COMPUTATION E INFORMATICA

Recibido 16 Oct. 2023 *ReCIBE, Año 12 No. 2, Nov. 2023*

Aceptado 31 Oct. 2023

Using Interactivity Metrics for interpretation of posts from an Institutional fan page on Facebook

Uso de métricas de interactividad para la interpretación de publicaciones de una fan page institucional en Facebook

Paola E. Velazquez-Solis [1]
Paola.velazquez@uabc.edu.mx

Jorge E. Ibarra-Esquer [2]
jorge.ibarra@uabc.edu.mx

M. Angélica Astorga-Vargas [2]
angelicaastorga@uabc.edu.mx

Brenda L. Flores-Rios [1]
brenda.flores@uabc.edu.mx

Mónica Carrillo Beltrán [1]
monica@uabc.edu.mx

Jesus Caro-Gutierrez [1]
Jesus.caro@uabc.edu.mx

Aguilar Vera Raúl Antonio [3]
avera@correo.uady.mx

[1]Instituto de Ingeniería, Universidad Autónoma de Baja California, Mexicali, Baja California. México.

[2]Facultad de Ingeniería, Universidad Autónoma de Baja California, Mexicali, Baja California. México.

[3] Facultad de Matemáticas, Universidad Autónoma de Yucatán, Mérida, Yucatán, México.

Abstract— Nowadays digital landscape, enhancing user engagement on social media platforms is crucial for elevating the visibility of universities within their respective communities. Through online presence, universities can transform their research activities and outcomes into valuable assets that extend well beyond the academia. As a result, the relationship with their community strengthens, opening new opportunities for collaboration. Interaction data from Social networks allows to identify the elements of User Engagement and, with this data, is possible to comprehend user behavior in order to create an online university community. A quantitative analysis of user engagement, viewed through university's commitment to its community, was conducted. The research findings offer a strategic tool to enhance the dissemination of scientific content geared toward user engagement. This study was carried out using data from a university's Facebook fan page, which served as a platform for sharing scientific content, news, and event updates. The research followed a methodology for analyzing social network data, which enabled the identification of key elements contributing to user engagement. The investigation revealed that variations in engagement for individual posts could be explained by a regression model, using the most correlated variables extracted from the fan page's interaction report. Further exploration, employing data mining techniques, underscored that self-generated content played the most pivotal role in driving user engagement.

Keywords— Social media, Interactivity, User Engagement, Facebook fan page, Spearman's Correlation Coefficient, Clustering

Resumen— Hoy en día, en el panorama digital el mejorar la participación de los usuarios en las plataformas de redes sociales es crucial para elevar la visibilidad de las universidades dentro de sus respectivas comunidades. A través de la presencia en línea, las universidades pueden transformar sus actividades y resultados de investigación en activos valiosos que se extienden mucho más allá del ámbito académico. Como resultado, la relación con su comunidad se fortalece, abriendo nuevas oportunidades de colaboración. Los datos de interacción de las redes sociales permiten identificar los elementos de User Engagement y, con estos datos, es posible comprender el comportamiento de los usuarios para crear una comunidad universitaria en línea. Se llevó a cabo un análisis cuantitativo de la participación de los usuarios, visto a través del compromiso de la universidad con su comunidad. Los resultados de la investigación ofrecen una herramienta estratégica para mejorar la difusión de contenido científico orientado a la participación de los usuarios. Este estudio se llevó a cabo utilizando datos de la página de fans de Facebook de una universidad, que sirvió como plataforma para compartir contenido científico, noticias y actualizaciones de eventos. La investigación siguió una metodología para analizar datos de redes sociales, que permitió identificar elementos clave que contribuyen a la participación de los usuarios. La investigación reveló que las variaciones en la participación en publicaciones individuales podrían explicarse mediante un modelo de regresión, utilizando las variables más correlacionadas extraídas del informe de interacción de la página de fans. Una mayor exploración, empleando técnicas de minería de datos, subrayó que el contenido autogenerado desempeñaba el papel más fundamental a la hora de impulsar la participación de los usuarios.

Palabras clave— Social media, interactividad, participación del usuario, página de fans de Facebook, coeficiente de correlación de Spearman, agrupamiento.

1. Introduction

Social networks are used by over 2.95 billion people (Liyanage & Premarathne, 2021), and have changed the way we create content, share information, communicate, and relate to each other (Froment et al., 2022). In the educational context, the appearance of Social networks makes the teaching-learning processes have a more social and customizable character in students, incorporating Social networks into academic activities (Shang et al., 2011). ResearchGate is a specialized academic Social network aimed at scientists and researchers to share publications, seek collaborators, communicate work in progress, and build a scholarly reputation (Sababi et al., 2017). It is a site that closely and realistically reflects the level of research activity of institutions, forming virtual Social networks centered on academic institutions (Yan & Zhang, 2018). But, the dynamics of scientific communication have changed, accelerating the knowledge feedback cycle (Campos-Freire & Rúas-Araújo, 2016). Facebook as an example of a generalist Social network, is used by most higher education institutional taking advantage of its benefits for the institutional dissemination of knowledge. In trying to increase the reach of their posts, higher educational institutions frequently vary the content of their posts, hoping to increase the engagement of their followers or users. In order to determine if there is an ideal content mix, it is important first to understand the current landscape (Peruta & Shields, 2017).

From a technological standpoint, interactivity is a representation of user preferences and the way they perceive information, feeding algorithms with data for future recommendations. It is a measure of how many users take part in the modification of the format and content of an environment in real time. Interpreting interactivity as a measurement strategy provides the potential to take advantage of data created from monitoring and analyzing posts on Social networks, benefiting from the expert knowledge (Steinmetz et al., 2020).

User activities and conversation metrics like reactions, which are described as the total number of "like", "love," "haha," "wow," "sad," and "angry" in the post; the number of times it is shared, and the total number of comments in a post are the most common benchmarks on Facebook, that describe interactivity and form the basis for its analysis and discussion (Furtado Avanza & Moreira Pinheiro, 2018; Sutanto, 2016). However, there are other metrics such as the number of people who viewed the post at least once. The number of people who viewed the content in News Feed, no need to interact with the post. And the number of clicks on posts that led to destinations or experiences on or off Facebook. From these metrics, it is assumed that a high level of interactivity generates various types of attitudes about the contents of fan pages or groups which exist on Facebook. These attitudes can be positive or negative (Kaye, 2021). All these are limited by the comparison between the results of interactivity in different posts of the same nature (Ballesteros Herencia, 2018; Sutanto, 2016).

