ATLANTIS PRESS

Atlantis Studies in Uncertainty Modelling, volume 1

11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)

Analysis of privacy profiles applying fuzzy clustering techniques

Aigul Kaskina[a] **and Oleksii K. Tyshchenko**[b]
[a]Department of Informatics, University of Fribourg,
Boulevard de Pérolles 90 , 1700 Fribourg, Switzerland,
aigul.kaskina@unifr.ch
[b]Institute for Research and Applications of Fuzzy Modeling, CE IT4Innovations,
University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic,
oleksii.tyshchenko@osu.cz

Abstract

Unsolved roots of "privacy paradox" motivate researchers to extricate the underlying reasons of such phenomenon. In the field of privacy research, the majority of empirical studies lack the availability of the real data collected from the actual platform instead of data collected from the experimental lab setup. This paper uses the real-world data set of user privacy behavior. Different fuzzy clustering algorithms (such as Fuzzy C-means (FCM), Gustafson-Kessel (GK) algorithm, and Fuzzy Partitioning Around Medoids (PAM)) are applied to the given dataset, and their outfits are compared. The analysis provides the clustering validity procedures applied to the data and then produces the partitioning results of the given set of data in the form of graphical visualizations. This work demonstrates how differently clustering algorithms behave with a given dataset producing various shapes and properties of clusters.

Keywords: privacy profiles, fuzzy clustering, membership degree, fuzzifier, cluster property

1 Introduction

Privacy has become a critical consideration for users while interacting with online services and various systems. These services collect and store users' personal information which can be further shared across other devices, manufactures, and third parties. Due to various reasons, people are not always able to effectively manage their privacy. A design of privacy controls turns out to be non-user-friendly and becomes a barrier for users in effectively maintaining their privacy settings. Consequently, users tend to have problems controlling what precisely they disclose to unintended audiences, thus adapting their online behavior by existing functionality of the privacy controls [6, 10, 17]. The difficulty in defining personal privacy in the online environment not only leads to frustration but may even make users neglect the implied consequences of their disclosure behavior [7]. For that reason, the research community is increasingly focused on developing intelligent privacy controls that aim to help users automate privacy configurations at their desired level. This includes the analysis of users disclosure behavior, privacy preferences and user privacy profiling [8, 13, 1, 4]. However, the development of these tools might fail in the face of the "privacy paradox", which argues that users' actual privacy behavior in most cases diverges from their privacy attitudes [15, 3].

This manuscript extends the previous study [8] and conducts the comparative analysis of users' privacy profiles with the help of different fuzzy clustering tools to illustrate in which way a user who possesses a multidimensional privacy profile may pertain to several clusters (groups) with some specific membership level. The research in [8] introduced a concept of fuzzy privacy profiles which have been derived by applying the Fuzzy C-means (FCM) clustering algorithm. The fuzzy privacy profiles demonstrate how some users' disclosure behavior can belong to several patterns at a time, meaning their ambiguous and more complex privacy decision-making. FCM detects well a spherical shape of clusters. However, the given data tended to be described by ellipsoidal forms. As the previous research lacked a comparative analysis of the clustering results, this paper addresses the classification of user's privacy profiles by applying various fuzzy-based algorithms compared to the results obtained by the initial FCM clustering.

The remainder of this manuscript is organized as described below: Section 2 provides a review of related works referring to the user disclosure behavior and user privacy profiling. Section 3 presents a structure of the given data set of users privacy preferences used for the analysis. Then Section 4 introduces the methods applied for an exploration of the users privacy preferences. In particular, fuzzy clustering validation is shown in Section 4.1, Section 4.2 presents the discussion and visualization of results. Finally, Section 5 outfits a summary and some ideas on future work.

2 Related works

Several surveys inspected users' privacy profiles through the use of clustering techniques [12], [19] due to several reasons. Since the clustering procedures serve as a sort of machine learning techniques, sufficient data of users' privacy profiles should be accumulated to develop an accurate and robust users privacy model. Also, it is absolutely imperative that the differentiation between users' privacy settings and privacy desires is carried out. The former speaks for the users' meaningful privacy behavior inside a platform while the latter states their demeanor for privacy. In the present case, considering an attitude-behavior gap (the so-called "privacy paradox" [15]) is important to avoid fallacious models.

