefully acknowledged.

http://www.siam.org/journals/siopt/18-1/63746.html

†Faculty of Logistics, University in Maribor, Mariborska cesta 3, Celje, Slovenia (janez.povh@uni-mb.si).

‡Institut für Mathematik, Universität Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt, Austria (franz.rendl@uni-klu.ac.at).

The MCP by itself may seem like an artificial optimization problem. It can however serve as a powerful tool to solve some fundamental graph optimization problems. It is connected to the (balanced) vertex separator problem, where the objective is to find a minimal subset of $V(G)$, whose removal disconnects the graph into two subgraphs of roughly equal size. If $OPT_{MC} = 0$, then the graph G underlying A has a *vertex separator* of size m_3 and its connectivity is at most m_3 (see [10] for more details). On the other hand, if $OPT_{MC} > 0$, then the *bandwidth* of the matrix A is at least $m_3 + 1$ (see [10]). The ability to solve (or at least approximate) the MCP therefore has strong impact on these graph problems. The MCP is a special instance of more general graph partitioning problems, where one is interested in a partition of $V(G)$ into k disjoint subsets $S_1, \ldots, S_k$ with cardinalities $m_1 \leq m_2 \leq \cdots \leq m_k$, $\sum_i m_i = |V(G)|$, such that the total weight of edges between some subsets is minimized. A survey on the graph partitioning problem and related problems is given in [13]. Graph partitioning, bandwidth minimization, and vertex separator problems appear in a wide range of applications, from numerical linear algebra to floor planning and analysis of bottlenecks in communication networks. In parallel computing, partitioning the set of tasks among processors in order to minimize the communication between processors is another instance of a graph partitioning problem. A comprehensive survey with results in this area up to 1995 is contained in [1]. Formulating the partitions using vertex variables leads to a quadratic cost function with linear and quadratic constraints in binary variables; see (1)–(5) below. Maintaining the orthogonality condition (2) leads to spectral relaxations based on the Hoffman–Wielandt inequality; see [10, 17]. In [10, 17], these relaxations are investigated for the MCP. The quality of this approach has also been studied in [8]. The spectral relaxation of the MCP from [10] is attractive because of the closed form optimal solution; see (15) in section 3 below. The drawback of this model however is that further refinements, like adding sign constraints, make it intractable. It is the purpose of the present paper to overcome these difficulties, and extend and strengthen the spectral approach. Here are our main contributions.

(i) We first formulate the MCP as a linear program over the cone of completely positive matrices; see section 2. This does not make the problem tractable, since linear optimization over this cone is NP-hard [14], but suggests a new family of tractable relaxations, which we get by approximating the copositive constraint with a tractable one, for example, by using the hierarchy of cones, suggested by Parrilo [15], which approximates the cone of completely positive matrices arbitrarily close.

(ii) The conic formulation of the MCP leads to various semidefinite programming (SDP) relaxations of increasing complexity and strength. In section 3 we show that the spectral model from [10] corresponds to a specific semidefinite program, obtained by approximating the cone of completely positive matrices by the cone of positive semidefinite matrices. The proof of this result is rather involved, and we break it down into several smaller steps. It is given in section 4. As in [10] we provide a closed form solution of this semidefinite program (subsection 4.3).

(iii) Finally, we investigate further tightenings of the SDP relaxation in section 5. This opens the way to powerful new approximations of the bandwidth and vertex separator problems. We provide some preliminary computational results which clearly indicate the potential of the new relaxations.

We point out that similar results have been shown recently for other combinatorial optimization problems. De Klerk and Pasechnik [11] have shown that computing the stability number of a graph is equivalent to solving a copositive program. Anstreicher and Wolkowicz [2] have shown that the spectral relaxation of the quadratic assignment problem can equivalently be formulated as a semidefinite program. SDP also turned out to be a useful tool to get tractable relaxations for the graph partitioning problem (see [19]) and the vertex separator problem (see [6]). Finally, de Sousa and Balas have recently proposed an integer linear programming approach combined with a branch and cut algorithm to get minimal balanced vertex separators; see [18].

1.1. Notation. We denote the ith standard unit vector by e_i, while the vector of all ones is $u_n \in R^n$ (or u if dimension n is obvious). The square matrix of all ones is J_n (or J) and the identity matrix is $I = (\delta_{ij})$. We set $E_{ij} = e_i e_j^T$ and its symmetrization is $B_{ij} = \frac{1}{2}(E_{ij} + E_{ji})$. In this paper we consider the following sets of matrices. The vector space of real symmetric $n \times n$ matrices is denoted by $\mathcal{S}_n = \{X \in R^{n\times n} \colon X = X^T\}$. The cone of $n \times n$ positive semidefinite matrices is $\mathcal{S}_n^+ = \{X \in \mathcal{S}_n \colon y^T X y \geq 0 \ \forall y \in R^n\}$. The cone of $n \times n$ copositive matrices is denoted by $\mathcal{C}_n = \{X \in \mathcal{S}_n \colon y^T X y \geq 0 \ \forall y \in R^n_+\}$, the cone of $n \times n$ completely positive matrices is $\mathcal{C}_n^* = \{X = \sum_{i=1}^k y_i y_i^T,\ k \geq 1,\ y_i \in R^n_+ \ \forall i = 1, \ldots, k\}$, and the cone of $n \times n$ symmetric nonnegative matrices is $\mathcal{N}_n = \{X \in \mathcal{S}_n \colon x_{ij} \geq 0 \ \forall i, j\}$. We also use $X \succeq 0$ for $X \in \mathcal{S}_n^+$ and $X \geq 0$ for an elementwise nonnegative matrix. A linear program over $\mathcal{S}_n^+$ is called a semidefinite program, while a linear program over $\mathcal{C}_n$ or $\mathcal{C}_n^*$ is called a copositive program.

The sign $\otimes$ stands for Kronecker product, while the matrices V_i and W_j denote $V_i = e_i u_3^T \in R^{3\times 3}$, $W_j = e_j u_n^T \in R^{n\times n}$, $1 \leq i \leq 3$, $1 \leq j \leq n$. When we consider a matrix $X \in R^{m\times n}$ as a vector from R^{mn}, we write this vector as $\mathrm{vec}(X)$ or x. The $\langle \cdot, \cdot \rangle$ denote the standard scalar product. For $u, v \in R^n$ we have $\langle u,\, v\rangle = u^T v$ and for $X, Y \in R^{m\times n}$ we have $\langle X,\, Y\rangle = \mathrm{trace}(X^T Y)$. For matrix columns and rows we will use MATLAB notation; hence $X(i,:)$ and $X(:,i)$ will stand for the ith row and column, respectively. If $a \in R^n$, then $\mathrm{Diag}(a)$ is an $n \times n$ diagonal matrix with a on the main diagonal. When P is the name of the optimization problem then OPT_P denotes its optimal value.

2. The MCP as a conic linear program. We first use the partition formulation of the MCP to express the MCP as a quadratic program in nonnegative variables. Following [10] we represent partitions (S_1, S_2, S_3) of $V(G)$ by $n \times 3$ matrices X, where

$$x_{ij} = \begin{cases} 1 & \text{if} \quad i \in S_j, \\ 0 & \text{if} \quad i \notin S_j. \end{cases}$$

It will also be useful to identify columns of X directly; hence we denote the ith column of X by x_i. Using X, we can easily express $\mathrm{cut}(S_1, S_2)$ as

$$\mathrm{cut}(S_1, S_2) = x_1^T A x_2 = \frac{1}{2}\langle X,\, AXB\rangle, \tag{1}$$

where

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In [17] it is shown that an $n \times 3$ matrix X represents a partition of $V(G)$ into subsets S_1, S_2, and S_3 of prescribed sizes $m = (m_1, m_2, m_3)^T$ if and only if X satisfies the following relations:

$$X^T X = \mathrm{Diag}(m) =: M, \tag{2}$$

$$X u_3 = u_n, \tag{3}$$

$$X \geq 0. \tag{4}$$

Note in particular that the constraint

$$X^T u_n = m, \tag{5}$$

asking that each partition block has the right number of elements, is implied by these conditions. The set of all $n \times 3$ matrices, representing some partition of $V(G)$ into sets of cardinalities, specified by m, will be denoted by $\mathcal{F}$. Using the above characterization of such partition matrices, we have

$$\mathcal{F} = \{X \in R^{n\times 3};\ X \text{ satisfies (2)–(4)}\}.$$

The MCP can equivalently be written as a quadratic program:

$$(MC_{QP}) \qquad \min\ \tfrac{1}{2}\langle X,\, AXB\rangle \text{ such that } X \in \mathcal{F}.$$