Analyzing structured and quantifiable data like the user ratings, number of clicks and links, and network connections, among others (Chen et al., 2013; Savelev et al., 2021) provides a numerical value to the data. User Engagement on Social media platforms is one of the most discussed metric in the scientific community (Molina et al., 2021) whose level can be estimated by analyzing interactivity (Kim & Yang, 2017; Shahbaznezhad et al., 2021) that considers the qualitative meaning of an individual’s cognitive state and the active relationship they have with the content (Furtado Avanza & Moreira Pinheiro, 2018). User Engagement is perceived as a significant factor for overseeing the use of Social networks, monitoring the acceptance of their contents, and understanding the level of interaction and interest that each publication generates (Furtado Avanza & Moreira Pinheiro, 2018).

Different approaches have recently been proposed for User Engagement calculation. This calculation requires formulas in which the total number of likes, comments and shares are added. In Facebook, the User Engagement is calculated with total of all interactions in the post (Figure 1). Due to the nature of the Facebook data, qualitative research methods require additional adaptation to capture visual, virtual, and textual interactions in Social networks accurately (Chen et al., 2013).

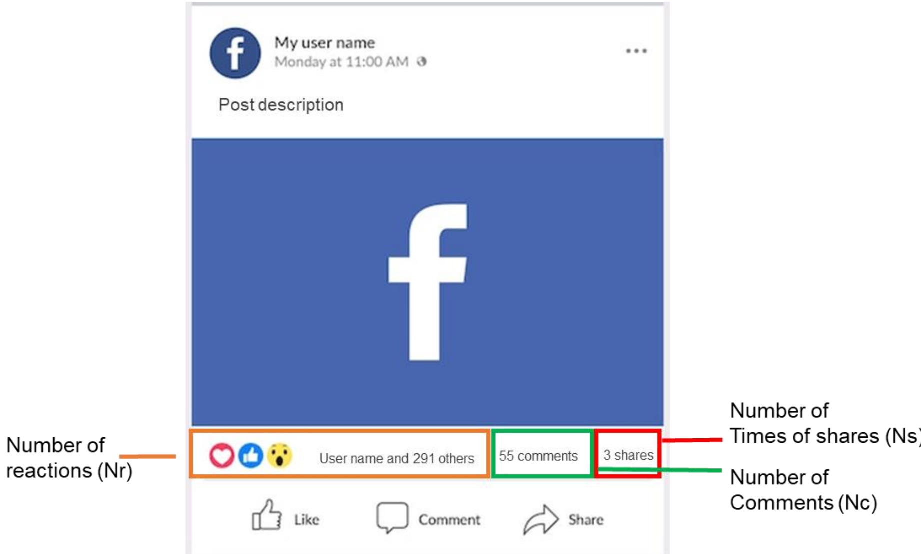

Figure 1. Formula for User Engagement in Facebook.

The scientific community considers the analysis of social network data as a multidisciplinary research area that enables the creation, extension, and adaptation of the data analysis models (Savelev et al., 2021). These models empirically provide quantitative and qualitative results using statistical methods based on evidence and objective data, which allow achieving improvements (Velazquez-Solis et al., 2022), performing a search exploration the information on Facebook for differentiating behavioral profiles of students and teachers (Karademir Coşkun et al., 2020). Various methods and techniques are implemented in the analysis of the Social networks data, like variable association (Eriksson et al., 2019; Phuntusil & Limpiyakorn, 2017), regression models (Purba et al., 2020), predictive models with process mining (Li & De Carvalho, 2019), and data mining techniques such as naive Bayes (Aviles & Esquivel, 2019), principal component analysis (Bharati et al., 2015), and K-means (Liyanage & Premarathne, 2021). Also, the unstructured textual information requires an analysis to understand the emotions, feelings, thoughts, opinions, and life experiences.

The objectives of this research are: 1) Use Interactivity metrics for interpretation how different types of content affect engagement in fan pages on Facebook. An Institutional fan page used for dissemination of scientific content, news, and events, created and maintained at a higher education institution, serves as the case of study 2) Create a regression model that predicts and explains variations in engagement for posts on the fan page based on the interaction reports from a two-year period.

The paper is organized as follows. In section II, related work about the qualitative analysis in Social networks and Facebook interactivity are described. Then, in section III we present the methodology; later in section IV data analysis and findings of the results are explained. Moreover, in section V conclusions and future work are presented.

2. Related work

2.1 Quantitative analysis on Social networks

Purba et al., (2020), a new metric for people who are not among the followers but liked a post from a user was proponed. For that, a regression model was developed using metadata, sentiment, hashtag to predict the outsider's percentage. On another hand, in the study of Eriksson et al., 2019, they take an explorative consumer engagement approach to analyze three interior design brands' postings on Facebook and Instagram. In order to, investigate differences in consumer brand post engagement regarding these postings.

The investigate of Aviles & Esquivel (2019) was centered on mining social networks data specifically on users' post from Facebook fan pages among selected Universities. The main objective of this study was mining social media data to identify their stakeholders' social networks behaviors and make knowledge-driven decisions using Naïve Bayes Classifier. The authors Liyanage & Premarathne (2021) introduce an approach to detect Facebook clone profiles by using a clustering technique on its profile attributes and network connections. With the Agglomerative hierarchical clustering algorithm and Jaccard similarity measurement, they list the possible clones with their percentages.

2.2 Facebook Interactivity

Measurement of User Engagement is calculated by three different methods: surveys and interviews, implicit measures, and web analytics (Chan-Olmsted et al., 2017). In web analytics, User Engagement is obtained from the behavior shown by users through their activity on digital platforms. Therefore, a series of metrics are required to quantify interaction-related variables, such as the total number of reactions on the post (Nr), the number of user comments on the post (Nc), and the number of times users shared the post (Ns) (Bonsón & Ratkai, 2013; Hu & Chen, 2016; Jayasingh, 2019; Pilař et al., 2021).

There exist diverse formulas and models for calculating User Engagement based on data obtained from Facebook fan pages or groups. Table 1 presents the evolution of metrics used to calculate User Engagement identified over 2013 - 2020. When reviewing each work, eight of them had no identifiable rationale for variables or mathematical verification about the origin of the formula they were implementing. Having information from the variables used by different formulas and models helps to identify an integrated measure of User Engagement, underscoring the importance of the parameters normalized (Eriksson et al., 2019).

