Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems (P. M. Pardalos, Editor), pp. 1-2

Block Exchange in Graph Partitioning

William W. Hager (hager@math.ufl.edu)
Department of Mathematics
University of Florida, Gainesville, FL 32611

Soon Chul Park (scp@math.ufl.edu)
Department of Mathematics
University of Florida, Gainesville, FL 32611

Timothy A. Davis (davis@cise.ufl.edu)
Department of Computer and Information Science and Engineering
University of Florida, Gainesville, FL 32611

Abstract

In a seminal paper (*An efficient heuristic procedure for partitioning graphs,* Bell System Technical Journal, **49** (1970), pp. 291–307), Kernighan and Lin propose a pair exchange algorithm for approximating the solution to min-cut graph partitioning problems. In their algorithm, a vertex from one set in the current partition is exchanged with a vertex in the other set to reduce the sum of the weights of cut edges. The exchanges continue until the total weight of the cut edges is no longer reduced. In this paper, we consider a block exchange algorithm in which a group of vertices from one set is exchanged with a group of vertices from the other set in order to minimize the sum of the weights of cut edges. An optimal choice for the exchanged vertices is the solution to a quadratic programming problem.

Keywords: Graph partitioning, min-cut, quadratic programming.

1 Introduction

In min-cut graph partitioning problems, we partition the vertices of a graph into disjoint sets satisfying specified size constraints, while minimizing the sum of the weights of (cut) edges connecting vertices in different sets. In their seminal paper [1], Kernighan and Lin propose an exchange algorithm for approximating the best partition. This algorithm determines a pair of vertices, one from each set, whose exchange decreases the weights of the edges connecting the sets as much as possible. Eventually, the algorithm achieves a partitioning of the vertices for which any exchange of a pair of vertices either increases or leaves unchanged the sum of the weights of the cut edges. Although this partition could be a locally optimal, it may not be globally optimal since the exchange of a collection of vertices could reduce the sum of the weights of the cut edges. In this paper, we show that the optimal set of vertices to exchange can be obtained from the solution to a quadratic programming problem, while in [2] we show that the solution to the graph partitioning problem is itself the solution to a related quadratic program. Iterative algorithms applied to this NP-hard program often converge to local minimizers that are not global minimizers. An approximate solution to the quadratic program associated with the block exchange problem yields a (nonlocal) change which can be used as a starting point in an algorithm to solve the quadratic programming formulation of the graph partitioning problem itself.

2 Quadratic programming formulation

Let G be a graph with n vertices V:

$$V = \{1, 2, \cdots, n\},$$

and let a_{ij} be a weight associated with the edge (i, j). For each i and j, we assume that $a_{ii} = 0$, $a_{ij} = a_{ji}$, and if there is no edge between i and j, then $a_{ij} = 0$. The sign of the weights is not restricted. Given a positive integer $m < n$, we wish to partition the vertices into two disjoints sets, one with m vertices and the other with $n - m$ vertices, while minimizing the sum of the weights associated with edges connecting vertices in different sets. This optimal partition is called a min-cut.

In [2] we show that for an appropriate choice of the diagonal matrix $\mathbf{D}$, the min-cut can be obtained by solving the following quadratic programming problem:

$$\begin{array}{rl} \text{minimize} & f(\mathbf{x}) := (\mathbf{1} - \mathbf{x})^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{x} \\ \text{subject to} & \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \quad \mathbf{1}^{\mathsf{T}}\mathbf{x} = m. \end{array} \tag{1}$$

More precisely, for an appropriate choice of $\mathbf{D}$, (1) has a solution $\mathbf{x}^*$ for which each component is either zero or one. The two sets V_1 and V_2 in an optimal partition are given by

$$V_1 = \{i : x_i^* = 1\} \quad \text{and} \quad V_2 = \{i : x_i^* = 0\}. \tag{2}$$

When $\mathbf{x}$ is a 0/1 vector, the cost function f in (1) reduces to the sum of those elements a_{ij} of $\mathbf{A}$ corresponding to rows i where x_i vanishes and columns j where x_j is one. Hence, when $\mathbf{x}$ is a 0/1 vector, $f(\mathbf{x})$ is the sum of the weights of edges connecting the sets V_1 and V_2 in (2). The following theorem from [2] shows how to choose $\mathbf{D}$.

