Social Transparency through Recommendation Engines and its Challenges: Looking Beyond Privacy

Remus TITIRIGA
Babeş-Bolyai University, Cluj-Napoca, Romania
titiriga_r@yahoo.com

Our knowledge society is quickly becoming a 'transparent' one. This transparency is acquired, among other means, by 'personalization' or 'profiling': ICT tools gathering contextualized information about individuals in men–computers interactions. The paper begins with an overview of these ICT tools (behavioral targeting, recommendation engines, 'personalization' through social networking). Based on these developments the analysis focus a case study of developments in social network (Facebook) and the trade-offs between 'personalization' and privacy constrains. A deeper analysis will reveal unexpected challenges and the need to overcome the privacy paradigm. Finally a draft of possible normative solutions will be depicted, grounded in new forms of individual rights.
Keywords: *Recommendation Engines, Profiling, Privacy, 'Sui Generis' Copyright*

1 Introduction

This paper is a work in progress building a framework to be deployed in future researches. It focuses mainly the world of advertising (E-advertising) where specific ICT tools are trying to connect potential consumers to products they would be interested to buy.

The economic mechanism behind is linked to 'the long tail'[1] a retailing strategy of selling large numbers of unique items in relatively small quantities. The total sales of that large number of these articles will form the long tail. Only if the costs of storage and distribution are insignificant, as in some forms of E-commerce, it becomes economically rational to sell relatively unpopular products. Chris Anderson, the creator of this concept in an October 2004 *Wired* magazine article, mentioned Amazon.com and Netflix as examples of businesses applying this 'long tail' strategy [2].

In order to accomplish this kind of strategy it is necessary to meet the buyer and those unique items he would be interested in. In fact it is necessary to understand, to 'personalize' or 'profile' the buying behavior of the individual. This is the place where the personalized men-computers interactions become important. As a matter of fact information gathered through online (Internet) interactions, contextualized by particular ICT instruments may aggregate knowledge about individuals resulting in a 'profile' or 'personalization' different from that seen in psychological thrillers. Most of all, its scope is limited to gathering components of 'homo oeconomicus' (individual seen as consumer with his tastes, habits, customs, etc). There is a certain evolution, to be explored below, of instruments allowing this economic 'personalization' or 'profiling'.

a) Behavioral targeting instruments

A first approach used *behavioral targeting* [3]. These instruments determined average trends of large groups of Internet users rather than of the actual individuals. They used information collected of an individual's web-browsing behavior, such as the pages they have visited or the searches they have made (using for example cookies) in order to select which advertisements should be displayed to a certain person. The unique dimension of the process was a vertical one grounded only on behavior of an individual.

The use of AdWords by Google is one of best known examples. This mechanism is very effective because it target users based on their search results [4]. AdWords Search analyzes each search request to determine which advertisement will receive the

sponsored links on each page of results. But at a deeper level the mechanism acts as a big, fast and automated auction for advertising.

b) Recommendation engines or systems
The next approach to personalized experience was related to the arrival of *recommendation engines or systems*. These systems try to make educated guesses and make recommendation about interesting items for a potential buyer.
Some types of recommendation systems are less interesting for the present discussion but worth being mentioned. First examples of this kind are personalized recommendation systems based on past behavior of the user. It is the case for Google (since December 2009) which recommends useful results based on user's search history.
Another kind is represented by content recommendation systems that rely on intrinsic qualities of an item. This is the method used by Pandora, an Internet radio service from Oakland, United States which developed a huge musical database of songs. A song can have up to 400 attributes that are completed (rated) by specialized musicologists. The whole effort has began some 10 years ago and until now it has been established a detailed profile for 740,000 pieces of music. .
The most interesting variety, for the present discussion, is the *recommendation system based on collaborative filtering*. This is a method of making automatic predictions (filtering) about the interests of a user by collecting taste information from many other users (collaborating). In everyday life, an individual asks his friends for advices about how to choose among newspapers, records, books, movies or other items. With the time being an individual can figure out which of his friends have tastes similar to his own, and which ones are different [5].
Computerized 'collaborative filtering' will automate the process [6] based on the idea that if two similar people like an item, there are probably several other items they would also find interesting. The collaborative software tries to find connections between

