
ĐỒ ÁN KẾT THÚC HỌC PHẦN

MÔ HÌNH HÓA THỐNG KÊ

Lê Nhật Nam - 23C24004, Phạm Thừa Tiểu Thành - 23C24005

26/02/2023

Contents

1	Thông tin chung	2
2	Cấu trúc mã nguồn	3
2.1	Tổ chức dữ liệu đồ thị của lớp Graph	3
2.1.1	Cấu trúc dữ liệu đồ thị	3
2.1.2	Các hàm thành phần	3
2.2	Tổ chức dữ liệu đỉnh của lớp Vertex	5
2.2.1	Cấu trúc dữ liệu của lớp đỉnh Vertex	6
2.2.2	Các hàm thành phần	6
2.3	Các hàm hỗ trợ	7
2.4	Mô tả thuật toán DFS, BFS	8
2.4.1	Thuật toán DFS	8
2.4.2	Thuật toán BFS	10
3	Tài liệu tham khảo	14

1 Thông tin chung

- Thành viên:
 - Trần Xuân Lộc - 22C11064
 - Nguyễn Bảo Long - 22C11065
- Bảng phân công công việc:

Công việc	Người thực hiện
Tái cấu trúc đồ thị theo yêu cầu của thầy	Xuân Lộc
Viết hàm và tài liệu cho thuật toán DFS	Bảo Long
Viết hàm và tài liệu cho thuật toán BFS	Xuân Lộc

2 Cấu trúc mã nguồn

- Tại đây mô tả về schema của các class và thông tin (tóm tắt, đầu vào, đầu ra) của các hàm.

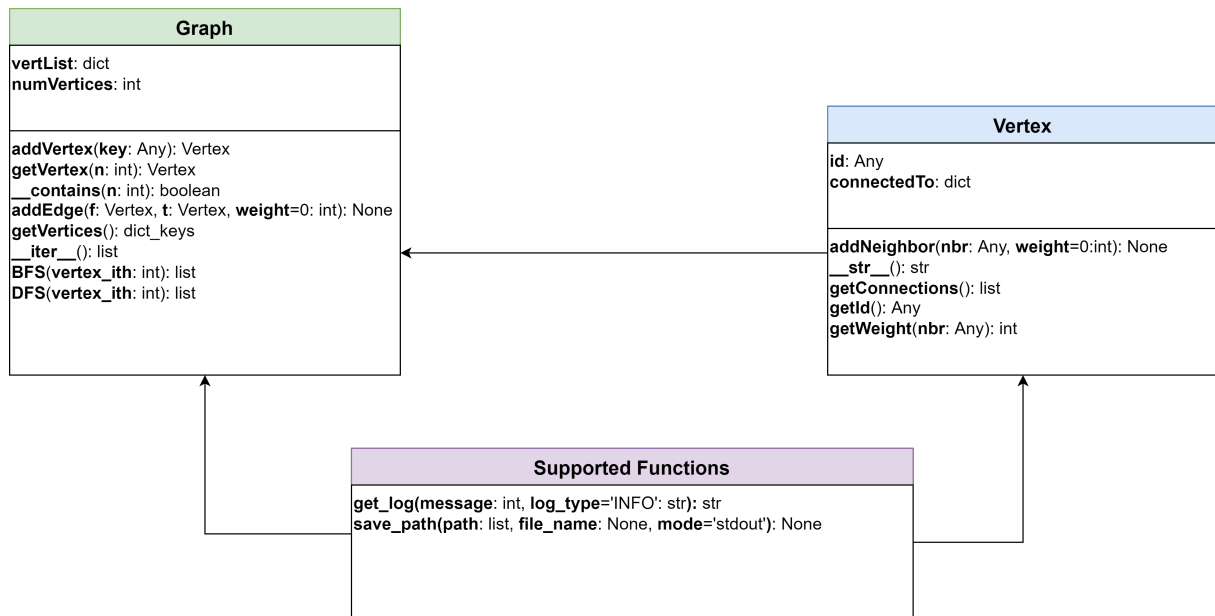


Figure 1: Tổ chức dữ liệu đồ thị.