Theorem 2.1 [2]. *If $\mathbf{D}$ is chosen so that*

$$d_{ii} + d_{jj} \geq 2a_{ij}$$

for each i and j, then (1) has a 0/1 solution $\mathbf{x}^*$ and the partition given by (2) is a min-cut. Moreover, if for each i and j,

$$d_{ii} + d_{jj} > 2a_{ij},$$

then every local minimizer of (1) is a 0/1 vector.

In the quadratic program (1), we emphasize that the variable $\mathbf{x}$ is continuous, with components taking values on the interval $[0, 1]$. Theorem 2.1 claims that this continuous quadratic program has a 0/1 solution which yields a min-cut. When continuous solution algorithms, such as the gradient projection method, are applied to (1), the iterates typically converge to an extreme point (either a local minimizer or a saddle point) that may not be the global optimum. In order to escape from this local optimum, we need to make a nonlocal change in $\mathbf{x}$. Exchanging l vertices in one set with l vertices in the other set is equivalent to replacing l ones in the $\mathbf{x}$ vector by zeros and l zeros in the $\mathbf{x}$ vector by ones. This change represents a movement in $\mathbf{x}$ of length $\sqrt{2l}$. Such a nonlocal change could potentially yield a starting point for an iterative method that would descend a deeper valley than that containing the current best approximation to a solution of (1).

Suppose that the vertices of V have been partitioned into two sets, V_1 of size m and V_2 of size $n - m$. We assume that the rows and columns of $\mathbf{A}$ are symmetrically permuted so that vertices in V_1 correspond to the first m rows and columns of $\mathbf{A}$ and the vertices of V_2 correspond to the last $n - m$ rows and columns of $\mathbf{A}$. Let us block partition $\mathbf{A}$ in the following way:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ corresponds to leading $m \times m$ and trailing $(n - m) \times (n - m)$ submatrices of $\mathbf{A}$. Given a positive integer l such that $l \leq \max\{m, n-m\}$, we consider the following quadratic programming problem:

$$\begin{aligned} \text{minimize} \quad & F(\mathbf{y}, \mathbf{z}) := \begin{pmatrix} \mathbf{1} - \mathbf{y} \\ \mathbf{1} - \mathbf{z} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \mathbf{A}_{11} + \mathbf{D}_1 & -\mathbf{A}_{12} \\ -\mathbf{A}_{21} & \mathbf{A}_{22} + \mathbf{D}_2 \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} \\ \text{subject to} \quad & \mathbf{0} \leq \mathbf{y} \leq \mathbf{1}, \;\; \mathbf{1}^{\mathsf{T}}\mathbf{y} = l, \;\; \mathbf{0} \leq \mathbf{z} \leq \mathbf{1}, \;\; \mathbf{1}^{\mathsf{T}}\mathbf{z} = l. \end{aligned} \tag{3}$$

Here $\mathbf{D}_1$ and $\mathbf{D}_2$ are the compatible diagonal blocks of $\mathbf{D}$.

