Joseph Heenan and Zainul Abi Din

1 Project Description

Our team has chosen to create an application which will assist in running a model to detect fake users, groups, pages and posts on the Facebook social media platform. It is motivated by recent research showing successful detection of large numbers of fake social media accounts [1]. There have also been reports of "paid likes" plaguing social media networks [2]. This application will assign a probabilistic "realness score" (e.g., a "truth likelihood" score) to Facebook users, groups, pages and posts based on the below general business logic.

1.1 Entity Sets, Relationship Sets, Business Rules

1.1.1 Content_Creator

This serves as the root of the class hierarchy. It is sublclassed into three subclasses. The primary key is id. Other attributes are name, about, email, is_verified, posts, profile_picture, contact_address.

1.1.2 User

This is a subclass of Content_Creator The primary key is id. Other attributes are friends_count, likes, groups, member_since.

1.1.3 Page

A page has an owner of type User. The primary key is id. Other attributes are fan_count, founded, overall_star_rating, category, were_here_count, location, owner. As a business rule, each page must be owned by exactly one user

1.1.4 Group

A group may make posts. Its membership cannot be directly queried. The primary Key is id. Other attributes are member_count, update_time, created, privacy. As a business rule each group must be owned by exactly one user.

1.1.5 Post

A post is a document posted to the social media network. Its primary key is id. Other attributes are created_time, owner, name, message, shares_count, reactions, update_time, with_tags. As a business rule, a post must be created by exactly one Content_Creator.

Scoring Logic

First, the program will download data regarding the public profiles of recently-created users. Next, the program will calculate a "realness score" to each public user based on:

1. Whether or not the biographical information and metadata (e.g., work history and photo) is verifiable via a query using the Google Search API, and

2. The density of relationships that these user has to other users; realness score will be positively influenced by connections to others users with a high realness score, and negatively influenced in the inverse scenario

Next, the program will examine groups and pages that these users have joined or created, respectively. Groups containing with a large number of users with a low realness score will be transitively assigned a low realness score as will pages created by such users. Finally, the program will examine publicly-available posts. Posts will be assigned a realness score based on:

1. The transitive realness score of their associated author users, linked groups, and linked pages, and

2. Density of dissenting comments found via natural language processing of comments related to the post

Application UI

The application UI will:

1. Contain a button to begin a new data import, allowing a user to specify the number of entities to import (within the parameters of the Facebook API's Terms of Service) as well as the Facebook API key the researcher will use and

2. Show a summary report of realness scores assigned to each of these entities, and allow a user to examine each entity (e.g., user, page, group or post) more closely if desired, and finally:

3. Allow a user to input a URI of a post, user or group and view a realness score on-demand according to the model contained in the program's business logic

References

[1] Echeverr and Zhou, "The star wars botnet with 350k twitter bots," *ArXiv e-prints*, Jan 2017.

[2] D. B. Clark, "The bot bubble." https://newrepublic.com/article/121551/bot-bubble-click-farms-have-inflated-social-media-currency, April 2015.