Paths and isomorphism

Introduction

Many problems can be modeled with paths formed by traveling along the edges of graphs.

For instance, the problem of determining whether a message can be sent between 2 computers using intermediate links can be studied with a graph model.

Problems of efficiently planning routes for mail delivery, garbage pickup, diagnostics in computer networks, and so on can be solved using models that involve paths in graphs.

Paths

Informally, a path (путь) is a sequence of edges that begins at a vertex of a graph and travels from vertex to vertex along edges of the graph.

As the path travels along its edges, it visits the vertices along this path, that is, the endpoints of these edges.

DEFINITION 1(Path in an undirected graph)

Let n be a nonnegative integer and G an undirected graph.

A path of length n (путь длины n) from u to v in G is a sequence of n edges $e_1, \ldots, e_n$ of G for which there exists a sequence of vertices $x_0 = u, x_1, \ldots, x_{n-1}, x_n = v$ such that e_i has, for $i = 1, \ldots, n$, the endpoints x_{i-1} and x_i .

When the graph is simple, we denote this path by its vertex sequence $x_0, x_1, \ldots, x_n$ (because listing these vertices uniquely determines the path).

The path is a circuit (цикл) if it begins and ends at the same vertex, that is, if $u = v$, and has length greater than zero.

The path or circuit is said to pass through the vertices $x_1, x_2, \ldots, x_{n-1}$ or traverse the edges $e_1, e_2, \ldots, e_n$.

A path or circuit is simple (простой путь, простой цикл) if it does not contain the same edge more than once.

When it is not necessary to distinguish between multiple edges, we will denote a path $e_1, e_2, \dots, e_n$, where e_i is associated with $\{x_{i-1}, x_i\}$ for i = 1, 2,...,n by its vertex sequence $x_0, x_1, \dots, x_n$.

This notation identifies a path only as far as which vertices it passes through.

Consequently, it does not specify a unique path when there is >1 path that passes through this sequence of vertices, which will happen ⇔ there are multiple edges between some successive vertices in the list.

Note that a path of length 0 consists of a single vertex.

Remark: There is considerable variation of terminology concerning the concepts defined in Definition 1.

For instance, in some books, the term walk (маршрут) is used instead of path,

where a walk is defined to be an alternating sequence of vertices and edges of a graph,

$v_0, e_1, v_1, e_2, \ldots, v_{n-1}, e_n, v_n$, where v_{i-1} and v_i are the endpoints of e_i for i = 1, 2, ..., n.

When this terminology is used,

closed walk (замкнутый маршрут) is used instead of circuit to indicate a walk that begins and ends at the same vertex,

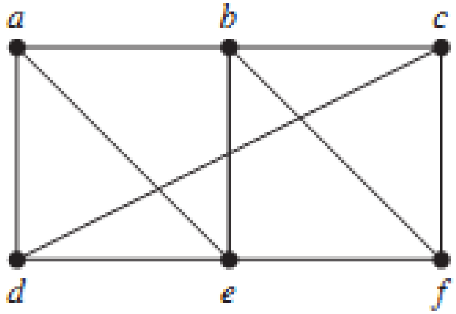
and trail (след, маршрут) is used to denote a walk that has no repeated edge (replacing the term simple path).

EXAMPLE 1 In the simple graph shown bellow, *a, d, c, f , e* is a simple path of length 4, because {*a, d*}, {*d, c*}, {*c, f* }, and {*f, e*} are all edges.

However, *d, e, c, a* is not a path, because {*e, c*} is not an edge.

Note that *b, c, f , e, b* is a circuit of length 4 because {*b, c*}, {*c, f* }, {*f, e*}, and {*e, b*} are edges, and this path begins and ends at *b*.

The path *a, b, e, d, a, b*, which is of length 5, is not simple because it contains the edge {*a, b*} twice.



A Simple Graph.

DEFINITION 2 (Path in an directed graph)

Let n be a nonnegative integer and G is a directed graph.

A path (путь) of length n from u to v in G is a sequence of edges $e_1, e_2, \ldots, e_n$ of G such that e_1 is associated with (x_0, x_1), e_2 is associated with (x_1, x_2), and so on, with e_n associated with (x_{n-1}, x_n), where $x_0 = u$ and $x_n = v$.

When there are no multiple edges in the directed graph, this path is denoted by its vertex sequence $x_0, x_1, x_2, \ldots, x_n$.

A path of length > 0 that begins and ends at the same vertex is called a circuit or cycle (цикл).

A path or circuit is called simple if it does not contain the same edge more than once.

Remark: Terminology other than that given in Definition 2 is often used for the concepts defined there.

In particular, the alternative terminology that uses walk, closed walk, trail, and path be used for directed graphs.

Note that the terminal vertex of an edge in a path is the initial vertex of the next edge in the path.

