Degrees and Distances

Stepan Kuznetsov

Computer Science Department, Higher School of Economics

Outline

Graph Invariants and Graph Isomorphism

Handshaking Lemma

Clustering Coefficients

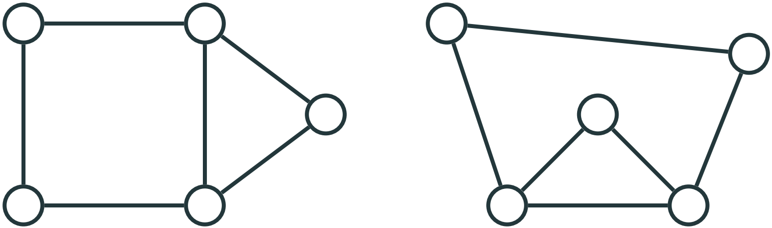
Distances. Diameter. Eccentricity

Graph Isomorphism

Sometimes graphs look different, but essentially are the same...

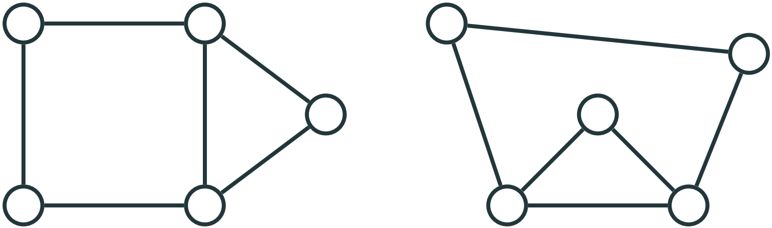
Graph Isomorphism

Sometimes graphs look different, but essentially are the same...



Graph Isomorphism

Sometimes graphs look different, but essentially are the same...



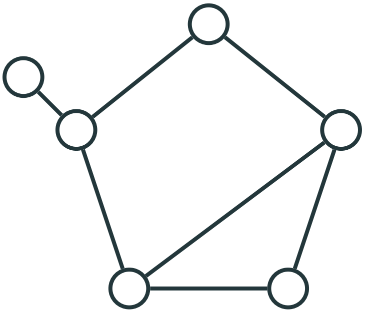
Here both graphs can be described as “a triangle and a quadrangle sharing a common edge.”

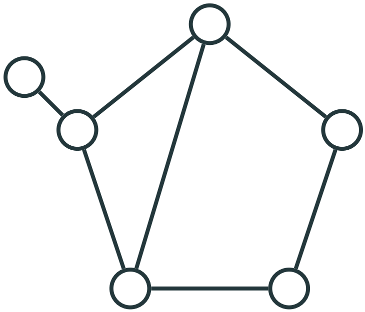
Graph Isomorphism

... and sometimes similarly looking graphs are different.

Graph Isomorphism

... and sometimes similarly looking graphs are different.





Graph Isomorphism

Isomorphic Graphs

Two graphs, $\mathcal{G}$ and $\mathcal{H}$, are *isomorphic,* if they have the same number n of vertices and vertices of each graph can be enumerated by numbers from 1 to n, so that vertices with numbers i and j are connected in $\mathcal{G}$ if and only if vertices with these numbers are connected in $\mathcal{H}$.

Graph Isomorphism

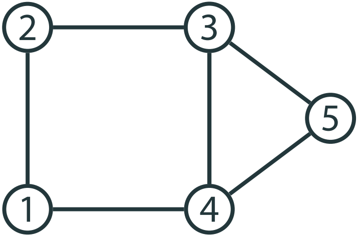
Isomorphic Graphs

Two graphs, $\mathcal{G}$ and $\mathcal{H}$, are *isomorphic,* if they have the same number n of vertices and vertices of each graph can be enumerated by numbers from 1 to n, so that vertices with numbers i and j are connected in $\mathcal{G}$ if and only if vertices with these numbers are connected in $\mathcal{H}$.

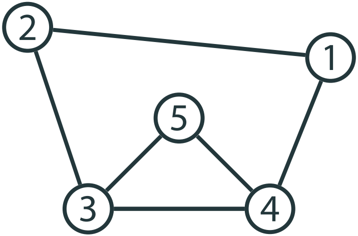
Isomorphic graphs can be seen as *different representations of the same graph*.

