Abusing a hypergraph partitioner for unweighted graph partitioning

B. O. Fagginger Auer R. H. Bisseling

Utrecht University

February 13, 2012

Main questions

1. Use: How good is Mondriaan partitioning for what it is meant to do, reducing communication volume?
2. Abuse: How much do we lose in speed and quality by using a more general hypergraph partitioner just for graph partitioning?

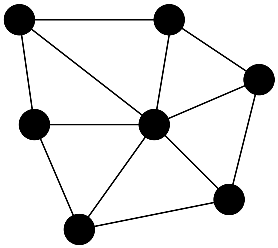
Introduction

- We will discuss applying the Mondriaan matrix partitioner to graph partitioning problems.
- We translate graphs → matrices.
- Mondriaan translates matrices → hypergraphs, and partitions these.
- Going back, this yields a partitioning of the original graph.

Adjacency set of a vertex

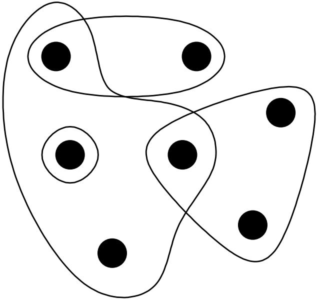
- An undirected graph is a pair (V, E) with vertices v and edges E.
- Edges $e \in E$ are of the form $e = \{u, v\}$ for $u, v \in V$.
- Define the set of neighbours of a vertex $v \in V$ as

$$V_v := \{u \in V \mid \{u, v\} \in E\}.$$



Hypergraphs

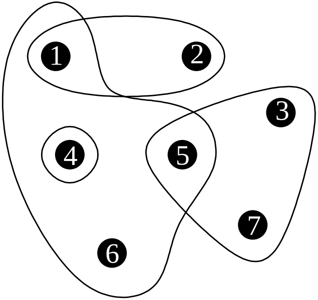
- Hypergraphs generalise graphs: edges can connect any number of vertices.
- A hypergraph is a pair $(\mathcal{V}, \mathcal{N})$ with vertices $\mathcal{V}$ and nets (or hyperedges) $\mathcal{N}$.
- Each net $n \in \mathcal{N}$ is a set of vertices: $n \subseteq \mathcal{V}$.



Matrices

- Pattern matrices $A \in \{0,1\}^{m \times n}$ can be represented by hypergraphs.
- Mondriaan supports three hypergraph representations:
  - row-net (Çatalyürek & Aykanat, 1999): columns are vertices, rows are nets.
  - column-net (Çatalyürek& Aykanat, 1999): rows are vertices, columns are nets.
  - finegrain (Çatalyürek& Aykanat, 2001): nonzeros are vertices, rows and columns are nets.
- We will use the row-net representation.

Row-net hypergraph



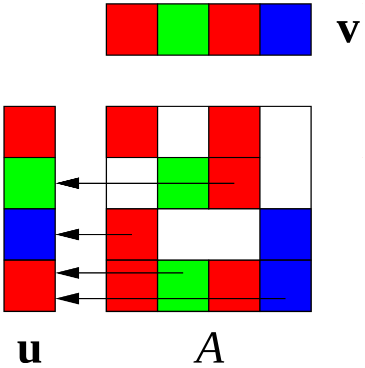
$$\mathcal{G} = (\mathcal{V}, \mathcal{N})$$
$$\mathcal{V} = \{1, 2, 3, 4, 5, 6, 7\}$$
$$\mathcal{N} = \{\{1, 2\}, \{1, 4, 5, 6\}, \{4\}, \{3, 5, 7\}\}$$

