Longitudinal Segmented Analysis of Internet Usage and Well-Being Among Older Adults

Alejandro Cervantes[1], David Quintana[2]*, Yago Sáez[2], Pedro Isasi[2]

[1] Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de La Rioja (UNIR), Logroño (Spain)

[2] Computer Science and Engineering Department, Universidad Carlos III de Madrid, Leganés (Spain)

* Corresponding author. dquintan@inf.uc3m.es

Received 8 June 2020 | Accepted 3 March 2022 | Early Access 2 May 2022

Abstract

The connection between digital literacy and the three core dimensions of psychological well-being is not yet well understood, and the evidence is controversial. We analyzed a sample of 2,314 individuals, aged 50 years and older, that participated in the English Longitudinal Study of Aging. Participants were clustered according to drivers of psychological well-being using Self-Organizing Maps. The resulting groups were subsequently studied separately using generalized estimating equations fitted on 2-year lagged repeated measures using three scales to capture the dimensions of well-being and Markov models. The clustering analysis suggested the existence of four different groups of participants. Statistical models found differences in the connection between internet use and psychological well-being depending on the group. The Markov models showed a clear association between internet use and the potential for transition among groups of the population characterized, among other things, by higher levels of psychological well-being.

Keywords

Aging, ELSA, Internet, Markov Models, Psychological Wellbeing, Self-Organizing Maps.

DOI: 10.9781/ijimai.2022.05.002

I. Introduction

The connection between psychological well-being and digital literacy at advanced age is an open research question at the core of a growing number of studies. Among them, only a few rely on large samples that track participants over long periods of time.

The aim of this study is providing further insights on the connection between Internet use and psychological well-being at advanced age using well-known artificial intelligence methods. The main contribution will be testing whether modeling the population as a homogeneous set causes a loss of relevant information that might be revealed by a more fine-grained segmented analysis.

We suggest clustering the population using machine-learning to subsequently fit more traditional statistical models on specific segments to assess the differential impact, if any, of Internet use on three core aspects of psychological well-being. This poses an innovation in this context that could potentially help identify connections that might have been overlooked in the literature.

We also intend to enrich the analysis exploring whether digital literacy results in differences in transition dynamics among the identified clusters over time using Markov models. This is relevant because it might show interesting patterns regarding the transition from clusters associated with higher degrees of psychological well-being to the ones with the lower one, and vice-versa.

The idea of the initial clustering, the dynamic analysis based on cluster transition, and the use of these algorithms all represent, to the best of our knowledge, technical innovations in the study of psychological well-being and digital literacy at advanced age.

The rest of the paper is organized as follows: section II will be devoted to related work and section III will introduce the materials and methods. Then, section IV will describe the experimental results. Section V will be used to discuss the results and, finally, we conclude in section VI summarizing the main conclusions and limitations.

II. Related Work

Psychological well-being is a complex construct that, according to different authors [1], [2], consists of three main dimensions: evaluative, hedonic and eudaimonic. Among these, the first one is related to the cognitive-judgmental aspect. The second one would be focused on the affective aspects, and covers feelings like happiness or sadness, and the last one would be centered on life purpose.

The expansion over last decades of Information Technologies and Communications in general, and the Internet in particular, has fostered the interest in the potential impact that these might have on psychological wellbeing. The evidence in this regard is mixed. Even though the initial studies identified an inverse association [3] subsequent ones questioned those results. Among these, some

Please cite this article as:
A. Cervantes, D. Quintana, Y. Sáez, P. Isasi, "Longitudinal Segmented Analysis of Internet Usage and Well-Being Among Older Adults", International Journal of Interactive Multimedia and Artificial Intelligence, vol. 8, no. 6, pp. 168-176, 2024, http://dx.doi.org/10.9781/ijimai.2022.05.002

suggested that connection might be weaker [4] or irrelevant [5]. Conversely, other studies [6], [7] report that using the Internet contributes positively to mental well-being.

The body of literature on the impact of Internet use among older adults is expanding [7]-[9]. Regarding psychosocial benefits, Forsman and Nordmyr [10] suggest that, in later adulthood, these might fall into three main categories: improved access to resources; empowered social inclusion and better interpersonal interaction.

Internet might function as a source of entertainment, which according to [8] has a direct connection to well-being among older-adults. Studies like [11] suggest that it might also double as a facilitator, fostering engagement with other activities. There is, however, a third possibility that might be especially relevant for older adults. According to some studies [12], [13], loneliness and social isolation are emerging risks in this population segment. For this reason, identifying tools to mitigate or solve these problems has the potential to have a major social impact. Having said that, the existence of positive features of digital literacy in this regard is a matter that is still under debate [14].

The most closely related study [15] explores the connection between the mentioned three dimensions of psychological well-being and Internet use in older adults relying on evidence gathered by the English Longitudinal Study of Aging (ELSA) [16].

This project has been collecting data on a bi-yearly basis from a representative sample of the English population aged 50 and older since 2002 to gain a better understanding of the ageing process. The study tracks a broad range of items that include aspects related to physical and mental health, economic position, or social participation, among others. The project is related to similar studies like the US Health and Retirement Study (HRS), the Survey of Health, Ageing and Retirement in Europe (SHARE) or the Japanese Study of Aging and Retirement (JSTAR), to mention a few.

According to these results, the connection between the main predictor and the scores on the scale used to measure the eudaimonic aspect was positive and statistically significant. However, they did not support the relevance of digital literacy on the evaluative and the hedonic components of psychological well-being.

This begs the question of whether the latter lack of connection is real or might be explained by the fact that the analysis considers the population to be homogeneous when that might not be case. Therefore, in this paper we perform a segmented analysis based on clusters identified using Self-Organizing Maps [17] instead of a global one; the objective is to confirm whether the conclusions in [15] are valid for the complete population or should be evaluated on the light of this segmentation. This research will contribute to gain a better understanding on how belonging to each of these groups impacts the connection between Internet literacy and psychological well-being. Finally, as users transition between groups over successive waves of the study, we shall also be able to perform an analysis on the evolution of users in the population.

III. Materials and Methods

A. Study Population

The data analyzed in this piece of research matches the one used in a previous study [15]. They were originated from the English Longitudinal Study of Aging. This survey tracks on a bi-yearly basis the evolution of aging and quality of life among older people in England. The sample covers community-dwelling population aged 50 and over. Among the wide range of aspects that included in the core component, we could mention health-related items; household and individual demographics; social participation; income and assets; expectations etc. Some waves supplement this information with one-off modules and questions.