Isomorphic Graphs

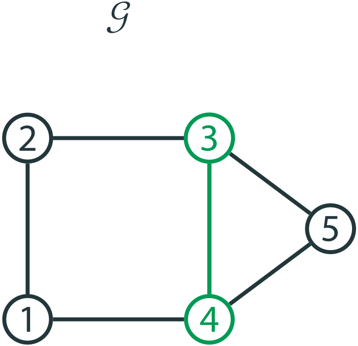
$\mathcal{G}$



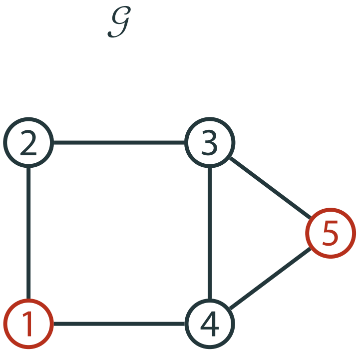
$\mathcal{H}$



Isomorphic Graphs



Isomorphic Graphs

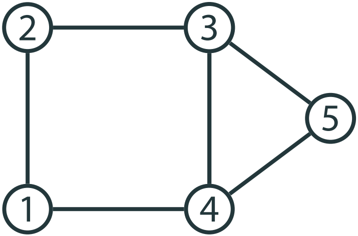


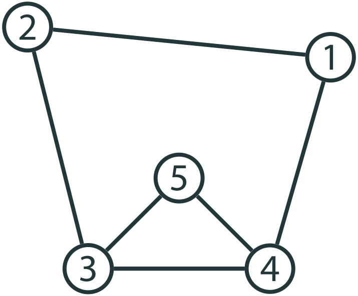
Isomorphism

The *isomorphism* itself is the correspondence between vertices with the same number.

Isomorphism

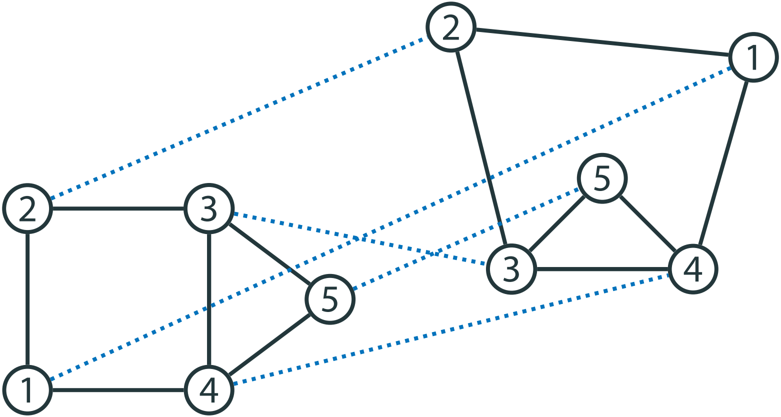
The *isomorphism* itself is the correspondence between vertices with the same number.





Isomorphism

The *isomorphism* itself is the correspondence between vertices with the same number.



Graph Invariants

- How to show that two graphs are **not** isomorphic?

Graph Invariants

- How to show that two graphs are **not** isomorphic?
- Let us calculate a *parameter* of the graph, which keeps the same among isomorphic versions of the same graph.

Graph Invariants

- How to show that two graphs are **not** isomorphic?
- Let us calculate a *parameter* of the graph, which keeps the same among isomorphic versions of the same graph.
- If the parameter is different for $\mathcal{G}$ and $\mathcal{H}$, then $\mathcal{G}$ and $\mathcal{H}$ cannot be isomorphic.

Graph Invariants

- How to show that two graphs are **not** isomorphic?
- Let us calculate a *parameter* of the graph, which keeps the same among isomorphic versions of the same graph.
- If the parameter is different for $\mathcal{G}$ and $\mathcal{H}$, then $\mathcal{G}$ and $\mathcal{H}$ cannot be isomorphic.
- Such a parameter is called a *graph invariant*.

Graph Invariants

- How to show that two graphs are **not** isomorphic?
- Let us calculate a *parameter* of the graph, which keeps the same among isomorphic versions of the same graph.
- If the parameter is different for $\mathcal{G}$ and $\mathcal{H}$, then $\mathcal{G}$ and $\mathcal{H}$ cannot be isomorphic.
- Such a parameter is called a *graph invariant*.
- Graph invariants could capture interesting *properties* of graphs.

Trivial Graph Invariants

- number of vertices

Trivial Graph Invariants

- number of vertices
- number of edges

Trivial Graph Invariants

- number of vertices
- number of edges
- acyclicity (is there any cyclic path in the graph?)

