SDP and Eigenvalue approaches to Bandwidth and Vertex-Separator problems in graphs

Franz Rendl

http://www.math.uni-klu.ac.at

Alpen-Adria-Universität Klagenfurt

Austria

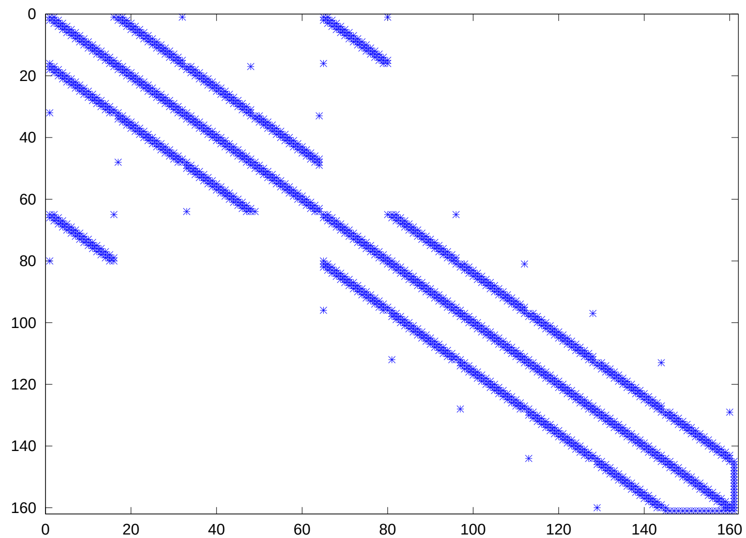
joint work (in progress) with A. Lisser (Paris Orsay) and M. Piacentini (Roma)

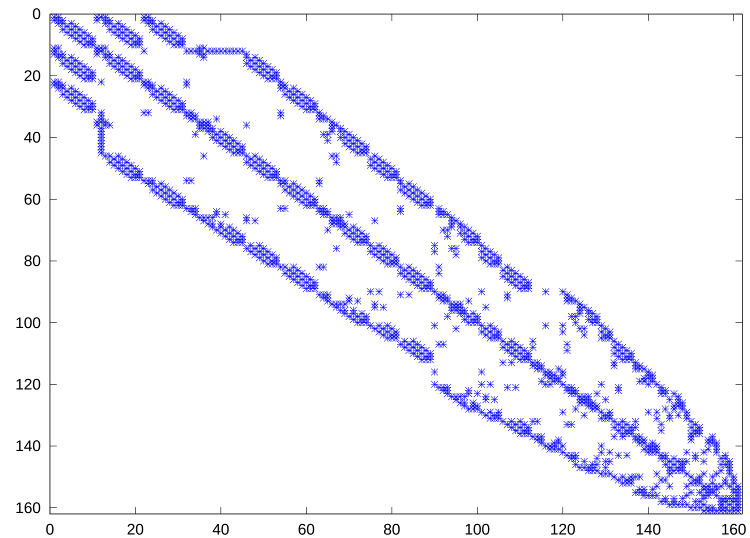
Overview

- Bandwidth and Separators
- Direct modeling (Quadratic Assignment)
- Indirect Model: (Partitioning)
- Weight redistribution
- Preliminary computations

Bandwidth

A graph from the Sparse Matrix Collection (Tim Davis, UFL)
$n = 161$, 1377 nonzeros





The same graph (right) after reordering the vertices. The nonzeros are close to the main diagonal.
Bandwidth left is 79, and right it is 34.

Small Bandwidth

Bandwidth of Matrix: largest distance of nonzero entry from main diagonal

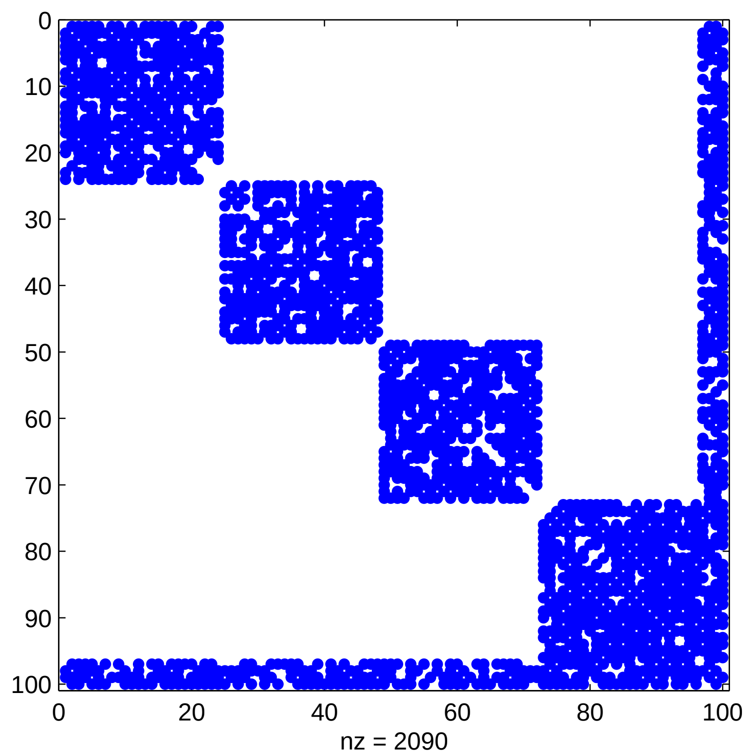
Small bandwidth saves computation time in numerical linear algebra

Typical Example: g3rmt3m3 (from Sparse Matrix Collection): $n = 5357, nz \approx 100,000$

Cholesky decomposition:
Matrix (as is) (dense): 7 sec.
Matrix (as is) (sparse format): 0.5 sec.
Matrix (after reorder) (sparse): 0.07 sec.

Vertex Separators

Given adjacency matrix A of a graph G. Does G have $S_k \subseteq V(G)$ such that $G \setminus S_k$ decomposes into $k-1$ pieces $S_1, \ldots, S_{k-1}$ of (roughly) equal size?



Here $k = 5$, last block separates the first four.

Vertex Separators: How they help

In numerical linear algebra, solve linear system

$$\begin{pmatrix} A_1 & 0 & A_{13} \\ 0 & A_2 & A_{23} \\ A_{13}^T & A_{23}^T & A_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Solve system on each subblock, then get final solution.