2.1 Tổ chức dữ liệu đồ thị của lớp Graph

Trong phần này, nhóm sẽ trình bày cách tổ chức lưu trữ dữ liệu đồ thị vào trong lớp **Graph** bao gồm các thành phần lưu trữ dữ liệu và mô tả các hàm thực thi phục vụ cho lớp. Tập `graph.py` lưu trữ thông tin cấu hình chi tiết và mã nguồn.

2.1.1 Cấu trúc dữ liệu đồ thị

- `self.vertList`: Biến có kiểu dữ liệu từ điển, chứa danh sách đỉnh của đồ thị. Mỗi phần tử trong từ điển có khóa là định danh của đỉnh (`id`) và giá trị là một đối tượng có kiểu dữ liệu `Vertex`.
- `self.numVertices`: Biến có kiểu dữ liệu là số nguyên, xác định số đỉnh của đồ thị.

2.1.2 Các hàm thành phần

- Hàm `addVertex(self, key):`

- Mô tả: Hàm thêm một đỉnh vào cấu trúc dữ liệu đồ thị.
 - Tham số:
 - * **key**: Định danh của một đỉnh.
 - Trả về: Đỉnh vừa được thêm vào dưới dạng một đối tượng **Vertex**.
 - Hàm **getVertex(self, n)**:
 - Mô tả: Hàm lấy thông tin của đỉnh có định danh **n** của đồ thị
 - Tham số:
 - * **n**: Định danh (**id**) của đỉnh trong đồ thị.
 - Trả về:
 - * Đối tượng **Vertex** có định danh **n**, nếu đỉnh **n** có tồn tại trong đồ thị.
 - * **None**, nếu trong đồ thị không tồn tại đỉnh có định danh **n**.
 - Hàm **__contains__(self, n)**:
 - Mô tả: Hàm kiểm tra đỉnh có định danh **n** có tồn tại trong đồ thị hay không.
 - Tham số:
 - * **n**: Định danh (**id**) của một đỉnh.
 - Trả về:
 - * **True**, nếu đỉnh **n** tồn tại trong đồ thị.
 - * **False**, nếu đỉnh **n** không tồn tại trong đồ thị.
 - Hàm **addEdge(self, f, t, weight=0)**:
 - Mô tả: Hàm thêm một cạnh có trọng số **weight** (mặc định bằng 0) đi từ đỉnh **f** đến đỉnh **t**. Nếu một trong hai đỉnh không tồn tại trong đồ thị thì thêm đỉnh đó vào đồ thị.
 - Tham số:
 - * **f**: Định danh của đỉnh xuất phát.
 - * **t**: Định danh của đỉnh đích.
 - * **weight**: Trọng số của đỉnh được thêm vào. Mặc định bằng 0.
 - Hàm **getVertices(self)**:
 - Mô tả: Hàm lấy thông tin của toàn bộ đỉnh trong đồ thị.
 - Tham số: Không.
 - Trả về:
 - * Trả về đối tượng **dict_key**, chứa danh sách định danh của toàn bộ đỉnh trong đồ thị.
 - Hàm **__iter__(self)**:
 - Mô tả: Hàm hỗ trợ việc duyệt qua mọi đỉnh trong đồ thị.
-