Trivial Graph Invariants

- number of vertices
- number of edges
- acyclicity (is there any cyclic path in the graph?)
- the number of isolated vertices

Trivial Graph Invariants

- number of vertices
- number of edges
- acyclicity (is there any cyclic path in the graph?)
- the number of isolated vertices
- ...

Outline

Graph Invariants and Graph Isomorphism

Handshaking Lemma

Clustering Coefficients

Distances. Diameter. Eccentricity

Vertex Degree

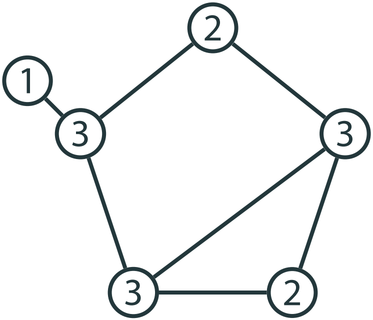
Degree

The *degree* of a vertex is the number of edges connected to it.

Vertex Degree

Degree

The *degree* of a vertex is the number of edges connected to it.



Degree as Invariant

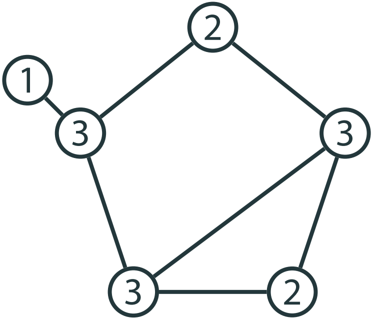
- Vertex degrees are a source of graph invariants.

Degree as Invariant

- Vertex degrees are a source of graph invariants.
- For example, if in $\mathcal{G}$ there are 14 vertices of degree 5 and in $\mathcal{H}$ there are 17 such vertices, then these graphs are not isomorphic.

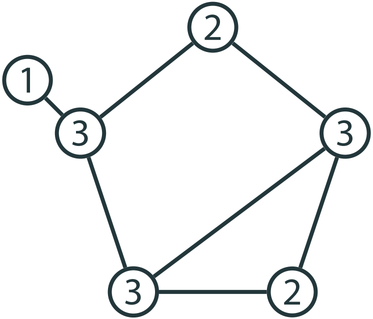
Sum of Degrees

Sum of Degrees



sum of vertex degrees $= 1 + 2 + 2 + 3 + 3 + 3 = 14$

Sum of Degrees



sum of vertex degrees $= 1 + 2 + 2 + 3 + 3 + 3 = 14$

number of edges $= 7$

Sum of Degrees

- This is always the case:

$$\text{number of edges} = \frac{\text{sum of vertex degrees}}{2}$$

Sum of Degrees

- This is always the case:

$$\text{number of edges} = \frac{\text{sum of vertex degrees}}{2}$$

- Proof: in the sum of degrees, each edge is counted exactly twice.

Handshaking Lemma

Handshaking Lemma

The number of people who shook hands with an odd number of other people is even.

Handshaking Lemma

Handshaking Lemma

The number of people who shook hands with an odd number of other people is even.

- Graph: vertices = people, edges = handshakes.

Handshaking Lemma

Handshaking Lemma

The number of people who shook hands with an odd number of other people is even.

- Graph: vertices = people, edges = handshakes.
- If the number of vertices with odd degree *(odd vertices)* is odd, then the sum of degrees of all vertices is also odd.

Handshaking Lemma

Handshaking Lemma

The number of people who shook hands with an odd number of other people is even.

- Graph: vertices = people, edges = handshakes.
- If the number of vertices with odd degree *(odd vertices)* is odd, then the sum of degrees of all vertices is also odd.
- This is a contradiction, since this sum is $2 \cdot (\text{number of edges})$.

Handshaking Lemma

- Notice that handshaking lemma is **not** a criterion of existence of a graph with given degrees of vertices.