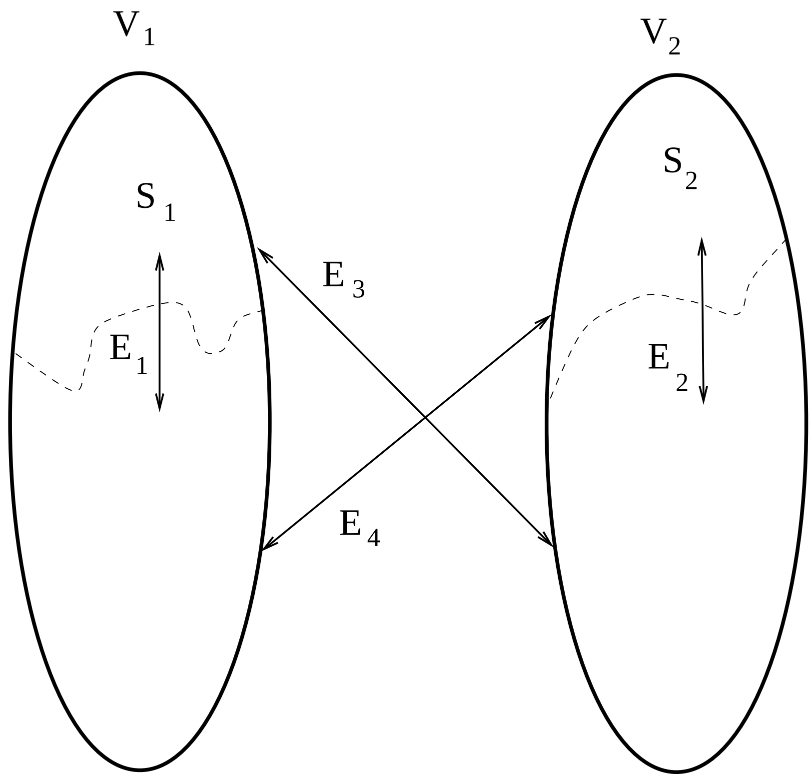

Figure 1: Interchange vertices S_1 in V_1 with S_2 in V_2

Suppose that $\mathbf{y}$ and $\mathbf{z}$ are 0/1 vectors with $\mathbf{1}^\mathsf{T}\mathbf{y} = l$ and $\mathbf{1}^\mathsf{T}\mathbf{z} = l$, and define the following sets:

$$S_1 = \{i : y_i = 1\} \quad \text{and} \quad S_2 = \{i + m : z_i = 1\}.$$

We now observe that $F(\mathbf{y}, \mathbf{z})$ measures the change in the sum of the weights of cut edges corresponding to the exchange of vertices S_1 in V_1 with the vertices S_2 in V_2. In particular, referring to Figure 1, the quantity $(\mathbf{1} - \mathbf{y})^\mathsf{T}\mathbf{A}_{11}\mathbf{y}$ is the sum of the weights of edges E_1 connecting the set S_1 with its complement in V_1; $(\mathbf{1} - \mathbf{z})^\mathsf{T}\mathbf{A}_{22}\mathbf{z}$ is the sum of the weights of edges E_2 connecting the set S_2 with its complement in V_2; $(\mathbf{1} - \mathbf{z})^\mathsf{T}\mathbf{A}_{21}\mathbf{y}$ is the sum of the weights of edges E_3 connecting the set S_1 with the complement of S_2 in V_2; $(\mathbf{1} - \mathbf{y})^\mathsf{T}\mathbf{A}_{12}\mathbf{z}$ is the sum of the weights of edges E_4 connecting the set S_2 with the complement of S_1 in V_1. When the vertices S_1 and S_2 are exchanged, the edges E_3 and E_4 change from external edges connecting V_1 and V_2 to internal edges, while the internal edges E_1 and E_2 change to external edges connecting V_1 and V_2. Hence, the change in the sum of the weights of cut edges is

$$(\mathbf{1} - \mathbf{y})^\mathsf{T}\mathbf{A}_{11}\mathbf{y} + (\mathbf{1} - \mathbf{z})^\mathsf{T}\mathbf{A}_{22}\mathbf{z} - (\mathbf{1} - \mathbf{z})^\mathsf{T}\mathbf{A}_{21}\mathbf{y} - (\mathbf{1} - \mathbf{y})^\mathsf{T}\mathbf{A}_{12}\mathbf{z},$$

the difference between the weights of the newly created external edges and the deleted external edges. This difference is precisely the cost function $F(\mathbf{y}, \mathbf{z})$ of (3) since the diagonal terms $(\mathbf{1} - \mathbf{y})^\mathsf{T}\mathbf{D}_1\mathbf{y}$ and $(\mathbf{1} - \mathbf{z})^\mathsf{T}\mathbf{D}_2\mathbf{z}$ vanish when $\mathbf{y}$ and $\mathbf{z}$ are 0/1 vectors.

If the components of $\mathbf{y}$ and $\mathbf{z}$ are restricted to be integers, then the quadratic program (3) is equivalent to minimizing the increase in the sum of the edge weights associated with the exchange of the sets S_1 and S_2. If the increase is negative, then the exchange of S_1 with S_2 will decrease the sum of the weights of the cuts edges. Note though that in (3), we do not restrict the components of $\mathbf{y}$ and $\mathbf{z}$ to be integers, and potentially, the minimum in this continuous problem is strictly smaller than the minimum in the discrete analogue where the variables are restricted to be 0/1. The following theorem, however, ensures that the continuous problem (3) has a (discrete) 0/1 solution.

