Construction of a Unified Graph Representation from Multiple Views

ISHA AGARWAL
14075063

Outline

- About the Dataset
- Purpose of the Project
- Introduction
- Concepts Used
- Steps followed
- Implementation & Result
- Constructed Graphs
- Observations
- References

DATASET -

- 3 Twitter datasets are being used for this project.
- Football : A collection of 248 English Premier League football players and clubs active on Twitter. The disjoint ground truth communities correspond to the 20 individual clubs in the league.
- Olympics: A dataset of 464 users 1 , covering athletes and organisations that were involved in the London 2012 Summer Olympics. The disjoint ground truth communities correspond to 28 different sports.
- Politics-ie: A collection of Irish politicians and political organisations, assigned to seven disjoint ground truth groups, according to their affiliation.
- Data taken from http://mlg.ucd.ie/networks .
- Each Dataset has 9 views : follows, followedby, mentions, mentionedby,retweets, retweetedby, listmerged, lists, tweets

Purpose of the Project -

- To produce a unified network representation from either feature-based or relational views on a set of social network users. SVD rank aggregation is applied to a matrix encoding multiple nearest neighbour sets for each user.
- The resulting aggregated per-user rankings are then combined to form a global graph covering all users.
- This sparse graph represents a unified summarisation of the strongest connections between users across all views.

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.

Concepts Used -

- Neighbour Set Identification for each user – Generated a ranked neighbour set of each user by combining ranked neighbour set of each view.
- 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
- Construction of Graph Using networkx library of python – It is library in python which can be used to construct graphs in python.
- Cosine similarity – It is a similarity measure used to find how much two users are similar.

Steps Followed -

This algorithm is taken from the research paper by Derek Greene and Pádraig

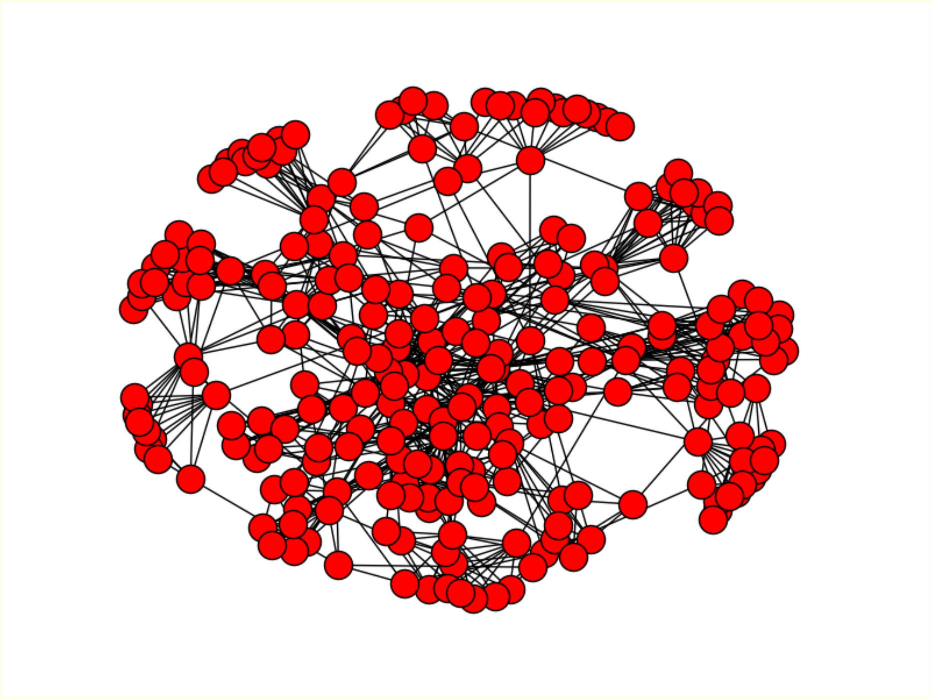
Cunningham ,"Producing a Unified Graph Representation from Multiple Social Network View.".