Handshaking Lemma

- Notice that handshaking lemma is **not** a criterion of existence of a graph with given degrees of vertices.
- For example (exercise!), there exists no graph with the following degrees of vertices:

Handshaking Lemma

- Notice that handshaking lemma is **not** a criterion of existence of a graph with given degrees of vertices.
- For example (exercise!), there exists no graph with the following degrees of vertices:
    - one vertex of degree 1;

Handshaking Lemma

- Notice that handshaking lemma is **not** a criterion of existence of a graph with given degrees of vertices.
- For example (exercise!), there exists no graph with the following degrees of vertices:
    - one vertex of degree 1;
    - two vertices of degree 2;

Handshaking Lemma

- Notice that handshaking lemma is **not** a criterion of existence of a graph with given degrees of vertices.
- For example (exercise!), there exists no graph with the following degrees of vertices:
  - one vertex of degree 1;
  - two vertices of degree 2;
  - one vertex of degree 3;

Handshaking Lemma

- Notice that handshaking lemma is **not** a criterion of existence of a graph with given degrees of vertices.
- For example (exercise!), there exists no graph with the following degrees of vertices:
    - one vertex of degree 1;
    - two vertices of degree 2;
    - one vertex of degree 3;
    - two vertices of degree 5.

Outline

Graph Invariants and Graph Isomorphism

Handshaking Lemma

Clustering Coefficients

Distances. Diameter. Eccentricity

Some Graphs Tend to Cluster

- Social network graph: vertices = users, edges = friendship relations.

Some Graphs Tend to Cluster

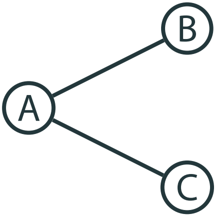
- Social network graph: vertices = users, edges = friendship relations.
- The probability, for two random vertices, to be connected, is generally quite low.

Some Graphs Tend to Cluster

- Social network graph: vertices = users, edges = friendship relations.
- The probability, for two random vertices, to be connected, is generally quite low.
- However, if Alex is a friend with Bob and Carl, a friendship relation between Bob and Carl becomes *much more probable*.

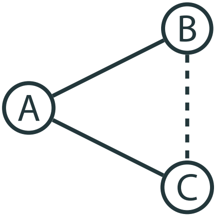
Some Graphs Tend to Cluster

- Social network graph: vertices = users, edges = friendship relations.
- The probability, for two random vertices, to be connected, is generally quite low.
- However, if Alex is a friend with Bob and Carl, a friendship relation between Bob and Carl becomes *much more probable*.



Some Graphs Tend to Cluster

- Social network graph: vertices = users, edges = friendship relations.
- The probability, for two random vertices, to be connected, is generally quite low.
- However, if Alex is a friend with Bob and Carl, a friendship relation between Bob and Carl becomes *much more probable*.

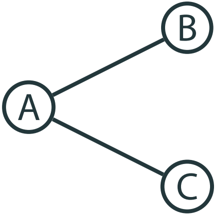


Global Clustering Coefficient

- One can measure clustering of a graph as a whole using the *global clustering coefficient*.

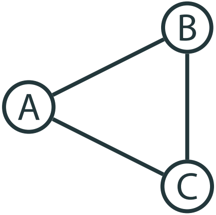
Global Clustering Coefficient

- One can measure clustering of a graph as a whole using the *global clustering coefficient*.
- A *triplet* is a pair of edges going from one vertex A:



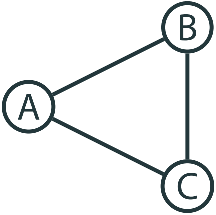
Global Clustering Coefficient

- A *triangle* is a triple of interconnected vertices:



Global Clustering Coefficient

- A *triangle* is a triple of interconnected vertices:



- $GCC(\mathcal{G}) = \dfrac{3 \cdot (\text{number of triangles})}{\text{number of triplets}}.$

Global Clustering Coefficient

- $GCC(\mathcal{G}) = \dfrac{3 \cdot (\text{number of triangles})}{\text{number of triplets}}.$

Global Clustering Coefficient

- $GCC(\mathcal{G}) = \dfrac{3 \cdot (\text{number of triangles})}{\text{number of triplets}}.$
- Why multiply by 3?

Global Clustering Coefficient

- $GCC(\mathcal{G}) = \dfrac{3 \cdot (\text{number of triangles})}{\text{number of triplets}}$.
- Why multiply by 3?
- Answer: each triangle includes three triplets.

Global Clustering Coefficient

- $GCC(\mathcal{G}) = \dfrac{3 \cdot (\text{number of triangles})}{\text{number of triplets}}$.
- Why multiply by 3?
- Answer: each triangle includes three triplets.
- Thus, the GCC is the *probability* for a random triplet A, B, C in $\mathcal{G}$ to be closed (that is, B and C connected).