people's interests in a very labor intensive approach using data mining, pattern recognition or other complex techniques.
An interesting example is the competition organized by Netflix (an online DVD rental service) to discover the best 'collaborative filtering' algorithms for predicting the rating of films based on previous assessments. Specific algorithms are usually a trade secret but in 2006 Netflix, unsatisfied with his own 'Cinematch' algorithm, made public a portion of his film database and offered a prize of one million dollars to anyone able to improve the initial result by 10%. The winner of 21 September 2009 was a consortium of seven people from four countries, called BellKor's Pragmatic Chaos. Under competition's conditions they provided an exclusive license to Netflix and the latter has published a brief description (but not the source code) of the algorithm [7].
Collaborative filtering can be used to deliver information targeted to consumers far more accurately than other technique-for example keyword searching seen in 'behavioral targeting' above. Predicting an act (pressing a click on Google AdWords for example) has an average of 1% probability to be true. A better model for a recommendation system can grow, in a context of accurate and rich set of information about user's 'profile', this rate to 3 or 4%. It seems to be a small variation but the aggregation of a multitude of individual's behaviors makes a true economic difference.
In collaborative filtering one can find two dimensions at work: a vertical one (based on past evolutions of an individual) but equally a horizontal one (based on comparisons with other individuals' behavior). The strength of the system is related to "the wisdom of crowds (a large community makes better decisions than a single individual) and "the law of large numbers" (the larger the database the better the decisions will be) [8].
Finally one can obtain educated and unpredicted guesses and a deeply personal experience as result of social interactions and their ICT processing

c) Personalized experience in social networking sites

This final approach can be greatly improved with new evolutions in social networking (the case of the new Facebook package). In fact the social dimension mentioned above is extremely developed in social networking sites as Facebook (which is the second most visited site of the Internet after Google, with 400 to 500 million users). A 'personalization' or 'personalized experience' as described above, existed already in Facebook which gives users the option of switching between a straight feed, which shows all their friends' news in chronological order, and an algorithmically created selection of the updates. And in the right-hand column, Facebook uses a different set of algorithms to recommend new friends.

Nevertheless a more remarkable evolution is related to the new package of Facebook. The package, released by Mark Zuckerberg at F8 conference[9] held in April 2010, includes a collection of "plugs-ins" (set of functionalities to be easily added on an existing site), a semantic mark-up language and an API for applications developers. Whenever one add a friend on Facebook one create a link to him, whenever one register a page or a group, one create a tie of belonging to that group. If one aggregates all these links for a set of groups and individuals, one obtains a graph of relationships[10]. And within the new package the graphs is linking more and more people and *objects they like* (Opengraph).

With the plugs-ins, buttons "Like" and "Likebox" add to Facebook's users, the ability to put a key on any content as a Facebook recommendation. The purpose is essentially to allow users to recommend content that will be published as link on Facebook and permanently stored on their profiles.

The new package imply that in foreseeable future Facebook will know more about user's preferences (since Facebook's users are not anonymous) than anyone else. And Facebook can use them to enhance the 'personalized' experience accordingly. Google has dominated the Internet in the era of Web pages connections, when each link from a Web site to another was a vote and determined the most relevant pages in relation to a searched item. Today Google pays to social networks the access at their public updates but lacks a number of key data, which would allow a deeper customization of search. Google lacks a full list of user's friends, combined with a list of their interests. Social networks like Facebook have all these lists available and are thus in a position to build a stronger type of recommendation system and search-engines that provide results based on friendships and interests related to items. On such a base Facebook could finally obtain (using on a large scale the mechanisms of *collaborative filtering*) the best personalized recommendation system yet ever seen.

2 Personalized experience and privacy constrains

We will focus our attention on the limits of privacy paradigm in relation to above mentioned mechanisms for personalized experience. The accent will be put on a case study centered on the most recent Facebook's developments.

a) Classical paradigm about privacy

The notion of a right to privacy entered the legal world in 1890 with Harvard Law Review's publication of Samuel Warren's and Louis Brandeis '*The Right to Privacy*': *"recent inventions and business methods can create new needs for protection of what courts as early as 1834 called a right 'to be let alone'"* (Warren & Brandeis, 1890, p. 208).

The passage of time has proved Warren and Brandeis as perceptive intellectuals. Technological inventions enhanced the need for privacy protection and the law's development to guarantee the protection. From telephone to magnetic tape recorder, photography, personal computer, wireless telephone, electronic payment systems and the Internet, technology has created new

challenges to privacy and law's capacity to protect it.

