ANSWER QUESTION PRIVACY 2017

1. Discuss the motivations in favour and against the reasoning "*I do have nothing to hide so I do not care about privacy*". In your discussion refer to the current technology advancements and the economic value of private information.

With the expression "I do have nothing to hide so I do not care about privacy" we usually refer to the behaviour adopted by some people that are based on a carelessness in providing their own personal information due to the fact that they are not involved in malicious facts. Sometimes the nothing to hide argument can be posed as a question: "If you have nothing to hide, then what do you have to fear?"

This kind of reasoning leads to a spread of personal information. Now, the key point is the way in which these information are used. The main advantages that derive from providing personal information can be, for example:

- *Personalized services* as a result of algorithms' computations that elaborate personal information.
- *Public national and international security*, based on surveillance and control. This can be performed at hardware level using closed circuit cameras or drones and at software level using advanced data mining techniques applied to call records, data gathered from internet, credit card records and so on. We can think about the NSA program that has taken place after the 11th September, when the Bush Administration has secretly allowed to engage in warrantless wiretapping of American citizens' telephone calls.

We can see the problem from another point of view: the fact that we have nothing to hide does not imply that some information shouldn't be remain secret. For example you wouldn't show your embarrassing (but not malicious) picture to a job recruiter. Therefore, by performing a deeper analysis of the problem we can understand that there are also disadvantages if we don't manage our privacy in an appropriate manner (also in the case we have nothing to hide). In fact the real problem is not protecting people from true but protect people from damaging conclusions drawn from misunderstood informations.

In conclusion, the nothing to hide arguments if related to terrorism prevention is very effective but focusing on other aspects of privacy, the surveillance can create chilling effects on free speech and free association.

Commento [DM1]: Non mi interessa la privacy

Commento [DM2]: Il comportamento adottato

Commento [DM3]: Trascuratezza nel fornire

Commento [DM4]: Dato che

Commento [DM5]: Questo tipo di ragionamento porta a una diffusione

Commento [DM6]: Fornire i dati personali

Commento [DM7]: Riuniti/trovati/messi insieme da

Commento [DM8]: Di impegnarsi in intercettazioni senza garanzia delle chiamate telefoniche dei cittadini americani

Commento [DM9]: Non implica che alcune informazioni non debbano rimanere segrete

Commento [DM10]: Eseguendo un'analisi approfondita del problema

Commento [DM11]: Infatti il vero problema non è proteggere le persone dalla verità ma proteggere le persone da conclusioni dannose tratte da informazioni fraintese.

Commento [DM12]: la sorveglianza può creare effetti frustranti sulla libertà di espressione e sulla libera associazione.

2. Discuss how people might release sensible data in exchange of an immediate advantage. Discuss examples that give evidence to your claims. Possibility gives a specific example in which you or a person you know was involved.

People often provide, without worrying too much, a lot of personal information in exchange of some services and/or immediate advantage. For example:

- Some restaurants require personal information in order to use public Wi-Fi (ex. Mc Donalds). These data are collected by the managers of the company.
- Some television companies require personal information (through registration) in order to let you stream legally from your computer (ex Mediaset). These data are collected and used to perform analysis (marketing purposes).
- Some websites require personal information in order to notify you promotional offers and to get advanced functionalities (ex. Groupon, Amazon).
- Releasing sensitive or personal data just for take part in a competition with prize.

I had a similar experience when I joined in a soccer tournament. They asked me some information as telephone number for promotional information about his future tournaments. In my case it was useful because I always intend to try to sign up for a new tournament with my team.

3. When do we claim that a social network has been deanonymized? Clearly it is not necessary that all data in the network have been deanonymized. Provide the different definitions that are used to claim that a network has been (possibly partially) deanonymised? Be specific and provide at least one example.

Operators of online social networks are increasingly sharing potentially sensitive information about users and their relationships with advertisers, application developers, and datamining researchers. To alleviate privacy concerns, the networks are anonymized, i.e., names and demographic information associated with individual nodes are suppressed. But despite there is the possibility to grab information about specific user with deanonymization processes.
In general, in order to deanonymize something the goal is: fixing some target record "r" in the original dataset learn as much about "r" as possible.
A network is partially or totally deanonymized when the attacker, starting from a reidentification of a fraction of nodes from an auxiliary network that partially overlaps with the original network, gains knowledge about the edges of the graph that represents the network. So in a more technical way determine a 1-1 mapping between two graphs.
As example we can consider the partial deanonymization performed on twitter graph passing through Flickr graph. The two companies both exposes: Mandatory username, optional name, optional location. As result researchers obtained: 27,000 mappings. Starting from a seed of 150 pairs of randomly selected mappings with the constraint that the degree of each mapped node in the auxiliary graph is at least 80 the result is: 30.8% of the mappings were re-identified correctly, 12.1% were identified incorrectly, and 57% were not identified.

"An example is for example deanonymization of Netflix users by using IMDB dataset, that is the auxiliary network."

Commento [DM13]: Le persone potrebbero rilasciare dati sensibili in cambio

Commento [DM14]: per farvi streaming legalmente

Commento [DM15]: Quando affermiamo che una rete sociale è stata deanonata?

Commento [DM16]: E' una strategia in data mining in cui i dati anonimi vengono "incrociati" con altre fonti di dati per ri- identificare l'origine dati anonima.

Commento [DM17]: Stanno diffondendo sempre più informazioni potenzialmente sensibili

Commento [DM18]: Per alleviare le preoccupazioni sulla privacy

Commento [DM19]: Per afferrare le informazioni

Commento [DM20]: Fissare/decidere alcuni record r nell'originale dataset e imparare/trovare informazioni il più possibile rispetto a questi

Commento [DM21]: Una rete sociale è parzialmente o completamente deanonimizzata quando un attaccante, partendo da l'identificazione di una frazione dei nodi della rete ausiliaria che parzialmente si sovrappone con quella originale, guadagno conoscenza sui lati del grafo che rappresenta la rete. In maniera tecnica, determina una mappatura 1-1 su entrambi i grafi

4. Discuss the techniques that an active adversary that is a member of the social network can use to deanonymize the network. Discuss how he/she can exploit the graph structure of the network to perform its attack.

An active adversary that is a member of a social network could de-anonymize the network itself in the following way: he may create a fake nodes or he has got already a real nodes in the graph, than he adds edges in the social network graph with features that easily help him to recognize himself in the anonymized version of graph, S. From an attacker point of view, he usually have access to different network, S_{aux}, whose structure membership's partially or complete overlaps with S. If not, it might be extracted from S (automatic crawling, malicious third-party application).

~~S_{aux}: a graph $G_{aux} = V_{aux}, E_{aux}$ and a set of probability distributions.~~

The process (deanonymize users of S using an auxiliary dataset S_{aux}) is done through re-identification algorithm. This algorithm is composed by two steps:

• **Seed identification:** identify a small number of seed nodes in both target graph and auxiliary graph, map them to each other. To find this mapping, assume a clique structure in an auxiliary graph, search the same clique on target graph, using common neighbours, in-degree and out-degree of a node. At the end connect the similar nodes. The output is a partial mapping.

• **Propagation**: attacker builds a deterministic 1-1 mapping between the two datasets. Each iteration starts with the accumulated mapping, then an unmapped node u in V1 is picked and a score for each unmapped node v in V2 is computed.

As example, we can present the de-anonymization of Twitter graph performed using Flickr graph. Running the algorithm proposed as results we have 27, 000 mappings, due to the overlapping between the two graphs.

5. Discuss two example on how it is possible to deanonymize a social networks by an adversary that receives an anonymized version of the network (to which he/she does not belongt to). With reference to the examples presented in class discusses the reasons why such approaches are successfully.

There are two main famous examples of this kind of deanonymization, one is the Netflix deanonymization attack, and the second one was about Mrs. Arnold on AOL.

Due to a contest for recommendation system, Netflix released a version of the database of ratings, with some level of anonymization (names removed, perturbation of information). Combined with background knowledge, which was IMDB, an attacker can perform a deanonymization attack with usage of another, similar, database.

In details, the objective was to: Fix some target record of original Netflix dataset and Try to learn as much about this record as possible. But background knowledge (IMDB dataset) was noisy, and Netflix dataset was perturbed (with only a sample of records released). Anyway, since ratings about not top100 movies are very personalized (is unusual for two users give same rating on same not so known movies), the researchers in this project found out that, with this cross references on movie ratings and date of ratings, some users turned out to be members of both IMDB and Netflix (with some personal informations voluntarily released on IMDB), and personal information was obtained with a very low percentage of error (in the experiment, just 4 ratings, in mean, were enough to uniquely identify the user).

Commento [DM22]: Può sfruttare la struttura del grafico

Commento [DM23]: La cui struttura parziale o completa si sovrappone a S ["Ovvero sono simili"]

Commento [DM24]: Scansionando in maniera automatica, applicazione malevola di una terza parte

Commento [DM25]: Mappatura accumulata/già trovata-effettuata ecc.

Commento [DM26]: è scelto

Commento [DM27]: Due esempi di applicazioni (attiva e passiva) o due esempi di casi avvenuti (netflix e AOL)? Di seguito riporto entrambi … da capire quale si intende

Commento [DM28]: A causa di un concorso per il sistema di raccomandazione Netflix ha rilasciato una versione del database delle valutazioni, con un certo livello di anonimizzazione (nomi rimossi, perturbazioni di informazioni).

Commento [DM29]: Comunque, dal momento che i voti di circa 100 film non sono molto personalizzati (è inusuale che due utenti danno la stessa valutazione a film non noti), i ricercatori di questo progetto hanno scoperto che, con questi riferimenti incrociati sui filmati e la data dei voti, alcuni utenti si sono rivelati membri di IMDB e Netflix (con alcune informazioni personali rilasciate volontariamente su IMDB) e le informazioni personali sono state ottenute con una percentuale di errore molto bassa (nell'esperimento, solo 4 valutazioni, in media, erano sufficienti a identificare in modo univoco l'utente).