The study that we present is based on data from waves 3 to 7, that is, the interviews were carried out in 2006-07, 2008-09, 2010-11, 2012-13 and 2014-15. At wave 3, 9,771 individuals took part in the study. Out of these, we lacked complete baseline data for 2,814, who were removed. Between waves 3 and 7, many individuals were lost to follow up. This further reduced the sample size to 3,547 participants by wave 7. Out of these, the information was complete over the five waves for 2,314 individuals over 50, who comprise the final dataset. We provide a table that summarizes the demographics in the Appendix.

Attrition was associated with less educated individuals that tended to show a higher degree of functional impairment. These participants, who were slightly older, also reported lower net wealth and lower digital literacy.

B. Measurements

As we mentioned, the study considers the three core dimensions of psychological well-being together with an indicator of internet literacy. The former was quantified using the indicators suggested in [18], and the second with a specific questionnaire item.

Evaluative well-being was evaluated using the Satisfaction with Life Scale (SWLS) [19]. This indicator is defined in the range 0-30, being the higher values associated with the greatest satisfaction with life. In regard to hedonic well-being, it was measured using the Enjoyment of Life Scale (EOLS). This indicator, already used in other studies [20], [21], is defined in the range 0-12. Once again, higher scores are associated with higher enjoyment of life. The instrument used to measure eudaimonic well-being consists of the items of CASP-19 not considered in EOLS (EDS). The minimum score is 0 and the maximum 45. As in EOLS, the higher the score, the greater is the eudaimonic well-being. Finally, the study proxies internet literacy through the answer to the question "I use internet/E-mail: yes/no". The resulting dichotomous variable was encoded using (0) for negative responses and (1) for positive ones.

In addition to these measurements, the analysis also controlled for socioeconomic and health indicators. The former included age, whether the individual was a woman (1) or a man (0) and the highest academic qualification using a 3-way split to represent whether the participant had no qualification, an intermediate one, or a degree or equivalent. Given that marriage tends to be associated with more well-being [22], [23], we also considered whether the participant was legally married (1) or not (0). This was complemented with net non-pension household wealth, identified as important in previous studies [24], a dichotomous variable that indicates whether the subject volunteered (1) during the previous year or not (0), relevant according to [25], [26] and associative interests. This proxy for social connectedness, identified to have an impact on psychological well-being in studies like [26], measured the diversity of organizations that the subject reported to be part of among eight possible broad categories including, among others political parties, religious groups.

Physical activity: self-reported physical activity encoded using four consecutive levels, from 0 to 3, that represent the categories sedentary, low, moderate, and high, respectively. Finally, it was considered high in case it involved heavy manual work or vigorous activity more than once a week. The study also considers whether the participant reported having suffered limitations in instrumental activities of daily living (IADSLs) or activities of daily living (ADLs) caused by mental, physical, memory or emotional problems for a period over three months (0) or not (1). Finally, the instrument used to measure cognitive ability was the learning recall test included in the Consortium to Establish a Registry for Alzheimer Disease (CERAD) Neuropsychological battery [27]. Here, higher scores are associated with higher cognitive abilities.

Table I reports the main descriptive statistics for all these baseline characteristics of the analytical sample measured at Wave 3 [16].

The factors and covariates were measured at all follow-up interviews except for age, sex, education, wealth, and marital status, which were only measured at baseline.

TABLE I. Baseline Characteristics of the Analytical Sample Measured at Wave 3. English Longitudinal Study of Aging 2006-07. Main Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
SWLS Score*	20.49	6.159	0	30
EOLS Score*	10.13	1.624	2	2
EDS Score*	32.94	6.589	6	45
Internet/Email User	0.66	0.475	0	1
Delayed Recall	5.32	1.778	0	10
Physical Activity	2.09	0.719	0	3
Org. membership	1.79	1.410	0	8
Voluntary Work	0.39	0.487	0	1
Age	61.62	7.668	50	99
Sex	0.55	0.498	0	1
Marital Status	0.75	0.431	0	1
Education	1.18	0.807	0	2
Lack of impair.	0.85	0.360	0	1
Net wealth**	412.7	785.9	-51.9	20.818

* Scales used to measure the tree core components of psychological wellbeing: evaluative (SWLS), hedonic (EOLS) and eudaimonic (EDS).

** Net wealth in thousands of pounds.

C. Analytical Approach

The analytical strategy followed in this study combines three different instruments that will be used sequentially. The initial step will be clustering the participants according to drivers of psychological well-being using Self Organizing Maps. Then, the influence of Internet/Email use on the three main dimensions of psychological well-being by group will be assessed generalized estimating equations. Finally, the dynamic aspects of transitions among clusters will be explored using Markov models.

D. Clustering Based on Self-Organizing Maps

Self-Organizing Maps (SOM) [17], [28] is a type or artificial neural network frequently used to perform clustering analysis. This method adapts a grid of "neurons" to a specific topology. These grids are a powerful representation technique because the multidimensional topology of data can be projected in the two-dimensional space defined by the relative positions of neurons on the grid. After training, each neuron is characterized by a set of features (codebook) that summarize the features of data in its vicinity.

SOM has been used in many applications due to its simplicity and accuracy in unsupervised learning. Its learning rule can be more efficient than competing network architectures for large datasets and high dimensionality. SOM can be quite efficient to perform an unsupervised preprocessing step that is later classified by another algorithm. For instance, SOM has been used to reduce noise and dimensionality on protein classification problems [29]. The output of this phase was processed by a Particle Swarm classification. Also, it has been proved useful as part of a methodology in Human Sentiment Classification [30]. In that work SOM is used to cluster the initial patterns into separate groups, that are later classified using a Convolutional Neural Network. This approach improves the accuracy of classification over the competing techniques.

The Self Organizing map is a discrete interconnecting network (that is, a map) of neurons (also called "units"). This map is adapted to a provided training set by minimizing a loss function, the quantization error. The learning process of SOM can be summarized as follows:

First, the map is initialized. Original SOM used random initialization, but recent versions typically use Principal Component Analysis for this task. Each input neuron has N inputs, as many as features as inputs. A vector of numerical weights w_i is associated to each neuron. Thus, each neuron has a representation in the N-dimensional input space.

