Root crawl sẽ là id của ManUtd.

Các bước:

1. Crawl tất cả các tweet của ManUtd trong tháng 4 ( được khoảng 700 tweet)
2. Crawl tất cả các user đã retweet của ManUtd, mỗi tweet lấy 20 user, sau khi loại trùng lặp được khoảng 4600 user. -> lưu vào level 1
3. Nếu làm tương tự 2 bước trên đối với từng user trong level 1 không khả thi, vì các user phần lớn không phải người nổi tiếng -> ít lượt retweet hoặc thậm chí là không có -> Thu thập các retweet của các user này (15 retweet/user), sau đó lấy level 2 là các user là tác giả gốc của các retweet này ( có nghĩa là user ở level đã retweet các tweet của user level2 này).

Sau khi loại trùng lặp ta được khoảng 45000 node đồ thị lưu ở file graph.text

Tất cả các file kết quả lưu ở thư mục data

Để chạy file được trên máy khác nên thay đổi soure\_file ( đường dẫn)