1 Minimum Spanning Trees

undirected Graph $G = (V, E)$. 

nodes $V$, $n = |V|$, e.g., $V = \{1, \ldots, n\}$

edges $e \in E$, $m = |E|$, two-element subsets of $V$.

edge weight $c(e)$, $c(e) \in \mathbb{R}_+$. 

$G$ is connected, i.e., $\exists$ path between any two nodes.

Find a tree $(V, T)$ with minimum weight $\sum_{e \in T} c(e)$ that connects all nodes.
MST: Overview

- Basics: Edge property and cycle property
- Jarník-Prim Algorithm
- Kruskals Algorithm
- Some tricks and comparison
- Advanced algorithms using the cycle property
- External MST

Applications: Clustering; subroutine in combinatorial optimization, e.g., Held-Karp lower bound for TSP. Challenging real world instances???
Anyway: almost ideal “fruit fly” problem
Selecting and Discarding MST Edges

The Cut Property

For any $S \subset V$ consider the cut edges

$$C = \{(u, v) \in E : u \in S, v \in V \setminus S\}$$

The lightest edge in $C$ can be used in an MST.

The Cycle Property

The heaviest edge on a cycle is not needed for an MST.
The Jarník-Prim Algorithm [Jarník 1930, Prim 1957]

Idea: grow a tree

\[ T := \emptyset \]
\[ S := \{s\} \text{ for arbitrary start node } s \]

repeat \(n-1\) times
  find \((u, v)\) fulfilling the cut property for \(S\)
  \[ S := S \cup \{v\} \]
  \[ T := T \cup \{(u, v)\} \]
Implementation Using Priority Queues

Function jpMST(V, E, w) : Set of Edge

\[ \text{dist} = [\infty, \ldots, \infty] : \text{Array} [1..n] \]  
\[ \text{dist}[v] \] is distance of \( v \) from the tree

\[ \text{pred} : \text{Array of Edge} \]  
\[ \text{pred}[v] \] is shortest edge between \( S \) and \( v \)

\[ q : \text{PriorityQueue of Node with dist[\cdot] as priority} \]

\[ \text{dist}[s] := 0; \quad \text{q.insert}(s) \] for any \( s \in V \)

for \( i := 1 \) to \( n - 1 \) do do

\[ u := \text{q.deleteMin()} \]  
\[ \text{new node for } S \]

\[ \text{dist}[u] := 0 \]

foreach \( (u, v) \in E \) do

if \( c((u, v)) < \text{dist}[v] \) then

\[ \text{dist}[v] := c((u, v)); \quad \text{pred}[v] := (u, v) \]

if \( v \in q \) then \( q.\text{decreaseKey}(v) \) else \( q.\text{insert}(v) \)

return \{ \text{pred}[v] : v \in V \setminus \{s\} \}
Graph Representation for Jarník-Prim

We need node → incident edges

+ fast (cache efficient)
+ more compact than linked lists
- difficult to change
- Edges are stored twice
Sanders: Algorithm Engineering

Analysis

□ $O(m+n)$ time outside priority queue

□ $n$ deleteMin (time $O(n \log n)$)

□ $O(m)$ decreaseKey (time $O(1)$ amortized)

$O(m+n \log n)$ using Fibonacci Heaps

practical implementation using simpler pairing heaps.
But analysis is still partly open!
Kruskal’s Algorithm [1956]

\[ T := 0 \]  \hspace{1cm} // \text{subforest of the MST}

\textbf{foreach} \ (u, v) \in E \ \text{in ascending order of weight} \ \textbf{do}

\hspace{1cm} \textbf{if} \ u \ \text{and} \ v \ \text{are in different subtrees of} \ T \ \textbf{then}

\hspace{2cm} T := T \cup \{(u, v)\} \hspace{1cm} // \text{Join two subtrees}

\textbf{return} \ T
The Union-Find Data Structure

Class `UnionFind(n : N)`  // Maintain a partition of 1..n

parent = [n + 1, ..., n + 1] : Array [1..n] of 1..n + ⌈log n⌉

Function `find(i : 1..n) : 1..n`

if parent[i] > n then return i
else
    i′ := find(parent[i])
    parent[i] := i′
return i′

Procedure `link(i, j : 1..n)`

assert i and j are leaders of different subsets
if parent[i] < parent[j] then parent[i] := j
else if parent[i] > parent[j] then parent[j] := i
else parent[j] := i; parent[i]++  // next generation

Procedure `union(i, j)` if find(i) ≠ find(j) then link(find(i), find(j))
Kruskal Using Union Find

\(T : \text{UnionFind}(n)\)

sort \(E\) in ascending order of weight

\text{kruskal}(E)

**Procedure** \(\text{kruskal}(E)\)

\[
\text{foreach } (u, v) \in E \text{ do}
\]

\[
u' := T.\text{find}(u)
\]

\[
v' := T.\text{find}(v)
\]

\[\text{if } u' \neq v' \text{ then}
\]

output \((u, v)\)

\[
T.\text{link}(u', v')
\]
Graph Representation for Kruskal

Just an edge sequence (array)!