- Tham số: Không.
- Trả về:
 - * Đối tượng có kiểu dữ liệu `iterator`, hỗ trợ việc duyệt qua mọi đỉnh trong đồ thị.
- Hàm `BFS(self, vertex_ith)`:
 - Mô tả: Hàm duyệt qua tất cả các đỉnh trong đồ thị bằng thuật toán `BFS` với đỉnh bắt đầu là `vertex_ith`.
 - Tham số:
 - * `vertex_ith`: Định danh (`id`) của đỉnh bắt đầu.
 - Trả về:
 - * Thứ tự đỉnh được duyệt qua bởi thuật toán `BFS`.
- Hàm `DFS(self, vertex_ith)`:
 - Mô tả: Hàm duyệt qua tất cả các đỉnh trong đồ thị bằng thuật toán `DFS` với đỉnh bắt đầu là `vertex_ith`.
 - Tham số:
 - * `vertex_ith`: Định danh (`id`) của đỉnh bắt đầu.
 - Trả về:
 - * Thứ tự đỉnh được duyệt qua bởi thuật toán `DFS`.
- Hàm `save_path(path: list, file_name=None, mode='stdout')`:
 - Mô tả: Hàm hỗ trợ lưu kết quả dưới dạng tệp hoặc hiển thị kết quả ra màn hình.
 - Tham số:
 - * `path`: Danh sách lưu lại các thứ tự duyệt các nút của các thuật toán.
 - * `file_name`: Biến lưu giá trị tên tệp lưu kết quả, nếu `mode='write_to_file'` nhưng giá trị biến rỗng sẽ báo lỗi cho người dùng.
 - * `mode`: Biến mang cấu hình hiển thị ra màn hình nếu `mode='stdout'` (giá trị mặc định) hoặc ghi kết quả vào tệp `mode='write_to_file'`.
 - Trả về: Không

2.2 Tổ chức dữ liệu đỉnh của lớp `Vertex`

Tiếp theo, nhóm sẽ trình bày cách tổ chức lưu trữ dữ liệu của các đỉnh bao gồm định danh đỉnh và tập hợp các láng giềng kề với đỉnh đó. Ngoài ra, chúng em cũng mô tả thêm các hàm thực thi hỗ trợ cho lớp này và các hàm này được lưu trong tệp `vertex.py`.

2.2.1 Cấu trúc dữ liệu của lớp đỉnh Vertex

- `self.id`: Biến lưu trữ định danh của một đỉnh, có kiểu dữ liệu bất kì - `Any`, không có ràng buộc có thể là kiểu số hoặc chuỗi tùy ý.
- `self.connectedTo`: Biến lưu trữ tập hợp các láng giềng có kề với đỉnh hiện tại, kiểu dữ liệu lưu trữ là từ điển - `dict`.

2.2.2 Các hàm thành phần

- Hàm `addNeighbor(self, nbr, weight=0)`:
 - Mô tả: Hàm thêm láng giềng `nbr` có liên kết với đỉnh hiện tại với trọng số mặc định `weight=0`.
 - Tham số:
 - * `nbr`: láng giềng của đỉnh đang xét.
 - * `weight`: giá trị thể hiện trọng số liên kết.
 - Trả về: Không
 - Hàm `__str__(self)`:
 - Mô tả: Hàm mô tả đỉnh thông qua các thông số lưu trữ.
 - Tham số: Không
 - Trả về: Chuỗi bao gồm định danh và tập hợp các đỉnh kề của đỉnh đó.
 - Hàm `getConnections(self)`:
 - Mô tả: hàm trả về
 - Tham số: Không
 - Trả về: mảng các đỉnh có liên kết với đỉnh hiện tại thông qua giá trị của biến `self.connectedTo`.
 - Hàm `getId(self)`:
 - Mô tả: Hàm trả về định danh của đỉnh hiện tại.
 - Tham số: Không
 - Trả về: Định danh của đỉnh hiện tại của đỉnh với kiểu dữ liệu bất kì `Any`.
 - Hàm `getWeight(self, nbr)`:
 - Mô tả: Hàm trả về trọng số liên kết của đỉnh `nbr` với đỉnh đang xét.
 - Tham số:
 - * `nbr`: láng giềng của đỉnh hiện tại.
 - Trả về: Trọng số của cạnh giữa đỉnh hiện tại và `nbr`.
-

2.3 Các hàm hỗ trợ

Phần này sẽ tập trung mô tả các hàm hỗ trợ ghi dữ liệu và hiển thị dữ liệu cho người dùng thông báo tình trạng thực thi các hàm của các lớp. Các hàm này được lưu trong tệp `support.py`.

