X. Évfolyam 4. szám - 2015. december

BALOGH Zoltán - RÉTALLÉR Orsolya
zoltan.balogh2@uni-corvinus.hu - o.retaller@rug.nl

SPECIALITIES OF PSYCHOLOGICAL TRAITS OF THE CITIZENS OF CORVINUS UNIVERSITY OF HUNGARY

Abstract

There are several papers dealing with the data fetched from Facebook by the myPersonality Project that contains psychological traits of Facebook users of the USA. During the spring of 2012 an embedded tracking script in the e-learning environment of the Corvinus University of Budapest was collecting all available attributes of the visitors and also saved their likes from Facebook with their prior consent. After this we parsed all the saved likes with the psychological API[1]. In 2015 we got permission to compare our dataset's characteristics to the properties of another database, gathered by the myPersonality project. We are revealing the similarities and differences between the two nations that can be deducted from the Facebook likes of visitors.

Sok kutatás született a myPersonality Project nevű kutatás eredményeképpen létrejött adatokból, amely főleg amerikai felhasználók pszichológiai jellemzőit kapcsolja össze a Facebook-os like-jaikkal. 2012 szeptemberében egy adatgyűjtő alkalmazást helyeztünk a Budapesti Corvinus Egyetem e-Learning felületére, amely lementette a felhasználók elérhető adatait, valamint előzetes hozzájárulásukkal a Facebook-on található like-jaikat is. Ebben a cikkben közöljük az adatok elemzésének eredményeit. A lementett like adatbázisukat feldolgoztattuk a myPersonality pszichológiai API[2]-val, majd a kapott eredményeket összehasonlítottuk a myPersonality projekt által közzétett amerikai adatokkal, melyek használatára 2015-ben kaptunk engedélyt.

Keywords: *myPersonality Project, psychological traits, Facebook ~ myPersonality Project, pszichológiai jellemzők, Facebook*

[1] Advanced Programmers Interface: a software component that defines functionalities that are independent of their respective implementations, a definition of software operations that can be used in other applications

[2] Alkalmazásiprogramozási felület: egy alkalmazás függvényeinek használatának lehetősége más alkalmazások számára, annak belső működésének megértése nélkül

INTRODUCTION

According to the latest statistics, Facebook has 1.49 billion active users, who spend more than 20 minutes daily browsing it. Mark Zuckerberg claims, that an average US user spends 40 minutes on Facebook every single day. 20% of them are basically always online. (D'Onfro, 2015) This number increases every year as Facebook is becoming available on more and more platforms and gadgets.

„If time is money, then the Facebook.com site represents
the most valuable Internet property on the web today"
1., Quote by a Needham analyst on Facebook

If companies knew more about the characteristics of their online visitors, they could send them better targeted advertisements or align their campaigns to better meet the needs of their potential buyers. (Escobido & Gillian, 2013) The visitors' personal characteristics are not likely to be collected from the public domain, because it is a personal property that can be deducted from online behavior. (Kosinksi, Stillwell, & Graepel, Private traits and attributes are predictable from digital records of human behavior, 2013)

In 2012 the researchers of The Psychometrics Center at the University of Cambridge analyzed a dataset of more than 58 000 volunteers who provided access to their Facebook likes and filled out several psychological tests. Then they used the collected data to develop a public API that can predict private properties of an individual from his Facebook likes. (Kosinksi, Stillwell, & Graepel, Private traits and attributes are predictable from digital records of human behavior, 2013)

In psychology, the Big Five personality traits are attributes that can describe human personality. These five factors are openness, conscientiousness, extraversion, agreeableness, and neuroticism. Computers made it possible to access and analyze large amounts of text samples, so that it can be used to identify personality types and predict potential reactions and behaviors. (Poria, Gelbukh, Agarwal, Cambria, & Howard, 2013)

The properties of the Big Five Traits are as follows:

- *Openness* is the willingness of seeking of new experiences and interest in culture, ideas, and aesthetics, also measures a person's imagination and curiosity.
- *Conscientiousness:* people like an organized approach to life. People who are like this are usually ambitious, resourceful and persistent.
- *Extraverts* are more outgoing, friendly, and socially active. They are usually energetic and talkative, they do not mind being at the center of attention, and they can make new friends more easily. Introverts are the opposite: they are more comfortable in their own company and tend to seek environments characterized by a lower level of external stimulation.
- *Agreeableness:* people are trusting, altruistic and tender-minded. High agreeableness scorers are usually friendly and compassionate and it is difficult for them to tell the hard truth.
- The trait of *Neuroticism* is associated with descriptive terms such as emotional liability and impulsiveness. The tendency to experience mood swings and negative emotions such as guilt, anger, anxiety, and depression and often referred to as emotional instability. (Kosinski, Bachrach, Kohli, Stillwell, & Graepel, 2013)