This problem has a nonconvex objective function, defined over a finite set. Our main goal in this section is to transform this problem into an equivalent linear program over the cone of completely positive matrices. We do this by expressing the linear constraints in an appropriate way as quadratic ones. Then we linearize the resulting quadratic terms. Specifically, we consider the following equations in the variable $X \in R^{n\times 3}$:

$$(e_i^T X u_3)^2 = \left(\sum_k X_{ik}\right)^2 = 1,\ 1 \leq i \leq n, \tag{6}$$

$$(u_n^T X e_j)(e_i^T X u_3) = \left(\sum_k X_{kj}\right)\left(\sum_k X_{ik}\right) = m_j,\ 1 \leq i \leq n,\ 1 \leq j \leq 3, \tag{7}$$

$$(u_n^T X e_i)(u_n^T X e_j) = \left(\sum_k X_{ki}\right)\left(\sum_k X_{kj}\right) = m_i m_j,\ 1 \leq i \leq j \leq 3. \tag{8}$$

Equations (6) are obtained by squaring the equations from (3). The equations (7) are obtained by elementwise multiplication of (3) and (5). The last set of equations is obtained from pairwise multiplication of (5). Clearly, any $X \in \mathcal{F}$ will satisfy (6)–(8). Using the Kronecker product and the property $\mathrm{vec}(PXQ) = (Q^T \otimes P)\,\mathrm{vec}(X)$ we get

$$\langle X,\, PXQ\rangle = \mathrm{vec}(X)^T \,\mathrm{vec}(PXQ) = x^T(Q^T \otimes P)x = \langle Q^T \otimes P,\, xx^T\rangle.$$

This helps us to reformulate the constraints (6)–(8) as follows:

$$\begin{cases} (e_i^T X u_3)^2 & = & \langle X,\, e_i e_i^T X u_3 u_3^T\rangle & = & \langle J_3 \otimes E_{ii},\, xx^T\rangle, \\ (u_n^T X e_i)(e_j^T X u_3) & = & \langle X,\, u_n e_j^T X u_3 e_i^T\rangle & = & \langle V_i \otimes W_j^T,\, xx^T\rangle, \\ (u_n^T X e_i)(u_n^T X e_j) & = & \langle X,\, u_n u_n^T X e_j e_i^T\rangle & = & \langle E_{ij} \otimes J_n,\, xx^T\rangle. \end{cases} \tag{9}$$

In the last term we may replace E_{ij} with B_{ij}, since xx^T is symmetric. Similarly, we can rewrite the (i, j)th component on the left-hand side of (2) as

$$e_i^T X^T X e_j \;=\; \langle Xe_i,\, Xe_j\rangle \;=\; \langle X,\, XE_{ji}\rangle \;=\; \langle E_{ij} \otimes I,\, xx^T\rangle.$$

Let us now introduce $Y = xx^T$. Then MC_{QP} can be equivalently formulated as follows:

$$\min \quad \frac{1}{2}\langle B^T \otimes A, Y\rangle$$

$$\text{s. t.} \quad \langle B_{ij} \otimes I, Y\rangle = m_i\delta_{ij}, \quad 1 \le i \le j \le 3, \tag{10}$$

$$\langle J_3 \otimes E_{ii}, Y\rangle = 1, \quad 1 \le i \le n, \tag{11}$$

$$\langle V_i \otimes W_j^T, Y\rangle = m_i, \quad 1 \le i \le 3,\ 1 \le j \le n, \tag{12}$$

$$\langle B_{ij} \otimes J_n, Y\rangle = m_i m_j, \quad 1 \le i \le j \le 3, \tag{13}$$

$$Y = xx^T, \quad x \in R_+^{3n}.$$

To see that this optimization problem is equivalent to MC_{QP}, we note that for any X feasible for MC_{QP}, we can take $x = \text{vec}(X)$ to get a feasible $Y = xx^T$ for this problem with the same objective value and vice versa. The above problem has linear objective, linear constraints, and the quadratic equation, coupling Y and x. As a final simplification, we replace the constraints $Y = xx^T$ and $x \ge 0$ by $Y \in \mathcal{C}_{3n}^*$. The new optimization problem, which is a copositive program, will be denoted by MC_{CP}:

$$(MC_{CP}) \qquad \min \frac{1}{2}\langle B^T \otimes A, Y\rangle \text{ such that } Y \in \mathcal{C}_{3n}^* \text{ satisfies (10)–(13).}$$

The following theorem explains the relation between the feasible sets of MC_{QP} and MC_{CP}.

THEOREM 1.

$$\text{CONV}\{xx^T;\, x \in R_+^{3n},\, xx^T \textit{ feasible for } (10)\text{–}(13)\}$$
$$= \{Y \in \mathcal{C}_{3n}^*;\, Y \textit{ feasible for } (10)\text{–}(13)\}.$$

Proof. The "$\subseteq$" inclusion is obvious. To show inclusion in the other direction, we have to prove that for any $Y \in \mathcal{C}_{3n}^*$, feasible for MC_{CP}, there exist finitely many vectors $y^1, y^2, \ldots \in R_+^{3n}$ and numbers $\lambda_k \in [0, 1]$ with $\sum_k \lambda_k = 1$ such that $y^k(y^k)^T$ are feasible for constraints (10)–(13) and $Y = \sum_k \lambda_k y^k(y^k)^T$. Let $Y \in \mathcal{C}_{3n}^*$. From the definition of the cone $\mathcal{C}_{3n}^*$ follows that there exist finitely many nonzero vectors $x^k \in R_+^{3n}$ such that $Y = \sum_k x^k(x^k)^T$. We can treat x^k as a vector representation of some matrix $X^k \in R^{n\times 3}$; therefore we will index the components of each x^k with two indices: $x^k = (x_{ij}^k)$, $i = 1, \ldots, n$ and $j = 1, 2, 3$ (components $x_{:1}^k$ are the first n components of x^k—the first "column" of x^k, etc.). Let us first fix i and j ($1 \le i \le n$, $1 \le j \le 3$). If we denote with $r_k = \sum_{s=1}^3 x_{is}^k$ the sum of the "ith row" of x^k and with $c_k = \sum_{s=1}^n x_{sj}^k$ the sum of the "jth column" of x^k, then we can rewrite the constraints (11)–(13) using (6)–(9) as

$$\textstyle\sum_k r_k^2 \;=\; 1,\ \sum_k r_k c_k \;=\; m_j,\ \sum_k c_k^2 \;=\; m_j^2.$$

The Cauchy inequality, applied to vectors $v_1 = (r_1, r_2, \ldots)$ and $v_2 = (c_1, c_2, \ldots)$, implies $r_k = c_k/m_j$, or equivalently

$$\sum_s x_{is}^k = \frac{\sum_s x_{sj}^k}{m_j}, \quad k = 1, 2, \ldots. \tag{14}$$

Since this is true for all i and j, we can see that the numbers $\sum_s x_{sj}^k/m_j$ are equal for all j. This means that in any vector x^k the sum of any "row" is equal to the sum of column j divided by m_j for all $j = 1, 2, 3$. Therefore we may take without loss of generality $j = 1$ and define $\alpha_k = \sum_s x^k_{s1}/m_1$. Since none of x^k is zero we have $\alpha_k > 0$ for all k, and we may define $\lambda_k = \alpha_k^2 = (\sum_s x^k_{s1})^2/m_1^2$ and $y^k = x^k/\alpha_k$. From (13) we get

$$\sum_k \lambda_k = \frac{1}{m_1^2}\sum_k \left(\sum_s x^k_{s1}\right)^2 = 1.$$

Equation (14) implies that $y^k(y^k)^T$ are feasible for (11)–(13) and $Y = \sum_k \lambda_k y^k(y^k)^T$. To finish the proof it remains to show that $y^k(y^k)^T$ is feasible for (10) for all k. Indeed, if there exist $i \neq j$ and k such that $\langle B_{ij} \otimes I,\, y^k(y^k)^T\rangle > 0$, then because of nonnegativity of y^k we have $\langle B_{ij} \otimes I,\, Y\rangle > 0$, but this is a contradiction to the feasibility of Y. In particular, this means that in each "row" of y^k there is only one nonzero component, which must be equal to 1 because of feasibility for (11). Hence y^k is a 0–1 vector. This implies together with (13) that $\langle E_{ii} \otimes I,\, y^k(y^k)^T\rangle = \sum_s (y^k_{si})^2 = \sum_s y^k_{si} = m_i$; hence $y^k(y^k)^T$ is feasible for (10), too. ∎

The feasible set of MC_{CP} is therefore a polytope, spanned by the rank 1 matrices of type xx^T, where x is a vector representation of matrix X, feasible for MC_{QP}. Since MC_{CP} is a linear program, it has a rank 1 optimal solution; hence $OPT_{CP} \geq OPT_{QP}$. The opposite direction is obvious; hence we have the following corollary.

