

= Directed acyclic graph =

In mathematics and computer science, a directed acyclic graph ( DAG / ?dæ? / ), is a finite directed graph with no directed cycles. That is, it consists of finitely many vertices and edges, with each edge directed from one vertex to another, such that there is no way to start at any vertex  $v$  and follow a consistently @-@ directed sequence of edges that eventually loops back to  $v$  again. Equivalently, a DAG is a directed graph that has a topological ordering, a sequence of the vertices such that every edge is directed from earlier to later in the sequence.

DAGs may be used to model many different kinds of information. The reachability relation in a DAG forms a partial order, and any finite partial order may be represented by a DAG using reachability. The program evaluation and review technique uses DAGs to model the milestones and activities of large human projects, and schedule these projects to use as little total time as possible. Scheduling of computer operations such as the updates to a spreadsheet or the compilation operations in a makefile can also be performed using topological orderings of DAGs. Combinational logic blocks in electronic circuit design, and the operations in dataflow programming languages, involve acyclic networks of processing elements. DAGs can also represent collections of events and their influence on each other, either in a probabilistic structure such as a Bayesian network or as a record of historical data such as family trees or the version histories of distributed revision control systems. DAGs can also be used as a compact representation of sequence data, such as the directed acyclic word graph representation of a collection of strings, or the binary decision diagram representation of sequences of binary choices.

Important polynomial time computational problems on DAGs include topological sorting ( finding a topological ordering ), construction of the transitive closure and transitive reduction ( the largest and smallest DAGs with the same reachability relation, respectively ), and the closure problem, in which the goal is to find a minimum @-@ weight subset of vertices with no edges connecting them to the rest of the graph. Transforming a directed graph with cycles into a DAG by deleting as few vertices or edges as possible ( the feedback vertex set and feedback edge set problem, respectively ) is NP @-@ hard, but any directed graph can be made into a DAG ( its condensation ) by contracting each strongly connected component into a single supervertex. The problems of finding shortest paths and longest paths can be solved on DAGs in linear time, in contrast to arbitrary graphs for which shortest path algorithms are slower and longest path problems are NP @-@ hard.

The corresponding concept for undirected graphs is a forest, an undirected graph without cycles. Choosing an orientation for a forest produces a special kind of directed acyclic graph called a polytree. However there are many other kinds of directed acyclic graph that are not formed by orienting the edges of an undirected acyclic graph. Moreover, every undirected graph has an acyclic orientation, an assignment of a direction for its edges that makes it into a directed acyclic graph. To emphasize that DAGs are not the same thing as directed versions of undirected acyclic graphs, some authors call them acyclic directed graphs or acyclic digraphs.

= = Definitions = =

A graph is formed by a collection of vertices and edges, where the vertices are structureless objects that are connected in pairs by edges. In the case of a directed graph, each edge has an orientation, from one vertex to another vertex. A path in a directed graph can be described by a sequence of edges having the property that the ending vertex of each edge in the sequence is the same as the starting vertex of the next edge in the sequence; a path forms a cycle if the starting vertex of its first edge equals the ending vertex of its last edge. A directed acyclic graph is a directed graph that has no cycles.

A vertex  $v$  of a directed graph is said to be reachable from another vertex  $u$  when there exists a path that starts at  $u$  and ends at  $v$ . As a special case, every vertex is considered to be reachable from itself ( by a path with zero edges ). If a vertex can reach itself via a nontrivial path ( a path with one or more edges ), then that path is a cycle, so another way to define directed acyclic graphs is

that they are the graphs in which no vertex can reach itself via a nontrivial path .

A topological ordering of a directed graph is an ordering of its vertices into a sequence , such that for every edge the start vertex of the edge occurs earlier in the sequence than the ending vertex of the edge . A graph that has a topological ordering cannot have any cycles , because the edge into the earliest vertex of a cycle would have to be oriented the wrong way . Therefore , every graph with a topological ordering is acyclic . Conversely , every directed acyclic graph has a topological ordering . Therefore , this property can be used as an alternative definition of the directed acyclic graphs : they are exactly the graphs that have topological orderings .