Another famous example occurred when AOL (America on Line – è un Internet Service Provider ISP) published anonymized search logs for research purposes, and one of the anonymized users was repeatedly searching for:

-Info about a small city in USA (Lilburn, CA)
-"60 single men"
-Several people with last name "Arnold"
-Numb fingers
-Dog behaviour.

Since in Lilburn area there were 14 citizens with last name Arnold. A researcher from New York Times, by contacting each one of these 14 people, managed to reach that anonymized user, a 62 y.o. lady

---

It is possible to have two different types of attack aimed to de-anonymize a social network graph: an active attack and a passive one. The first type of attack is feasible even if the attacker is not part of the graph structure of the network, therefore he/she does not belong to it.
In active attack, the adversary chooses an arbitrary set of users whose privacy it wishes to violate, then it creates a feasible number of new user accounts with edges pointing to these targeted users, and finally creates a pattern of links among the new accounts with the goal of making it stand out in the anonymized graph structure.
The adversary then efficiently finds these new accounts together with the targeted users in the anonymized network that is released. At a theoretical level, the creation of O(p log n) nodes by the attacker in an n-node network can begin compromising the privacy of arbitrary targeted nodes, with high probability for any network. Experimental evidence suggests that it may be very difficult to determine whether a social network has been compromised by such an active attack.

Commento [DM30]: fattibile

Commento [DM31]: dei quali intende violare la privacy

Commento [DM32]: L'obiettivo di farlo apparire nella struttura del grafico anonimo.

Commento [DM33]: Le prove sperimentali suggeriscono che può essere molto difficile determinare se un network sociale è stato compromesso da un attacco così attivo

Two examples of active attacks are:

- Walk-based attack.
- Cut-based attack.

For the first attack, we show that with k = O(log n) new accounts, a randomly generated sub-graph H will be unique with high probability, regardless of what G looks like and regardless of how H is attached to the rest of G. Moreover, if the maximum node degree in H is O(log n), then H is recoverable efficiently, together with the identities of up to b = O(log2 n) targeted nodes to whom the attacker created links from H. The recovery algorithm uses a search over short walks in the anonymized graph G, and accordingly we call it the walk-based attack.
The second active attack is similar in flavour; it also constructs H by including edges at random, but it attaches H to G using very few edges and recovers it using a more complex computation based on Gomory-Hu cut trees. Hence we will refer to it as the cut-based attack.
The walk-based attack comes with an extremely fast recovery algorithm that easily scales to millions of nodes, and it appears to be very hard to detect. The cut-based attack has the advantage of matching the tight theoretical bound on the number of nodes needed, it is possible to show that an attacker must create at least (p log n) new nodes in the worst case to begin compromising the privacy of arbitrary targeted nodes. The use of Gomory-Hu trees in the cut-based attack makes its recovery algorithm more expensive than that of the walk-based attack. Finally, the walk-based attack has the potential to compromise O(k2) users,
while the cut-based attack can only compromise O(k), and it also appears easier to detect that the cut-based attack has taken place.

Commento [DM34]: Indipendentemente da ciò che G sembra e indipendentemente da come H sia attaccato al resto di G

6. Explain what is k-anonymity of a data base and provide one example where a k-anonymous data base does not leak any information (i.e. it preserves privacy) and one in which a k-anonymous data base can leak information.

K-anonymity is a property of dataset containing info about people, which does not allow attackers to track a specific user that have put his own data inside the dataset. A dataset is said to have the k-anonymity property if the information of each person inside the dataset is not distinguishable at least from k-1 other people whose information also appears in the dataset.
A quasi-identifier ($\mathbf{QT_{RT}}$) is a set of information that could be used for linking with external information or for unique identifying individuals. They are the attributes available to an adversary.
The goal of k-anonymity is to release person-specific data such that the ability to link to other information using the quasi-identifier is limited.

There are two common methods for achieving k-anonymity for some value of k:

- **Suppression**: some values of the attributes are replaced by an asterisk '*'.
- **Generalization**: individual values of attributes are replaced by a broader category.

Examples

a k-anonymous data base does not leak any information (i.e. it preserves privacy)

ZIP code	Age	Sex	Disease	Name
47677	29	F	Headache	Alice
47905	43	M	Heart Disease	Bob
47602	22	F	Cancer	Lia
47678	27	F	Flu	Carmen
47918	48	M	Hurt arm	Ross

ZIP code	Age	Sex	Disease	Name
476**	2*	F	Headache	*
479**	4*	M	Heart Disease	*
476**	2*	F	Cancer	*
476**	2*	F	Flu	*
479**	4*	M	Hurt arm	*

The above table RT(ZIPcode, Age, Sex,Disease,Name) achieves **2-anonymity** by using the suppression technique, by selecting as quasi-identifiers $\mathbf{QT_{RT}}$ **= (ZIPcode,Age,Name).**

a k-anonymous data base can leak information.

ZIP code	Age	Sex	Disease	Name
34782	23	M	Cancer	Bob
34795	28	F	Flu	Alice
34785	29	M	Headache	Mark
45832	52	M	Heart disease	Lucas
47918	48	M	Hurt arm	Ross

ZIP code	Age	Sex	Disease	Name
347**	2*	M	Cancer	*
347**	2*	F	Flu	*
347**	2*	M	Headache	*
458**	5*	M	Heart disease	*
479**	4*	M	Hurt arm	*

The above table RT(ZIPcode, Age, Sex,Disease,Name) **does not achieve k-anonymity**, by selecting as quasi-identifiers $\mathbf{QT_{RT}}$ **= (ZIPcode, Age,Name).** An attacker can easily identify both Alice and Lucas.

Commento [DM35]: non perde alcuna informazione

Commento [DM36]: di tracciare, o meglio sarebbe definire/identificare

Commento [DM37]: Per rilasciare dati specifici per la persona in modo tale che la capacità di collegarsi ad altre informazioni utilizzando il quasi identificatore è limitata.

Commento [DM38]: Il raggiungimento

7. Consider the following quasi-identifiers. Does Dataset1, Dataset2, Dataset3 satisfy k-anonymity?
If so: what is the maximal k for which it satisfies k-anonymity? Explain your answer.

Dataset$_1$

Age	Gender	Fav.Show
19-25	female	Friends!
19-25	male	Friends!
19-25	male	Friends!
12-15	female	Friends!
19-25	male	G.o.T.
19-25	female	G.o.T.
19-25	male	G.o.T.

Dataset$_2$

Age	Gender	Fav.Show
19-25	female	Grey's A.
19-25	female	Simpsons
19-25	female	Futurama
19-25	female	Friends!
19-25	female	G.o.T.
19-25	female	C.Minds
19-25	female	Br.Ba.

Dataset$_3$

Age	Gender	Fav.Show
19	female	Friends!
19	male	Friends!
19	male	Friends!
19	female	Friends!
20	male	G.o.T.
20	male	G.o.T.
20	male	G.o.T.

CASE: Quasi-Identifier = Age,Gender, Fav.Show
Dataset1 doesn't satisfy k-anonymity. There is only one "12-25, female, Friends!" tuple.
Dataset2 doesn't satisfy k-anonymity. All tuples have a different value for "Fav.Show".
Dataset3 satisfies k-anonymity and k=2.

CASE: Quasi-Identifier = Age,Gender. Sensible data = Fav.Show
Dataset1 doesn't satisfy k-anonimity. There is only one "12-25, female, Friends!" tuple.
Dataset2 satisfies 7-anonimity.
Dataset3 satisfies 2-anonimity.

8. K-anonymity is not sufficient for maintaining anonymity. Provide other formulations that enforce the concept. Discuss the approach and provide one example that shows that this approach might be better than K-anonymity; show one negative example that shows that this approach is not adequate.

k-Anonymity is not sufficient for maintaining anonymity for different reasons:
- Generalization: is difficult to do in database of huge attribute dimensions (Curse of dimensionality), and reducing the number of characteristics let the dataset useless.
- Anonymity is not sufficient if:
  - -If the rows of dataset are not changed.
  - -If the elements of the same equivalence class are all the same (homogeneity attack).
  - -If the attacker has background knowledge, with which it can reduce the candidate set of possible targets

Thanks to this example we can show that, even if k-anonymity is achieved, an attacker can easily find sensitive information about a user.

Commento [DM39]: Con attributi di enormi dimensioni
Curse of dimensionality → Maledizione della dimensionalità

Commento [DM40]: Riducendo il numero delle caratteristiche si rende il db inutilizzabile

ZIP code	Age	Sex	Disease	Name
47677	29	F	Cancer	Alice
47905	43	M	Heart Disease	Bob
47602	22	F	Cancer	Lia
47678	27	F	Cancer	Carmen
47918	48	M	Hurt arm	Ross

ZIP code	Age	Sex	Disease	Name
476**	2*	F	Cancer	*
479**	4*	M	Heart Disease	*
476**	2*	F	Cancer	*
476**	2*	F	Cancer	*
479**	4*	M	Hurt arm	*

We can notice that 2-anonymity is achieved with the quasi-identifier (ZIP Code, Age, Sex). If the attacker knows that Alice is in the dataset, he can learn her immediately that she has cancer, since all females have the same disease.

A better approach is to following **the UK confidentiality** guidance that shows us how to reach a good level of anonymization. There are three main methods that we have to apply on the dataset:
1. Table redesign: is recommended as a simple method that will minimise the number of unsafe cells and preserve original counts;
2. Modify cell values: as Cell suppression, rounding (involves adjusting the values in all cells in a table to a specified base.) or barnardisation(cells of every table are adjusted by +1, 0 or -1, according to probabilities)
3. Adjust the data: Swap pairs of records within a micro-dataset that are partially matched to alter the geographic locations attached to the records but leave all other aspects unchanged.
Thanks to this method we can modify the above example in the following one:

Sex	Cancer	Heart disease
M	0	2
F	3	0

Now, it is possible to notice that the table does not leak any information about the clients because we have grouped them into similar categories, in such a way that the table does not present similar records any more. The value in each cell of the table represents the number of clients having a certain disease.

