Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: 0005-1144 (Print) 1848-3380 (Online) Journal homepage: http://www.tandfonline.com/loi/taut20

Exploratory analysis of pairwise interactions in online social networks

Luka Humski, Damir Pintar & Mihaela Vranić

To cite this article: Luka Humski, Damir Pintar & Mihaela Vranić (2017) Exploratory analysis of pairwise interactions in online social networks, Automatika, 58:4, 422-428, DOI: 10.1080/00051144.2018.1468162

To link to this article: https://doi.org/10.1080/00051144.2018.1468162

AUTOMATIKA, 2017
VOL. 58, NO. 4, 422–428
https://doi.org/10.1080/00051144.2018.1468162

Taylor & Francis
Taylor & Francis Group

SHORT PAPER

OPEN ACCESS Check for updates

Exploratory analysis of pairwise interactions in online social networks

Luka Humski, Damir Pintar and Mihaela Vranić

University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

ABSTRACT

In the last few decades sociologists were trying to explain human behaviour by analysing social networks, which requires access to data about interpersonal relationships. This represented a big obstacle in this research field until the emergence of online social networks (OSNs), which vastly facilitated the process of collecting such data. Nowadays, by crawling public profiles on OSNs, it is possible to build a social graph where "friends" on OSN become represented as connected nodes. OSN connection does not necessarily indicate a close real-life relationship, but using OSN interaction records may reveal real-life relationship intensities, a topic which inspired a number of recent researches. Still, published research currently lacks an extensive exploratory analysis of OSN interaction records, i.e. a comprehensive overview of users' interaction via different ways of OSN interaction. In this paper, we provide such an overview by leveraging results of conducted extensive social experiment which managed to collect records for over 3200 Facebook users interacting with over 1,400,000 of their friends. Our exploratory analysis focuses on extracting population distributions and correlation parameters for 13 interaction parameters, providing valuable insight into OSN interaction for future researches aimed at this field of study.

ARTICLE HISTORY
Received 21 December 2017
Accepted 19 April 2018

KEYWORDS
SNA; synthetic data; online social networks; Facebook; feature collection

Introduction and related work

A social network is a structure composed of nodes and edges which represent people and their relationships, such as family bonds, friendships, etc. Social network analysis (SNA) is a research field which deals with analysing such networks and extracting useful information about people described within, with the analysis being mostly focused on user interactions. There are numerous possible applications: by analysing social networks sociologists and social psychologists are trying to explain how people's thoughts, feelings and behaviours are influenced by presence of others [1,2]; recommender systems can use it to make customized and novel recommendations [3,4]; corporations are trying to improve relations between employees and their working effect [5–7]; telecoms want to prevent users churn [8–10]; in the educational domain information about connectedness between students may be used to enhance the learning process [11–13], etc.

Modern online social networks (OSNs) such as Facebook or Twitter are widely accepted as platforms for exchanging messages, sharing photos, links and other kinds of information. We can treat these OSNs as applications for social networks management. Due to their nature as digital platforms, information about connectedness and interaction between users is usually stored in a structured fashion and is becoming more accessible than ever, which has vastly facilitated the ability to observe social networks for research purposes. One of the basic methods of gathering OSN information is creating software which uses the OSN's API to crawl public profiles and construct a social graph based on publicly available "friendship" information contained within [14–16]. In that way it is possible to create a social graph with information whether two users are connected, but usually not the details about the nature or intensity of their real-life relationship. There are however some researches that introduce various models and algorithms which enable calculating friendship intensity and picking out real-life relationships from ego-users' total OSN friends by considering their interaction on OSN [17–26]. Some papers aim simply to differentiate between strong and weak friendships of the ego-user [17–19], others classify ego-user's friends in more than just two basic classes [20,21] while some aim to determine the connection strengths between all OSN users and express it in a numerical fashion [20,22–26].