All the formulas shown in Table I are based on an addition of the different types of interactions, expanding from the formula used by Facebook (equation 1).

$$User\ Engagement\ on\ Facebook = N_r + N_c + N_s \qquad (1)$$

A difference is established in the formula proposed by (Jayasingh, 2019), where coefficients, or weights, are used for each type of interaction as a representation of the effort needed. In addition, engagement rate is calculated over a fraction of the total followers of the page, represented by the 0.8 exponent.

Year	Metrics	Proposed Formula	Authors
2013	Likes Comments Sharing Followers Messages	$\dfrac{N_L + N_C + N_S}{\dfrac{N_{Messages}}{N_F}}$	Bonsón & Ratkai Hoffmann
2014	Likes Comments Sharing Followers	$\dfrac{N_L + N_C + N_S}{N_F}$	Niciporuc
	Likes Comments Sharing Followers Scope	$\dfrac{N_L + N_C + N_S}{\dfrac{N_F}{\mu_{Scope}}}$	Oviedo-García et al.
2015	Reactions Comments Sharing Likes on fan page	$\dfrac{N_R + N_C + N_S}{N_{Likes\ fan\ page}}$	Vadivu & Neelamalar
2017	Reactions Comments Sharing Followers	$\dfrac{N_R + N_C + N_S}{N_F}$	Herrera-Torres et al.
	Reactions Comments Sharing Likes on fan page	$\dfrac{N_R + N_C + N_S}{N_{Likes\ fan\ page}}$	Peruta & Shields
	Reactions Comments Sharing Likes on fan page	$\dfrac{N_R + N_C + N_S}{N_{Likes\ fanpage}^{0.8}}$	Phuntusil & Limpiyakorn
2018	Reactions Comments Sharing Scope	$\dfrac{N_R+N_C+N_S}{Scope\ post}$	Ballesteros-Herencia
2019	Reactions Comments Sharing	$N_R + 5N_C + 10N_S$	Ge & Grezel Eriksson et al.
	Reactions Comments Sharing Likes on fan page	$\dfrac{N_R + 5N_C + 10N_S}{N_{Likes\ fanpage}^{0.8}}$	Jayasingh Sudarsan
2020	Reactions Comments Sharing Posts Followers	$\dfrac{N_R + N_C + N_S}{N_{Post} \times N_{Followers}}$	Martínez-Sala & Segarra-Saavedra

Table 1. Formulas for User Engagement.

2.3 Qualitative analysis on Social networks

There are two directions in which qualitative analysis is recommended. The first, commonly known as qualitative research, is focused on the fields of sociology, medicine, psychology, and linguistics.

The second direction, called intellectual, is used in the field of Artificial Intelligence for qualitative analysis and is characterized by reasoning, simulation, and data analysis (Tsvetkov, 2021). A qualitative approach to social network data analysis supports the creation of meaning by placing technology in specific social contexts, places, and times. On the other hand, there is a need to implement inductive methods that allow us to delve deeper into the culture of digital spaces, considering users' perspectives as complex and heterogeneous (Voorveld et al., 2018).

A review of related work identified the application of various types of methods, which established data empirically using accepted statistical methods (He et al., 2017; Karademir Coşkun et al., 2020; Phuntusil & Limpiyakorn, 2017), such as nonparametric tests (Bharati et al., 2015; Molina et al., 2021), analysis of variance (ANOVA) and coefficient correlation analysis for metrics analysis (Gil-quintana et al., 2022; Velazquez-Solis et al., 2022), statistical analysis models (Voorveld et al., 2018), regression models for interpretate and predicate metrics (Wheatley & Vatnoey, 2020), association models, predictive or data mining models like Naive Bayes (He et al., 2017), Principal Component Analysis (Gil-quintana et al., 2022), and K-means (Liyanage & Premarathne, 2021). The Pearson's correlation coefficients, ANOVA test, chi-square test, and linear discriminant were used for a dataset analysis (Agusriandi et al., 2020).

The models employed are based on the extraction of knowledge from the data sets as by the ANOVA and text mining, to analyze the data later with the goal of understanding virtual social behavior, and also the feelings and emotions of interacting persons (Tsvetkov, 2021; Wheatley & Vatnoey, 2020). There are methodological proposals that present applications of algorithms for content analysis of text data (Chan-Olmsted et al., 2017), PageRank algorithm and HITS algorithm, Facebook API (Tsinovoi, 2020), JavaScript and REST API (Dabbagh & Kitsantas, 2012). In some of them, data mining and cleansing activities in the process of social network data analysis are similar (Eriksson et al., 2019; Phuntusil & Limpiyakorn, 2017; Pilař et al., 2021; Stieglitz et al., 2018; Surya Gunawan et al., 2020).

In the work of (Franz et al., 2019), qualitative research was conducted on Facebook users and their activity on the social network. They sought options to address health issues through the analysis of social network generated texts, concluding that qualitative and quantitative analysis on Facebook involves participation of the researcher in the study of observed patterns of information to identify relational themes, grammatical elements and the valence of sentiments contained in Facebook posts and associated comments. In trying to increase the reach of posts, organizations often vary the content, hoping to increase the User Engagement of their followers. The identified methods only rely on users' Social networks or their behavior, while the role of content contributed by fan page owners on Social networks is ignored (Al-Dheleai & Tasir, 2017).

3. Methodology

An adaptation of the methodologies for statistical and qualitative analysis of social network data proposed by (Agusriandi et al., 2020; Peruta & Shields, 2017; Pilař et al., 2021; Stieglitz et al., 2018; Surya Gunawan et al., 2020) was performed establishing four phases: 1) Sample selection and data collection, 2) Data preprocessing, 3) Application of methods and techniques, and 4) Visualization and Analysis of results. Fig 2 displays the adapted methodology. Each of the phases is explained in further detail in the following sections.

A research question was posed as a guide for the study: What is the content type that helps to increase interactivity of users interest in scientific dissemination and scientific work?