In previous studies [2, 14, 18], privacy profiles were examined from a uni-dimensional representation. In the example given, clustering would be sufficient; however, it might bring weaker results and lead to oversimplification of the users' privacy behavior, both actual and judgemental. To that, [12] stated that the privacy profiles are more than the uni-dimensional representations where, the other way round, they have a multidimensional configuration. The multidimensionality provides additional information for modeling the users' privacy profiles. Also, in accordance with the given set of data, especially when it is real-world data, an initial clusters' quantity is unknown in the majority of cases.

Among a variety of machine learning techniques that help conduct user profiling, the question remains open: "are these techniques providing "smart privacy profiles" smart enough?" These techniques usually identify some n-number of the privacy profiles and assign users to the privacy settings associated with the closest profile. This leads to the uni-dimensional modeling approach which oversimplifies users privacy concerns with the implied issues of an inexact classification and discrimination. Having inherent characteristics of multiple profiles, not every user can perfectly fit the assigned class, thus making the use of traditional profiling techniques inadequate. This study proposes applying the fuzzy clustering tools to avoid the discriminating classification providing more accurate users' representation. The fuzzy clustering can well detect nuances of user privacy preferences. The associated risks may boost by attributing a cluster tag to a crisply distinct cluster and overlooking those inquiries that differentiate drastically from others within that group of points. The multidimensionality of the user's privacy profile indicates that the users privacy decision fluctuates per different data items. On this subject, the fuzzy separation of a data frame may lead to a more precise description of clusters. The excellent, distinctive quality of the fuzzy clustering is that it tags a membership degree to every sample, demonstrating to what measure the user acquires fundamental aspects of each cluster in the data collection. In this research, fuzzy clustering methods are picked against crisp clustering methods.

3 Dataset description

The data used for this study represents users' actual privacy settings configured for their accounts in the voting platform called Participa Inteligente[1] during presidential elections in Ecuador. The information was extrated from the platform for three months between December 2016 and February 2017. We examined the information revelation behavior of users, precisely, with which target group they opt for sharing the data rates of their profile. Inside the voting platform, the users' privacy settings are held inside a MySQL database. It stands to mention that users who had not tuned their inclinations, privacy settings in the default mode were set up to be publicly visible. The data set embodies a collection of users vectors containing six dimensions affiliated with the data grades - {*MyActivity, ContactMe, MyRelations, MyTopics, PersonalInformation and VoteIntention*}. A particular data grade can possess four rates by appointing a sharing position to a specific audiences level in the sets {*1 means "OnlyMe", 2 – Friends, 3 – FriendsOfFriends, 4 –Public*}. Following the data processing and cleaning, the resulting data frame consists of 391 users' privacy profiles. It contains 131 females, 253 males, and seven users with unconfirmed gender proof. The median age of users was 28 years old (but mostly in the range of 23 to 36 years old).

4 Methods

4.1 Clustering validation

Clustering data is an unsupervised tool when a primary number of groups in the dataset is obscure; in this way, defining the right quantity of possible separations in the data becomes the crucial point. In order to validate the fact whether the clustering approach has produced an acute partitioning of the dataset, a validity function can estimate the clustering result. In

[1] https://participacioninteligente.org/

Index name	Criteria	Input parameters			Evaluation properties	
		$U = [u_{ij}]^1$	V^2	X^3	Compact	Separation
Partition Coefficient	$\max(V_{PC})$	x			x	
Partition Entropy	$\min(V_{PE})$	x			x	
Modified Partition Coefficient	$\max(V_{MPC})$	x			x	
Xie and Beni	$\min(V_{XBI})$	x	x	x	x	x
Crisp Silhouette	$\max(V_{CS})$		x	x		x
Fuzzy Silhouette	$\max(V_{FS})$	x	x	x	x	x