- very fast (cache efficient)
- Edges are stored only once

~ more compact than adjacency array
Analysis

\[ O(\text{sort}(m) + m\alpha(m,n)) = O(m \log m) \] where \( \alpha \) is the inverse Ackermann function
Kruskal versus Jarník-Prim I

- Kruskal wins for very sparse graphs
- Prim seems to win for denser graphs
- Switching point is unclear
  - How is the input represented?
  - How many decreaseKeys are performed by JP? (average case: $n \log \frac{m}{n}$ [Noshita 85])
  - Experimental studies are quite old [?], use slow graph representation for both algs, and artificial inputs
Better Version For Dense Graphs?

Procedure quickKruskal \((E : \text{Sequence of Edge})\)

\[
\begin{align*}
\text{if } m &\leq \beta n \text{ then kruskal}(E) & \quad & \text{// for some constant } \beta \\
\text{else} & \\
& \text{pick a pivot } p \in E \\
& E_{\leq} := \langle e \in E : e \leq E \rangle & \quad & \text{// partitioning a la quicksort} \\
& E_{> :} := \langle e \in E : e > E \rangle \\
& \text{quickKruskal}(E_{\leq}) \\
& E'_{> :=} \text{ filter}(E_{>}) \\
& \text{quickKruskal}(E'_{>})
\end{align*}
\]

Function filter \((E)\)

\[
\text{return } \langle (u, v) \in E : \text{leader}[u] \neq \text{leader}[v] \rangle
\]
1.1 Attempted Average-Case Analysis

Assume different random edge weights, arbitrary graphs
Assume pivot $p$ has median weight
Let $T(m)$ denote the expected execution time for $m$ edges

$m \leq \beta n$: $O(n \log n)$

Partitioning, Filtering: $O(m + n)$

$m > \beta n$: $T(m) = \Omega(m) + T(m/2) + T(2n)$ [Chan 98]

Solves to $O \left( m + n \log(n) \cdot \log \frac{m}{n} \right) \leq O(m + n \log(n) \cdot \log \log n)$

Open Problem: I know of no graph family with $T(n) = \omega(m + n \log(n))$
Kruskal versus Jarník-Prim II

Things are even less clear.

Kruskal may be better even for dense graphs

Experiments would be interesting.

Even for artificial graphs.
1.2 Filtering by Sampling Rather Than Sorting

\[ R := \text{random sample of } r \text{ edges from } E \]
\[ F := \text{MST}(R) \]  \hspace{1cm} \text{// Wlog assume that } F \text{ spans } V
\[ L := \emptyset \]  \hspace{1cm} \text{// “light edges” with respect to } R
\textbf{foreach } e \in E \hspace{1cm} \text{// Filter}

\[ C := \text{the unique cycle in } \{e\} \cup F \]
\textbf{if } e \text{ is not heaviest in } C \textbf{ then}
\[ L := L \cup \{e\} \]
\textbf{return } \text{MST}((L \cup F))
1.2.1 Analysis

[Chan 98, KKK 95]

Observation: \( e \in L \) only if \( e \in \text{MST}(R \cup \{e\}) \).
(Otherwise \( e \) could replace some heavier edge in \( F \)).

**Lemma 1.** \( \mathbb{E}[|L \cup F|] \leq \frac{mn}{\rho} \)
MST Verification by Interval Maxima

- Number the nodes by the order they were added to the MST by Prim’s algorithm.
- $w_i =$ weight of the edge that inserted node $i$.
- Largest weight on path $(u,v) = \max\{w_j | u < j \leq v\}$. 