Corollary 2. *Problems MC_{QP} and MC_{CP} have the same optimal value; therefore the MCP can be equivalently formulated as a linear program in completely positive matrices.*

Remark 2. *This copositive representation again confirms the importance of copositive programming in combinatorial optimization which was revealed by de Klerk and Pasechnik* [11]*, who proved that computing the stability number of a graph is equivalent to solving a copositive program and then presented a hierarchy of positive semidefinite relaxations, which follow from this approach and are strongly connected with the ϑ-function.*

3. The spectral relaxation as a semidefinite program. Helmberg et al. have derived in [10] a lower bound for OPT_{MC} which is easy to compute. They have omitted the nonnegativity constraint (4) in MC_{QP} and added constraint (5), yielding the problem

$$OPT_{HW} \quad = \quad \min\ \tfrac{1}{2}\langle X,\, \hat{A}XB\rangle \ \text{ such that } X \text{ satisfies (2), (3), and (5).}$$

In the above formulation we introduced $\hat{A} = A + D$ with $D = \frac{s(A)}{n} I - \mathrm{Diag}(r(A))$ and $s(A) = u^T A u$, $r(A) = Au$. This is a quadratic problem defined over a nonconvex set described by linear and quadratic equations. If we replace in the models MC_{QP} and MC_{CP} matrix A with $\hat{A}$, then the optimal values of these models do not change, since matrix XBX^T in the model MC_{QP} has only zeros on the main diagonal and similarly any feasible matrix Y in model MC_{CP} has only zeros on the main diagonals of off-diagonal blocks, as follows from (10) and complete positiveness of Y. Therefore $OPT_{MC} \geq OPT_{HW}$. Helmberg et al. [10] have in fact shown that OPT_{HW} has the explicit form

$$OPT_{HW} = -\frac{1}{2}\mu_2\lambda_2 - \frac{1}{2}\mu_1\lambda_n, \tag{15}$$

where λ_2 and λ_n are second smallest and the largest Laplacian eigenvalues of the graph G (i.e., the eigenvalues of matrix $L = \mathrm{Diag}(r(A)) - A = \frac{s(A)}{n} I - \hat{A}$) and $\mu_1 \geq \mu_2$ are defined as

$$\mu_{1,2} = \frac{1}{n}\Big(-m_1 m_2 \pm \sqrt{m_1 m_2 (n-m_1)(n-m_2)}\Big). \tag{16}$$

The key tool to get this result was the Hoffman–Wielandt inequality [9] combined with a projection technique for partitioning the nodes of a graph from [17]. It is an attractive feature of this bound that the closed form solution (15) is quite easy to compute, as it involves only the computation of the extreme Laplacian eigenvalues. On the other hand, the relaxation OPT_{HW}, as described above, does not permit the inclusion of further constraints, like, for instance, $X \geq 0$, without losing tractability. One of the main motivations for the current research was in fact the search for a new equivalent formulation of OPT_{HW} which is suitable for further tractable refinements. We now propose such a refinement. As already mentioned, we do not change the optimal value by replacing A with $\hat{A}$ in the models MC_{QP} and MC_{CP}. Let us consider the model, obtained from MC_{CP} by this replacement and relaxing the constraint $Y \in \mathcal{C}^*_{3n}$ to $Y \in \mathcal{S}^+_{3n}$. We will denote it by MC_{SDP} and its optimal value by OPT_{SDP}. Hence

$$(MC_{SDP}) \qquad \begin{array}{rcll} OPT_{SDP} & = & \min & \frac{1}{2}\langle B^T \otimes \hat{A},\, Y\rangle \\ & & \text{s. t.} & Y \;\in\; \mathcal{S}^+_{3n} \\ & & & Y \text{ satisfies } (10)\text{–}(13). \end{array}$$

In the next section we will show that the value OPT_{SDP} is equal to OPT_{HW}. First we have the easy part.

LEMMA 3.

$$OPT_{HW} \geq OPT_{SDP}.$$

Proof. If X satisfies (2), (3), and (5), then X satisfies constraints (2) and (6)–(8). The matrix $Y = xx^T$ satisfies (10)–(13) and is in $\mathcal{S}^+_{3n}$ hence is feasible for MC_{SDP}. Since $\frac{1}{2}\langle X, \hat{A}XB\rangle = \frac{1}{2}\langle B^T \otimes \hat{A}, Y\rangle$, the lemma follows. ◻

The main result of this section is the following theorem.

THEOREM 4.

$$OPT_{HW} = OPT_{SDP}.$$

From Lemma 3 it follows that we need to prove that $OPT_{HW} \leq OPT_{SDP}$. Our proof of this result is rather involved and consists of two major steps. In the first step we reformulate the semidefinite program MC_{SDP} in a new coordinate system, obtained by diagonalizing the cost matrix $B^T \otimes \hat{A}$. This is the content of subsection 4.1, which ends with the main result of the first step, i.e., with the semidefinite program (18). The second part of the proof is more subtle. We extract a subproblem of (18) by leaving out some constraints and projecting the feasible set to a proper hyperplane. In subsection 4.2 we show that this subproblem, denoted by MC_{SDPa}, in fact captures the essential part of MC_{SDP}, and its optimal solution OPT_{SDPa} satisfies $OPT_{SDP} \geq OPT_{SDPa} \geq OPT_{HW}$. This closes the chain of inequalities and the proof is finished.

4. Proof of Theorem 4.

4.1. Diagonalization of the cost matrix. Let $\frac{1}{2}\hat{A} = PSP^T$, $B = QTQ^T$, where P and Q are orthonormal matrices whose columns are eigenvectors of $\frac{1}{2}\hat{A}$ and B, respectively, and S, T are diagonal matrices with eigenvalues on the diagonal. We take the factorizations where the eigenvalues are in nondecreasing order; hence we have

$$Q = \frac{1}{2}\begin{bmatrix} -\sqrt{2} & 0 & \sqrt{2} \\ \sqrt{2} & 0 & \sqrt{2} \\ 0 & 2 & 0 \end{bmatrix},\ T = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If we denote with $\ell_i = s_{ii}$, then from $\hat{A} = \frac{s(A)}{n} I - L$ (see the beginning of the previous section) follows $\ell_i = \frac{s(A)}{2n} - \frac{\lambda_{n-i+1}(L)}{2}$, in particular $\ell_1 = \frac{s(A)}{2n} - \frac{\lambda_n(L)}{2}$ and $\ell_n = \frac{s(A)}{2n}$. We choose P in such a way that the last column of P is equal to $u/\sqrt{n}$. This can be done since u is an eigenvector of $\hat{A}$ corresponding to the largest eigenvalue of $\hat{A}$. In the following lemmas we investigate what happens if we substitute in the model MC_{SDP} the matrix variable Y with matrix variable Z, which are related by

$$Y = (Q \otimes P)\, Z\, (Q \otimes P)^T. \tag{17}$$

This substitution simplifies the objective function, which becomes $\langle T \otimes S,\, Z\rangle$; hence only diagonal elements of Z will determine the objective value. If $Y \in \mathcal{S}_{3n}^+$, then the new matrix variable Z is from $\mathcal{S}_{3n}^+$, too. We will often for the sake of simplicity write matrix Z as a block matrix: $Z = [Z^{ij}]_{1\le i,j\le 3}$, where $Z^{ij} \in R^{n\times n}$. This actually means that

$$Z = \sum_{1\le i,j\le 3} E_{ij} \otimes Z^{ij} = \begin{bmatrix} Z^{11} & Z^{12} & Z^{13} \\ Z^{21} & Z^{22} & Z^{23} \\ Z^{31} & Z^{32} & Z^{33} \end{bmatrix}.$$

We will denote with Z^{ij}_{kl} the (k,l)th component of matrix Z^{ij}.