Local Clustering Coefficient

- In the local clustering coefficient, we count only triplets with a given A as the central vertex.

Local Clustering Coefficient

- In the local clustering coefficient, we count only triplets with a given A as the central vertex.
- Suppose, the degree of A is k.

Local Clustering Coefficient

- In the local clustering coefficient, we count only triplets with a given A as the central vertex.
- Suppose, the degree of A is k.
- The total number of triplets with A as the central is $k \cdot (k-1)/2$.

Local Clustering Coefficient

- In the local clustering coefficient, we count only triplets with a given A as the central vertex.
- Suppose, the degree of A is k.
- The total number of triplets with A as the central is $k \cdot (k-1)/2$.
- $LCC(A)$ is the ratio

$$\frac{\text{number of pairs (B, C) which form a triangle with A}}{k \cdot (k-1)/2}$$

Local Clustering Coefficient

- In the local clustering coefficient, we count only triplets with a given A as the central vertex.
- Suppose, the degree of A is k.
- The total number of triplets with A as the central is $k \cdot (k-1)/2$.
- $LCC(A)$ is the ratio

$$\frac{2 \cdot (\text{number of pairs (B, C) which form a triangle with A})}{k \cdot (k-1)}$$

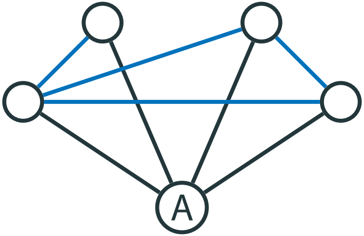
Local Clustering Coefficient

- In the local clustering coefficient, we count only triplets with a given A as the central vertex.
- Suppose, the degree of A is k.
- The total number of triplets with A as the central is $k \cdot (k-1)/2$.
- $LCC(A)$ is the ratio

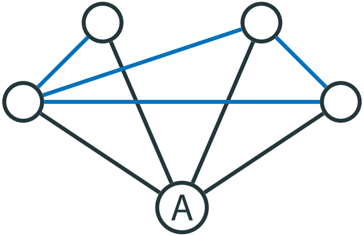
$$\frac{2 \cdot (\text{number of pairs (B, C) which form a triangle with A})}{k \cdot (k-1)}$$

- If A is an isolated vertex (degree = 0), then $LCC(A)$ is undefined (zero-by-zero division).

Local Clustering Coefficient



Local Clustering Coefficient



In this example, $LCC(A) = \frac{2 \cdot 4}{4 \cdot 3} = \frac{2}{3}$.

Outline

Graph Invariants and Graph Isomorphism

Handshaking Lemma

Clustering Coefficients

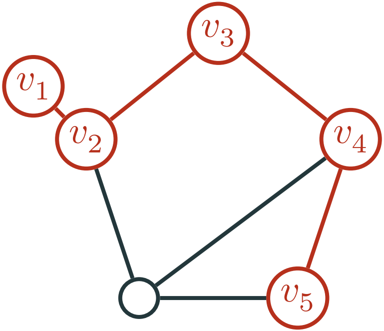
Distances. Diameter. Eccentricity

Paths

- A *path* in a graph is a sequence of vertices $v_1, v_2, \ldots, v_n$, such that v_i is connected to v_{i+1}

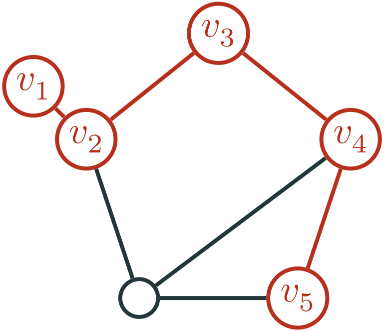
Paths

- A *path* in a graph is a sequence of vertices $v_1, v_2, \dots, v_n$, such that v_i is connected to v_{i+1}



Paths

- A *path* in a graph is a sequence of vertices $v_1, v_2, \ldots, v_n$, such that v_i is connected to v_{i+1}



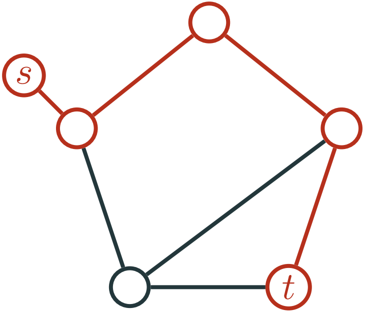
- The *length* of the path is the number of *edges* in it: $n - 1$ (here the length is 4).

