Construction of a Unified Graph Representation from Multiple Views

Stream Project

B.Tech PART – III

Submitted By:

Isha Agarwal 14075063

Under the guidance of

Dr. R.S.Singh

Department of Computer Science & Engineering

Indian Institute of Technology (Banaras Hindu University)

Varanasi – 221005 Uttar Pradesh

Certificate

This is to certify that the work in this report entitled "**Construction of a Unified Graph Representation from Multiple Views**" submitted by **Isha Agarwal(14075063)** in partial fulfillment of the requirements of the degree of **Bachelor of Technology** in **Computer Science and Engineering**, 2014-2018, in the Department of Computer Science and Engineering, **Indian Institute of Technology, Banaras Hindu University, Varanasi** (IIT-BHU, Varanasi), is an authentic work carried out by her under my supervision and guidance.

To the best of my knowledge the report embodies the literature from various reputed resources and has not been submitted to any other University/Institute for the award of any degree.

Dr. R.S.Singh
Department of Computer Science and Engineering
IIT(BHU), Varanasi

Dated:_____

Candidate’s Declaration

I hereby, declare that the work embodied in this project report submitted in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering is my own bonafide work carried out by me under the supervision of Dr. R.S.Singh, Computer Science and Engineering, IIT (BHU), Varanasi.

I declare that I have faithfully acknowledged, given credit to and referred to the research workers wherever their works have been cited in the text and body of the report.

Isha Agarwal

14075063

Dated:
Place : Varanasi

Acknowledgement

I would like to convey our deepest gratitude to Dr. R.S.Singh whose guidance, motivation, support and suggestions were a reason for this project becoming success.

Thank you so much, Sir.

I would also like to thank our friends for their support and suggestions without which completion of this project would not be possible.

Isha Agarwal

14075063

B.Tech Part – III

Department of Computer Science and Engineering

IIT(BHU), Varanasi

Index

1. Abstract 6

2. Introduction 7

3. Concepts Used 8

4. Algorithm Used 9

5. About Dataset 10

6. Implementation & Result 12

7. Observation 19

8. References 20

1. Abstract

In social networks, one user is related to many other users. A user may be related to a group of users on basis of different grounds(also called views here). Combining all such type of data into one poses a problem. This project uses a project combine such different views to find the set of k nearest neighbours for a particular user. And construct a single graph after combining such different views.This id done on multiview Twitter dataset.

2. Introduction

A social network is modeled by a graph, where the nodes represent individuals, and an edge between nodes indicates that a direct relationship between the individuals [3]. This graph is to be constructed using multiple views. One graph can be made per view. It is preferable to work with one graph only which can represent all the views. Multiple views are integrated to construct one graph. This is done by generating the ranked neighbour sets for each individual for each view and then constructing a nearest neighbour graph from the local neighbour sets.

3. Concepts Used

1. Neighbour Set Identification for each user – Generated a ranked neighbour set of each user by combining ranked neighbour set of each view

2. SVD (Sngular Value Decomposition) using Numpy library in python – SVD is an aggregation method. It is a well established technique for projecting high dimensional data into a lower dimension space. It produces an aggregated ranking.[4]

3. Construction of Graph Using networkx library of python – It is library in python which can be used to construct graphs in python.

4. Cosine similarity – It is a similarity measure used to find how much two users are similar.

4. Algorithm Used

This algorithm is taken from the research paper by Derek Greene and Pádraig Cunningham ,“Producing a Unified Graph Representation from Multiple Social Network View “.
The first phase of the aggregation process is as follows, for each user u i :

1. For each view j = 1 to l, compute a similarity vector v ij between u i and all other users present in that view, using the similarity measure provided for the view.

2. From the values in v ij , produce a rank vector of all other (n−1) users relative to u i , denoted r ij . In cases where not all users are present in view j, missing users are assigned a rank of (n 0 j + 1), where n 0 j is the number of users present in the view.

3. Stack all l rank vectors as columns, to form the (n − 1) × l rank matrix R i ,and normalise the columns of this matrix to unit length.

4. Compute the SVD of R i , and extract the first left singular vector. Arrange the entries in this vector in descending order, to produce a ranking of all other (n − 1) users. Then select the k highest ranked users as the neighbour set of u i .