In USA the word privacy does not appear in the Constitution, but it is has been constructed by legal scholars based on Bill of Rights, such as the Fourth Amendment (which forbids unreasonable search and seizure). U.S. citizens prize the right to be left alone in their homes, and courts have been generally compliant of this right. However, once a person leaves the home the right to privacy quickly disappears [11].

There are as well special laws that protect privacy in the U.S. while the main concern here is the government. The corporate word is not very much concerned about privacy regulation whereas is considered that a healthy competition will impose the minimal standards by it own.

The situation is largely different in Europe where the spotlight is on corporate world while the Government with its regulations is perceived as savior of privacy. The privacy become in Europe a right, often broadly regulated on the base of The European Union Directive on Data Protection of 1995. This instrument commanded that each EU nation elaborate a national privacy law and create a Data Protection Authority to protect citizens' privacy and investigate attacks against it.

b) The privacy and the trade-offs with personalized experience: the case of Facebook

Privacy measures were designed by social networking sites to provide users a protection of their personal information. The available settings on Facebook include for example the ability to block certain individuals from seeing one's profile, the ability to choose one's friends and the ability to limit who has access to one's pictures and videos. It is the user's prerogative to apply such settings when providing personal information on the Internet.

Some incidents relating to privacy concerns in Facebook[12] highlight the intensification of conflict between privacy advocates, who perceive it as a fundamental right of citizens, and the Web 2.0 business, to whom the marketing of these data to third parties is a potential gold mine capable of compensating the huge costs involved.

The privacy discussion had an even bigger significance in front of new Facebook package that links, for the first time, huge and highly structured social environment with objects and items. With appropriate progress in algorithms, semantic applications and alike the collaborative filtering technique could provide here as mentioned above a 'personalized experience' never seen before.

There is a general trade-off implied at this point: too much privacy (right 'to be left alone') in this context implies almost no 'personalized experience' (personalized recommendation or research). In other words one had to leave freely his personal data as a condition to enjoy a maximum personal experience as a consumer or searcher. As a result of the trade-off, in a future word (highly praising 'personalized experiences') the privacy concern should be left aside. It is an interesting dialectic between personal data (privacy as right to be let alone) and the need to give personal information required by deep social interaction. Being part of the crowd and being as 'transparent' as possible means one can be fulfilled as an individuality as soon as the 'personalized experience' will be stretched to the limit. From the point of view of 'profiling' a no privacy world associated with IT means will assure the deepest 'personal experience'.

c) The limits of privacy in social media

Maybe this is the significance of the statement made by Mark Zuckerberg and announcing that the era of privacy was gone [13] and that we are living in times worried only by those who have something to hide.

Maybe we are not that far away from the 'transparent society 'described by David Brin [14]. The author warned that the biggest threat to our freedom, lay in the surveillance technology being used by few people, rather than too many. He considers that the true privacy will be lost in this 'transparent society' of tomorrow. Hence it would be better for the society as a whole that the

surveillance will be equal for all, as the public should have the same information access as those in power (everybody should and will watch everybody else regardless of social or political status). Here the context seem to be quite different (a trade-offs between privacy and the 'personalized experience') but the result is the same: 'social transparency'.

We can in addition agree with some authors [15] that classical protection by means of privacy protection attended a different limit, a physical or structural limit. Facebook's users can control their privacy settings and choose the highest protection. But there are still other concerns. For example beyond the age, sex or place of life, the real value that Facebook seeks to exploit can be find in the simple network structure composed by all the friends and groups to which user is registered. Then it becomes possible for a third person to measure and analyze these relations on the whole network and identify the precise profile of individuals (even if they have chosen the maximum privacy protection in Facebook).

In fact there is some inherent knowledge in Facebook's structure that allows retrieving information that crosses privacy barriers. The social relationships graph shows the location of individuals within their social network and group membership contextualizes this network by the interests of individuals. This is an expression of 'system effect' where the whole is more than the sum of its parts. Therefore this specific technological environment will deceive the privacy settings which become irrelevant.

3 Personalization and the overcome of privacy paradigm; some hypothetical remedies in the field of E-governance

No privacy (either as free choice or as an evolution of social expectations) would mean that social networking sites will acquire through high tech instruments a great amount of 'profiles', information that largely overcome that of each individual about himself. These are very valuable information in the knowledge economy of today. In a different context some time ago Facebook proposed to sell personal information it has gathered but stepped back after the vast discussions that emerged. The power acquired through Facebook's new package will create far more opportunities of this kind.