When it is not necessary to distinguish between multiple edges, we will denote a path $e_1, e_2, \ldots, e_n$, where e_i is associated with (x_{i-1}, x_i) for $i = 1, 2, \ldots, n$, by its vertex sequence $x_0, x_1, \ldots, x_n$.

The notation identifies a path only as far as which the vertices it passes through.

There may be >1 path that passes through this sequence of vertices, which will happen <=> there are multiple edges between 2 successive vertices in the list.

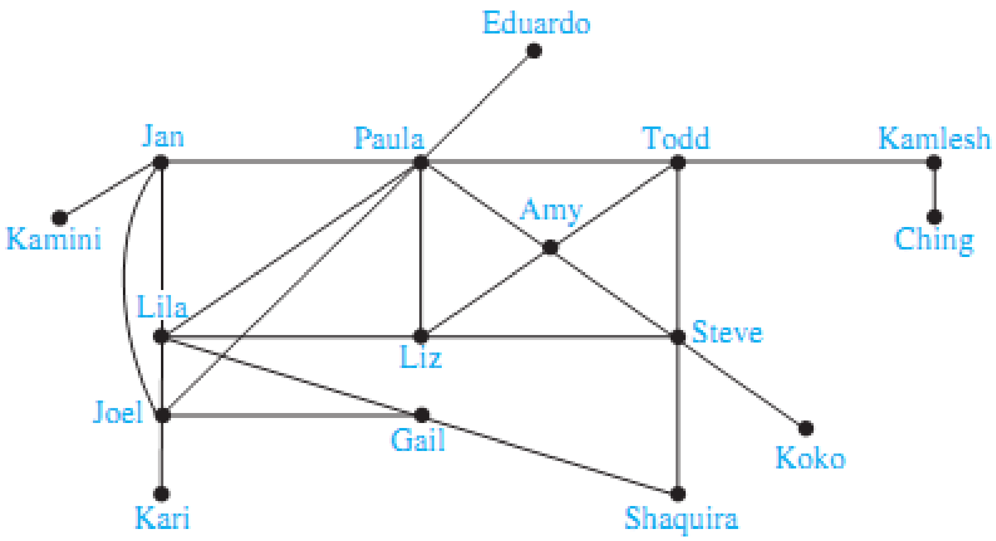
EXAMPLE Paths in Acquaintanceship Graphs

In an acquaintanceship graph there is a path between two people if there is a chain of people linking these people, where two people adjacent in the chain know one another.

For example, in Figure bellow, there is a chain of 6 people linking Kamini and Ching.

Many social scientists have conjectured that almost every pair of people in the world are linked by a small chain of people, perhaps containing just 5 or fewer people.

This would mean that almost every pair of vertices in the acquaintanceship graph containing all people in the world is linked by a path of length not exceeding 4.



EXAMPLE Paths in Collaboration Graphs

In a collaboration graph, 2 people *a* and *b* are connected by a path when there is a sequence of people starting with *a* and ending with *b* such that the endpoints of each edge in the path are people who have collaborated.

In the academic collaboration graph of people who have written papers in mathematics, the Erd˝os number of a person *m* is the length of the shortest path between *m* and the extremely prolific mathematician Paul Erd˝os (who died in 1996).

That is, the Erd˝os number of a mathematician is the length of the shortest chain of mathematicians that begins with Paul Erd˝os and ends with this mathematician, where each adjacent pair of mathematicians have written a joint paper. The number of mathematicians with each Erd˝os number as of early 2006, according to the Erd˝os Number Project, is shown in Table 1.

TABLE 1 The Number of Mathematicians with a Given Erdős Number (as of early 2006).

Erdős Number	*Number of People*
0	1
1	504
2	6,593
3	33,605
4	83,642
5	87,760
6	40,014
7	11,591
8	3,146
9	819
10	244
11	68
12	23
13	5

In the Hollywood graph 2 actors *a* and *b* are linked when

- there is a chain of actors linking a and b, where every 2 actors adjacent in the chain have acted in the same movie.

In the Hollywood graph, the *Bacon number* of an actor *c* is defined to be the length of the shortest path connecting *c* and the well-known actor *Kevin Bacon*.

As new movies are made, including new ones with Kevin Bacon, the Bacon number of actors can change.

In Table 2 the number of actors with each Bacon number as of early 2011 are shown.

The origins of the Bacon number of an actor dates back to the early 1990s, when Kevin Bacon remarked that he had worked with everyone in Hollywood or someone who worked with them.

This lead some people to invent a party game where participants where challenged to find a sequence of movies leading from each actor named to Kevin Bacon.

We can find a number similar to a Bacon number using any actor as the center of the acting universe.

TABLE 2 The Number of Actors with a Given Bacon Number (as of early 2011).