(1) Fuzzy partition matrix; (2) Set of cluster centroids; (3) Initial dataset

Table 1: Overview of applied validation indexes

related sources, multiple validity measures have been considered for the validation of the dataset partition produced by the fuzzy clustering algorithms. Performing the analysis of validity indexes, an index is independent of the results of a clustering algorithm. The validity indexes explore all defined number of c clusters to estimate a rate of compactness and separation properties; thus, the data partitioning with the optimal number of clusters c is compact, and clusters are well-separated [20]. The *compactness* measures the proximity of cluster's members, in particular, their variance where a low variance indicates higher compactness. The *separation* quantifies the distinction of two different clusters and calculates a “distance” between them. The validation indexes aim to discover the partitioning of clustering methods which minimizes the compactness and maximizes the separation [11].

Table 1 summarizes computed indexes for this scrutiny. As presented, each validity index has a defined set of parameters used for its estimation, inherent detection of characteristics such as compactness or\and separation of clusters including the criteria of the optimal clusters' quantity. Table 2 displays the validation results presented by the Fuzzy C-means (FCM[2]) [16], Partitioning Around Medoids (PAM[3]) [9] and Gustafson-Kessel (GK[4]) [5] clustering algorithms. In practical solutions, choosing an optimal number of clusters is not a subject to a particular set of rules or unambiguous guidelines. The results in Table 2 show that indexes' values do not come to an agreement on the identical number of clusters. Thus, an optimal number of clusters c was chosen based on the value agreed by most of the validation indexes, where a bold value stands for the optimal values of c selected by each index in agreement with its criteria.

The results of the validation procedure received through the FCM algorithm showcased the agreement of PC, PE, CS and XBI indexes indicating the optimal partition for $c = 2$, whereas FS' and MPC' output results in an optimal number of clusters c that equals 10 and 15 correspondingly. It is worth to be mentioned that execution of CS and FS with $c = 13$ and $c = 14$ brings (NaN) values. The detailed analysis of the (NaN)- produced data partitioning demonstrated that the cluster centers with this number of clusters contain identical values. After the investigation of the clustering outcomes with these number of centroids, it was revealed that FCM with the Euclidean distance generates cluster centroids with identical values. That is to say that FCM with $c = 13$ and $c = 14$ detects two centroids at the same data point. Thus, CS and FS indexes failed to be calculated when the defined parameters contain cluster centroids with the same vectors. Equivalently, the XBI index produced an infinity (Inf) value with $c = 13$ and $c = 14$. These results demonstrate the incapability of the FCM algorithm to find the valid dataset partition for this quantity of clusters as well as emphasizing the existence of the indistinguishable structure of the data under consideration.

Differently to FCM results, the executed validation of PAM clustering yielded correct indexes' values. The reason for such clustering behavior is that essentially the PAM algorithm assigns initial cluster centers to data points taken from the given frame of data, whereas the FCM algorithm artificially calculates initial cluster centers as average means. Table 2 displays that the PAM clustering reaches the optimal partitioning of the dataset with $c = 2$ according to PC, PE, and CS validation indexes. However, MPC and FS also indicated the best partition with $c = 14$ and $c = 15$, as well as XBI performed the best one with $c = 13$.

Similarly, the validation procedures were successfully performed on the GK clustering algorithm without handling any error exceptions. The optimal number of $c = 3$ and $c = 4$ has been produced on the agreement of PC, PE, MPC validation indexes. In contrast, CS and FS considered the more granular partition of the data to be optimal indicating $c = 12$. The XBI