= = Mathematical properties = =

= = Reachability , transitive closure , and transitive reduction = = =

The reachability relationship in any directed acyclic graph can be formalized as a partial order  $\preceq$  on the vertices of the DAG . In this partial order , two vertices  $u$  and  $v$  are ordered as  $u \preceq v$  exactly when there exists a directed path from  $u$  to  $v$  in the DAG ; that is , when  $v$  is reachable from  $u$  . However , different DAGs may give rise to the same reachability relation and the same partial order . For example , the DAG with two edges  $a \preceq b$  and  $b \preceq c$  has the same reachability relation as the graph with three edges  $a \preceq b$  ,  $b \preceq c$  , and  $a \preceq c$  . Both of these DAGS produce the same partial order , in which the vertices are ordered as  $a \preceq b \preceq c$  .

If  $G$  is a DAG , its transitive closure is the graph with the most edges that represents the same reachability relation . It has an edge  $u \preceq v$  whenever  $u$  can reach  $v$  . That is , it has an edge for every related pair  $u \preceq v$  of distinct elements in the reachability relation of  $G$  , and may therefore be thought of as a direct translation of the reachability relation  $\preceq$  into graph @-@ theoretic terms . The same method of translating partial orders into DAGs works more generally : for every finite partially ordered set  $(S, \preceq)$  , the graph that has a vertex for each member of  $S$  and an edge for each pair of elements related by  $u \preceq v$  is automatically a transitively closed DAG , and has  $(S, \preceq)$  as its reachability relation . In this way , every finite partially ordered set can be represented as the reachability relation of a DAG .

The transitive reduction of a DAG  $G$  is the graph with the fewest edges that represents the same reachability relation as  $G$  . It is a subgraph of  $G$  , formed by discarding the edges  $u \preceq v$  for which  $G$  also contains a longer path connecting the same two vertices . Like the transitive closure , the transitive reduction is uniquely defined for DAGs . In contrast , for a directed graph that is not acyclic , there can be more than one minimal subgraph with the same reachability relation .

If a DAG  $G$  has a reachability relation described by the partial order  $\preceq$  , then the transitive reduction of  $G$  is a subgraph of  $G$  that has an edge  $u \preceq v$  for every pair in the covering relation of  $\preceq$  . Transitive reductions are useful in visualizing the partial orders they represent , because they have fewer edges than other graphs representing the same orders and therefore lead to simpler graph drawings . A Hasse diagram of a partial order is a drawing of the transitive reduction in which the orientation of each edge is shown by placing the starting vertex of the edge in a lower position than its ending vertex .

= = Topological ordering = = =

Every directed acyclic graph has a topological ordering , an ordering of the vertices such that the starting endpoint of every edge occurs earlier in the ordering than the ending endpoint of the edge . The existence of such an ordering can be used to characterize DAGs : a directed graph is a DAG if and only if it has a topological ordering . In general , this ordering is not unique ; a DAG has a unique topological ordering if and only if it has a directed path containing all the vertices , in which case the ordering is the same as the order in which the vertices appear in the path .

The family of topological orderings of a DAG is the same as the family of linear extensions of the reachability relation for the DAG , so any two graphs representing the same partial order have the

same set of topological orders .

== Combinatorial enumeration ==

The graph enumeration problem of counting directed acyclic graphs was studied by Robinson ( 1973 ) . The number of DAGs on  $n$  labeled vertices , for  $n = 0, 1, 2, 3, \dots$  ( without restrictions on the order in which these numbers appear in a topological ordering of the DAG ) is

1, 1, 3, 25, 543, 29281, 3781503, ... ( sequence A003024 in the OEIS ) .

