CSE 598 - Project 1

Social Media Mining

Twitter Data Analysis

By:

Nitesh Kedia

ASU ID: 1206342939

**P0**

* This project is about Analysis of Twitter Networks by traversing the “Following” i.e. Out-degrees.
* BFS traversal approach has been used to crawl.
* 5000 Nodes have been visited with sampling done to get 1000 edges uniformly if the numbers of edges are greater than 1000.
* Dataset has a total of 2276957 edges and 1172297 nodes.
* MATLAB is used as the language to crawl which uses an API called twitty.m.

**Deliverables**

1. non\_annonymized\_edge\_List.csv [P0\non\_annonymized\_edge\_List.csv](P0/non_annonymized_edge_List.csv)
2. mapping\_Table.csv
3. [P0\mapping\_Table.csv](P0/mapping_Table.csv)
4. annonymized\_edge\_List.csv [P0\annonymized\_edge\_List.csv](P0/annonymized_edge_List.csv)
5. TwitterCrawl.m
6. [P0\TwitterCrawl.m](P0/TwitterCrawl.m)
7. TwitterAnalysis.cpp SNAP program used to do P1,P2,P3 measures
8. [P0\TwitterAnalysis.cpp](P0/TwitterAnalysis.cpp)

P1

* Power Law Degree Distribution

[P1\PowerLaw.jpg](P1/PowerLaw.jpg)

* Log-Log Plot of Power Law

[P1\loglog.jpg](P1/loglog.jpg)

* Power Law exponent value is 0.6
* Number of Bridges = 921417

[P1\Bridges.txt](P1/Bridges.txt)

* Number of 3-cycles ignoring the direction = 2649678
* Diameter of the graph = 14
* Average Path Length = 4.00181

[P1\P1\_Measures.txt](P1/P1_Measures.txt)

P2

* Average Local Clustering Coefficient = .0261 and is calculated using MATLAB program
* [P2\clustering\_coefficients\_program.m](P2/clustering_coefficients_program.m)
* Global Clustering Coefficient is calculated Using SNAP library and its value is 0.0387352
* Top 3 Page Rank values are calculated using SNAP
  + 1st : 2.13346e-006
  + 2nd : 8.59598e-007
  + 3rd : 8.65355e-007
* Top 3 Eigen-Vector Centralities values are calculated using SNAP
  + 1st : 0.00802788
  + 2nd : 0.006776
  + 3rd : 0.00124344
* Since we are doing sampling on out degrees, Maximum out degree is 1000 for those having the original out degree>1000.

[P2\P2\_measures.txt](P2/P2_measures.txt)

P3

* Random Graph is simulated using SNAP with number of nodes 5000 and average degree as 4
  + Average Path Length : 7.09642
  + Clustering Coefficient : 0.0006
  + Degree Distribution

[P3\RandomGraph.jpg](P3/RandomGraph.jpg)

* Small World Graph is simulated using SNAP with number of nodes 5000 and average degree as 4 and Beta = 0.5
  + Average Path Length : 7.93692
  + Clustering Coefficient : 0.553583
  + Degree Distribution

[P3\SmallWorld.jpg](P3/SmallWorld.jpg)

* Preferential Graph is simulated using SNAP with number of nodes 5000 and average degree as 4
  + Average Path Length : 3.66392
  + Clustering Coefficient : 0.0118301
  + Degree Distribution

[P3\PrefrentialGraph.jpg](P3/PrefrentialGraph.jpg)

[P3\P3\_measures.txt](P3/P3_measures.txt)

Comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Twitter | Random | Small World | Preferential |
| Average Path Length | 4.00181 | 7.09642 | 7.93692 | 3.66392 |
| Clustering Coefficient | 0.0387352 | 0.0006 | 0.553583 | 0.0118301 |
| Degree Distribution | Power Law | Poisson distribution | Poisson distribution | Power Law |