Empowering users regarding the sensitivity of their data in social networks through nudge mechanisms

J. Alemany
Universitat Politècnica de València
jalemany1@dsic.upv.es

E. del Val
Universidad de Zaragoza
edelval@unizar.es

A. García-Fornes
Universitat Politècnica de València
agarcia@dsic.upv.es

Abstract

The use of online social networks (OSNs) is a continuous trade-off between relinquishing some privacy in exchange for getting some social benefits like maintaining (or creating new) relationships, getting support, influencing others' opinions, etc. OSN users are faced with this decision each time they share information. The amount of information or its sensitivity is directly related to the amount of users' loss of privacy. Currently, there are several approaches for assessing the sensitivity of the information based on the willingness of users to provide them, the monetary benefits derived from extracting knowledge of them, the amount of information they provide, etc. In this work, we focus on quantifying data sensitivity as the combination of all of the approaches and adapting them to the OSN domain. Furthermore, we propose a way of scoring publication sensitivity as the accumulative value of the sensitivity of the information types included in it. Finally, an experiment with 196 teenagers was carried out to assess the effectiveness of empowering users regarding the sensitivity of the publication. The results show a significant effect on users' privacy behavior by the nudge message and the sensitivity included in it.

1. Introduction

Online social networks have become a popular tool, being one of the main Internet activities among users [1]. OSN users[1] interact and socialize with each other by sharing their opinions and comments, supporting their friends and favorite groups, and posting their information, activities, etc. As a result, a huge amount of traffic of personal data is produced daily. The way to control the access and use of the data is via privacy policies. However, privacy is a very complex concept for users due to its diffuse nature and the number of factors that must be taken into account [2] (e.g., the sensitivity of the message, the properties/context of the receiver, the scope of the action, etc.). Even so, users are constantly confronted with the privacy decision-making process for each piece of data they share, which may produce privacy issues. These occur not only due to the users' lack of knowledge about privacy or the service provider's data usage but also because other users may have access to user data [3]. As a consequence, some users may be exposed to situations such as losing their reputations, experiences that make them feel uncomfortable, or publications that unintentionally become available to a broader audience than the audience initially expected [4]. For these reasons, users' concerns regarding the vulnerability of their personal data have been raised.

Although users state that they are concerned about their privacy [5, 6], there are works that highlight the difference between users' attitude and their actual behavior towards privacy [7, 8]. This phenomenon is known as the Privacy Paradox. A way to explain this phenomenon is the users' perception at the moment of privacy decision making [9]. When they are going to share personal data, users assess the benefits and risks of sharing personal data. If users perceive the benefits to be higher than the risks, they will share. However, privacy risks are perceived as being abstract and psychologically distant, and more related to the distant future, while the social rewards are perceived to be psychologically near and more concrete, and related to the short term. If users had informative and personalized metrics available about the risks, they could better assess their privacy decisions. The nudge mechanisms are a great solution since it may minimize regret and align the behavior with the stated preferences [10].

The information shared by users, especially personal data, has different levels of sensitivity ranging from totally trivial to extremely intimate data. Legislation such as the GDPR[2], NIST[3], or UKAN[4] (which emerged

[1] Over the course of this paper, when we use user and network concepts, we refer to OSN users and OSNs, respectively.

[2] European General Data Protection Regulation

[3] National Institute of Standards and Technology

[4] UK Anonymisation Network

URI: https://hdl.handle.net/10125/64052
978-0-9981331-3-3

from the need to protect users' data) distinguishes different levels of data sensitivity. Companies that buy, sell, and exchange users' data as an economic resource also consider different values of data based on the kind of information provided and whether they can link it to other data [11]. Companies have included the data as part of their business model in a data-driven economy. However, users are not completely aware of the value or sensitivity of their data. Moreover, we have different perceptions of sensitivity for our personal data depending on socio-cultural factors [12]. Therefore, if we empower users by making them aware of the value and/or sensitivity of their personal data through nudge mechanisms, they will be able to choose more suitable privacy policies [10].

In this work, our main contributions are the following: (i) a literature review about the different approaches to estimate the sensitivity of personal data (law, market, individuals, linguistic, and social networks); (ii) a proposal for a ranking/metric of sensitivity as the combination of all of the approaches that adapts them to the OSN domain; and (iii) a validation experiment that tests the effect of sensitivity nudges (based on our metric proposal) on real users' behavior. The paper is organized as follows. Section 2 analyzes and reviews previous works on establishing a sensitivity value for data. Section 3 presents our proposal for calculating the sensitivity value of social network publications. Section 4 includes the results of the experiment carried out to assess the effect of informing the user (via a nudge message) about the sensitivity of their publications before sharing them. We discuss our findings in Section 5 and provide our final conclusions in Section 6.

2. Literature review

2.1. Definition of personal data

First, we define what we mean when we speak about data, information, or personal data. Data is the raw material that is processed and refined to generate information that provides meaning. Individually, a single piece of data is rarely useful. For example a single date may be an appointment, a holiday, or an anniversary [13]. However, data is often used to specifically mean digitally stored quantified information. In this paper, we use the terms data and information to refer to the same concept. Personal data is information that can be linked directly or indirectly to an individual and can specifically identify him/her.

