Identifying The Influential Users Of A Location In Twitter

Sara Aein **Farzaneh Mahmoudi**

Maryam Sadat Hashemi

Iran University of Sience and Technology
Department of Computer Engineering

Social media usage has increased in the last decade and it is still continuing to grow. Companies, data scientists, and researchers are trying to infer meaningful information from this vast amount of data. One of the most important target applications is to find influential people in these networks. This information can serve many purposes such as; product recommendation, public opinion analysis, voting. In this study, a network-based method on twitter is proposed to find influential users in a location including countries and cities.

1 Problem Statement

In this section, we explain definitions and problems.

Definition 1. (user) user is defined to person who is signed up in Twitter. For unified notations, let u_i be a user, and U be the set of all users.

Definition 2. (users of a particular location) U_l denotes the users of a particular city or country and it is a subset of U.

Definition 3. (graph of Twitter network) We correspond Twitter network to directed graph $G(V,E)$. Each user $u_i \in V$ in U_l set is a vertex in graph $G(V,E)$. There is a directional edge $e_{ij} \in E$ from the u_i to u_j if user u_i follows user u_j.

Given a set U and its subset U_l and the directed graph $G(V,E)$, we aim at designing an Algorithm to find influential users in $G(V,E)$.

2 Method

In this section, first we explain about how we generate an appropriate dataset. Next, we detail two simple but efficient algorithms: Betweenness Centrality algorithm(Naive) and.

2.1 Dataset

We consider generating 50 graphs as a test case with the different number of vertex which is gradually increasing. The number of vertexes At least is 5 and at most is 1000. Then we assign a random number to each vertex v_i as the degree of vertex which illustrate the number of edges connect to v_i. With regard to the degree of each vertex, we choose randomly another vertex v_j to be the end of the edge e_{ij}.

2.2 Betweenness Centrality algorithm(Naive)

In graph theory and network analysis, indicators of centrality identify the most important vertices within a graph.

betweenness centrality is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex.

we inspire from the idea of betweenness central-

ity to address this problem. next, it is enough to find a vertex with maximum betweenness centrality in the graph as an influential user.

Algorithm 1 Delete item from Hash Table (with mCBF and aCBF)

```
1: procedure DeleteItem(item x)
2:     for i = 1 → k do
3:         if mCBF.C_{f_i(x)%N} == 0 then
4:             return
5:     for i = 1 → k do
6:         if CBF_query(aCBF_i, x) == TRUE then
7:             j ← C_{f_i(x)%N}
8:             for item y ∈ B_j do
9:                 if y.key == x then                ▷ found
10:                    remove y from hash table
```