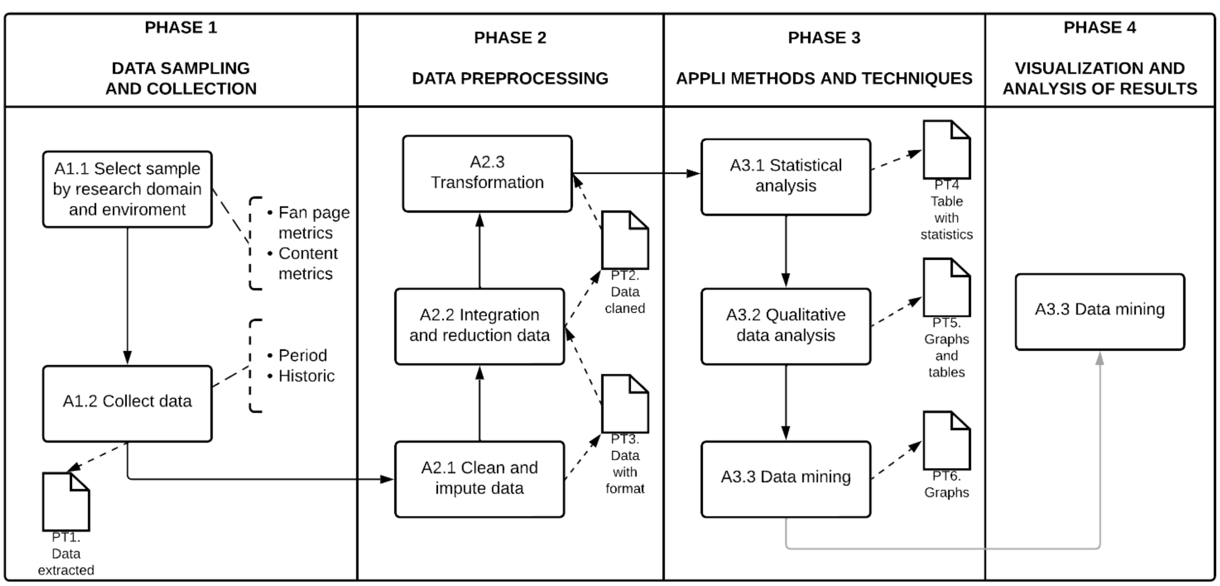

Fig 2. Methodology adapted from Peruta & Shields (2017); Stieglitz et al. (2018); Agusriandi et al. (2020); Surya Gunawan et al. (2020); Pilař et al. (2021).

3.1 Phase 1 - data sampling and data collection

The case study approach was adopted for Mujeres en la Investigación - UABC Facebook's fan page (www.facebook.com/mujeresinvestigacion) managed by the community-university program of a Higher Education Institution. Since 2013, this fan page shares institutional information and results of scientific research activities carried out with the participation of 11 female researchers from the areas of Engineering Physics, Chemistry and Environment. To date, the fan page has a total of 1,570 followers and its publications have accumulated 1,393 likes.

The strategy defined for a two years period was that the selection of the extracted data would be a series of scientific demonstration and dissemination events in an online format, organized and published through this fan page. The data were generated between January 1, 2021 and December 31, 2022, corresponding to 345 posts. Taking into account that the number of likes, comments, and shares continuously changes in the publications, an update of the data sample was performed as of May 30 to have the most recent number of reactions, comments, clicks, and shares in the same 345 publications. Authors Agusriandi et al., (2020) and Niciporuc (2014) recommend allowing a period for users to reflect on the content after it has been posted. It was decided that one month after the last publication should be sufficient, since most reactions are often received within the first 48 hours. The result of this phase generated an XLSX or CSV document.

3.2 Phase 2 - data preprocessing

The calculation of User Engagement (E) of each publication was performed using the formula proposed by Jayasingh (Bonsón & Ratkai, 2013; Oviedo-García et al., 2014).

$$E = \frac{N_R + 5N_C + 10N_S}{N_{Likes\,fanpage}^{0.8}} \qquad (2)$$

User Engagement results were placed as a new variable to analyze the User Engagement index of the followers towards the fan page. Subsequently, the processes of cleaning and imputation of missing data were performed. When applying statistical techniques, a preliminary analysis was performed to verify the normality distributions and homogeneity of variance in the sample with a Kolmogorov-Smirnov test. SPSS and RStudio tools were used to support the results processing and graphic visualization.

3.3 Phase 3 - apply methods and techniques

The main analysis included descriptive statistics of publications in general and type of content, and a nonparametric chi-square analysis. A chi-square analysis was chosen because there is a set of variables, where both the dependent and independent variables are categorical. An analysis of the variables was carried out by means of a Spearman's Rank correlation coefficient, with the intention of creating a proposed multiple regression model to explain the dependent variable (User Engagement). Section IV presents the results of phase 4.

4. Results and discussion

Sample initial exploration included characterizing the set of followers of the fan page. Although the content generated on that page is from women Figure 3 presents the 1,532 followers, 90% belong to locations in Mexico, where 83% correspond to women in the age range between 25 - 50 years and 17% of male followers distributed between 25 - 45 years mostly. The followers distribution is understandable due to the fan page objective which is to position the diffusion of scientific content created by women researchers.

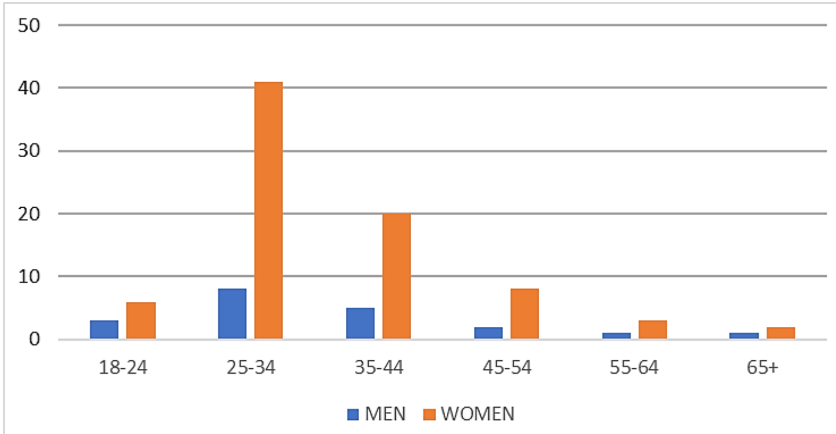

Fig 3. Age distribution of fan page followers.

Over the 24 months period from January 2021 to December 2022, the sample recorded a total of 345 posts. Table 2 shows the distribution of three quantitative metrics selected and extracted from the dataset: number of comments, number of reactions, and number of times a post was shared, separated by months of each year. Focusing on all-User Engagement metrics, there were an increase. This means that reactions increased by 951.5%, comments by 1528.9% and shares by 425.8%. The 3rd month (March 2022) shown the greater interactivity. This was due to the realization of an event in live streaming to dissemination science with 7289 interactions in posts, unlike March 2021 with 93 interactions. Among the most interesting findings is that in 2021, 7th month was the one that had the most interactions with a single publication made. Month 4 in both years maintained similar interactivity, since there are activities carried out by the fan page every year. In most months increase is observed, but there are particular cases where a decrease was observed, such as months 9, 10 and 11 of the year 2022.

