Facebook research databases

ELTE_facebook_kutatas_2019.07._compl_20190930.sav
The .sav file of the survey

friend.sav
This file contains the FB friends (the hashed name of the friend is in the **friend_hash_code** column). During preprocessing, the gender of friends was predicted based on their names (**friend_gender;** 1-male, 2-female).

Dimension: 115,955 x 8

Columns: id, year, month, day, hour, min, friend_hash_code, friend_gender

reactions.sav
All user reactions with date (**year, month, day, hour, min**). The **target** includes who/what the participant reacted to. If the target is an FB friend, then his/her hash code is displayed. If it is a content on a page, then the name of the page is displayed. The **target_group** indicates whether the target was a friend or a page. Only pages that the user liked are listed. If the target of the reaction is not a liked page or a friend, then the field is empty. The **reaction_type** records what the user reacted to. The exact content is not visible, but the type of content is (e.g. image, post, video). The **reaction** variable contains the type of reaction.

Dimension: 1,802,430 x 10

Columns: id, year, month, day, hour, min, reaction, target, reaction_type, target_group

pages.sav
Pages (**page**) liked by the user with date (**year, month, day, hour, min**). If the page was also visible on the user's profile page, the **page_type** variable shows what page group FB classifies the page. If this variable is empty, it could be due to a character encoding problem, but it is more likely that the page has already been discontinued or the user has disliked it. In page names, emojis are in unicode format.

Dimension : 83,232 x 8

Columns : id, year, month, day, hour, min, page, page_type

ad_interest.sav

Interest list created by FB. If the area of interest begins with a lowercase letter, it is a more general category, if it is capitalized, it is a more specific one (eg: NoSalty → specific, friendship → more general).

Dimension: 105,642 x 2

Columns: id, ad_interest

ad_click.sav

The list of ads that the user has clicked on. The date of the click (**year, month, date, min**) is also included in the file. The **ad_click** variable shows what kind of advertisement the user clicked on.

Dimension: 9,674 x 7

Columns: id, year, month, day, min, ad_click

friend_meta.sav

The file contains summary information about the user's FB friends. In addition to the number of FB friends (**friend_num**), the gender (**friend_male, friend_female, friend_unknown_gender**) is displayed (based on estimates from names) as well as the number of invitations sent and received (**received, rejected, sent**) (without confirmation) and the number of removed friends (**removed**). The file contains the date the user made his/her first friendship (**first_friend_date_year, first_friend_date_month, first_friend_date_day**) and how many days this is away from October 1, 2019 (**first_friend_day_diff**). This variable should be used to determine how long the user has been an FB member.

Dimension: 150 x 13

Columns: id, friend_num, friend_male, friend_female, friend_unknown_gender, received, sent, removed, first_friend_date_year, first_friend_date_month, first_friend_date_day, first_friend_day_diff

fb_posts_proc.sav

Contains users' posting activity with dates (**year, month, day, hour, min**). The **type** variable contains what activity the user did (e.g. he wrote an entry → 1, updated his status → 99, etc.). The **person1** and **person2** variables show the people, the **group** variable the group, and the **page, game, music** variables show the pages, games, and music related to the the activity (e.g. "wrote on John Doe's timeline "→ John Doe → *hash code,* "posted an entry in group XY "→ group XY). The **text** variable contains the text that the user wrote for the activity (e.g., "Rated the XY game" → "I loved"). The **route** variable contains the path to the files associated with the post (if a file is attached). The **event** variable contains the name of the event associated with the post. The **link** variable contains the link of the domain that can be linked to the activity (e.g. played a game via FB → https://goodgames.com/randomgame). The **sum_comment** variable shows the total number of comments received for a given post, and the **tagnum** is the number of people tagged.

Dimension: 346,352 x 19

Columns: id, year, month, day, hour, min, text, route, sum_comment, tagnum, event, link, type, person1, person2, group, page, game, music

fb_posts_proc_trimmed.sav

The fb_posts_proc database without the **text, event,** and **route** columns.

Dimension: 346,352 x 16

Columns: id, year, month, day, hour, min, sum_comment, tagnum, link, type, person1, person2, group, page, game, music

fb_survey.sav

Questionnaire related to the research. One can merge with the other data tables using the **id** variable.

Dimension: 150 x 160

Columns: caseid, id, hash_code, ...

Files

List of .Rdata files.

Files_sav

List of .sav files.

friend_remove.sav

The dates of friend removals.

Dimension: 34,151 x 6

Columns: id, year, month, day, hour, min

post_comments.sav

The data table contains user comments posted on friends' timeline. The variables prefixed with **comment_** (**comment_year, comment_month, comment_day, ...**) show when the user commented; and the **year, month, day, ...** (without the **comment_** part) variables indicate when the post was created. The author of the post is contained in the **hash_code** variable, and the text of the comment is contained in **text**.

Dimension: 2,302 x 13

Columns: id, comment_year, comment_month, comment_day, comment_hour, comment_minute, year, month, day, hour, minute, hash_code, text

post_comments_trimmed.sav

The post_comments database without the **text** column.

Dimension: 2,302 x 12

Columns: id, comment_year, comment_month, comment_day, comment_hour, comment_minute, year, month, day, hour, minute, hash_code

post_timeline.sav

This database contains the activities on the user's timeline with dates (**year, month, day**). The **activity** variable shows the activity itself, and the **activity_author** shows who performed the activity (the friend's name is hashed), and the **activity_cat** show the category of the activity. The **text** variable contains the text related to the activity, and the **link** contains the attached link.

Dimension: 91,821 x 9

Columns: id, year, month, day, activity_author, activity, activity_cat, text, link

post_timeline_trimmed.sav

The post_timeline database without the text column.

Dimension: 91,821 x 8

Columns: id, year, month, day, activity_author, activity, activity_cat, link