At the same time, the sensitivity of information is the potential loss that is associated with the disclosure of that information. This definition allows for the fact that sensitive information is perceived as being riskier and more uncomfortable to divulge [14]. Generally, by definition, personal data is more sensitive than data.

2.2. Quantifying the value of personal data

According to Acquisiti et al. [15], there is not just one method for properly establishing the value of privacy and personal data. Different references could be considered to establish this value, such as the money users would be willing to accept for their data, the money they would be willing to pay to protect their data, the cost of making their data public, etc. For this purpose, we reviewed the relevant research studies that proposed rankings and metrics. We have detected four different approaches for sensitivity that are based on (i) laws and regulations; (ii) market valuation; (iii) individuals' valuation; and (iv) linguistics. Below, we present and discuss the solutions provided by each one.

Law & regulation. In this approach, countries have been forced to regulate company activities that collect, store, and manage personal data. These regulations distinguish between different levels of sensitivity of the data that requires more protection than other data. The starting point for defining sensitive data under EU law is the list of "special categories of data" in the GDPR, which is based on the concept of privacy as a fundamental right. According to Article 9 of this regulation, sensitive data includes personal data revealing racial origin, political opinions, or religious or other beliefs as well as personal data on health, sex life, or criminal convictions [16]. Personal data that does not match these categories is also protected but is considered to be less sensitive, so companies do not have so many controls. In the UK, the UK Anonymization Network (UKAN) classifies data following two criteria: whether or not data is personal, and whether data may or may not be identifiable. It is interesting to consider the data that is not personal but that may be used to identify an individual such as vehicle registration or a dynamic IP address, since it may be strongly associated with an individual [17]. In US law, there is no comprehensive data protection regulation and no clear starting point for defining sensitive data that is analogous to the special categories of personal data found in the EU Data Protection Directive [18]. Agencies such as the National Institute of Standards and Technology (NIST) and the Department of Homeland Security have struggled to provide a precise definition of personally identifiable information, but they have not completed the next step of defining different categories of sensitivity or developing a topology of personal data that quantifies personal data. Current US laws and regulations cover only the use of certain types of personal data, such as financial and medical information.

Based on laws that are related to the protection of individual privacy in personal information record-keeping systems, Turn proposes a sensitivity scale and classification for personal information [19]. This scale consists of six levels that are based on the potential adverse effects on the individual, which may range from a mild annoyance to physical harm. The levels are: AS, public (by statute); A, public; B, limited; C, restricted; D, confidential (by statute); E, sensitive (by statute); and F, secret (by statute). The work also provides a simplified classification on three levels merging the ones above: basic (AS, A, B); medium (C, D); and high (E, F). The main problem with the law and regulations approach is that it groups information into broad, abstract categories without providing a scale or a ranking that indicates sensitivity in a fine-grained way.

Market valuation. This approach focuses on the information value for companies. Companies generate economic benefits from users' data and have decided to include users' data in their business models. User data and knowledge derived from it are sold and bought by companies for different purposes such as developing new features, offering new services, customizing an advertising campaign, etc. A report elaborated by the OECD[5] [20] analyzes different methodologies for measuring and estimating the value of personal data from a purely monetary perspective (i.e., without taking into account the indirect impacts of the use of personal data on the economy or society). This report analyzes approaches from two perspectives based on the market's valuation and individual's valuation. From the market perspective, the report assesses the value of data from indicators such as the market revenues obtained per data record, the market prices for data, the cost of a data breach, and data prices in illegal markets. From the individual perspective, the report assesses the value of data from indicators such as an individual's willingness to pay to protect data. The result is several rankings based on different indicators. An example of ranking based on the indicator of the market prices for data is made up of the following types of data that are ordered from highest to lowest cost: bankruptcy information, felony, employment history, sex offender, education background, unpublished phone number, business ownership, credit history, marriage/divorce, past address, social security number, address, voter registration. Another example of ranking based on the prices that individuals are willing to pay to protect their information is made up of the following criteria: the top tier includes social security numbers (national identity numbers) and credit card information, which most people value highly (USD 150–240 per entry); the middle tier contains digital communication history, such as web browsing history as well as location and health information (around USD 50); the last tier of information contains facts about users, including online purchasing history and online advertising click history, to which individuals attach little value (USD 3–6).

The Financial Times newspaper developed a calculator app based on the analysis of industry pricing data from a range of sources in the US [21]. Malgieri et al. [11] distinguish the following categories according to their economic value (from lower to higher): general (mainly demographic) information about a person; shopping, financial, or vacation intentions; personal data of people going through certain important life events (such as getting married, having a baby, etc.); and personal data containing specific health conditions or information on taking certain prescription (the highest value). In their work, they found that all of the data of a single person is not much valuable economically (approximately less than one dollar). The authors emphasize that the price of personal data has followed a declining trend in recent years. Conversely, companies collect the personal data of more and more people and this data can be resold several times, increasing the profits generated. Another important factor that is highlighted is that there is a positive relationship between the sensitivity of data and its economic value (i.e., the more sensitive the personal data, the higher its economic value).

Individuals' valuation. Another interesting approach is the individuals' perception of the data. After the analysis of the responses of 310 adults in a national survey, Milne et al. [22] detected six groups of personal information and established a ranking based on the consumers' perceived sensitivity. The groups detected (ordered from lowest to highest sensitivity) were: basic demographics, personal preferences, contact information, community interaction, financial information, secure identifiers. In addition, the authors detected that the perception of risk is multidimensional. They considered that there is not just one type of risk. They differentiated four types of risk where the six information groups could be classified: physical risk (secure identifiers); monetary risk (financial information and secure identifiers); social risk (community interaction); and psychological risk (community interaction and secure identifiers).