2021				**2022**			
Month	**Metrics**			**Month**	**Metrics**		
	Reactions	**Comments**	**Share**		**Reactions**	**Comments**	**Share**
1	29	0	11	1	771	165	59
2	61	2	17	2	472	56	57
3	79	2	12	3	5996	994	299
4	254	25	52	4	393	507	71
5	48	1	2	5	122	181	11
6	7	0	1	6	373	415	4
7	252	103	23	7	30	1	2
8	0	0	0	8	43	1	0
9	40	1	0	9	22	1	5
10	7	1	0	10	11	2	0
11	83	16	2	11	7	0	2
12	8	1	0	12	19	1	1
TOTAL	868	152	120	TOTAL	8259	2324	511

Table 2. Distribution of reactions, comments and shares extracted from the dataset.

Five different categories of posts were identified: 1) Video, 2) Photo, 3) Link, 4) Live video, and 5) Text. Table 3 describes the distribution of posts according to the categories created. Looking more in depth at the different categories, significant findings emerge: In 2021, the month with the highest number of interactions (Table 2) was July, but it only had one Photo-type publication. In the case of the year 2022, the concentration of more interactions is in March, where about 85% of the posts were elements of the photo category. In Fig 4 it was identified that the number of Text-type posts decreased considerably. Video-type posts also decreased, but links to other external elements increased. Posts with photos remained constant. It is separated per month with the purpose of visualizing the behavior of the sets of categories in an ordered way. Table III compares the results of both years separated by month. The difference between the values can be observed, standing out the first six months of the year 2022. It was calculated with both formulas in the two periods and they were compared between them. In general, if there is an increase from one year to the next, regardless of the formula. In month eleven there is a decrease only in engagement on Facebook. As there are more interactions, more User Engagement is generated. It can be seen that in 2021 the

Engagement with the formula, as calculated by Facebook, is higher than that obtained with the 2022 posts. In month 8 of 2021, the publications had no interactions, impacting User Engagement. While month 7 obtained the highest value in this period.

2021						2022					
Month	Categories					Month	Categories				
	1	2	3	4	5		1	2	3	4	5
1	6.66	7.76	17.65	4.17	0	1	0	14.73	15.79	0	0
2	13.33	24.14	29.41	29.17	16.67	2	0	11.63	15.79	25.00	0
3	6.66	12.93	23.53	8.33	16.67	3	40	48.84	15.79	12.50	25
4	46.66	17.24	11.76	29.17	0	4	0	7.75	10.53	25.00	25
5	13.33	12.07	5.88	16.67	0	5	20	5.43	5.26	37.50	25
6	6.66	1.72	5.88	0	0	6	20	3.88	5.26	0	0
7	0	0.86	0	0	0	7	0	0.78	10.53	0	0
8	0	0	0	0	16.67	8	0	1.55	0	0	0
9	0	8.62	5.88	4.17	0	9	0	1.55	5.26	0	0
10	6.66	4.31	0	4.17	0	10	0	0.78	10.53	0	25
11	0	7.76	0	4.17	33.33	11	0	1.55	0	0	0
12	0	2.59	0	0	16.67	12	20	1.55	5.26	0	0

Table 3 Percetage of post with the categories.

Month	Engagement 2021		Engagement 2022	
	Facebook	Jayasigh	Facebook	Jayasigh
1	2.8571	2.9460	45.2272	29.4831
2	2.3488	1.8079	29.25	19.6131
3	3.2608	2.4640	102.662	58.3242
4	9.1944	7.4097	64.7333	71.9644
5	2.4285	1.0314	24.1538	25.9515
6	2	1.2610	113.1429	105.4626
7	378	295.8294	11	5.4398
8	0	0	22	7.1212
9	3.4166	1.1126	9.3333	7.6158
10	1.1428	0.5086	3.25	1.5577
11	6.3	3.1155	4.5	4.0057
12	2.25	0.9643	5.25	2.5221

Table 4. Comparison of the two years between the formulas by month

Table 5 allowed us to observe the correlations between the different categories extracted from the data set. This helped to verify the relationship between content type and period of time as related variables, where the incidence is visible in the elements of type photo. Also, in Table 5 we can see that in the calculation of the User Engagement with formula proposed by Jayasigh (2019) and Facebook. The Text-type contents would be the ones that have the least influence; while Video content and Links generate a medium impact on the relationship.

	Categories				
User Engagement Formulas	**1) Video**	**2) Photo**	**3) Link**	**4) Live video**	**5) Text**
Facebook	0.41	0.78	0.42	0.47	0.16
Jayasigh	0.33	0.77	0.32	0.48	0.21

Table 5. Comparison of correlations by categories

When making a cluster with a total of posts that were made distributed by the various calculations of the implemented equations. Fig 5 distributes the posts from lowest to highest User Engagement with both formulas. By the elbow method it will be reduced that the appropriate number of groups is three. The first group (blue) corresponds to publications that are low due to interactivity and for consequence have poor User Engagement. Group two (green) has the publications with a commitment between 100 and 300. Finally, group three (red) represents the publications with a commitment greater than 300 due have more interactivity. It can be seen in the graph that there is a greater concentration of publications in group one, denoting that most of the commitment in the publications does not have the expected response. It is important to mention that group two contains the publications corresponding to a scientific dissemination event. In a more detailed analysis of the publications, Fig 6 provides that the clusters identified are the same as those obtained in the previous analysis; i.e., group 1 of the graph in Fig 5 is the same group 1 in Fig 6.