It is already proven that there is a connection between the personality of the actual user and his Facebook profile. (Golbeck, Robles, & Turner, Predicting personality with social media, 2011)

LITERATURE REVIEW

Apparently the data collected by the myPersonality Project gave a boost to the researchers. There has been a wide range of research since the initial publication of the data. Some papers had surprising results or they confirmed a presumption:

- Marketing research has proved that individuals with high in Openness are likely to be innovators and might influence others, while they tend to have less restrictive privacy. (Kosinski, et al., 2012)
- Facebook's Gross National Happiness (FGNH) indexes the positive and negative words used in the millions of status updates submitted daily by Facebook users. It has been pointed out that FGNH peaks during events, such as Thanksgiving and Christmas, while troughs appear when depressing events or commemoration days occur, such as after Michael Jackson's sudden death. (Stillwell, Kosinki, Rust, & Wang, 2012)

Meanwhile, others are aiming to build a better profile of the online citizens:

- In the paper called „Predicting personality with social media" a method is presented by which a user's personality can be accurately predicted through the publicly available information on their Facebook profile. Researches have shown that there is a connection between general personality trait and online behavior. (Golbeck, Robles, & Turner, Predicting personality with social media, 2011)
- In „Facebook and Privacy: The Balancing Act of Personality, Gender and Relationship Currency" paper the users were put into categories based on their personality traits and it was shown that - the more privacy-conscious the users are, the higher traits of Openness and Extraversion they have. Men and women share an equal amount of personal information; however, women tend to be more cautious and make information less visible. (Kosinski, et al., 2012)

In this nomenclature, this paper belongs to the second category as we try to reveal hidden characteristics of the Hungarian university citizens compared to data from the myPersonality Project.

REASEARCH

Data collection

In April 2012 we embedded a visitor tracker script into the e-learning system of the Corvinus University of Budapest. It saved all the available attributes of the visitors' software and hardware environment, their geolocation and data on Facebook with their prior consent. The tracker script ran for 1 month: during the first 2 weeks of testing period we were debugging and fine-tuning the script, then we ran it for another 2 weeks to collect data. The properties of the collected data:

- *647 242 records* collected, one per each page load (2,4 GB)
- *32 529* unique sessions
- According to the logs, *8 169 users* logged into of the e-learning environment
- Sharing private data including Facebook data and Geolocation
    - The users were motivated with sweepstakes
    - *139* visitors shared their *Facebook data (1,6%)*
    - *303* visitors shared their *Geolocation (3,1%)*

The collected raw data had be parsed before the statistical analysis. It is necessary, because some of the columns were empty or not available.

Then the collected Facebook likes were sent to the myPersonality API and the results were saved into the database linked to the visitors. The API needs a certain number of likes to calculate reliable prediction. *48* of the *total 139 visitors* had enough likes to fulfill this criteria and these resulted in a valid output. We compared the results of the 48 visitors against the data of USA citizens downloaded from myPersonality.org.

Control data from myPersonality

We obtained the Big Five, demographical data, iq, satisfaction with life, religion and political views data from myPersonality. According to the descriptive statistics available on their website the mean of the age of their sample is 23,55 and the top 5 contributors were from mainly English speaking countries (USA, UK, Canada, Australia, England). They provide 79,06% of the whole data. (Kosinksi, myPersonality Project, 2012)

Statistical findings

Some descriptive statistics

We started our analysis with a comparison of the variables. Table 1.-Table 4. illustrate the means and their 95% confidence intervals for both the United States and Hungary.