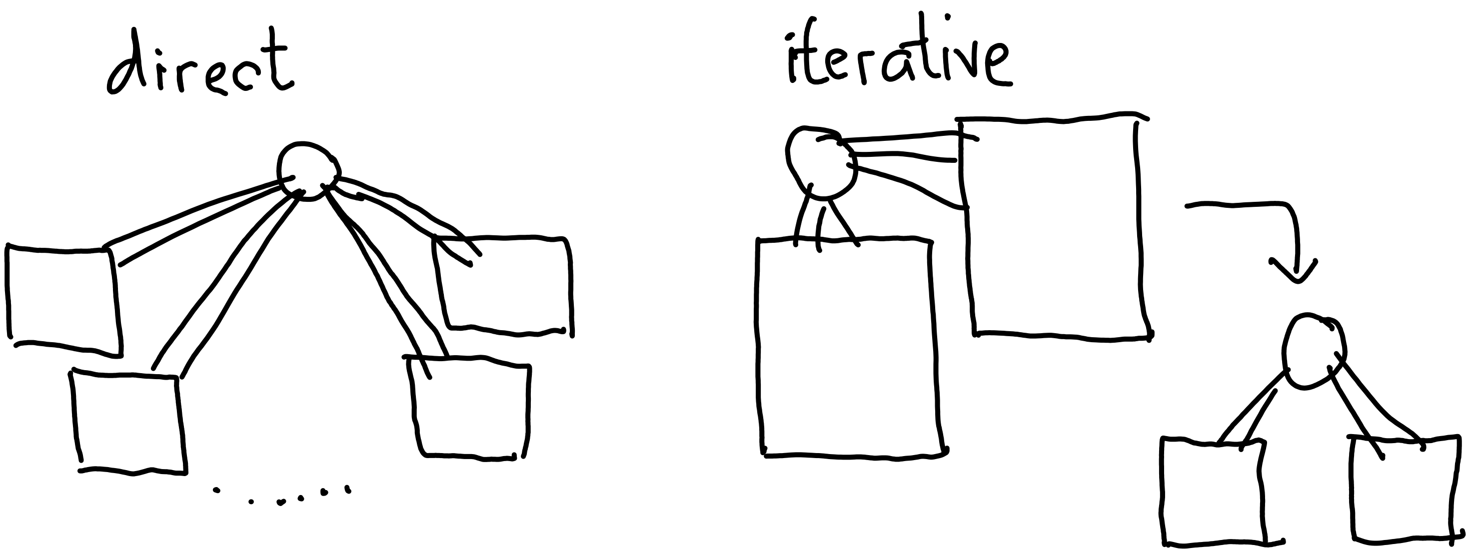
$$x_1 = A_1^{-1} b_1 - A_1^{-1} A_{13} x_3, \;\; x_2 = A_2^{-1} b_2 - A_2^{-1} A_{23} x_3,$$

$$(A_3 - \sum_i A_{i3}^T A_i^{-1} A_{i3}) x_3 = rhs.$$

Instead of n^3, needs roughly $\frac{1}{k^2} n^3$, if separator leaves k pieces (of roughly equal size).

Hierarchical versus direct Partition

It may be better to directly look for 2^k separator blocks instead of k recursive simple bisectors. If separation is bad in initial blocks, this may not be fixed later.



Challenge: How find a good separator to get 2^k blocks directly?

Complexity

- Bandwidth and Vertex Separator are NP-hard
- Bandwidth NP-complete even for trees with maximal degree 3
- Bandwidth approximation by Blum, Konjevod, Ravi, Vempala (2000) using hyperplane rounding
- Approximation for vertex separators by Feige, Hajiaghayi, Lee (2005)
- $O(\sqrt{n})$ vertex separators in planar graphs Lipton, Tarjan (1979)
- Eigenvalue model by Helmberg, Mohar, Poljak, R. (1995)
- semidefinite and copositive model by Povh, R. (2007)

Matrix Reorderings

There is a variety of matrix reordering problems.

- Minimize Bandwidth
- Minimize Cholesky fill-in
- Minimize 1-sum or 2-sum of reordering
- Reorder into k blocks with small interblock connectivity
- Find small separator leaving two roughly equal parts

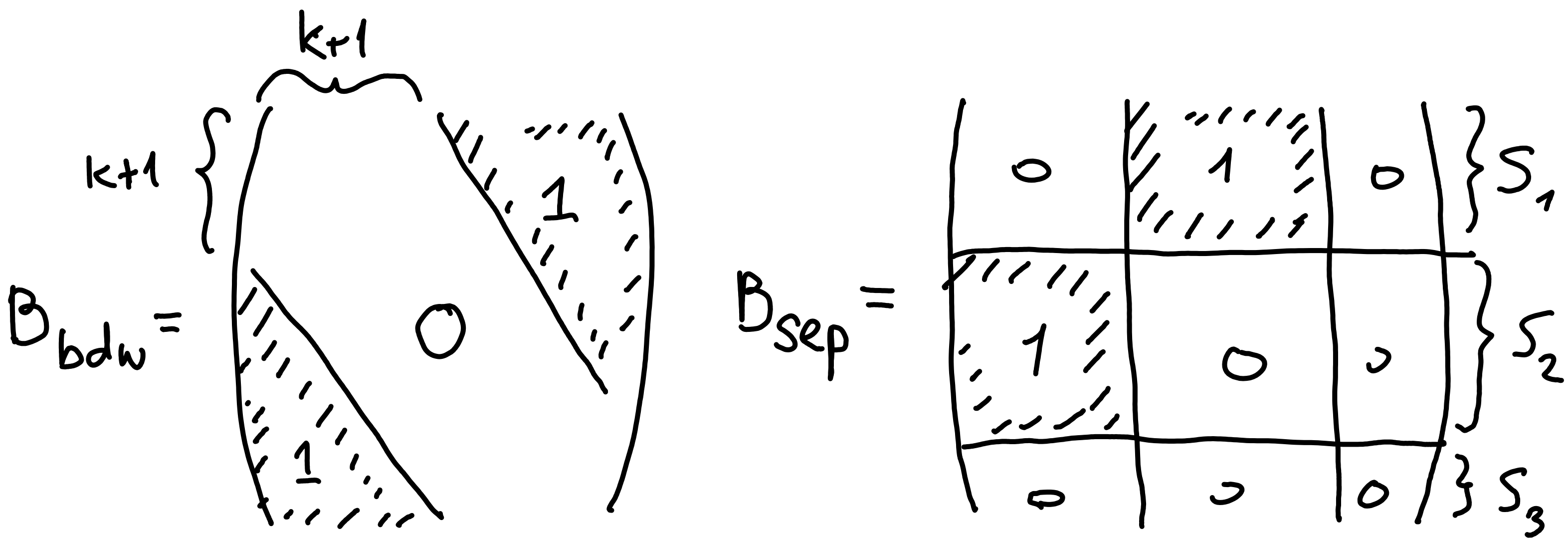
Matlab has several (fast) heuristics to get good reorderings: symrcm (Symmetric reverse Cuthill-McKee permutation), see also symamd, colamd, colperm.

Overview

- Bandwidth and Separators
- Direct modeling (Quadratic Assignment)
- Indirect Model: (Partitioning)
- Weight redistribution
- Preliminary computations

Quadratic Assignment Model

A adjacency matrix of graph, the matrix B is one of the following



Permute A, so that it fits into the zero-pattern of B_{sep} or B_{bdw}. This is Quadratic Assignment Problem.

Quadratic Assignment Model (2)

QAP model:

minimize tr$(X^T AX)B$ over permutation matrices X.

There exists a separator of the required size, or a reordering having the required bandwidth $\iff$ opt. value=0.

If minimal value > 0, then we have information on how many entries of the matrix (=edges of the graph) would have to be fixed.

Could use weighted version of B_{bdw} to force permuted matrix closer to main diagonal.

Instead of A, we use Laplacian L of A: $L = diag(Ae) - A$, because diag$(B) = 0$, so tr$X^T AXB =$ tr$X^T(-L)XB$.

Quadratic Assignment relaxations

There are several ways to get relaxations:
• $X^TX = I$ (orthogonal matrices), $Xe = X^Te = e$ (constant row and column sums):
This leads to projected eigenvalue bound, and uses the Hoffman-Wielandt theorem

$$\min_{X^TX=I} \text{tr}(AXBX^T) = \sum_i \lambda_i(A)\lambda_{n+1-i}(B),$$

• use Semidefinite relaxations: Work in the space of $n^2 \times n^2$ matrices. Computationally expensive!
DeKlerk, Nagy, Sotirov (2011): Apply symmetry reduction to SDP relaxation of the Quadratic Assignment formulation.