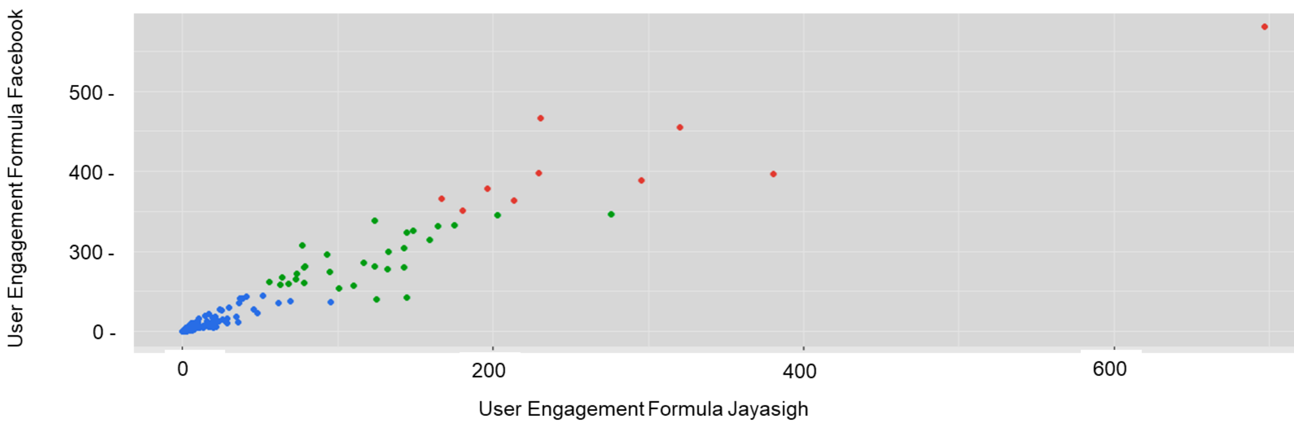

Fig 5. User Engagement with both formulas

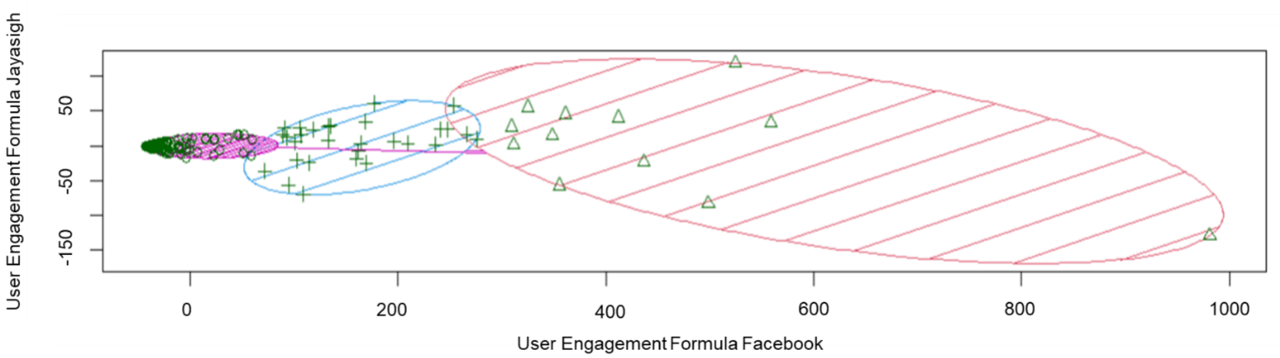

Fig 6. User Engagement clustering with both formulas

A visualization of the groups can be made over the space of the first two principal components. The first cluster explains the elements that Jayasigh (2019) formula and the one provided by Facebook have low interaction and therefore little User Engagement. In the second, we have a User Engagement of 100 to 300 from Facebook, but with the Jayasigh proposal at 50. From the data set obtained for the fan page behaviour, the variables that were used to be classified per type and number of interactions in the calculation of User Engagement were selected. Therefore, it is inferred that the greater interactions and depends content type, the value User Engagement is affected.

Applied qualitative analysis in unstructured data, relationships and groups are extracted that allowed us to take advantage not only of the descriptions obtained but also of the relationships and rules. In this way, qualitative analysis supports the identification of various types of relationships within data interactivity in the Facebook social network. A regression model analysis was performed taking into account interactions and content type as independent variables, resulting in a multiple linear regression model to predict User Engagement, and is defined as follows:

$$Regression\ model\ of\ User\ Engagement = 0.56_{Interations} - 0.29_{Photo} - 1.58_{Text} + 172.18_{Video} + 22.23_{Live} \quad (3)$$

The regression model defined (equation 3) obtained a coefficient of multiple determination (R^2) of 0.59 of influence of the independent variables (p-value of 2.2e-16). The equation shows that User Engagement is highly influenced by the video type content, no matter if it's live or not. It also indicates that User Engagement decreases for each photo or text content type tends. This is considered highly representative and allows interpreting User Engagement in future publications based on content type and number of interactions. In calculating the value of User Engagement, a fan page with consistency in ranges of User Engagement in its posts is considered to have engaged users (Ballesteros Herencia, 2018). The regression model allowed an interpretation of the impact of the categories by type of content. This is considered highly representative and allows predicting User Engagement in future publications based on content type

5. Conclusions and future work

It was possible to create scientific outreach strategies for increasing the visibility of scientific content and improve interest of the educational community, taking advantage of Facebook's versatility. In this work, it is mostly highlighted the User Engagement formula for obtaining metrics of great value because it includes the visualization of content and the interactivity that users have on the Facebook fan page. It is also relevant that there is a positive correlation between the User Engagement a publication receives and the content type, especially in scientific videos created by the case study fan page.

In the analysis and visualization of the variables of each post, clustering was also applied as a data mining technique, demonstrating that self-produced content provides a higher Engagement. And finally, a regression model was establish to User Engagement percentage of a sample of posts about scientific dissemination on a Facebook fan page. The creation of new models for the interpretation of Facebook variables resulted in a proprietary multiple linear regression model that allows to identify User Engagement based on the different categories in each post.

Consequently, it is advisable to use qualitative analysis of Facebook data, methodologies for social network data analysis, and a linear regression model as strategic tools to improve the practices of creation and dissemination of content focused on Science and Engineering topics. In order to highlight the importance of this research an adequate treatment and analysis for social network data, was determined by the type of followers and consumption of the posts created.

As future work, these results will be used to conduct more comparative studies under different periods, where better practices are applied in the posts of fan pages and groups of different undergraduate programs contents and, through these tests, the different user behaviors will be analyzed providing data that generalizes the possibilities of Facebook in Higher Education.

Acknowledgment

This research was supported by Consejo Nacional de Ciencia y Tecnología (CONACyT) under grant 668484 for the first author.

References

Agusriandi, Sukaesih Sitanggang, I., & Hartono Wijaya, S. (2020). *Student Performance Based on Activity Log on Social Network and e-Learning*. *10*(6).

Al-Dheleai, Y. M., & Tasir, Z. (2017). Using facebook for the purpose of students' interaction and its correlation with students' academic performance. *Turkish Online Journal of Educational Technology*, *16*(4), 170–178.