To enforce the concept of privacy we can introduce l-diversity and t- closeness.

-L-diversity states that each equivalent class must have at least l sensitive attributes. But it has a problem: It does not consider the semantics of the sensitive attributes

-T-closeness states that the distribution of sensitive values within each equivalence class should be similar to the distribution of the original dataset

9. Consider the following tables

Name	Gender	Age	City of birth	Favorite TV Series	Relationship Status
*	male	19-25	Saarbrücken	Game of Thrones	single
*	female	16-18	Trier	Big Bang Theory	in relationship
*	male	12-15	München	Big Bang Theory	in relationship
*	female	19-25	Berlin	Game of Thrones	in relationship
*	female	19-25	Hamburg	Big Bang Theory	single
*	female	19-25	Saarbrücken	Game of Thrones	single
*	male	16-18	Trier	Game of Thrones	single
*	female	12-15	München	Big Bang Theory	in relationship
*	male	19-25	Berlin	Big Bang Theory	single

Name	Email	TV Show	Rating (1=bad, 5=great)
Alice	alice1993@email.com	Friends!	1
Bob	bobbybob@email.com	Friends!	3
Charlie	s9charchar@email.com	Friends!	2
Eve	evelyn@myhighscool.com	Friends!	1
Bob	bobbybob@email.com	Game of Thrones	5
Alice	alice1993@email.com	Game of Thrones	2
Charlie	s9charchar@email.com	Game of Thrones	2
Bob	bobbybob@email.com	Big Bang Theory	5
Charlie	s9charchar@email.com	Big Bang Theory	1
Alice	alice1993@email.com	Big Bang Theory	5
Eve	evelyn@myhighscool.com	Big Bang Theory	5

Commento [DM41]: È consigliato come metodo semplice **che ridurrà al minimo il numero di celle non sicure** e conserva i conteggi originali;

Commento [DM42]: Come soppressione delle cellule, arrotondamento (coinvolge la regolazione dei valori in tutte le celle in una tabella a una base specificata.) O barnardizzazione (le cellule di ogni tabella sono regolate da +1, 0 o -1, secondo le probabilità)

Commento [DM43]: Scambia coppie di record in un micro database parzialmente corrispondente per alterare le posizioni geografiche collegate ai record ma lasciare invariati tutti gli altri aspetti.

Commento [DM44]: Afferma che ogni classe equivalente deve avere almeno l attributi sensibili

Commento [DM45]: Afferma che la distribuzione dei valori sensibili all'interno di ciascuna classe di equivalenza dovrebbe essere simile alla distribuzione del set di dati originale

Assume you know that Alice's data is within the released dataset. Where is she most likely born and what is most likely her relationship status? Describe how you deanonymized her.
Can you deanonymize Charlie as well? If so: describe how. If not: describe why.

In the first, I say that Alice is a female (from the name). For this reasons, the tuple which I have to consider are:

Name	Gender	Age	City of birth	Favorite TV Series	Relationship Status
*	male	19-25	Saarbrücken	Game of Thrones	single
*	female	16-18	Trier	Big Bang Theory	in relationship
*	male	12-15	München	Big Bang Theory	in relationship
*	female	19-25	Berlin	Game of Thrones	in relationship
*	female	19-25	Hamburg	Big Bang Theory	single
*	female	19-25	Saarbrücken	Game of Thrones	single
*	male	16-18	Trier	Game of Thrones	single
*	female	12-15	München	Big Bang Theory	in relationship
*	male	19-25	Berlin	Big Bang Theory	single

In the second table, there is an email of Alice. So, the email is: alice1993@email.com. From this, I can derive that Alice is between 19 and 25 y.o. (that is a unique quasi identifier). So, the tuple which I have to consider are:

Name	Gender	Age	City of birth	Favorite TV Series	Relationship Status
*	male	19-25	Saarbrücken	Game of Thrones	single
*	female	16-18	Trier	Big Bang Theory	in relationship
*	male	12-15	München	Big Bang Theory	in relationship
*	female	19-25	Berlin	Game of Thrones	in relationship
*	female	19-25	Hamburg	Big Bang Theory	single
*	female	19-25	Saarbrücken	Game of Thrones	single
*	male	16-18	Trier	Game of Thrones	single
*	female	12-15	München	Big Bang Theory	in relationship
*	male	19-25	Berlin	Big Bang Theory	single

Then, the second table tells me that the TV series with the highest rating for Alice is "Big Bang Theory"

Name	Email	TV Show	Rating (1=bad, 5=great)
Alice	alice1993@email.com	Friends!	1
Bob	bobbybob@email.com	Friends!	3
Charlie	s9charchar@email.com	Friends!	2
Eve	evelyn@myhighscool.com	Friends!	1
Bob	bobbybob@email.com	Game of Thrones	5
Alice	alice1993@email.com	Game of Thrones	2
Charlie	s9charchar@email.com	Game of Thrones	2
Bob	bobbybob@email.com	Big Bang Theory	5
Charlie	s9charchar@email.com	Big Bang Theory	1
Alice	alice1993@email.com	Big Bang Theory	5
Eve	evelyn@myhighscool.com	Big Bang Theory	5

So, in the first table, Alice is:

Name	Gender	Age	City of birth	Favorite TV Series	Relationship Status
*	male	19-25	Saarbrücken	Game of Thrones	single
*	female	16-18	Trier	Big Bang Theory	in relationship
*	male	12-15	München	Big Bang Theory	in relationship
*	female	19-25	Berlin	Game of Thrones	in relationship
*	female	19-25	Hamburg	Big Bang Theory	single
*	female	19-25	Saarbrücken	Game of Thrones	single
*	male	16-18	Trier	Game of Thrones	single
*	female	12-15	München	Big Bang Theory	in relationship
*	male	19-25	Berlin	Big Bang Theory	single

She is born in Hamburg and She is Single!

In the first, I say that Charlie is a male (from the name). For this reasons, the tuple which I have to consider are:

Name	Gender	Age	City of birth	Favorite TV Series	Relationship Status
*	male	19-25	Saarbrücken	Game of Thrones	single
*	female	16-18	Trier	Big Bang Theory	in relationship
*	male	12-15	München	Big Bang Theory	in relationship
*	female	19-25	Berlin	Game of Thrones	in relationship
*	female	19-25	Hamburg	Big Bang Theory	single
*	female	19-25	Saarbrücken	Game of Thrones	single
*	male	16-18	Trier	Game of Thrones	single
*	female	12-15	München	Big Bang Theory	in relationship
*	male	19-25	Berlin	Big Bang Theory	single

From the second table, I cannot deduce important information about Charlie.

Name	Email	TV Show	Rating (1=bad, 5=great)
Alice	alice1993@email.com	Friends!	1
Bob	bobbybob@email.com	Friends!	3
Charlie	s9charchar@email.com	Friends!	2
Eve	evelyn@myhighscool.com	Friends!	1
Bob	bobbybob@email.com	Game of Thrones	5
Alice	alice1993@email.com	Game of Thrones	2
Charlie	s9charchar@email.com	Game of Thrones	2
Bob	bobbybob@email.com	Big Bang Theory	5
Charlie	s9charchar@email.com	Big Bang Theory	1
Alice	alice1993@email.com	Big Bang Theory	5
Eve	evelyn@myhighscool.com	Big Bang Theory	5

For this reasons, I cannot deanonymize Charlie!

10. Present and discuss the technique known as input perturbation in accessing a data base; discuss its limitations.

Input perturbation is a technique of Data Sanitization (Data Sanitization is the process of disguising sensitive information in databases by overwriting it with realistic looking but false data of a similar type) consists of applying some function R that will modify the data with perturbed entries.

Commento [DM46]: Nascondere informazioni riservate

Commento [DM47]: Con dati realistici simili ma falsi

The function R (***y = R(x)***, where y is the perturbed data and x the original data) that perform the perturbation has to be publicly known, in order to give an idea about how much the dataset is perturbed. In this way, a researcher that needs to work with almost real data won't use a high perturbed dataset, while another researcher could use it if the value of the perturbation is not so important for his study

Perturbation must be done such that the KL distance between two distributions is small but not too much, where KL distance is defined as

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

This because if the distance is small distributions are very similar and can leak information, very different otherwise. This can be a limitation because if we make a low perturbation an attacker

can easily understand data, but if we make a high perturbation data can change too much and the information of the database could become useless for initial purposes.
The limitation of the input perturbation technique is also come from the function R used to perturb the data. Since it must be public, some background knowledge or other info can help an attacker to recognize some people in the dataset.
For example:

• If the perturbation of the age is a function f(x) = [x+ε], (ε = [−20, +20]);
• There is a limit about the possible values of the attribute age, which goes from 0 to 90;
• There is an entry in the perturbed dataset with age # 91.

If an attacker knows the limitation about the attribute age, he can easily recognize that the entry with age 110 is in reality 90.

11.Given the following table modify it to verify 2-anonimity

Name	Age	Gender	State of domicile	Religion	Disease
Ramsha	29	Female	Tamil Nadu	Hindu	Cancer
Yadu	24	Female	Kerala	Hindu	Viral infection
Salima	28	Female	Tamil Nadu	Muslim	TB
sunny	27	Male	Karnataka	Parsi	No illness
Joan	24	Female	Kerala	Christian	Heart-related
Bahuksana	23	Male	Karnataka	Buddhist	TB
Rambha	19	Male	Kerala	Hindu	Cancer
Kishor	29	Male	Karnataka	Hindu	Heart-related
Johnson	17	Male	Kerala	Christian	Heart-related
John	19	Male	Kerala	Christian	Viral infection

There are 6 attributes and 10 records in this data. There are two common methods for achieving k -anonymity for some value of k:

- **Suppression** : In this method, certain values of the attributes are replaced by an asterisk '*'. All or some values of a column may be replaced by '*'. In the anonymized table below, we have replaced all the values in the 'Name' attribute and all the values in the 'Religion' attribute with a '*'.
- **Generalization** : In this method, individual values of attributes are replaced by with a broader category. For example, the value '19' of the attribute 'Age' may be replaced by ' ≤ 20', the value '23' by '19 < Age ≤ 30' , etc.

