How Tinder and Facebook make stalkers' life easier

T. Delemazure - P. Bourhis - R. Rouvoy - W. Rudametkin

Abstract

Edit: Tinder patched this breach and does not use Facebook informations anymore. Everything explained here is outdated. Launched in 2012, Tinder is the first mobile app in several countries with more than 100 millions download. This dating app enable its users to meet people with close geographical proximity. In this paper, we'll use tools given by Facebook and Tinder to show that it is possible to find on Facebook almost every Tinder user who linked is Tinder account with his Facebook account.

1 Introduction

Tinder is a free dating app available on Android and iOS with more than 100 millions downloads on the play store.

The challenge of this kind of app is to enable date between persons with common interests, and for that, Tinder use an easy solution : The Facebook profile of the user. Indeed, Facebook profiles say a lot about who we are : Where do we live, who are our friends, what are my interests, what pictures I liked, who are my classmates, etc.

By promoting the connection between your Tinder account and Facebook account, Tinder get a lot of information about you, and consequently every Tinder user do.

Stalkers know that more than other, they even have a tool for stalking their matches on Tinder [1].

Today people are more concerned by their personal information, by who have access to them, and by data gathering more globally. However, they rarely secure these data and social networks like Facebook don't really help them.

2 Tools

To create a Tinder account, people have two possibilities : Use their phone number, or use their Facebook account. In any case, they have to give a phone number to the app.

We are interested in the case where the user use his Facebook account to log in. In that case, we may have access to several information about the user, according to his privacy settings : The first name is mandatory and unchangeable, the age and the distance from me are often displayed but can be hidden by the user. The gender can be chosen by the user and we can deduce it from our own settings (interested in). We also have access to some pictures, very often those picture are taken from the user's Facebook account. The user can show where he studies / where he works and add a personal description. Finally, a user can link is Tinder account to an Instagram account and a Spotify account. An example of what a Tinder profile look like is presented in Figure 1.

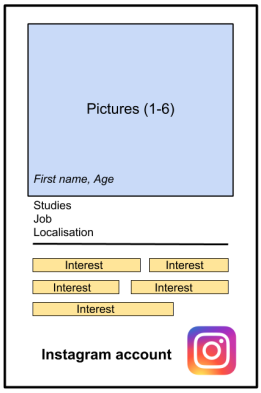

Figure 1: A Tinder profile

All these information can be shared publicly, even with people you don't match with.

In our experiment we will use Facebook interests. This information contains the 100 last Facebook pages liked by the users and display the intersection of my 100 in-

terests and the 100 interests of the user of who I see the Tinder profile.

In its side, Facebook give to the stalkers a lot of useful tools. First of all, Facebook enable precise queries on its Search bar tool. To simplify the queries, we will use the Facebook Query Language 2 [5].

This tool enables to select Facebook users with their first name, their gender, their interests (if they like a page), their work place, their university, their city. Most importantly, one can do the intersection of several selection like these.

Since the first name is a part of the public profile, we are sure to find the user on the result of a query selecting only according to the first name. But there is billions users on Facebook and their might be too many results for a stalker. We can then use Facebook interest, job, studies and city, but the user have the possibility to hide those information on Facebook.

Finally, we want to know how Facebook choose in which order it will display results of your query. We will see that it shows you people you may know first. Those people are : your friends, friends of friends and people with whom you have a common attribute (city/job/study/interests). We also want to know if this order is done even if those attributes are hidden.

3 The restrictive power of interests

It is known that our Facebook interests (i.e. pages we like on Facebook) tells many things about us. Actually, they can also say who we are, like a fingerprint. To be precise enough, we don't need to know every page you like on Facebook. Indeed, with your first name and around 10 interests, there is a really high probability to find you. First of all, what happen if we assume probabilities of like are random and independents ?

3.1 A bit of probabilities

Let F be the number of Local users of Facebook (targets of the page). Let $\mathscr{P}$ be a set of local Facebook pages, each page $P \in \mathscr{P}$ have $\phi(P)$ likes. Let $\mathscr{N}$ be a set of first name.

If we assume every like is independent, the probability that a user u like a page $P \in \mathscr{P}$ is $\mathbb{P}(u,P) = \frac{\phi(P)}{F}$. If we assume that $F > 30000000$ and $\forall P \in \mathscr{P}, \phi(P) < 5000000$, then $\frac{\phi(P)}{F} < 1/6$.