- Hàm `save_path(path: list, file_name=None, mode='stdout')`:
 - Mô tả: Hàm hỗ trợ lưu kết quả dưới dạng tệp hoặc hiển thị kết quả ra màn hình.
 - Tham số:
 - * `path`: Danh sách lưu lại các thứ tự duyệt các nút của các thuật toán.
 - * `file_name`: Biến lưu giá trị tên tệp lưu kết quả, nếu `mode='write_to_file'` nhưng giá trị biến rỗng sẽ báo lỗi cho người dùng.
 - * `mode`: Biến mang cấu hình hiển thị ra màn hình nếu `mode='stdout'` (giá trị mặc định) hoặc ghi kết quả vào tệp `mode='write_to_file'`.
 - Trả về: Không
- Hàm `get_log(message, log_type='INFO')`:
 - Mô tả: Hàm hỗ trợ ghi thông tin thực thi của hàm bao gồm loại nhật kí ghi, thời gian thực thi và thông điệp muốn ghi lại.
 - Tham số:
 - * `path`: Danh sách lưu lại các thứ tự duyệt các nút của các thuật toán.
 - * `log_type`: Biến mang cấu hình loại nhật kí được thực thi với giá trị mặc định là `log_type='stdout'`, một số loại nhật kí khác như `WARNING`, `ERROR`, `DEBUG`
 - Trả về: Chuỗi lưu trữ nhật kí hoặc thông tin thực thi tại thời điểm gọi hàm.

2.4 Mô tả thuật toán DFS, BFS

2.4.1 Thuật toán DFS

- Ý tưởng thuật toán: Bắt đầu từ đỉnh xuất phát đi xa nhất có thể, đến khi không thể đi được nữa thì quay lui (backtracking). Chính vì vậy, có thể cài đặt thuật toán này bằng đệ quy hoặc sử dụng một ngăn xếp.
- Thuật toán được cài đặt như sau:

```
1 def DFS(self, vertex_ith: int):
2     """depth first search function, start from `vertex_ith`
3     Args: vertex_ith (int): key of vertex in graph
4     Raises: ValueError: can't find a vertex with given key
5     Returns: list[int]: the path that DFS agent has gone through
6     """
7     vertex: Vertex = self.getVertex(vertex_ith)
8     if vertex is None:
9         message = 'Invalid vertex id, could not found vertex id `'+
10            + str(vertex_ith) + '` in Graph'
11         raise ValueError(get_log(message, log_type='ERROR'))
12
13     closed_set: list[int] = []
14     open_set: list[int] = [vertex.getId()]
15
16     while open_set:
17         cur_vertex: Vertex = self.getVertex(open_set.pop())
18         cur_vertex_id = cur_vertex.getId()
19
20         if cur_vertex_id not in closed_set:
21             closed_set.append(cur_vertex_id)
22             neighbors = [x.id for x in cur_vertex.getConnections()]
23
24             for neighbor in neighbors:
25                 if neighbor not in closed_set:
26                     open_set.append(neighbor)
27
28     return closed_set
```

- Minh họa thuật toán:
 - Đồ thị:

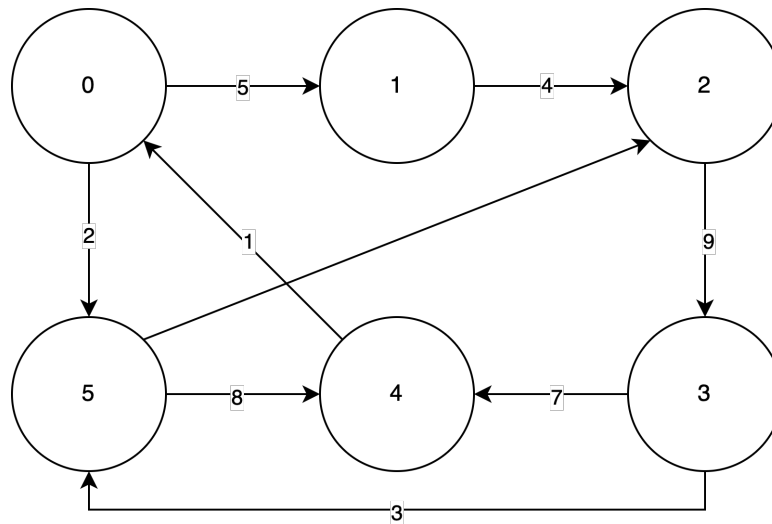


Figure 2: Minh họa đồ thị

– Quá trình duyệt đồ thị:

current node	stack	visited
0	{1,5}	{0}
5	{1,4,2}	{0,5}
2	{1,4,3}	{0,5,2}
3	{1,4}	{0,5,2,3}
4	{1}	{0,5,2,3,4}
1	{}	{0,5,2,3,4,1}

– Kết quả: Thứ tự duyệt của đồ thị là {0, 5, 2, 3, 4, 1}

– Minh họa bằng cây tìm kiếm:

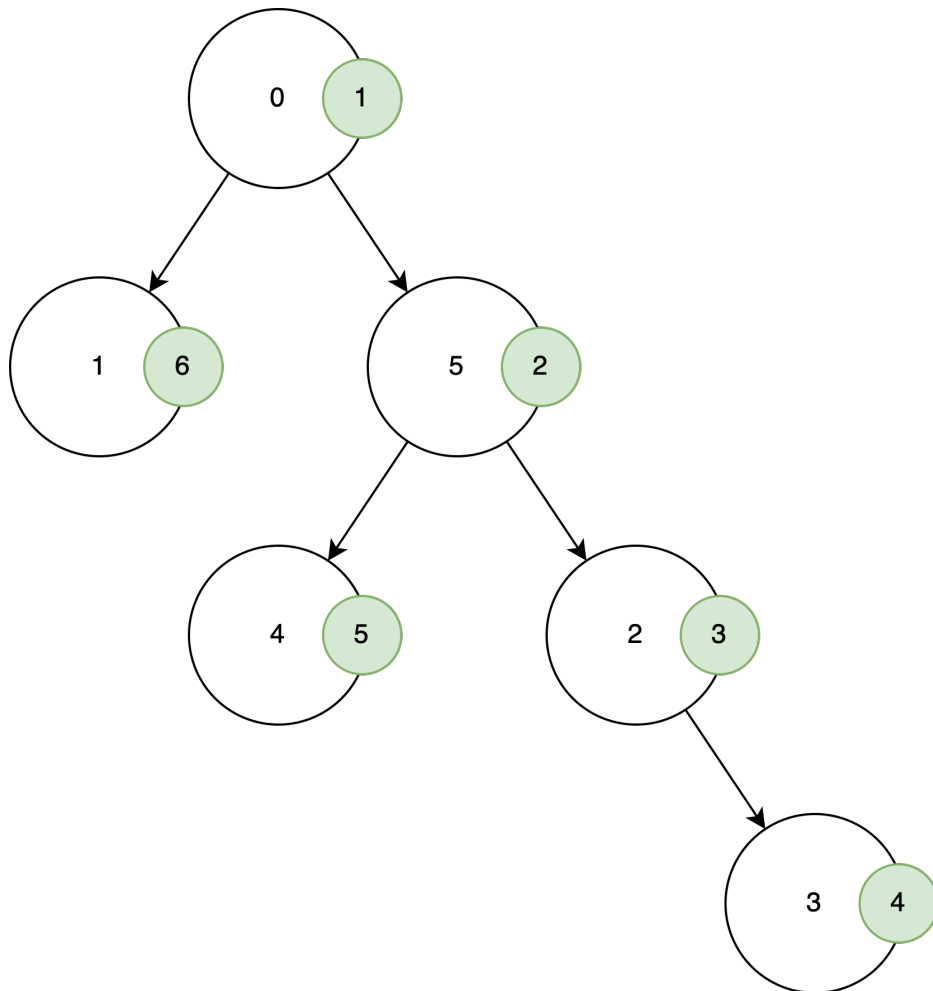


Figure 3: Minh họa thuật toán DFS bằng cây tìm kiếm. Thứ tự duyệt được thể hiện trong hình tròn màu xanh.