Table 1., Means of Big Five

Big Five	**United States of America**			**Hungary**		
	Mean	95% confidence interval		Mean	95% confidence interval	
Agreeableness	,6281	,6207	,6354	,3413	,3030	,3795
Conscientiousness	,5756	,5676	,5835	,4728	,4235	,5220
Openness	,7736	,7673	,7800	,4509	,4196	,4822
Neuroticism	,4473	,4377	,4569	,3245	,2892	,3597
Extraversion	,5323	,5227	,5418	,3765	,3431	,4098

Table 2., Means of relationship status

Relationship status	**United States of America**			**Hungary**		
	Mean	95% confidence interval		Mean	95% confidence interval	
Not in a relationship	,4651	,4429	,4873	,5480	,5355	,5605
Married	,0857	,0733	,0982	,1449	,1349	,1550
In a relationship	,1828	,1656	,1999	,3071	,2980	,3161

Table 3., Means of political views

Political views	**United States of America**			**Hungary**		
	Mean	95% confidence interval		Mean	95% confidence interval	
Conservative	,0103	,0058	,0147	,3171	,2917	,3425
Liberal	,0277	,0204	,0350	,4072	,3783	,4361
Uninvolved	,0154	,0099	,0209	,1676	,1538	,1814
Libertarian	,0062	,0027	,0096	,1081	,0979	,1183

Table 4., Means of the rest of the variables

Other variables	**United States of America**			**Hungary**		
	Mean	95% confidence interval		Mean	95% confidence interval	
Age	28,09	27,61	28,58	26,11	25,67	26,54
Number of friends	106,10	97,63	114,57	594,50	513,35	675,65
Intelligence	114,77	114,12	115,4	104,99	103,09	106,97
Satisfaction with life	,5337	,5231	,5442	,4536	,4180	,4891

The most important finding is probably that the confidence intervals do not overlap anywhere. Therefore, the variables mentioned above have significantly different mean values in the two countries.

The target group of this research is somewhat older in the US than in Hungary; they seem to be more intelligent, and they are more satisfied with their lives. On the contrary, Hungarian students have more Facebook friends. Moreover, the mean values of all Big Five variables are higher in the United States.

It is a bit surprising that Hungarian students seem to be more actively communicating their political views: the mean values in each political view is significantly higher than their American equivalents. It is also clearer, if someone is not involved in politics.

The API scores for the relationship status are again higher of Hungarian students, which suggests that they communicate it more clearly on Facebook, if they are in a relationship, got married, or if they are single.

We were also interested whether there is a relationship between these variables. By creating scatter-plots we could easily see that all of them look rather independent, even if we split them by gender. As an illustration, figure Figure 1. presents the independence of intelligence and satisfaction with life first in the United States and then in Hungary:

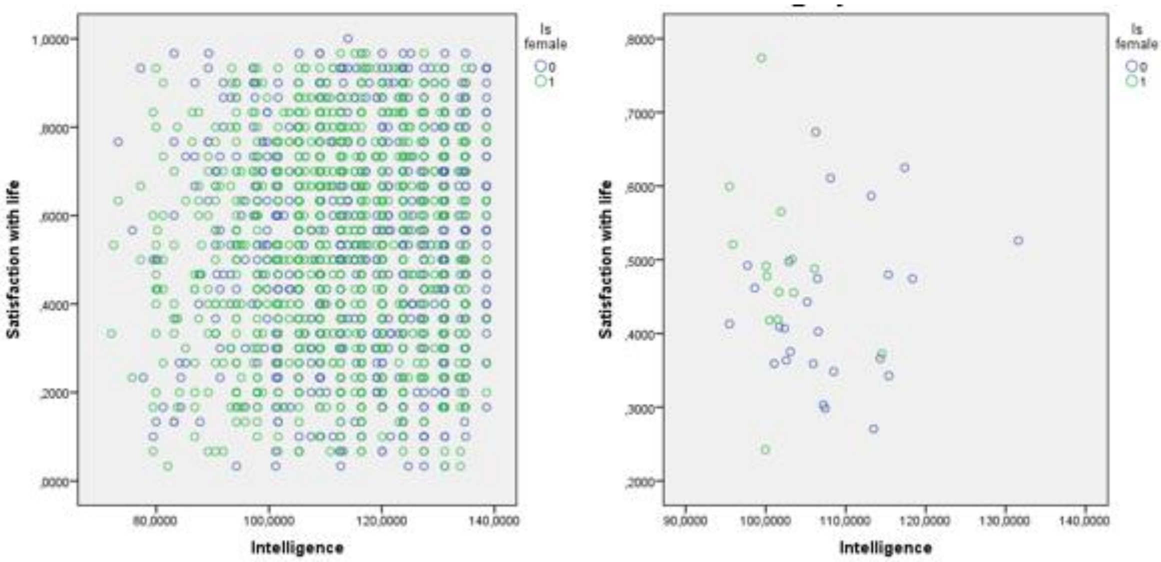

Figure 1., Scatter-plot diagram of satisfaction with life and intelligence

Clustering

Our goal was to discover the underlying structure of potential groups in the data, and observe whether they differ in the two countries. Therefore we used clustering algorithms with unsupervised learning in both cases, and compared the results of different clustering methods.

