High Performance Computing

Network Topology

Martin Raum

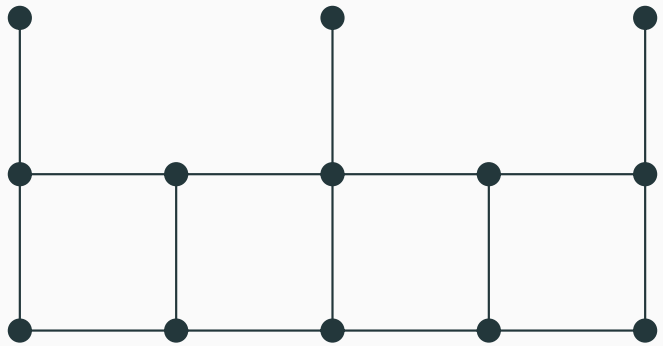
Nodes and their connection

Latency and bandwidth, as in memory access, impact communication among nodes of a distributed computation network.

The “cloud” has usually no planed or definable structure, but dedicated HPC systems often do: This is reflected in the network topology.

Encoding network topology

Network topology is usually encoded as a graph, in which vertices stand for compute nodes and edges between two nodes represent a connection.



Neighbors and proximity

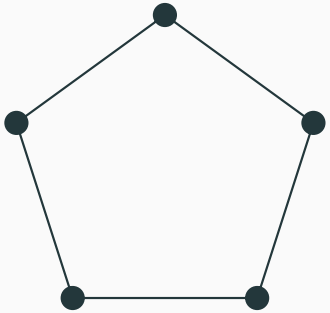
Classical network topology does not encode connections of different quality. The usual graph distance encodes all information required to optimize communication.

Some modern computation networks are not homogeneous. Their topologies are more accurately modeled by metrized graphs, in which edges have varying lengths (or weights).