LEMMA 5. *Let $Y, Z \in \mathcal{S}_{3n}$ satisfy* (17). *The matrix Y satisfies constraint* (13) *if and only if the matrix Z satisfies*

$$Z^{ij}_{nn} = f_{ij},\ 1 \le i \le j \le 3, \tag{13a}$$

where matrix $F = (f_{ij}) \in \mathcal{S}_3^+$ is as follows:

$$F = \frac{1}{2n}\begin{bmatrix} m_2 - m_1 \\ \sqrt{2} m_3 \\ m_2 + m_1 \end{bmatrix} \cdot \begin{bmatrix} m_2 - m_1 \\ \sqrt{2} m_3 \\ m_2 + m_1 \end{bmatrix}^T.$$

Proof. Here we use the fact that $P(:,n) = u/\sqrt{n}$. Constraint (13) becomes $\langle (Q^T B_{ij} Q) \otimes (P^T J_n P),\, Z\rangle = m_i m_j$. Since all columns of P are orthogonal, we have $P^T J_n P = P^T W_n^T = n E_{nn}$. We also get matrices $\tilde{B}_{ij} := Q^T B_{ij} Q$:

$$\tilde{B}_{11} = \tfrac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad \tilde{B}_{12} = \tfrac{1}{2}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{B}_{13} = \tfrac{\sqrt{2}}{4}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \tilde{B}_{22} = \tfrac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\tilde{B}_{23} = \tfrac{\sqrt{2}}{4}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \tilde{B}_{33} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore we get $\langle B_{11} \otimes J_n,\, Y\rangle = n\langle \tilde{B}_{11} \otimes E_{nn},\, Z\rangle = \frac{n}{2}(Z^{11}_{nn} - 2Z^{13}_{nn} + Z^{33}_{nn})$. Equation $\langle B_{11} \otimes J_n,\, Y\rangle = m_1^2$ is thus equivalent to

$$Z^{11}_{nn} - 2Z^{13}_{nn} + Z^{33}_{nn} = \frac{2m_1^2}{n}.$$

Similarly, we rewrite the other equations from constraint (13) into

$$\begin{array}{rclrcl} -Z^{11}_{nn} + Z^{33}_{nn} &=& \frac{2m_1m_2}{n}, & -Z^{12}_{nn} + Z^{23}_{nn} &=& \frac{\sqrt{2}m_1m_3}{n}, \\ Z^{11}_{nn} + 2Z^{13}_{nn} + Z^{33}_{nn} &=& \frac{2m_2^2}{n}, & Z^{12}_{nn} + Z^{23}_{nn} &=& \frac{\sqrt{2}m_2m_3}{n}, \\ Z^{22}_{nn} &=& \frac{m_3^2}{n}. & & & \end{array}$$

The solution of this system of six linear equations in six variables is $Z^{ij}_{nn} = f_{ij}$. □

LEMMA 6. *Let $Y, Z \in \mathcal{S}_{3n}$ satisfy* (17). *The matrix Y satisfies constraint* (10) *if and only if the matrix Z satisfies constraint*

$$\text{trace}(Z^{ij}) = h_{ij},\ 1 \le i \le j \le 3, \tag{10a}$$

where matrix $H = (h_{ij}) \in \mathcal{S}_3$ is defined as

$$H = \frac{1}{2}\begin{bmatrix} m_1 + m_2 & 0 & m_2 - m_1 \\ 0 & 2m_3 & 0 \\ m_2 - m_1 & 0 & m_1 + m_2 \end{bmatrix}.$$

Proof. From $P^T I P = I$ follows $\langle B_{ij} \otimes I,\, Y\rangle = \langle \tilde{B}_{ij} \otimes I,\, Z\rangle$. If $i = j = 1$, then $\langle \tilde{B}_{11} \otimes I,\, Z\rangle = (\text{trace}(Z^{11}) - 2\,\text{trace}(Z^{13}) + \text{trace}(Z^{33}))/2$, so the first equation from (10) could be rewritten as

$$\frac{1}{2}\Big(\text{trace}(Z^{11}) - 2\,\text{trace}(Z^{13}) + \text{trace}(Z^{33})\Big) = m_1.$$

Similarly, we get the other five linear equations in six variables $\text{trace}(Z^{ij})$. The unique solution is given by $\text{trace}(Z^{ij}) = h_{ij}$, $1 \le i \le j \le 3$. □

LEMMA 7. *Let $Y, Z \in \mathcal{S}_{3n}$ satisfy* (17).

(a) *The matrix Y satisfies constraint* (11) *if and only if the matrix Z satisfies the constraint*

$$\Big\langle U \otimes P(i,:)^T P(i,:),\, Z\Big\rangle = 1,\ 1 \le i \le n, \tag{11a}$$

where

$$U = Q^T J_3 Q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & \sqrt{2} \\ 0 & \sqrt{2} & 2 \end{bmatrix}.$$

(b) *The matrix Y satisfies constraint* (12) *if and only if the matrix Z satisfies the constraint*

$$\langle \tilde{V}_i \otimes (e_n \cdot P(j,:)),\, Z\rangle = \frac{m_i}{\sqrt{n}},\ 1 \le i \le 3,\ 1 \le j \le n, \tag{12a}$$

where $\tilde{V}_i = Q^T V_i Q$.

Proof. (a) This statement follows immediately from $P^T E_{ii} P = P(i,:)^T P(i,:)$.

(b) After the substitution the left-hand side of constraint (12) becomes $\langle (Q^T V_i Q) \otimes (P^T W_j^T P),\, Z\rangle = m_i$. A short calculation shows

$$\tilde{V}_1 = \frac{1}{2}\begin{bmatrix} 0 & -\sqrt{2} & -2 \\ 0 & 0 & 0 \\ 0 & \sqrt{2} & 2 \end{bmatrix}, \quad \tilde{V}_2 = \frac{1}{2}\begin{bmatrix} 0 & \sqrt{2} & 2 \\ 0 & 0 & 0 \\ 0 & \sqrt{2} & 2 \end{bmatrix}, \quad \tilde{V}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 0 \end{bmatrix}.$$

The term $P^T W_j^T P$ simplifies because of the choice of the last column of P into $\sqrt{n} E_{nj} P = \sqrt{n} e_n e_j^T P = \sqrt{n} e_n P(j,:)$. ∎

By introducing the set

$$\mathcal{G} = \{Z \in \mathcal{S}_{3n}^+,\ Z \text{ satisfies constraints (10a), (11a), (12a), and (13a)}\},$$

we can see that the problem MC_{SDP} is equivalent to the problem

$$\min \langle T \otimes S,\, Z\rangle \quad \text{such that } Z \in \mathcal{G}, \tag{18}$$

since for any feasible solution Y for MC_{SDP} we can find a solution $Z \in \mathcal{G}$ via (17) with the same value of the objective value and vice versa. It should be noted that the cost function in (18) simplifies to $\langle T \otimes S,\, Z\rangle = \sum_{i=1}^n \ell_i (Z_{ii}^{33} - Z_{ii}^{11})$.

4.2. A block-diagonal subproblem. The semidefinite program (18) is still quite complicated. Since Lemmas 5–7 show that feasibility for constraints (10a)–(13a) is mostly determined with the diagonal entries of blocks Z^{ij}, we are going to study the following semidefinite program, which we obtain by keeping in the program (18) only constraints (10a) and (13a) and ignoring all nondiagonal components in any block Z^{ij}. We also omit blocks Z^{2i} and Z^{i2}, $i = 1, 2, 3$, since they do not contribute to the cost function.

$$(MC_{SDPa}) \qquad \begin{array}{lrcl} \min & \sum_{i=1}^{n-1} \ell_i (r_i - p_i) + \ell_n (f_{33} - f_{11}) & & \\ \text{s. t.} & \sum_{i=1}^{n-1} p_i & = & h_{11} - f_{11} := b_1, \\ & \sum_{i=1}^{n-1} r_i & = & h_{33} - f_{33} := b_2, \\ & \sum_{i=1}^{n-1} q_i & = & h_{13} - f_{13} := b_3, \\ & U_i & = & \begin{bmatrix} p_i & q_i \\ q_i & r_i \end{bmatrix} \succeq 0. \end{array}$$

The constants b_i are

$$b_1 = \frac{4m_1 m_2 + m_1 m_3 + m_2 m_3}{2n}, \quad b_2 = \frac{(m_1 + m_2) m_3}{2n}, \quad \text{and} \quad b_3 = \frac{(m_2 - m_1) m_3}{2n}.$$

In the following lemma we compare the optimal values of MC_{SDP} and MC_{SDPa}.

LEMMA 8.

$$OPT_{SDP} \geq OPT_{SDPa}.$$

Proof. We will show that any feasible solution for (18) implies a feasible solution for MC_{SDPa}. Let $Z = [Z^{ij}]$ be a feasible solution for (18) and let us define $p_i = Z_{ii}^{11}$, $r_i = Z_{ii}^{33}$, and $q_i = Z_{ii}^{13}$ for $1 \leq i \leq n-1$. From Lemmas 5 and 6 follows $\sum_{i=1}^{n-1} p_i = \operatorname{trace}(Z^{11}) - Z_{nn}^{11} = b_1$, and hence p_i are feasible for the first equation in MC_{SDPa}. Similarly we can show that the other two constraints are satisfied and that the matrices $U_i = \left[\begin{smallmatrix} p_i & q_i \\ q_i & r_i \end{smallmatrix}\right]$ are positive semidefinite, following from $Z \succeq 0$. The objective value of the MC_{SDPa} is exactly $\sum_{i=1}^{n} \ell_i(Z_{ii}^{11} - Z_{ii}^{33}) = \langle T \otimes S, Z\rangle$; hence the lemma follows. □

Here is the dual semidefinite program for MC_{SDPa}:

$$(DMC_{SDPa}) \qquad \begin{aligned} &\max && b_1y_1 + b_2y_2 + 2b_3y_3 + \ell_n(f_{33} - f_{11}) \\ &\text{s. t.} && V_i = \begin{bmatrix} -\ell_i - y_1 & -y_3 \\ -y_3 & \ell_i - y_2 \end{bmatrix} \succeq 0, \quad 1 \le i \le n-1. \end{aligned}$$

First let us introduce the number

$$\delta = \frac{2m_1m_2 + m_1m_3 + m_2m_3}{2\sqrt{m_1m_2(n-m_1)(n-m_2)}} = \frac{m_1(n-m_1) + m_2(n-m_2)}{2\sqrt{m_1m_2(n-m_1)(n-m_2)}}.$$

This number is well defined in view of Remark 1. Note also that δ is of the form

$$\frac{1}{2}\left(u + \frac{1}{u}\right) \text{ with } u = \sqrt{\frac{m_1(n-m_1)}{m_2(n-m_2)}} > 0.$$

Therefore $\delta \ge 1$. The next lemma allows us to finish the proof of Theorem 4. We need the following simple observation for its proof.