2.4.2 Thuật toán BFS

- Ý tưởng thuật toán: Bắt đầu từ đỉnh xuất phát đi rộng nhất có thể, đến khi không thể đi được nữa thì quay lại đi xuống 1 bậc đồ thị để tiếp tục quá trình tương tự. Do đó, ta có thể cài đặt thuật toán này bằng 1 hàng đợi và 1 mảng đánh dấu đã duyệt là đủ.
- Cấu hình thuật toán được thể hiện ở bên dưới.

```
1 def BFS(self, vertex_ith: int):  
2     """  
3     Module applying Breadth First Search Algorithm.  
4  
5     :param vertex_ith: the vertex id in Graph
```

```
6 :return: path computed by BFS
7 """
8 # get the vertex `vertex_ith`.
9 vertex = self.getVertex(vertex_ith)
10
11 # checking if not exist `vertex_ith` in Graph then raise error
12 if not vertex:
13     message = 'Invalid vertex id, could not found vertex id `'
14             + str(vertex_ith) + '` in Graph'
15     raise ValueError(get_log(message, log_type='ERROR'))
16
17 # get the number of vertices.
18 n = self.numVertices
19
20 # bool array for marking visited or not.
21 visited = [False] * n
22
23 # get the vertex_id for easy management.
24 vertex_id = vertex.getId()
25
26 # initializing a queue to handling which vertex is remaining.
27 queue = [vertex_id]
28
29 # marking the `vertex_id` is visited due to the beginning
30 vertex.
31 visited[vertex_id] = True
32
33 # path to track the working state of BFS.
34 path = []
35 while queue:
36     # handling current vertex before removing out of queue.
37     cur_pos = queue[0]
38
39     # appending to path to track.
40     path.append(cur_pos)
41     # remove it out of queue
42     queue.pop(0)
43     # get all neighbors id of current vertex.
44     neighbor_cur_pos = [x.id for x in self.getVertex(cur_pos).
45                         getConnections()]
46
47     # loop over the neighbor of current vertex.
48     for neighborId in neighbor_cur_pos:
49         # if not visited then push that vertex into queue.
50         if not visited[neighborId]:
51             visited[neighborId] = True
52             queue.append(neighborId)
53
54 return path
```

- Minh họa thuật toán:

- Đồ thị:

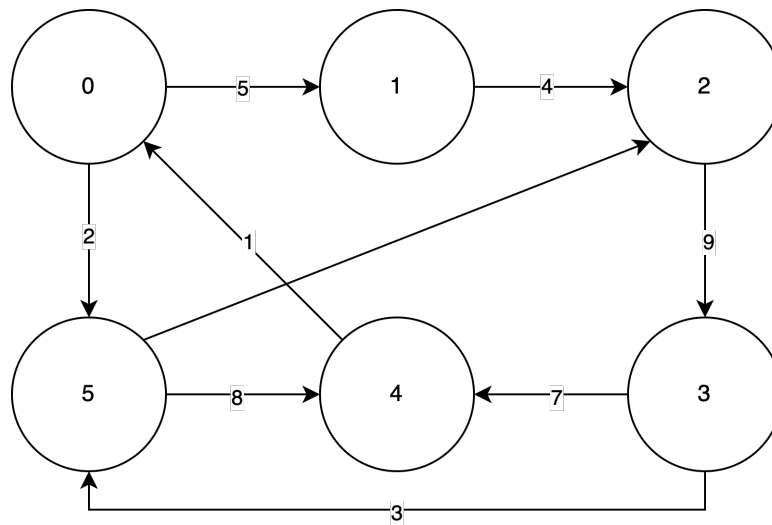


Figure 4: Minh họa đồ thị

- Quá trình duyệt đồ thị:

current node	stack	visited
0	{1,5}	{0}
1	{5,2}	{0,1}
5	{2,4}	{0,1,5}
2	{4,3}	{0,1,5,2}
4	{3}	{0,1,5,2,4}
3	{}	{0,1,5,2,4,3}