The next table shows the anonymized database.

Name	Age	Gender	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	Tamil Nadu	*	TB
*	20 < Age ≤ 30	Male	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male	Karnataka	*	TB
*	Age ≤ 20	Male	Kerala	*	Cancer
*	20 < Age ≤ 30	Male	Karnataka	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Viral infection

This data has 2-anonymity with respect to the attributes 'Age', 'Gender' and 'State of domicile' **since for any combination of these attributes found in any row of the table there are always at least 2 rows with those exact attributes**. The attributes available to an adversary are called "quasi-identifiers". Each "quasi-identifier" tuple occurs in at least k records for a dataset with k-anonymity.

Commento [DM48]: poiché per qualsiasi combinazione di questi attributi trovati in una qualsiasi riga della tabella ci sono sempre almeno 2 righe con quegli attributi esatti.

12. Present and discuss the technique known as output perturbation in accessing a data base; discuss its limitations.

Output perturbation is an example of Data Sanitization(Data Sanitization is the process of disguising sensitive information in databases by overwriting it with realistic looking but false data of a similar type) . Input and output perturbations are similar techniques used to sanitize a dataset. With the output perturbation, the db is not published, but it is possible to do query to it. Every answer of the query is perturbed before arriving to the guy who asked. So, the difference between the two operations lies in when the perturbation is performed. While the input perturbation is performed before the query is executed , the output it is performed after the query evaluation.

Commento [DM49]: Nascondere informazioni riservate

Commento [DM50]: Con dati realistici simili ma falsi

Commento [DM51]: Ma è possibile fare una domanda ad esso

Commento [DM52]: La differenza fra le due operazioni si trova / ce l'abbiamo quando viene eseguita la pertubazione.

As the input perturbation, it presents some weakness:
• Let n be the size of the database (total number of entries);
• If $O(n^2)$ perturbation applied, adversary can extract entire database after poly(n) queries;
• but even with $O(n^2 \rightarrow \log n)$ perturbation, it is unlikely that user can learn anything useful from the perturbed answers (too much noise).
So at the end the problem is that a weak output perturbation can allow an attacker to discover information after making a certain number of queries, but a strong output perturbation makes the data useless since there is too much noise.

Commento [DM53]: È improbabile che l'utente possa imparare qualcosa di utile dalle risposte perturbed (troppo rumore) SE QUESTA E' ELEVATA.

Commento [DM54]: Debole perturbazione dell'uscita

13. What is differential privacy and discuss its advantages and disadvantages with respect to other approaches.

In cryptography, differential privacy aims to provide means to maximize the accuracy of queries from statistical databases while minimizing the chances of identifying its records. Therefore the goal is to release statistical information without compromising the privacy of the individual respondent. Differential privacy ensures that the removal or addition of an item in the database does not affect the outcome of any analysis, so there are no risks by joining the database.

Commento [DM55]: La privacy differenziale mira a fornire mezzi per massimizzare l'accuratezza delle query dei db statistici riducendo al contempo le possibilità di identificare i propri record. + ACCURATEZZA - INFORMAZIONI

Commento [DM56]: Assicura che la rimozione o l'aggiunta di un elemento nel database non influenzino l'esito di un'analisi, quindi non ci sono rischi associandosi al database.

DEF: Given a real number ε , we say that a function K gives e-differential privacy, if for all datasets D1, D2 differing on at most one element and for all S belonging to range(K):

$$Pr[K(D_1) \in S] \leq e^{\varepsilon} \times Pr[K(D_2) \in S]$$

It means that if a participant add or remove his data to/from the dataset, no outputs would become significantly more or less likely.
(PRO)The powerful of differential privacy is that any event (anything an adversary might do to you) has nearly same probability if you join or don't join, lie or tell the truth. So, if an attacker has a lot

Commento [DM57]: Ha quasi la stessa probabilità se si unisce o non si unisce, mentire o dire la verità

of information that are not in the dataset about the user, he never find out information that are in the dataset about that user, while for instance with k-anonymity if we know exactly all the information about the user we can discover the sensitive information.
(CONTRO)The problem with differential privacy is that the parameter ε is public and the value of this parameter is still uncertain, because given a high value of ε implies that the dataset protection is very strong but it happen that the answer to a query is meaningless, instead a low value of ε can allow an attacker to discover data.

Commento [DM58]: Mentre per esempio con k-anonimato se conosciamo esattamente tutte le informazioni sull'utente, possiamo scoprire le informazioni sensibili

Commento [DM59]: Priva di significato

14. Discuss how Bitcoin allows to preserve privacy of users and at the same time it does not allow double spending of bitcoins. Which is the computational hard problem that is used to guarantee the privacy of users and why we trust bitcoin?

Bitcoin is a digital asset and a payment system invented by Satoshi Nakamoto.
The system relays on decentralized peer-to-peer architecture where all transactions take place between users directly, without an intermediary. To reach a good level of privacy and security Bitcoin uses three cryptographic tools:
• **Hash functions**: they take as input any string and produce a fixed size output."""" They are collision resistant, hiding and puzzle friendliness, so is infeasible given the output to find the input."""
• **Digital signatures**: are unforgeable, and are used to create a pair of keys (one private and one public), to sign the transactions and to verify the signature of the transaction. They guarantee authentication.
• **Public/Private keys**: the private key is used to sign the message, the public key is used to verify the signature and the hash of the public keys the address of the user. They guarantee anonymity.

Commento [DM60]: Sono effettuate

Commento [DM61]: Sono imprevedibili e vengono utilizzati per creare un paio di chiavi (un privato e uno pubblico), per firmare le transazioni e per verificare la firma della transazione. Garantiscono l'autenticazione.

Bitcoin allows to preserve privacy thanks to two main concepts:

1. **All transactions take place directly between users**, without an intermediary.
2. Users **use public and private keys and a bitcoin address** (hash of the public key) to send and receive payments.

From the first point we have the advantage that there is not the need of a central authority which manages transactions. This allows to avoid also the need of someone that stores the correspondence between wallets and physical users.
From the second point we have the advantage that each user need only a pair of public and private keys in order to execute a transaction. This is important because each user can randomly generate different new pairs of keys and then use them to execute new transactions. It's clear that in this way is really difficult to track the traffic of a user.

The protocol used by Bitcoin to verify if a transaction is valid prevents also the double-spending attack: in order to validate transactions, we have to check if each input is already spent or not. .
This is due to the fact that by simply broadcasting transactions through the network it is not possible to achieve also an ordering between them. A consequence is that we cannot have consistency because each node can accept a different set of transactions. We solve this problem by using the so called blockchain that is a sequence of blocks. A block contains: an hash pointer to the previous block that it is extending, a set of transactions and a nonce.

Commento [DM62]: . In the Bitcoin system this is avoided because every transaction is signed with the private key of the user and sent along with its address to the network, in order to be accepted by the community. Thanks to the block chain the system can control if the bitcoins of the transaction in a block were already sent by the same user in previous transactions

Commento [DM63]: Ciò è dovuto al fatto che semplicemente trasmettendo transazioni attraverso la rete non è possibile ottenere anche un ordine tra di essi

The algorithm executed by each node is the following:

- It maintains a set of unconfirmed transactions received in broadcasting.
- It selects a subset of valid transactions (among unconfirmed) and, in order to create a new block, it starts the so called puzzle game.
- If it succeeded in the mathematical challenge, it broadcasts the new block to the network.

A **node accepts a block if** all transactions in it are valid. A **transaction is valid if** and only if its inputs are not already spent. An **input is not spent if** we can go from the block which contains it to the current block.

Commento [DM64]: Possiamo andare dal blocco che lo contiene al blocco corrente.

In the case in which **two blocks are generated at the same** time the chain is divided in two branches. An important rule is that each node has to consider as the correct one the **longest chain**.

Commento [DM65]: La regola è che ogni nodo deve considerare come quello giusto la catena più lunga.

All the transactions of the discarded blocks come back to the **unconfirmed set**. This feature seems to expose the system to a double spending attack. In practice this is not true because it is n**ot possible to precompute a long enough chain** that obtain acceptance. This is due to two main aspects:

Commento [DM66]: Tutte le operazioni dei blocchi scartati tornano al set non confermato

- Each block contains an **hash pointer to the previous** accepted block and then it cannot be solved before that the previous one is available
- The **puzzle game** requires a lot of computations in order to be solved

This means that also **if we have the 50% of the computational power** of the whole network we have only the 50% of probability to solve the block before another node. This probability decreases exponentially if we want to solve the challenge several times consecutively.

Commento [DM67]: Questa probabilità diminuisce esponenzialmente se vogliamo risolvere la sfida più volte consecutivamente

Bitcoin is a digital reliable crypto-currency because it uses a mathematical challenge, which is a computational hard problem, to create a new blocks, and so to put transactions into it. This mathematical challenge is called mining. The idea of the mining is that the miner have to find a value that when hashed, for instance with SHA-256, it falls in a specific domain. This domain is a set of numbers, having some zero bits at the beginning.

Commento [DM68]: Bitcoin è un cripto-valuta affidabile digitale poiché usa una sfida matematica

Commento [DM69]: L'idea della miniera è che il minatore deve trovare un valore che, quando viene sparso, ad esempio con SHA-256, rientra in un dominio specifico.

In conclusion, we can say that **we trust bitcoin because** we have:

- **Privacy** thanks to the cryptographic hash function.
- **Authentication and data integrity** thanks to the digital signature.
- **Consistency** thanks to blockchain.