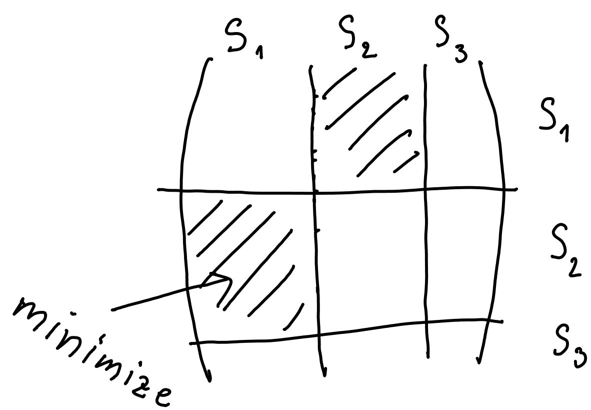
Overview

- Bandwidth and Separators
- Direct modeling (Quadratic Assignment)
- Indirect Model: (Partitioning)
- Weight redistribution
- Preliminary computations

Indirect Approach through Partition

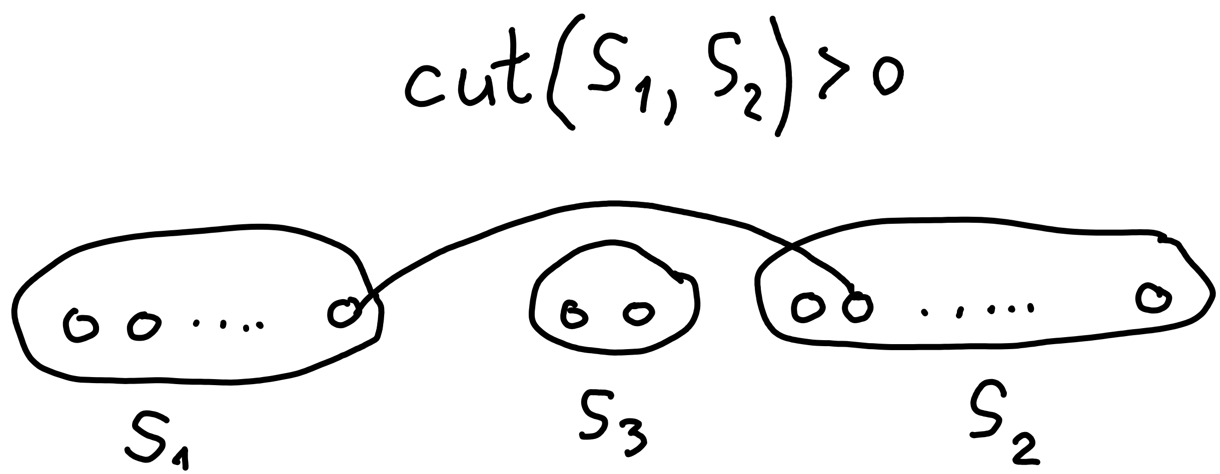
Some notation:
We consider partitions $S = (S_1, S_2, S_3)$ of $N = \{1, \ldots, n\}$ (rows/columns of A) for given cardinalities $m = (m_1, m_2, m_3)$ with $\sum_i m_i = n$ and $|S_i| = m_i$.

Auxiliary Problem: For given m, find partition S such that $\delta(S_1, S_2)$, the number of edges between S_1 and S_2 is minimized.
Denote the minimal value over all $S = (S_1, S_2, S_3)$ by $cut(m)$.



Bounding Bandwidth

- If $cut(m) > 0$ then bandwidth is at least $m_3 + 1$, because removal of any set of m_3 vertices leaves edges between the remaining two parts.



In any ordering of the vertices, there is an edge between S_1 and S_2, passing over S_3, hence bandwidth$\geq |S_3| + 1$.

Bounding Separators

Simple observation:
• If $cut(m) = 0$ then there exists separator with cardinalities given in m
• If $cut(m) > 0$ then no such separator exists.

Idea: We replace $cut(m)$ by lower bound, and check whether it is > 0.
In this case we know that $cut(m) > 0$.

This gives lower bound on size of separator.

If lower bound > 0, then $cut(m) > 0$ and we have bound on bandwidth.

Partition Model

We model the k-partitions as $n \times k$ 0-1 matrices $X = (x_1, \ldots, x_k)$ such that $Xe = e, X^T e = m$, where $m^T = (m_1, \ldots, m_k)$ and $\sum_i m_i = n$.

$$P_m := \{X : Xe = e,\ X^T e = m,\ x_{ij} \in 0, 1,\ X \ldots n \times k\}.$$

Optimizing over the following superset F_m of P_m is tractable, see Helmberg, Mohar, Poljak, R. (1995).
We set $M = \text{diag}(m)$.

$$F_m := \{X : Xe = e,\ X^T e = m,\ X^T X = M\}.$$

Partition model (2)

Lemma (HMPR (1995)):

$$X \in F_m \iff X = \frac{1}{n} e m^T + V Z W \tilde{M}, \; Z^T Z = I_{k-1}$$

where V orthonormal basis to $e^\perp$, $\tilde{m} = (\sqrt{m_1}, \ldots, \sqrt{m_k})^T$ and W is orthogonal basis to $\tilde{m}^\perp$, $\tilde{M} = \text{diag}(\tilde{m})$.

Easy fact: $X \in P_m \iff X \in F_m, \; X \geq 0$.

Modeling $\delta(S_1, S_2)$

Let $m = (m_1, m_2, m_3)$ and $B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

For $X = (x_1, x_2, x_3) \in P_m$ we get

$$XBX^T = x_1 x_2^T + x_2 x_1^T.$$

Hence, for $X \in P_m$ the matrix XBX^T is the adjacency matrix of the edge set $\delta(S_1, S_2)$ corresponding to partition X, (Matrix B_{sep} from before!)

Cost function

tr$A(XBX^T)$ gives twice the number of edges in $\delta(S_1, S_2)$.

Since diag$(X^TBX) = 0$ and Laplacian L of A has $L = -A$ outside main diagonal, we get tr$AXBX^T =$ tr$(-L)XBX^T$.

Cost function

tr$A(XBX^T)$ gives twice the number of edges in $\delta(S_1, S_2)$.

Since diag$(X^TBX) = 0$ and Laplacian L of A has $L = -A$ outside main diagonal, we get tr$AXBX^T =$ tr$(-L)XBX^T$.

Now substitute $X = \frac{1}{n}em^T + VZW\tilde{M},\ Z^TZ = I_{k-1}$ from lemma to get

$$\text{tr}LXBX^T = \text{tr}(V^TLV)Z(W\tilde{M}B\tilde{M}W^T)Z^T$$

Minimizing over F_m asks for $Z^TZ = I_{k-1}$ in view of previous lemma.

Orthogonal relaxation

$$2cut(m) := \min_{X \in P_m} \mathrm{tr}(-L)XBX^T \geq \min_{X \in F_m} \mathrm{tr}(-L)XBX^T$$

$$= \min_{Z^TZ=I_{k-1}} \mathrm{tr} - (V^TLV)Z(W\tilde{M}B\tilde{M}W^T)Z^T := {\color{red}2f(m)}.$$

Orthogonal relaxation

$$2cut(m) := \min_{X \in P_m} \mathrm{tr}(-L)XBX^T \geq \min_{X \in F_m} \mathrm{tr}(-L)XBX^T$$

$$= \min_{Z^TZ = I_{k-1}} \mathrm{tr} - (V^TLV)Z(W\tilde{M}B\tilde{M}W^T)Z^T := 2f(m).$$

Explicit solution using the Hoffman-Wielandt theorem:

$$\min\{\langle A, XBX^T\rangle : \; X^TX = I_k\} =$$

$$= \min\{\sum_i \lambda_i(B)\lambda_{\phi(i)}(A) : \; \phi : \{1,\ldots,k\} \mapsto \{1,\ldots,n\} \text{ injection}\}$$

(Here A and B are symmetric, of orders n and k with $n \geq k$. The minimizer X is determined through eigenvectors $X = PQ^T$.)