K-MEANS[3] CLUSTERING – THE UNITED STATES OF AMERICA

The disadvantage of K-means clustering is that without actually having any further knowledge on the data structure we already have to assume the existence of a certain number of clusters. In the beginning we chose 2 clusters, but with later analyses we used 3 and 4 clusters as well. The algorithm used the squared Euclidean distance as a distance measure, with starting points assigned by SPSS[4].

We first did clustering by only using the Big Five variables. As the following ANOVA[5] table, table Table 5 illustrates, these two clusters significantly differ from each other by each dimension, due to the low p-values of the corresponding F-tests.

[3] K-means clustering, or Lloyd's algorithm, is an iterative, data-partitioning algorithm that assigns n observations to exactly one of k clusters defined by centroids, where k is chosen before the algorithm starts. (The MathWorks Inc., 2015)

[4] IBM SPSS Statistics: Statistical Software Package used for statistical analysis

[5] ANOVA: analysis of variance, collection of statistical models used to analyze the differences among group means

Table 5., ANOVA table of Big Five

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Big Five Agreeableness	8,910	1	,020	1946	434,660	,000
Big Five Conscientiousness	8,901	1	,024	1946	366,586	,000
Big Five Openness	1,892	1	,018	1946	106,342	,000
Big Five Neuroticism	43,019	1	,022	1946	1945,223	,000
Big Five Extraversion	30,532	1	,026	1946	1158,872	,000

Next we tried clustering with respect to political views. The results suggest that these variables do not separate the clusters significantly, since the p-values of the F-tests exceed the 5% significance level in every dimension:

Table 6., ANOVA table of politics

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Conservative in politics	,006	1	,010	1946	,576	,448
Liberal in politics	52,503	1	,000	1946	.	.
Uninvolved in politics	,013	1	,015	1946	,868	,352
Libertanian in politics	,002	1	,006	1946	,344	,558

The only exception is the Liberal view, since all individuals with a non-liberal view were assigned to one cluster, and the ones with liberal views were all assigned to the other cluster.

The relationship status seems to distinguish between two clusters once again:

Table 7., ANOVA table of relationship

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Not in relationship	39,511	1	,229	1946	172,739	,000
Married	152,683	1	,000	1946	.	.
In relationship	6,100	1	,146	1946	41,678	,000

While creating clusters we also saved the cluster memberships so that we can check how they relate to each other. The following cross-tab illustrates the clusters created by using the Big Five variables and the clusters which were created by using the Social factors (B_clusters and S_clusters):

Table 8., Cluster membership crossvalidation

B_clusters * S_clusters Cross-tabulation

Count

		S_clusters		
		1	2	Total
B_clusters	1	872	72	944
	2	909	95	1004
Total		1781	167	1948

The data suggest that these clusters are independent from each other, which we could verify by calculating the Pearson Chi-Square (2,090) and its corresponding p-value (0,148). Therefore the null-hypothesis of independence cannot be rejected. This is even more visible if we plot the data into a 3D histogram:

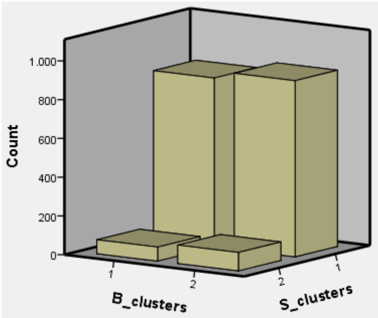

Figure 2., 3D histogram of clusters

Since the political views did not seem to distinguish enough between the clusters, they were not used for our final clustering algorithm. Due to the independence of the Big Five and the Social characteristics we decided to do K-means clustering by these variables. We ran the algorithm for 2, 3 and 4 clusters as well. We found that 4 clusters seem to be the best choice, since the variables used distinguish between these clusters the best. This is also visible in the following table:

Table 9., ANOVA table of clusters of Big Five & relationship

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Big Five Agreeableness	,065	3	,025	1944	2,612	,050
Big Five Conscientiousness	,236	3	,029	1944	8,265	,000
Big Five Openness	,106	3	,019	1944	5,683	,001
Big Five Neuroticism	,070	3	,044	1944	1,586	,191
Big Five Extraversion	,151	3	,042	1944	3,610	,013
Not in relationship	161,542	3	,000	1944	.	.
Married	50,894	3	,000	1944	.	.
In relationship	96,980	3	,000	1944	.	.