Distance

- The *distance* between two vertices is the length of the *shortest* path connecting them.

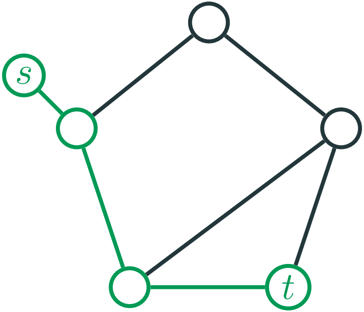
Distance

- The *distance* between two vertices is the length of the *shortest* path connecting them.



Distance

- The *distance* between two vertices is the length of the *shortest* path connecting them.



- Here the distance between s and t is 3:

$$d(s,t) = 3.$$

Distance

- If there is no path from s to t, then $d(s,t) = \infty$.

Distance

- If there is no path from s to t, then $d(s,t) = \infty$.
- $d(s,s) = 0$.

Distance

- If there is no path from s to t, then $d(s,t) = \infty$.
- $d(s,s) = 0$.
- **Triangle inequality:**

$$d(s,t) \leq d(s,q) + d(q,t),$$

for any vertices s, t, q.

Distance

- If there is no path from s to t, then $d(s,t) = \infty$.
- $d(s,s) = 0$.
- **Triangle inequality:**

$$d(s,t) \leq d(s,q) + d(q,t),$$

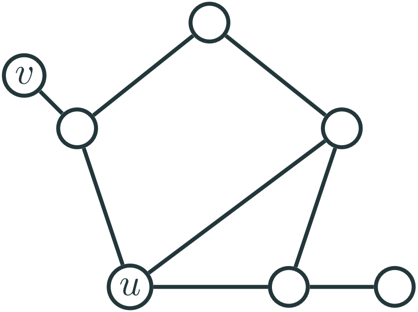
  for any vertices s, t, q.
- Proof: the *shortest* path between s and t is shorter, or at least not longer, then *any* path from s to t via q.

Eccentricity

- The *eccentricity* of a given vertex is the maximal distance from this vertex to another one: $ecc(u) = \max_{v \in V} d(u, v)$.

Eccentricity

- The *eccentricity* of a given vertex is the maximal distance from this vertex to another one: $ecc(u) = \max_{v \in V} d(u, v)$.
- In this example, the eccentricity of u is 2 and the eccentricity of v is 4:



Calculating Distance and Eccentricity

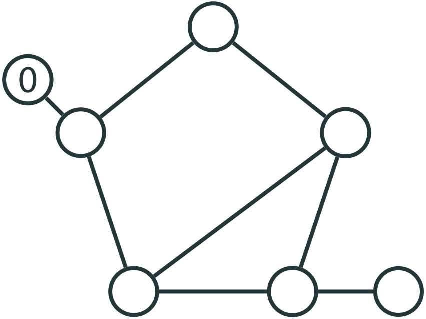
- Calculating distance is an application of BFS.

Calculating Distance and Eccentricity

- Calculating distance is an application of BFS.
- The BFS algorithm, starting from a given vertex s, yields the vertices layer-by-layer: first with distance 1 (directly connected), then 2, then 3, etc.

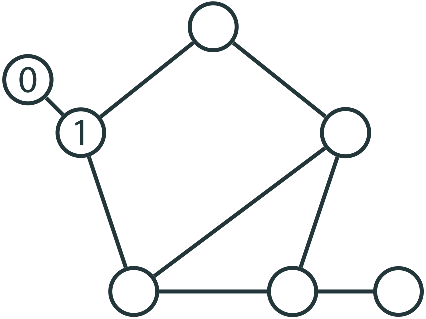
Calculating Distance and Eccentricity

- Calculating distance is an application of BFS.
- The BFS algorithm, starting from a given vertex s, yields the vertices layer-by-layer: first with distance 1 (directly connected), then 2, then 3, etc.



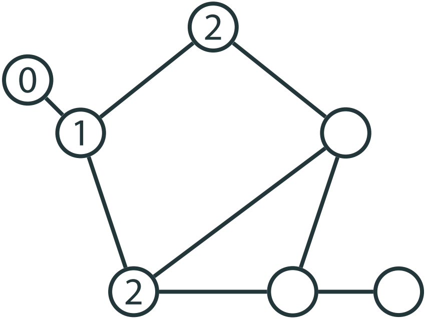
Calculating Distance and Eccentricity

- Calculating distance is an application of BFS.
- The BFS algorithm, starting from a given vertex s, yields the vertices layer-by-layer: first with distance 1 (directly connected), then 2, then 3, etc.