These numbers may be computed by the recurrence relation

<formula>

Eric W. Weisstein conjectured , and McKay et al . ( 2004 ) proved , that the same numbers count the  $(0,1)$  matrices for which all eigenvalues are positive real numbers . The proof is bijective : a matrix  $A$  is an adjacency matrix of a DAG if and only if  $A + I$  is a  $(0,1)$  matrix with all eigenvalues positive , where  $I$  denotes the identity matrix . Because a DAG cannot have self-loops , its adjacency matrix must have a zero diagonal , so adding  $I$  preserves the property that all matrix coefficients are 0 or 1 .

== Related families of graphs ==

A polytree is a directed graph formed by orienting the edges of a free tree . Every polytree is a DAG . In particular , this is true of the arborescences formed by directing all edges outwards from the roots of a tree .

A multitree ( also called a strongly unambiguous graph or a mangrove ) is a directed graph in which there is at most one directed path ( in either direction ) between any two vertices ; equivalently , it is a DAG in which , for every vertex  $v$  , the subgraph reachable from  $v$  forms a tree .

== Computational problems ==

== Topological sorting and recognition ==

Topological sorting is the algorithmic problem of finding a topological ordering of a given DAG . It can be solved in linear time . Kahn 's algorithm for topological sorting builds the vertex ordering directly . It maintains a list of vertices that have no incoming edges from other vertices that have not already been included in the partially constructed topological ordering ; initially this list consists of the vertices with no incoming edges at all . Then , it repeatedly adds one vertex from this list to the end of the partially constructed topological ordering , and checks whether its neighbors should be added to the list . The algorithm terminates when all vertices have been processed in this way . Alternatively , a topological ordering may be constructed by reversing a postorder numbering of a depth-first search graph traversal .

It is also possible to check whether a given directed graph is a DAG in linear time , either by attempting to find a topological ordering and then testing for each edge whether the resulting ordering is valid or alternatively , for some topological sorting algorithms , by verifying that the algorithm successfully orders all the vertices without meeting an error condition .

== Construction from cyclic graphs ==

Any undirected graph may be made into a DAG by choosing a total order for its vertices and directing every edge from the earlier endpoint in the order to the later endpoint . The resulting orientation of the edges is called an acyclic orientation . Different total orders may lead to the same acyclic orientation , so an  $n$ -vertex graph can have fewer than  $n!$  acyclic orientations . The number of acyclic orientations is equal to  $|\chi(G)|$  , where  $\chi$  is the chromatic polynomial of the given graph .

Any directed graph may be made into a DAG by removing a feedback vertex set or a feedback arc set, a set of vertices or edges ( respectively ) that touches all cycles . However , the smallest such set is NP @-@ hard to find . An arbitrary directed graph may also be transformed into a DAG , called its condensation , by contracting each of its strongly connected components into a single supervertex . When the graph is already acyclic , its smallest feedback vertex sets and feedback arc sets are empty , and its condensation is the graph itself .

== Transitive closure and transitive reduction ==

The transitive closure of a given DAG , with  $n$  vertices and  $m$  edges , may be constructed in time  $O(mn)$  by using either breadth @-@ first search or depth @-@ first search to test reachability from each vertex . Alternatively , it can be solved in time  $O(n^{\omega})$  where  $\omega < 2.373$  is the exponent for fast matrix multiplication algorithms ; this is a theoretical improvement over the  $O(mn)$  bound for dense graphs .

In all of these transitive closure algorithms , it is possible to distinguish pairs of vertices that are reachable by at least one path of length two or more from pairs that can only be connected by a length @-@ one path . The transitive reduction consists of the edges that form length @-@ one paths that are the only paths connecting their endpoints . Therefore , the transitive reduction can be constructed in the same asymptotic time bounds as the transitive closure .