[2] https://cran.r-project.org/web/packages/e1071/

[3] https://cran.r-project.org/web/packages/ClusterR/

[4] https://cran.r-project.org/web/packages/advclust/

Index	FCM with the Euclidean norm, $\mu = 2$, $\epsilon = 1e-3$, $T = 100$, $V_0 = c$ is randomly generated													
	2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
V_{PC}	**0.77**	0.64	0.57	0.54	0.53	0.52	0.55	0.54	0.56	0.58	0.59	0.57	0.57	0.60
V_{PE}	**0.36**	0.61	0.80	0.91	0.99	1.07	1.04	1.09	1.09	1.06	1.08	1.16	1.19	1.12
V_{MPC}	0.54	0.46	0.43	0.43	0.44	0.44	0.49	0.49	0.51	0.54	0.55	0.53	0.53	**0.57**
V_{CS}	**0.62**	0.52	0.44	0.46	0.51	0.52	0.50	0.48	0.52	0.56	0.56	NaN	NaN	0.55
V_{FS}	0.73	0.74	0.77	0.77	0.77	0.79	0.77	0.77	**0.80**	0.79	0.80	NaN	NaN	**0.80**
V_{XBI}	**0.16**	0.30	0.77	1.56	0.40	0.48	0.80	10.54	122.06	0.19	5.81	*	**	641
Index	PAM with the Euclidean norm, $c_{max} = 15$													
	2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
V_{PC}	**0.70**	0.60	0.55	0.58	0.56	0.56	0.56	0.56	0.56	0.59	0.59	0.60	0.60	0.60
V_{PE}	**0.43**	0.66	0.83	0.85	0.94	0.99	1.04	1.08	1.11	1.06	1.09	1.10	1.12	1.14
V_{MPC}	0.40	0.41	0.40	0.47	0.47	0.48	0.49	0.50	0.51	0.55	0.56	0.56	**0.57**	**0.57**
V_{CS}	**0.60**	0.42	0.45	0.43	0.45	0.46	0.47	0.50	0.54	0.55	0.57	0.57	0.57	0.58
V_{FS}	0.77	0.75	0.77	0.71	0.72	0.75	0.72	0.79	0.80	0.80	0.82	0.81	0.82	**0.83**
V_{XBI}	0.18	0.42	0.29	0.35	0.27	0.22	0.19	0.17	0.14	0.13	0.11	**0.10**	0.12	0.12
Index	Gustafson-Kessel $\mu = 2$, $\epsilon = 1e-3$, $T = 100$, $V_0 = c$ is randomly generated													
	2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
V_{PC}	0.73	**1.00**	**1.00**	0.99	0.99	0.89	0.92	0.89	0.93	0.96	0.95	0.94	0.93	0.95
V_{PE}	0.42	**0.00**	**0.00**	0.01	0.03	0.17	0.14	0.19	0.12	0.08	0.08	0.10	0.13	0.09
V_{MPC}	0.46	**1.00**	**1.00**	0.99	0.98	0.88	0.91	0.88	0.93	0.95	0.95	0.94	0.92	0.94
V_{CS}	0.06	0.11	0.17	0.26	0.32	0.20	0.13	0.32	-0.17	0.24	**0.35**	0.26	0.25	0.25
V_{FS}	0.07	0.11	0.17	0.26	0.33	0.25	0.12	0.36	-0.20	0.26	**0.37**	0.28	0.29	0.27
V_{XBI}	2.73	2.71	2.73	3.20	3.09	3.90	3.60	**1.19**	66.18	4.48	17.85	4.62	2.37	2.37

(1) μ fuzzification parameter; (2) ϵ threshold of convergence criteria; (3) T maximum number of iterations; (4) V_0 initial values for cluster centers; (5) c number of cluster centers

Table 2: Cluster validation of FCM, PAM and GK clustering algorithms

index found the best-case partition to be $c = 9$.

4.2 Clustering results

A crucial point in clustering involves a detection procedure of how the similarity between observations should be detected in order to generate clusters with higher similarity inside the same cluster and with lower similarity to observations from external groups. The proximity between a pair of points establishes closeness to be expressed regarding similarity, dissimilarity or a distance between a pair of points. In this way, a pair of points is considered to be close if their dissimilarity/ distance is small enough, or their similarity is considerable. To visualize the suggested partitions, the graphs given in this section are acquired by means of *fviz_cluster()* function of the *R* package *factoextra*[5]. It is the 2D visualization where *fviz_cluster()* conducts the *principle component analysis* for the provided data and maps its first two principal components to the graph.