[5] Organization for Economic Cooperation and Development

Schomakers et al. [12] established a sensitivity ranking of 40 different data types. The authors compared their results with Brazilian, EEUU [23], and German individuals. Based on the ranking, they grouped data into three categories (high, medium, and low) using a linear clustering based on the sensitive value of data. Rumbold et al. [13] propose six categories of data (based on the UKAN): non-personal data; human-machine interactions; human demographics, behavior, thoughts and opinions; human characteristics (unprotected); human characteristics (protected); and medical or healthcare data. The authors propose a spectrum of sensitivity for these six categories and subcategories inside them, where the relative frequency with which data would occur is given by individuals.

Although these previous works highlight that the perceived sensitivity of a specific type of data varies depending on socio-cultural factors (i.e., religion is highly sensitive in areas where there is a high degree of sectarian conflict), the ranking of data based on sensitivity is similar among individuals [12, 13].

Linguistics. The last approach considered is centered on linguistics and the use of words. According to Viejo et al. [24], the terms that provide/disclose a large amount of information are also likely to be sensitive. In this respect, several privacy-protection methods for textual data and empirical studies have shown the close relationship between the informativeness of textual terms and their sensitivity [25, 26]. Therefore, Viejo et al. [24] measure the informativeness of a term according to its Information Content (IC), which is computed as the inverse of the term's probability of appearance in a corpus. To that end, they use the largest and most up-to-date electronic repository available: the Web. Other works such as Imran et al. [27] also consider the same idea of linguistics properties to quantify the sensitivity of data, but they use the ontological properties of DBPedia[6] resources to create taxonomic generalizations of words. To do this, they use SPARQL as a query language and the Semantic Web API. Thus, the deeper a word is in the taxonomic tree of generalization, the higher the sensitivity of the word.

2.3. Sensitivity in the OSN domain

Some of the categories and information types analyzed in the previous approaches may make no sense in the social network domain (e.g., DNA profiles or bankruptcy information). Conversely, other categories that were not included in the above approaches could appropriately be considered as sensitive due to the risks or consequences for the post's owner, such as personal attacks (which are very common in Twitter) [28].

Information types from regrets in OSNs	Source
Location data	[31]
Personal and Family issues	[32, 28, 33, 31]
Work and Company data	[32, 28, 31]
Religious issues	[32, 34, 28, 33]
Political issues	[32, 34, 28, 33]
Health and Medical	[28, 33, 31]
Alcohol consumption	[32, 28, 33, 31]
Illegal drug use	[32, 28, 33, 31]
Sexual content	[32, 28, 33]
Negative emotions	[32, 34, 28, 31]
Positive emotions	[34, 31]
Attacks on individuals	[32, 34, 28, 33]
Attacks on collectives	[32, 34, 28, 33]
Lies and Secrets	[32, 34]

Table 1. Summary of the most common information types that cause regrets in social networks

An important aspect in the OSN domain is identification. Depending on the network platform, users need to provide a minimum amount of information about themselves in order to have a profile. Even when this information is not required, users upload information about themselves to be identifiable to others (e.g., their real name, birthday, a photo of themselves, etc.) [29]. This effect emerges from the nature of OSNs for communicating and socializing with others. When users are identifiable, they are easily included in the social network structure as friends or followers, and, in addition, they increase their social rewards with the interactions [30]. Therefore, information that users share could be personal by default.

As a consequence, there are a lot of works in the literature that collect and group different kinds of content based on users' regrets caused by sharing data on social networks [32, 34, 28, 33] (see Table 1). Most of them consider the most common regret as revealing too much information. Based on this regret, users usually highlight posting about categories like personal and family issues, religion, politics, health, work and company issues, and location data. These categories fully match the categories of personal data from the previous approaches analyzed. As an example, although it is not a common practice, posting lists of defaulters involves a high risk for the publisher and the defaulters due to other users' reactions. On the other hand, network usage has generated regrets that are related to self-presentation and reputation. The information types that could cause these regrets such as publications

[6] http://dbpedia.org/About

about alcohol or illegal drug use, obscenity, personal attacks, complaints, and curses were not considered in the approaches analyzed. Since these information types can change the other's perceptions towards the user that publishes, they are also sensitive for the users' privacy. Extreme emotions are also included in this same group. Most of them are negative emotions, but there are some cases of extreme happiness that can cause regrets by the reactions of others (in this case, moved by jealousy). Finally, this research also highlights regrets caused by posting lies and secrets, but no one (to the best of our knowledge) has enough information to detect them.

In fact, some works such as [31] have tried to identify some of these categories, but the only thing they did was to reveal the habits of users. They did not extend their work to enhance the users' awareness of the sensitivity of this type of information or privacy-seeking behavior. Our goal is to propose a quantification value for users' posts and to use it to improve users' privacy.

3. Proposal

In the proposal of this work, we address three issues: (i) providing a sensitivity value for each information type that might be present in the OSN domain; (ii) providing and justifying the sensitivity value for the regret-based information types; and (iii) the representation of the total value of sensitivity for a publication, taking into consideration that multiple information types could appear in the same publication.