Theorem 2.2. *If* $\mathbf{D}$ *is chosen so that*

$$d_{ii} + d_{jj} \geq 2a_{ij} \tag{4}$$

for each i *and* j *in* $[1, m]$ *and for each* i *and* j *in* $[m+1, n]$*, then* (3) *has a* 0/1 *solution* $(\mathbf{y}^*, \mathbf{z}^*)$*. Let us define the sets* $V_1 = \{1, 2, \ldots, m\}$, $V_2 = \{m+1, m+2, \ldots, n\}$,

$$S_1 = \{i : y_i^* = 1\} \quad \textit{and} \quad S_2 = \{i + m : z_i^* = 1\}.$$

Exchanging the vertices S_1 *of* V_1 *with the vertices* S_2 *of* V_2 *leads to the smallest possible increase in the sum of the weights of cut edges among all possible* l *element subsets of* V_1 *and* V_2*. Moreover, if for each* i *and* j *in* $[1, m]$ *and for each* i *and* j *in* $[m+1, n]$*, we have*

$$d_{ii} + d_{jj} > 2a_{ij}, \tag{5}$$

then every local minimizer of (3) *is a* 0/1 *vector.*

Proof. Our proof is basically the same as that given in [2] for Theorem 2.1. Given a solution $(\mathbf{y}, \mathbf{z})$ to (3), we construct a piecewise linear path taking us from $(\mathbf{y}, \mathbf{z})$ to a solution $(\mathbf{y}^*, \mathbf{z}^*)$ of (3) whose components are either 0 or 1. Let $\mathcal{F}(\mathbf{y})$ be the inactive (or free) components of the vector $\mathbf{y}$:

$$\mathcal{F}(\mathbf{y}) = \{i : 0 < y_i < 1\}. \tag{6}$$

Either $\mathcal{F}(\mathbf{y})$ is empty, and $\mathbf{y}^* = \mathbf{y}$, or $\mathcal{F}(\mathbf{y})$ has two or more elements since the constraint $\mathbf{1}^\mathsf{T}\mathbf{y} = l$ of (3), where l is integer, cannot be satisfied when $\mathbf{y}$ has a single noninteger component. If $\mathcal{F}(\mathbf{y})$ has two or more elements, we show that there exists another minimizing point $\bar{\mathbf{y}}$ with $\mathcal{F}(\bar{\mathbf{y}})$ strictly contained in $\mathcal{F}(\mathbf{y})$, and $F(\mathbf{x}, \mathbf{z}) = F(\mathbf{y}, \mathbf{z})$ for all $\mathbf{x}$ on the line segment connecting $\mathbf{y}$ and $\bar{\mathbf{y}}$. Utilizing this property in an inductive fashion, we conclude that there exists a piecewise linear path taking us from any given minimizer $(\mathbf{y}, \mathbf{z})$ to another minimizer $(\mathbf{y}^*, \mathbf{z})$ with $\mathcal{F}(\mathbf{y}^*) = \emptyset$ (that is, all the components of $\mathbf{y}^*$ are either 0 or 1), and $F(\mathbf{x}, \mathbf{z}) = F(\mathbf{y}, \mathbf{z})$ for all $\mathbf{x}$ on this path. The same argument applied to the $\mathbf{z}$ component of the solution $(\mathbf{y}^*, \mathbf{z})$ shows that there exists a $\mathbf{z}^*$ that is feasible in (3) with $\mathcal{F}(\mathbf{z}^*) = \emptyset$ and $(\mathbf{y}^*, \mathbf{z}^*)$ optimal in (3).