We can examine a hypothesis based on the actual situation. If users are giving freely all the information concerning themselves, no privacy restrictions are met. From that moment on Facebook can used them with the help of 'profiling' instruments and acquire a most valuable set of data. Facebook had to cover the expenses needed for the functioning of their site (research, development, maintenance and so on). They gain value by doing what they do the best (linking people). But the value of the great amount of profiles updated in real time would be very huge indeed and incommensurably with the costs.

Intuitively we can think here that it is a problem of justice. If the information statistically obtained (based on free work and free information provided by the users) might be retailed by social sites some new instruments (normative or economic) seems to be needed.

For Benavent [16] at stake is the right of controlling the data: *Not for the knowledge of a new type, but just because the data are the keys to a value that is pirated. Therefore at stake is the right to income. The only accusation that could be addressed to social media is to pirate the value of personal data. There is no problem of privacy in the sense that fundamental rights are violated, but there is a problem of economic control of personal data. The right, for which all must fight, is not the right to control the use of traces that one leave, but the right to income it can generate* (free translation from French).

One can identify here an expression of commutative justice. If one gives freely one's data it doesn't mean that one's profile (the aggregation of data obtained by a social networking site) doesn't have a certain value as soon as there is a market for that

(advertising market). From the perspective of commutative justice (exchange of 'do ut des'-'you give me I give you') there had to be also a value in the hand of the users which is transformed (and augmented accordingly) by social networking sites as Facebook. And wherever is value there are equally rights (and even ownership rights).

Benavent defends a 'sui generis' ownership over personal data. He considers that Facebook's explicit agreement with its users, indicating that the rights to use free his services in exchange of the statistical data they give up, is not unreasonable, nor immoral. But the informant (the user) must enjoy the same benefit that gives the journalists the right to protect their sources of information (vis-à-vis government of corporate world). This is what social media should respect, not just privacy, but a specific contract, one about statistical data.

We can see here that the privacy paradigm was largely overcome and a new kind of (normative) remedies was found. These remedies seem to be an adaptation of ancient rights used in a different context. The author proposed an ownership on data, which he qualified as copyright (over statistical data resulting from a method accomplished by Facebook). In fact he needed to create an ownership right over statistical data resulted from processing individual elements (not submitted by themselves to the same ownership protection). It is not the normal property right here but a kind of immaterial property, something keen to 'copyright'. It is the reason for the author to qualify them this way (with some of their special attributes discussed above).

We can notice at this point that a normative intervention in the sphere of E-governance would be necessary to answer the problem discovered by an analysis from commutative justice point of view.

In the same vein there is an almost identical solution proposed by Garfinkel [17]. In the context of behavioral tracking software (first level of personalized experience seen in first pages of this paper) the author proposed a kind of 'compilation copyright'.

The American law identifies three distinct elements which should be met for a work to qualify as a 'copyrightable compilation': the collection of and assembly of pre-existing material, facts, or data; the selection, coordination, or arrangement of those materials; and the creation, by virtue of the particular selection, coordination, or arrangement of an original work of authorship. But collecting and assembling facts and information is not enough. Compilations, just as any other work, may only be copyrighted if the originality requirement, 'an original work of authorship', is met.

Garfinkel is proposing an adaptation of this concept: *"...This copyright protects newspapers, compact discs, and other sorts of information-rich media from illegal copying even when the individual items they contain are not subject to copyright protection. The doctrine of compilation copyright could be extended to cover individual components of a person's life. You might not have copyright protection on each sentence you say, each product you buy, or the names of each of the streets you've lived on since you were born. But when these facts are assembled into a whole, they might be held to be an unacceptable appropriation of your mortal essence".*

'Mutatis mutandis' this solution can be adopted in the case of statistically determined data (profiling or personalization through aggregation).

As a matter of fact this solution is very akin to the former one of Benavent. In the context of our discussion it can be seen as copyright of a user over (statistically) data gathered by Facebook with its software. The user doesn't have a copyright of his initial data but over the aggregation (the building of a 'profile') based on his data.