Upon initialization, neurons are interconnected in a specific way (for example a two-dimensional matrix), where grid distances can be measured to gauge the degree of influence among neurons.

During training, the following adaptation process is applied for each input vector x.

1. Determine the best matching neuron (BMU), as the one with the minimum Euclidean distance to the input vector (1):

$$BMU = argmin_i \| x - w_i \|^2 \tag{1}$$

2. Adjust the weights of the BMU and also some or all the neurons in the map are "moved" closer to **x**; this is summarized using the following learning rule for all neurons *j* (2):

$$w'_j = w_j + \eta \, H(BMU, j) \; (x - w_j) \tag{2}$$

3. The process is repeated for all the training values over a certain number of epochs or iterations.

In the learning rule above, η is a small learning rate that is used to tune the speed of the algorithm convergence. The matrix **H** is called *function of lateral interaction*, and is composed of positive numbers that determine how intense is the modification of the weights of neurons that are neighbors of the BMU on the grid. Neurons directly connected to the BMU are "dragged" more than neurons re connected at distance 2, etc. Different versions of this function **H** exist, but they all have in common that **H** is dynamic during learning. At the start of the SOM training, influence is more global: influence is significant for a certain radius of influence from the BMU; and at the end of the process, adaptation is mostly local, that is, only the BMU and maybe its direct neighbors on the grid are adapted.

One of the advantages of SOM is that is a model-based clustering method. Once trained, the SOM network weights are retained and can be saved as a model that can be used later when new patterns become available. Thus, a SOM can be trained with the first wave of data, and later the evolution of each customer's record can be followed to check if the cluster to which it is assigned changes in later waves, thus tracking its evolution over time.

Adaptation of the SOM is only the first step of the process. The resulting model is an approximate representation of the distribution on the original data, but with a much more limited number of elements. These elements are further grouped down to generate a manageable number of clusters. This latter stage relied on Hierarchical Agglomerative clustering (HAC) [31]. This method starts with as many seeds as initial elements to be clustered. Then, recursively selects the two closest seeds in terms of the desired criterion and generates a conglomerate element with averaged values for its features. The process continues until the number of elements matches the number of desired clusters, at which point the algorithm stops.

1. Cluster-Specific Analysis Using Generalized Estimating Equations

The clusters identified in the first step of the process represent the main broad categories in which we can classify the participants according to the socioeconomic and health drivers of psychological well-being introduced in section 2.2. Once the individuals were assigned to the different groups, it was possible to perform an exploratory segmented analysis with the potential to reveal dependencies that might be difficult to identify when one considers the whole population.

The second part of the analysis relied on generalized estimating equations. These models, which are closely related to generalized linear models, provide the capability to study population-averaged effects across repeated measurements [32]. This allowed us to assess the influence of digital literacy on psychological well-being by dimension and segment of the population, controlling for potentially relevant covariates.

2. Markov Models

Markov models are commonly used for analysis of the temporal dynamics, though specific techniques used depend facts and assumptions on available data. In [33], the authors presented a generic framework for this analysis. In literature the most referenced method for parameter estimation is Expectation-Maximization (EM) [34].

A characteristic of panel data is that it can be more properly described as a mixture of models, where different groups of participants show different behavior in the temporal dimension. These models are called Latent Segment Markov Chain models. For instance, in [35] authors address how market segmentation helps providing insight on the different models that apply to each segment, without prior knowledge of the segment to which specific users belong. Some reviews of usage of these techniques in analysis can be found in [36].

Markov modelling has been used for the variation in social network structures [37] or psychological evaluation of patients [38] where specific randomization techniques are introduced to take into account interpersonal variations in the population.

In general, these models are fitted to provide both quantitative predictions on the unobserved variables (model states) and qualitative descriptions of the temporal variation on data. This type of study may provide insight even in cases where data are insufficient to provide statistically reliable predictions.

HMMs were fitted and plotted using the R package seqHMM [39].

IV. Results

This section reports the experimental results. To that end, it starts describing the experimental setup. Then, it focuses on the cluster analysis and the group-specific statistical analysis. Finally, it discusses the dynamic analysis.

A. Experimental Setup

The purpose of this work is to examine in depth the relationship between Internet usage and each of the three measures of well-being. We suspect that these relationships can't be properly assessed by analyzing the joint population of participants in the study. Thus, a more detailed analysis was performed by introducing a preliminary stage that groups participants based on the values of the covariates, the aforementioned Self-Organized Map based clustering.

Generation of the SOM maps and construction of the dynamic models was performed on a standard Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz machine with 16 GB of memory.

The overall experimental procedure was a sequence of three steps, that we cover in the following sections. As a summary, these three steps were:

Cluster analysis. The values of indicator and predictor variables were removed from the data set.

The quality of a SOM can be measured using several metrics. In this analysis we used two: topographic error and quantization error. Topographic error counts the number of times that, for any given sample vector, the second-closest neuron is not located in the immediate neighborhood of the closest neuron. If this error is low, the SOM is accurately representing the topology of data. Quantization error is calculated as the average distance from each vector to the closest neuron codebook. We tested three different grid configurations (5×5, 10×10 and 15×15). We verified that as number of neurons grew, topographical error increased, while quantization error decreased. Thus, for practical considerations we have considered that an intermediate map of 10 × 10 neurons was enough to provide a topographically accurate representation ($T_{error} = 8 \cdot 10^{-4}$) with adequate quantization error ($Q_{error} = 7.43$).

The second step was to apply the HAC algorithm to generate a small number of clusters. From the cluster u-matrix representation we chose four clusters as target. The result is an assignment of all individuals in the first wave to one of these four groups.

The software used for this was SOMbrero [40] an R package. Data loading and filtering, and generation of the SOM and clusters took 15 sec in our platform.

Group-specific statistical analysis. This analysis was performed using generalized estimating equations, so we can determine whether the conclusions are dependent on the group to which customers belong.

We fitted *c*x3 models, where *c* is the number of clusters identified by SOM. For every cluster, there were three different models, each of them targeting the scores of the relevant scales as the dependent variable (SWLS for the evaluative dimension, and EOLS and EDS for the hedonic and eudaimonic ones, respectively). All models shared the main predictor and the covariates. Following [15] and other previous studies, age and wealth were stratified. The former considered the intervals 50-59, 60-69, 70-79 and >79, and the latter quintiles.