If $\mathcal{F}(\mathbf{y})$ has two or more elements, then choose two elements i and $j \in \mathcal{F}(\mathbf{y})$, and let $\mathbf{v} \in \mathbf{R}^m$ be the vector whose entries are all zero except that $v_i = 1$ and $v_j = -1$. For ϵ sufficiently small, $(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z})$ is feasible in (3). Expanding F in a Taylor series around $(\mathbf{y}, \mathbf{z})$, we have

$$F(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z}) = F(\mathbf{y}, \mathbf{z}) - \epsilon^2 \mathbf{v}^{\mathsf{T}}(\mathbf{A}_{11} + \mathbf{D}_1)\mathbf{v}. \tag{7}$$

The $O(\epsilon)$ term in this expansion disappears since $F(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z})$ achieves a minimum at $\epsilon = 0$, and the first derivative with respect to ϵ vanishes at $\epsilon = 0$. In addition, from the inequality

$$F(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z}) \geq F(\mathbf{y}, \mathbf{z}) \quad \text{for all } \epsilon \text{ near } 0,$$

we conclude that the quadratic term in (7) is nonnegative, or equivalently,

$$\mathbf{v}^{\mathsf{T}}(\mathbf{A}_{11} + \mathbf{D}_1)\mathbf{v} \;=\; d_{ii}v_i^2 + d_{jj}v_j^2 + 2a_{ij}v_iv_j \;=\; d_{ii} + d_{jj} - 2a_{ij} \;\leq\; 0. \tag{8}$$

Since $d_{ii} + d_{jj} - 2a_{ij} \geq 0$ by (4), it follows that $d_{ii} + d_{jj} - 2a_{ij} = 0$ and $F(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z}) = F(\mathbf{y}, \mathbf{z})$ for each choice of ϵ. Let $\bar{\epsilon}$ be the largest value of ϵ for which $(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z})$ is feasible in (3). Defining $\bar{\mathbf{y}} = \mathbf{y} + \bar{\epsilon}\mathbf{v}$, $\mathcal{F}(\bar{\mathbf{y}})$ is strictly contained in $\mathcal{F}(\mathbf{y})$ and $(\bar{\mathbf{y}}, \mathbf{z})$ achieves the minimum in (3) since $F(\mathbf{y} + \epsilon\mathbf{v}, \mathbf{z}) = F(\mathbf{y}, \mathbf{z})$ for all ϵ. In summary, for any given solution $(\mathbf{y}, \mathbf{z})$ to (3), we can find a point $\bar{\mathbf{y}}$ with $\mathcal{F}(\bar{\mathbf{y}})$ strictly contained in $\mathcal{F}(\mathbf{y})$ and $F(\mathbf{x}, \mathbf{z}) = F(\mathbf{y}, \mathbf{z})$ for all $\mathbf{x}$ on the line segment connecting $\mathbf{y}$ and $\bar{\mathbf{y}}$. Proceeding by induction, there exists a solution $(\mathbf{y}^*, \mathbf{z})$ of (3) where $\mathbf{y}^*$ is 0/1. The same argument applied to $\mathbf{z}$ shows that there exists a solution $(\mathbf{y}^*, \mathbf{z}^*)$ of (3) where $\mathbf{z}^*$ is 0/1.

Finally, suppose that (5) holds, that $(\mathbf{y}, \mathbf{z})$ is a local minimizer for (3), and $\mathbf{y}$ is not a 0/1 vector. As noted above, $\mathcal{F}(\mathbf{y})$ has two or more elements, and the expansion (7) holds where the quadratic term satisfies (8), contradicting (5). We conclude that $\mathcal{F}(\mathbf{y})$ is empty and $\mathbf{y}$ is a 0/1 vector. By the same argument, $\mathbf{z}$ is 0/1 as well. □

3 Numerical illustrations

As an application , let us consider the case where the edge weights are all one, $m = n/2$ (the bisection problem), and $\mathbf{D} = \mathbf{I}$, the identity matrix. Our first example is the matrix msc01050 in the Boeing test problem library found on Tim Davis' web page at

www.cise.ufl.edu/˜davis/sparse/Boeing.