- 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.
- 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 (nj + 1), where nj is the number of users presentin the view.
- 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.
- 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 .
- Draw the graph using those k highest ranked users as the neighbour set.

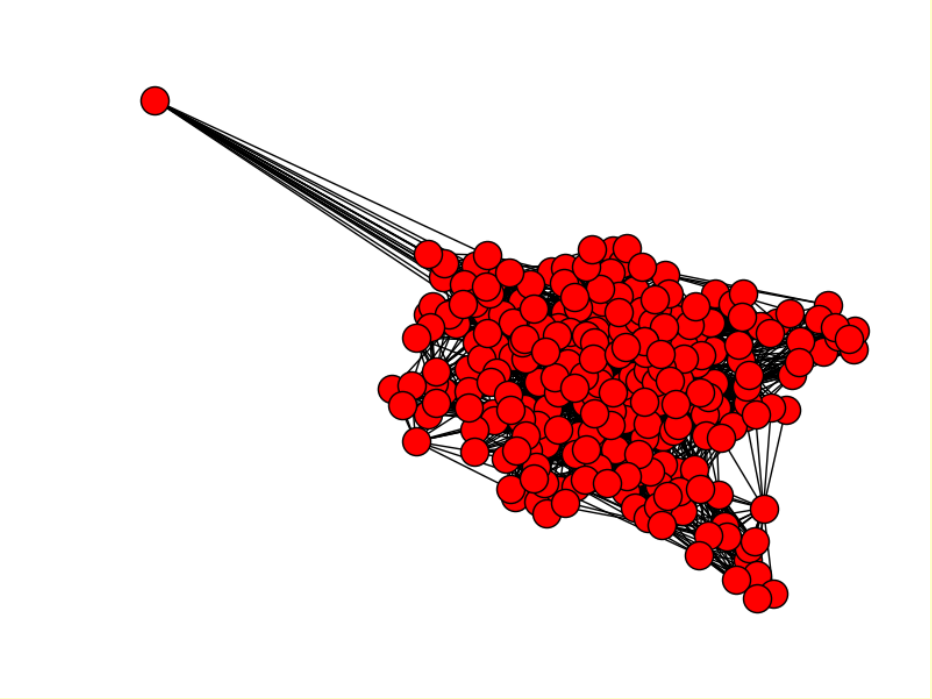
Implementation & Result

- After generating the ranked neighbour set for each user, we construct the graph.
- Networkx library is imported.
- Each user is represented by a node.
- An edge is drawn between the user and the users which are there in the ranked k-nearest neighbour set.
- Graphs are constructed for neighbours for each community for 5 and 15 nearest neighbours.
- The link to the github repository for code is : https://github.com/IshaAg/Constructing_Unified_Graph_Representation_from_Multiple_Views-.git