Although OSN interaction records are frequently used as basis for various research purposes, so far a comprehensive exploratory analysis of users' OSN interaction has not yet been published. Taken this into consideration, we have decided to invest a great effort in collecting a representative real-life OSN interaction dataset, followed by performing an extensive exploratory analysis in order to extract and describe its key properties. As Facebook is arguably the most

CONTACT Mihaela Vranić mihaela.vranic@fer.hr University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

popular OSN today with over 2 billion active users [27], we decided to focus on this particular social network. We have conducted a comprehensive Facebook social experiment *NajFrend* where we collected records that describe interaction between almost a million and a half pairs of Facebook users. We have then performed an exploratory analysis where we focused on extracting population distributions and correlation matrices for 13 Facebook interaction parameters such as posts, likes, comments, mutual photos, etc. (which we will call *interaction parameters* in the following sections). All these parameters were collected and summarized on pairwise levels – e.g. total likes, total comments, etc. between pairs of Facebook friends. The results of this user interaction exploratory analysis based on huge empirical dataset represent the pivotal contribution of this paper.

The paper is organized as follows: in the *Methodology* section we provide details about the conducted social experiment, present the collected dataset and describe in detail the process of extracting population distributions and constructing the correlation matrix; the *Results* section contains tabular and visual results of the exploratory analysis; *Discussion* provides insight and interpretations of gained results; finally, in *Conclusion* we give final remarks on this research.

Methodology

This section will provide a brief description of the conducted social experiment *NajFrend* and the dataset collected in that experiment, which is a core dataset for our exploratory analysis. Also, we will explain the steps undertaken in the exploratory analysis itself.

Social experiment NajFrend and the collected dataset

NajFrend is a comprehensive social experiment held in April and May of 2015. It has involved 3277 examinees, mostly from Croatia and neighbouring countries. Majority of examinees were between 18 and 30 years old. Close to 80% of examinees were high school and university students. 57.7% of examinees were men and 42.3% were women. This experiment collected a dataset about interactions between 3277 examinees and over 1,400,000 of their Facebook friends. All examinees gave explicit permission to allow using collected data about their Facebook interaction for this research.

For the following exploratory analysis, we have chosen 13 Facebook interaction parameters to describe user interactions, whose list and explanations can be found in Table 1. Additionally, for each attribute in the table we have included an abbreviation which will be used in the certain following figures with insufficient space for the full attribute names.

Exploratory analysis

Main goal of our exploratory analysis was to analyse behaviour of the collected Facebook interaction parameters. We focused on extracting population distributions for each of the observed 13 interaction parameters and calculating *Pearson's correlation coefficient* for each pair of interaction parameters. For each distribution, we have provided a detailed quantile table and a theoretical distribution which has shown to be the best approximation for an empirical distribution of each interaction parameter. Since most Facebook users interact very little with a large portion of their Facebook friends, our dataset contains a lot of zero values. Taking this into consideration, we have chosen to focus on the best approximative theoretical distribution for the non-zero values and present ratios of zero values for each interaction parameter. The following candidate distributions were tested for each parameter: *beta*, *gamma*, *inverse gamma*, *normal*, *log-normal*, *skewed normal*, *geometrical* and *uniform*. For the theoretical distributions which are defined only on interval [0,1] we have first normalized the data according to Equation (1),

$$x_{norm} = \frac{x_{empirical} - x_{min}}{x_{max} - x_{min}}. \quad (1)$$

Maximum likelihood estimation (MLE) was used for each listed distribution to find the distribution

Table 1. Available interaction parameters.

Interaction parameter name	Abbreviation	Description
friend_mutual	fm	Number of mutual friends between ego-user and his observed Facebook friend
feed_like	fl	Number of observed friend's "likes" on ego-user's posts
feed_comment	fc	Number of observed friend's comments on posts on the ego-user's timeline
feed_addressed	fa	Number of observed friend's posts on the ego-user's timeline
feed_together_in_post	ftp	Number of times when ego-user and his observed Facebook friend are tagged together in posts
mutual_photo_published_by_user	mpu	Number of mutual photos of ego-user and his observed Facebook friend published by ego-user
mutual_photo_published_by_friend	mpf	Number of mutual photos of ego-user and his observed Facebook friend published by observed friend
mutual_photo_published_by_others	mpo	Number of mutual photos of ego-user and his observed Facebook friend published by some other
photo_like	pl	Number of "likes" of observed Facebook friend on ego-user's photos
photo_comment	pc	Number of comments by observed Facebook friend on ego-user's photos
inbox_chat	ic	Number of exchanged private messages between ego-user and his observed Facebook friend (taking into account only last 50 from ego-user's total conversations)
my_photo_likes	mpl	Number of ego-user's "likes" on observed Facebook friend's photos
my_link_likes	mll	Number of ego-user's "likes" on observed Facebook friend's links