This is another example of adapting a copyright concept to cover new technological evolutions. It is not the first time that such a development imposes a copyright adjustment (in the same way the derivate copyrights were elaborated with regard to broadcasting, interpretation, databases, etc). This 'sui

generis' copyright allows a certain mastery of the user over his statistically determined data with an associated legal regime (relating to selling, copying, licensing and so on). Once again the intervention of the legislator (an E-governance intervention) seems to be necessary.

4 Conclusion
The paper offered some basic hints, a first approach of the challenges the new social media and social transparency tools are posing to current normative framework regarding privacy constrains. The paper examined also some ideas about possible normative remedies (as 'sui generis' copyrights) to challenges overcoming privacy borders. The research shall be developed in the future through an analysis and a deconstruction of copyright regimes. Only this kind of future analysis may find a clear answer to some of the challenges depicted above. It is a highly exigent topic but equally a most interesting one.

References
[1] "Long tail" *Wikipedia*. Available: http://en.wikipedia.org/wiki/Long_Tail.
[2] C, Anderson, "The Long Tail", *Wired*, October 2004.
[3] "Behavioral targeting" *Wikipedia*. Available at: http://en.wikipedia.org/wiki/Behavioral_t argeting.
[4] S. Levy,"How Google's Algorithm Rules the Web", *Wired*, March 2010.
[5] Simson Garfinkel, *Database Nation. The Death of Privacy in the 21st Century*, O'Reilly & Associates, 2000, p 308-315.
[6] G Adomavicius; A.Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering 17 (6): June 2005, p. 734–749. Available at: http://portal.acm. org/citation.cfm?id=1070611.1070751.

[7] L. Grossman, "How Computers Know What We Want — Before We Do", *Time,* June 7, 2010, p. 44.
[8] S. Muhammad, *Collaborative Filtering: Lifeblood of The Social Web*, June 30, Available: http://www.readwriteweb.com/archives/c ollaborative_filtering_social_web.php.
[9] Video available at: http://apps.facebook.com/feightlive/ on Facebook.
[10] C. Benavent, *Les effets limités des médias sociaux*. 23 April 2010. Available: http://i-marketing. blogspot.com/2010/04/les-effets-limites-des-medias-sociaux.html.
[11] J. Thomas "Is it safe to talk,yet?The evolution of electronic privacy law", *Computer Security, Privacy, and Politics: Current Issues, Challenges, and Solutions*, Ramesh Subramanian (ed), IRM Press, 2008, p. 23-29.
[12] B Stone, "Facebook to Settle Thorny Lawsuit Over Its Origins". *The New York Times* (blog). April 7, 2008. Available: http://bits.blogs.nytimes.com/2008/04/07 /facebook-to-settle-thorny-lawsuit-over-its-origins/.
[13] M. Kirkpatrick, *Facebook's Zuckerberg says the age of privacy is over*, January 9, 2010, Available at: http://www.readwriteweb.com/archives/f acebooks_zuckerberg_says_the_age_of_ privacy_is_ov.php.
[14] D. Brin, *The transparent society,* Reading, MA Perseus Books, 1998.
[15] G.Fouetillou, "Facebook:de la nécessité de protéger ses données 'relationnelles'", *Lemonde.fr*, April 22, 2010.
[16] C Benavent, *Les effets limités des médias sociaux.* Available: http://i-marketing.blogspot.com/2010/04/les-effets-limites-des-medias-sociaux.html, 23 April, 2010.
[17] S. Garfinkel *Database nation: The Death of privacy in the 21st Century*, O'Reilly & Associates, 2000, p 315-316.

Remus TITIRIGA holds a PhD in European Law from the University Nancy 2, France. He has a Masters degree in European Law from the University Nancy 2, France. He also received a special degree in Comparative Law from the International Faculty of Comparative Law of Strasbourg. He holds degrees in Law (Honours) from the University of Timisoara. From 2006 to 2008 he was Teaching Assistant of Constitutional Law and European Institutional Law at the University Nancy 2 in France. He also holds a Bachelor and a Master of Applied Science degree in Computer Engineering. Remus Titiriga was a visiting fellow in Law at KU Leuven Belgium specializing in Law of Information and Communications Technology (ICT). He was admitted as Barrister on probation in October 1998 at the Bar Association of Timisoara, Romania and practiced as a civil and penal law litigator until October 1999. Remus has a number of publications and has presented papers at international or national conferences about topics like Computers, European Law and ICT Law.