The computation of the regression coefficients of the 2-year lagged models and their associated confidence intervals relied on model-based estimations of the covariance matrices. The statistical contrast used to assess the significance con the coefficients was the Wald test.

The software package used both to fit the models and evaluate the results was SPSS 23.

Dynamic Analysis. This analysis was performed using the map generated initially for the initial wave, but applying it to the successive waves. We separated users between in two populations: those who started with an affirmative answer to the question on Internet use (Internet users), and those who answered negatively. Then, we identified the group to which each customer belonged for different study waves. Then we constructed Markov Models that represent the transitions found between successive waves for all users. This was performed independently for both possibilities of the "Internet Use" variable in the initial wave.

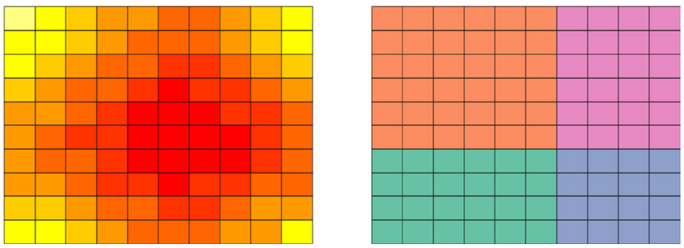
The construction of Markov models using data partitioned into groups is a mechanical task, given that we already had identified the number of possible states (4) for each customer. It is performed by calculating the conditional transition probability $P(C_{w=i+1}|C_{w=i})$ for all members of the study, and the starting $P(C_{w=0})$ for every individual.

The software used for this was self-programmed R code. Generating this model took less than 1 sec in our experimental platform.

B. Cluster Analysis

Automatic clustering was first performed on wave 3 of the available data. The resulting clusters, (hereafter called "groups") identified in this first wave are later used as reference to monitor evolution over time, by incorporating waves 4 to 7. Clustering was performed in two stages: first, a Self-Organizing Map (SOM) was used to generate a set of neurons that closely represent distribution of the covariate attributes of each participant; secondly, a hierarchical clustering algorithm was used to separate neurons in four groups. Indicator (well-being scores) and predictor variables (Internet usage) are then examined on the different groups to see if their inter-relationships are of a different nature depending on the group.

Fitting the basic SOM to the sample results in the structures illustrated in Fig. 1a. There we can see an *umatrix* representation of the SOM grid obtained for a SOM with 10 × 10 neurons. This representation depicts Euclidean distance between the neuron codebooks by using lighter colors for closer distances and darker colors for longer distances Thus, darker (red) areas can be used to separate groups of neurons that represent individuals whose features (values of the covariates in our case) are more distant.



(a) Umatrix for the 10 x 10 grid.

(b) Cluster assignments.

Fig. 1. SOM umatrix charts for wave 3 and corresponding groups after clustering.

The second step was to apply the HAC algorithm to generate a small number of clusters. Visual inspection of Fig. 1a clearly suggests the existence of 4 clusters, as four distinct areas are identified (in yellow) where neurons are grouped, separated from the rest with darker areas. Given this number as input, the HAC super-clustering process divided the map as indicated in Fig.1b, where each color represents a different cluster.

TABLE II. Averages for Each Group on Indicator and Predictor Variables, Sorted by Average Value of Internet Use

	SWLS		EOLS		EDS		Int. Use	
	Score	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
Group 1	21.90	1	10.50	1	34.66	1	0.84	1
Group 2	20.65	3	10.26	2	33.61	2	0.77	2
Group 3	21.06	2	10.13	3	32.76	3	0.48	3
Group 4	18.53	4	9.57	4	30.41	4	0.43	4

In Table II we average and rank the values of indicator and predictor variables for each group of individuals. Group numbers have been selected in order to match the degree of Internet use from Group 1 (highest) to 4 (lowest). It is immediate to see that values for two of the three components of psychological well-being scores (EOLS and EDS) are sorted in the same descending order. Therefore, regarding these two measures of well-being, Group 1 has the highest level of psychological well-being and also the highest level of Internet use, and Group 4 has the lowest level for both measures.

However, this generic correlation is not true for all groups regarding SWLS: in this case, Group 3 ranks second in the SWLS score, over Group 2 which is third, though differences are not great.

Average Internet use in Groups 1 and 2 is very similar, and there is also a small difference between Groups 3 and 4. However, well-being scores do not show such division, and present a more gradual distribution.

In order to evaluate the properties of each of the groups, we have averaged the values of all covariates for each cluster in Table III. Here we see that some of the covariates follow the same ordering and might be equally correlated to psychological well-being: Cognitive ability (Delay Recall), qualification (Education) and net non-pension income (Net Wealth). Others, such as degree of membership to organizations and Volunteer work seem to account for much of the distinction between clusters 1 and 2, and also between clusters 3 and 4.

C. Group-Specific Statistical Analysis

The statistical analysis results are summarized in Table IV. There, we report the main coefficients of the 12 GEE models, one per combination of cluster and scale, together with the associated p-values obtained using the Wald test. These represent the differential impact of digital literacy on the results of the three scales used to proxy the three key dimensions, controlling for all the covariates, as discussed in 2.3.2.

If we focus our attention on those that are statistically significant at the 5% conventional level, the use of Internet/Email had a generally positive impact. However, the beta coefficients of the 2-year lagged models reveal the existence of differences among clusters. This is especially noticeable in cluster 3, as the model suggests an inverse relationship between Internet literacy and the hedonic dimension.

TABLE IV . Summary of GEE Analysis of Connection Between Internet Use and Psychological Well-Being Test Scores by Cluster, English Longitudinal Study of Aging 2006-14

	SWLS Score[a]		EOLS Score[a]		EDS Score[a]	
	Coeff.[b]	**P[c]**	**Coeff.[b]**	**P[c]**	**Coeff.[b]**	**P[c]**
Group 1	0.24	.45	0.12	.17	0.44	.22
Group 2	0.47	.06	0.04	.59	0.60	.03
Group 3	-0.29	.57	-0.24	.002	-0.25	.43
Group 4	0.50	.04	0.07	.006	0.87	.001

[a] Scales used to measure the tree core components of psychological well-being: evaluative (SWLS), hedonic (EOLS) and eudaimonic (EDS).
[b] Beta regression coeffcients estimated through 2-year lagged generalized estimating equations.
[c] P values from Wald test.