Again, for the social characteristics the F-values and corresponding p-values are missing, because these are completely distinguished, as the final cluster centers also illustrate:

Table 10., Final cluster centers

	Cluster			
	1	2	3	4
Big Five Agreeableness	,6126	,6346	,6259	,6418
Big Five Conscientiousness	,5712	,5715	,5566	,6334
Big Five Openness	,7811	,7678	,7888	,7410
Big Five Neuroticism	,4566	,4385	,4639	,4453
Big Five Extraversion	,5250	,5218	,5583	,5550
Not in relationship	0	1	0	0
Married	0	0	0	1
In relationship	0	0	1	0

Hierarchical Clustering – the United States of America

To confirm our results we decided to run another clustering algorithm on the data. We chose Ward's method with the squared Euclidean distance measure. The advantage of hierarchical clustering is that it is not necessary to predefine the number of clusters, because the dendrogram will provide us with a great tool to identify the underlying structure by ourselves – if there is any. In our case, 4 clusters clearly revealed themselves. By saving the cluster memberships, it is possible to check how different these results are from the K-means clustering results. The following cross-tab proves that the hierarchical clustering created exactly the same groups, which suggests that these underlying groups are considerably stable:

Table 11., Comparison between the two clustering algorithms

Cluster Number of Case * Ward Method Cross-tabulation						
Count						
		Ward Method				
		1	2	3	4	Total
Cluster Number of Case	1	0	519	0	0	519
	2	906	0	0	0	906
	3	0	0	0	356	356
	4	0	0	167	0	167
Total		906	519	167	356	1948

K-means Clustering – Hungary

Using the same methodology as previously, we created 2-2-2 clusters using the Big Five, political view and social characteristics. The results were somewhat different than in the USA's case: we found that in the case of Hungarian students we can distinguish between clusters according to political views, except for the variable of "uninvolved in politics". Since a comparison with the US is desirable, we decided not to use this variable when creating the final clusters. Nevertheless, we disclose the ANOVA tables of the clustering algorithms.

Table 12., ANOVA table of Big Five clusters

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Big Five Agreeableness	,067	1	,013	38	5,140	,029
Big Five Conscientiousness	,611	1	,008	38	74,019	,000
Big Five Openness	,038	1	,009	38	4,317	,045
Big Five Neuroticism	,069	1	,011	38	6,454	,015
Big Five Extraversion	,064	1	,009	38	6,747	,013

Table 13., ANOVA table of politics clusters

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Conservative in politics	,122	1	,003	38	37,399	,000
Liberal in politics	,226	1	,002	38	93,156	,000
Uninvolved in politics	,002	1	,002	38	1,074	,307
Libertanian in politics	,007	1	,001	38	7,680	,009

Table 14., ANOVA table of relationship

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Not in relationship	,032	1	,001	38	43,480	,000
Married	,013	1	,001	38	18,541	,000
In relationship	,004	1	,001	38	6,064	,018

Once again we wanted to see how different the created clusters are when using the Big Five variables and the social variables, therefore the following cross-tabulation was made:

Table 15., Cluster cross-tabulation

B_clusters * S_clusters Cross-tabulation				
Count				
		S_clusters		
		1	2	Total
B_clusters	1	5	5	10
	2	9	21	30
Total		14	26	40

The table suggests independence, but due to the low number of data the expected count in one cell does not reach up to 5. We still decided to conduct the test, since this expected count is 3,5, which cannot be considered as extremely low. According to our results the hypothesis of independence cannot be rejected, we calculated a value of 1,319 for the Pearson Chi-Square, which corresponds with a p-value of 0,251.

Once again we ran the K-means clustering algorithm with 2, 3 and 4 clusters and checked out the results. Interestingly, 3 clusters seem to fit better to the Hungarian data than 4, since in this case all the used variables distinguish well between the clusters. Later we will check this using hierarchical clustering.