For the first issue, we propose a ranking that combines the sensitivity values and the information types that appear in the works reviewed (see Figure 1). The value for each information type was normalized on a scale of 0 to 1, where 0 means no sensitive data and 1 means the maximum sensitive value of data. We added new information types based on users' regrets (see Table 1). For each of these information types, we proposed a potential sensitivity value considering the nature of the regret and its proximity to other information types. For the *Illegal drug use* information type, we propose positioning this information type between *Medication* and *Law enforcement files* (immediately next to *Medication*) since it could be considered as medication, but, depending on the kind of drug, it may lead to legal consequences. For the *Alcohol consumption* information type, we propose positioning this information type as individuals' behaviors, especially since it represents health-damaging behaviors [35] (below *Religion* and close to *Lifestyle*). For strong sentiments, works about regrets concurred that negative emotions are more regrettable than positive ones [32, 34]. Therefore, we propose positioning the positive

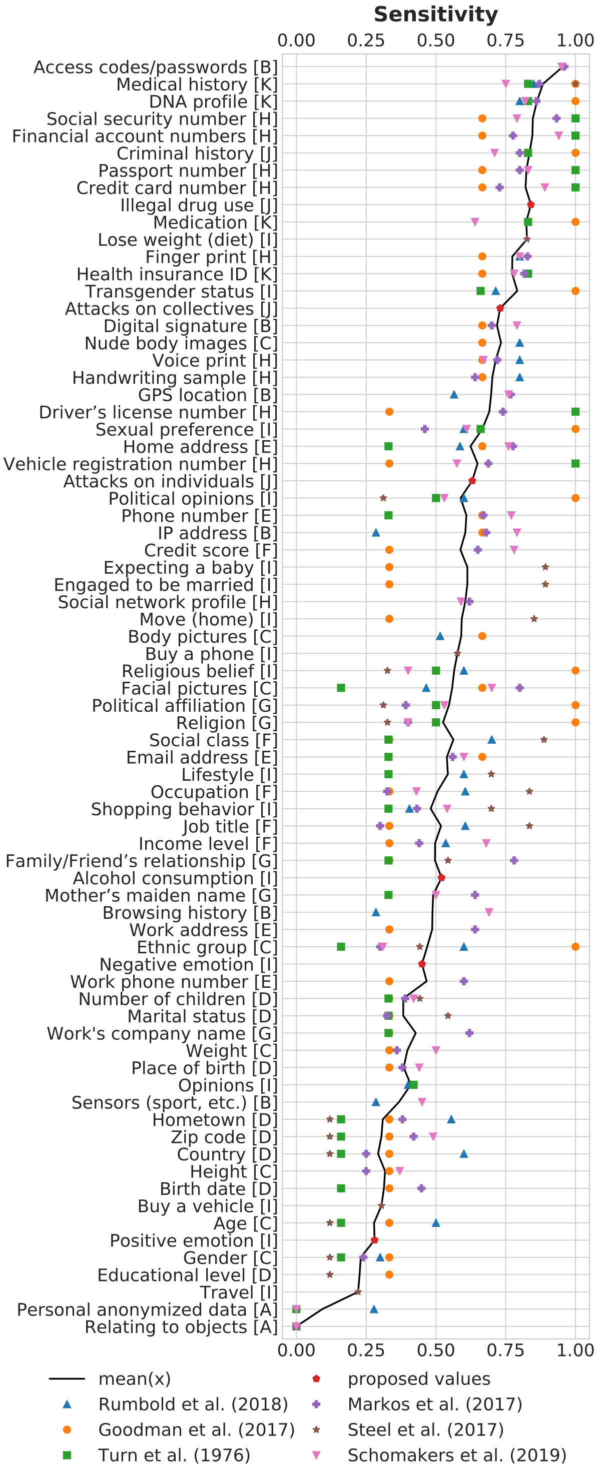


Figure 1. Sensitivity for information types. Categories are A=Non-personal, B=Human-machine, C=Human characteristics, D=Demographics, E=Contact info., F=Status/Financial, G=Association, H=Identifiable info., I=Behaviors/Intentions, J=Law-related, and K=Medical. The color of each point corresponds to the research paper that considered the data.

emotions in the same place as demographic data, and the negative emotions over the *Opinions* information type. Finally, the information type related to attacks, curses, offensive comments, or profanity depends on the kind of target. On the one hand, if the target is an individual (i.e., *Attacks on individuals*), we propose positioning the information type between *Negative emotion* and types related to laws due to its possible legal consequences. On the other hand, if the target is a group or association (i.e., *Attacks on collectives*), the sensitivity is greater than the sensitivity of *Attacks on individuals*, and, in some countries, it could be considered as a hate crime against a collective.

