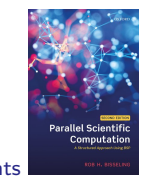
Further Improvements

Section 5.8 of Parallel Scientific Computation, 2nd edition

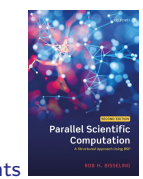
Rob H. Bisseling

Utrecht University

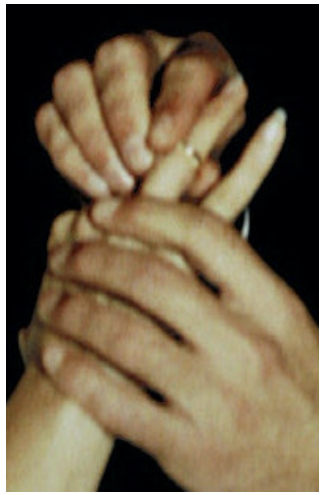


Reducing the number of proposals

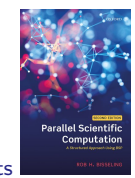
- The parallel matching algorithm can be improved by trying to reduce the number of proposals.
- Adding a motivation to each rejection of a proposal may help the receiving processor avoid sending futile proposals in the future.
- If vertex v proposes to u, but u rejects because it has a:
  - better suitor x, then processor $P(\phi(u))$ can send back the weight $\omega_{\text{suitor}}(u) = \omega(u, x)$ to $P(\phi(v))$ as a motivation;
  - match, then the processor can send back $\omega_{\text{suitor}}(u) = \infty$.
- No vertex residing on $P(\phi(v))$ with a weight $< \omega_{\text{suitor}}(u)$ will ever propose again to vertex u.
- This might save many proposals, especially at a later stage, when the suitor weights have increased.



Wedding rings

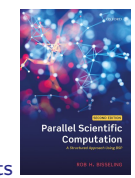


- Wedding rings are commonly exchanged as a symbol of love and to tell all the world about a marriage.
- In algorithms, the equivalent is to broadcast a match to all processors in the vicinity.
- Proposing marriage to someone already wearing a wedding ring or proposing a match to an already matched vertex is useless.
- Thus, a further improvement is to broadcast $\omega_{\text{suitor}}(v) = \infty$ to the whole set $\{P(\phi(u)) \ : \ (u, v) \in \mathcal{E}_s\}$, once $v \in \mathcal{V}_s$ matches.



Additional communication volume of broadcasts

- To get an upper bound on the extra costs, we assume that
  - all vertices are matched during the algorithm;
  - all matches are broadcast.
- In the worst case, this broadcast is based on the original edge set $\mathcal{E}_s$, before any edges have been removed.
- The additional communication volume then depends completely on the partitioning ϕ of the vertices.