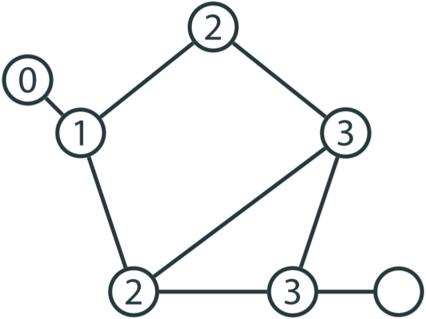
Calculating Distance and Eccentricity

- Calculating distance is an application of BFS.
- The BFS algorithm, starting from a given vertex s, yields the vertices layer-by-layer: first with distance 1 (directly connected), then 2, then 3, etc.



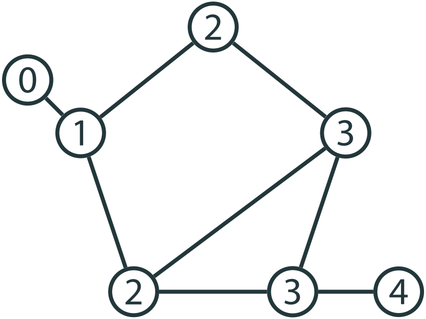
Calculating Distance and Eccentricity

- Calculating distance is an application of BFS.
- The BFS algorithm, starting from a given vertex s, yields the vertices layer-by-layer: first with distance 1 (directly connected), then 2, then 3, etc.



Calculating Distance and Eccentricity

- Calculating distance is an application of BFS.
- The BFS algorithm, starting from a given vertex s, yields the vertices layer-by-layer: first with distance 1 (directly connected), then 2, then 3, etc.



Calculating Distance and Eccentricity

- Once BFS reaches the other given vertex t, we know the distance $d(s, t)$.

Calculating Distance and Eccentricity

- Once BFS reaches the other given vertex t, we know the distance $d(s, t)$.
- If BFS finishes without reaching t, then t is not reachable from s, i.e., $d(s, t) = \infty$.

Calculating Distance and Eccentricity

- Once BFS reaches the other given vertex t, we know the distance $d(s,t)$.
- If BFS finishes without reaching t, then t is not reachable from s, i.e., $d(s,t) = \infty$.
- If BFS reaches all vertices, then $ecc(s)$ is the number of layers produced by BFS (not counting the 0-th one).

Calculating Distance and Eccentricity

- Once BFS reaches the other given vertex t, we know the distance $d(s, t)$.
- If BFS finishes without reaching t, then t is not reachable from s, i.e., $d(s, t) = \infty$.
- If BFS reaches all vertices, then $ecc(s)$ is the number of layers produced by BFS (not counting the 0-th one).
- Otherwise, $ecc(s) = \infty$.

Eccentricity and Diameter

- Eccentricity measures how far a given vertex is from the “central part” of the graph.

Eccentricity and Diameter

- Eccentricity measures how far a given vertex is from the “central part” of the graph.
- The *diameter* of a graph is the maximal distance possible in this graph:

$$diam(\mathcal{G}) = \max_{u,v \in V} d(u, v).$$

Eccentricity and Diameter

- Eccentricity measures how far a given vertex is from the “central part” of the graph.
- The *diameter* of a graph is the maximal distance possible in this graph:

$$diam(\mathcal{G}) = \max_{u,v \in V} d(u, v).$$

- Equivalently, the diameter is the maximal eccentricity:

$$diam(\mathcal{G}) = \max_{u \in V} ecc(u).$$

Diameter and Radius

- The diameter is the maximal eccentricity.

Diameter and Radius

- The diameter is the maximal eccentricity.
- The *radius* is the minimal eccentricity:

$$r(\mathcal{G}) = \min_{u \in V} ecc(u).$$

Diameter and Radius

- The diameter is the maximal eccentricity.
- The *radius* is the minimal eccentricity:

$$r(\mathcal{G}) = \min_{u \in V} ecc(u).$$

- It is easy to see (exercise!) that

$$r(\mathcal{G}) \leq diam(\mathcal{G}) \leq 2 \cdot r(\mathcal{G}).$$