![Diagram of MST and interval maxima](image)
Interval Maxima

Preprocessing: build 

\[ n \log n \text{ size array PreSuf.} \]

To find \( \text{max} \ a[i..j] \):

- Find the level of the LCA: \( \ell = \lceil \log_2 (i \oplus j) \rceil \).
- Return \( \text{max}(\text{PreSuf}[\ell][i], \text{PreSuf}[\ell][j]) \).
- Example: \( 2 \oplus 6 = 010 \oplus 110 = 100 \): \( \ell = 2 \)
A Simple Filter Based Algorithm

Choose \( r = \sqrt{mn} \).

We get expected time
\[
T_{\text{Prim}}(\sqrt{mn}) + O(n \log n + m) + T_{\text{Prim}}(\frac{mn}{\sqrt{mn}}) = O(n \log n + m)
\]

The constant factor in front of the \( m \) is very small.
Results

10 000 nodes, SUN-Fire-15000, 900 MHz UltraSPARC-III+
Results

10 000 nodes, SUN-Fire-15000, 900 MHz UltraSPARC-III+
10,000 nodes, NEC SX-5
Vector Machine
“worst case”
Edge Contraction

Let \{u, v\} denote an MST edge.
Eliminate \( v \):

\[
\text{forall } (w, v) \in E \text{ do}
\]

\[
E := E \setminus (w, v) \cup \{(w, u)\} \quad // \text{ but remember orignal terminals}
\]
Boruvka’s Node Reduction Algorithm

For each node find the lightest incident edge. Include them into the MST (cut property) contract these edges,

Time $O(m)$

At least halves the number of remaining nodes
1.3 Simpler and Faster Node Reduction

for $i := n$ downto $n' + 1$ do

pick a random node $v$

find the lightest edge $(u, v)$ out of $v$ and output it

contract $(u, v)$

$E[\text{degree}(v)] \leq 2m/i$

$$
\sum_{n'<i\leq n} \frac{2m}{i} = 2m \left( \sum_{0<i\leq n} \frac{1}{i} - \sum_{0<i\leq n'} \frac{1}{i} \right) \approx 2m(\ln n - \ln n') = 2m \ln \frac{n}{n'}
$$

output {1,2}  

7 (was {2,3})

output {2,3}

2 (was {4,3})
1.4 Randomized Linear Time Algorithm

1. Factor 8 node reduction (3× Boruvka or sweep algorithm)
   \(O(m + n)\).

2. \(R \leftarrow m/2\) random edges.  \(O(m + n)\).

3. \(F \leftarrow MST(R)\) [Recursively].

4. Find light edges \(L\) (edge reduction).  \(O(m + n)\)
   \[E[|L|] \leq \frac{mn/8}{m/2} = n/4.\]

5. \(T \leftarrow MST(L \cup F)\) [Recursively].

\[T(n,m) \leq T(n/8, m/2) + T(n/8, n/4) + c(n + m)\]
\[T(n,m) \leq 2c(n + m)\] fulfills this recurrence.
1.5 External MSTs

Semiexternal Algorithms

Assume $n \leq M - 2B$:
run Kruskal’s algorithm using external sorting
Streaming MSTs

If $M$ is yet a bit larger we can even do it with $m/B$ I/Os:

$$T := \emptyset \quad \text{// current approximation of MST}$$

while there are any unprocessed edges do

load any $\Theta(M)$ unprocessed edges $E'$

$$T := \text{MST}(T \cup E') \quad \text{// for any internal MST alg.}$$

Corollary: we can do it with linear expected internal work

Disadvantages to Kruskal:
Slower in practice
Smaller max. $n$
General External MST

while $n > M - 2B$ do
    perform some node reduction
use semi-external Kruskal

Theory: $O(sort(m))$ expected I/Os by externalizing the linear time algorithm.
(i.e., node reduction + edge reduction)
External Implementation I: Sweeping

$\pi$: random permutation $V \rightarrow V$

sort edges $(u, v)$ by $\min(\pi(u), \pi(v))$

\textbf{for} $i := n \textbf{ downto } n' + 1 \textbf{ do}$

pick the node $v$ with $\pi(v) = i$

find the lightest edge $(u, v)$ out of $v$ and output it

contract $(u, v)$

Problem: how to implement relinking?
Relinking Using Priority Queues

Q: priority queue  // Order: max node, then min edge weight

\[ \text{foreach } (\{u, v\}, c) \in E \text{ do } \text{Q.insert}((\{\pi(u), \pi(v)\}, c, \{u, v\})) \]

current := n + 1

\[ \text{loop} \]

\[ (\{u, v\}, c, \{u_0, v_0\}) := \text{Q.deleteMin}() \]

\[ \text{if } \text{current} \neq \text{max } \{u, v\} \text{ then} \]

\[ \text{if } \text{current} = M + 1 \text{ then return} \]

\[ \text{output } \{u_0, v_0\}, c \]

\[ \text{current} := \text{max } \{u, v\} \]

\[ \text{connect} := \text{min } \{u, v\} \]

\[ \text{else } \text{Q.insert}((\{\text{min } \{u, v\}, \text{connect}\}, c, \{u_0, v_0\})) \]

\[ \approx \text{sort}(10m \ln \frac{n}{M}) \text{ I/Os with opt. priority queues} \] [Sanders 00]