PROPOSITION 9. *If $a + b = c + d$ and $|a - b| \le |c - d|$, then $ab \ge cd$.*

Proof. We can write $(a-b)^2 = (a+b)^2 - 4ab$ and $(c-d)^2 = (c+d)^2 - 4cd$. Using the assumptions of the proposition we get $(a+b)^2 - 4ab \ge (c+d)^2 - 4cd$ and the result follows. □

LEMMA 10. *The numbers*

$$\begin{aligned} y_1 &= -\frac{\ell_1 + \ell_{n-1}}{2} - \frac{\delta}{2}(\ell_{n-1} - \ell_1), \\ y_2 &= \frac{\ell_1 + \ell_{n-1}}{2} - \frac{\delta}{2}(\ell_{n-1} - \ell_1), \\ y_3 &= \sqrt{(-\ell_1 - y_1)(\ell_1 - y_2)} \end{aligned}$$

form an optimal solution for the dual problem DMC_{SDPa} with objective value equal to OPT_{HW}.

Proof. First note that $\delta \ge 1$ implies $y_1 \le -\ell_{n-1}$ and $y_2 \le \ell_1$. This shows that in the definition of y_3 we take the square root of a nonnegative number; hence y_3 is well defined. To see that $V_i \succeq 0$, we first note that the numbers ℓ_i are in nondecreasing order; therefore $-\ell_i - y_1 \ge 0$, $\ell_i - y_2 \ge 0$. Using that $y_2 = y_1 + \ell_1 + \ell_{n-1}$ we get $y_3^2 = (-\ell_1 - y_1)(\ell_1 - y_2) = (-\ell_{n-1} - y_1)(\ell_{n-1} - y_2)$ and $(-\ell_i - y_1) + (\ell_i - y_2) = -y_1 - y_2 = \delta(\ell_{n-1} - \ell_1)$. Since $|(-\ell_i - y_1) - (\ell_i - y_2)| = |\ell_1 + \ell_{n-1} - 2\ell_i| \le |\ell_{n-1} - \ell_1| = |(-\ell_1 - y_1) - (\ell_1 - y_2)|$, we get by Proposition 9

$$(-\ell_i - y_1)(\ell_i - y_2) \ge (-\ell_1 - y_1)(\ell_1 - y_2) = (-\ell_{n-1} - y_1)(\ell_{n-1} - y_2) = y_3^2;$$

hence $\det(V_i) \ge 0$ and positive semidefiniteness of V_i follows. Second we will show the optimality of (y_1, y_2, y_3). It is sufficient to prove that

$$b_1y_1 + b_2y_2 + 2b_3y_3 + \ell_n(f_{33} - f_{11}) = -\frac{1}{2}\mu_2\lambda_2 - \frac{1}{2}\mu_1\lambda_n = OPT_{HW},$$

since from Lemmas 3 and 8 and the weak duality property we know that the optimal value of DMC_{SDPa} is at most $OPT_{SDP} \le OPT_{HW}$. Using the fact that

$\ell_1 = \frac{s(A)}{2n} - \frac{\lambda_n}{2}$, $\ell_{n-1} = \frac{s(A)}{2n} - \frac{\lambda_2}{2}$, and $\ell_n(f_{33} - f_{11}) = \frac{s(A)m_1m_2}{n^2}$, it remains to show that

$$b_1y_1 + b_2y_2 + 2b_3y_3 = \mu_2\ell_{n-1} + \mu_1\ell_1.$$

One can derive that

$$y_3 = \sqrt{(-\ell_1 - y_1)(\ell_1 - y_2)} = \frac{\ell_{n-1}-\ell_1}{2}\sqrt{\delta^2 - 1} = \frac{m_3(m_2-m_1)(\ell_{n-1}-\ell_1)}{4\sqrt{m_1m_2(n-m_1)(n-m_2)}}$$

and show

$$\begin{aligned} b_1y_1 + b_2y_2 + 2b_3y_3 &= \tfrac{\ell_1}{2}\Big(\delta(b_1 + b_2) - b_1 + b_2 - 2b_3\frac{m_3(m_2-m_1)}{2\sqrt{m_1m_2(n-m_1)(n-m_2)}}\Big) \\ &\quad - \tfrac{\ell_{n-1}}{2}\Big(\delta(b_1 + b_2) + b_1 - b_2 - 2b_3\frac{m_3(m_2-m_1)}{2\sqrt{m_1m_2(n-m_1)(n-m_2)}}\Big) \\ &= \mu_1\ell_1 + \mu_2\ell_{n-1}. \end{aligned}$$

Checking the last equality involves tedious but straightforward algebraic manipulation. ∎

Proof of Theorem 4. From Lemmas 3, 8, and 10 and the weak duality property for semidefinite program MC_{SDPa} follows

$$OPT_{HW} \geq OPT_{SDP} \geq OPT_{SDPa} \geq OPT_{DSDPa} = OPT_{HW};$$

hence equality holds throughout. ∎

4.3. Reconstructing the optimal solution of the problem MC_{SDP}. Once we know the optimal solution of the dual problem DMC_{SDPa}, we can reconstruct the optimal solution of MC_{SDP} by tracing the procedure from the previous subsection and using the structural information about the feasible set $\mathcal{G}$. We will first compute the optimal solution of MC_{SDPa} from the optimal solution of MC_{DSDPa} and then will extend it to the optimal solution of MC_{SDP}. Let $U^* = \text{diag}(U_1, \ldots, U_{n-1})$ be the optimal solution of MC_{SDPa} and (y_1, y_2, y_3) the optimal solution for DMC_{SDPa} from Lemma 8. We define matrix $V^* = \text{diag}(V_1, \ldots, V_{n-1})$ with

$$V_i = \begin{bmatrix} -\ell_i - y_1 & -y_3 \\ -y_3 & \ell_i - y_2 \end{bmatrix}, \quad 1 \leq i \leq n-1. \tag{19}$$

From the feasibility of (y_1, y_2, y_3) it follows that $V_i \succeq 0$ and any matrix V_i is in fact the dual matrix to U_i for $1 \leq i \leq n-1$. Since V^* is actually optimal for DMC_{SDPa}, the strong duality property implies $\langle U_i, V_i\rangle = 0$ for $1 \leq i \leq n-1$. Suppose first that $\ell_1 < \ell_{n-1}$ and V_1 and V_{n-1} are the only singular matrices in V^* (hence V_1 and V_{n-1} are rank one matrices). Let

$$U_1 = \begin{bmatrix} p_1 & q_1 \\ q_1 & r_1 \end{bmatrix},\ U_{n-1} = \begin{bmatrix} p_2 & q_2 \\ q_2 & r_2 \end{bmatrix},\ V_1 = \begin{bmatrix} v_1 & z_1 \\ z_1 & w_1 \end{bmatrix}, \text{ and } V_{n-1} = \begin{bmatrix} v_2 & z_2 \\ z_2 & w_2 \end{bmatrix}.$$

Using (19) we see that $v_1 = -\ell_1 - y_1$, $z_1 = -y_3$, $w_1 = \ell_1 - y_2$, etc. From the strong duality property it follows that $U_2, U_3, \ldots, U_{n-2}$ are zero matrices and U_1, U_{n-1} are singular. Since U_1, U_{n-1}, V_1, and V_{n-1} are singular, the following must be true:

$$\begin{aligned} z_1^2 &= v_1w_1, & z_2^2 &= v_2w_2, \\ q_1^2 &= p_1r_1, & q_2^2 &= p_2r_2. \end{aligned}$$