This matrix is 1050×1050 with 29,156 nonzero entries. The pattern of the nonzero entries appears in Figure 2. If the Kernighan/Lin exchange algorithm is applied, starting from the initial partition

$$V_1 = \{1, 2, \ldots, 525\} \quad \text{and} \quad V_2 = \{526, 527, \ldots, 1050\},$$

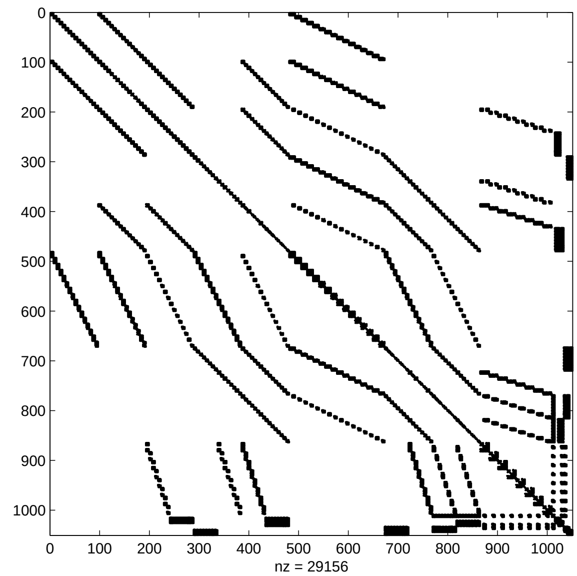

Figure 2: Sparsity pattern for the test problem Boeing/msc01050, 29,156 nonzeros.

then it converges to a partition with 2202 cut edges.

On the other hand, when the gradient projection algorithm was applied to the quadratic program (1), we converged to a local minimizer with 1493 cut edges. Now consider the following choices for l in (3): l = 262, 183, 128, 89, 62, 43, 30, 21, 14, 9, 6, 4, 2. These values for l are obtained by initializing $l = \text{floor}(m/4)$ and then successively multiplying l by .7. For each choice of l, we approximate the solution to (3) using the gradient projection algorithm. After making the interchange, we treat the resulting point as the intial guess in the gradient projection algorithm applied to (1). We only retain the resulting local minimizer if it yields fewer cut edges. The number of cut edges after the exchange (swap) of the sets of size l and after the subsequent gradient projection (GP) steps appears in Figure 3. Hence, by exchanging blocks of size 183 and later of size 4, the number of cut edges is reduced from 1493 down to 1455. If $\mathbf{A}$ is permuted so that the first 525 columns correspond to the vertices in one set in the best computed partition, then the sparsity pattern of the resulting matrix appears in Figure 4. In this figure, the 1455 cut edges appear in the lower left corner. For comparison, the pmetis code [4] of Karypis and Kumar generates a partitioning with 1491 cut edges with a slightly unbalanced partitioning (524/526). The Chaco package of Hendrickson and Rothberg gives the following number of cut edges for the various implemented algorithms: 1578 (multi), 1565 (spectral), 1574 (linear), 1574 (random), and 1544 (scattered).

For another example, we consider the less structured test problem G38 in Ye's test problem collection found at

ftp://dollar.biz.uiowa.edu/pub/yyye/Gset/.

l	Cut edges after swap	Cut edges after GP
262	1562	1562
183	1458	1456
128	1512	1511
89	1520	1503
62	1493	1461
43	1481	1457
30	1467	1461

l	Cut edges after swap	Cut edges after GP
21	1465	1458
14	1459	1458
9	1464	1457
6	1457	1457
4	1455	1455
2	1455	1455

Figure 3: Block exchange for the matrix of Figure 2

The sparsity pattern of this matrix is in Figure 5, while the sparsity pattern of the permuted matrix, associated with the best computed partition (containing 2690 cut edges), obtained by the block exchange approach, appears in Figure 6. During the computation of this partition, blocks of size 500, 350, 244, 170, and 118 were exchanged. In constrast, if pairs of vertices are exchanged rather than blocks of vertices, then starting from the initial partition $V_1 = \{1, \ldots, 1000\}$ and $V_2 = \{1001, \ldots, 2000\}$, the iterates converge to a partition with 3063 cut edges. The number of cut edges for other codes were the following: 2902 (pmetis), 2831 (chaco/multi), 2838 (chaco/-spectral), 2941 (chaco/linear), 2990 (chaco/random), and 2896 (chaco/scattered).