15. Discuss how Bitcoin achieves consensus and discuss how the protocol does not allow i) Stealing Bitcoins, ii) Denial of service attack, and iii) Double spending Bitcoins.

Consensus in bitcoin system is reached following different steps:

- Each new transaction is broadcasted from a node to every other nodes
- Each node collects new transactions into a block
- In each round a miner node tries to find a nounce that, combined with the hash of the last block in the ledger and new transactions, allows him to find a valid block to propose (valid means that

H (H (previousBlockInLedgerHash) | nonce | newTransaction) = Y,

where Y is the set of good possible values).

- Once he has found this nounce, he broadcasts his block, together with the nounce
- Other nodes accept the block only if all transaction in it are valid and the hash function with his nounce is correct (find a nonce is hard, but having it is easy to prove that the result of the hashing function is correct).
- Nodes express their acceptance of the block by including its hash in the next block they create.

This is one of the reason of why the nodes are honest. If some of them accept a malicious block with fake transactions and starts to mining from them, and then the block is not approved from the entire community, they have to restart the mining from the beginning, so it's not convenient for them to accept the malicious block.

1. Stealing bitcoin: a miner that proposes a block with new transaction cannot steal money from them because all that he knows from users in its transactions is their public address (that is the hash of their public key), while to spend their bitcoin he should know their private key.

2. Denial of Service: a miner that proposes a new block can decide to exclude from his block all the transaction that come from a specific user A, but since transaction are sent in broadcast, there will be sooner or later a miner that will keep the transaction of A in his approved block.

3. Double spending (vedi anche sopra): Double spending is the possibility to make transactions with different addresses using the same bitcoin. In the Bitcoin system this is avoided because every transaction is signed with the private key of the user and sent along with its address to the network, in order to be accepted by the community. Thanks to the block chain the system can control if the bitcoins of the transaction in a block were already sent by the same user in previous transactions in any of the previous blocks and if there's something irregular, the consensus on the transaction is negated, otherwise it is accepted and the transaction goes with other valid transactions to the creation of a new block.

Commento [DM70]: raggiunto

Commento [DM71]: "libro mastro" / registro

Commento [DM72]: RUBARE IL BITCOIN- inutile poiché si conosce solo l'indirizzo mentre per spenderlo serve anche la chiave privata

Commento [DM73]: Rubare denaro da loro

Commento [DM74]: RIFIUTO DEL SERVIZIO – Anche se uno decide di eliminare tutte le transazione di un'utente specifico, ci sarà sempre uno che se lo terrà, questo perché le transazioni sono inviate in broadcast

Commento [DM75]: Ci sarà prima o poi un minatore che manterrà la transazione di A nel suo blocco approvato

Commento [DM76]: Inviato insieme al suo indirizzo alla rete

16. i) Explain what is the meaining of the following claim"Bitcoins are immutable"; ii) the number of bitcoins available increases every day (for several years). Discuss this claim explaining why the designer of Bitcoin decided to include such possibility in the protocol.

The claim "Bitcoins are immutable" means that they can never be changed, subdivided, or combined. Each coin is created, once, in one transaction and later consumed in some other transaction. The only way to subdivide a coin is to create a new transaction that consumes a coin, and then produces two new coins with the same total value of the previous one. Those two new coins could be assigned back to the previous owner. This immutability is one of the main features that adds trust to the system, since the ledger is intact and users can trust it to control previous transaction in order to avoid different attacks.

Commento [DM77]: Libro mastro / registro

The main activity of the Bitcoin network is the so called "mining". A miner is rewarded with bitcoins when he creates a new block because in order to perform this operation, a lot of computational and energetic resources are required. A reward is paid only if a block ends up on the long-term consensus branch, therefore miners are encouraged to behave honestly. We can note that this procedure is also a way to initially distribute coins into circulation (since there is no central authority to issue them).

Commento [DM78]: ricompensato

Today the value of the block reward is 25 Bitcoin but it actually halves every 210000 blocks then we have a geometric series which converges to a finite sum. This means that the rewards system will work out to a total of 21 million bitcoins. When the bitcoins dedicated to rewards will end, the system will continue to work thanks to transaction fees. They are voluntary incentives associated to each single transaction. They will become more and more important, almost mandatory in order to get a reasonable quality of service.
The steady addition of a constant of amount of new coins is analogous to gold miners expending resources to add gold to circulation (in our case we talk about CPU time and electricity that is expended).

Commento [DM79]: PER QUESTA PARTE VEDERE ANCHE LA DOMANDA 20

Commento [DM80]: Forse 12.5 CONTROLLARE

Commento [DM81]: Quando i bitcoins dedicati ai premi saranno terminati, il sistema continuerà a lavorare grazie alle commissioni di transazione. Sono incentivi volontari associati ad ogni singola transazione. Diventeranno sempre più importanti, quasi obbligatori per ottenere una qualità ragionevole di servizio.

Thanks to this reward, miners want to be honest in order to have their block approved, and people will approve honest blocks in order to start mining from that block (that it's honest, so it will have more probability to be approved than a malicious one).

17.Bitcoin network: i) describe the network and how a transaction is propagated and validated; ii) describe the different kind of nodes.

i) The Bitcoin relays over a peer-to-peer network, which is a network in which all nodes are equal without hierarchy or special nodes. It runs over TCP and has a random topology. In each moment, a new node can join to the network having equal rights and capabilities as every other node. The new node should send a message to present itself to one node in order to know nodes that it knows. After that, the new node repeats the process with the new known nodes and then chooses which ones to peer with. A node has the possibility to leave the network. Other nodes start to forget it after a period in which they have not received any signals from it. To publish a new transaction, this kind of network use a simple flooding algorithm, presented below in few steps:

• a node that wants to complete an action, send the transaction to all nodes it's peered with and each of them checks if it is valid;

• if it is valid, the node in turn sends it to all of its peer nodes;

• nodes that hear about a transaction put them in a pool of transactions that they have heard about but are not on the block chain yet.

ii) In the Bitcoin network there are two kinds of nodes: fully validating nodes and lightweight nodes.

The first ones must store the entire blockchain ledger and they are permanently connected and fully validate every transaction they heard.

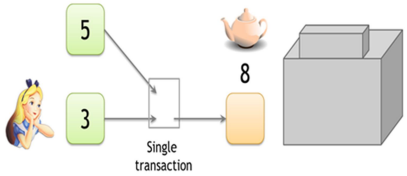
The second ones are the vast majority on the Bitcoin network and only store pieces of the ledger that they need to verify specific transactions.

18. Bitcoin network: Bitcoin is completely anonymous but all transactions are public. Discuss how the network of Bitcoin transactions looks like and how it is possible to use this knowledge to deanonymize the Bitcoin network. Discuss at least one example.

In order to achieve **anonymity** we must have pseudonymity and unlinkability.
Bitcoin network i**s pseudonymous** because addresses are hashes of the public key instead of real identities but it is **not unlinkable**.
In general, we have **unlinkability** when different interactions of the same user with the system are linkable to each other. This means that it is:

- Hard to link different addresses of the same user
- Hard to link different transactions of the same user
- Hard to link sender of a payment to its recipient

The **first point is not reached** because in a situation like this:



with a high probability, the input transactions used to execute this payment belong to the same user even if they have different public keys.

The **second point is not reached** because all transactions that have the same public key also belong to the same user.

Commento [DM82]: Un nodo che vuole completare un'azione, inviare la transazione a tutti i nodi con cui viene guardata e ciascuno di essi controlla se è valido;

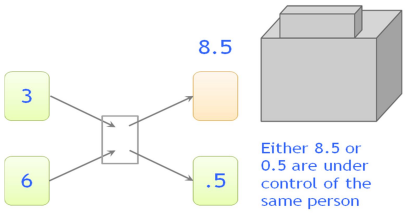
Commento [DM83]: Sono la stragrande maggioranza della rete Bitcoin

Commento [DM84]: Per la prima parte, vedere anche il punto 1 della domanda precedente (come si effettuano le transazioni dei BitCoin)

Commento [DM85]: chiedere

Commento [DM86]: • È difficile collegare diversi indirizzi dello stesso utente

The **third point is not reached** because in a situation like this:



we know that the receiver of 0.5 or 8.5 belong to the same user as the sender.

Despite the popular belief, Bitcoin is not an anonymous network.

Using these methods and some external information (if available) it is possible to deanonymize the bitcoin network.

There are basically three ways to de-anonymize Bitcoin users.

First of all, even though Bitcoin transactions are randomly transmitted over the peer-to-peer network, this system is not airtight. If an attacker, for instance, has the means to connect multiple nodes to the Bitcoin network, the combined data collected from these different nodes might be enough to determine where a transaction originated.

Second, Bitcoin addresses can be linked to real identities if these real identities are used in combination with the Bitcoin addresses in some way. This includes addresses used to deposit or withdraw money to or from a (regulated) exchange or wallet service, publicly exposed donation addresses, or addresses simply used to send bitcoin to someone (including the online store) when using a real identity.

But perhaps most importantly, all transactions over the Bitcoin network are completely transparent and traceable by anyone. It's typically this complete transparency that allows multiple Bitcoin addresses to be clustered together, and be tied to the same user. Therefore, if just one of these clustered addresses is linked to a real-world identity through one or several of the other de-anonymizing methods, all clustered addresses can be.

The most common example is the **deanonymization of silk road**, that was the largest market for illegal drugs based on a tor hidden service and bitcoin as currency. Even if the creator tried to cover his activities, he was arrested 2 years later because **police oversee movements in silk road** until they reach to the conclusion that he was the owner.

Commento [DM87]: Ermetico / chiuso ermeticamente

Commento [DM88]: Se un attaccante, ad esempio, dispone dei mezzi per collegare più nodi alla rete Bitcoin, i dati combinati raccolti da questi diversi nodi potrebbero essere sufficienti per determinare la provenienza di una transazione.