D. Dynamic Analysis

In this section we are concerned by how individuals migrate from one group to another during the successive waves of the study.

In Table V we show the variation on the population in the formerly calculated groups. These figures were calculated on the individuals

TABLE III. Average Values of Covariates by Group. Group 1 Has the Highest Average Internet Use, and Group 4 the Lowest

	N[a]	Age[b]	Sex[c]	MrS.[d]	Edu.[e]	NW[f]	Imp.[g]	Phys.[h]	Rec.[i]	Org.[k]	Vol.[l]
Group 1	590	59.81	0.55	0.84	1.68	624.1	0.94	2.38	6.12	2.93	0.94
Group 2	798	58.51	0.43	0.83	1.41	419.0	0.96	2.39	5.66	1.46	0.00
Group 3	336	68.21	0.61	0.68	0.99	326.5	0.71	1.83	4.32	2.09	0.93
Group 4	590	63.86	0.68	0.59	0.49	241.6	0.67	1.56	4.65	0.94	0.03

[a] Participants; [b] Age; [c] Sex; [d] Marital status; [e] Education; [f] Net Wealth (Thousands of pounds); [g] lack of impairments; [h] Physical activity; [i] Delayed recall; [k] Organization membership; [l] Voluntary work

that were present in the five waves of the study. It is obvious that ageing must have an overall impact that is easily shown in the total number of individuals that compose each of the groups. The number of participants in Group 3 increases by a factor of 1.76 between wave 3 and wave 7, and Group 4 increases by a factor of 1.34. On the other hand, participants in groups of younger average age decrease with time: Group 1 decreases by factor of 0.65, and Group 2 by 0.68.

TABLE V. Evolution of the Number of Individuals Per Cluster. Groups Are Calculated for Wave 3. For Other Waves, Each Individual Is Assigned to the Group to Which the Closest Codebook in the SOM Grid Belongs

	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
Group 1	590	522	498	455	385
Group 2	798	749	649	587	543
Group 3	336	401	471	528	593
Group 4	590	642	696	744	793

As we can see in Table VI, with 2314 participants, we have 9256 possible transitions. The diagonal totals 6374, that is in 68.86% of the cases individuals do not change the group to which they belonged to at the start of the study.

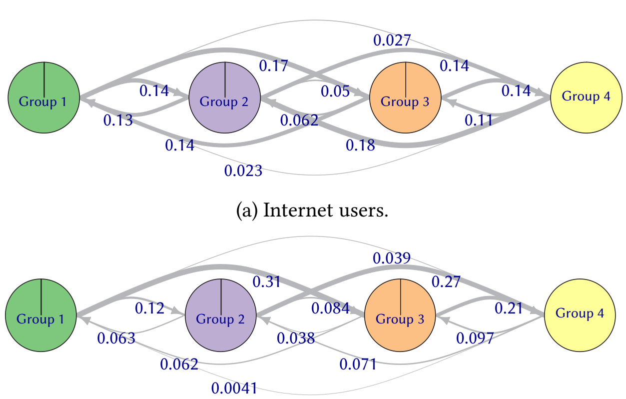
TABLE VI. Transition Table in Number of Cases. In Rows, the Starting Group (in Any Wave From 3 to 6); in Columns the Destination Group in the Following Wave (4 to 7)

	Group 1	Group 2	Group 3	Group 4
Group 1	1328	288	389	60
Group 2	321	1828	159	475
Group 3	177	89	1174	296
Group 4	34	323	271	2044

This information may be used to construct a transition table that details the probability of a user to either move to, or stay in any group, depending on the original group. This is equivalent to constructing two independent first-order Markov chain models of group labels.

We have constructed the global Markov chain for group transitions using the classical EM method [34] on the full chains of four elements. We must point out that these figures are only approximate, as we are averaging results for the four transitions between different waves that can't be modelled as time-homogeneous.

The resulting transition probabilities are shown graphically in Fig. 2. In this figure, arrow width is a measure of the transition probability value.



(a) Internet users.

(b) Non-internet users.

Fig. 2: Cluster transition diagrams. Group 1 has the highest scores in wellbeing, while Group 4 has the lowest. Transition probabilities printed on edges.

For clarity, in Table VII and VIII we show the transition probabilities in these figures.

TABLE VII. Transition Probabilities for Internet Users. In Rows, the Starting Group (in Any Wave From 3 to 6); in Columns the Destination Group in the Following Wave (4 to 7)

	Group 1	Group 2	Group 3	Group 4
Group 1	0.662	0.143	0.168	0.027
Group 2	0.130	0.678	0.050	0.142
Group 3	0.135	0.062	0.664	0.139
Group 4	0.023	0.182	0.106	0.688

TABLE VIII. Transition Probabilities for Non-Internet Users. In Rows, the Starting Group (in Any Wave From 3 to 6); in Columns the Destination Group in the Following Wave (4 to 7)

	Group 1	Group 2	Group 3	Group 4
Group 1	0.533	0.121	0.307	0.039
Group 2	0.063	0.580	0.084	0.237
Group 3	0.062	0.038	0.692	0.208
Group 4	0.004	0.071	0.097	0.828

V. Discussion

The SOM analysis resulted in the identification of four groups of participants whose main characteristics were summarized in Table III. The first one showed high values for all the well-being scores and the highest internet use. This group had the second lowest average age. It showed a high level of physical activity that corresponded to the lack of impairments. This group also showed high level of participation in organizations. The second group had a high use of internet and high values for all the well-being scores except for the SWLS score. Level of activity and lack of impairments levels were the same as in group 1. However, there was a clear difference in net wealth, level of organization membership and voluntary work. There was also a preponderance of males compared to group 1. Group 3 had the highest average age, and covariates such as marriage status, education, physical activity, and net wealth were lower than the previous groups. It also showed a much lower internet use. Conversely, it was characterized by higher organizational membership and voluntary work. Finally, the fourth group was characterized by the least internet use and low well-being scores, even though average age was lower than members of group 3. This group was also characterized by the lowest education and physical activity levels, the highest preponderance of females, and lowest score in marital status. This group shows a very low degree of membership in organizations and a very low score in voluntary work.