Closed-form solution

The Hoffman-Wielandt theorem yields closed form solutions. We need the eigenvalues of

- $V^T L V$, these are $\lambda_2(L), \ldots, \lambda_n(L)$.
- $W\tilde{M}B\tilde{M}W^T$ as $\mu_1 \leq \ldots \leq \mu_{k-1}$ as functions of m. For any partition m it holds that

$$\mu_1 \leq \ldots \leq \mu_{k-2} < 0 < \mu_{k-1}$$

This gives

$$f(m) = \sum_{i=1}^{k-2} (-\mu_i)\lambda_{i+1}(L) - \mu_{k-1}\lambda_n(L).$$

Closed form in case of equal blocks

Consider partitions with $m = (s, \ldots, s, t)$ where $(k-1)s + t = n$ (k partition blocks, the separator has size t, all other $k-1$ sets have equal size s).

Closed form in case of equal blocks

Consider partitions with $m = (s, \ldots, s, t)$ where $(k-1)s + t = n$
(k partition blocks, the separator has size t, all other $k-1$ sets have equal size s).

Lemma: The eigenvalues of $W\tilde{M}B\tilde{M}W^T$ with $m = (s, \ldots, s, t)$ as above are

$$\mu_1 = \ldots = \mu_{k-2} = -s, \ \mu_{k-1} = \frac{k-2}{n}st.$$

Closed form in case of equal blocks

Consider partitions with $m = (s, \dots, s, t)$ where $(k-1)s + t = n$
(k partition blocks, the separator has size t, all other $k-1$ sets have equal size s).

Lemma: The eigenvalues of $W\tilde{M}B\tilde{M}W^T$ with $m = (s, \dots, s, t)$ as above are

$$\mu_1 = \dots = \mu_{k-2} = -s, \ \mu_{k-1} = \frac{k-2}{n}st.$$

This leads to the following lower bound on $cut(m)$:

$$cut(m) \geq \frac{1}{2}(s \sum_{i=2}^{k-1} \lambda_i(L) - \frac{k-2}{n} st \lambda_n(L)).$$

Improvements ?

How could this bound be improved ?

(a) use SDP formulation of the Hoffman-Wielandt Theorem
initial SDP can be warmstarted using HW Thm.

(b) use weight redistribution
works also for large instances

Overview

- Bandwidth and Separators
- Direct modeling (Quadratic Assignment)
- Indirect Model: (Partitioning)
- Weight redistribution
- Preliminary computations

Eigenvalue bounds

The eigenvalues $\mu_1 \leq \ldots \leq \mu_{k-1}$ of the $(k-1) \times (k-1)$ matrix $W\tilde{M}B\tilde{M}W^T$ can easily computed, sometimes even explicitely, see before.

The (extreme) eigenvalues of L, $\lambda_n(L)$ and $\lambda_2(L)$,. . .,$\lambda_{k-1}(L)$ can be computed by iterative methods, in case A is a reasonably sparse matrix.

The eigenvalue lower bound is in general rather weak, so we need some improvement.

We use the idea of weight redistribution, as employed by Boyd, Diaconis and Xiao (2004) for rapidly mixing Markov chains, and Göring, Helmberg and Wappler (2008) for geometric embedding problems of graphs.

Weight redistribution

The lower bound can be interpreted as a function of the Laplacian L. We consider a whole family of graphs, defined through their Laplacians $L(x)$ as follows.

Let E be the edge set of the underlying graph. We define a family of Laplacians as

$$L(x) := \sum_{[i,j]\in E} x_{ij} E_{ij}$$

where $x_{ij} \geq \epsilon, \ [i,j] \in E, \ \sum_{[i,j]\in E} x_{ij} = |E|$.

These graphs have edges $[i,j]$ exactly if $[i,j]$ is an edge of G, the weight of all edges is constant ($= |E|$) and positive ($\geq \epsilon > 0$).

$E_{ij} := (e_i - e_j)(e_i - e_j)^T$, (Laplacian of edge $[i,j]$).

Weight redistribution (2)

From before we have, with

$$\mu_1 \leq \ldots \leq \mu_{k-2} < 0 < \mu_{k-1}$$

(holds for any partition m)

$$f(L(x)) = \sum_{i=1}^{k-2} (-\mu_i)\lambda_{i+1}(L(x)) - \mu_{k-1}\lambda_n(L(x)).$$

Due to the signs of μ_i this is a concave function in x.

Weight redistribution (2)

From before we have, with

$$\mu_1 \le \ldots \le \mu_{k-2} < 0 < \mu_{k-1}$$

(holds for any partition m)

$$f(L(x)) = \sum_{i=1}^{k-2} (-\mu_i)\lambda_{i+1}(L(x)) - \mu_{k-1}\lambda_n(L(x)).$$

Due to the signs of μ_i this is a concave function in x.

In case that we can find x such that $f(L(x)) > 0$, then the bandwidth is greater than $t+1$, and no separator of size t exists. Hence we want to maximize the lower bound $f(L(x))$ over all admissable $x > 0, \ \sum_{e \in E} x_e = |E|$.

Maximizing $f(L(x))$

Maximizing f can be done by

- Eigenvalue optimization
(suitable for large problems)

- Semidefinite optimization
(exact optimum, but computationally expensive)

Maximizing f as SDP

Anstreicher-Wolkowicz (2000) show the following:

$$\text{For } A, B \in S_n \quad \min_X \{\langle A, XBX^T \rangle : X^T X = I\} =$$

$$\max_{S,T} \{\text{tr} S + \text{tr} T : B \otimes A - S \otimes I - I \otimes T \succeq 0\}.$$

Maximizing f as SDP

Anstreicher-Wolkowicz (2000) show the following:

$$\text{For } A, B \in S_n \quad \min_X \{\langle A, XBX^T \rangle : X^T X = I\} =$$

$$\max_{S,T} \{\text{tr}S + \text{tr}T : B \otimes A - S \otimes I - I \otimes T \succeq 0\}.$$

In case that order k of B is $\leq n$, this becomes:

$$\text{For } B \in S_k, k \leq n \quad \min_X \{\langle A, XBX^T \rangle : X^T X = I_k\} =$$

$$= \max_{S,T} \text{tr}S + \text{tr}T \text{ such that}$$

$$B \otimes A - S \otimes I - I \otimes T \succeq 0, \; S \in S_k, \; T \in S_n, -T \succeq 0.$$

SDP formulation (2)

From before we have

$$\min_{X \in F_m} \text{tr}(-L)XBX^T = \min_{Z^TZ=I_{k-1}} \text{tr}(-V^TLV)Z(\tilde{B})Z^T$$

$$= \max_{S,T} \text{tr}S + \text{tr}T \text{ such that}$$

$$\tilde{B} \otimes (-V^TLV) - S \otimes I - I \otimes T \succeq 0, \; S \in S_k, \; T \in S_n, -T \succeq 0.$$

SDP formulation (2)

From before we have

$$\min_{X \in F_m} \text{tr}(-L)XBX^T = \min_{Z^T Z = I_{k-1}} \text{tr}(-V^T LV)Z(\tilde{B})Z^T$$

$$= \max_{S,T} \text{tr}S + \text{tr}T \text{ such that}$$

$$\tilde{B} \otimes (-V^T LV) - S \otimes I - I \otimes T \succeq 0, \; S \in S_k, \; T \in S_n, -T \succeq 0.$$

Now substitute $L(x) = \sum_{[i,j] \in E} x_{ij} E_{ij}$ and we get

$$\max_{S,T,x} \text{tr}S + \text{tr}T : \sum_{ij} x_{ij} F_{ij} - S \otimes I - I \otimes T \succeq 0, \; S \in S_k, -T \succeq 0.$$

Here $F_{ij} = -\tilde{B} \otimes (V^T E_{ij} V)$.