== Closure problem ==

The closure problem takes as input a directed acyclic graph with weights on its vertices and seeks the minimum ( or maximum ) weight of a closure , a set of vertices with no outgoing edges . ( The problem may be formulated for directed graphs without the assumption of acyclicity , but with no greater generality , because in this case it is equivalent to the same problem on the condensation of the graph . ) It may be solved in polynomial time using a reduction to the maximum flow problem .

== Path algorithms ==

Some algorithms become simpler when used on DAGs instead of general graphs , based on the principle of topological ordering . For example , it is possible to find shortest paths and longest paths from a given starting vertex in DAGs in linear time by processing the vertices in a topological order , and calculating the path length for each vertex to be the minimum or maximum length obtained via any of its incoming edges . In contrast , for arbitrary graphs the shortest path may require slower algorithms such as Dijkstra 's algorithm or the Bellman ? Ford algorithm , and longest paths in arbitrary graphs are NP @-@ hard to find .

== Applications ==

== Scheduling ==

Directed acyclic graphs representations of partial orderings have many applications in scheduling for systems of tasks with ordering constraints . An important class of problems of this type concern collections of objects that need to be updated , such as the cells of a spreadsheet after one of the cells has been changed , or the object files of a piece of computer software after its source code has been changed . In this context , a dependency graph is a graph that has a vertex for each object to be updated , and an edge connecting two objects whenever one of them needs to be updated earlier than the other . A cycle in this graph is called a circular dependency , and is generally not allowed , because there would be no way to consistently schedule the tasks involved in the cycle . Dependency graphs without circular dependencies form DAGs .

For instance , when one cell of a spreadsheet changes , it is necessary to recalculate the values of

other cells that depend directly or indirectly on the changed cell . For this problem , the tasks to be scheduled are the recalculations of the values of individual cells of the spreadsheet . Dependencies arise when an expression in one cell uses a value from another cell . In such a case , the value that is used must be recalculated earlier than the expression that uses it . Topologically ordering the dependency graph , and using this topological order to schedule the cell updates , allows the whole spreadsheet to be updated with only a single evaluation per cell . Similar problems of task ordering arise in makefiles for program compilation and instruction scheduling for low @-@ level computer program optimization .

A somewhat different DAG @-@ based formulation of scheduling constraints is used by the program evaluation and review technique ( PERT ) , a method for management of large human projects that was one of the first applications of DAGs . In this method , the vertices of a DAG represent milestones of a project rather than specific tasks to be performed . Instead , a task or activity is represented by an edge of a DAG , connecting two milestones that mark the beginning and completion of the task . Each such edge is labeled with an estimate for the amount of time that it will take a team of workers to perform the task . The longest path in this DAG represents the critical path of the project , the one that controls the total time for the project . Individual milestones can be scheduled according to the lengths of the longest paths ending at their vertices .

= = = Data processing networks = = =

A directed acyclic graph may be used to represent a network of processing elements . In this representation , data enters a processing element through its incoming edges and leaves the element through its outgoing edges .

For instance , in electronic circuit design , static combinational logic blocks can be represented as an acyclic system of logic gates that computes a function of an input , where the input and output of the function are represented as individual bits . In general , the output of these blocks cannot be used as the input unless it is captured by a register or state element which maintains its acyclic properties . Electronic circuit schematics either on paper or in a database are a form of directed acyclic graphs using instances or components to form a directed reference to a lower level component . Electronic circuits themselves are not necessarily acyclic or directed .

Dataflow programming languages describe systems of operations on data streams , and the connections between the outputs of some operations and the inputs of others . These languages can be convenient for describing repetitive data processing tasks , in which the same acyclically @-@ connected collection of operations is applied to many data items . They can be executed as a parallel algorithm in which each operation is performed by a parallel process as soon as another set of inputs becomes available to it .

In compilers , straight line code ( that is , sequences of statements without loops or conditional branches ) may be represented by a DAG describing the inputs and outputs of each of the arithmetic operations performed within the code . This representation allows the compiler to perform common subexpression elimination efficiently .