The results of the GEE models supported the basic hypothesis: studying the population as a whole causes a loss of relevant information vs. a more fine-grained segmented analysis.

Internet literacy does not seem to have any significant connection with psychological well-being for the participants in group 1. The rest of the groups present different associations with the different scales of well-being:

Internet use seems to have an association with the SWLS score for groups 2 and 4, though this is only significant at 5% for group 4.

For the eudeimonic (EDS) score, this connection is significant at 1% for both groups 2 and 4.

Results for the hedonic (EOLS) score were particularly interesting. We found a positive association from Internet use to EOLS for group 4. However, the most interesting result in this regard is the existence of a clear negative relationship with enjoyment of life for group 3.

There are some differences with results obtained by Quintana et al. when analyzing the whole sample [15]: first, in that work neither SWLS nor the EOLS score could be associated with Internet use at conventional levels, whereas we have found groups for which that association can be significant; secondly, the significant association of Internet use and the score in the EDS scale reported in the work on the aggregated data is now shown to be related (and significant at 1%) to participants in groups 2 and 4 (approximately 60% of the sample); finally, the negative association in group 3 was not detected when considering the whole sample.

The dynamics captured by the Markov models show markedly differentiated transition probabilities among groups depending on digital literacy. In Fig. 2, arrows from left to right mean (in general) decrease in the wellbeing scores. Those transitions are much more balanced for Internet users: internet users are more likely to move in both directions, where non internet users are more likely to decrease their levels of well-being over time. In fact, non-Internet users have an 83% probability of staying in the lowest-scoring group (4) once they fall in it, and their probabilities of reaching that state from groups 2 and 3 are much higher.

The findings derived from both the static and the dynamic analysis open the door for more targeted research that could provide more insights on the connection between Internet use and well-being and implementation of better targeted intervention programs.

Current literature is yet to reach a definitive conclusion on the relation between aspects of psychological well-being and Internet literacy among the older population [14]. There are several important problems to common approaches, both from the methodological and from the data collection point of view, especially in terms of the impacts of different types of Internet use [8]. Our work confirms that segmentation of data may provide significant insights that help comprehend seemingly contradictory results. For instance, some studies report positive relationships between these variables: use of Internet and self-reported life satisfaction in [7], or with the hedonic dimension of well-being [6], as it reduces the probability of depression very significantly. On the other hand, negative or mixed relationships among certain uses of the Internet and perception of wellbeing have also been documented. In [41] authors use of Internet for communication with unknown people is a symptom of feeling of loneliness, while communication with family members reduces those feelings. Also [42] points out that frequent use of Internet may have a positive association with depression.

We must point out some challenges of this study that have to do with the representativity of the sample. Our data is based on ELSA and thus has been recorded on various geographical locations in England. Evidence from previous literature suggest that most research conclusions on ageing studies can only be generalized to countries on with similar levels of development. What is more, some studies on aging point out differences among results from European elderly population and American counterparts [43]. The sample included participants aged 50 years or older not living in assisted living or nursing homes, and those speciffic living conditions might not be directly extrapolated to the general population.

Finally, we will point out a limitation regarding data availability. Most of the waves considered in this study provide very limited information on the type of use of Internet: only the dichotomous response used as independent variable is available. More details on intensity and specific uses would leave room for more detailed analysis, along the lines the one described by Hofer et al. [44] on online information seeking. We hope that we will be able to go deeper into the analysis and get a clearer picture once data on new waves gets released over the next years.

VI. Conclusions

This study provides new insights on the connection between Internet use and three core dimensions of psychological well-being at advanced age.

The results support our two initial hypothesis: the existence of a segmented population in terms of the main drivers of well-being, and the importance of performing a fine-grained segmented analysis.

The existence of four clusters and the differential impact of digital literacy depending on the group opens the door to further research and the development of specific interventions. The latter is especially relevant in the light of the dynamic analysis, as there seems to be a clear association between this factor and the potential for transition to segments of the population characterized by higher levels of psychological well-being.

From an instrumental point of view, the results support a high potential for Self-Organizing Maps and Markov models in this domain.

Appendix

Baseline Characteristics of the Analytical Sample by Psychological Well-Being Indicator. English Longitudinal Study of Aging 2006–2007

	n (%)	SWLS Score *		EOLS Score *		EDS Score *	
		Mean	Std.	Mean	Std.	Mean	Std.
Internet/Email User							
No	792 (34%)	19.77	6.368	9.93	1.781	31.20	6.983
Yes	1522 (66%)	20.86	6.015	10.24	1.526	33.59	6.280
Age							
50–59	1059 (46%)	19.58	6.559	10.05	1.708	32.74	6.721
60–69	839 (36%)	20.82	5.983	10.20	1.539	33.28	6.719
70–79	385 (17%)	21.01	5.355	10.16	1.577	32.62	6.076
>79	31 (1%)	22.52	4.434	10.03	1.791	34.10	6.156
Sex							
Male	1041 (45%)	20.64	5.934	10.04	1.648	32.85	6.370
Female	1273 (55%)	20.36	6.336	10.12	1.600	33.01	6.765
Education							
None	579 (25%)	20.11	6.220	9.94	1.699	31.65	6.984
Intermediate	729 (32%)	19.98	6.391	10.02	1.640	32.78	6.639
Degree	1006 (43%)	21.07	5.904	10.22	1.625	33.79	6.183
Lack of Impairments							
Yes	354 (15%)	16.64	7.169	9.15	1.806	28.27	7.152
No	1960 (85%)	21.05	5.786	10.31	1.523	33.78	6.115
Marital Status							
Single	572 (25%)	18.05	6.896	9.86	1.659	32.49	6.884
Married	1742 (75%)	21.29	5.676	10.22	1.602	33.08	6.485
Physical Activity							
Sedentary	42 (2%)	16.57	7.979	8.86	1.555	27.51	8.055
Low	373 (16%)	19.34	6.709	9.65	1.750	30.50	7.065
Moderate	1225 (53%)	20.48	5.975	10.12	1.614	32.99	6.331
High	674 (29%)	21.37	5.849	10.49	2.125	34.52	6.099
Voluntary Work							
No	423 (61%)	19.80	6.438	9.99	1.689	32.34	6.875
Yes	891 (39%)	21.59	5.511	10.35	1.489	33.89	5.988
Wealth Quintile †							
Q1	290 (12%)	17.27	7.517	9.42	1.863	29.48	7.655
Q2	383 (16%)	19.34	6.539	9.80	1.728	31.24	7.075
Q3	479 (21%)	20.38	5.919	10.19	2.307	32.51	6.360
Q4	548 (24%)	21.27	5.393	10.34	1.497	33.73	5.896
Q5	614 (27%)	22.10	5.260	10.43	1.489	35.25	5.379
Org. membership Delayed Recall	2314 (100%)	20.49	6.159	10.13	1.624	32.94	6.589

* Scales used to measure the tree core components of psychological well-being: evaluative (SWLS), hedonic(EOLS) and eudaimonic (EDS).