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Mondriaan partitioner

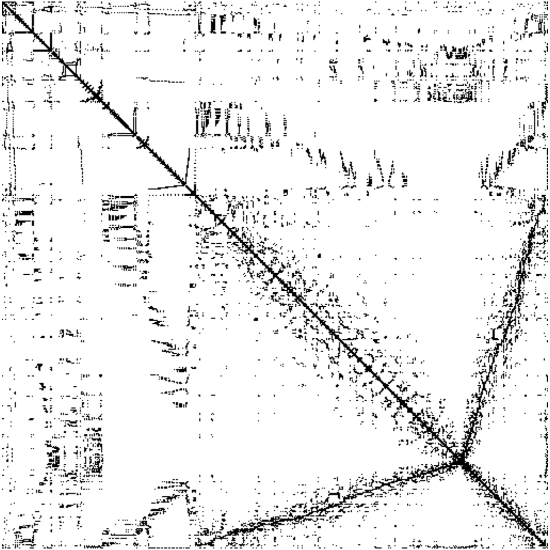
- Mondriaan is a matrix partitioner for parallel sparse matrix–dense vector multiplication.
- Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{v} \in \mathbb{R}^n$ be the matrix and input vector, and define the output vector by
$$\mathbf{u} := A\mathbf{v}.$$
- This multiplication is done in parallel: nonzeros of A and vector components of $\mathbf{u}$ and $\mathbf{v}$ are distributed among processors.
- Mondriaan tries to minimize parallel communication, while balancing the number of nonzeros per processor.

Column partitioning (Mondriaan in 1D mode)

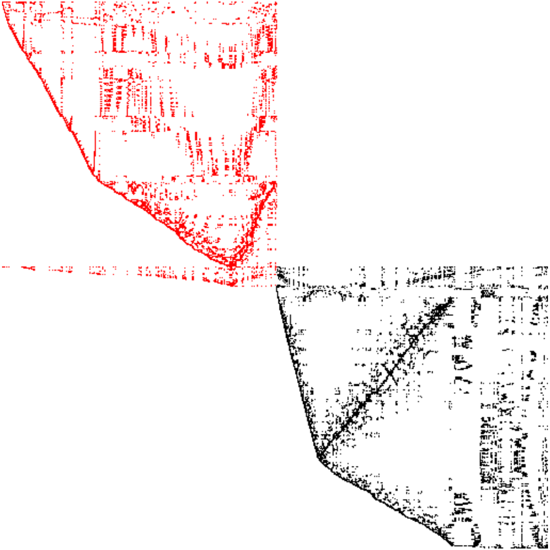


Partitioning among $k = 3$ processors with communication volume 4.

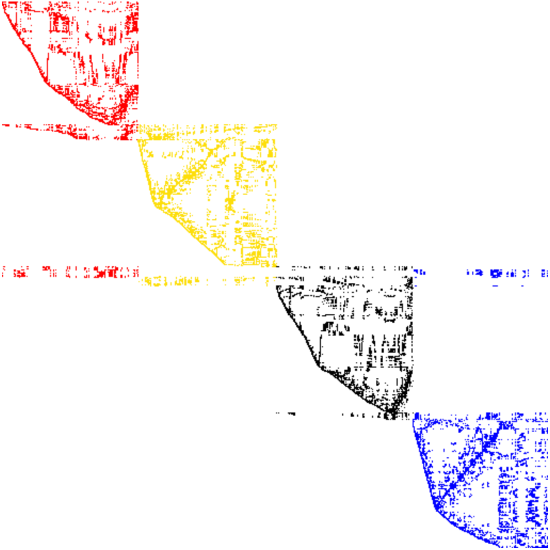
Column partitioning (`fe_tooth`)