Together with the strong duality property $\langle U_1, V_1\rangle = p_1v_1 + 2q_1z_1 + r_1w_1 = 0$ this implies that

$$\frac{p_1v_1 + r_1w_1}{2} = |q_1z_1| = \sqrt{p_1v_1r_1w_1}.$$

From the arithmetic-geometric inequality it follows that $p_1v_1 = r_1w_1$ and similarly $p_2v_2 = r_2w_2$. Components of U_1 and U_{n-1} must also satisfy linear constraints from MC_{SDPa}: $p_1 + p_2 = b_1$, $r_1 + r_2 = b_2$, and $q_1 + q_2 = b_3$. All these equations uniquely determine the components of U_1 and U_{n-1} as

$$\begin{array}{lllllllll} p_1 & = & \alpha w_1, & q_1 & = & -\alpha z_1, & r_1 & = & \alpha v_1, \\ p_2 & = & \beta w_2, & q_2 & = & -\beta z_2, & r_2 & = & \beta v_2, \end{array} \tag{20}$$

where

$$\begin{array}{lllllll} \alpha & = & \frac{b_2w_2-b_1v_2}{v_1w_2-v_2w_1} & = \frac{-m_1m_2+\sqrt{m_1m_2(n-m_1)(n-m_2)}}{(\ell_{n-1}-\ell_1)n} & = & \frac{\mu_1}{\ell_{n-1}-\ell_1}, \\ \beta & = & \frac{b_1v_1-b_2w_1}{v_1w_2-v_2w_1} & = \frac{m_1m_2+\sqrt{m_1m_2(n-m_1)(n-m_2)}}{(\ell_{n-1}-\ell_1)n} & = & -\frac{\mu_2}{\ell_{n-1}-\ell_1}. \end{array}$$

If we have $\ell_1 < \ell_{n-1}$ and there exists $1 < i < n-1$ such that V_i is a rank one matrix, then the matrix $U^* = \text{diag}(U_1, \ldots, U_{n-1})$, where $U_2, \ldots, U_{n-2}$ are zero matrices and components of U_1 and U_{n-1} are those from (20), is still the (nonunique) optimal solution of MC_{SDPa}. The last case is that $\ell_1 = \ell_{n-1}$. In this case we cannot use U_1 and U_{n-1}, defined with (20), because α and β are not defined. We will find the optimal solution of MC_{SDPa} directly. Let us define U_1 and U_{n-1} with

$$p_1 = p_2 = \tfrac{b_1}{2}, \quad r_1 = r_2 = \tfrac{b_2}{2}, \quad q_1 = q_2 = \tfrac{b_3}{2}, \tag{21}$$

and let U_i be zero matrices for $2 \le i \le n-2$. The matrix $U = \text{Diag}(U_1, \ldots, U_{n-1})$ is feasible for MC_{SDPa} and $\sum_{i=1}^{n-1} \ell_i(r_i - p_i) + \ell_n(f_{33} - f_{11}) = \ell_1(b_2 - b_1) + \ell_n(f_{33} - f_{11}) = -\frac{2m_1m_2\ell_1}{n} + \frac{s(A)m_1m_2}{n^2} = OPT_{HW}$; hence U is optimal for MC_{SDPa}. However, the MCP is trivial if $\ell_1 = \ell_{n-1}$, since in this case the underlying graph is the complete graph K_n. Let us introduce the matrices

$$Z_1 = \begin{bmatrix} p_1 & -\sqrt{2}q_1 & q_1 \\ -\sqrt{2}q_1 & 2r_1 & -\sqrt{2}r_1 \\ q_1 & -\sqrt{2}r_1 & r_1 \end{bmatrix}, \; Z_{n-1} = \begin{bmatrix} p_2 & -\sqrt{2}q_2 & q_2 \\ -\sqrt{2}q_2 & 2r_2 & -\sqrt{2}r_2 \\ q_2 & -\sqrt{2}r_2 & r_2 \end{bmatrix},$$

and $Z_n = F$, where $F \in \mathcal{S}_3^+$ is from Lemma 5, and p_i, r_i, and q_i are either from (20) or from (21).

PROPOSITION 11. *The matrix*

$$Z^* = Z_1 \otimes E_{11} + Z_{n-1} \otimes E_{n-1,n-1} + Z_n \otimes E_{nn} \tag{22}$$

is the optimal solution for (18) *and the matrix*

$$Y^* = (Q \otimes P)\, Z^*\, (Q \otimes P)^T$$

is the optimal solution for MC_{SDP}.

Proof. The structure of Z^* for the case $n = 3$ can be seen in Figure 1. From the construction of Z^*, Theorem 4, and Proposition 10 follows that $\langle T \otimes S, Z^*\rangle = \ell_1(r_1 - p_1) + \ell_{n-1}(r_2 - p_2) + \ell_n(f_{33} - f_{11}) = OPT_{HW} = OPT_{SDP}$; hence Z^* gives the optimal value of (18). Therefore it remains to show that Z^* is feasible for the problem (18). Positive semidefiniteness of Z^* follows from positive semidefiniteness

$$Z^* = \left[\begin{array}{ccc|ccc|ccc} p_1 & 0 & 0 & -\sqrt{2}q_1 & 0 & 0 & q_1 & 0 & 0 \\ 0 & p_2 & 0 & 0 & -\sqrt{2}q_2 & 0 & 0 & q_2 & 0 \\ 0 & 0 & f_{11} & 0 & 0 & f_{12} & 0 & 0 & f_{13} \\ \hline -\sqrt{2}q_1 & 0 & 0 & 2r_1 & 0 & 0 & -\sqrt{2}r_1 & 0 & 0 \\ 0 & -\sqrt{2}q_2 & 0 & 0 & 2r_2 & 0 & 0 & -\sqrt{2}r_2 & 0 \\ 0 & 0 & f_{12} & 0 & 0 & f_{22} & 0 & 0 & f_{23} \\ \hline q_1 & 0 & 0 & -\sqrt{2}r_1 & 0 & 0 & r_1 & 0 & 0 \\ 0 & q_2 & 0 & 0 & -\sqrt{2}r_2 & 0 & 0 & r_2 & 0 \\ 0 & 0 & f_{13} & 0 & 0 & f_{23} & 0 & 0 & f_{33} \end{array}\right]$$

FIG. 1. *Structure of Z^* for $n = 3$.*

of matrices U_1, U_{n-1}, and F. Feasibility for the constraints (10a) and (13a) follows immediately from the feasibility of U_1 and U_{n-1} for the problem MC_{SDPa} and the structure of Z^*.

To check the feasibility for (11a) we need to compute for all $1 \le i \le n$

$$\begin{aligned} \langle U \otimes P(i,:)^T P(i,:),\, Z^* \rangle &= P(i,n)^2(f_{22} + 2\sqrt{2}f_{23} + 2f_{33}) \\ &\quad + (P(i,1)^2 + P(i,n-1)^2)(2r_1 + 2r_2 - 4r_1 - 4r_2 + 2r_1 + 2r_2) \\ &= (f_{22} + 2\sqrt{2}f_{23} + 2f_{33})/n \;+\; 0 \;=\; 1, \end{aligned}$$

so Z^* is feasible for (11a). The last constraint (12a) reduces for $i = 1$ and arbitrary $1 \le j \le n$ to

$$\begin{aligned} \langle \tilde{V}_1 \otimes e_n P(j,:),\, Z^* \rangle &= \frac{P(j,n)}{2}\left(-\sqrt{2}f_{12} - 2f_{13} + \sqrt{2}f_{23} + 2f_{33}\right) \\ &= \frac{1}{2\sqrt{n}} 2m_1 \;=\; \frac{m_1}{\sqrt{n}}. \end{aligned}$$

Similarly, we check the feasibility for (12a) for $i = 2, 3$. Once we know that Z^* is optimal for (18), the optimality of Y^* follows from Lemmas 5–7 and the fact that $\langle T \otimes S,\, Z^* \rangle = \frac{1}{2}\langle B \otimes \hat{A},\, Y^* \rangle$. □

A simple implication of Proposition 11 is the following closed form formula for the optimal solution of the semidefinite program MC_{SDP}:

$$\begin{aligned} (23) \qquad Y^* &= (Q \otimes P)\, Z^* \,(Q \otimes P)^T = (QZ_1Q^T) \otimes (P(:,1)P(:,1)^T) \\ &\quad + (QZ_{n-1}Q^T) \otimes (P(:,n-1)P(:,n-1)^T) + \frac{1}{n}(QZ_nQ^T) \otimes J_n. \end{aligned}$$

We can see that for any graph and fixed m, Y^* is completely determined by (y_1, y_2, y_3) from Lemma 8 and hence with the first and the second to last eigenvalues of $\hat{A}$ and corresponding eigenvectors, which are determined by the second and the last eigenvalues of the graph Laplacian (λ_2 and λ_n) and corresponding eigenvectors.