Table 16., ANOVA table of Big Five & relationship clusters

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Big Five Agreeableness	,145	2	,007	37	20,077	,000
Big Five Conscientiousness	,343	2	,006	37	52,967	,000
Big Five Openness	,066	2	,007	37	10,097	,000
Big Five Neuroticism	,019	2	,012	37	1,631	,210
Big Five Extraversion	,020	2	,010	37	1,949	,157
Not in relationship	,004	2	,001	37	3,096	,057
Married	,003	2	,001	37	3,630	,036
In relationship	,004	2	,001	37	7,272	,002

Hierarchical Clustering – Hungary

The dendrogram of the hierarchical clustering clearly reveals that 3 clusters seems to be a better choice for our data than 4, although 2 clusters should also be considered. Yet we concluded that there are four underlying clusters in the data, since with 2 clusters the ANOVA table showed that the variable "Married" does not distinguish well enough.

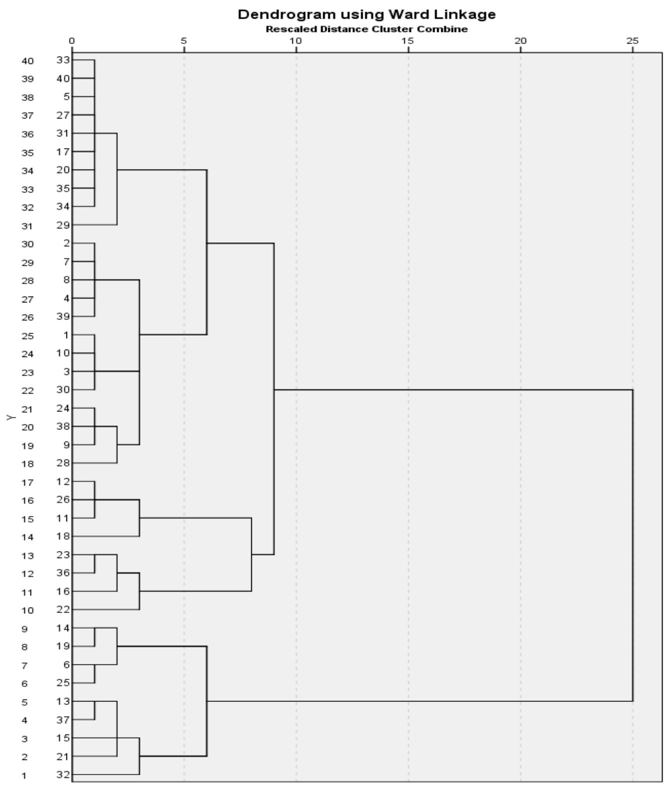

Figure 3., Dendrogram of clusters from the Hungarian data

To verify the stability of the clusters we have saved the cluster memberships of the hierarchical clustering with 3 clusters and compared them to the K-means case. It seems that the two algorithms create slightly different, but almost the same clusters. All in all we can still conclude that these groups are stable.

Table 17., Comparison between the two clustering algorithms

Ward Method * Cluster Number of Case Cross-tabulation					
Count					
		Cluster Number of Case			Total
		1	2	3	
Ward Method	1	0	22	1	23
	2	8	1	0	9
	3	0	1	7	8
Total		8	24	8	40

Comparison between the USA and Hungary

Using the final results of the clustering algorithms we can make a comparison between the underlying groups of the USA and Hungary. In order to do so, we disclose the final cluster centers of the K-means clustering in both cases:

Table 18., USA final cluster centers

	Cluster			
	1	2	3	4
Big Five Agreeableness	,6126	,6346	,6259	,6418
Big Five Conscientiousness	,5712	,5715	,5566	,6334
Big Five Openness	,7811	,7678	,7888	,7410
Big Five Neuroticism	,4566	,4385	,4639	,4453
Big Five Extraversion	,5250	,5218	,5583	,5550
Not in relationship	0	1	0	0
Married	0	0	0	1
In relationship	0	0	1	0

Table 19., Hungarian final cluster centers

	Cluster		
	1	2	3
Big Five Agreeableness	,2284	,4107	,2457
Big Five Conscientiousness	,7207	,4379	,3293
Big Five Openness	,4235	,4217	,5659
Big Five Neuroticism	,2667	,3317	,3605
Big Five Extraversion	,4400	,3597	,3632
Not in relationship	,5749	,5376	,5522
Married	,1442	,1532	,1207
In relationship	,2809	,3091	,3271

The interpretations of the clusters are also presented. In the case of the USA we can identify the following groups:

– Cluster 1: *The ones who live their life for real* - Although these people seem to be quite open, they do not reveal much about themselves on Facebook, probably they rather live their lives in reality than in virtual reality.
– Cluster 2: *The loners* - Most likely these people are not in a relationship, and they also do not have a very vivid social life on Facebook.
– Cluster 3: *The chatty ones* – These people are the ones that have the highest values for openness, neuroticism and extraversion at the same time, so they feel the need to share the most with others. Also, they are in a relationship most of the times.

