You Are Who You Know: Inferring User Profiles In Online Social Networks

Shaurya Rajat Dewan
2019101017
CSE, UG2k19

Independent Study Application Task
Repository URL:

Overview

- Aim is to infer a user's profile attributes even if hidden using the user graph and other user's revealed attributes
- Premise is that users/people generally interact more with like-minded people with similar attributes
- Moreover, stronger communities are formed among people having at least 1 common attribute

Peer Reviewer Role

Positives

- Basic premises are promising and reasonable
- Detailed data analysis done to confirm the validity of both promises
- Well observed that the 2nd premise is not necessarily true just because 1st one is

Positives (Continued)

- Good catch to use a variation of conductance as a metric
- Comprehensive evaluation and comparison of local approach

Shortcomings

- Datasets are biased
- Did not run global approach on 1 dataset
- Did not compare global approach with other approaches
- No comparison done between modularity and conductance
- Time complexity is a concern

Suggestions

- Could use other data (if available) such as timestamps, etc. to apply weights to graph edges
- Could try a more localised approach by skipping some vertices/communities in each iteration of expanding the community

Systems Designer Role

Applications

- Most applications involving some form of graph with some attributes/properties for each node
- Use cases involving grouping and recommender systems
- Auto-filling of profile details
- Social networks, streaming services, e-commerce, search

Issues

- Privacy concerns
- Lack of strong guidelines and rules for such implicit data
- Could be misused just like explicit data
- Such systems could be misused for inferring information like political and religous alignments, etc.

Implementation

- Will work better for stronger/more connected graphs and communities
- Practically usable mostly in subnetworks and smaller applications
- Can apply weighting on inferred attributes based on accuracy

Future Work

- Goes a long way in improving clustering and recommender systems
- Main implementation issue to be addressed is robustness for sparse data
- Currently infeasible for larger graph applications due to time complexity

13

Hacker Role

Code and Analysis Ideas

- Implementing the global approach for a subset of the New Orleans dataset
- Comparing modularity and conductance as metrics since both have very similar logic
- Comparing global and local approaches with the same metrics

Important Links

- [Dummy dataset](#)
- [Implementation repo](#)
- [Google Drive folder](#)
-

Doubts

THANK YOU