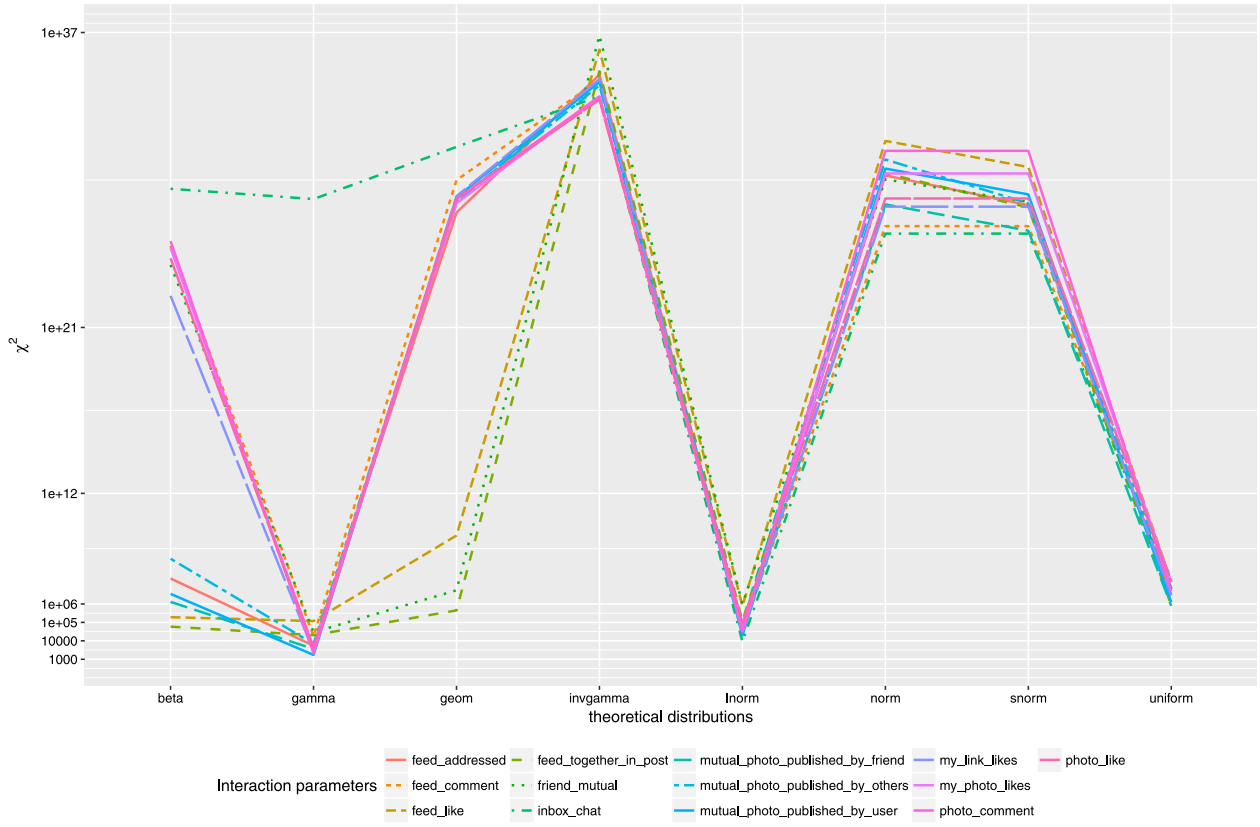

Figure 1. Results of chi-square test per parameters per distributions.