5. A new family of relaxations for the MCP. In the previous section we have seen that relaxing the constraint $Y \in \mathcal{C}^*_{3n}$ in model MC_{CP} to $Y \in \mathcal{S}^+_{3n}$ leads to the lower bound OPT_{HW}. To get a better lower bound it is therefore natural to use a (tractable) set $\mathcal{K}$ with $\mathcal{C}^*_{3n} \subset \mathcal{K} \subset \mathcal{S}^+_{3n}$. Specifically, let $OPT_{\mathcal{K}}$ be defined by

$$OPT_{\mathcal{K}} = \min \frac{1}{2}\langle B^T \otimes \hat{A},\, Y \rangle \text{ such that } Y \in \mathcal{K} \text{ and } Y \text{ satisfies (10)–(13);}$$

then $OPT_{MC} \ge OPT_{\mathcal{K}} \ge OPT_{HW}$. A simple (and tractable) candidate for the set $\mathcal{K}$ is $\mathcal{K}_0 = \mathcal{S}^+_{3n} \cap \mathcal{N}_{3n}$. This is actually the first member in the hierarchy of cones introduced by Parrilo in [15] and used also by de Klerk and Pasechnik in their work about the stability number in [11]. We may also replace it with any other member of this hierarchy, but already the second cone $\mathcal{K}_1$ leads to a very expensive semidefinite program. $OPT_{\mathcal{K}_0}$ is already quite expensive, since each sign constraint contributes one linear equation and one slack variable and we have approximately $9n^2/2$ of them. We get cheaper models if we take for $\mathcal{K}$ the cone

$$\mathcal{K}_0^a = \{X \in \mathcal{S}_{3n}^+,\ \mathcal{Z}(X) = 0, \text{ and } X_{ij}^{12} \geq 0 \text{ for any } (i,j) \text{ with } a_{ij} > 0\},$$

where $\mathcal{Z}(X) = 0$ means that all diagonal entries in all nondiagonal blocks must be zero, which corresponds to componentwise orthogonality of columns of partition matrices. Taking the last cone makes sense, since the matrix $B^T \otimes \hat{A}$ in the model MC_{CP} is nonzero only in those positions of $(1,2)$th and $(2,1)$th blocks, where $a_{ij} > 0$, and the constraint $\mathcal{Z}(X) = 0$ is satisfied by any feasible solution for MC_{CP}. Table 1 shows numerical results, which we obtained by optimizing over the cones $\mathcal{K}_0$ and $\mathcal{K}_0^a$. Table 1 contains computational results on small graphs: $P_6 \times P_4$ is the product of two paths, i.e., a 6×4 grid graph, $K_{6,9}$ is the complete bipartite graph on 15 nodes, and rand(15, 0.5) is a random graph on 15 nodes with edge density 0.5. We partition them in several different ways, given by m in column 2. The vectors m are exactly those for which $m_2/2 \leq m_1 \leq m_2$ and m_3 fixed (later we will see that this is useful when considering the balanced vertex separators of a graph). For all these graphs except the random graph we can determine OPT_{MC} by inspection; see column 3. The last three columns contain the original bound OPT_{HW} from [10] and improvements obtained by optimizing over the cones $\mathcal{K}_0$ and $\mathcal{K}_0^a$.

TABLE 1
MCP and the relaxations on some small graphs.

Graph	$m_1,\ m_2,\ m_3$	OPT_{MC}	OPT_{HW}	$OPT_{\mathcal{K}_0}$	$OPT_{\mathcal{K}_0^a}$
$P_6 \times P_4$	7 14 3	1	-3.36	0.26	0.00
$P_6 \times P_4$	8 13 3	1	-3.32	0.22	0.00
$P_6 \times P_4$	9 12 3	1	-3.31	0.15	0.00
$P_6 \times P_4$	10 11 3	1	-3.29	0.10	0.00
$P_6 \times P_4$	8 14 2	2	-1.88	1.12	0.00
$P_6 \times P_4$	9 13 2	2	-1.84	1.09	0.00
$P_6 \times P_4$	10 12 2	2	-1.81	0.94	0.05
$P_6 \times P_4$	11 11 2	2	-1.80	0.85	0.06
rand(15, 0.5)	5 6 4	7	0.19	6.65	6.07
$K_{6,9}$	4 6 5	4	2.18	4.00	3.61
$K_{6,9}$	5 5 5	5	2.50	4.79	3.99
$K_{6,9}$	4 7 4	8	4.71	8.00	6.87
$K_{6,9}$	5 6 4	9	5.41	8.99	7.75

While the relaxation over $\mathcal{K}_0$ provides a substantial improvement as compared to OPT_{HW}, this bound is also rather expensive: we need to solve a semidefinite program in matrices of order $3n$ with approximately $9n^2/2$ additional constraints. Looking at $OPT_{\mathcal{K}_0^a}$ we see that this relaxation is slightly weaker than $OPT_{\mathcal{K}_0}$ but is less expensive since it includes only approximately $m = |E|$ additional constraints, if E is the edge set of the graph. When it is positive, then it is significantly better than OPT_{HW}. We note that $OPT_{\mathcal{K}_0}$ rounded up gives the exact value OPT_{MC} in almost all cases.

To explore the potential of our approach, we also generated some bigger instances with up to 100 nodes. We generated random graphs with edge probability p (g50.1, g50.4, g100) and also random graphs where the entries in the (1,2) block of size (m_1, m_2) are chosen with probability $q < p$. This should result in "easier" instances, as the partition $S_1 = \{1, \ldots, m_1\}$, $S_2 = \{m_1+1, \ldots, m_1+m_2\}$ has a smaller expected

Table 2
Some random graphs on n nodes.

Graph	n	$\|E\|$	p	q	m_1	m_2
g50.1	50	247	0.2			
g50.2	50	237	0.2	0.15	20	20
g50.3	50	198	0.2	0.10	25	20
g50.4	50	114	0.1			
g100	100	1000	0.2			

Table 3
Min-Cut approximation for the graphs from Table 2.

Graph	$m_1,\ m_2,\ m_3$	ubd	OPT_{HW}	$OPT_{\mathcal{K}_0^a}$	$m_3 + \lceil\sqrt{2\alpha}\rceil - 1$
g50.1	20 25 5	49	16.4	41.8	14
	20 20 10	29	-4.3	22.9	16
	15 20 15	14	-20.9	8.3	19
g50.2	20 25 5	45	17.2	36.7	13
	20 20 10	26	-4.9	19.1	16
	15 20 15	12	-22.6	5.8	18
g50.3	20 25 5	29	-1.1	24.1	12
	20 20 10	15	-17.5	10.1	14
	15 20 15	12	-22.6	5.8	16
g50.4	20 25 5	43	7.2	35.1	8
g100	40 50 10	251	161.7	225.1	31
	40 40 20	173	80.2	147.4	37
	35 35 30	107	12.5	84.7	43

number of edges (g50.2, g50.3). In Table 2 we provide some specifics about these graphs.

For these graphs the computation of the relaxation over $\mathcal{K}_0$ is beyond the possibilities of our computing facilities. The simpler relaxation over $\mathcal{K}_0^a$ can still be calculated rather easily. In Table 3 we summarize the results. The column labeled ubd gives an upper bound on OPT_{MC}, obtained by a simulated annealing heuristic. Then we compare $OPT_{\mathcal{K}_0^a}$ to the spectral bound OPT_{HW}. We partition the graphs in several ways, indicated by the vector m. We note also here that the new relaxation provides a substantial improvement over the original spectral bound, which in case of negative values does not give any relevant information at all. Further, more detailed, computational experiments will be reported elsewhere; see the dissertation [16].

6. Advances to the bandwidth and the vertex separator problem. For a graph G on n vertices we define a labeling of vertices as a bijection $\Phi\colon V = \{v_1,\dots,v_n\} \to \{1,2,\dots,n\}$. The labeling bandwidth $\sigma_\infty(G,\Phi)$ of the labeling Φ is the maximal difference over all graph edges:

$$\sigma_\infty(G,\Phi) := \max_{(i,j)\in E} |\Phi(v_i) - \Phi(v_j)|.$$

The bandwidth of a graph G is the minimum of the labeling bandwidth over all labelings:

$$\sigma_\infty(G) := \min_{\Phi} \sigma_\infty(G,\Phi).$$

The bandwidth problem is an NP-hard problem and remains NP-hard even if the graph G is a tree with maximal degree at most 3 or a caterpillar with hairlength ≤ 3. Even approximating the bandwidth is an extremely difficult task. Blache et al. have shown that there is no polynomial time algorithm with an approximation ratio smaller than 1.5 unless P = NP (for more results about the bandwidth problem and its complexity see [3, 4, 5, 12]). In [10] several lower bounds for σ_∞ have been established for an unweighted graph, using Laplacian eigenvalues of the graph. The basic tool the authors used was showing that $OPT_{MC} > 0$. If this is the case, then $\sigma_\infty(G) \geq m_3+1$. This is generalized in the following proposition.