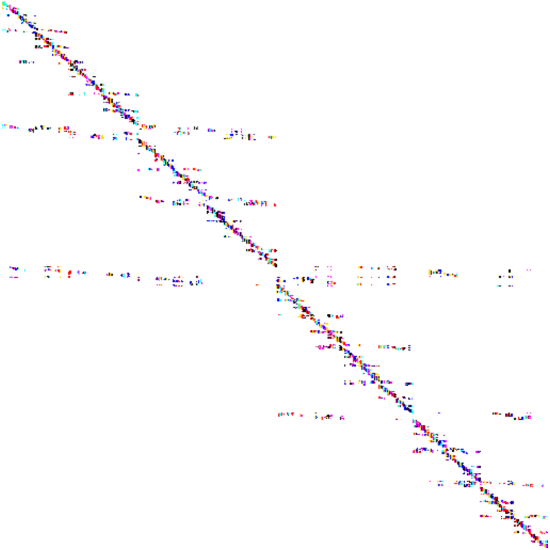
Column partitioning (fe_tooth)



Column partitioning (fe_tooth)



Column partitioning (`fe_tooth`)



Hypergraph partitioning problem

- Mondriaan solves the following problem internally.

Problem

Let $\mathcal{G} = (\mathcal{V}, \mathcal{N})$ be a hypergraph, with vertex weights $\zeta : \mathcal{V} \to \mathbb{R}_{>0}$, $k \in \mathbb{N}$ the number of parts, and $\epsilon > 0$ an imbalance factor.
Find a partitioning $\Pi : \mathcal{V} \to \{1, \ldots, k\}$ that is balanced:

$$\sum_{\substack{v \in V \\ \Pi(v) = i}} \zeta(v) \leq (1 + \epsilon) \left\lceil \frac{\zeta(\mathcal{V})}{k} \right\rceil, \qquad (1 \leq i \leq k), \tag{1}$$

which minimises the $(\lambda - 1)$-metric

$$LV(\Pi) := \sum_{n \in \mathcal{N}} (|\Pi(n)| - 1). \tag{2}$$

Use: Graph partitioning problem (Communication Volume)

- For DIMACS, the communication volume problem is similar.

Problem

Let $G = (V, E)$ be a graph, $k \in \mathbb{N}$ the number of parts, and $\epsilon > 0$ an imbalance factor.
Find a partitioning $\Pi : V \to \{1, \ldots, k\}$ that is balanced:

$$|\{v \in V \ : \ \Pi(v) = i\}| \leq (1 + \epsilon) \left\lceil \frac{|V|}{k} \right\rceil, \qquad (1 \leq i \leq k), \tag{3}$$

which minimises the communication volume

$$CV(\Pi) := \max_{1 \leq i \leq k} \sum_{\substack{v \in V \\ \Pi(v) = i}} |\Pi(V_v) \setminus \{\Pi(v)\}|. \tag{4}$$

Graph partitioning problem (Communication Volume)

- How to use Mondriaan to solve this DIMACS problem?
- Balancing and partitioning: pick $\mathcal{V} := V$ and $\zeta(v) := 1$ for all $v \in V$.
- We should choose nets $\mathcal{N}$ to match

$$\sum_{n \in \mathcal{N}} (|\Pi(n)| - 1) \qquad \text{and} \qquad \max_{1 \le i \le k} \sum_{\substack{v \in V \\ \Pi(v) = i}} |\Pi(V_v) \setminus \{\Pi(v)\}|.$$

- Note that

$$|\Pi(V_v) \setminus \{\Pi(v)\}| = |\Pi(V_v \cup \{v\})| - 1.$$

- Hence, $\mathcal{N} := \{V_v \cup \{v\} \mid v \in V\}$ is a good choice.

Graph partitioning problem (Communication Volume)

- With $\mathcal{V} := V$ and $\mathcal{N} := \{V_v \cup \{v\} \mid v \in V\}$, we have

$$\mathrm{LV}(\Pi) = \sum_{v \in V} |\Pi(V_v) \setminus \{\Pi(v)\}|.$$

- Comparing this with

$$\mathrm{CV}(\Pi) = \max_{1 \leq i \leq k} \sum_{\substack{v \in V \\ \Pi(v) = i}} |\Pi(V_v) \setminus \{\Pi(v)\}|,$$

we find that

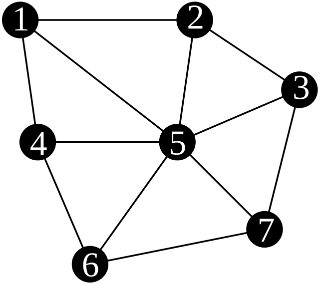
$$\frac{1}{k} \mathrm{LV}(\Pi) \leq \mathrm{CV}(\Pi) \leq \mathrm{LV}(\Pi).$$

- Thus we minimise $\mathrm{LV}(\Pi)$, in the hope of minimising $\mathrm{CV}(\Pi)$.