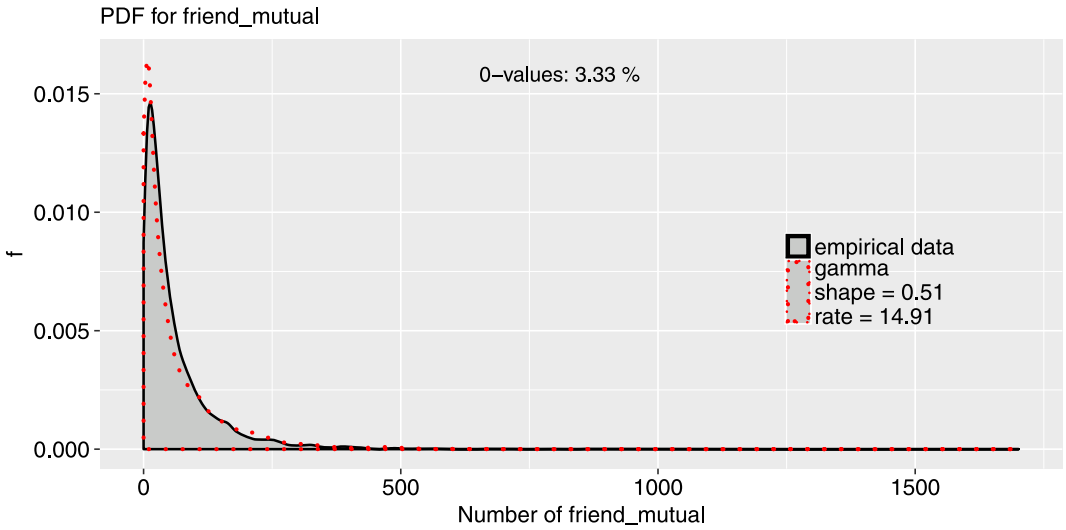

Figure 2. Comparison of empirical distribution of friend_mutual parameter and the approximative gamma distribution as the best approximative theoretical distribution.

parameters which show the best fit. Using the *chi-square test* we have decided on the final theoretical distributions with lowest corresponding *chi-square* values.

Results

Analysis of the underlying distributions

In this section, we will present the results of our underlying distributions analysis for each interaction parameter. Detailed quantile tables with over 10,000 records for each interaction parameter empirical distribution are not included in this paper due to obvious size constrains, but can be found at *r.lukahumski.iz.hr/EAPIOSN/quantiles.csv*. For each interaction parameter, we have found out the best approximative theoretical distribution of non-zero values and presented the ratio of zero values. Figure 1 shows the results of *chi-square tests* (with the number of bins set to 50) for each interaction parameter for different distributions. To show a simple graphical illustration of differences between empirical distributions and the best approximate theoretical distributions we also include a representative probability density function (PDF) of empirical and approximative theoretical distribution

Table 2. The best approximative theoretical distributions.

Interaction parameter name	Best approximative theoretical distribution	Theoretical distribution parameters	Chi-square value	Ratio of zero values (%)
friend_mutual	gamma	shape = 0.51 rate = 14.91	29,369	3.33
feed_like	gamma	shape = 0.12 rate = 5	118,155	55.29
feed_comment	gamma	shape = 0.09 rate = 17.23	17,905	86.83
feed_addressed	gamma	shape = 0.07 rate = 6.25	5467	93.19
feed_together_in_post	gamma	shape = 0.06 rate = 4.35	20,297	97.58
mutual_photo_published_by_user	gamma	shape = 0.07 rate = 5.72	1726	96.94
mutual_photo_published_by_friend	gamma	shape = 0.07 rate = 5.45	3618	97.88
mutual_photo_published_by_others	gamma	shape = 0.08 rate = 6.98	6851	87.35
photo_like	gamma	shape = 0.09 rate = 12.07	4359	71.67
photo_comment	gamma	shape = 0.09 rate = 11.11	2081	90.46
inbox_chat	log-normal	meanlog = -8.83 sdlog = 4.26	9024	91.86
my_photo_likes	gamma	shape = 0.08 rate = 14.52	4098	82.03
my_link_likes	gamma	shape = 0.07 rate = 8.71	2688	94.34