SDP formulation (3)

We have (for $x \geq 0, \sum x_{ij} = |E|$)

$$\max_{x} f(x) = \sum_{i=1}^{k-2} (-\mu_i)\lambda_{i+1}(L(x)) - \mu_{k-1}\lambda_n(L(x)) =$$

$$\max_{S,T,x} \text{tr} S + \text{tr} T : \sum_{ij} x_{ij} F_{ij} - S \otimes I - I \otimes T \succeq 0, \; S \in S_k, -T \succeq 0.$$

The SDP is difficult to solve computationally.
The dual slack matrix is of order nk, so there are $O((nk)^2)$ equality constraints.

We work with the dual instead.

Computational limits for $n \approx 100$ and $k \leq 4$.

Eigenvalue maximization approach

We recall the situation. From the Hoffman-Wielandt we have

$$f(x) = \sum_{i=1}^{k-2} (-\mu_i)\lambda_{i+1}(L(x)) - \mu_{k-1}\lambda_n(L(x)).$$

This is concave in x (but nonsmooth) and can therefore be approached using subgradient techniques.

We use bundle methods to maximize this function.

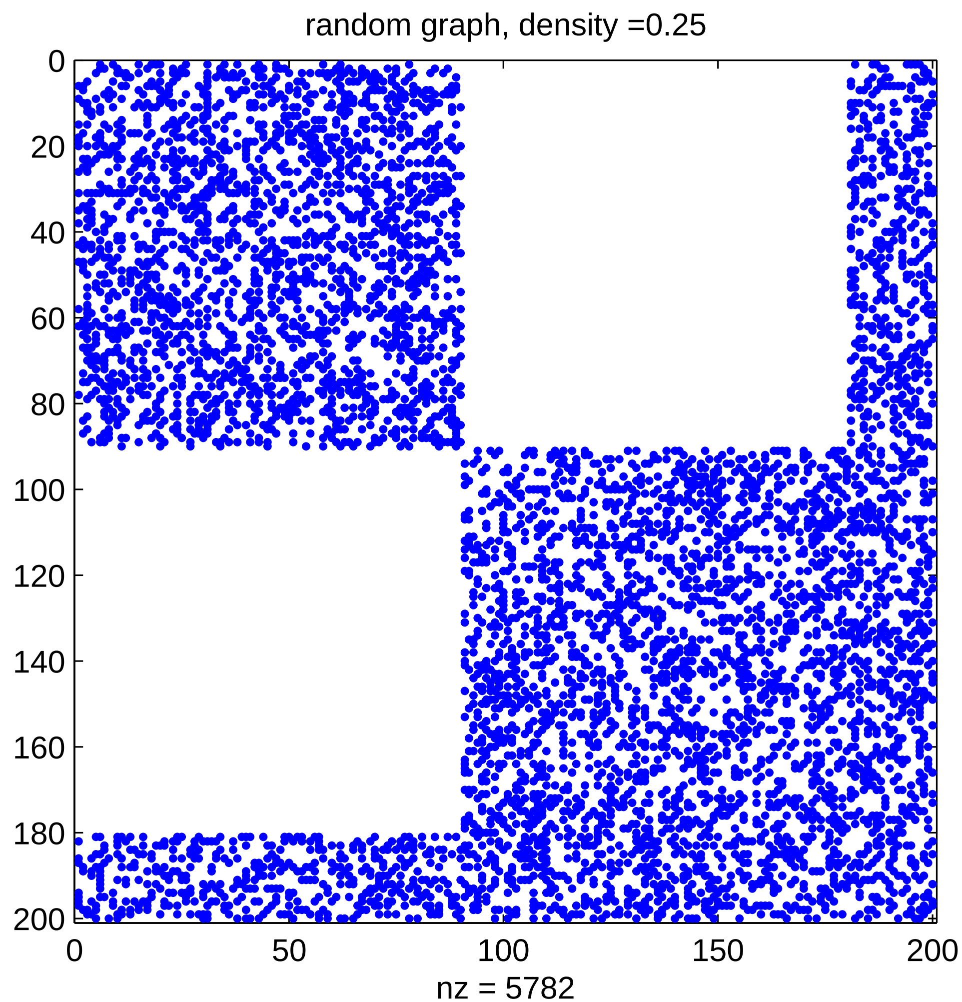
Main interest: Can we get lower bound from <0 to >0

Overview

- Bandwidth and Separators
- Direct modeling (Quadratic Assignment)
- Indirect Model: (Partitioning)
- Weight redistribution
- Preliminary computations

