

RGG social feed proposal

1 Introduction

As stated in the RGG proposal, we are now filling the RGG with social data feed. Before starting development we carried out a survey on existing solutions and available statistics on Facebook. Among publications the most useful to use seems [1], and among statistics resources we mostly used [2], as well as official Facebook blog [3].

2 Model

2.1 Structure

For every node in RGG we will create three files *i_profile.txt*, *i_feed.txt* and *i_likes.txt* where *i* is the number of the node. *i_profile.txt* is a model of Facebook profile: for this moment it will contain full name and age of the person. *i_feed.txt* is a model of Facebook timeline: for this moment it will contain random text messages with generated message IDs. The message id will look like: *i.j*, where *i* is the user ID and *j* is the number of the message (starting from 0). *i_likes.txt* strings will have the following structure:

message ID: total number of likes User IDs who liked this message

Obviously, this file will contain as much strings as the number of messages of the user # *i*.

2.2 Data generation

Messages will be generated by the *nltk* Python library [4] using Markov chains. From our survey we found out that likes are also distributed by power law, so we will use our model to generate them. Names will be taken out from available trusted Facebook dumps. As far as ages are concerned, from Facebook statistics we learned that the most active ages are 18-45. So the connected core [5] of the generated graph will mostly consist of users with ages from this range, while indense part of the graph will mostly consist of older users.

2.3 Soundness

We suppose the chosen model to be sound enough since it is based on Facebook statistics and implements basic Facebook communication techniques such as text messages and likes.

References

- [1] A. Nazir, S. Raza, C.-N. Chuah. Unveiling Facebook: a measurement study of social network based applications. IMC'08.
- [2] <http://www.socialbakers.com/facebook-statistics/>
- [3] <http://blog.facebook.com/>
- [4] <http://nltk.org/>
- [5] J. Leskovec *et al.* Statistical Properties of Community Structure in Large Social and Information Networks. WWW, 2008.