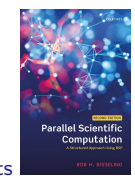


Correspondence between graph and adjacency matrix

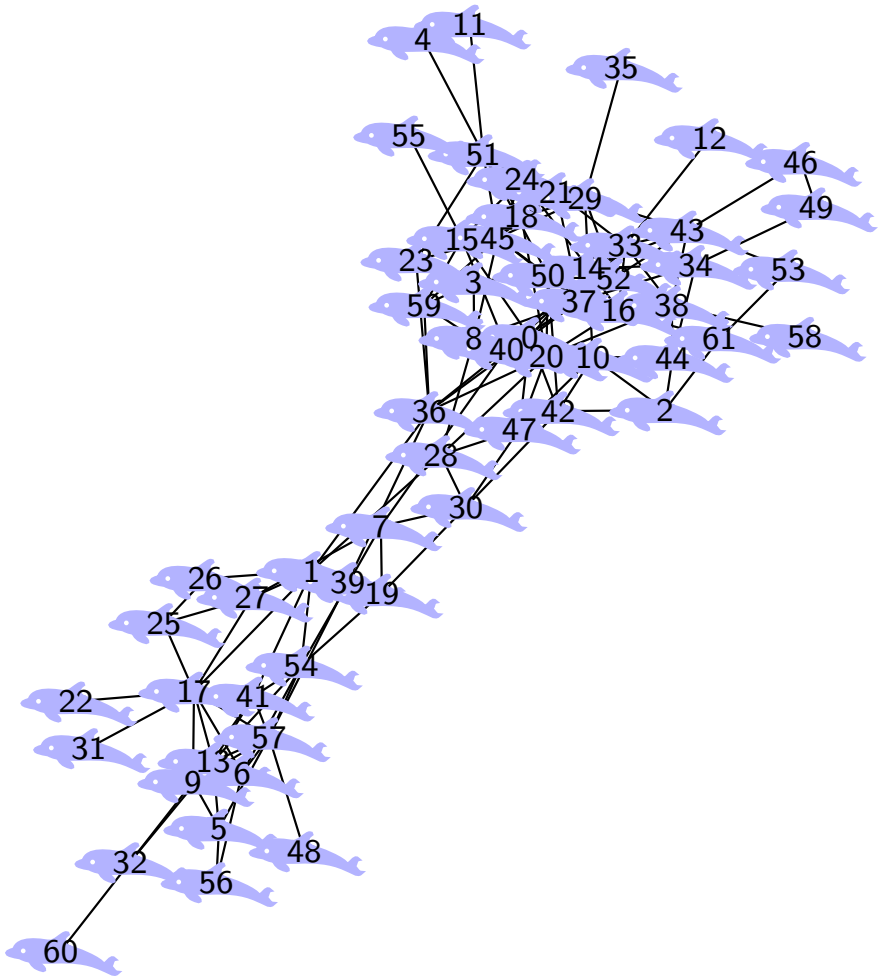
- The worst-case additional communication volume of the broadcasts equals the volume of a corresponding parallel SpMV for a suitably defined matrix and vector.
- The adjacency matrix $A = A(\mathcal{G})$ of an undirected simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined by

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } 0 \leq i,j < n.$$

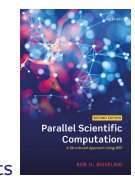
- The adjacency matrix is binary (having only values 0 and 1), sparse, symmetric, and it has a zero diagonal.