Translate hypergraph to adjacency matrix

- Row-net matrix representation $A \in \{0,1\}^{|V|\times|V|}$:

$$a_{u\,v} := \begin{cases} 1 & u = v \text{ or } \{u, v\} \in E, \\ 0 & \text{otherwise}, \end{cases} \qquad (u, v \in V).$$



$$A = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Abuse: Graph partitioning problem (Edge Cut)

- The edge cut problem is easier.

Problem

Let $G = (V, E)$ be a graph, $k \in \mathbb{N}$ the number of parts, and $\epsilon > 0$ an imbalance factor.
Find a partitioning $\Pi : V \to \{1, \dots, k\}$ that is balanced:

$$|\{v \in V \ : \ \Pi(v) = i\}| \leq (1 + \epsilon) \left\lceil \frac{|V|}{k} \right\rceil, \qquad (1 \leq i \leq k), \tag{5}$$

which minimises the edge cut

$$EC(\Pi) := |\{\{u, v\} \in E \mid \Pi(u) \neq \Pi(v)\}|. \tag{6}$$

Graph partitioning problem (Edge Cut)

- Again: pick $\mathcal{V} := V$ and $\zeta(v) := 1$ for all $v \in V$.
- Note that for vertices $u, v \in V$,

$$|\Pi(\{u, v\})| - 1 = \begin{cases} 1 & \Pi(u) \neq \Pi(v), \\ 0 & \text{otherwise.} \end{cases}$$

- Pick $\mathcal{N} := E$, then

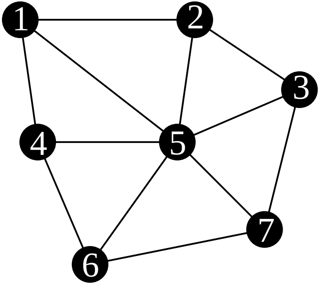
$$\mathrm{EC}(\Pi) = \mathrm{LV}(\Pi).$$

- Thus, every edge becomes a hyperedge.

Translate hypergraph to incidence matrix

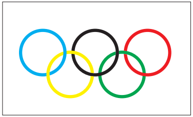
- Row-net matrix representation $A \in \{0,1\}^{|E|\times|V|}$:

$$a_{e\,v} := \begin{cases} 1 & v \in e, \\ 0 & \text{otherwise}, \end{cases} \qquad (e \in E, v \in V).$$



$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Olympic spirit



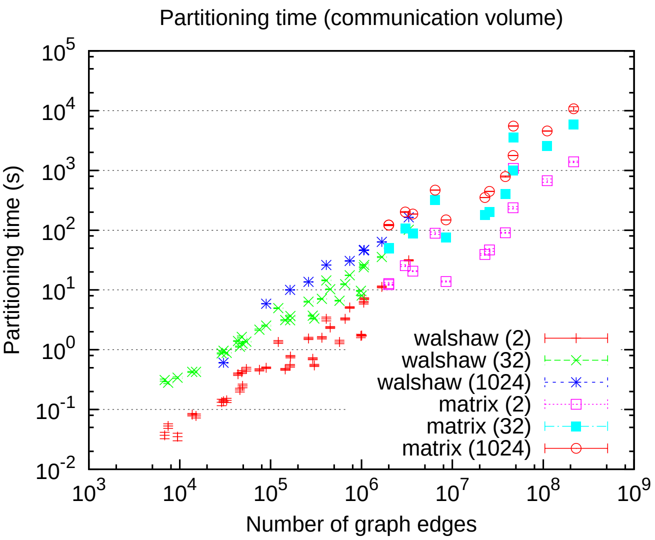
"The most important thing in the Olympic Games is not to win but to take part, just as the most important thing in life is not the triumph but the struggle. The essential thing is not to have conquered but to have fought well." *(Pierre de Coubertin)*