Aviles, J., & Esquivel, R. (2019). Mining Social Media Data of Philippine Higher Education Institutions Using Naïve Bayes Classifier Algorithm. *SSRN*. https://doi.org/Aviles, Joey and Esquivel, Rosanna, Mining Social Media Data of Philippine Higher Education Institutions Using Naive Bayes Classifier Algorithm (April 27, 2019). Available at SSRN: https://ssrn.com/abstract=3379025 or http://dx.doi.org/10.2139/ssrn.3379025

Ballesteros Herencia, C. A. (2018). El índice de engagement en redes sociales, una medición emergente en la Comunicación académica y organizacional. *RAZÓN Y PALABRA Primera Revista Electrónica En Iberoamérica Especializada En Comunicación*, *22*(3_102), 96–124.

Bharati, P., Zhang, W., & Chaudhury, A. (2015). Better knowledge with social media? Exploring the roles of social capital and organizational knowledge management. *Journal of Knowledge Management*, *19*(3), 456–475. https://doi.org/10.1108/JKM-11-2014-0467

Bonsón, E., & Ratkai, M. (2013). A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page. *Online Information Review*, *37*(5), 787–803. https://doi.org/10.1108/OIR-03-2012-0054

Campos-Freire, F., & Rúas-Araújo, R. (2016). Uso de las redes sociales digitales profesionales y científicas: el caso de las 3 universidades gallegas. *El Profesional de La Información*, *25*(3), 431. https://doi.org/10.3145/epi.2016.may.13

Chan-Olmsted, S., Wolter, L. C. , & Wang, R. (2017). Toward a multidimensional framework of media engagement: Conceptualizing consumer experience and connection with media content in a digital environment. *Emma Conf 2017. European Media Management Associafion.*, *585*.

Chen, X., Madhavan, K., & Vorvoreanu, M. (2013). A web-based tool for collaborative social media data analysis. *Proceedings - 2013 IEEE 3rd International Conference on Cloud and Green Computing, CGC 2013 and 2013 IEEE 3rd International Conference on Social Computing and Its Applications, SCA 2013*, 383–388. https://doi.org/10.1109/CGC.2013.67

Dabbagh, N., & Kitsantas, A. (2012). Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *Internet and Higher Education*, *15*(1), 3–8. https://doi.org/10.1016/j.iheduc.2011.06.002

Eriksson, N., Sjöberg, A., Rosenbröijer, C.-J., & Fagerstrøm, A. (2019). Consumer brand post engagement on Facebook and Instagram-A study of three interior design brands. *International Conference on Electronic Business (ICEB) at AIS Electronic Library (AISeL)*, 12–20. https://aisel.aisnet.org/iceb2019

Franz, D., Marsh, H. E., Chen, J. I., & Teo, A. R. (2019). Using facebook for qualitative research: A brief primer. *Journal of Medical Internet Research*, *21*(8), 1–12. https://doi.org/10.2196/13544

Froment, F., García-González, A. J., & Cabero, J. (2022). The relationship of Twitter with teacher credibility and motivation in university students. *Comunicar*, *30*(71), 1–12. https://doi.org/10.3916/C71-2022-10

Furtado Avanza, M., & Moreira Pinheiro, W. (2018). Audience Measurement in Press Through Social Interactivity on Facebook in Brazil. *Mediaciones*, *14*(20), 48–69. https://doi.org/10.26620/uniminuto.mediaciones.14.20.2018.48-69

Gil-quintana, J., de León, E. V., Osuna-acedo, S., & Marta-lazo, C. (2022). Nano-Influencers Edutubers: Perspective of Centennial Generation Families in Spain. *Media and Communication*, *10*(1), 247–258. https://doi.org/10.17645/mac.v10i1.4760

He, Y., Wang, C., & Jiang, C. (2017). Mining Coherent Topics With Pre-Learned Interest Knowledge in Twitter. *IEEE Access*, *5*, 10515–10525. https://doi.org/10.1109/ACCESS.2017.2696558

Hu, Y. H., & Chen, K. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management*, *36*(6), 929–944. https://doi.org/10.1016/j.ijinfomgt.2016.06.003

Jayasingh, S. (2019). Consumer brand engagement in social networking sites and its effect on brand loyalty. *Cogent Business and Management*, *6*(1). https://doi.org/10.1080/23311975.2019.1698793

Karademir Coşkun, T., Erdoğdu, F., & Kokoç, M. (2020). Students' Profiles of Information Seeking Behaviors on Facebook: Relations to Personality, Time Management, Gender, and Facebook Use. *International Journal of Technology in Education and Science (IJTES)*, *4*(2), 144–159. www.ijtes.net

Kaye, L. K. (2021). Exploring the "socialness" of social media. *Computers in Human Behavior Reports*, *3*, 100083. https://doi.org/10.1016/j.chbr.2021.100083

Kim, C., & Yang, S. U. (2017). Like, comment, and share on Facebook: How each behavior differs from the other. *Public Relations Review*, *43*(2), 441–449. https://doi.org/10.1016/j.pubrev.2017.02.006

Li, G., & De Carvalho, R. M. (2019). Process Mining in Social Media: Applying Object-Centric Behavioral Constraint Models. *IEEE Access*, *7*, 84360–84373. https://doi.org/10.1109/ACCESS.2019.2925105

Liyanage, C. R., & Premarathne, S. C. (2021). Clustered Approach for Clone Detection in Social Media. *International Journal on Advanced Science, Engineering and Information Technology*, *11*(1), 99–104. https://doi.org/10.18517/ijaseit.11.1.9272

Molina, O. E., Reynaldo, D., & Cancell, F. (2021). Engagement and desertion in MOOCs: Systematic review Engagement and desertion in MOOCs: Systematic review El engagement y la deserción en los MOOCs: Revisión sistemática. *Comunicar*, *October*, 111–124.