† Quintile distribution based on the initial unfiltered sample, not the analytical one. Higher quartiles represent more wealth.

Acknowledgment

The data were made available through the UK Data Archive. ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute for Fiscal Studies. The data were collected by the National Centre for Social Research. The funding is provided by the National Institute of Aging in the United States and a consortium of UK government departments coordinated by the Office for National Statistics. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here.

This work has been supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3MXX), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

References

[1] D. Kahneman and A. Deaton, "High income improves evaluation of life but not emotional well-being," In *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, pp. 16489–93, 2010.

[2] P. Dolan, R. Layard and R. Metcalfe, "*Measuring Subjective Wellbeing for Public Policy: Recommendations on Measures*," CEP Special Papers Centre for Economic Performance, LSE, 2011.

[3] R. Kraut, M. Patterson, V. Lundmark, S. Kiesler, T. Mukopadhyay and W. Scherlis, "Internet paradox: A social technology that reduces social involvement and psychological well-being?," *The American psychologist,* vol. *53*, pp. 1017–31, 1998.

[4] R. Kraut, S. Kiesler, B. Boneva, J. Cummings, V. Helgeson and A. Crawford, "Internet paradox revisited," *Journal of Social Issues,* vol. *58*, pp. 49–74, 2002.

[5] L. A. Jackson, A. von Eye, G. Barbatsis, F. Biocca, H. E., Fitzgerald and Y. Zhao, "The impact of internet use on the other side of the digital divide," *Communications of the ACM,* vol. 47, pp. 43–47, 2004.

[6] S. R. Cotten, G. Ford, S. Ford and T. M. Hale, "Internet use and depression among older adults," *Computers in Human Behavior,* vol. 28, pp. 496–499, 2012.

[7] O. Lelkes, "Happier and less isolated: internet use in old age," *Journal of Poverty and Social Justice,* vol. 21, pp. 33–46, 2013.

[8] R. Lifshitz, G. Nimrod and Y. G. Bachner, "Internet use and well-being in later life: a functional approach," *Aging & Mental Health,* vol. *22*, pp. 85–91, 2018.

[9] B. B. Neves, R. Franz, R. Judges, C. Beermann and R. Baecker, "Can digital technology enhance social connectedness among older adults? A feasibility study," *Journal of Applied Gerontology,* vol. 38, no. 1, pp. 49–72, 2017.

[10] A. K. Forsman and J. Nordmyr, "Psychosocial links between internet use and mental health in later life: A systematic review of quantitative and qualitative evidence," *Journal of Applied Gerontology*, vol. 36, pp. 1471–1518, 2017.

[11] M. Näsi, P. Räsänen and O. Sarpila, "ICT activity in later life: Internet use and leisure activities amongst senior citizens in Finland," *European Journal of Ageing,* vol. 9, pp. 169–176, 2012.

[12] N, Savikko, P. Routasalo, R. Tilvis, T. Strandberg and K. Pitälä, "Predictors and subjective causes of loneliness in an aged population," *Archives of Gerontology and Geriatrics,* vol. 41, pp. 223–233, 2005.

[13] P. A. Dykstra, T.G. van Tilburg and J. de Jong Gierveld, "Changes in older adult loneliness: Results from a seven-year longitudinal study," *Research on Aging,* vol. *27*, pp. 725–747, 2005.

[14] R. Beneito-Montagut, Y de, N. Cassián and A. Begueria, "What do we know about the relationship between internet mediated interaction and social isolation and loneliness in later life?," *Quality in Ageing and Older Adults*, vol. 19, no. 1, pp. 14–30, 2018.

[15] D. Quintana, A. Cervantes, Y. Sáez and P. Isasi, "Internet use and psychological well-being at advanced age: Evidence from the English longitudinal study of aging," *International Journal of Environmental Research and Public Health,* vol. 15, no. 3, art. 480, 2018.

[16] J. Banks, G. D. Batty, K. Coughlin, P. Dangerfield, M. Marmot, J. Nazroo, Z. Oldfield, N. Steel, N., Steptoe, M. Wood and P. Zaninotto, "English Longitudinal Study of Ageing: Waves 0-9, 1998-2019". [data collection]. 33rd Edition. UK Data Service. SN: 5050, 2019, http://doi.org/10.5255/UKDA-SN-5050-20

[17] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics,* vol. 43, pp. 59–69, 1982.

[18] A. Steptoe, P. Panayotes and C. de Oliveira, "The psychological wellbeing, health and functioning of older people in England," [in] the dynamics of ageing: evidence from the English longitudinal study of ageing 2002-10 (Wave 5). In *The dynamics of ageing: evidence from the English longitudinal study of ageing 2002-10 (Wave 5)* (pp. 98–182). Institute for Fiscal Studies, 2012.

[19] E. Diener, R. A. Emmons, R. J. Larsen and S. Griffin, "The satisfaction with life scale," *Journal of Personality Assessment,* vol. 49, pp. 71–75, 1985.

[20] A. Steptoe and J. Wardle, "Positive affect measured using ecological momentary assessment and survival in older men and women," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 18244–18248, 2011.

[21] A. Steptoe, P. Demakakos, C. de Oliveira and J. Wardle, "Distinctive biological correlates of positive psychological well-being in older men and women," *Psychosomatic medicine,* vol. 74, pp. 501–8, 2012.

[22] M. L. Diener and M. B. Diener McGavran, "What makes people happy?: A developmental approach to the literature on family relationships and wellbeing," In *The science of subjective well-being* (pp. 347–375). Guilford Press, 2008.

[23] D. Carr, V. A. Freedman, J. C. Cornman and N. Schwarz, "Happy marriage, happy life? Marital quality and subjective well-being in later life," *Journal of marriage and the family,* vol. *76*, pp. 930–948, 2014.