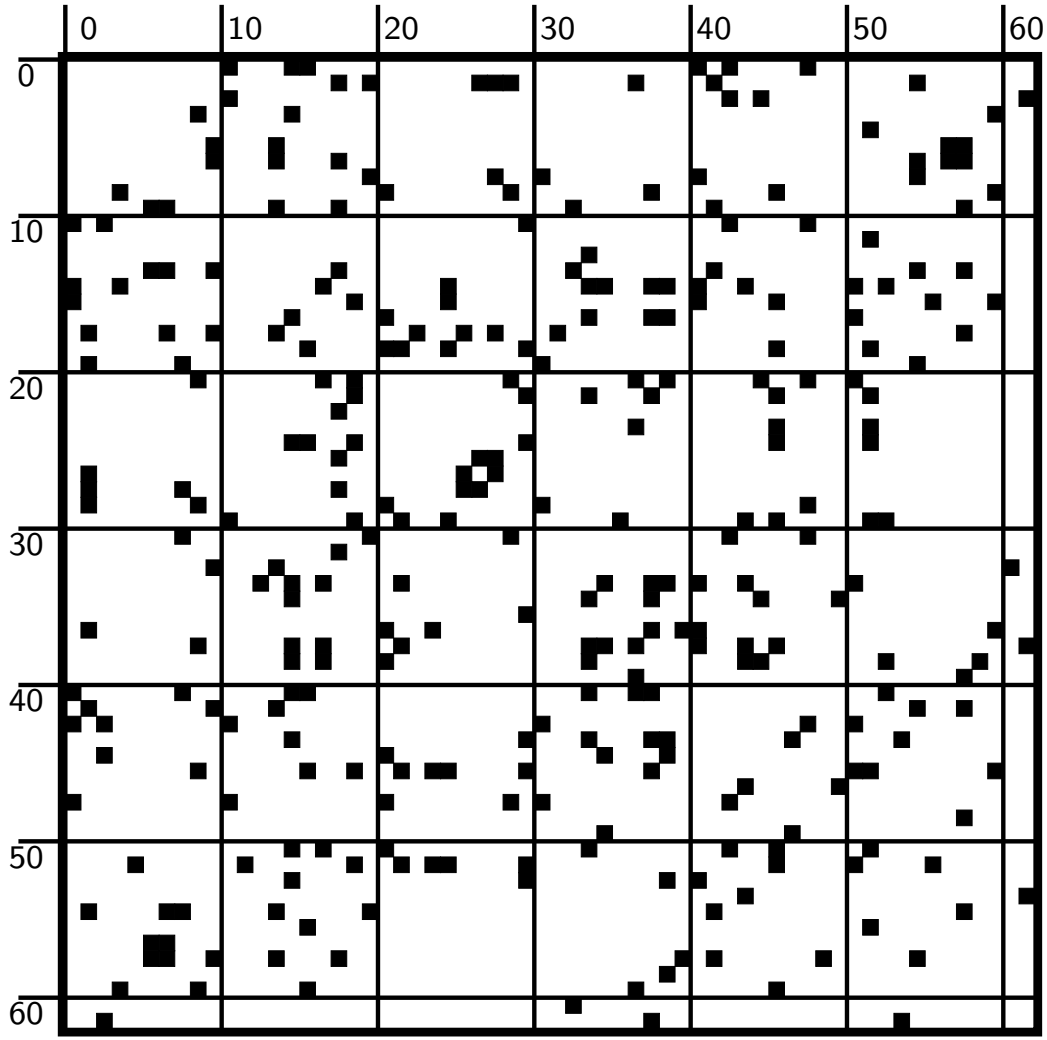
Social network of dolphins



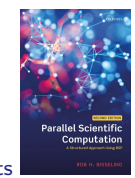
- A social network of 62 bottlenose dolphins with 159 frequent associations between them in a community living off the Doubtful Sound fjord, New Zealand.



Adjacency matrix of dolphins network



- 62×62 symmetric sparse matrix A.
  $nz(A) = 318$.