5. Construct the graph.

5. About Dataset

3 Twitter Datasets are used.

Football : A collection of 248 English Premier League football players and clubs active on Twitter. It corresponds to the 20 individual clubs in the league.

Olympics: A dataset of 464 users, covering athletes and organisations that were involved in the London 2012 Summer Olympics. Users belong to to 28 different sports.

Politics-ie: A collection of 348 Irish politicians and political organisations, assigned to 7 groups, according to their affiliation.

Each dataset has 9 views :
1. tweet content: User content profiles, constructed from the concatenation of the 500 most recently-posted tweets for each user.

2. list text: List content profiles, constructed from the concatenation of both the names and the descriptions of the 500 Twitter lists to which each user has most recently been assigned.

3. follows: From the unweighted directed follower graph, construct binary user profile vectors based on the users whom they follow (i.e. out-going links).

4. followed-by: From the unweighted directed follower graph, construct binary user profile vectors based on the users that follow them (i.e. incoming links).

A pair of users are deemed to be similar if they are frequently “co-followed” by the same users.

5. mentions: From the weighted directed mention graph, construct user profile vectors based on the users whom they mention.

6. mentioned-by: From the weighted directed mention graph, construct binary user profile vectors based on the users that mention them. A pair of users are deemed to be similar if they are frequently “co-mentioned” by the same users.

7. retweets: From the weighted directed retweet graph, construct user profile vectors based on the users whom they retweet.

8. retweeted-by: From the weighted directed retweet graph, construct user profile vectors based on the users that retweet them. Users are deemed to be similar if they are frequently “co-retweeted” by the same users.

9. co-listed: We only consider the 500 lists to which each user has been assigned most recently assigned.

6. Implementation & Result

Implementation -

1. After generating the ranked neighbour set for each user, we construct the graph.

2. Networkx library is imported.

3. Each user is represented by a node.

4. An edge is drawn between the user and the users which are there in the ranked k-nearest neighbour set.

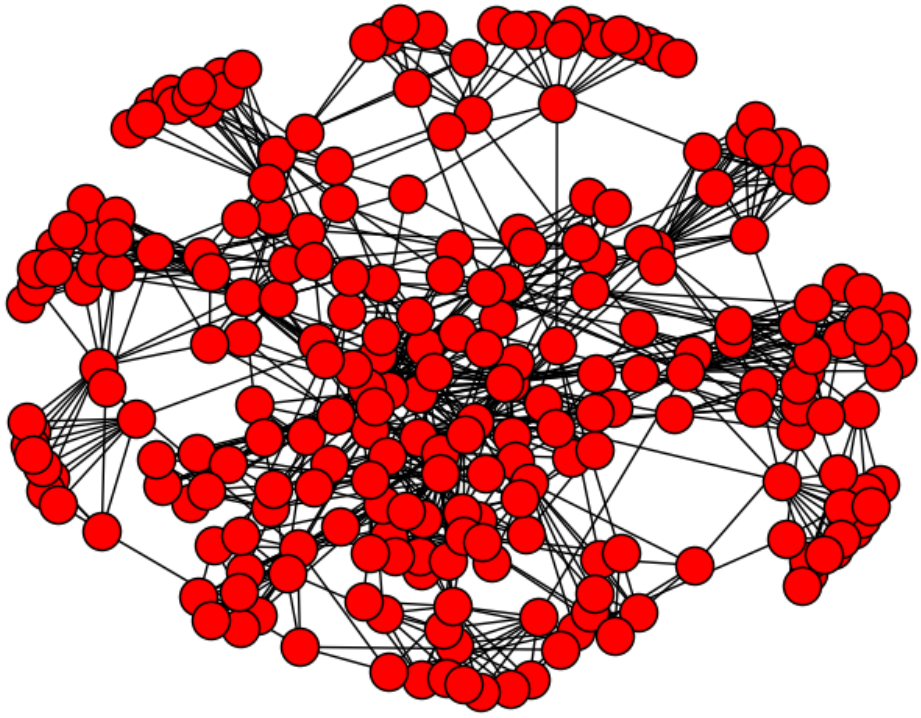
5. Graphs are constructed for 5 and 15 nearest neighbours for each community.