Graph for Football Community

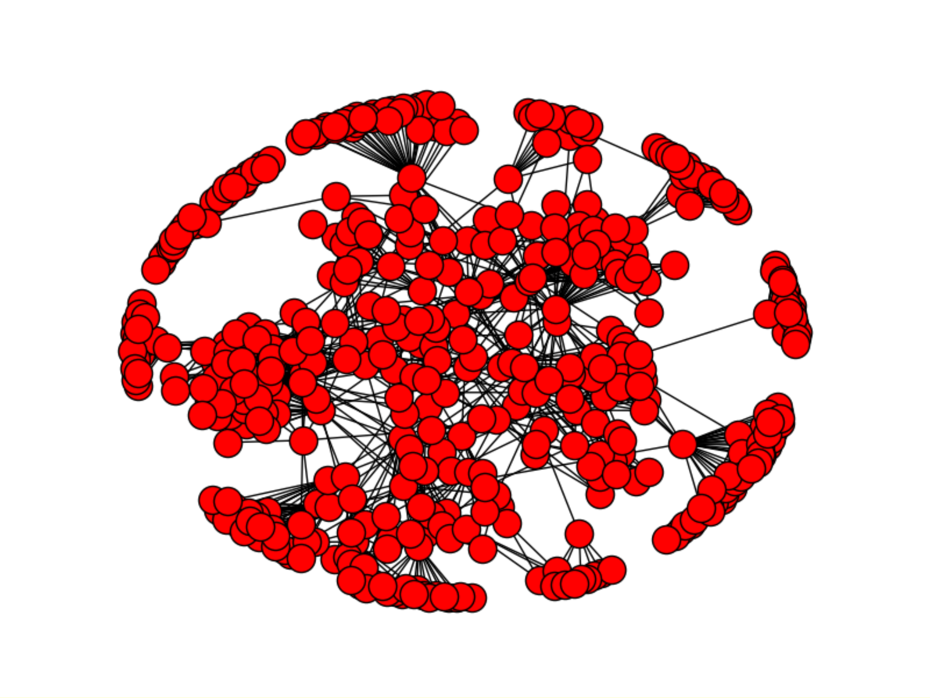


For 5 nearest neighbour users

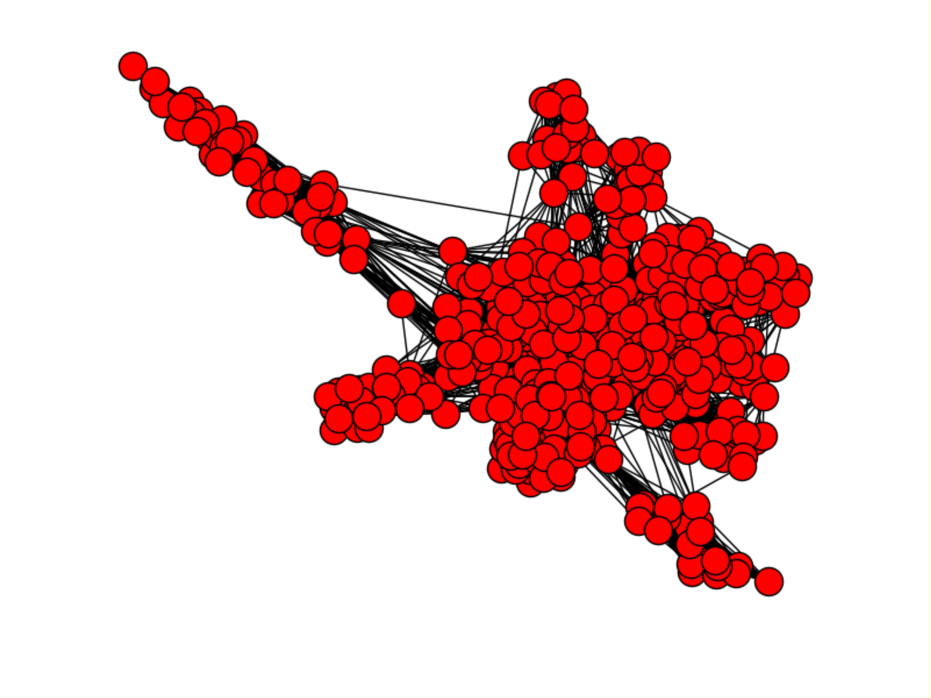


For 15 nearest neighbour users

Graph for Olympics Community

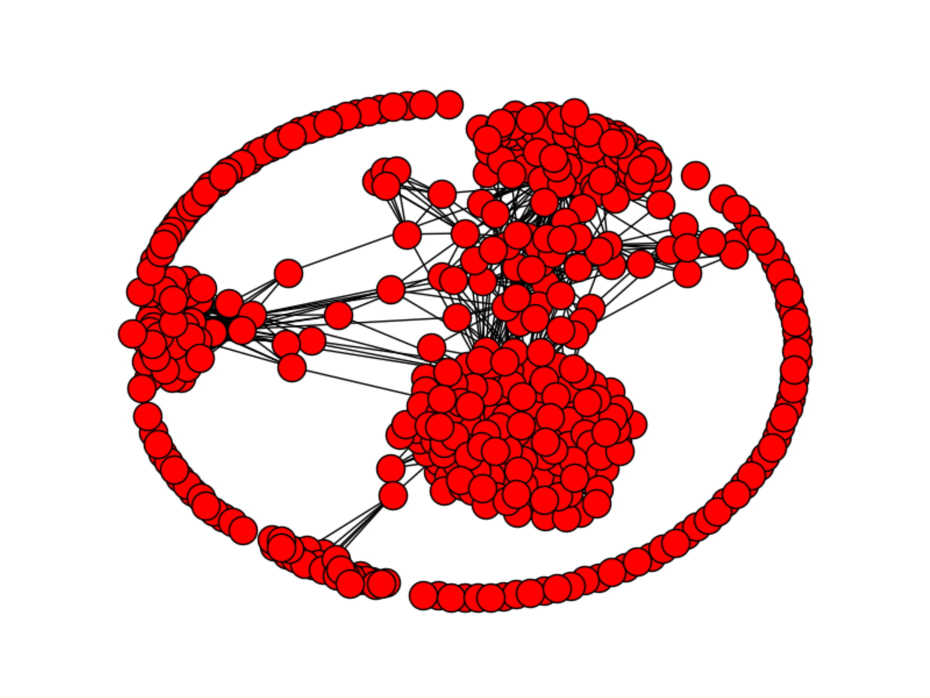


For 5 nearest neighbour users

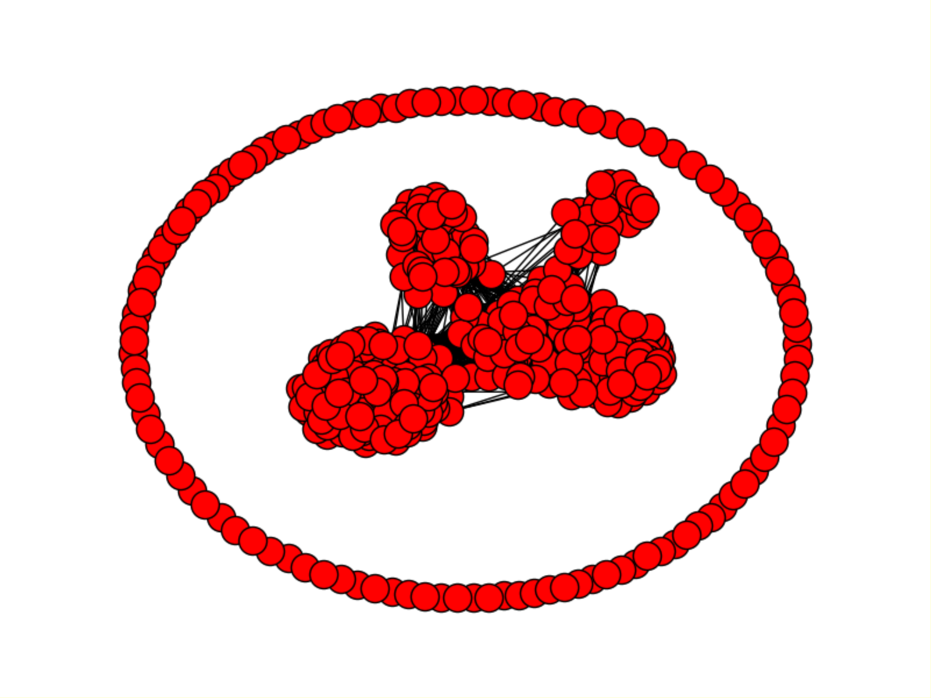


For 15 nearest neighbour users

Graph for Politics-ie Community



For 5 nearest neighbour users



For 15 nearest neighbour users

Observations

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

References

- Derek Greene and Pádraig Cunningham. Producing a Unified Graph Representation from Multiple Social Network View.
- http://mlg.ucd.ie/networks (last accessed on 19 Oct 2016)
- 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.
- G. Wu, D. Greene, and P. Cunningham. Merging multiple criteria to identify suspicious reviews. In Proc. 4th ACM Conference on Recommender Systems (RecSys'10), pages 241–244. ACM, 2010.
- 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.
- S. Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010.

Thank you