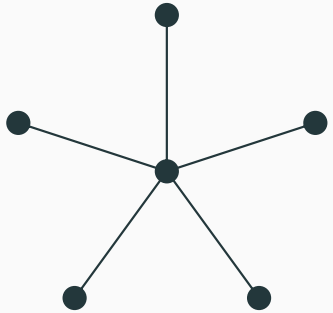
Archetypes of network topologies

Line

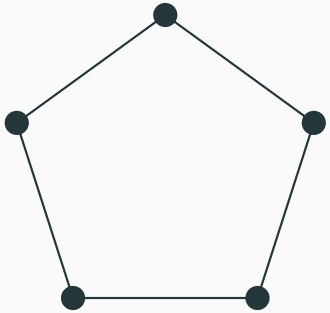
Ring



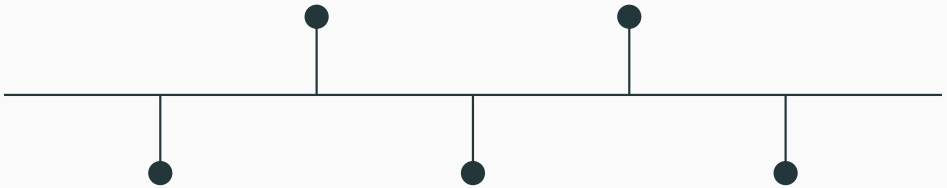
Star



Fully connected

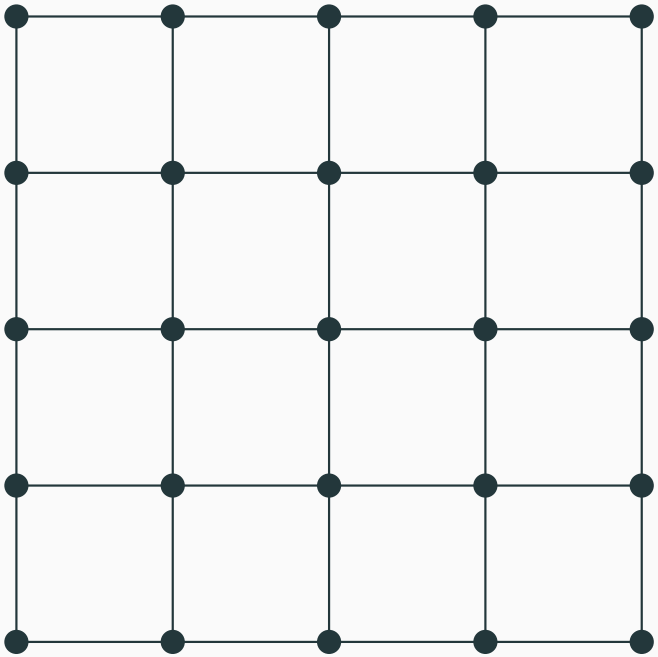


Bus

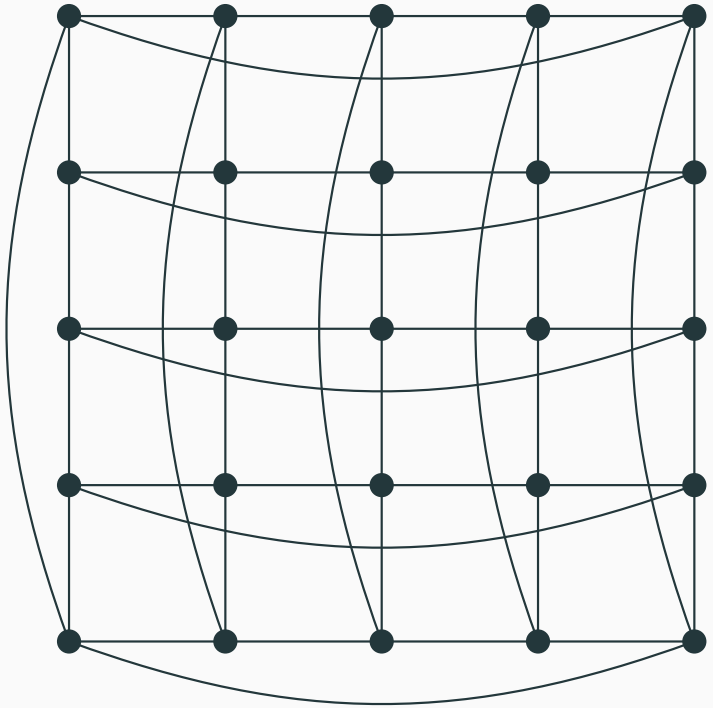


Archetypes of network topologies

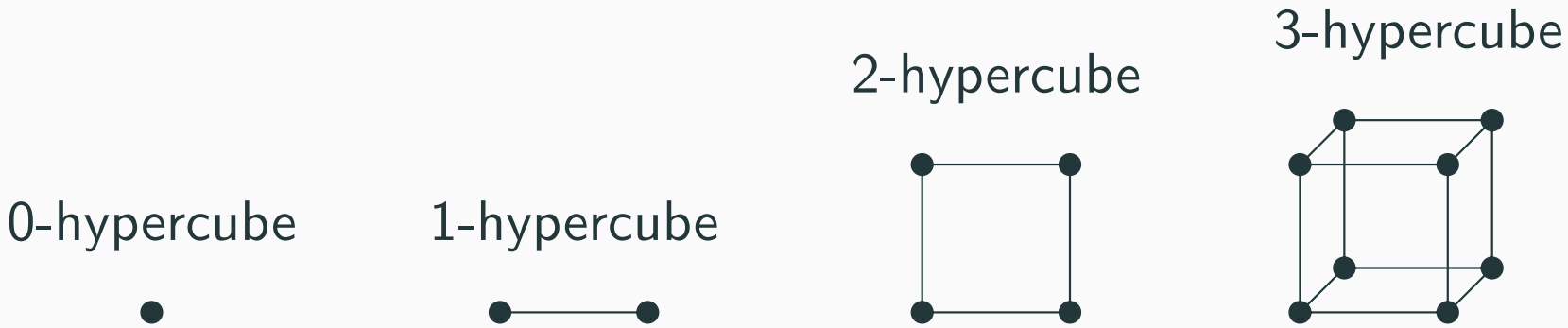
Regular mesh



Torus



Archetypes of network topologies



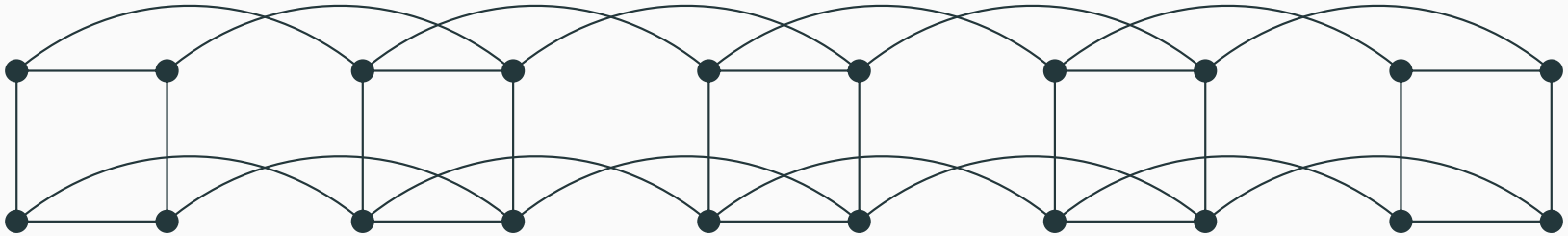
n-hypercube have maximal distance n and 2^n nodes.

Maximal distance growth by the logarithm of the number of nodes.

Archetypes of network topologies

Using network hubs with $m+n$ connections each, a typical network topology can be a set of m-hypercubes that appear as vertices in another network topology.

2-hypercube arranged on a line



MPI Communication respecting topology

MPI-3.0 introduces sparse communication patterns:

MPI_Neighbor_allgather, MPI_Neighbor_allgatherv.

MPI_Neighbor_alltoall, MPI_Neighbor_alltoallv.

These work like the corresponding standard functions, but communication between two nodes n and n' only happens if there is a direct connection between n and n'.

They can be used to create communicators and then use standard directives.