Problem: Compute bound
Sweeping with linear internal work

- Assume $m = O\left(\frac{M^2}{B}\right)$
- $k = \Theta(M/B)$ external buckets with $n/k$ nodes each
- $M$ nodes for last “semiexternal” bucket
- split current bucket into internal buckets for each node

Sweeping:
Scan current internal bucket twice:

1. Find minimum
2. Relink

New external bucket: scan and put in internal buckets
Large degree nodes: move to semiexternal bucket
Experiments

Instances from “classical” MST study [Moret Shapiro 1994]

- sparse random graphs
- random geometric graphs
- grids
  \( O(\text{sort}(m)) \) I/Os
  for planar graphs by removing parallel edges!

Other instances are rather dense or designed to fool specific algorithms.
$m \approx 2n$

Graph showing the comparison of algorithms for different values of $m / 1,000,000$. The algorithms include Kruskal, Prim, random, geometric, and grid. The $y$-axis represents $t / m$ in microseconds, and the $x$-axis represents $m / 1,000,000$. The graph shows the performance of each algorithm as $m$ increases.
\[ m \approx 4n \]
$m \approx 8n$
MST Summary

- Edge reduction helps for very dense, “hard” graphs
- A fast and simple node reduction algorithm
  \( \sim 4 \times \) less I/Os than previous algorithms
- Refined semiexternal MST, use as base case
- Simple pseudo random permutations (no I/Os)
- A fast implementation
- Experiments with huge graphs (up to \( n = 4 \cdot 10^9 \) nodes)

External MST is feasible
Open Problems

- New experiments for (improved) Kruskal versus Jarník-Prim
- Realistic (huge) inputs
- Parallel external algorithms
- Implementations for other graph problems
Conclusions

- Even fundamental, “simple” algorithmic problems still raise interesting questions
- Implementation and experiments are important and were neglected by parts of the algorithms community
- **Theory** an (at least) equally important, essential component of the algorithm design process
More Algorithm Engineering on Graphs

- Count triangles in very large graphs. Interesting as a measure of clusteredness. (Cooperation with AG Wagner)

- External BFS (Master thesis Deepak Ajwani)

- Maximum flows: Is the theoretical best algorithm any good? (Jein)

- Approximate max. weighted matching (Studienarbeit Jens Maue)
Maximal Flows

Theory: $O(m \Lambda \log(n^2/m) \log U)$ binary blocking flow-algorithm mit $\Lambda = \min\{m^{1/2}, n^{2/3}\}$ [Goldberg-Rao-97].

Problem: best case $\approx$ worst case

[Hagerup Sanders Träff WAE 98]:

- Implementable generalization
- best case $\ll$ worst case
- best algorithms for some “difficult” instances
Ergebnis

- Einfach extern implementierbar
- \( n' = M \sim \text{semiexterner Kruskal Algorithmus} \)
- Insgesamt \( O\left(\text{sort}(m\ln \frac{n}{m})\right) \) erwartete I/Os
- Für realistische Eingaben mindestens \( 4 \times \) bisher als bisher bekannte Algorithmen
- Implementierung in \(<\text{stxxl}>\) mit bis zu \( 96 \text{ GByte} \) großen Graphen läuft „über Nacht“
Presenting Data from Experiments in Algorithmics

Restrictions

- black and white $\implies$ easy and cheap printing
- 2D (stay tuned)
- no animation
- no realism desired
Not here

☐ ensuring reproducibility

☐ describing the setup

☐ finding sources of measurement errors

☐ reducing measurement errors (averaging, median, unloaded machine . . . )

☐ measurements in the creative phase of experimental algorithmics.
The Starting Point

- (Several) Algorithm(s)

- A few quantities to be measured: time, space, solution quality, comparisons, cache faults, . . . There may also be measurement errors.