Finally, using the sensitivity information types from previous works and the proposed sensitivity values for regrets, we created a ranking (see Figure 1) of 74 information types (y-axis). We grouped the information types into the following categories: (A) Non-personal, related to anonymized data or object data; (B) Human-machine, data generated from technology interactions; (C) Human characteristics, related to physical aspects; (D) Demographics, related to common features that are not identifiable; (E) Contact info., any information that allows others to contact you; (F) Status/Financial, related to monetary status; (G) Association, data able to link users with other individuals or collectives; (H) Identifiable info., data that can directly identify an individual; (I) Behaviors/Intentions, related to past and future actions; (J) Law-related, that can or have caused legal consequences; and (K) Medical, related to health data. Once we quantified and normalized the value for each information type, we placed them in order from the most sensitive types to the least sensitive types. The sort criterion was the mean value of the information type and the number of research works that assessed that information type. From the resultant ranking, we observed that there is a consensus in most cases for the information types at the extremes of sensitivity (the lowest and highest value). We also observe that information with low values of sensitivity is mainly demographic, anonymized, or related to objects, while information with the maximum values of sensitivity is mainly passwords/access codes (because they can give access to other sensitive information), health information, and identifiable and unique data of an individual. In contrast, the rest of the information types have less consensus with a huge sensitivity variability among the works. We highlight the information types of facial pictures, ethnic/race, and behavior/intentions as being of greatest variance. Furthermore, we also observe that behaviors/intentions of individuals (such as losing weight, expecting a baby, engaging to be married, etc.) are especially valuable for companies, while other approaches give them values that are significantly lower. For this reason, it is so important to consider all of the approaches involved.

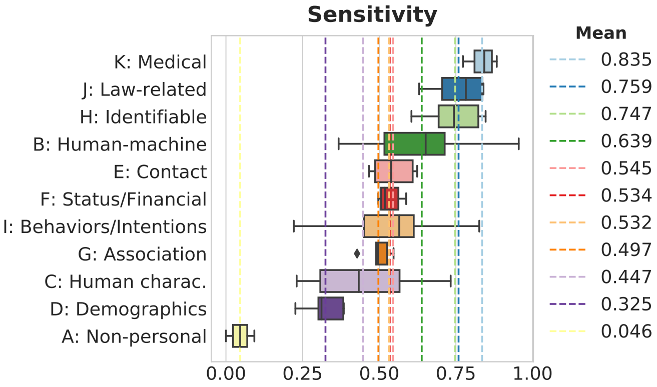


Figure 2. Distribution of sensitivity values for the identified categories of information types.

In order to provide an estimated value by category, we selected the mean as the representative value for the category. Figure 2 depicts the distribution values per category, which includes a Box-plot to enhance visual comprehension. The mean values are included in the legend, which reflects the conclusions extracted in the ranking. Thus, if a relevant information type was not included in our proposal, other researchers could derive an approximate sensitivity value by classifying the new information type in a category and/or comparing it with the other information types.

We illustrate a common scenario for proposing a value of sensitivity for a publication. A user posts on a social network. The post (e.g., W) may consist of a media item (e.g., a photo), a textual message, or both [32]. The question that may arise is: What value of sensitivity s should the publication W have (i.e., $s(W)$)? In fact, the combination of information types ($t_i \in W$) actually creates value. When the attribute name is provided as “John” or the attribute gender is provided as “male”, these are meaningless. Single attributes without any further context have no monetary value. Only when they are combined (i.e., when John is male) do these attributes create value [11]. In OSN, the profile provides a linkable space for new attributes. Therefore, assessing personal data sensitivity is not about assessing individual information types, but rather assessing combinations of personal data. Using the approach of the market pricing valuation of personal data [21], where data is bought and sold in a combined way, we propose using the same system (i.e., summing their values) to assess combinations of information types.

$$s(W) = \sum_{t_i \in W} s(t_i) \tag{1}$$

Age [years]	mean(SD)	13.03 (0.70)
	12 years	22.45%
	13 years	51.53%
	14 years	26.02%
Gender	male	53.57%
	female	46.43%

Table 2. Demographics of participants (N=196).

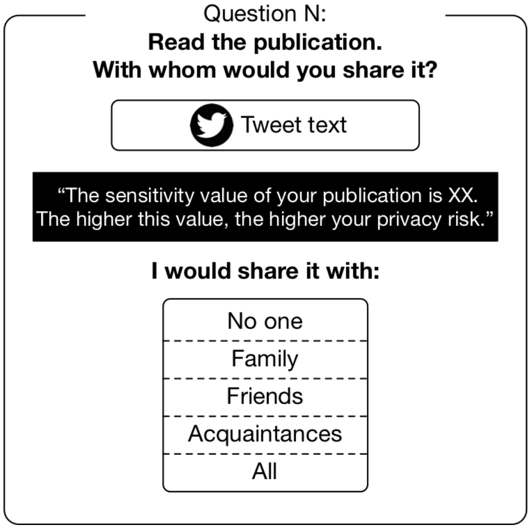
Thus, the more information types the post has and the more sensitive they are, the riskier it is to share the post.

4. Experiment

We ran an experiment to test the effect of informing the user (via a message) about the sensitivity of their publications before sharing them. The message acts as a nudge for users. Nudges attempt to influence decision making in order to improve individual well-being without actually limiting users' ability to choose freely, thus, preserving freedom of choice [10]. For this reason, we consider that nudges could reduce users' regrets.