If we assume every like is independent and we know for a user u 10 liked pages, then we have $\mathbb{P}(u,P_1,..P_{10}) < \frac{1}{6}^{10} < 1,6 \times 10^{-8}$.

In France, the most common first name is "Marie" and there is less than 3% french citizen with this name. Then the probability for $n \in \mathscr{N}$ is $\mathbb{P}(u,n,P_1,...,P_{10}) < 3.5 \times 10^{-10}$

It means that the probability that 2 french citizen have the first name n and that their interests contains $P_1,...,P_{10}$ is around $1-(1-\mathbb{P}(u,n,P_1,...,P_{10}))^{50000000} = 1,7\%$. With 15 interests, there is almost no probability of false positive.

3.2 What happen in reality ?

Unfortunately, we can't say that every like is independent. Indeed, some pages have really similar communities. For instance, if one likes the page of a rap group, there is more probability that he also likes the page of a rap singer than the page of reality show. The figure 2 summarize this example by doing the intersection of audiences of different pages with users named "Manon" (in order to reduce the number of results)

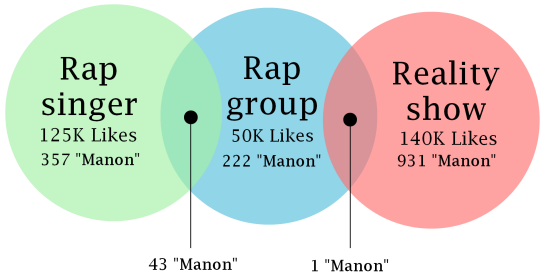

Figure 2: Intersection of audiences

In the second example shown Figure 3, we can see that it is useless to do the intersection between pages with identical audiences, for instance between a Youtuber 1 (2M Likes) and a Youtuber 2 (4,5M Likes). At the opposite, the intersection with a different kind of audience is very restrictive, for instance between the same Youtuber 1 (2M Likes) and a newspaper (4,2M Likers).

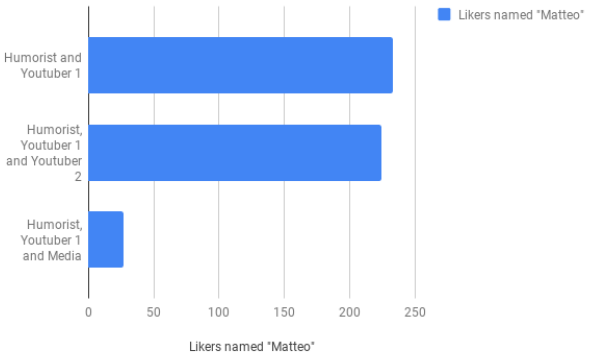

Figure 3: Intersection of identical/opposite audiences

Moreover, first name are not equally distributed. For instance, a page with 96K likes will have ten times more likers named "Alexandre" (668) than "Marcel" (59). Another issue with the Facebook search tool when you search user by name is that it return not only user who have exactly this first name but also user who have this name as last name or who have a name similar to this one. For instance, with the same page with 96K likes, if you want likers named "Jean", the Facebook tool will return more than 2,2K results, because of every "Jean-Pierre", "Jean-Jacques", etc., even if the actual number of "Jean" who liked this page does not exceed 300.

Another important point is to adapt the selection of pages to the kind of audience we want to target. Two important criteria are the gender and the age range. For instance, some pages have 90% of their like which are from women or from people under 25 years old. It is an important point because in Tinder you can choose the gender and the age range of your target.

Finally, if we choose pages with audiences which are not so similar, we estimate that 15 interests are sufficient to reduce the results to 1 output. We can also do a function that map to each first name an estimation of the minimal number of page required to reduce the number of result to one output.

4 The Facebook Query Language 2

This query language uses Facebook Graph Search which enables to ask complicated query on the Facebook query. From the search bar, Facebook Graph Search is available to the user only in the US, but it is actually possible to use Facebook Graph Search with the url.

For instance, if you want to sees photos liked by you this week, you just have to type this url : **http://facebook.com/search/me/photos-liked/this-week/date/photos/intersect**. The FQL2 language get an understandable code as input and output a link like this one.

Here, we are going to use some possibilities offered by this language and formilize them.

A view of users with a particular first name : In FQL2, we write `users named "name"`, the corresponding link is **str/name/users-named**.