- An unlimited number of potential inputs. $\sim \rightarrow$ condense to a few characteristic ones (size, $|V|$, $|E|$, . . . or problem instances from applications)

Usually there is not a lack but an abundance of data $\neq$ many other sciences
The Process

Waterfall model?

1. Design
2. Measurement
3. Interpretation

Perhaps the paper should at least look like that.
The Process

- Eventually stop asking questions (Advisors/Referees listen!)
- Build measurement tools
- Automate (re)measurements
- Choice of Experiments driven by risk and opportunity

Distinguish mode

- **Exploratory**: many different parameter settings, interactive, short turnaround times
- **Consolidating**: many large instances, standardized measurement conditions, batch mode, many machines
Of Risks and Opportunities

Example: Hypothesis = my algorithm is the best

big risk: untried main competitor

small risk: tuning of a subroutine that takes 20% of the time.

big opportunity: use algorithm for a new application
  ↦ new input instances
Basic Principles

- Minimize nondata ink
  (form follows function, not a beauty contest,...)

- Letter size $\approx$ surrounding text

- Avoid clutter and overwhelming complexity

- Avoid boredom (too little data per $m^2$.

- Make the conclusions evident
Tables

+ easy
  - easy $\rightsquigarrow$ overuse
+ accurate values ($\neq 3D$)
+ more compact than bar chart
+ good for unrelated instances (e.g. solution quality)
  - boring
  - no visual processing

rule of thumb that “tables usually outperform a graph for small data sets of 20 numbers or less” [Tufte 83]

Curves in main paper, tables in appendix?
2D Figures

default: $x = \text{input size}, \ y = f(\text{execution time})$
Choose unit to eliminate a parameter?

Length $k$ fractional tree broadcasting. Latency $t_0 + k$
x Axis

logarithmic scale?

yes if x range is wide
$x$ Axis

logarithmic scale, powers of two (or $\sqrt{2}$)

with tic marks, (plus a few small ones)
gnuplot

set xlabel "N"
set ylabel "(time per operation)/log N  [ns]"
set xtics (256, 1024, 4096, 16384, 65536, "2^18" 262144)
set size 0.66, 0.33
set logscale x 2
set data style linespoints
set key left
set terminal postscript portrait enhanced 10
set output "r10000timenew.eps"
plot [1024:10000000][0:220] \n  "h2r10000new.log" using 1:3 title "bottom up binary heap" ,
  "h4r10000new.log" using 1:3 title "bottom up aligned 4-ary" ,
  "knr10000new.log" using 1:3 title "sequence heap" with linespoints
## Data File

<table>
<thead>
<tr>
<th>Size</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>703.125</td>
<td>87.8906</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>729.167</td>
<td>81.0185</td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td>768.229</td>
<td>76.8229</td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td>830.078</td>
<td>75.4616</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>846.354</td>
<td>70.5295</td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td>878.906</td>
<td>67.6082</td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td>915.527</td>
<td>65.3948</td>
<td></td>
</tr>
<tr>
<td>32768</td>
<td>925.7</td>
<td>61.7133</td>
<td></td>
</tr>
<tr>
<td>65536</td>
<td>946.045</td>
<td>59.1278</td>
<td></td>
</tr>
<tr>
<td>131072</td>
<td>971.476</td>
<td>57.1457</td>
<td></td>
</tr>
<tr>
<td>262144</td>
<td>1009.62</td>
<td>56.0902</td>
<td></td>
</tr>
<tr>
<td>524288</td>
<td>1035.69</td>
<td>54.51</td>
<td></td>
</tr>
<tr>
<td>1048576</td>
<td>1055.08</td>
<td>52.7541</td>
<td></td>
</tr>
<tr>
<td>2097152</td>
<td>1113.73</td>
<td>53.0349</td>
<td></td>
</tr>
<tr>
<td>4194304</td>
<td>1150.29</td>
<td>52.2859</td>
<td></td>
</tr>
<tr>
<td>8388608</td>
<td>1172.62</td>
<td>50.9836</td>
<td></td>
</tr>
</tbody>
</table>
$x$ Axis

linear scale for ratios or small ranges (#processor,...)
\( x \) Axis

An exotic scale: arrival rate \( 1 - \varepsilon \) of saturation point
y Axis

Avoid log scale! Scale such that theory gives \( \sim \) horizontal lines

but give easy interpretation of the scaling function
**y Axis**

Give units

![Graph showing performance comparison between different heap structures. The x-axis represents the size of the heap (N), and the y-axis represents the time complexity (T(deleteMin) + T(insert))/log N in nanoseconds. Different heap structures are compared, including bottom up binary heap, bottom up aligned 4-ary heap, and sequence heap.]
**y Axis**

start from 0 if this does not waste too much space

You may assume readers to be out of Kindergarten
**y Axis**

clip outclassed algorithms
y Axis

vertical size: weighted average of the slants of the line segments in the figure should be about $45^\circ$ [Cleveland 94]
y Axis

graph a bit wider than high, e.g., golden ratio [Tufte 83]
Multiple Curves

+ high information density
+ better than 3D (reading off values)
  - Easily overdone