= = = Causal structures = = =

Graphs that have vertices representing events , and edges representing causal relations between events , are often acyclic ? arranging the vertices in linear order of time , all arrows point in the same direction as time , from parent to child ( due to causality affecting the future , not the past ) , and thus have no loops .

For instance , a Bayesian network represents a system of probabilistic events as vertices in a directed acyclic graph , in which the likelihood of an event may be calculated from the likelihoods of its predecessors in the DAG . In this context , the moral graph of a DAG is the undirected graph created by adding an ( undirected ) edge between all parents of the same vertex ( sometimes called marrying ) , and then replacing all directed edges by undirected edges .

Another type of graph with a similar causal structure is an influence diagram , the vertices of which

represent either decisions to be made or unknown information , and the edges of which represent causal influences from one vertex to another . In epidemiology , for instance , these diagrams are often used to estimate the expected value of different choices for intervention .

== Genealogy and version history ==

Family trees may be seen as directed acyclic graphs , with a vertex for each family member and an edge for each parent @-@ child relationship . Despite the name , these graphs are not necessarily trees because of the possibility of marriages between relatives ( so a child has a common ancestor on both the mother 's and father 's side ) causing pedigree collapse . The graphs of matrilineal descent ( " mother " relationships between women ) and patrilineal descent ( " father " relationships between men ) are trees within this graph . Because no one can become their own ancestor , family trees are acyclic .

For the same reason , the version history of a distributed revision control system generally has the structure of a directed acyclic graph , in which there is a vertex for each revision and an edge connecting pairs of revisions that were directly derived from each other . These are not trees in general due to merges .

In many randomized algorithms in computational geometry , the algorithm maintains a history DAG representing the version history of a geometric structure over the course of a sequence of changes to the structure . For instance in a randomized incremental algorithm for Delaunay triangulation , the triangulation changes by replacing one triangle by three smaller triangles when each point is added , and by " flip " operations that replace pairs of triangles by a different pair of triangles . The history DAG for this algorithm has a vertex for each triangle constructed as part of the algorithm , and edges from each triangle to the two or three other triangles that replace it . This structure allows point location queries to be answered efficiently : to find the location of a query point q in the Delaunay triangulation , follow a path in the history DAG , at each step moving to the replacement triangle that contains q . The final triangle reached in this path must be the Delaunay triangle that contains q .

== Data compression ==

Directed acyclic graphs may also be used as a compact representation of a collection of sequences . In this type of application , one finds a DAG in which the paths form the given sequences . When many of the sequences share the same subsequences , these shared subsequences can be represented by a shared part of the DAG , allowing the representation to use less space than it would take to list out all of the sequences separately . For example , the directed acyclic word graph is a data structure in computer science formed by a directed acyclic graph with a single source and with edges labeled by letters or symbols ; the paths from the source to the sinks in this graph represent a set of strings , such as English words . Any set of sequences can be represented as paths in a tree , by forming a tree vertex for every prefix of a sequence and making the parent of one of these vertices represent the sequence with one fewer element ; the tree formed in this way for a set of strings is called a trie . A directed acyclic word graph saves space over a trie by allowing paths to diverge and rejoin , so that a set of words with the same possible suffixes can be represented by a single tree vertex .

The same idea of using a DAG to represent a family of paths occurs in the binary decision diagram , a DAG @-@ based data structure for representing binary functions . In a binary decision diagram , each non @-@ sink vertex is labeled by the name of a binary variable , and each sink and each edge is labeled by a 0 or 1 . The function value for any truth assignment to the variables is the value at the sink found by following a path , starting from the single source vertex , that at each non @-@ sink vertex follows the outgoing edge labeled with the value of that vertex 's variable . Just as directed acyclic word graphs can be viewed as a compressed form of tries , binary decision diagrams can be viewed as compressed forms of decision trees that save space by allowing paths to rejoin when they agree on the results of all remaining decisions .