Figure 3. Correlation between attributes available in dataset.

for the *friend_mutual* parameter on Figure 2. Theoretical distribution is depicted as a dotted line, while the empirical distribution is shown with a solid line. In Table 2 we list all the interaction parameters, their best approximative theoretical distribution name, parameters for best fit, resulting *chi-square* value and the ratio of zero values. It is important to emphasize that according to the *chi-square test* it is not unequivocally proven for any interaction parameter to be distributed according to a specific theoretical distribution, but highlighted theoretical distributions are the best approximation for observed empirical distributions considering the scope of observed theoretical distributions.

Analysis of correlations between interaction parameters

Pearson's correlation coefficients between attributes in the dataset are shown in Figure 3. Upper part of the figure shows correlation intensity using the size and colour of the squares, while the lower part shows exact numerical values. Due to reasons of clarity, all attributes have abbreviated names (according to Table 1).

Discussion

Previous section presented the results of exploratory analysis done by using the dataset gained in the conducted social experiment. In the following paragraphs, we will briefly review gained results and try to provide some interpretations.

Correlations show which interaction parameters are connected and how strong that connection is. Our analysis shows that *feed_comment* and *feed_addressed* have the strongest correlation. It is interesting to note that people who make a lot of comments on friend's posts will also write many standalone posts on their respective timelines. Analysis also shows high correlation between parameters *photo_like* and *feed_like*, which is logical concerning the nature of these parameters, i.e. users treat reacting to textual posts and pictures very similarly. High correlation between attributes *photo_comment* and *feed_comment* also supports this assumption.

Low correlation between parameters that show the numbers of mutual photos is slightly surprising. We previously expected to see a relative similarity between parameters *mutual_photo_published_by_user*, *mutual_photo_published_by_friend* and *mutual_photo_published_by_others* because all these parameters count the number of mutual photos between ego-users and their observed Facebook friend, with the only difference being the person who published the photo. Analysis, however, showed that photo sharing habits vary significantly between users.

Another interesting find is that there is no correlation between number of mutual friends and the level of interaction on OSN via observed interaction parameters. An assumption can be made that people who have more friends in common belong to a certain clique which will be reflected in a more intensive online communication, but our analysis showed this is not corroborated by facts gained by the survey results.

When looking at various distributions, the large number of zero values is apparent, meaning that ego-users generally interact very little with most of their Facebook friends. This is not so surprising if we refer to the *Dunbar's number* [28] which states that people can comfortably maintain only 150 stable relationships, compared to the average number of Facebook "friends" in our survey which was 429. The total lack of interaction further affirms this supposition, and this fact additionally motivates researches which aim to distinguish OSN friends which truly are digital representations of actual real-life relationships.

Finally, if one wants to model interaction paramater behaviour using theoretical distributions, the overall best approximative theoretical distribution for all interaction parameters is the *gamma distribution*, the sole exception being the *inbox_chat* parameter for which the *log-normal* distribution gives the best results.

Conclusion

In this paper, we have presented the results of our exploratory analysis aimed to extract key properties of the data which describes interactions between pairs of connected Facebook users. For each interaction parameter, we have provided an empirical distribution as a detailed quantile table. Also, we discovered the best approximative theoretical distributions and associated parameters for all observed interaction parameters. For all pairs of interaction parameters, we presented the level of correlation by calculating the *Pearson's correlation coefficient*.

The presented dataset was obtained in a massive social experiment *NajFrend* which involved over 3000 participants and collected more than 1,400,000 records with summarized frequencies of interaction parameters between ego-users and their Facebook friends. The interaction records were collected using Facebook API 1.0. This dataset will also be the mainstay of our future research involving methods for discovering and visualizing real-life relationships based on observed social network interaction parameters.

Acknowledgement

We would like to thank our ex-student Juraj Ilić who developed PHP application "NajFrend" for conducting the survey on which this research is based and helped us in survey conducting.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research team would like to thank Croatian Science Foundation (*Hrvatska zaklada za znanost* – www.hrzz.hr). The work has been fully supported by Croatian Science Foundation under the project UIP-2014-09-2051 eduMINE – Leveraging data mining methods and open technologies for enhancement of the e-learning infrastructure.