6. The link to the Github repository for the code is :

https://github.com/IshaAg/Constructing_Unified_Graph_Representation_from_Multiple_Views-.git

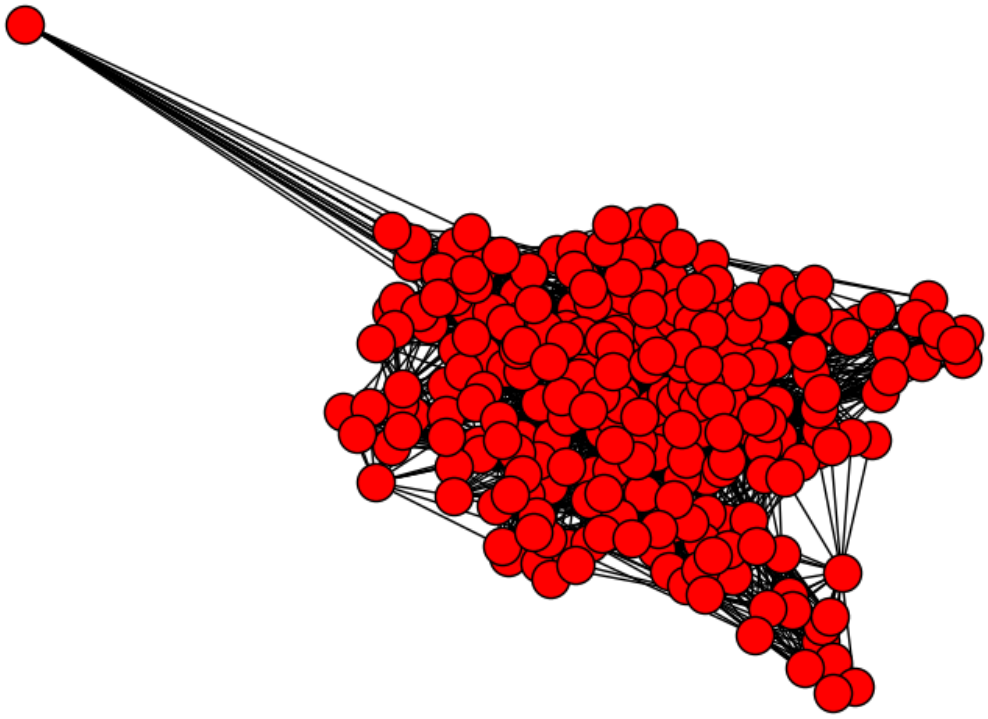
Football:

5 nearest neighbour users.



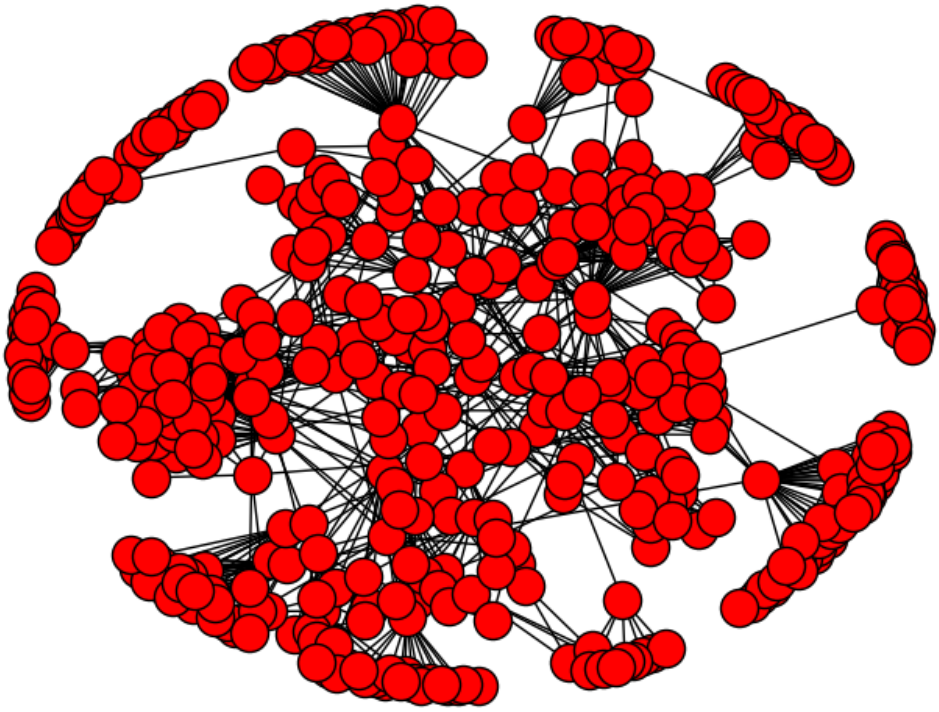
Football:

15 nearest neighbour users.