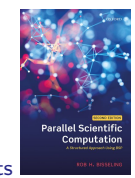


Distributing the vertices as rows of $A + I$

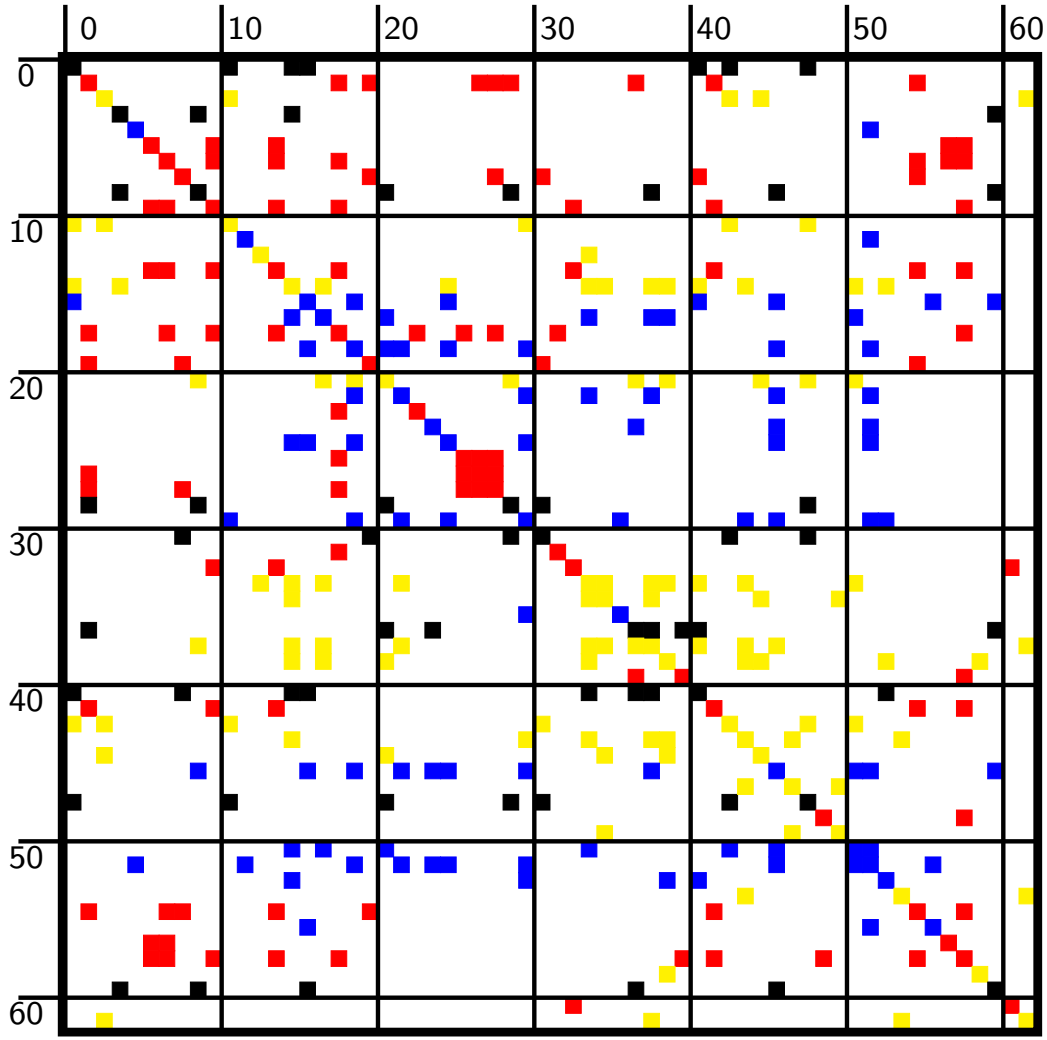
- We provide the matrix with a unit diagonal by adding the identity matrix $I = I_n$.
- We define a row distribution ϕ_{A+I} of the matrix $A + I$ corresponding to the vertex partitioning ϕ of the graph $\mathcal{G}$ by

$$\phi_{A+I}(i,j) = \phi(i), \quad \text{for } a_{ij} \neq 0 \vee i = j.$$

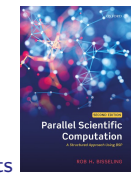
- We define a corresponding vector distribution by setting $\phi_{\mathbf{v}} = \phi_{\mathbf{u}} = \phi$.
- Here, $\mathbf{v}$ is the input vector and $\mathbf{u}$ the output vector of the parallel SpMV $\mathbf{u} := (A + I)\mathbf{v}$.



Partitioning of the matrix $A + I$ for $p = 4$ and $\epsilon = 0.2$

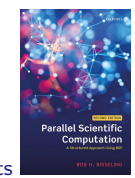


- Partitioning with $V = 47$, $EC = 47$ (coincidence!) obtained by using Mondriaan in 1D row mode.