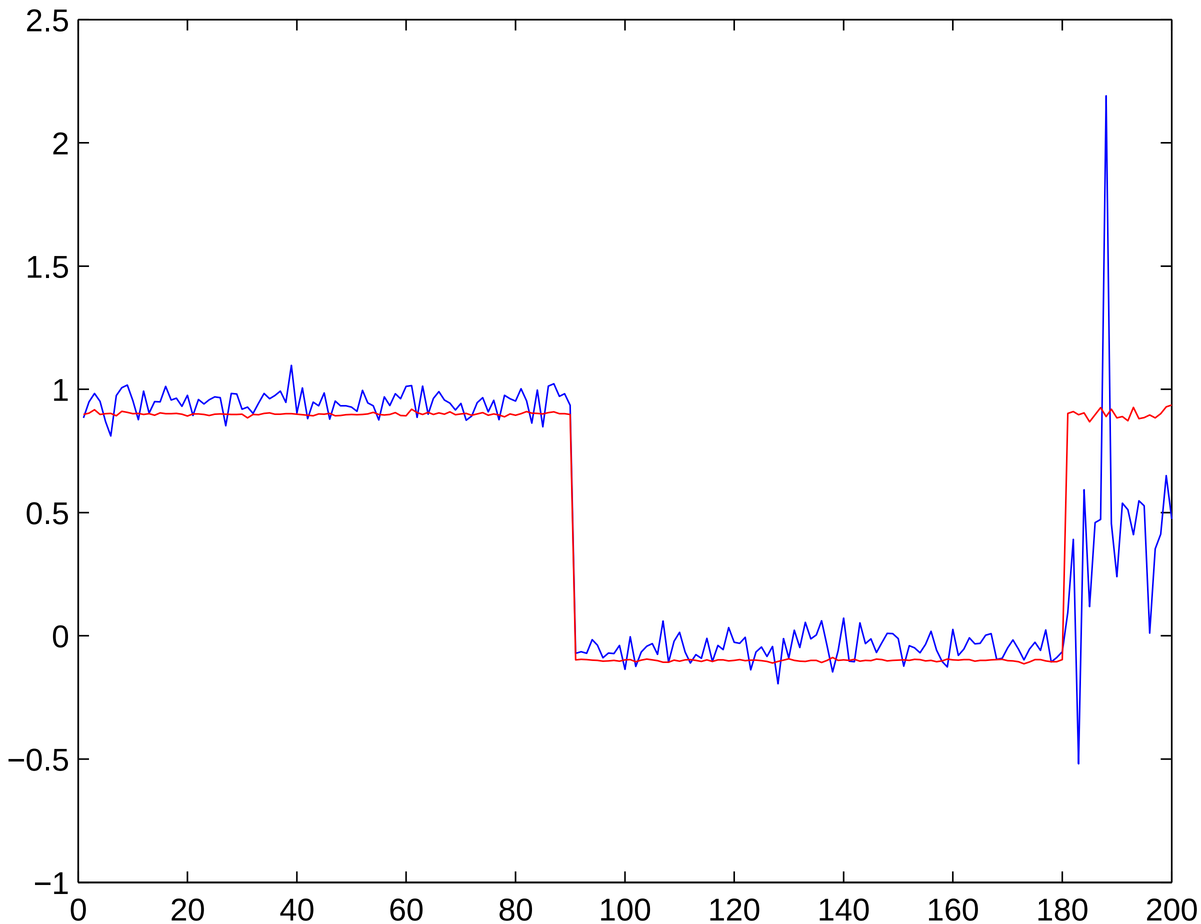
First computational experience

Random graph, with obvious structure, $n = 200$, block size=90



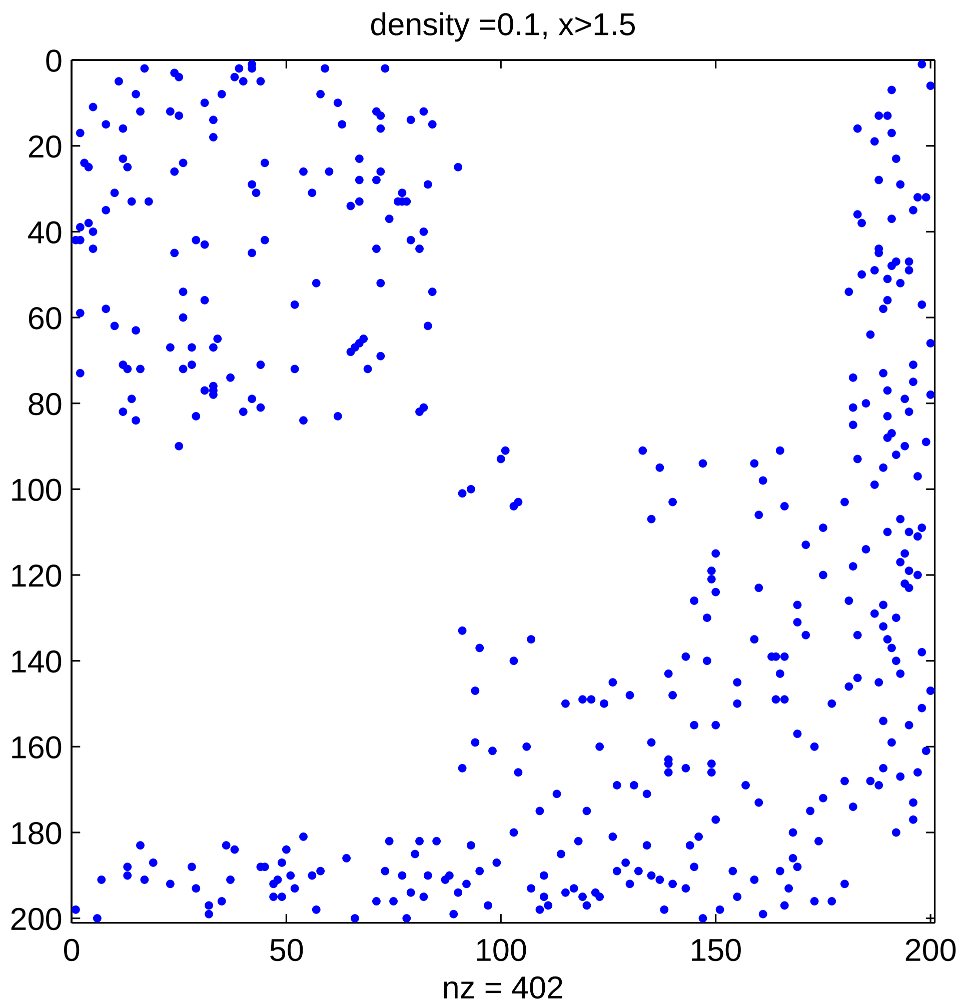
Columns of X

column of X before (blue) and after (red) weight redistribution



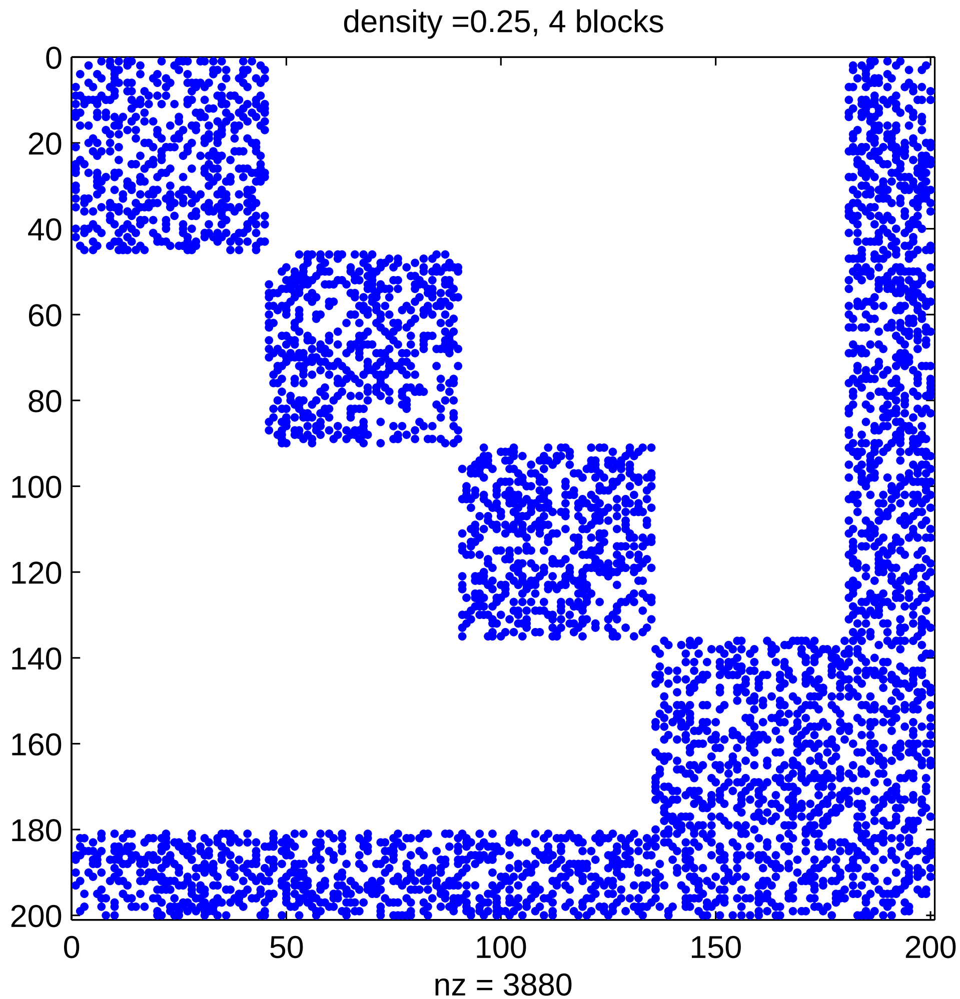
Where are the big weights ?