In Figure 1, there is the data partition by FCM (for a case $c = 2$) in strict adherence to the initiated values of indexes. As is evident from the preceding, having the partition containing only two clusters may basically grant a high level of clusters' separation; even though, while trying to capture spherical shapes of partitions, the clusters drastically suffer on the compactness. As it can be seen from Figure 1, the cluster-2 center is located far from the majority of data points which are located by two different polar zones from the centroid.

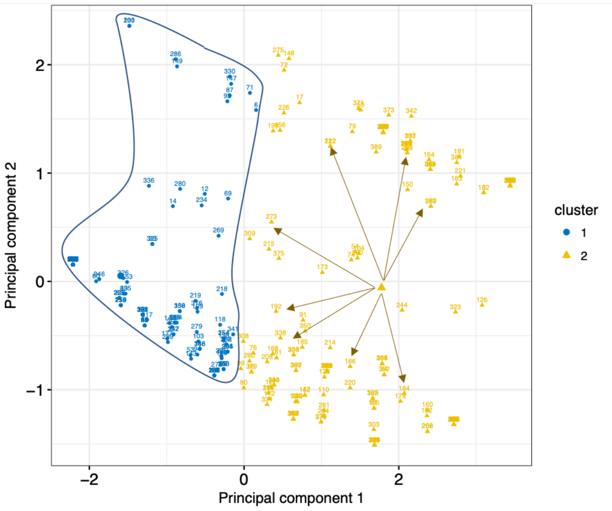

Figure 1: FCM with 2 cluster partition

Figure 2 gives a demonstration of a 15-cluster solution obtained by the PAM procedure. In contrast to

[5] https://cran.r-project.org/web/packages/factoextra

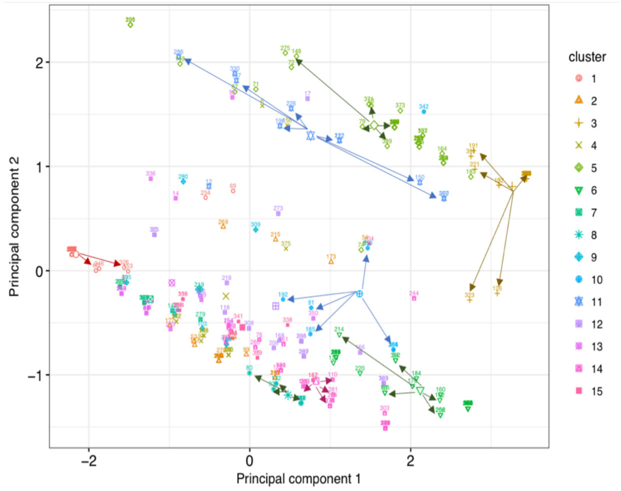

Figure 2: PAM with 15 cluster partition

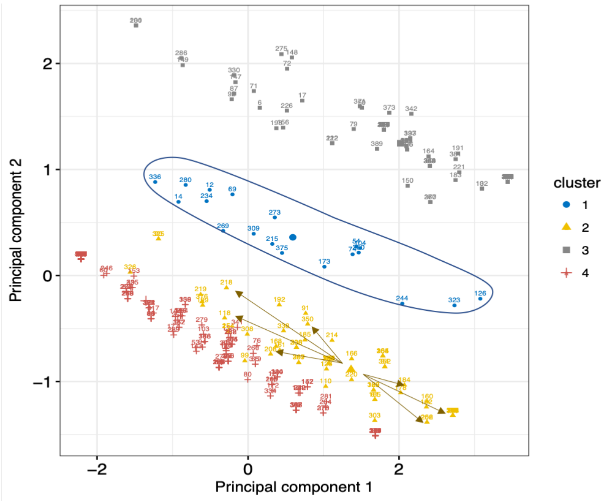

Figure 3: GK with 4 cluster partition

the previous partition, this figure demonstrates more granular data division, which was suggested according to the validation results. It can be discovered that the compactness of clusters is improved to a certain extent, while the separation of cluster centers suffers cardinally. For example, clusters 1, 6, 8 and 14 are relatively compact and small. However, they can be barely separated from other neighboring clusters whereas cluster 11 is separated clearly, but its data points are poorly compact. It is worth mentioning that while the FCM method has a tendency to gain only spherical clusters, PAM produces spherical clusters like clusters 3 and 10, as well as more ellipsoid clusters like clusters 8 and 11.