[24] V. Lorant, D. Deliège, W. Eaton, A. Robert, P. Philippot, P. and M. Ansseau, "Socioeconomic inequalities in depression: A meta-analysis," *American Journal of Epidemiology,* vol. 157, pp. 98–112, 2003.

[25] F. Borgonovi, "Doing well by doing good. the relationship between formal volunteering and self-reported health and happiness," *Social Science & Medicine*, vol. 66, pp. 2321–2334, 2008.

[26] A. M. Creaven, A. Healy and S. Howard, "Social connectedness and depression: Is there added value in volunteering?," *Journal of Social and Personal Relationships,* vol. 35, no. 10, pp. 1400–1417, 2017.

[27] J. C. Moms, A. Heyman, R. C. Mohs, J. P. Hughes, G. van Belle, G. Fillenbaum, E. D. Mellits and C. Clark, "The consortium to establish a registry for alzheimer's disease (CERAD). Part I. clinical and neuropsychological assessment of Alzheimer's disease," *Neurology,* vol. *39*, pp. 1159–1159, 1989.

[28] T. Kohonen, M. Schroeder and T. S Huang, T., *Self-Organizing Maps*. Springer-Verlag Berlin Heidelberg, 2001.

[29] M. S. Kamal, M. G. Sarowar, N. Dey, A. S. Ashour, S. H. Ripon, B. K. Panigrahi and J. M. R. S. Tavares, "Self-organizing mapping based swarm intelligence for secondary and tertiary proteins classification," *International Journal of Machine Learning and Cybernetics,* vol. 10, pp 229–252, 2019.

[30] M. Ali, M.G. Sarowar, M. Rahman, M. L. Rahman, J. Chaki, N. Dey and J. M. R. S. Tavares, "Adam Deep Learning with SOM for Human Sentiment Classification," *International Journal of Ambient Computing and Intelligence,* vol. 10, pp. 92–116, 2019.

[31] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika,* vol. *32*, pp. 241–254, 1967.

[32] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika,* vol. *73*, pp. 13–22, 1986.

[33] J. D. Kalbfleisch and J. F. Lawless, "The analysis of panel data under a Markov assumption," *Journal of the American Statistical Association,* vol. 80, pp. 863–871, 1985.

[34] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B,* vol. 39, pp. 1–38, 1977.

[35] J. G. Dias and J. K. Vermunt, "Latent class modeling of website users' search patterns: Implications for online market segmentation," *Journal of Retailing and Consumer Services*, vol. 14, pp. 359–368, 2007.

[36] J. K. Vermunt, "Longitudinal research using mixture models," In *Longitudinal Research with Latent Variables* (pp. 119–152). Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[37] T. Snijders *Network Analysis, Longitudinal Methods of* Analysis. Springer Verlag New York, 2013.

[38] S. de Haan-Rietdijk, P. Kuppens, C. S. Bergeman, L. B. Sheeber, N. B. Allen and E. L Hamaker, "On the use of mixed Markov models for intensive longitudinal data," *Multivariate Behavioral Research,* vol. 52, pp. 747–767, 2017.
[39] J. Helske and S. Helske, *Mixture hidden Markov models for sequence data: the seqHMM package in R.* R package version 1.0.8-1.
[40] N. Villa-Vialaneix, J. Mariette, M. Olteanu, F., Rossi, L. Bendhaiba and J. Boelaert (2018). *SOMbrero: SOM Bound to Realize Euclidean and Relational Outputs*. R package version 1.2-3, 2018.
[41] S. Sum, R. M. Mathews, I. Hughes and A. Campbell, "Internet use and loneliness in older adults," *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, vol. 11, pp. 208–11, 2008.
[42] C. Morrison and H. Gore, "The relationship between excessive internet use and depression: A questionnaire-based study of 1,319 young people and adults," *Psychopathology,* vol. 43, no. 2, pp. 121–126, 2010.
[43] B. van Landeghem, "A test for the convexity of human well-being over the life cycle: Longitudinal evidence from a 20-year panel," Journal of Economic Behavior & Organization, vol. 81, pp. 571–582, 2012.
[44] M. Hofer, E. Hargittai, M. Büchi and A. Seifert, "Older adults' online information seeking and subjective well-being: The moderating role of internet skills," *International Journal of Communication* vol. 13, pp. 4426–4443, 2019.

Alejandro Cervantes

Alejandro Cervantes graduated as Telecommunications Engineer at Universidad Politecnica of Madrid (Spain), in 1993. He received his PhD in Computer Science at Carlos III of Madrid in 2007. He is currently an assistant professor at the Computer Science Department at this same University. His current interests focus on algorithms for classification of non-stationary data, large multi-objective optimization problems, swarm intelligence algorithms for data mining and AI systems for steganalysis.

David Quintana

Holds Bachelor degrees in Psychology and Computer Science. He has an M.S. in Intelligent Systems from Universidad Carlos III de Madrid and a Ph.D. in Finance from Universidad Pontificia Comillas (ICADE). He is currently an Associate Professor with the Department of Computer Science at Universidad Carlos III de Madrid. There, he is part the bio-inspired algorithms group EVANNAI. His current research interests are mainly focused on applications of computational intelligence in finance and education.

Yago Sáez

Received the degree in computer engineering in 1999. He got his Ph.D. in Computer Science from the Universidad Politécnica de Madrid, Spain, in 2005. Since 2007 till 2015 he was vice-head of the Computer Science Department from the Carlos III University of Madrid, where he got a tenure, and is nowadays associate professor. He belongs to the Evolutionary Computation, Neural Networks and Artificial Intelligence research group (EVANNAI).

Pedro Isasi

Graduate and Doctor in Computer science by the Polytechnic University of Madrid since 1994. Currently he is University professor and head of the Evolutionary Computation and Neural Networks Laboratory in the Carlos III of Madrid University. Dr. Isasi has been Chair of the Computational Finance and Economics Technical Committee (CFETC) of the IEEE Computational Intelligence Society (CIS), Head of the Computer Science Department and Vice-chancellor in the Carlos III University among others. His research is centered in the field of the artificial intelligence, focusing on problems of Classification, Optimization and Machine Learning, fundamentally in Evolutionary Systems, Metaheuristics and artificial neural networks.