ORCID

Luka Humski http://orcid.org/0000-0002-6819-8899

References

[1] Biancini A. Social psychology testing platform leveraging Facebook and SNA techniques. 2012 Eighth Internal Conference on Signal Image Technol and Internet-based Systems. [Internet]. Naples, Italy: IEEE; 2012 [cited 2017 Nov 18]. p. 776–783. Available from: http://ieeexplore.ieee.org/document/6395170/.

[2] Smith PB, Bond MH, Kâğıtçıbaşı C. Understanding social psychology across cultures: living and working in a changing world. Thousand Oaks (CA): Sage; 2006.

[3] Diaby M, Viennet E, Launay T. Toward the next generation of recruitment tools: an online social network-based job recommender system Mamadou. Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. – ASONAM '13 [Internet]. Niagara, Canada: ACM Press; 2013 [cited 2017 Nov 8]. p. 821–828. Available from: http://dl.acm.org/citation.cfm?doid=2492517.2500266.

[4] Nie B, Zhang H, Liu Y. Social interaction based video recommendation: recommending YouTube videos to Facebook users. 2014 IEEE Conference on Computer Communications. (INFOCOM WKSHPS) [Internet]. Toronto, Canada: IEEE; 2014 [cited 2017 Nov 8]. p. 97–102. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6849175.

[5] Humski L, Štriga D, Podobnik V, et al. Building implicit corporate social networks: the case of a multinational company. In: Pripužić K, Banek M, editors. Proceedings of the 12th International Conference on Telecommunications. Zagreb, Croatia: University of Zagreb; 2013. p. 31–38.

[6] Lin C-Y, Wu L, Wen Z, et al. Social network analysis in enterprise. Proc IEEE. 2012;100:2759–2776.

[7] Lin C-Y, Ehrlich K, Griffiths-Fisher V, et al. Smallblue: people mining for expertise search. IEEE Multimed. 2008;15:78–84.

[8] Dasgupta K, Singh R, Viswanathan B, et al. Social ties and their relevance to churn in mobile telecom networks. Proceedings of the 12th International Conference on Extending Database Technology. – EDBT '08 [Internet]. New York, USA: ACM Press; 2008 [cited 2017 Nov 8]. p. 668–677. Available from: http://portal.acm.org/citation.cfm?doid=1353343.1353424.

[9] Phadke C, Uzunalioglu H, Mendiratta VB, et al. Prediction of subscriber churn using social network analysis. Bell Labs Tech. J. [Internet]. 2013 [cited 2017 Nov 8];17:63–75. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6768308.

[10] Benedek G, Lublóy Á, Vastag G. The importance of social embeddedness: churn models at mobile providers. Decis Sci. [Internet]. 2014 [cited 2017 Nov 8];45:175–201. Available from: http://doi.wiley.com/10.1111/deci.12057.

[11] Romero C, Ventura S. Data mining in education. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. [Internet]. 2013 [cited 2017 Nov 8];3:12–27. Available from: http://doi.wiley.com/10.1002/widm.1075.

[12] Rabbany R, Takaffoli M, Zaiane OR. Analyzing participation of students in online courses using social network analysis techniques. Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, The Netherlands; 2011. p. 21–30.

[13] Carolan B V. Social network analysis and education: theory, methods & applications [Internet]. Sage Publications; 2014 [cited 2017 Nov 8]. Available from: https://uk.sagepub.com/en-gb/eur/social-network-analysis-and-education/book236900.

[14] Catanese S, De Meo P, Ferrara E, et al. Analyzing the Facebook friendship graph. Proceedings of the 1st International Workshop on Mining the Future Internet [Internet]. Berlin, Germany; 2010 [cited 2017 Nov 18]. Available from: http://arxiv.org/abs/1011.5168.