- Cluster 4: *The serious ones* – The most agreeable, conscientious people, who are most likely living in a marriage. Therefore it is not very surprising that they have such strong skills.

In case of the Hungarian people we could identify 3 clusters, but these clusters somewhat differ from the ones we defined previously for the USA:

- Cluster 1: *The big thinkers* – They are most likely not in a relationship, but they are quite extravert and conscientious people. Maybe dealing with singleness is different than in the United States, so this group might as well correspond to the previously defined "loners".
- Cluster 2: *The serious ones* – They are the most agreeable people, quite conscientious ones, most likely to be married. Very similar to the American cluster 4.
- Cluster 3: *The chatty ones* – They are very open and quite extravert, most likely to be in a relationship. Very similar to the American cluster 3.

CONCLUSION

Although the geographical distance is quite enormous between Hungary and the United States of America, it appears so that the behavior of young people on Facebook is not that different from each other. We should not ignore the fact that the Hungarian data is gathered from only Hungarian university citizens and does not represent the Hungarian population. We could identify 4 groups in the US, the ones who live their life for real, the loners, the chatty ones and the serious ones. Except for the first group, all of these groups could be identified among the Hungarian students too, even though they were slightly younger than the American participants. The only question is why there is no group in this community, which is similar to "the ones who live their life for real" group in the USA. Most likely these people also exist in Hungary, but they are probably bringing this even more to the extreme, and they are just simply not using Facebook at all.

Our research also reveals the importance of social media: geographical distance seems to be less important nowadays, what and how we communicate in the virtual reality matters more and more.

References

[1] Apply Magic Sauce. (2013). University of Cambridge - The Psychometrics Centre. Retrieved from Apply Magic Sauce - Prediction API: http://applymagicsauce.com/documentation.htm

[2] D'Onfro, J. (2015, july 8). Business Insider. (Here's how much time people spend on Facebook per day) Retrieved august 16, 2015, from http://www.businessinsider.com/how-much-time-people-spend-on-facebook-per-day-2015-7#ixzz3fKtewifD

[3] Escobido, M., & Gillian, S. (2013). Can Personality Type be Predicted by Social Media Network Structures? *The Asian Conference on Psychology & the Behavioral Sciences.* Osaka: The International Academic Forum.

[4] Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 253-262). New York: ACM New York.

[5] Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (pp. 253-262). Vancouver: ACM New York, NY, USA.

[6] Kosinksi, M. (2012, october 12). *myPersonality Project.* Retrieved from Available data: http://mypersonality.org/wiki/doku.php?id=available_data

[7] Kosinksi, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 5802-5805.

[8] Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2013, october 19). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning, June 2014, Volume 95*, pp. 357-380.

[9] Kosinski, M., Las Casas, D., Paulo Pesce, J., Quercia, D., Stillwell, D., Almeida, V., & Crowcroft, J. (2012). Facebook and Privacy: The Balancing Act of Personality, Gender, and Relationship Currency. *Sixth International AAAI Conference on Weblogs and Social Media.* Dublin: ICWSM.

[10] Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., & Howard, N. (2013). Common Sense Knowledge Based Personality Recognition from Text. *Advances in Soft Computing and Its Applications* (pp. 484-496). Mexico City: Springer Berlin Heidelberg.

[11] Stillwell, D., & Kosinski, M. (2012). myPersonality project: Example of successful utilization of online social networks for large-scale social research. *Mobile systems for Computational Social Science.* Low Wood Bay.

[12] Stillwell, D., Kosinki, M., Rust, J., & Wang, N. (2012, february 3). Can Well-Being be Measured Using Facebook Status Updates? Validation of Facebook's Gross National Happiness Index. *Social Indicators Research vol 115, issue 1*, pp. 483-491.

[13] The MathWorks Inc. (2015). *kmeans*. Retrieved from MathWorks: http://www.mathworks.com/help/stats/kmeans.html