The behavior of looking for ellipsoid shapes is a feature of the GK clustering algorithm [5]. It can be seen from Figure 3 displaying 4-cluster partition of the data based on the validation results. A clear separation of all clusters is observed, for example, clusters 1, 2 and 3 are well separated from each other, at the same time each cluster has better compactness of its members. Clusters 2 and 4 have a lower degree of separation but a higher degree of compactness, which indeed could be grouped into one cluster, finally producing a three-cluster solution, like the PC, PE, MPC validation indexes (see Table 1) recommended.

The proposed optimal number of FCM and PAM clusters does not hold both qualities of compactness and separation for the partition performed. This effect may have become possible having regard to the above: the first reason is the essence of the given data array and a lack of its configuration. One gets the impression that privacy profiles exhibit quite similar characteristics, and the dataset's dimensions are highly correlated. To that, Pal et al. [16] suggest that "the unique minimum V_{PC} (or maximum V_{PE}) are rather significant in making final decisions when a pattern is not detected." In the presented data sample, the offered values of validity indexes of clusters $c = 2$ or 15 do not supply any sufficient knowledge, which indicates a fuzzy structure and scarcely detectable observations in the data sample. Accordingly, the $min(PC)$, $min(MPC)$ and $max(PE)$ values are examined. Subsequently, the $max(PE)$ values constantly work for a maximum number of clusters which does not guarantee to achieve any substantial classification. The $min(PC)$ illustrated for FCM $c = 7$, and PAM $c = 4$, while GK validation results initially agreed on the optimal $c = 4$. It is fair to assume that for the provided collection of privacy profiles, it is rather hard to detect a trade-off between compactness and separation of clusters while carrying out a worthwhile classification. Depending on the defined problem and system's requirements, a trade between compactness and separation as an optimal solution can be determined.

5 Conclusions

Previous research studies performed first attempts to measure users' privacy behavior in the multidimensional form, which was further presented as a user privacy profile. Dealing with social problems such as people's privacy behavior can involve potential issues from a data science perspective. While artificial objects like books, music, and others can be represented crisply as a set of numerical dimensions, it becomes more challenging when a similar approach is used to estimate people's behavior numerically. In many cases, human's behavior is uncertain, vague or unclear, especially when it comes to privacy decision-making. Therefore, inherently handling the uncertain nature, the fuzzy clustering algorithms become the relevant solution to be applied.

This research is an extension of the previous work [8], which introduced the concept of fuzzy privacy profiles for the first time. By applying the Fuzzy C-means clustering algorithm, people's privacy behavior was defined in the form of fuzzy profiles. As the follow-up study, this work aimed to conduct an analysis of different fuzzy clustering algorithms (PAM, GK) compared to the results of the FCM clustering. Visualization of all three clustering algorithms demonstrates the behavior of each algorithm and the differences in the suggested partitioning of the given set of data. As the application case represents users' real privacy behavior expressed in the form of privacy settings, it is harder to estimate the most optimal partitioning solution inherent to the given nature of the data.

No previous research has been done yet in order to define what μ parameter would be appropriate when dealing with people's privacy decision-making. Therefore, for future scrutiny, it will be relevant to tweak the algorithms' parameter μ with various values, which is the main fuzzification parameter in the fuzzy clustering process. Another point for future research can be considered in regards to the chosen distance norm for clustering, which can influence the final results of the data partitioning as well. Besides the commonly used Euclidean distance norm, Manhattan and Mahalanobis metrics have different measuring paths between two data points, thus, applying these norms to our data can provide additional analytical insights.