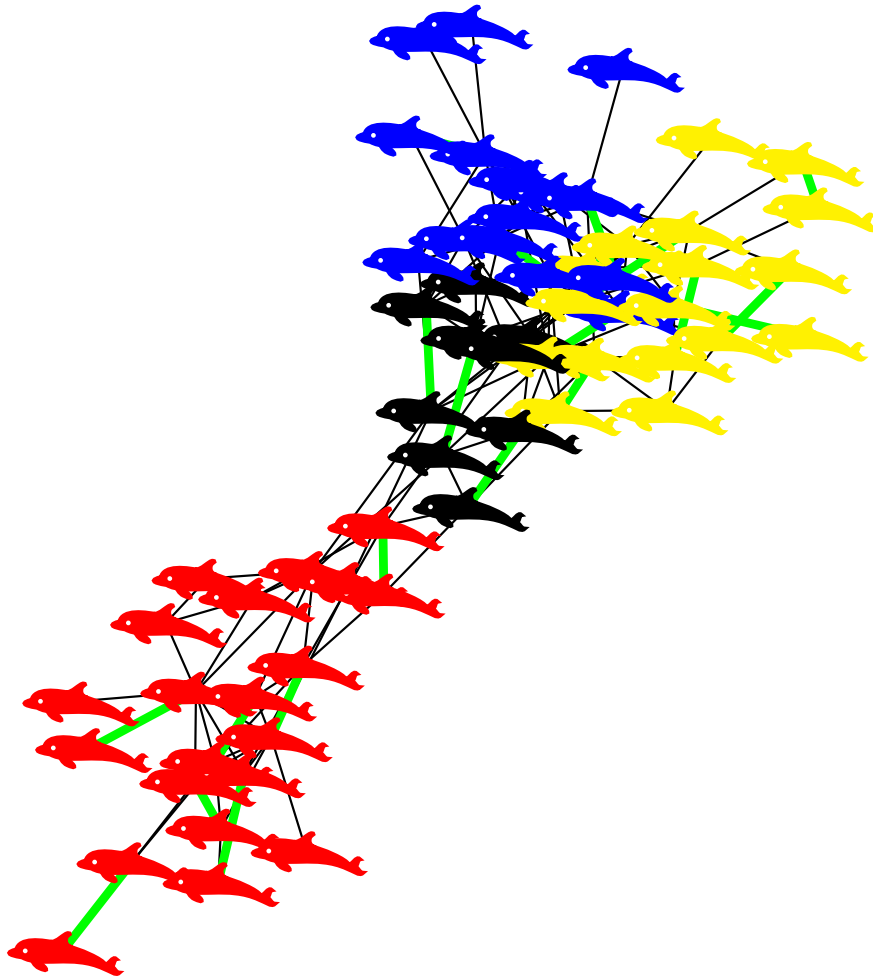


Why add the unit diagonal?

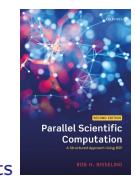
- This ensures that the processor $P(\phi(j))$ owning vertex j and hence matrix row j of $A + I$ is also represented in matrix column j.
- The remote processors owning nonzeros in column j of $A + I$ are exactly the processors connected by a cut edge to vertex j, which receive the broadcast of the matching of j.
- Therefore, the communication volume $V_{\phi_{A+I}}$ of the fanout of the parallel SpMV (and hence the whole SpMV) equals that of the match broadcasts.



Matched social network of dolphins

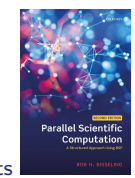


- The dolphins are matched on 4 processors based on uniform edge weights and with preference for local matches, giving 23 internal matches and 1 external match.



Load balancing constraint

- When partitioning the vertices of the graph, we can either balance the number of vertices, or the number of edges.
- Both are only approximate indications of the amount of work of the matching algorithm, which is more dynamic in nature than the SpMV, where the nonzero balance represents the true work balance.
- Because processors with more edges will have more work searching and splitting adjacency lists, balancing the edges better reflects the load balance of the matching algorithm.
- This also balances the memory requirements.
- Using Mondriaan, we can add a unit diagonal to the input matrix A without affecting load balance by setting the options `SquareMatrix_DistributeVectorsEqual` and `SquareMatrix_DistributeVectorsEqual_AddDummies`.