Commento [DM89]: Gli indirizzi Bitcoin possono essere collegati a identità reali se queste identità reali vengono utilizzate in combinazione con gli indirizzi Bitcoin in qualche modo.

Commento [DM90]: Ma forse più importante, tutte le transazioni sulla rete Bitcoin sono completamente **trasparenti e tracciabili da chiunque. Di solito questa trasparenza completa consente di raggruppare insieme più indirizzi Bitcoin e essere legati allo stesso utente**. Pertanto, se solo uno di questi indirizzi cluster è collegato ad un'identità del mondo reale attraverso uno o più degli altri metodi di anonimizzazione, tutti gli indirizzi raggruppati possono essere.

Commento [DM91]: L'esempio più comune è la deanonymizzazione della silk road, il più grande mercato di droga illegale basato su un tor hidden e bitcoin come valuta. Anche se il creatore ha cercato di coprire le sue attività, è stato arrestato due anni dopo perché la polizia sorveglia i movimenti della silk road fino a giungere alla conclusione che egli era il proprietario.

19. Robustness of Bitcoin against malicious adversaries. If a malicious ISP completely controls a user's connections, can it launch a double-spending attack against the user? How much computational effort would this take?

Even if a malicious ISP controls a user connection, it can't do a double spending attack because the

success of the attack doesn't depend by the fact that the user is controlled, but it depends from the consensus of the network. The ISP is treated as normal node, so each of his transaction must be published within all over his neighbours. One of the transactions of the double spending attack won't be accepted by the community, and that is independent from the fact that the ISP completely controls the user connection. The double spending attack could be done theoretically if the attacker computationally controls over the 51% of the network nodes, so that he can control the consensus process and approve a double spending. This is infeasible from the financial point of view, and even if an attacker realizes it, people who get money stolen will leave the bitcoin system, the value of bitcoin will decrease and the attacker won't earn nothing to cover the expenses done to have 51% computational power.

20. Discuss which are the cost and the incentives for miners. Discuss how the incentives favour good behaviour in block chain. What is the transaction cost that is paid to miners and why it is becoming more frequent?

Miner's cost is the effort required to create a new block; in particular they must find a nonce n such that *H(H(previousBlock), N, new transactions) = Y* , where Y is the set of good possible values (tries to find a nounce that, combined with the hash of the last block in the ledger and new transactions, allows him to find a valid block to propose) . The system tries to maintain the average time required to find the nonce in about 10 minutes for a random user in the network. If users, in average, solve the hashing function in less/more than 10 minutes, the difficulty of the hash will increase/decrease. The financial cost for users is the hardware required to find the nonce (nowadays there are dedicated hardware to do that) and the power to use it 24h/7.

The incentives are that an approved block will give to the miner a x number (25 now) of new bitcoins and the good behaviour is guaranteed because a miner not only needs to find the nonce, but all the transactions that he envelope must be approved by the community, otherwise the block won't be approved and he will lose money. Moreover, miner could be rewarded with transaction fees of the transaction in the block, and this reward is becoming more frequent because the "x" flat bitcoin received from the approved block are halved every tot block created.

A **transaction fee** is the difference between the total values of coins that go into a transaction minus the total value of coins that come out.

21. Which are the possibility to store bitcoins; discuss the risks that are associated.

Since the possession of bitcoin is related to the possession of a private key, in many cases the management of the couple *<public key, private key>* is done by some software, called "wallet" software, who care about all the operations of signature of outgoing transactions and blockchain reading (even if exist some "lite" versions, which avoid to download the entire blockchain), and care about generation and storing of private keys files

In order to store bitcoins, we have two main types of storages:

- **Hot storage**, storing bitcoins on your computer is like carrying money around in your wallet, convenient but also risky. It performed by software connected to internet, which allows users to easily spend and receive bitcoins. Main risk is that private keys are usually

Commento [DM92]: Questo è impraticabile dal punto di vista finanziario, e anche se un aggressore lo rende conto, le persone che riceveranno soldi rubati lasciano il sistema bitcoin, il valore del bitcoin diminuirà e l'aggressore non guadagnerà nulla per coprire le spese per avere Il potere computazionale del 51%.

Commento [DM93]: Per la prima parte, vedere anche il secondo punto della domanda 16

Commento [DM94]: "libro mastro" / registro

Commento [DM95]: Forse 12.5, come nella 16

Commento [DM96]: Inoltre, il minatore potrebbe essere premiato con le commissioni di transazione dell'operazione nel blocco e questa ricompensa sta diventando più frequente perché il bitcoin piatto "x" ricevuto dal blocco approvato viene dimezzato ogni blocco totale creato.

Una commissione di transazione è la differenza tra i valori totali di monete che entrano in una transazione meno il valore totale delle monete che vengono fuori.

Commento [DM97]: Che si preoccupano di tutte le operazioni di firma delle transazioni uscenti e della lettura di blocchi

stored into a file, which can be stolen using a malware, or can be lost if hardware damage happens, like carrying your real money in your real wallet.

- **Cold storage**, which is an approach that generates public and private key for storing them **offline** in a safe place (usb sticks, even in paper, through IQ codes). They are secure from the point of view of digital intrusions/damages, but, since they are offline, the relative transactions require more time because the keys need to be physically accessed to make payments. In addition, like every physical good, they present some **risks** (loss of storage drive, disaster, etc.) and bitcoins can be lost forever.

Other less used possibilities are:

- Tamper resistant card
- Online wallet, you give your keys to a trust entity that acts as a Bank and stores your Bitcoin. You give your money but you trust that they are expert in security.
- Secret sharing mechanism, you divide the secret in pieces.

Commento [DM98]: carte resistenti alla manomisssione

22. Even when all nodes of Bitcoin are honest, blocks will occasionally get orphaned: if two miners Minnie and Mynie discover blocks nearly simultaneously, neither will have time to hear about the other's block before broadcasting hers.
-1: What determines whose block will end up on the consensus branch?
-2: What factors affect the rate of orphan blocks?
-3: if Mynie hears about Minnie's block just before she's about to discover hers, does that mean she wasted her effort?

Commento [DM99]: Se Mynie ascolta il blocco di Minnie poco prima che sta per scoprirlo, significa che ha sprecato il suo sforzo?

-1: In general, if we have at the same time two blocks A and B, we can have the following situations:

- **A valid and B not valid**. Obviously in this case a miner will continue the chain choosing A.
- **A valid and B valid**. The general rule provides that a miner should continue on the longest branch. In this situation we have two branches with the same length therefore a miner can choose randomly one of them. This means that the next block will determine the longest chain. Nodes try to extend the transaction they are aware of: the accepted transaction is the one that will be the longest; time to extend is random. The accepted transaction is decided within some time window: more time, more certainty.

Commento [DM100]: I nodi tentano di estendere la transazione di cui sono consapevoli: la transazione accettata è quella più lunga, il tempo per estendere è casuale. La transazione accettata viene decisa all'interno
Qualche finestra di tempo: più tempo, più certezza

-2: Essentially there are two main factors which affects the rate of orphan blocks:

Commento [DM101]: Influenzano nel senso di provocano

- **Network delay**. More network delay → More time to reach consensus.
- **Time to find a block**. Easier hash puzzle → Less average time to build a block → Higher probability to have different blocks at the same time → More chains → More orphan blocks.

Commento [DM102]: • Tempo per trovare un blocco. Puzzle hash più facile → Minore tempo medio per costruire un blocco → Maggiore probabilità di avere blocchi diversi allo stesso tempo → Più catene → Più blocchi orfani.

-3: Mynie does not waste time **if Minnie's block is not valid**. In fact, in this case Mynie can discard the received block and continue to solve the hash puzzle.

Instead, if **Minnie's block is valid**, Mynie has no incentive to continue working on her block because with a high probability it will become an orphan block.
No, her block could be the one with the longest valid branch and be approved by the consensus before the other one. (or search for a new block starting from the Minnie's one) (??????).

23.Even when all nodes of Bitcoin are honest, blocks will occasionally get orphaned. In 2013 a major fork was observed. Discuss the reasons why it happened and how it was solved. Do you think that it might happen again?

Commento [DM103]: Presa totalmente da Pic

A **hard/major fork** is a software upgrade that introduces a new rule to the network that isn't compatible with the older software. You can think of a hard fork as an expansion of the rules.

In August 2013, after a software upgrade, a **block that had a larger number of total transaction** inputs than previously seen was mined and broadcasted. Bitcoin 0.8 nodes were able to handle this, but some pre-0.8 Bitcoin nodes rejected it, causing an unexpected hard fork of the blockchain. The **pre-0.8 incompatible chain at that point had around 60% of the hash power** ensuring the split did not automatically resolve.

Commento [DM104]: La catena incompatibile pre-0.8 a quel punto aveva circa il 60% del potere hash che garantiva che la spaccatura non si risolvesse automaticamente.

In order to restore a canonical chain as soon as possible, **miners BTCGuild and Slush downgraded their Bitcoin 0.8 nodes to 0.7**. They quickly downgraded their nodes to restore a pre-0.8 chain as canonical, despite the fact that this caused them to sacrifice significant amounts of money. This placed majority hashpower on the chain without the larger block, thus eventually causing the 0.8 nodes to reorganise to the pre-0.8 chain.

During this time there was one large double spend; it was doen by someone experimenting to see if it was possible (not a malicious double spending)

[NB] Per dettagli tecnici sulla causa e su come è stato risolto vedi slide 96-97 (4-final)

In my opinion it can be happen again for several reasons:

- technological evolution may require a system upgrade
- political empasses
- Bu vs SegWit dispute

Commento [DM105]: Emendamenti politici

24. What is a zero knowledge proof; discuss why a zero knowledge proof of identity is an interesting concept also for security reasons. Which are the practical problems when using zero knowledge proofs?