A view of likers of a particular page : In FQL2 we write `id(page_id)`$\rightarrow$ `likers`, the corresponding link is **page_id/likers**

A view of users having a particular job : In FQL2 we write `pages named "job_name"` $\rightarrow$ `employees`, the corresponding link is **str/job_name/pages-named/employees**

A view of users from a particular profession : In FQL2 we write `workers (profession_code)`, the corresponding link is **profession_code/job-liker-union/employees**

An intersection of same-type views : In FQL2 we write $\texttt{view}_1$ `INTER` $\texttt{view}_2$, the corresponding link is **view$_1$/view$_2$/intersect-2**. To be more understandable, we will write it $V_1 \bigcap V_2$.

An union of same-type view : In FQL2 we write $\texttt{view}_1$ `UNION` $\texttt{view}_2$, the corresponding link is **view$_1$/view$_2$/union-2**. To be more understandable, we will write it $V_1 \bigcup V_2$.

variable corresponding to a view : In FQL2 it is `VIEW ?view_name = view_code` then we can use the view `?view_name` in following queries/view declarations.

5 Attacks

5.1 Past attacks

In several papers about privacy on social network [2], the main attack was reverse search on Google Image. This attack seems to don't work anymore.

Another attack presented on the paper [2] seems to be a little more efficient : it consists in just getting keywords appearing on the Tinder profile and search them together on Google. We were able to link less than 10% of Tinder profiles to a last name with this attack.

We tried these attack on 30 women profiles, none of them where found with the "search by image" tool powered by Google. In 4 cases the person linked an Instagram account (and it often contains the last name). In 2 other cases we found the LinkedIn profiles of the person with the job and studies keyword and the first name. And in one other case we found the Facebook profile because the name of an associations appeared on one picture and the person seems to be part of this association (See Figure 4). However, this attack does not work really well and seem harder to implement because there is often a lot of results when you search by keywords.

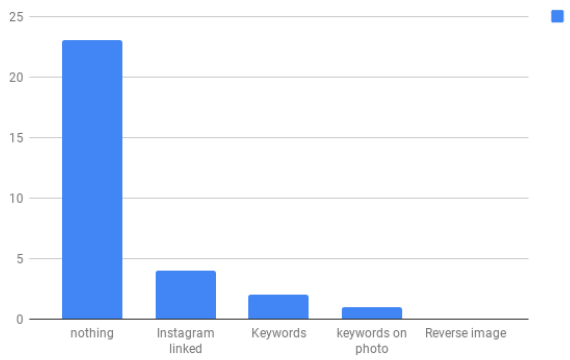

Figure 4: Statistics on 30 Tinder profiles

Reasons why this attack does not work anymore is probably the recent changes on Google reverse search algorithm, or maybe the improvement in security of social networks against indexing bots.

There exists other attacks presented on [3], [4] which seems to have been solved, but anyway attacks we will present here are available for any stalker with a Facebook account.

5.2 Stalker attack

The first attack we will present could be done manually by every stalker. It enables to find Tinder users on Facebook. We didn't save any personal data during the following experiments.

The protocol is the following : First of all, we need to create a Facebook account and localize it at the city where we want to do the attack (here, the city was Lille, France). In a second time, we liked with this profile 100 popular/rising Facebook pages which might be liked by many people of the target age-range/localization. We use for that website with data about Facebook pages popularity. (https://www.socialbakers.com/).

Indeed, we chose the pages according to 3 criteria : First, the page must have between 100K and 5M likes on Facebook. If there is more, then it will not reduce the number of results as much as wanted. If there is less, the probability that we find a user who liked this page is too low. Second, the page must be specific to a country (For instance, famous singer of the country), and if possible, a state/county (For instance a local newspaper). Third, the pages must be in an upward slope, because Tinder uses the last 100 pages liked, so the page of a famous singer who did nothing since 2012 will be useless.

Actually, we can sometimes Notwithstanding those rules with around 10% of the pages liked. If we add pages with more than 5M likes, we will probably have more users with common interest and if we add pages with less than 100K likes, maybe they will not appear often, but they will be useful when appeared. If we use only local pages, tourists/foreigners will not be found. Finally, we have to take in account that it is the last 100 pages liked when the Tinder account was created, so if one create his account in 2012, the interest shown will probably be famous pages of 2012.

When the Facebook account was ready, we created a Tinder account linked to this Facebook account. Meanwhile, a script get the Facebook id of our 100 liked pages and create one FQL view for each : `VIEW ?page_name = id(123456789)->likers;`

Once the Tinder account is created, we can begin to link Tinder profile to Facebook account : For each Tinder profile shown, we looked at the common interests (i.e. common Facebook pages liked, see Figure 5). If there were no common interests, we did nothing (even if the user was probably vulnerable, if he linked his Instagram account for instance). If there were at least one common interests, then it mean that the user linked his Facebook account to his Tinder account. We can try to find him.