[15] Akhtar N, Javed H, Sengar G. Analysis of Facebook social network. 2013 5th International Conference on Computational Intelligence and Communication Networks [Internet]. Mathura, India: IEEE; 2013 [cited 2017 Nov 18]. p. 451–454. Available from: http://ieeexplore.ieee.org/document/6658034/.

[16] Catanese S, De Meo P, Ferrara E, et al. Crawling Facebook for social network analysis purposes. Proceedings of the International Conference on Web Intelligence, Mining and Semantics – WIMS '11 [Internet]. Sogndal, Norway: ACM Press; 2011 [cited 2017 Nov 18]. Available from: http://portal.acm.org/citation.cfm?doid{\mathsurround=\opskip$=$}1988688.1988749.

[17] Kahanda I, Neville J. Using transactional information to predict link strength in online social networks. Proceedings of the Third International Conference Weblogs and Social Media – ICWSM 2009 [Internet]. San Jose, California (USA); 2009. p. 74–81. Available from: http://aaai.org/ocs/index.php/ICWSM/09/paper/view/213.

[18] Stupalo M, Ilić J, Humski L, et al. Applying the binary classification methods for discovering the best friends on an online social network. In: Dobrijević O, Džanko M, editors. 2017 14th Internation Conference on Telecommunications. Zagreb: Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva; 2017. p. 155–162.

[19] Gilbert E, Karahalios K. Predicting tie strength with social media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. – CHI 09 [Internet]. New York, USA: ACM Press; 2009 [cited 2017 Nov 8]. p. 211–220. Available from: http://dl.acm.org/citation.cfm?doid=1518701.1518736.

[20] Ilić J, Humski L, Pintar D, et al. Proof of concept for comparison and classification of online social network friends based on tie strength calculation model. In: Zdravković M, Trajnović M, Konjović Z, editors. ICIST 2016 Proceedings. [Internet]. Beograd: Society for Information Systems and Computer Networks; 2016 [cited 2017 Nov 23]. p. 159–164. Available from: http://www.eventiotic.com/eventiotic/library/paper/37.

[21] Sever N, Humski L, Ilić J, et al. Applying the multiclass classification methods for the classification of online social network friends. In: Rožić N, Lorenz P, editors. 2017 25th International Conference on Software, Telecommunications and Computer Networks [Internet]. Split: IEEE; 2017 [cited 2018 Jan 23]. p. 1–6. Available from: http://ieeexplore.ieee.org/document/8115508/.

[22] Podobnik V, Štriga D, Jandras A, et al. How to calculate trust between social network users? In: Rožić N, Begušić D, editors. SoftCOM 2012, 20th International Conference on Software, Telecommunications and Computer Networks. Split, Croatia: FESB; 2012. p. 1–6.

[23] Majić M, Skorin J, Humski L, et al. Using the interaction on social networks to predict real life friendship. In: Rožić N, Begušić D, editors. 2014 22nd International

Conference on Software, Telecommunications and Computer Networks. Split: FESB; 2014. p. 378–382.
[24] Pappalardo L, Rossetti G, Pedreschi D. "How well do we know each other?" Detecting tie strength in multidimensional social networks. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. [Internet]. Istanbul, Turkey: IEEE; 2012 [cited 2017 Nov 8]. p. 1040–1045. Available from: http://ieeexplore.ieee.org/document/6425622/.
[25] Krakan S, Humski L, Skočir Z. Determination of friendship intensity between online social network users based on their interaction. Teh Vjesn / Tech Gaz. 2018;25(3). DOI:10.17559/TV-20170124144723.
[26] Li X, Yang Q, Lin X, et al. Itrust: interpersonal trust measurements from social interactions. IEEE Network. [Internet]. 2016 [cited 2017 Nov 18];30:54–58. Available from: http://ieeexplore.ieee.org/document/7513864/.
[27] Global social media ranking 2018 – statistic [Internet]. [cited 2018 Mar 5]. Available from: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.
[28] Dunbar RIM. Neocortex size as a constraint on group size in primates. J Hum Evol. [Internet]. 1992 [cited 2017 Nov 18];22:469–493. Available from: https://www.sciencedirect.com/science/article/pii/004724849290081J.