Edges with weight > 1.5



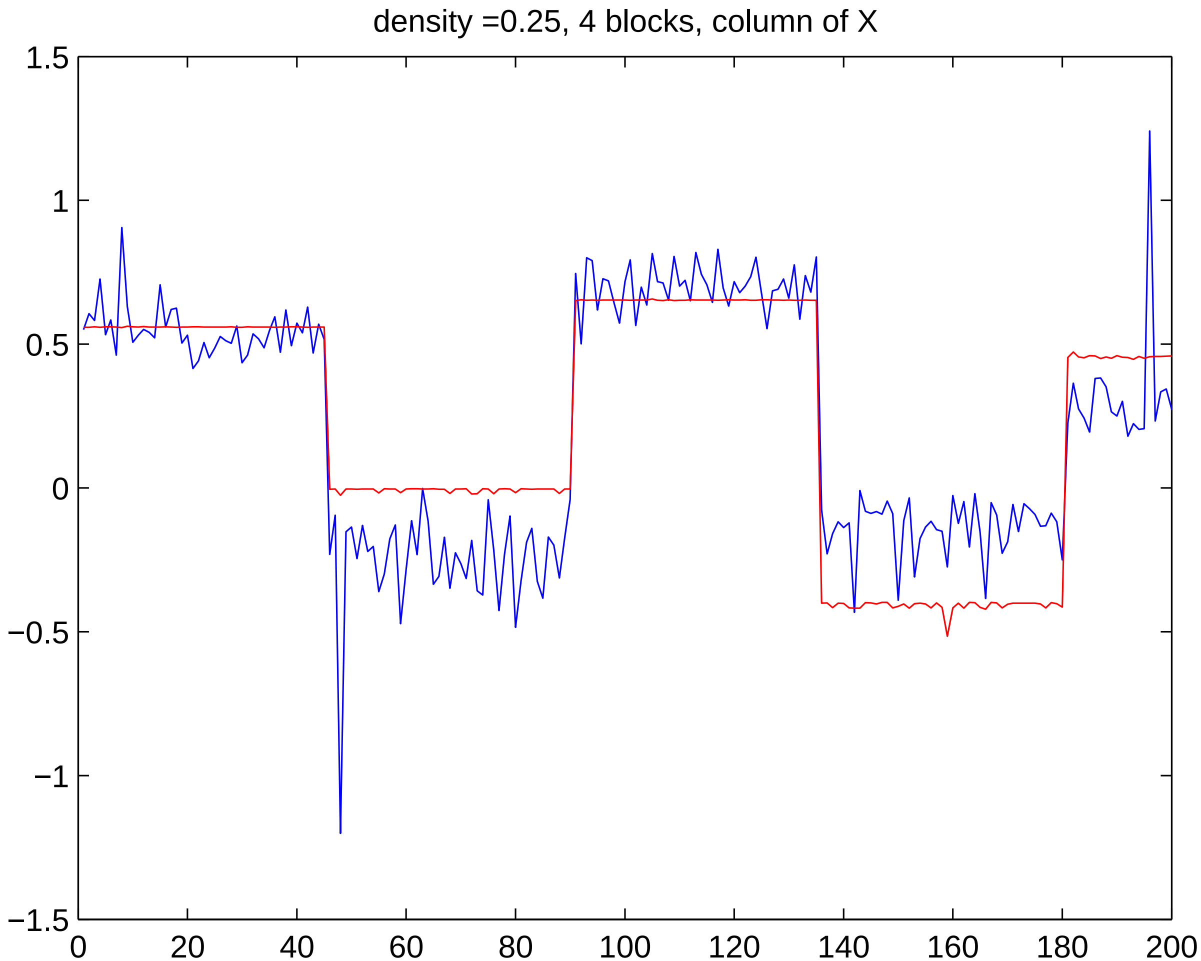
A 5-block example

Random graph, four blocks, $n = 200$, block size=45



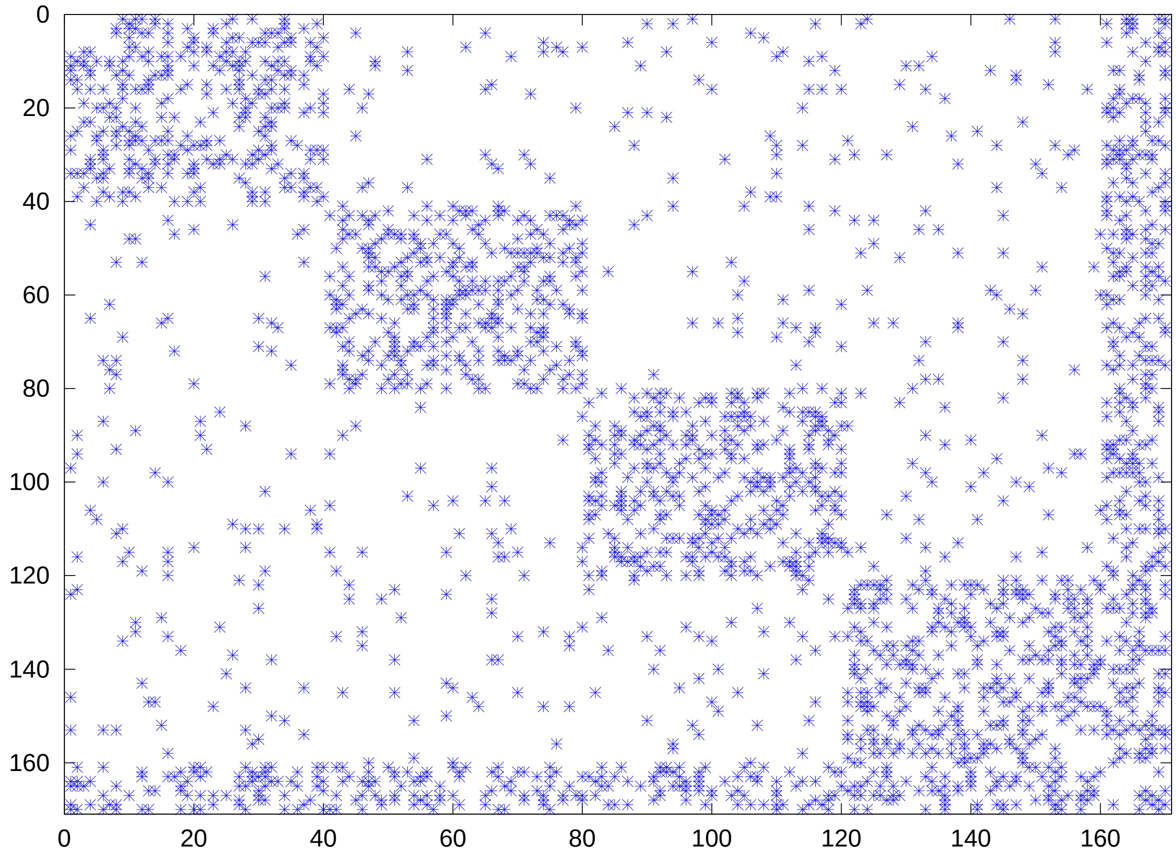
Columns of X

column of X before (blue) and after (red) weight redistribution



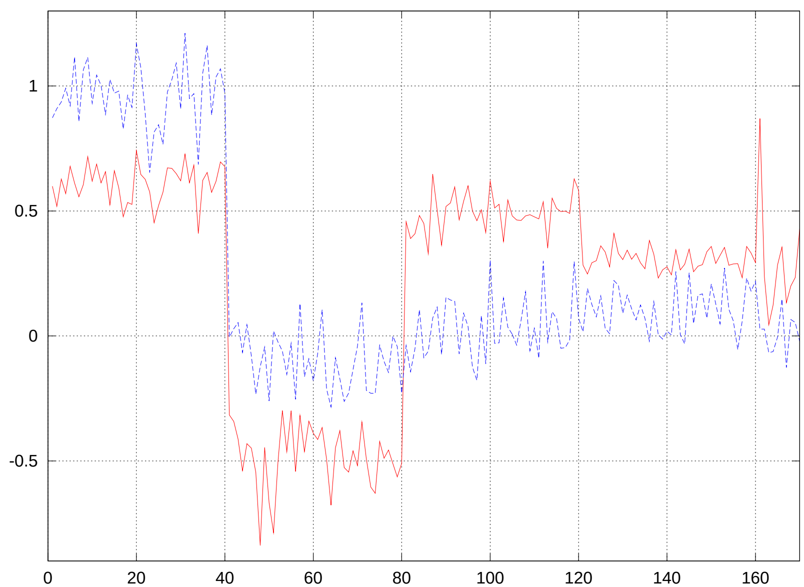
A 5-block example with noise

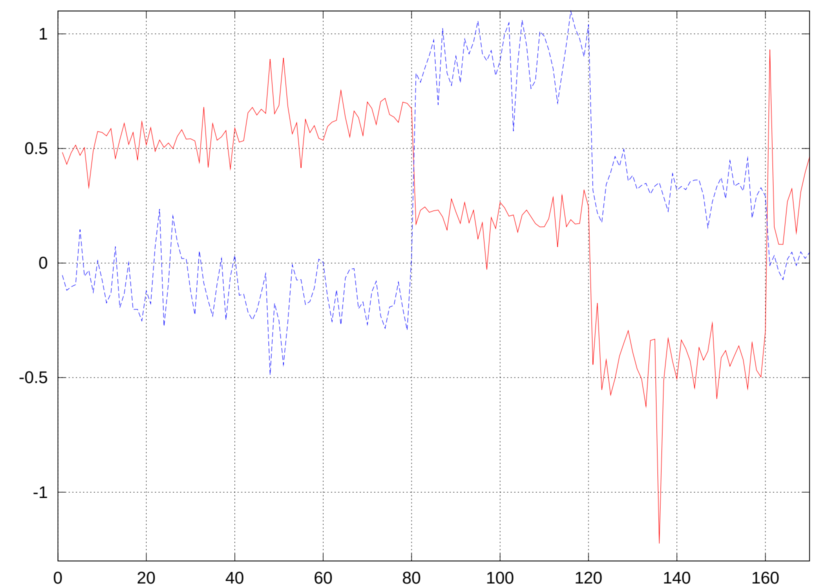
$m = (40,\ 40, 40,\ 40,\ 10)$.



Columns of X

column of X before (red) and after (blue) weight redistribution

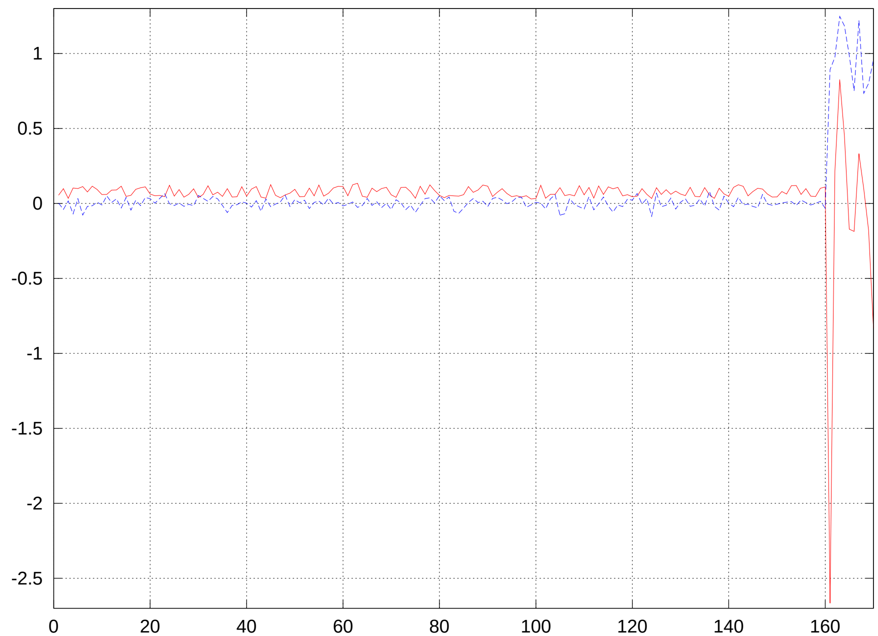




The final (blue) curves give better estimates for partition blocks.