- Kết quả: Thứ tự duyệt của đồ thị là {0, 1, 5, 2, 4, 3}

- Minh họa bằng cây tìm kiếm:

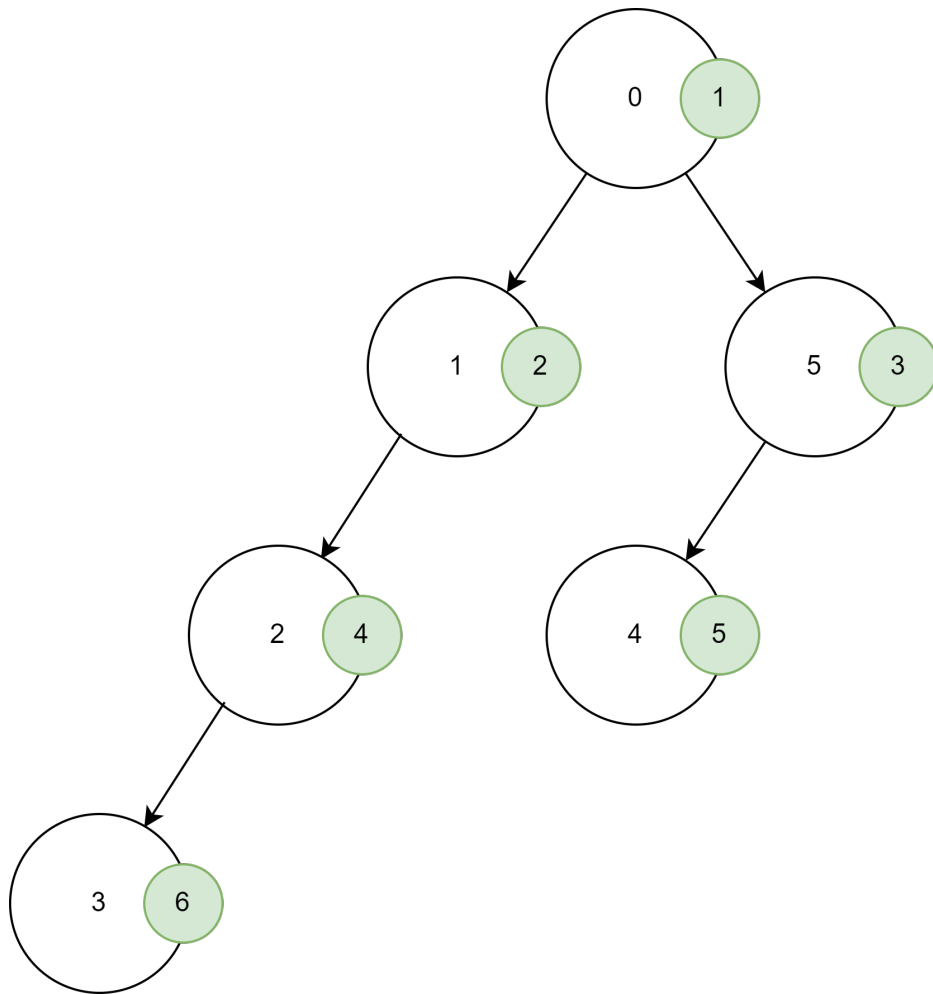


Figure 5: Minh họa thuật toán BFS bằng thuật toán duyệt theo chiều rộng với cây tìm kiếm. Thứ tự duyệt được thể hiện trong hình tròn màu xanh.

3 Tài liệu tham khảo

- Một số bài toán cơ bản trong phân tích dữ liệu. (n.d.). Thuc Nguyen Dinh.