A zeroknowledge proof is a method in which a party (prover) can prove to another party (verifier) that a certain statement is true, without giving any information about the statement apart from the fact that it is indeed true.
Interactive Zero Knowledge proof is made by 3 steps:

Commento [DM106]: Senza dare alcuna informazione sulla dichiarazione, a prescindere dal fatto che sia veramente vero.

1. a Commitment phase by Prover to Verifier;
2. a Challenge phase by Verifier to Prover;
3. a Response phase by Prover to Verifier.

A ZKP must satisfy three properties:

- Completeness: given a honest prover and honest verifier, the protocol succeeds with overwhelming probability. The verifier accepts the prover's claim.
- Soundness: a dishonest prover can't convince a honest verifier except with a small probability.
- Zero-knowledge: prover does not reveal any information about the secret and participation in the protocol does not increase the chances of subsequent impersonation.

It's interesting by the point of view of security because the prover does not release any information about the secret knowledge. *Participation in the protocol does not increase the chances of subsequent impersonation.* ZKP eliminates the need for a preliminary secure interaction between parties, because it implies the authentication between the parties involved.

Practical problems are that the protocol must be repeated a good number of times in order to be statistically reliable (a dishonest user has the probability to cheat about knowing a proof equals to 1/(2^t) where t is the number of times he has to show the ZKP to the verifier) and it is possible to apply the Zero Knowledge method to a specific subset of problems.

Commento [DM107]: completezza

Commento [DM108]: Con incredibile probabilità

Commento [DM109]: Solidità

Commento [DM110]: Il prover non rivela alcuna informazione sul segreto e la partecipazione al protocollo non aumenta le probabilità di successiva rappresentazione.

Commento [DM111]: La partecipazione al protocollo non aumenta le possibilità di una successiva rappresentazione.

Commento [DM112]: ZKP elimina la necessità di un'interazione preliminare protetta tra le parti, perché implica l'autenticazione tra le parti coinvolte.

Commento [DM113]: Per essere statisticamente affidabili

Commento [DM114]: imbrogliare

25. What are according to you the main concerns of privacy protection in social networks?

Third party applications, as social network, doesn't have term of conditions and they can easily steal personal informations. The main challenge of data privacy is to utilize data while protecting individual's privacy preferences and their personally identifiable information. Social network is a structure made up of a set of actors (such as individuals or organizations), and interactions between these actors. Privacy concerns with social networking services is a subset of data privacy. Social network privacy issues result from the amounts of informations these sites process each day. Features that invite users to participate in messages, invitations, photos, open platform applications and other applications are often the venues for others to gain access to a user's private information.
As there are so much informations provided by social networks, other things can be deduced, such as the person's social security number (fiscal code), which can then be used as part of identity theft. Many studies shown that it is possible to predict most and sometimes all of an individual's 9-digit Social Security number (16 for fiscal code) using information gained from social networks and on-line databases.

Due to the high content of personal informations placed on social networking sites, as well as the ability to hide behind a pseudo-identity, such sites have become increasingly popular for sexual predators. Further, the lack of age verification mechanisms is a cause of concerns in these social networking platforms.
It has been noted that the number of sexual predators caught using social networking sites has been increasing, and has now reached an almost weekly basis. In worst cases children have become victims of pedophiles or lured to meet strangers.

Commento [DM115]: La principale sfida della privacy dei dati è quella di utilizzare i dati proteggendo le preferenze di privacy individuali e le loro informazioni personali

Commento [DM116]: Preoccupazioni relative alla privacy

Commento [DM117]: I problemi di privacy di rete sociale derivano dalle quantità di informazioni che questi siti gestiscono ogni giorno.

Commento [DM118]: Sedi/Luoghi

Commento [DM119]: Numero di previdenza sociale: è assegnato negli Stati Uniti ad ogni cittadino, per tenere traccia dei contributi versati e della posizione occupazionale. Stessa funzione, anzi più ampia, è assolta in Italia dal codice fiscale.

Commento [DM120]: A causa dell'elevato contenuto di informazioni personali collocate sui siti di social networking e della possibilità di nascondersi dietro una pseudo-identità, tali siti sono diventati sempre più popolari per i predatori sessuali.

Commento [DM121]: Mancanza verifica età

Commento [DM122]: Nel peggiore dei casi i bambini sono diventati vittime di pedofili o attirati per incontrare estranei.

With the amount of information that users post about themselves on-line, it is easy for users to become a victim of stalking without even being aware of the risk. 63% of Facebook profiles are visible to the public, meaning that if you Google someone's name and adding "+Facebook" in the search bar, you pretty much will see most of the person profiles.

Twitter, Facebook, and others, are application which allow users to share their current location information. However, the disclosure of location information within these networks can cause privacy concerns among mobile users. Although there are algorithms using encryption, k-anonymity and noise injection algorithms, its better to understand how the location sharing works in these applications to see if they have good algorithms in place to protect location privacy.

Commento [DM123]: La divulgazione delle informazioni sulla localizzazione all'interno di queste reti può causare preoccupazioni sulla privacy tra gli utenti mobili

26. Discuss the difference between a semi-honest attacker and a malicious one in a social network; give examples of attacks.

The social media sites are constant targets for spam, scams and other attacks. In this scenario it is possible to distinguish two types of attacker: semi-honest attacker and malicious attacker. The difference between them is in the way the two attackers executing their intent.

Commento [DM124]: truffe

A semi-honest attacker always follows the protocol of the attacked entities, but it tries to learn information that he would not learn in an ideal execution of the protocol. The adversary passively attempts to learn the inputs of honest agents by using intermediate informations received during the protocol and any other information that it can gain through other legitimate means.

Commento [DM125]: Un attaccante semi-onesto segue sempre il protocollo delle entità attaccate, ma cerca di apprendere informazioni che non avrebbe imparato in un'esecuzione ideale del protocollo. L'avversario tenta in modo passivo di apprendere gli input di agenti onesti utilizzando informazioni intermedie ricevute durante il protocollo e qualsiasi altra informazione che possa ottenere attraverso altri mezzi legittimi

Instead, a malicious attacker deviates from the protocol in arbitrary ways. They may refuse to participate in the protocol, provide out of range values as inputs, selectively drop messages that they are supposed to send, prematurely abort the protocol, distort informations, and tamper with all communication channels. Objectives of malicious attacker: learn the input of honest agents, disrupt the protocol of honest agents.

Commento [DM126]: Invece, un attaccante dannoso devia dal protocollo in modo arbitrario. Possono rifiutarsi di partecipare al protocollo, fornire i valori di intervallo come input, eliminare selettivamente i messaggi che dovrebbero essere inviati, abortire il protocollo, distorcere le informazioni e provvedere a mancare tutti i canali di comunicazione. Obiettivi di attaccanti maligni: imparare l'ingresso di agenti onesti, interrompere il protocollo di agenti onesti

To explain better this difference it is possible to consider the case of two attackers, a malicious one and a semi-honest, in a social network. The first attacker, malicious, may post malicious link or injects malicious code in chat messages, with the aim of gaining control of the adversary machine or steal credential.

Commento [DM127]: Interrompere

The latter (semi-honest) instead try to gain more informations about the victim by searching throughout the social network for pictures, friends, posts and all the other stuff related to a social network.

Commento [DM128]: cose

So, a semi-honest attacker can learn nothing more about a victim than it reveals by itself therefore if nothing is revealed by the victim the attacker has no possibility to gain information; instead a malicious one may always try to attack social network user's.

Commento [DM129]: Vittima di quanto non si rivela di per sé, quindi se nulla viene rivelato dalla vittima l'attaccante non ha possibilità di ottenere informazioni

27. Present and discuss how the Netflix network was deanonymized.

Commento [DM130]: Parte giù inserito nella domanda 5

Due to a contest for recommendation system Netflix released a version of the database of ratings, with some level of anonymization (names removed, perturbation of information). Combined with background knowledge, which was IMDB, an attacker can perform a deanonymization attack with usage of another, similar, database.

In details, the objective was to:

-Fix some target record of original Netflix dataset

-Try to learn as much about this record as possible

But background knowledge (IMDB dataset) was noisy, and Netflix dataset was perturbed (with only a sample of records released).
Anyway, since ratings about not top100 movies are very personalized (is unusual for two users give same rating on same not so known movies), the researchers in this project found out that, with this cross references on movie ratings and date of ratings, some users turned out to be members of both IMDB and Netflix (with some personal informations voluntarily released on IMDB), and personal information was obtained with a very low percentage of error (in the experiment, just 4 ratings, in mean, were enough to uniquely identify the user).
With only eight movie ratings, and dates that may be up to two weeks in error, they can uniquely identify 99% of the records in the dataset. After all, all they need is a little bit of identifiable data: from the IMDb, from your blog, from anywhere.
The moral is that it takes only a small named database for someone to pry the anonymity off a much larger anonymous database

Commento [DM131]: Con solo otto valutazioni di film e date che possono durare fino a due settimane in errore, possono identificare in modo univoco il 99% dei record nel set di dati. Dopo tutto, tutto quello di cui hanno bisogno è un po 'di dati identificabili: dagli IMDb, dal tuo blog, da qualsiasi parte.
La morale è che occorre solo un database di dimensioni ridotte per consentire a qualcuno di fare l'anonimato su un database anonimo molto più grande

28. Which are the main characteristics of the new European regulations for privacy for individuals?

With the new European regulations, the European Commission intends to strengthen and unify data protection. It also addresses export of personal data outside the EU. Individuals must retain effective control over their personal data. This is a fundamental right for everyone in the EU and must be safeguarded.

Commento [DM132]: Rafforzare e unificare

Commento [DM133]: Essa si occupa anche dell'esportazione di dati personali al di fuori dell'UE

Commento [DM134]: conservano

The main characteristics regulations for privacy for individuals are:

• a **right to be forgotten** will help you manage data protection risks online. When you no longer want your data to be processed and there are no legitimate grounds for retaining it, the data will be deleted.