Bacon Number	*Number of People*
0	1
1	2,367
2	242,407
3	785,389
4	200,602
5	14,048
6	1,277
7	114
8	16

Connectedness in Undirected Graphs

When does a computer network have the property that every pair of computers can share information, if messages can be sent through one or more intermediate computers?

When a graph is used to represent this computer network, where vertices represent the computers and edges represent the communication links, this question becomes:

When is there always a path between 2 vertices in the graph?

DEFINITION 3(connected graph)

An undirected graph is called connected (связный) if there is a path between every pair of distinct vertices of the graph.

An undirected graph that is not connected is called disconnected (несвязный).

We say that we disconnect a graph when we remove vertices or edges, or both, to produce a disconnected subgraph.

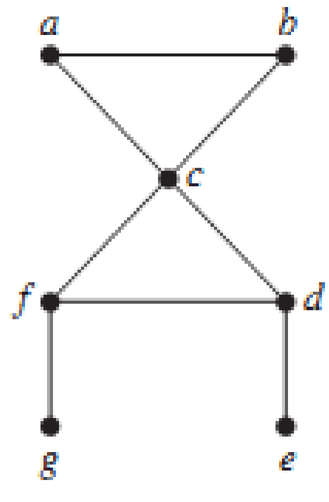
Thus, any 2 computers in the network can communicate ⇔ the graph of this network is connected.

EXAMPLE

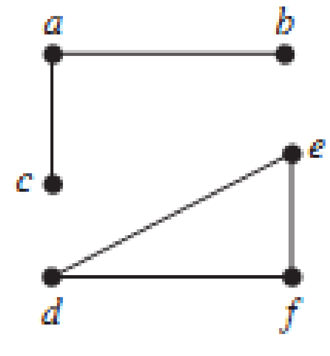
The graph G_1 in is connected, because for every pair of distinct vertices there is a path between them

However, the graph G_2 is not connected.

For instance, there is no path in G_2 between vertices a and d.



G_1



G_2

THEOREM 1 There is a simple path between every pair of distinct vertices of a connected undirected graph.

Proof: Let u and v be 2 distinct vertices of the connected undirected graph $G = (V, E)$.

Because G is connected, there is ≥ 1 path between u and v.

Let $x_0, x_1, \ldots, x_n$, where $x_0 = u$ and $x_n = v$, be the vertex sequence of a path of least length.

This path of least length is simple.

To see this, suppose it is not simple.

Then $x_i = x_j$ for some i and j with $0 \leq i < j$.

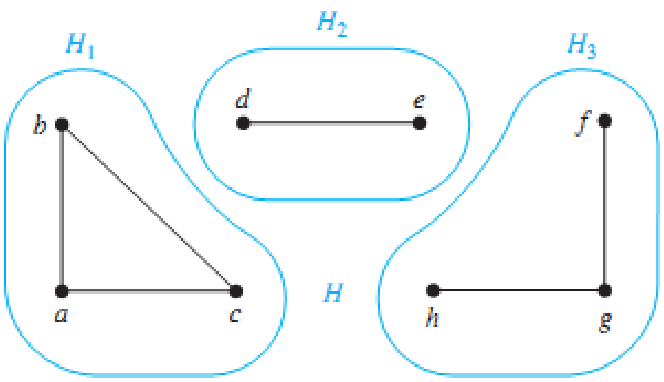
This means that there is a path from u to v of shorter length with vertex sequence $x_0, x_1, \ldots, x_{i-1}, x_j, \ldots, x_n$ obtained by deleting the edges corresponding to the vertex sequence $x_i, \ldots, x_{j-1}$.

CONNECTED COMPONENTS

A connected component of a graph G is a connected subgraph of G that is not a proper subgraph of another connected subgraph of G.

That is, a connected component of a graph G is a maximal connected subgraph of G.

A graph G that is not connected has ≥ 2 connected components that are disjoint and have G as their union.



EXAMPLE What are the connected components of the graph H?

Solution: The graph H is the union of 3 disjoint connected subgraphs H_1, H_2, and H_3.

These 3 subgraphs are the connected components of H.

EXAMPLE Connected Components of Call Graphs

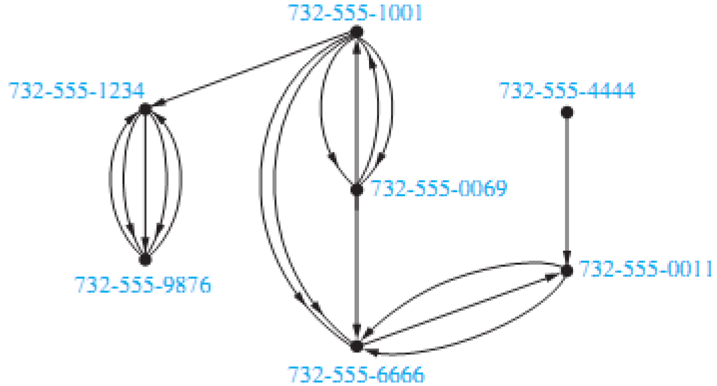
2 vertices *x* and *y* are in the same component of a telephone call graph when there is a sequence of telephone calls beginning at *x* and ending at *y*.