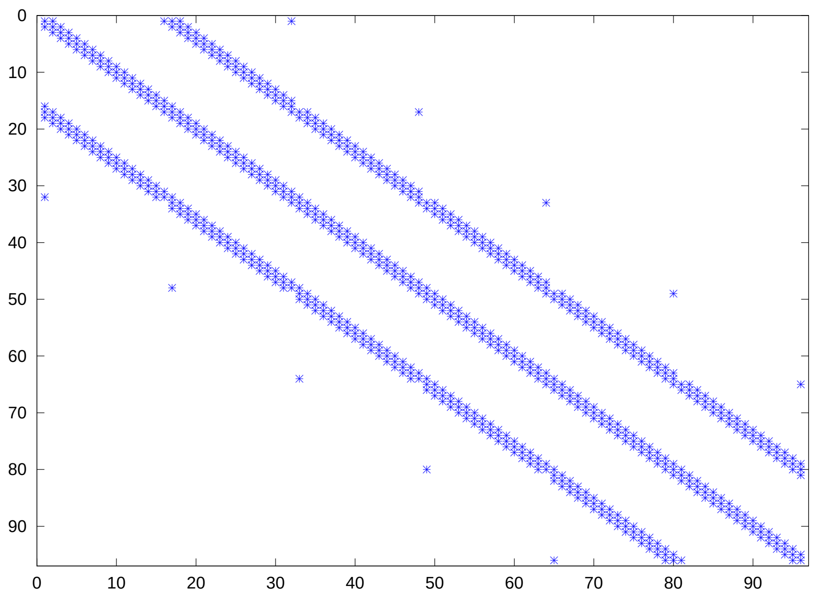
Columns of X

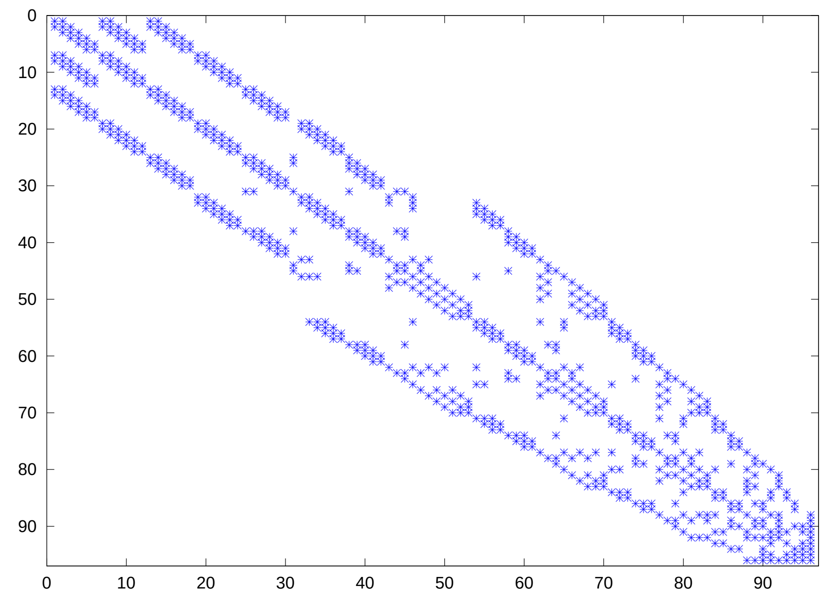
column of X before (red) and after (blue) weight redistribution



An example from the Davis collection

Graph Can96:





Original bandwidth is 31, after reodering it is 21.
Optimization moves lower bound from 3 to 5.

Last Slide

- Rich mathematical structure (Eigenvalues, SDP).
- First results look encouraging.
- Can be done for large graphs using iterative eigenvalue computation
- Can be used for rounding using optimizer X
- Work in progress, jointly done with Abdel Lisser (Orsay, Paris) and Mauro Piacentini (Roma)