Acknowledgement

The support of Aigul Kaskina was provided by the Information System Research Group under the direction of Prof. Dr. Andreas Meier, University of Fribourg, Switzerland.

The support of Oleksii K. Tyshchenko was provided by the National Science Agency of the Czech Republic within the project TAČR TL01000351.

References

[1] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, et al., Nudges for privacy and security: Understanding and assisting users? choices online, ACM Computing Surveys (CSUR) 50 (3) (2017) 44.

[2] E. Aïmeur, S. Gambs, A. H, Upp: User privacy policy for social networking sites, in: 2009 Fourth International Conference on Internet and Web Applications and Services, 2009.

[3] S. Athey, C. Catalini, C. Tucker, The digital privacy paradox: Small money, small costs, small talk, Tech. rep., National Bureau of Economic Research (2017).

[4] C. Dong, H. Jin, B. P. Knijnenburg, PPM: A privacy prediction model for online social networks, in: International Conference on Social Informatics, Springer, 2016, pp. 400–420.

[5] D. E. Gustafson, W. C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on, IEEE, 1979, pp. 761–766.

[6] M. Johnson, S. Egelman, S. M. Bellovin, Facebook and privacy: It's complicated, in: Proceedings of the eighth symposium on usable privacy and security, 2012.

[7] R. Kang, L. Dabbish, N. Fruchter, S. Kiesler, My data just goes everywhere? User mental models of the internet and implications for privacy and security, in: Symposium on Usable Privacy and Security (SOUPS), USENIX Association Berkeley, CA, 2015, pp. 39–52.

[8] A. Kaskina, Exploring nuances of user privacy preferences on a platform for political participation, in: 2018 International Conference on eDemocracy eGovernment (ICEDEG), IEEE, 2018, pp. 89–94.

[9] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, Vol. 344, John Wiley & Sons, 2009.

[10] P. Kelley, R. Brewer, Y. Mayer, L.F.Cranor, N. Sadeh, An investigation into Facebook friend grouping., in: IFIP Conference on Human-Computer Interaction, Springer Berlin Heidelberg, 2011, pp. pp. 216–233.

[11] D.-W. Kim, K. H. Lee, D. Lee, On cluster validity index for estimation of the optimal number of fuzzy clusters, Pattern Recognition 37 (10) (2004) 2009–2025.

[12] B. Knijnenburg, A.Kobsa, H. Jin, Dimensionality of information disclosure behavior, Journal of Human-Computer Studies 71 (12) (2013) 1144–1162.

[13] B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. Zhang, N. Sadeh, A. Acquisti, Y. Agarwal, Follow my recommendations: A personalized privacy assistant for mobile app permissions, in: Symposium on Usable Privacy and Security, 2016.

[14] K. Liu, E.Terzi, A framework for computing the privacy scores of users in online social networks, ACM Transactions on Knowledge Discovery from Data (TKDD) 5 (1) (2010) 6.

[15] P. A. Norberg, D. R. Horne, D. A. Horne, The privacy paradox: Personal information disclosure intentions versus behaviors, Journal of Consumer Affairs 41 (1) (2007) 100–126.

[16] N. R. Pal, J. C. Bezdek, On cluster validity for the fuzzy c-means model, in: IEEE TRANSACTIONS ON FUZZY SYSTEMS, Vol. 3, IEEE, 1995.

[17] J. Watson, A. Besmer, H. Lipford, + Your circles: Sharing behavior on Google +, in: Proceedings of the Eighth Symposium on Usable Privacy and Security, ACM, 2012, p. 12.

[18] A. F. Westin, Privacy and freedom, Washington and Lee Law Review 25 (1) (1968) 166.

[19] P. Wisniewski, B. Knijnenburg, H. Lipford, Making privacy personal: Profiling social network users to inform privacy education and nudging, International Journal of Human-Computer Studies.

[20] K.-L. Wu, M.-S. Yang, A cluster validity index for fuzzy clustering, Pattern Recognition Letters 26 (9) (2005) 1275–1291.