When a call graph for telephone calls made during a particular day in the AT&T network was analyzed, this graph was found to have 53,767,087 vertices, more than 170 million edges, and more than 3.7 million connected components.

Most of these components were small; approximately ¾ consisted of 2 vertices representing pairs of telephone numbers that called only each other.

This graph has one huge connected component with 44,989,297 vertices comprising more than 80% of the total.

Furthermore, every vertex in this component can be linked to any other vertex by a chain of $\leq$ 20 calls.



(a)

Paths and Isomorphism

There are several ways that paths and circuits can help determine whether 2 graphs are isomorphic.

For example, the existence of a simple circuit of a particular length is a useful invariant that can be used to show that 2 graphs are not isomorphic.

In addition, paths can be used to construct mappings that may be isomorphisms.

As we mentioned, a useful isomorphic invariant for simple graphs is the existence of a simple circuit of length k, where k is a positive integer > 2.

EXAMPLE

Determine whether the graphs *G* and *H* are isomorphic.

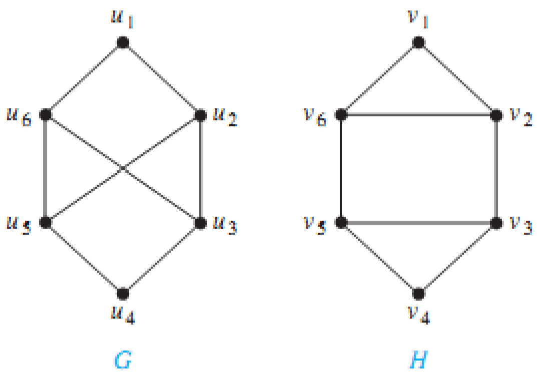
Solution: Both *G* and *H* have 6 vertices and 8 edges.

Each has 4 vertices of degree 3, and 2 vertices of degree 2.

So, the 3 invariants—number of vertices, number of edges, and degrees of vertices—all agree for the 2 graphs.

However, *H* has a simple circuit of length *3*, namely, v_1, v_2, v_6, v_1, whereas *G* has no simple circuit of length 3, as can be determined by inspection (all simple circuits in G have length at least 4).

Because the existence of a simple circuit of length 3 is an isomorphic invariant, *G* and *H* are not isomorphic.

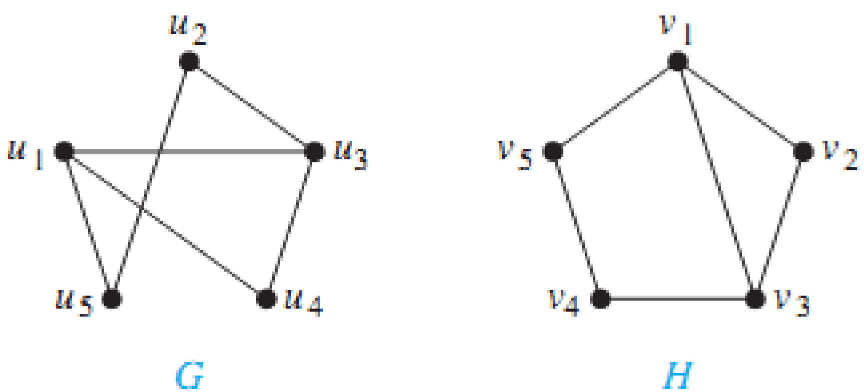


EXAMPLE

Determine whether the graphs G and H are isomorphic.

Solution: Both G and H have 5 vertices and 6 edges, both have 2 vertices of degree 3 and 3 vertices of degree 2, and both have a simple circuit of length 3, a simple circuit of length 4, and a simple circuit of length 5.

Because all these isomorphic invariants agree, G and H may be isomorphic.



To find a possible isomorphism, we can follow paths that go through all vertices so that the corresponding vertices in the 2 graphs have the same degree.

For example, the paths u_1, u_4, u_3, u_2, u_5 in G and $v_3, v_2, \underline{v}_1, v_5, v_4$ in H both go through every vertex in the graph; start at a vertex of degree 3; go through vertices of degrees 2, 3, and 2, respectively; and end at a vertex of degree 2.

By following these paths through the graphs, we define the mapping f with $f(u_1) = v_3$, $f(u_4) = v_2$, $f(u_3) = v_1$, $f(u_2) = v_5$, and $f(u_5) = v_4$.

It is possible to show that f is an isomorphism, so G and H are isomorphic, either by showing that f preserves edges or by showing that with the appropriate orderings of vertices the adjacency matrices of G and H are the same.