|  |
| --- |
| **Stack exchange network analysis** |

**Miroslav KrálAgus Nur Hidayat**

MSc Business Analytics MSc Business Analytics

University College London University College London

*Miroslav.kral.15@ucl.ac.uk Agus.hidayat.15@ucl.ac.uk*

**Abstract**

The Abstract paragraph should be indented ½ inch (3 picas) on both left and right-hand margins. Use 10 point type, with a vertical spacing of 11 points. **Abstract** must be centered, bold and in point size 12. Two line spaces precede the Abstract. The Abstract must be limited to one paragraph.

1. **Introduction**
2. **Dataset description**

In this paper we work with online community data, specifically with data from sports stack exchange community. Stack exchange is a network of question answers web communities on various topics. One of the pages (communities) that belongs to stack exchange portfolio is page on sports.

Sports stack exchange is question - answer web page, that means that users can specify questions and other members of community can answer these questions. There can be multiple answers to one question, however, user, who specify question can then mark answer that satisfied his information need as accepted one. Users are allowed not only writing a posts, but comments too. Comments can be related either to question and answer. Stack exchange also implemented voting system that enables to all users to rate questions and answers (both together are referred as a posts) by plus and negative points. These points are then transformed into reputation of user who wrote the post and add a gamification element to the page. Great benefit of Stack exchange pages is that they are opened to everyone, not only for registered user.

We acquired a dataset that contains all the data about activity of the users of Sport stack exchange web page. This dataset is released periodically by Stack exchange through following link: <https://archive.org/details/stackexchange>. It contains all the data since beginning of this community (February 2012) until March 2016 (date of the last release).

Sports stack exchange community belongs to the smaller communities Stack exchange portfolio; therefore, the dataset is reasonable big. Community consists of **5018** users whose wrote a **2676** questions and **4353** answers. Data set consists of 5 tables. The most important tables are table Users and table Posts.

Table Users capture information about users of stack exchange web site. All the user related data, such as name, location of user, when joins community, when he last access page and other more. However, the most important feature of user is reputation that capture how are the user's contributions accepted by community.

Table Post capture all the post that users wrote. It contains text of the post, who wrote it, when wrote it, how many positive and negative voting post received and most importantly it contains feature whether it is a question or answer. In case it is a question it is also captured accepted answer (if exists) and in case of answer is denoted to which question belongs.

Other tables are called Comments, Tags and Votes. Detailed description of all tables and columns of the tables are presented in Appendix A.

1. **Question – Answer network**
2. **Question – Answer network in time**
3. **Conclusion**