Community detection and labeling in an Instagram Network

Ahmed Furkan Ozkalay, Eda Bayram, Orçun Gümüs, Osman Berk Satir

November 28, 2017

Our objective is to detect communities in an Instagram Network using spectral clustering. The clusters can also be obtained and compared with the other graph based clustering methods mentioned in [1], such as Fast Greedy Modularity Optimization. Following the detection of communities, we aim to assign labels by benefiting the endpoint information attached with the users. Here, we need to pay attention that the inferred labels are so discriminatory and representative that the assigned labels should be different for each community but they should encompass the sufficient portion of the community agents at the same time. We expect that this step justify our clustering result and permit us to designate the clusters as "communities". A possible pipline can be given as follows:

1. **Data acquisition:** We acquire some users using Instagram API (A library for Instagram API can be found here). We can start by choosing a location in Lausanne, and gather some users who added this location in their recent photos. We create a network by adding some other users who liked these photos, and extend it by pursuing the "like" relations, hence, we simultaneously acquire users and determine the relationship between them. Later, we remove the users we gathered in the last hop (outer shell of the network) to be sure that we keep the inner connections between users.

2. **Data Exploration:** In a data frame, we store some endpoint information of acquired users, such as, bio of users or tags and locations that are attached with their photos. The endpoints we can access are given here. We intend to leverage this data frame later for labeling the clusters we obtain.

3. **Data Exploitation:** We employ spectral clustering algorithm on the network we created to detect the communities. We aim to label each community by taking advantage of the data frame we stored. For a community, we try to pick a label that explains the states of the users belonging to it at best in terms of the correlation of the label with the group. We can also consider to pick a user that shows a hub property in the community, if there is one.

In the above framework, we propose to construct our weighted graph using the "like" relationships. However, relations in a social network have multiple nature, in other words, we can construct different layer of graphs for the same set of users by focusing on the other endpoints, i.e., locations, tags, bio etc. Merging these layers may yield other communities whose users present some other type of common interest. For that purpose, we can apply a multi-modal spectral clustering method proposed in [2]. Then, we can compare single layer and multi-layer spectral clustering results.

Ps. Our group member Orçun may present some visualizations for data we acquire in this project to be evaluated in their project in Benzi's Data Visualization course. We will explicitly state which visualizations they provide us.

References

[1] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

[2] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov. Clustering on multi-layer graphs via subspace analysis on grassmann manifolds. *IEEE Transactions on signal processing*, 62(4):905–918, 2014.