≤ 7 smooth curves
Reducing the Number of Curves

use ratios

\[
\text{improvement } \min \left( \frac{T_1^* + T_\infty^*}{T^*} \right)
\]
Reducing the Number of Curves

omit curves

- outclassed algorithms (for case shown)
- equivalent algorithms (for case shown)
Reducing the Number of Curves

split into two graphs
Reducing the Number of Curves

split into two graphs

average delay vs. \(1/\varepsilon\)
Keeping Curves apart: log y scale

The diagram shows the time per edge in nanoseconds (ns) for different algorithms and edge densities. The algorithms compared include Prim (Scalar), Prim (Vectorized), I-Max (Scalar), and I-Max (Vectorized). The y-axis represents the time per edge, while the x-axis shows the edge density.
Keeping Curves apart: smoothing

- Nonredundant
- Mirror
- Ring shortest queue
- Ring with matching
- Shortest queue

average delay

$1/\varepsilon$
same order as curves
Sanders: Algorithm Engineering

Keys

place in white space

consistent in different figures
Todsünden

1. forget explaining the axes
2. connecting unrelated points by lines
3. mindless use/overinterpretation of double-log plot
4. cryptic abbreviations
5. microscopic lettering
6. excessive complexity
7. pie charts
Arranging Instances

- bar charts
- stack components of execution time
- careful with shading

![Bar Chart](image)

- preprocessing
- phase 1
- phase 2
- postprocessing
Arranging Instances

scatter plots
Measurements and Connections

- straight line between points do not imply claim of linear interpolation
- different with higher order curves
- no points imply an even stronger claim. Good for very dense smooth measurements.
Grids and Ticks

- Avoid grids or make it light gray
- Usually round numbers for tic marks!
- Sometimes plot important values on the axis

Usually avoidable for randomized algorithms. Median $\neq$ average,...
errors may not be of statistical nature!
3D

- you cannot read off absolute values
- interesting parts may be hidden
- only one surface
- good impression of shape
Caption

what  is displayed

how  has the date been obtained

surrounding text has more.
Check List

☐ Should the experimental setup from the exploratory phase be redesigned to increase conciseness or accuracy?

☐ What parameters should be varied? What variables should be measured? How are parameters chosen that cannot be varied?

☐ Can tables be converted into curves, bar charts, scatter plots or any other useful graphics?

☐ Should tables be added in an appendix or on a web page?

☐ Should a 3D-plot be replaced by collections of 2D-curves?

☐ Can we reduce the number of curves to be displayed?

☐ How many figures are needed?
- Scale the $x$-axis to make $y$-values independent of some parameters?
- Should the $x$-axis have a logarithmic scale? If so, do the $x$-values used for measuring have the same basis as the tick marks?
- Should the $x$-axis be transformed to magnify interesting subranges?
- Is the range of $x$-values adequate?
- Do we have measurements for the right $x$-values, i.e., nowhere too dense or too sparse?
- Should the $y$-axis be transformed to make the interesting part of the data more visible?
- Should the $y$-axis have a logarithmic scale?
□ Is it be misleading to start the $y$-range at the smallest measured value?

□ Clip the range of $y$-values to exclude useless parts of curves?

□ Can we use banking to 45°?

□ Are all curves sufficiently well separated?

□ Can noise be reduced using more accurate measurements?

□ Are error bars needed? If so, what should they indicate? Remember that measurement errors are usually not random variables.

□ Use points to indicate for which $x$-values actual data is available.

□ Connect points belonging to the same curve.
Only use splines for connecting points if interpolation is sensible.

Do not connect points belonging to unrelated problem instances.

Use different point and line styles for different curves.

Use the same styles for corresponding curves in different graphs.

Place labels defining point and line styles in the right order and without concealing the curves.

Captions should make figures self contained.

Give enough information to make experiments reproducible.