**LAB 5**

***Phân tích mạng xã hội***

Cho file văn bản ***data/twitter\_following.txt*** chứa thông tin về việc tài khoản người dùng theo dõi tài khoản khác trên cùng mạng xã hội. Mỗi dòng file văn bản có dạng:

**<user\_id1> <user\_id2>**

cho biết người dùng với user\_id1 theo dõi người dùng user\_id2. Nếu user\_id2 cũng theo dõi user\_id1 thì cặp tài khoản này được gọi là theo dõi lẫn nhau (mutual followers).

**Yêu cầu:**

Vận dụng cả 2 phương pháp (1) truyền thống và (2) dữ liệu lớn thực hiện các xử lý sau:

* Thống kê số lượng người theo dõi (followers) của mỗi tài khoản người dùng.
* Liệt kê top 5 người dùng có nhiều theo dõi nhất.
* Liệt kê tất cả các cặp người dùng theo dõi lẫn nhau trong file dữ liệu được cho.

**Hướng dẫn:**

* Có thể biểu diễn việc theo dõi nhau của các tài khoản Twitter bằng một đồ thị có hướng (directed graph). Gọi G = (V, E) là đồ thị biểu diễn mối quan hệ theo dõi trên Twitter, trong đó V là tập hợp các đỉnh (vertices/nodes) - mỗi đỉnh tương ứng một tài khoản người dùng, E là tập các cạnh (edges) - mỗi cạnh là một cặp có thứ tự đỉnh <u, v> thể hiện "u theo dõi v".
* Đồ thị như trên có thể biểu diễn, xử lý bằng thư viện ***NetworkX***.
* Đoạn code sau đây minh họa đọc dữ liệu từ file ***data/twitter\_following.txt***, tạo đồ thị thể hiện mối quan hệ theo dõi, tìm và liệt kê các cặp tài khoản theo dõi lẫn nhau (mutual followers):

***import networkx as nx***

***edges = []***

***f = open('data/twitter\_following.txt', 'r')***

***for line in f.readlines():***

***id1, id2 = line.strip().split()***

***edges.append((id1,id2))***

***# create directed graph from the edge pairs using NetworkX Library***

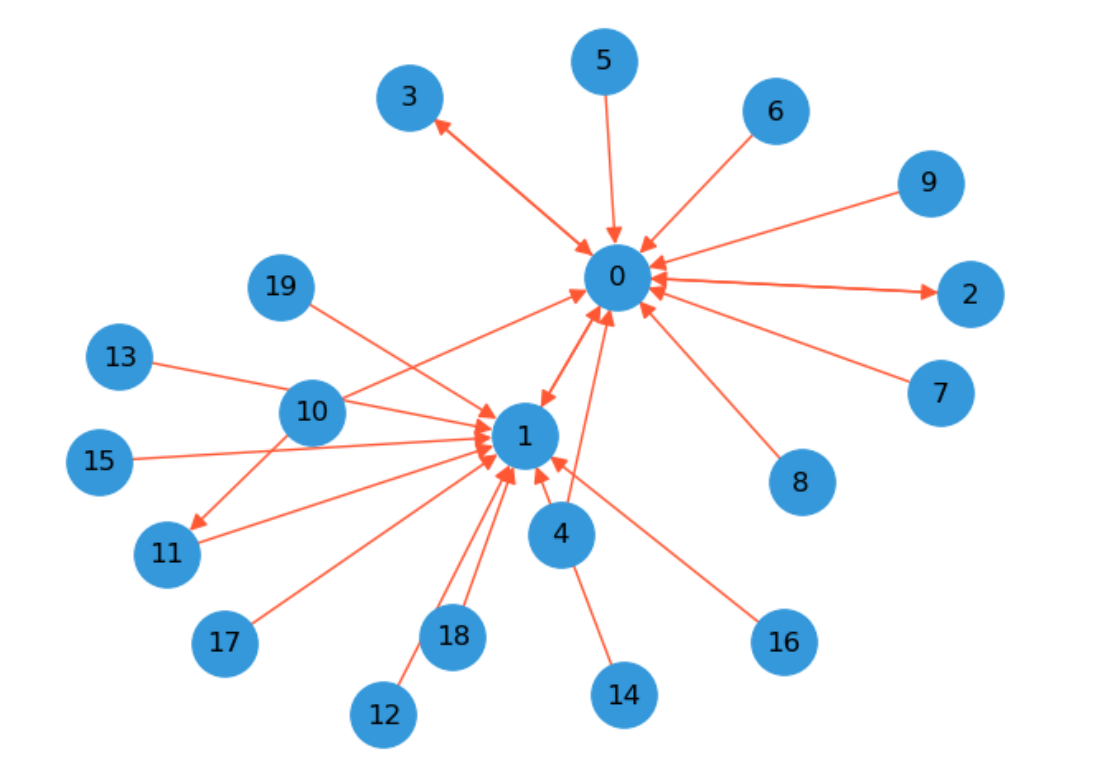
***G = nx.DiGraph(edges)***

***# visualize the graph***

***nx.draw(G, with\_labels=True, arrows=True, arrowstyle='-|>', arrowsize=15,***

***node\_color='#3498DB', node\_size=800,***

***edge\_color='#FF5733')***

******

***mutual\_followers = []***

***# Xét tất cả các cạnh trên đồ thị***

***for u, v in G.edges():***

***# Nếu có kết nối u->v và v->u thì (u,v) là cặp theo dõi lẫn nhau***

***if G.has\_edge(u,v) and G.has\_edge(v,u):***

***# Tránh việc liệt kê 1 cặp tài khoản 2 lần***

***if (v, u) not in mutual\_followers:***

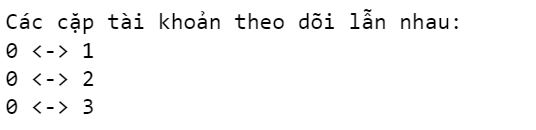
***mutual\_followers.append((u, v))***

***# List mutual follower pairs:***

***print('Các cặp tài khoản theo dõi lẫn nhau:')***

***for u, v in mutual\_followers:***

***print(u, '<->', v)***

******