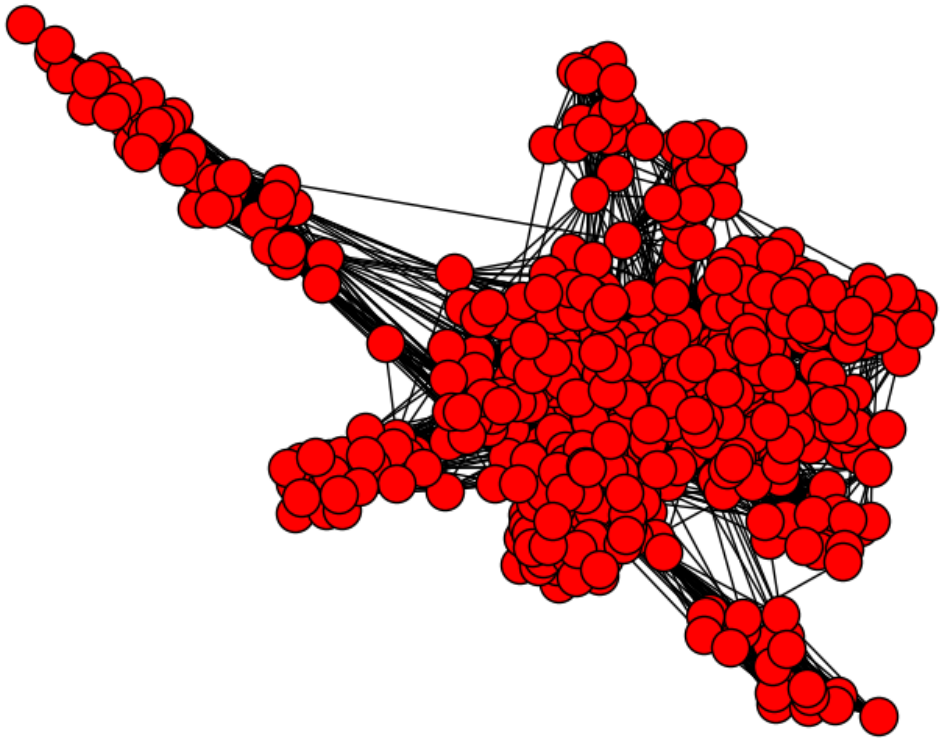
Olympics:

5 nearest neighbour users.



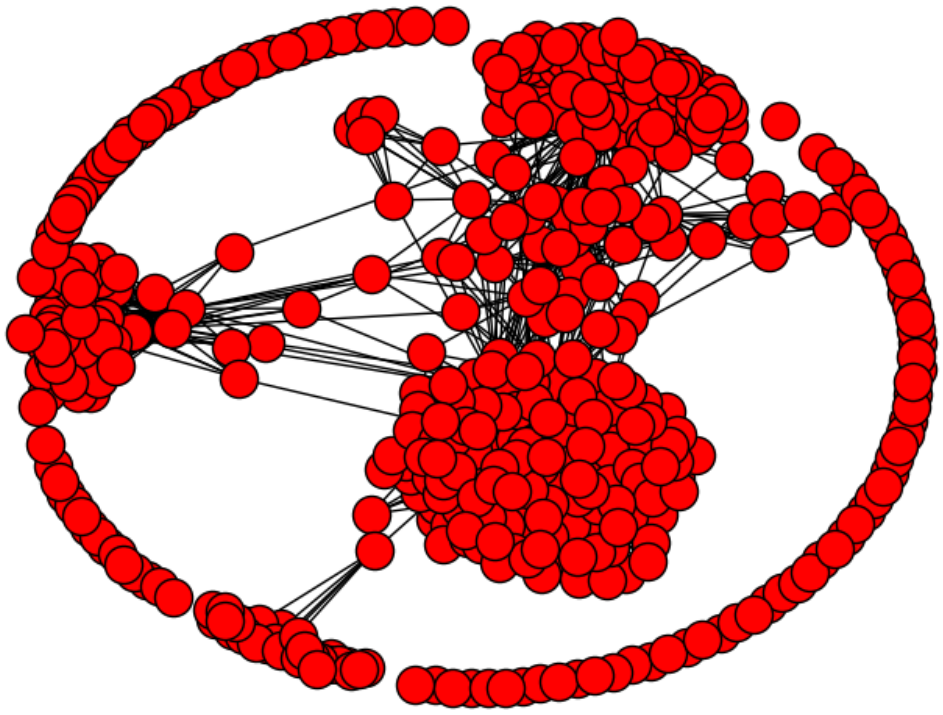
Olympics:

15 nearest neighbour users



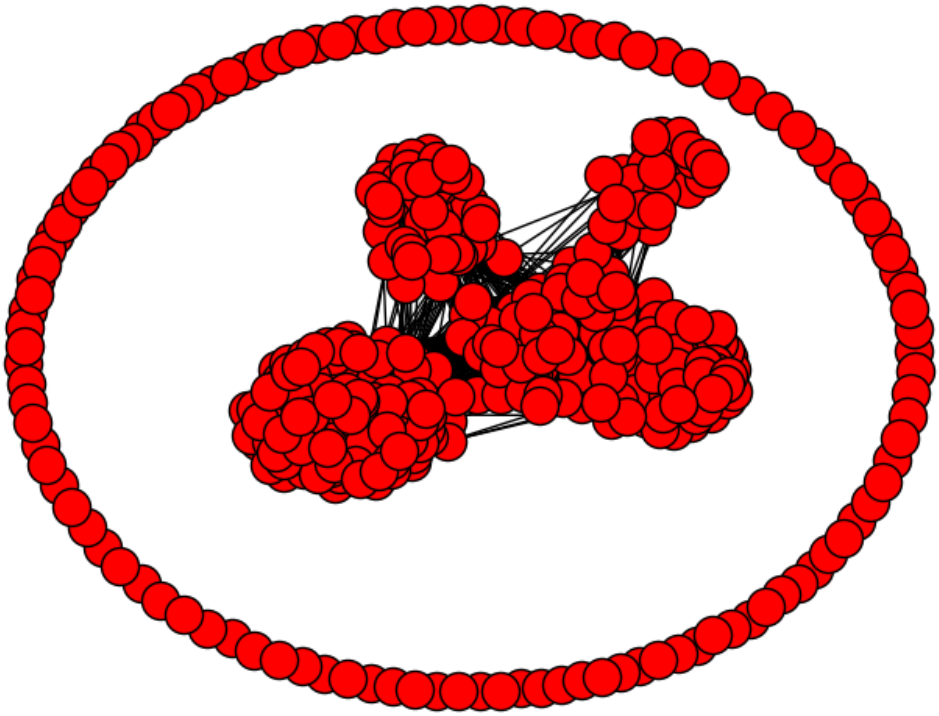
Politics-ie:

5 nearest neighbour users.



Politics-ie:

15 nearest neighbour users



7. Observation

The constructed graph is spare for k=5 nearest neighbours and the constructed graph is dense for k=15 nearest neighbour users.

8. References

[1] Derek Greene and Pádraig Cunningham. Producing a Unified Graph Representation from Multiple Social Network View.

[2] http://mlg.ucd.ie/networks (last accessed on 19 Oct 2016)

[3] D. Cai, Z. Shao, X. He, X. Yan, and J. Han. Mining hidden community in heterogeneous social networks. In Proc. 3rd International Workshop on Link Discovery,pages 58–65, 2005.

[4] G. Wu, D. Greene, and P. Cunningham. Merging multiple criteria to identify suspicious reviews. In Proc. 4th ACM Conference on Recommender Systems (Rec-Sys'10), pages 241–244. ACM, 2010.

[5] C. Ding and X. He. K-nearest-neighbor consistency in data clustering: Incorporating local information into global optimization. In Proc. ACM Symposium on Applied Computing (SAC'04), pages 584–589, 2004.

[6] S. Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010.