PROPOSITION 12. *Let G be an undirected and unweighted graph. If for some $m = (m_1, m_2, m_3)$ it holds that $OPT_{MC} \geq \alpha > 0$, then*

$$\sigma_\infty(G) \geq \max\{m_3 + 1,\, m_3 + \lceil\sqrt{2\alpha}\rceil - 1\}.$$

Proof. Let Φ be the optimal labeling of G. We may assume that the vertices of G are initially labeled such that Φ is identity, i.e., $\Phi(i) = i$. Let (S_1, S_2, S_3) be a partition of $V(G)$, defined by $S_1 = \{1, \dots, m_1\}$ and $S_2 = \{m_1 + m_3 + 1, \dots, n\}$, Δ the maximal difference of end numbers over all edges, connecting sets S_1 and S_2, and $\delta = \Delta - m_3$. We have $\delta \geq 1$ since $OPT_{MC} > 0$. The only vertices from S_1 that might have a neighbor in S_2 are $m_1 - \delta + 1, \dots, m_1$, since otherwise the difference of end vertices is greater than Δ. The same argument implies that the vertex m_1 has δ neighbors at most in S_2 , the vertex $m_1 - 1$ has $\delta - 1$ neighbors at most, etc. The last vertex $m_1 - \delta + 1$ has 1 neighbor at most in S_2. The number of edges between S_1 and S_2 is therefore $\delta + (\delta - 1) + \cdots + 1 = \delta(\delta + 1)/2$ at most; hence we get the inequality $\delta(\delta + 1) \geq 2\alpha$, which implies $\delta \geq \lceil\sqrt{2\alpha}\rceil - 1$. Since we also know that $\delta \geq 1$, the proposition follows from $\sigma_\infty(G) \geq \Delta$. ∎

Table 4 demonstrates the tightness of this lower bound on the graph instances from Table 1. The third column contains the bandwidth of the graph (for graphs $P_m \times P_n$ and $K_{m,n}$ we can compute it using the closed form formula, e.g., $\sigma_\infty(P_m \times P_n)$ $= \min\{m, n\}$). In the fourth column we have α, the lower bound for OPT_{MC}, obtained by rounding up the best $OPT_{\mathcal{K}_0}$ from Table 1, and the last column shows the lower bound for $\sigma_\infty(G)$ from Proposition 12. We can see that we might get good information about the bandwidth using a good lower bound for the OPT_{MC}, and this is very important according to the complexity hardness of the bandwidth problem.

TABLE 4
Lower bounds for bandwidth for the graphs from Table 1.

Graph	m_3	$\sigma_\infty(G)$	α	$m_3 + \lceil\sqrt{2\alpha}\rceil - 1$
$P_6 \times P_4$	3	4	1	4
$P_6 \times P_4$	2	4	2	3
rand(15, 0.5)	4	10	7	7
$K_{6,9}$	5	10	5	8
$K_{6,9}$	4	10	9	8

Finally, we also compare the lower bound on the bandwidth from [10], which is based on the spectral bound OPT_{HW}, and the bound from Proposition 12 on the graphs from Table 2. The results in Table 5 show again that the new model provides a significant improvement over the spectral bound.

A set $S_3 \subset V$ is a vertex separator if removing these vertices disconnects the graph. It is a balanced vertex separator if the resulting graph has two components of sizes between $s/3$ and $2s/3$, where $s = |V| - |S_3|$. Helmberg et al. have derived in [10] several lower bounds on the size of a minimal vertex separator. They have used the fact that if $OPT_{MC} = 0$ then $OPT_{HW} \leq 0$ and from this have derived lower bounds on the size of the vertex separator. By using the fact that for fixed m_3 is OPT_{HW}

TABLE 5
Lower bounds on the bandwidth for the graphs from Table 2.

Graph	Bound from [10]	Proposition 12
g50.1	9	19
g50.2	9	18
g50.3	5	16
g50.4	5	8
g100	33	43

maximal if m_1 and m_2 are equal (or differs by 1 if $n - m_3$ is an odd number) they have extended the result to balanced vertex separators. The optimal values $OPT_{\mathcal{K}}$ for $\mathcal{K}$ as above give information about the vertex separators only if they are positive, since in this case we know that the graph does not have a vertex separator of size m_3 whose removal divides the graph vertices into sets of sizes m_1 and m_2. Table 1 shows that on the test instances we always detected the nonexistence of the appropriate vertex separator. However, since in general the value $OPT_{\mathcal{K}}$ does not monotonically change with the difference $|m_1 - m_2|$ as is the case for OPT_{HW}, we can get the information about the balanced vertex separator only by checking all possible pairs $m_2/2 \leq m_1 \leq m_2$ with $m_1 + m_2 = n - m_3$. This might be time consuming so it is worth trying to change the model MC_{CP} in order to include the balanced cardinality constraint and then relaxing this model. We have already done some promising steps and the results appear in [16].

7. Conclusions. We have shown that the MCP can be formulated as a linear program over the cone of completely positive matrices of order $3n \times 3n$. Replacing the cone of completely positive matrices with any cone for which we are able to solve the separation problem gives a tractable approximate model. We have analyzed the relaxation, obtained by using the cone of positive semidefinite matrices, and we showed that this model gives the eigenvalue lower bound, originally found by Helmberg et al. in [10]. We provided the closed form solution of this relaxation and showed that it is determined with the second and the largest eigenvalues of graph Laplacian and corresponding eigenvectors. We also proposed some other relaxations, using the hierarchy of cones, proposed by Parrilo in [15]. Numerical results in section 5 show that the lower bounds, obtained this way, may be very tight. At this point we want to emphasize that our approach may be easily extended to a general graph partitioning problem. We finished with the study of the impact that the new results have on approximation of some other combinatorial problems. A reasonable good lower bound for the bandwidth problem may be obtained this way as well as the certificate that the graph does not have a separator of specified size, whose removal disconnects the graph into two sets of prescribed sizes. A preliminary study in modeling the balanced vertex separator problem by copositive programming has also been done and the results together with an extension to the general graph partitioning problem will be reported elsewhere; see also the dissertation [16].

Acknowledgments. We thank two anonymous referees for several suggestions to improve an earlier version of the paper.

REFERENCES

[1] C. J. ALPERT AND A. B. KAHNG, *Recent directions in netlist partition: A survey*, Integration, the VLSI Journal, 19 (1995), pp. 1–81.

[2] K. Anstreicher and H. Wolkowicz, *On Lagrangian relaxation of quadratic matrix constraints*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 41–55.

[3] G. Blache, M. Karpinski, and W. Jürgen, *On approximation intractability of the bandwidth problem*, in Electronic Colloquium on Computation Complexity, University of Trier, Trier, Germany, 1998, TR98-014.

[4] F. R. K. Chung, *Labelings of graphs*, in Selected Topics in Graph Theory 3, Academic Press, San Diego, 1988, pp. 151–168.

[5] J. Díaz, J. Petit, and M. Serna, *A survey on graph layout problems*, ACM Comput. Surveys, 34 (2002), pp. 313–356.

[6] U. Feige, M. Hajiaghayi, and J. R. Lee, *Improved approximation algorithms for minimum-weight vertex separators*, in Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC), Baltimore, MD, 2005, pp. 563–572.

[7] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, 1979.

[8] S. Guattery and G. L. Miller, *On the quality of spectral separators*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 701–719.

[9] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J., 20 (1953), pp. 37–39.

[10] C. Helmberg, F. Rendl, B. Mohar, and S. Poljak, *A spectral approach to bandwidth and separator problems in graphs*, Linear and Multilinear Algebra, 39 (1995), pp. 73–90.

[11] E. de Klerk and D. V. Pasechnik, *Approximation of the stability number of a graph via copositive programming*, SIAM J. Optim., 12 (2002), pp. 875–892.

[12] Y.-L. Lai and K. Williams, *A survey of solved problems and applications on bandwidth, edgesum and profiles of graphs*, J. Graph Theory, 31 (1999), pp. 75–94.

[13] T. Lengauer, *Combinatorial Algorithms for Integrated Circuit Layout*, Wiley, Chichester, UK, 1990.

[14] K. G. Murty and S. N. Kabadi, *Some NP-complete problems in quadratic and nonlinear programming*, Math. Programming, 39 (1987), pp. 117–129.

[15] P. A. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2000.

[16] J. Povh, *Application of Semidefinite and Copositive Programming in Combinatorial Optimization*, Ph.D. thesis, University of Ljubljana, Ljubljana, Slovenia, 2006.

[17] F. Rendl and H. Wolkowicz, *A projection technique for partitioning the nodes of a graph*, Ann. Oper. Res., 58 (1995), pp. 155–179.

[18] C. de Sousa and E. Balas, *The vertex separator problem, algorithms and computations*, Math. Program., 103 (2005), pp. 609–631.

[19] H. Wolkowicz and Q. Zhao, *Semidefinite programming relaxations for the graph partitioning problem*, Discrete Appl. Math., 96–97 (1999), pp. 461–479.