To find the user, we need his first name (public), his gender (public) and every common interests (sometimes private). With those information we can then do a Facebook query like this one : `SEARCH users named "Alice" INTER women INTER ?stromae INTER ?macron INTER ?griezmann;`. A can also add job and studies if they are given.

Formally, we write

$$\texttt{SEARCH } user\_named(\text{"}name_{user}\text{"}) \cap \Big(\bigcap_{i \in interests_{user}} i->likers \Big)$$

If the target gives his location to Facebook, then it will probably appear on top of the results, making the work of the stalker easier. If their is too many results for a human stalker, then we can also add a selection on the job, studies and city (but their is more probability for those information to be private)

If the target does not appear in the results, then he probably secure his account on one of the criteria used. Another explanation can be that the user unliked this page since he create his account. But if the first name is not common, the stalker might find the target even if he does not have information on the target's interests.

Some criteria enable to reduce significantly the number of results : an uncommon first name, an uncommon interest, several incompatible interests (for instance `?donaldtrump` and `?hillaryclinton`).

5.3 Data gathering

Now that we know it is possible for a human to stalk manually Tinder users, what about bots ? Actually, we can totally imagine a script which link each Tinder account with a Facebook account.

For that, we need to create hundred of Facebook account and like 100 different Facebook pages for each tinder account. With 100 account we can cover 10000 Facebook pages (we know for instance that only 417 french Facebook pages exceed one million likes). In a second time, we create a Tinder account for each Facebook account.

For each profile shown, our bot will save in a database the tuple (id, first name, age, first photo, bio) of the user and in another table it will save interests of each user and gather interests collected by each fake account.

When we have more than 10 interests for one user, we can do a query to find this user on Facebook. The

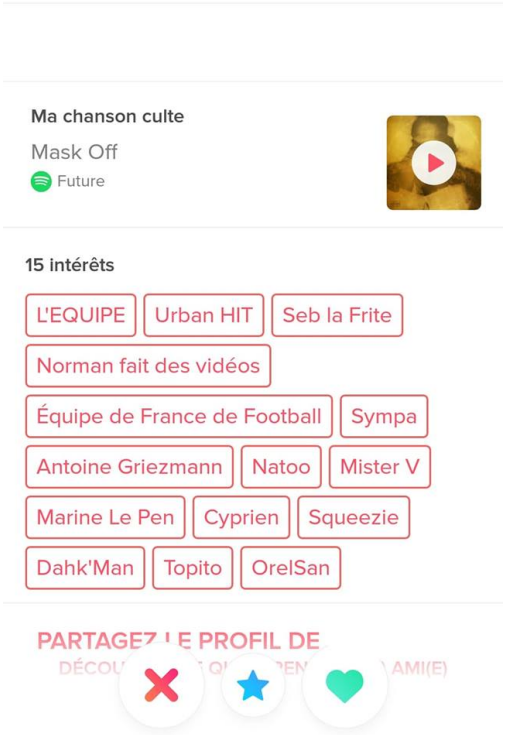

Figure 5: Example of a profile with many Facebook interest

probability that there is a false positive is really low (see section 7), so if there is one result, we can assume we have find our user.

If there is more than one result, then we have a false positive, so we can wait to have more interests for this user.

If there is no results, either the user have secured his account, either he had unlike some pages and we can try with a subset of the interests we have.

If we want to find n users on Facebook and we know for each user enough interests to avoid false positives, we can minimize the number of queries by doing the union of all the queries :

$$\texttt{SEARCH} \bigcup_{user} query_{user}$$

5.4 Who would have an interest to do that ?

Actually, it is easy to found an interest on those information. First of all, a firm can get a database of Facebook user who have or had a Tinder account. Then they can sell those information. They can also associated information appearing on the Tinder account to the user and sell those information.

It also enables stalker to found their target really easily.

We can also imagine that once you have associated a Tinder account to a identity, thanks to the geographic location you can have the location of someone at any time. [7]

A more fanciful scenario can be that a firm create IA with attractive profiles. When they have a match, they will ask questions about him/her, like if it was part of the seduction game but it is just to collect data.

6 Results of experiments

We did two slots of the stalker attack presented in section 3.2.

6.1 First experiment

In the first experiment we create a profile with gender male, interested in females. No city added on Facebook, no profile picture and 100 pages liked (see References). We did the experiment on 300 profiles shown.