Oviedo-García, M. Á., Muñoz-Expósito, M., Castellanos-Verdugo, M., & Sancho-Mejías, M. (2014). Metric proposal for customer engagement in Facebook. *Journal of Research in Interactive Marketing*, *8*(4), 327–344. https://doi.org/10.1108/JRIM-05-2014-0028

Peruta, A., & Shields, A. B. (2017). Social media in higher education: understanding how colleges and universities use Facebook. *Journal of Marketing for Higher Education*, *27*(1), 131–143. https://doi.org/10.1080/08841241.2016.1212451

Phuntusil, N., & Limpiyakorn, Y. (2017). Predicting engaging content for increasing organic reach on facebook. *Lecture Notes in Electrical Engineering*, *424*, 637–644. https://doi.org/10.1007/978-981-10-4154-9_73

Pilař, L., Kvasničková Stanislavská, L., Kvasnička, R., Bouda, P., & Pitrová, J. (2021). Framework for social media analysis based on hashtag research. *Applied Sciences (Switzerland)*, *11*(8). https://doi.org/10.3390/app11083697

Purba, K. R., Asirvatham, D., & Murugesan, R. K. (2020). An analysis and prediction model of outsiders percentage as a new popularity metric on Instagram. *ICT Express*, *6*(3), 243–248. https://doi.org/10.1016/j.icte.2020.07.001

Sababi, M., Marashi, S. A., Pourmajidian, M., Pourtabatabaei, S. S., Darki, F., & Sadrzadeh, M. R. (2017). View of How accessibility influences citation counts_ The case of citations to the full text articles available from ResearchGate.pdf. *Ournal on Research Policy & Evaluation*, *1*.

Savelev, A. O., Karpova, A. Y., Chaykovskiy, D. V., Vilnin, A. D., Kaida, A. Y., Kuznetsov, S. A., Igumnov, L. O., & Maksimova, N. G. (2021). The high-level overview of social media content search engine. *IOP Conference Series: Materials Science and Engineering*, *1019*(1). https://doi.org/10.1088/1757-899X/1019/1/012097

Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The Role of Social Media Content Format and Platform in Users' Engagement Behavior. *Journal of Interactive Marketing*, *53*, 47–65. https://doi.org/10.1016/j.intmar.2020.05.001

Shang, S. S. C., Li, E. Y., Wu, Y. L., & Hou, O. C. L. (2011). Understanding Web 2.0 service models: A knowledge-creating perspective. *Information and Management*, *48*(4–5), 178–184. https://doi.org/10.1016/j.im.2011.01.005

Steinmetz, C., Rahmat, H., Marshall, N., Bishop, K., Thompson, S., Park, M., Corkery, L., & Tietz, C. (2020). Liking, Tweeting and Posting: An Analysis of Community Engagement through Social Media Platforms. *Urban Policy and Research*, 1–21. https://doi.org/10.1080/08111146.2020.1792283

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*(October 2017), 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Surya Gunawan, T., Aleah Jehan Abdullah, N., Kartiwi, M., & Ihsanto, E. (2020). Social Network Analysis using Python Data Mining. *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*. https://doi.org/10.1109/CITSM50537.2020.9268866

Sutanto, S. M. (2016). *The Use of Facebook Fanpage as Promotional Media for the Novel Book fans by using the Model of SMCR Berlo*.

Tsinovoi, A. (2020). The management of visibility in digital diplomacy: Infrastructures and techniques. *First Monday*, *May*. https://doi.org/10.5210/fm.v25i5.10116

Tsvetkov, V. Y. (2021). Qualitative analysis relations. *IOP Conference Series: Materials Science and Engineering*, *1047*(1). https://doi.org/10.1088/1757-899X/1047/1/012127

Velazquez-Solis, P. E., Flores-Rios, B. L., Astorga-Vargas, M. A., Ibarra-Esquer, J. E., González-Navarro, F. F., & Aguilar Vera, R. A. (2022). Analysis of scientific dissemination posts on Facebook from a social media approach. In In press (Ed.), *CISTI'2022 - 17ª Conferencia Ibérica de Sistemas y Tecnologías de Información*.

Voorveld, H. A. M., van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with Social Media and Social Media Advertising: The Differentiating Role of Platform Type. *Journal of Advertising*, *47*(1), 38–54. https://doi.org/10.1080/00913367.2017.1405754

Wheatley, D., & Vatnoey, E. (2020). 'It's Twitter, a bear pit, not a debating society': A qualitative analysis of contrasting attitudes towards social media blocklists. *New Media and Society*, *22*(1), 5–25. https://doi.org/10.1177/1461444819858278

Yan, W., & Zhang, Y. (2018). Research universities on the ResearchGate social networking site: An examination of institutional differences, research activity level, and social networks formed. *Journal of Informetrics*, *12*(1), 385–400. https://doi.org/10.1016/j.joi.2017.08.002

Authors

P. Velazquez-Solis She is a student of Ph.D. at the Universidad Autónoma de Baja California in an Engineering Institute. Her research interests include social media, data mining, data Analysis and she is also consulting in private and public sector companies and dedicated to teaching, giving classes in local universities.

J. E. Ibarra-Esquer He is a professor at the Universidad Autónoma de Baja California in the Faculty of Engineering. His research interests include the Internet of Things and Data Mining and has taken part in research projects related to improving the learning processes for students in the Computer Engineering major.

M. Astorga-Vargas She is a professor of software process improvement at the Universidad Autónoma de Baja California/ Faculty of Engineering. Her research interests are Software Engineering with a focus in Software Process Improvement in small companies, and the effectiveness in software development teams.

B. L. Flores-Rios She is a researcher of knowledge management and software process improvement at the Universidad Autónoma de Baja California/Engineering Institute. Her research interests are Software Process Improvement and quality in small enterprises and the impact of Knowledge Engineering on Software Engineering.

M. Carrillo-Beltrán She is a researcher of Food Biotechnology at the Universidad Autónoma de Baja California/ Engineering Institute. Her research interests include innovation ecosystem where emerging technologies are applied for the development, sensory evaluation and marketing of healthier and safer foods and beverages with real benefits and recognizable added values, focusing on human nutrition, sustainability and planetary health.

Jesús Caro Gutiérrez He is a researcher attached to the UABC Engineering Institute. He has been recognized as a Candidate for the National System of Researchers and Recognition of the Performance of Academic Staff (PREDEPA-UABC). He is a member of the Mexican Society of Computer Science and his main lines of work are computer vision and data science, applied to the characterization of nanomaterials.

Raúl Antonio Aguilar Vera obtained the degree of Doctor from the Polytechnic University of Madrid, Spain and the Master's degree in Software Engineering from the same Institution, he also has the degree of Master in Higher Education from the Autonomous University of Yucatan. He is currently a professor at the Faculty of Mathematics of the Universidad Autónoma de Yucatán. He has been a member since 2014 of the CONAHCYT National Researchers System (Level 1). His research work includes the areas of Software Engineering and Educational Computing.