4.1. Methodology

The experiment consisted of a questionnaire that was distributed online within the context of a one-month workshop about social networks for teenagers. The questionnaire was embedded in the social network platform that they used in the workshop. A total of 196 Spanish participants (from the Valencia area) ranging in age from 12-14 years old completed the experiment. The sample shows heterogeneous distributions regarding age (M = 13.05, SD = 0.71), and gender (53.57% male teenagers) (see Table 2). The questionnaire consisted of asking participants to choose an audience (i.e., the privacy policy) for real publications (selected previously from Twitter) as if they had written them. The privacy options available were based on the social circles defined by [36]. We removed the social circles that made no sense for teenagers (such as coworkers), and we combined the first and second level of family into a single social circle. The final options were: *No one*, *Family*, *Friends*, *Acquaintances*, and *All*. We collected 53 tweets that were classified by raters taking into account the information types that the tweets had (Figure 1). We finally chose the 30 tweets with the highest level of agreement with the information types identified to be included in the questionnaire. From the manual classification, we calculated the sensitivity value of the tweet using our proposal (i.e., accumulating the sensitivity of the different information types, Eq. 1).


Question N:

Read the publication.
With whom would you share it?

Tweet text

"The sensitivity value of your publication is XX.
The higher this value, the higher your privacy risk."

I would share it with:

No one
Family
Friends
Acquaintances
All

Figure 3. Template of the survey questions.

The questionnaire had two stages, which took place in different weeks, with 15 questions per stage. The questions were designed following the structure depicted in Figure 3. The difference in the questions between stages was the nudging message that was hidden during the first 15 questions. In the first stage, the nudges were not activated, so the participants did not receive any kind of advice concerning the privacy decision. In contrast, in the second stage, the nudges were activated, assisting the participants with the privacy decision. Thus, we observed and assessed whether meaningful changes in their behaviors were produced with nudges about the sensitivity of the publication.

4.2. Results

Once the experiment had ended, a total of 5880 privacy decisions (196 participants x 30 questions) were collected. Each entry consisted of the participant identifier, the tweet identifier, whether the sensitivity nudge was enabled, the sensitivity value, and the privacy policy choice. We codified the data following the next criteria: the sensitivity nudge variable as a binary value (representing whether it was enabled); the sensitivity value was discretized into four grades (*none*, *low*, *medium*, and *high* sensitivity); and the privacy policy choice was normalized taking into account how restrictive the choice was, considering *No one* as the minimum value (0), and *All* as the maximum value (4). Since participants did not repeat their choices for the same tweet, we considered running an independent sample test to assess the effect of the nudge messages on the participants' decisions. Moreover, we also wanted to evaluate the effect taking into account the information

Source	*df*	Mean Squares	*F*	*p*	Effect Size
(A) Nudge	1	10.55	4.84	.028*	.001
(B) Sensitivity	3	165.26	75.87	.00**	.038
A × B	3	13.68	6.28	.00**	.003
Error	5873	2.17			

Table 3. ANOVA analysis for the privacy policy chosen ($\alpha = .05$). $^{*}p < .05$, $^{**}p < .01$.

sensitivity value of the tweet. Therefore, we used an ANOVA test ($\alpha = .05$) using the *privacy policy* value as the dependent variable and using the *sensitivity* value and whether the *nudge* was activated as the fixed factors (see Table 3).

The results of the statistical tests reveal a significant difference in users' behavior regarding privacy policy choices. On the one hand, the presence of nudge messages was significant to the privacy decision made (p-value = .028). On the other hand, the test results also showed a significant difference in the privacy decision made by participants regarding the sensitivity value shown in the nudge message (p-value = .00). Finally, the combination of both variables (nudge and sensitivity) also revealed a significant difference in their decisions (p-value = .00). Table 4 depicts the comparison of the mean value of the privacy policy decisions organized by sensitivity and nudge variables. The table shows the differences in the mean of privacy policy choices, that the ANOVA statistical test confirms that are significant. From the privacy decisions made by the participants, we can extract some remarkable facts. First, the participants were able to slightly identify the sensitivity of the information contained in the tweet. They showed more restrictive behavior for the information with higher sensitivity (see the *Mean* column for rows where the *nudge* variable is 0), except at the low sensitivity level. We explored the information types identified in those tweets. Most were from the Intentions category, which was one of the categories with the lowest level of agreement. Therefore, we believe that that may have had some effect on their initial decision-making. Second, the participants were less restrictive in their privacy policy choices when the nudge message confirmed that the message was not sensitive (see the *Mean* column for rows with non-sensitivity), while they were more restrictive with sensitive content. Finally, the difference in the mean of the privacy policy choices of the participants was higher when the sensitivity value of the message was higher.

Sensitivity	Nudge	N	Mean	Std. Error
none	0	784	1.858	.061
	1	784	2.042	.061
low	0	784	1.166	.048
	1	784	1.130	.047
medium	0	686	1.507	.056
	1	686	1.302	.058
high	0	686	1.408	.057
	1	686	1.117	.057

Table 4. Comparison of privacy policy mean values by sensitivity grade and nudges (0: the most restrictive policy; 4: the less restrictive policy).

5. Discussion

Depending on the information type, data is more or less sensitive/valuable due to: the cost of storing it [20]; the willingness of users to provide it [22]; the monetary benefits derived from extracting knowledge from it [21]; the loss of users' privacy [12]; the amount of information it provides [16]; etc. After the analysis and review of previous works that deal with the assignment of a sensitivity value to information types, we identified that some information types have small variability of value among the different works (especially in the case of information types that are located at the extremes of the sensitivity values). In this work, we have identified that categories such as demographics and human characteristics have a high degree of agreement among works that evaluate this data as being of low sensitivity. We have also identified that medical, legal and personally identifiable data categories also have a high degree of agreement, evaluating this data as being highly sensitive. For information types with less agreement among approaches, we highlight user behaviors and intentions. These may be valuable to companies, but the other approaches (laws & regulations and individuals' valuation) give them low sensitive value or they do not even assess the value of these types.