- For DIMACS challenge paper, we measure average volumes from a set of 5 or 10 runs, not the best from the set.
- Updated version of Mondriaan 3.11 with its own partitioner.
- Aim: investigate how standard Mondriaan fares on average.
- For DIMACS solution challenge, we submitted the first run.

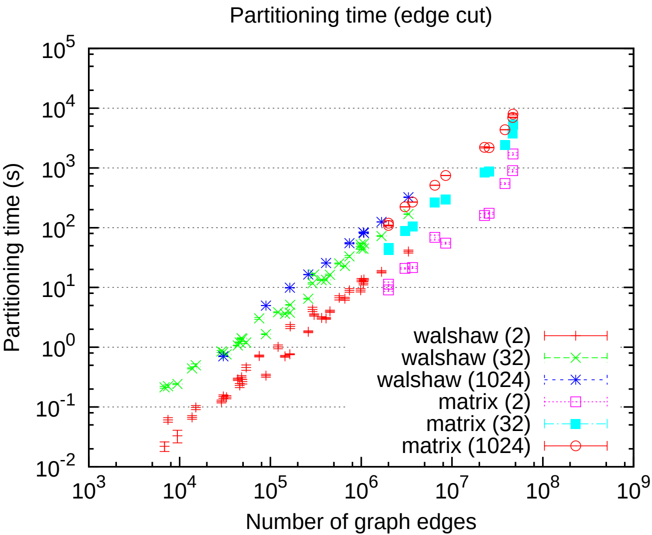
Test setup

- Test set: 10th DIMACS challenge `walshaw/` and `matrix/` categories.
- Imbalance is set to $\epsilon = 0.03$ and $k = 2, 4, \ldots, 1024$.
- Test hardware: dual quad-core AMD Opteron 2378 (Mondriaan uses one core).

Results (Communication Volume)



Results (Edge Cut)

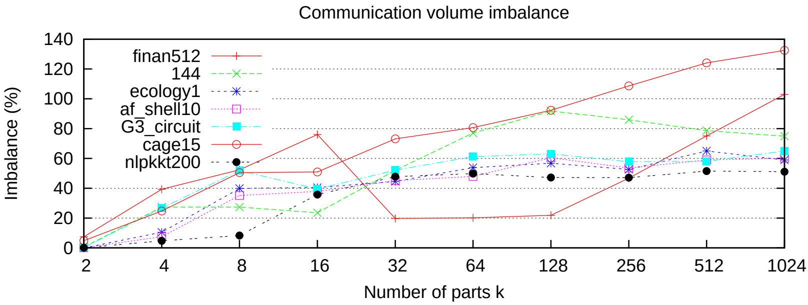


Results: how do we fare?

- Edge cut is, on average, 21% worse than best known (Walshaw).
- Communication volume imbalance,

$$\frac{\mathrm{CV}(\Pi)}{\mathrm{LV}(\Pi)/k} - 1,$$

is very small compared to its maximum $k - 1$. We attribute this to unbiased tie breaking in Mondriaan.



Conclusion

- It is possible to abuse Mondriaan to perform graph partitioning.
- Mondriaan's average edge cut is 21% higher than the best known (Walshaw).
- Partitionings generated by Mondriaan have reasonably balanced communication volume.
- We identified future additions to Mondriaan:
  - net weights to allow for full generality of DIMACS challenge;
  - volume equaliser in matrix partitioner for a little extra gain in communication balance.