Theorem 5.3: communication volume and edge cut

Theorem

Let V_ϕ be the communication volume of a broadcast induced by vertex partitioning ϕ of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

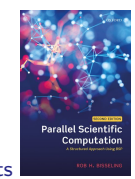
$$V_\phi = \sum_{v \in \mathcal{V}} |\{\phi(u) \ : \ (u, v) \in \mathcal{E} \ \wedge \ \phi(u) \neq \phi(v)\}|.$$

Let EC_ϕ be the edge cut of the vertex partitioning,

$$EC_\phi = |\{(u, v) \in \mathcal{E} \ : \ \phi(u) \neq \phi(v)\}|.$$

Then

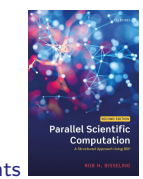
$$V_\phi \leq 2 \cdot EC_\phi.$$



Proof of Theorem 5.3

- Define the indicator function $\mathbf{1}_X$ as 1 if the statement X holds, and 0 otherwise.
- Because every edge is connected to two vertices, we have that

$$
\begin{aligned}
2 \cdot EC_\phi &= \sum_{v \in \mathcal{V}} \sum_{\substack{(u,v) \in \mathcal{E} \\ \phi(u) \neq \phi(v)}} 1 \\
&= \sum_{v \in \mathcal{V}} \sum_{\substack{s=0 \\ s \neq \phi(v)}}^{p-1} \sum_{\substack{(u,v) \in \mathcal{E} \\ \phi(u) = s}} 1 \\
&\geq \sum_{v \in \mathcal{V}} \sum_{\substack{s=0 \\ s \neq \phi(v)}}^{p-1} \mathbf{1}_{\exists u : (u,v) \in \mathcal{E} \wedge \phi(u) = s} \\
&= \sum_{v \in \mathcal{V}} |\{\phi(u) \ : \ (u,v) \in \mathcal{E} \wedge \phi(u) \neq \phi(v)\}| \\
&= V_\phi. \qquad \square
\end{aligned}
$$

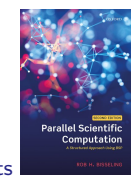


Potential gains of broadcasting matches

- In the worst case, the communication cost is doubled when a broadcast volume of $V_\phi = 2 \cdot EC_\phi$ is added to the volume of EC_ϕ proposals and EC_ϕ answers.
- But the worst case is highly unlikely, because then all cut edges of each vertex must be connected to different processors. Usually, however,

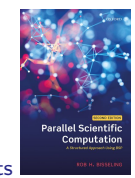
$$V_\phi \ll EC_\phi.$$

- In the best case, most proposals are prevented and the total communication volume is close to V_ϕ.
- In that case, the hypergraph-based partitioning obtained for parallel SpMV is also the right partitioning for the graph matching algorithm.
- The potential gains of broadcasting matches outweigh the potential losses.



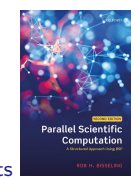
Final possible improvement: 2D partitioning

- We can use a 2D matrix distribution instead of a 1D row distribution, partitioning the matrix nonzeros instead of the rows.
- For the graph, this means partitioning the edges instead of the vertices.
- The 2D approach is more general and gives partitionings with lower communication volume and better load balance.
- The 2D approach may especially be beneficial for graphs with widely varying vertex degrees, such as power-law graphs, where the fraction of vertices of degree d scales as $\mathcal{O}(d^{-\alpha})$.
- In a 1D approach, a high-degree vertex with a long adjacency list requires much more computation than other vertices, thus harming load balance.



2D matching algorithm

- A 2D matching algorithm is more complicated than a 1D matching algorithm.
- For 2D, both the vertices and edges will be distributed.
- We need to find the preferences in a distributed manner, because the edges connected to a vertex v will not all be stored on processor $P(\phi(v))$.
- Thus, $P(\phi(v))$ will first have to request candidates for its new preference and only then can it decide on the best.



Summary

- The number of proposals sent in the parallel matching algorithm can be reduced by sending motivated rejections and by broadcasting matches,
- The adjacency matrix $A = A(\mathcal{G})$ of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined by

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } 0 \leq i,j < n.$$

- The additional communication volume of the match broadcasts is at most the volume V_ϕ of a parallel SpMV for the row-distributed matrix $A + I$,
- An upper bound on V_ϕ is

$$V_\phi \leq 2 \cdot EC_\phi,$$

  where EC_ϕ is the edge cut. But usually, $V_\phi \ll EC_\phi$.
- 2D matrix partitioning can further reduce communication volume and improve load balance.