Regarding the proposal presented in this work, for estimating a sensitivity value for each information type, we decided to accumulate the values of all the works and calculate the mean value for each information type. Based on these values, we create a ranking. This ranking could be extended considering new values for a certain information type and including new information types. We included new information types from OSNs regrets positioning them by proximity to others types. However, this order could be also validated through questionnaires to users and companies about their perception of sensitivity/value of these new information types in comparison with the existing ones. Moreover, the value of estimated sensitivity introduced to users could be provided to users in different ways (e.g., as a monetary value, as a color scale, etc.) for testing which representation has a greater effect. Another aspect to assess other alternatives to estimate the aggregated value of data. Only few works calculate the sensitivity value of the information and, to the best of our knowledge, there is a lack of proposals that consider the combined value of information.

As some works point out [10, 11], empowering users during complex decisions with valuable information has two direct effects on them: (i) it raises their awareness; and (ii) it nudges their behaviors toward controlled decisions (i.e., with expected consequences). Wang et al. [37] proved that identifying the imagined audience before making decisions about posting changed the users' privacy decisions. Schöning et al. [38] tested significant differences by personalizing the styles of the nudges shown. Alemany et al. [39] tested personalized nudges based on an estimation of the final audience before posting. In that work, the authors reported that teenagers used more restrictive policies when they were aware of the potential audience. In this work, we empower users with nudges that contain information about the sensitivity value of the information they would share in OSNs. We assess how teenage users choose the privacy policy for a given publication when we nudge about its sensitivity. Through the experiment, we found out that the teenagers of the experiment had some previous knowledge about the sensitivity of information, because they chose restrictive privacy policies for the most sensitive posts when nudges were not activated. We also figured out that nudge messages about sensitivity had a significant effect on their behavior as well as the sensitivity level shown on the nudge message. The effect on teenagers' privacy behavior was more significant the greater the sensitivity value included in the nudge message (i.e., the privacy policy mean value decreased more for *high* sensitive posts than for *low* sensitive posts). From the results, we conclude that the teenagers were able to understand the nudge message that contained information about the sensitivity of their publications and they used them to have less risky behaviors on social networks.

6. Conclusions

This paper proposes a combined ranking using sensitive information types collected from an extensive literature review as well as a set of newly proposed information types for the OSN domain based on the most common regrets. The ranking provides the quantification of the sensitivity value of the different information types and could also be used to approximate the value of new information types that are not included in this work. Our proposal for assessing the sensitivity of a publication uses the ranking to estimate its value by accumulating the values of the different information types identified in the publication. The sensitivity value associated with a publication was used in nudges that were tested in an experiment with 196 teenagers. In this experiment, the teenagers had to choose a privacy policy for a set of publications with different degrees of sensitivity. The information provided by the nudges made them more aware of the privacy risk before choosing a privacy policy. The results of the experiment showed the relevance of empowering users with information about the sensitivity of their publications in order to make informed decisions that protect their privacy.

As future work, we plan to include our proposal in a real social network; thus, we could apply the nudges to the users' generated content in their daily usage. We also think that it would be interesting to match the topics of the publication with the sensitive categories developed in our proposal in order to improve its performance.

7. Acknowledgement

This work is partially supported by the Spanish Government project TIN2017-89156-R and the FPI grant BES-2015-074498.

References

[1] S. Survey, "Most popular daily online activities of adult internet users in the united states," 2017.

[2] D. J. Solove, *Understanding privacy*, vol. 173. Harvard University Press Cambridge, May, 2008.

[3] C. Richthammer, M. Netter, M. Riesner, J. Sänger, and G. Pernul, "Taxonomy of social network data types," *EURASIP Journal on Information Security*, vol. 2014, no. 1, p. 11, 2014.

[4] J. Alemany, E. del Val, J. Alberola, and A. García-Fornes, "Estimation of privacy risk through centrality metrics," *FGCS*, vol. 82, pp. 63–76, 2018.

[5] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015.

[6] Y. Jeong and Y. Kim, "Privacy concerns on social networking sites: Interplay among posting types, content, and audiences," *Computers in Human Behavior*, vol. 69, pp. 302–310, 2017.

[7] S. Kokolakis, "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon," *Computers & security*, vol. 64, pp. 122–134, 2017.

[8] S. Barth and M. D. De Jong, "The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior–a systematic

literature review," *Telematics and Informatics*, vol. 34, no. 7, pp. 1038–1058, 2017.

[9] C. Hallam and G. Zanella, "Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards," *Computers in Human Behavior*, vol. 68, pp. 217–227, 2017.

[10] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, *et al.*, "Nudges for privacy and security: Understanding and assisting users' choices online," *ACM Computing Surveys*, vol. 50, p. 44, 2017.

[11] G. Malgieri and B. Custers, "Pricing privacy–the right to know the value of your personal data," *Computer Law & Security Review*, vol. 34, no. 2, pp. 289–303, 2018.