Commento [DM135]: Motivi legittimi per mantenerlo

• **easier access** to your own personal data;
• **a right to data portability**: it will be easier to transfer personal data from one service provider to another. In addition, the data must be provided by the controller in a structured and commonly used electronic format;
• when **your consent is required**, you must be asked to give it by means of a clear affirmative action;

Commento [DM136]: Bisogna chiederlo per un'azione chiara e affermativa;

• **more transparency** about how your data is handled, with easy-to-understand information, especially for children;
• **the right to know when your data has been hacked :** businesses and organizations will need to inform you about data breaches;
- **Stronger enforcement of the rules:** Data protection authorities will fine companies which are not compliant with EU rules.

Commento [DM137]: • Maggiore attuazione delle norme: : Le autorità di protezione dei dati beneficeranno di aziende non conformi alle norme dell'UE.

• two important principles:
  - **'data protection by design'**, which requires that data protection is designed into the development of business processes for products and services

Commento [DM138]: Che richiede che la protezione dei dati sia progettata nello sviluppo di processi aziendali per prodotti e servizi

  - **'data protection by default'**, which means that the default settings should be those that provide the most privacy.

Commento [DM139]: Le impostazioni predefinite dovrebbero essere quelle che forniscono la massima privacy.

29. Which are the main characteristics of the new European regulations for privacy for business?

With the new European regulations, the European Commission intends to strengthen and unify

data protection. It also addresses export of personal data outside the EU. A high level of data protection is essential to foster people's trust in on-line services and in the digital economy in general. Privacy concerns are among the top reasons for people not buying goods and services on-line. With the technology sector directly contributing to 20% of overall productivity growth in Europe individual trust in on-line services is vital for stimulating economic growth in the EU.

The main characteristics regulations for privacy for business are:

- **One continent, one law**: The regulation will establish one single set of rules which will make it simpler and cheaper for companies to do business in the EU.
- **One-stop-shop**: Businesses will only have to deal with one single supervisory authority. They will profit from faster decisions, from one single interlocutor, and from less red tape.
- **The same rules for all companies** regardless of where they are established. Companies based outside of Europe will have to apply the same rules when they offer goods or services on the EU market.
- **Risk-based approach**: the rules will avoid a "one-size-fits-all" obligation and rather tailor them to the respective risks.
- **Rules fit for innovation**: Data protection by design: will guarantee that data protection safeguards are built into products and services from the earliest stage of development. Privacy-friendly techniques such as pseudonomysation will be encouraged.

30. According to the new European rules companies and organizations need a Data Protection Officer (DPO); which are the task and the obligations related to this person?

Data protection officers are a designated person within an organization that collects the personal data of Union citizens who is responsible for making sure that the organization follows the new regulations.

DPO tasks are:

- **inform and advise the owner** of the treatment about their obligations under the European Regulation;
- **check the implementation** and application of the Regulations , provide opinions on the assessment of the impact on data protection;
- **act as a contact point** about any issue related to the owners of treatments' processing of data or the exercise of their rights;
- **act as a contact point** for the Authority for the protection of personal data or, refer to the Guarantor as its own initiative.

DPO requirements are:

- **Adequate knowledge** of legislation and data management practices personal;
- **Carry out** its functions in full independence and in the absence of conflicts of interest;
- **Operate in the employ** of the owner or responsible on the basis of a contract. Will make available to the head of the personal data protection human and financial

31. The regulations and law that protect sensible data are evolving. With reference to the Italian regulations discuss which data are considered sensible and which are the main rules to protect such data. Discuss which are the main problems in processing information according to the regulations and which kind of difficulties/problems such regulations poses to companies and administrations.

Commento [DM140]: Un livello elevato di protezione dei dati è essenziale per promuovere la fiducia delle persone nei servizi on-line e nell'economia digitale in generale.

Commento [DM141]: Non comprano merci

Commento [DM142]: Con il settore tecnologico che contribuisce direttamente al 20% della crescita complessiva della produttività in Europa, la fiducia individuale nei servizi on-line è fondamentale per stimolare la crescita economica nell'UE

Commento [DM143]: Che renderà più semplice e meno costoso per le imprese di fare affari nell'UE.

Commento [DM144]: Le imprese dovranno affrontare solo un'unica autorità di vigilanza.

Commento [DM145]: Da meno burocrazia

Commento [DM146]: Indipendentem ente da dove sono stabiliti

Commento [DM147]: • Approccio basato sui rischi: le regole eviteranno un obbligo "unificato" e piuttosto adeguarli ai rispettivi rischi.

Commento [DM148]: Le garanzie di protezione dei dati sono integrate in prodotti e servizi fin dalla prima fase di sviluppo

Commento [DM149]: Consigliare

Commento [DM150]: Sulla valutazione

Commento [DM151]: agire come punto di contatto su qualsiasi problema relativo ai proprietari dei trattamenti di trattamento dei dati o dell'esercizio dei loro diritti;

Commento [DM152]: • agire come punto di contatto per l'Autorità per la protezione dei dati personali o, riferirsi al Garante come propria iniziativa.

Commento [DM153]: • svolgere le proprie funzioni in piena indipendenza e in assenza di conflitti di interesse;

Commento [DM154]: Operare nell'impiego del proprietario o responsabile in base a un contratto. Metterà a disposizione del capo della protezione dei dati personali e finanziari

In Italy with the legislative decree of June 30th 2003, starts the regulation and law that protect sensible data according to the fast spreading of application technologies in the world. By considering section 4 of the legislative decree with respect to the definition of sensible data and personal data, they define:

- **Sensitive data**: personal data revealing racial or ethnic origin, religious beliefs , philosophical or other beliefs, political opinions , membership of parties , unions, associations or organizations of a religious , philosophical, political or trade union, as well as personal data disclosing health and sex life

- **Judicial data**: personal data relating to criminal records, the register of offense-related administrative sanctions and the relevant current charges , or as an accused or suspected person of the criminal procedure code

These data are processed following three main rules:

1. Anyone has the right to protection of personal data concerning him
2. The processing of personal data will respect human rights and fundamental freedoms and dignity, with particular reference to privacy
3. The information systems and programs are configured to minimize the use of personal data and identification data

The rights of individuals data are:

- The **right of access to personal data**, including the right to obtain confirmation of the existence of data concerning him, the right to have their communication in intelligible form
- **The updating**, correction or the integration of data
- **The cancellation**, anonymization or blocking of data processed in violation of the law
- **The right to oppose**
- *For **legitimate** and **documented reasons** to the treatment , even if carried out in a legitimate way*
- *If the **processing carried out** for the purpose of sending advertising materials or direct selling or for carrying out market research . In this case, you do not need motivation.*

- *The Authority: in Italy there is a collegial Ministerial authority (Garante) 4 people, expert in law and computer science*
- *Obligations of individuals: obligation in disclosure of information, obligation of authorization and consensus of personal data treatment, external information policy (e.g. using of data for some purpose, for websites, must require previous auth by the users) ←this could ve the problem in processing informations*

The problems for companies and administrations:

Commento [DM155]: Veloce diffusione

Commento [DM156]: rivelando

Commento [DM157]: oneri

Commento [DM158]: dati che lo riguardano

Commento [DM159]: Per motivi legittimi e documentati al trattamento

Commento [DM160]: Se l'elaborazione è effettuata allo scopo di inviare materiale pubblicitario o vendita diretta o per effettuare ricerche di mercato. In questo caso, non hai bisogno di motivazione.

Commento [DM161]: divulgazione di informazioni

They have to follow a strict set of rules regarding nondisclosure, information to users, consens, authorization to the treatment of data, long and difficult process in order to be compliant with Italian laws.
Some complications may come from BYOD **(Bring your own device)** . In corporations monitoring of employees and their devices is done for security reasons, but it must be done in accordance with the law: also, in case of BYOD monitoring activity on the personal device must be done within the employment context, and not outside.

Commento [DM162]: Deve seguire una serie rigorosa di regole in materia di **non divulgazione**, informazione agli utenti, consenso, autorizzazione al trattamento dei dati, processo lungo e difficile per essere conformi alle leggi italiane

Commento [DM163]: Porta il tuo dispositivo personale

32. What are the main challenges and difficulties of making a job as an expert in privacy protection? Do you think that there will be good opportunities for such a job? Which are the main issues that you foresee. Explain why yes or not.

Since we are now in an evolution of technology such that there are lot of effort and development about big data analysis, efficient algorithm of predictions, data mining; and in a scenario where information about people has a great economic value (spreading of personalized advertisement, market investigations etc.) paired with the great spread of social network in our daily lives (with are pure collectors of personal data) makes the privacy protection a big issue to be cared about.
Also in a corporate environment, such as companies, factories, and sensitive organizations, privacy of data is one of the most important focus in order to preserve safety of the company, and in some cases safety of people as well.

Commento [DM164]: Sono molti sforzi e sviluppo

Due to this growing need of privacy protection, yes, an expert in privacy protection is becoming a relevant figure both in private and corporate environment, indeed, in EU regulations, there is the obligation for companies which deal with personal data to hire a DPO (Data Protection Officer).
This person is the one that regulates and controls all aspects concerning data protection and data privacy, therefore due to this kind of activities that it is charged it must be an expert of either legislative regulation and cybersecurity aspects related to the IT of the companies for which it has been employed.

Commento [DM165]: Si richiede di essere un esperto di regolamentazione legislativa e di aspetti della cybersecurity legati all'IT delle società per le quali è stata occupata.

With this growing trend of sensitive data all across the digital world, this kind of figure will become more important, because is an expert from the point of view of the legal aspects and the impact of data protection, and operating with a more technical figure also expert in data protection is a fundamental aspect of his work.
In add to this situation, citizens are becoming also well informed in data privacy issues (or, let's hope so), and the same concerns of private citizens can be put in practice, forming a virtuous cycle driven by the customers, making the figure of data protection expert more relevant in the future with a growing need of it.

Commento [DM166]: Le stesse preoccupazioni dei cittadini privati possono essere messe in pratica