Network Analytics
Group Assignment 2

INSTRUCTIONS

- This is a group assignment.
- Submit your answer digitally as two files through Moodle:
  - An R markdown file (extension **Rmd**). Use the template provided to you and provide your answers (both code and text) below each question.
  - An **HTML** file "knitted" by RStudio including all the results and plots. More details on how to create these files were provided in class on week 3.
- Follow the Style Guide (available on Moodle). You can be penalized on up to 20% in each question for which you do not follow the Style Guide.
- Questions regarding the assignment should be posted exclusively on the respective discussion forum on Moodle.

Deadline: Monday, March 10 at 23:59.

- Late submissions are not allowed

Warning: The detection of any form of plagiarism in your work means the assignment will be graded with ZERO points.

Dating Platforms

Online dating platforms enable people to find new connections with the goal of developing personal, romantic, or sexual relationships. In some platforms individuals express their interest in others in multiple ways, including by “liking” them. “Liking” someone is an indication of interest that, in most platforms, is revealed to both individuals only if the “liking” is mutual, i.e., if the target individual also stated interest. In such cases we say there is a match.

For this assignment we are going to analyze “likes” issued by members in an online dating platform using tools learned in class. Answer each of the questions below the best you can. In case you find any ambiguity in the question, assume the most sensible option in you opinion, state your assumptions and proceed. In most cases ambiguity is by design.

Data Description

The file `nda-dating-likes.RData` contains two `data.table`, each of them is described below.

`dt.users`

Field	Type	Description
`user_id`	integer	user identification
`inviter_id`	integer	id of the user that invited this user to the platform
`gender`	string	gender
`birth_year`	integer	year of birth
`education`	integer	education
`approved_week`	date	week in which this user was approved to the platform
`height`	integer	height (in cm)
`children`	boolean	whether the user has children
`smoker`	boolean	whether the user smokes
`n_fb_friends`	integer	how many Facebook friends does the user have

`dt.likes`

Field	Type	Description
`sender_user_id`	integer	user identification of the sender of the like
`receiver_user_id`	integer	user identification of the receiver of the like
`week`	date	week in which the like was issued

Setup and data loading

Start by loading the required libraries and loading the likes data

```
library(data.table)
library(ggplot2)
library(igraph)

load("nda-dating-likes.RData")
```

Questions

Invites Network `[10 points]`

This section contains questions related to the invites network.

1. *(Easy)* Build a directed graph representing the invites network: an individual A is connected to individual B if A invited B to the platform. What is the size of the longest chain? What is the clustering coefficient of this network? Justify. `[3 points]`
2. *(Easy, once you get what you need to do)* Goel, Watts and Goldstein, in their paper entitled "The structure of online diffusion networks", state that long cascades in diffusion networks happen very rarely in online networks. Do you think is also the case with the invites network in this online dating platform? Perform the analyses you deem necessary to answer this question with reasonable confidence. `[3 points]`
3. *(Moderate)* Do you think the invites network exhibits homophily? In other words, do you think men are more likely to invite other men and women more likely to invite other women? What about in terms of age? Do members tend to invite members of a similar age? Please describe your overall approach to answering these questions and justify and explain each step of your answer. `[4 points]`
4. *(Make up question: You can skip any of the previous questions and answer this question instead.)* Describe and perform any additional analysis you want **using the invites network**. You will be graded by the creativity and correctness of the analysis, but also as a comparison to the best answers provided by other students. The best answers will have full points, while the other answers will be judged by comparison with the best answers. `[3 points]`

Likes Network `[10 points]`

This section contains questions related to the likes network.

1. *(Easy)* Build a directed graph representing the likes network: an individual A is connected to individual B if A liked B. What is the clustering coefficient of this network? Justify. `[2.5 points]`

2. *(Moderate)* How many individuals "like" individuals of the same gender? Does this happen more between women or between men? `[2.5 points]`

3. *(Easy)* Build an undirected graph representing the matches network: an individual A is connected to individual B if A liked B and B liked A. What is the clustering coefficient of this network? Justify. Hint: the function `as.undirected` transforms a directed graph into an undirected graph; it can transform only the mutually directed edges in undirected edges. `[2.5 points]`

4. *(Moderate)* A recent research article gathered significant attention for claiming that, contrary to decades of prior research, most real life networks are not scale-free (i.e., the degree distribution does no not follow a power law). Based on the Quanta Magazine article (not on the original research paper) linked below, how do you classify the network of matches? Does it resemble a random network, a free-scale network, or a mixed network? Justify your answer. `[2.5 points]`

   Article: `http://bit.ly/2BwobR1`

5. *(Make-up question: You can skip any of the previous questions and answer this question instead.)* Describe and perform any additional analysis you want **using the any network originated from the likes data.** You will be graded by the creativity and correctness of the analysis, but also as a comparison to the best answers provided by other students. The best answers will have full points, while the other answers will be judged by comparison with the best answers. `[2.5 points]`