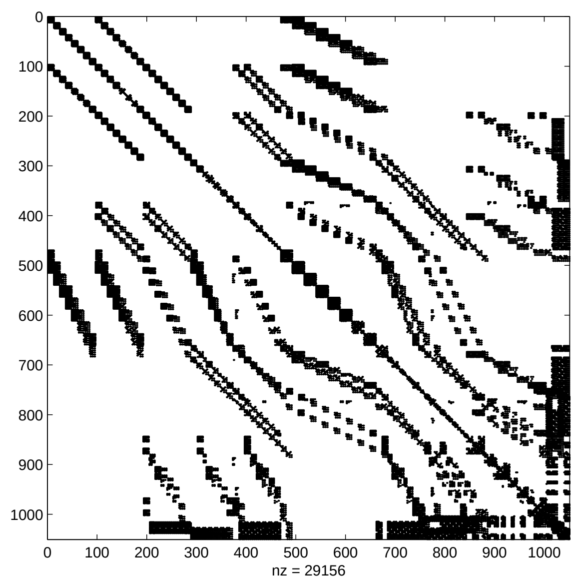

Figure 4: The pattern of the permuted matrix of Figure 2 associated with best computed partition.

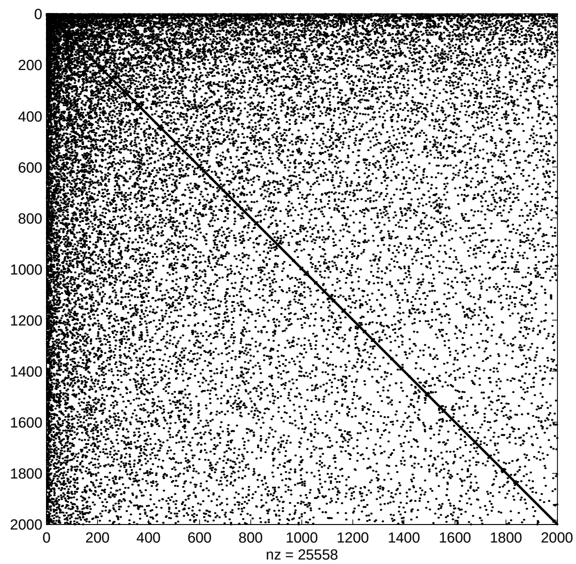

Figure 5: Pattern of the 25,558 nonzero elements for matrix G38 in Ye's test problem set.

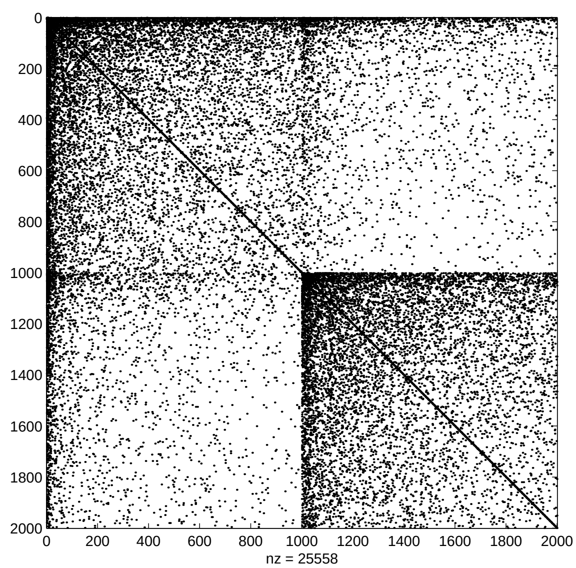

Figure 6: The pattern of the permuted matrix of Figure 5 associated with the best computed partition.

References

[1] B. W. Kernighan and S. Lin (1970), "An efficient heuristic procedure for partitioning graphs," *Bell System Technical Journal*, vol. 49, pp. 291–307.

[2] W. W. Hager and Y. Krylyuk (to appear), "Graph partitioning and continuous quadratic programming," *SIAM J. Discrete Math.*

[3] B. Hendrickson and E. Rothberg (1997), "Improving the runtime and quality of nested dissection ordering," Technical report, Sandia National Laboratories, Albuquerque, NM.

[4] G. Karypis and V. Kumar (1995), "METIS: Unstructured graph partitioning and sparse matrix ordering system," Technical report, Dept. of Computer Science, Univ. of Minnesota.