[12] E.-M. Schomakers, C. Lidynia, D. Mllmann, and M. Ziefle, "Internet users' perceptions of information sensitivity – insights from germany," *Inter. Journal of Information Management*, vol. 46, pp. 142 – 150, 2019.

[13] J. M. Rumbold and B. K. Pierscionek, "What are data? a categorization of the data sensitivity spectrum," *Big data research*, vol. 12, pp. 49–59, 2018.

[14] D. L. Mothersbaugh, W. K. Foxx, S. E. Beatty, and S. Wang, "Disclosure antecedents in an online service context: The role of sensitivity of information," *Journal of service research*, vol. 15, no. 1, pp. 76–98, 2012.

[15] A. Acquisti, C. Taylor, and L. Wagman, "The economics of privacy," *Journal of Economic Literature*, vol. 54, no. 2, pp. 442–92, 2016.

[16] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a right to explanation," *AI Magazine*, vol. 38, pp. 50–57, 2017.

[17] M. Elliot, E. Mackey, K. O'Hara, and C. Tudor, *The anonymisation decision-making framework*. UKAN, 2016.

[18] N. J. King and V. Raja, "Protecting the privacy and security of sensitive customer data in the cloud," *Computer Law & Security Review*, vol. 28, no. 3, pp. 308–319, 2012.

[19] R. Turn, "Classification of personal information for privacy protection purposes," in *Proc. of national computer conference*, pp. 301–307, 1976.

[20] OECD, "Exploring the economics of personal data," no. 220, 2013.

[21] E. Steel, C. Locke, E. Cadman, and B. Freese, "How much is your personal data worth," *Financial Times*, vol. 12, 2013.

[22] G. R. Milne, G. Pettinico, F. M. Hajjat, and E. Markos, "Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing," *Journal of Consumer Affairs*, vol. 51, no. 1, pp. 133–161, 2017.

[23] E. Markos, G. R. Milne, and J. W. Peltier, "Information sensitivity and willingness to provide continua: a comparative privacy study of the united states and brazil," *Journal of Public Policy & Marketing*, vol. 36, no. 1, pp. 79–96, 2017.

[24] A. Viejo and D. Sánchez, "Enforcing transparent access to private content in social networks by means of automatic sanitization," *Expert Systems with Applications*, vol. 62, pp. 148–160, 2016.

[25] D. Abril, G. Navarro-Arribas, and V. Torra, "On the declassification of confidential documents," in *Int. Conf. on Modeling Decisions for Artificial Intelligence*, pp. 235–246, 2011.

[26] D. Sanchez, M. Batet, and A. Viejo, "Automatic general-purpose sanitization of textual documents," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 853–862, 2013.

[27] M. Imran-Daud, D. Sánchez, and A. Viejo, "Privacy-driven access control in social networks by means of automatic semantic annotation," *Computer Communications*, vol. 76, pp. 12–25, 2016.

[28] L. Zhou, W. Wang, and K. Chen, "Tweet properly: Analyzing deleted tweets to understand and identify regrettable ones," in *Proc. of the 25th Int. Conf. on WWW*, pp. 603–612, 2016.

[29] W. Xie and C. Kang, "See you, see me: Teenagers self-disclosure and regret of posting on social network site," *Computers in Human Behavior*, vol. 52, pp. 398–407, 2015.

[30] C.-M. Chiu, M.-H. Hsu, and E. T. Wang, "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories," *Decision support systems*, vol. 42, no. 3, pp. 1872–1888, 2006.

[31] A. Caliskan Islam, J. Walsh, and R. Greenstadt, "Privacy detective: Detecting private information and collective privacy behavior in a large social network," in *Proc. of the 13th Workshop on Privacy in the Electronic Society*, pp. 35–46, ACM, 2014.

[32] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor, "I regretted the minute i pressed share: A qualitative study of regrets on facebook," in *Proc. of the 7th symposium on usable privacy and security*, p. 10, ACM, 2011.

[33] Q. Wang, J. Bhandal, S. Huang, and B. Luo, "Classification of private tweets using tweet content," in *Proc. of the 11th ICSC*, pp. 65–68, IEEE, 2017.

[34] M. Sleeper, J. Cranshaw, P. G. Kelley, B. Ur, A. Acquisti, L. F. Cranor, and N. Sadeh, "I read my twitter the next morning and was astonished: A conversational perspective on twitter regrets," in *Proc. of the SIGCHI*, pp. 3277–3286, ACM, 2013.

[35] B. J. Everitt and T. W. Robbins, "Neural systems of reinforcement for drug addiction: from actions to habits to compulsion," *Nat. neuroscience*, vol. 8, p. 1481, 2005.

[36] R. Dunbar, *How many friends does one person need?: Dunbar's number and other evolutionary quirks*. Faber & Faber, 2010.

[37] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, and L. F. Cranor, "Privacy nudges for social media: an exploratory facebook study," in *Proc. of the 22nd Int. Conf. on WWW*, pp. 763–770, ACM, 2013.

[38] C. Schöning, C. Matt, and T. Hess, "Personalised nudging for more data disclosure? on the adaption of data usage policies format to cognitive styles," in *Proc. of the 52nd HICCS*, 2019.

[39] J. Alemany, E. del Val, J. Alberola, and A. García-Fornes, "Enhancing the privacy risk awareness of teenagers in online social networks through soft-paternalism mechanisms," *International Journal of Human-Computer Studies*, 2019.