117 of the 300 profiles shown had at least one common interest with those I chose for my fake account. 37 had more than four (See Figure 6).

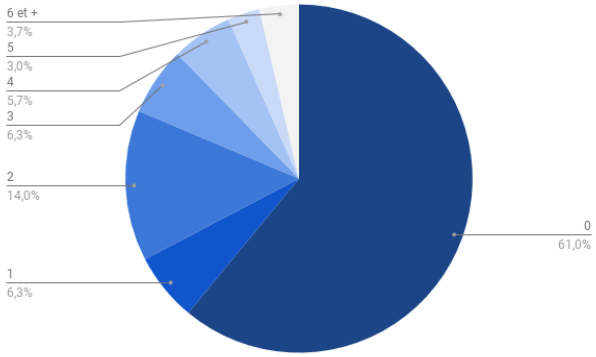

Figure 6: Number of user in term of number of common interests

38 of those 117 profiles were easily find (32 %). At the end, we find the Facebook id of 12% of the profiles shown (See Figure 7).

78 of the 100 Facebook pages appeared at least one time and 35 of them appeared on at least 1% of the profile shown.

Actually, 12% is not bad, but we made many mistake in this first experiment : The interests were not good enough and we forget to provide a city of residence to Facebook, in order to obtain people living near on the top of the results.

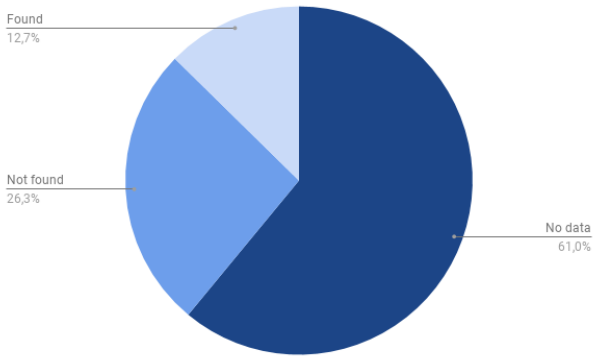

Figure 7: Statistics on the 300 profiles shown

6.2 Second experiment

In our second experiment, we create a profile with the gender female interested in males. We add the city where we did the experiment as residence city and 100 new Facebook interests (see References)

144 of the 300 profiles shown had at least one common interest (27 more than in the first experiment) and 63 of them had at least four (26 more than in the first experiment). See Figure 8.

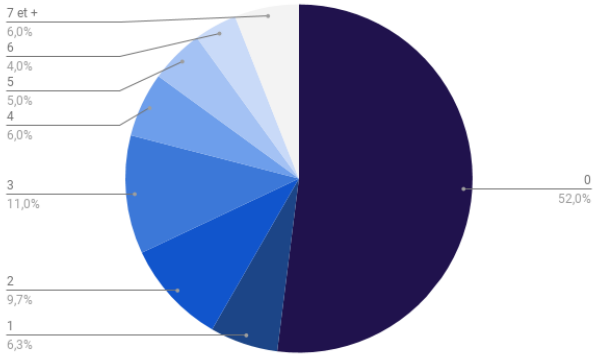

Figure 8: Number of users in terms of number of common interests

87 of the 144 profiles were easily found (60%), 33 were not found because of a there were too many results (23%), 18 were not identifiable (13%) and in 6 other case there were a doubt on the identity (4%). See Figure 9.

only 4 of the 30 profiles with at least 6 common interests where not found (success rate of 86%). See Figure 10.

Finally, we were able to find 29% of the profiles shown.

83 of the 100 pages appeared at least one time

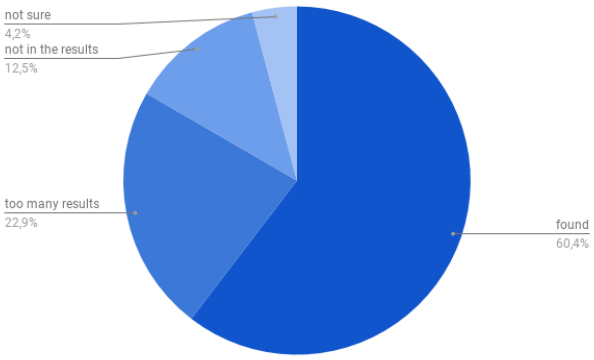

Figure 9: Statistics on the 144 profiles with common interests

Common Interests	Users	Found	Too many results	Not ident -ifiable	Doubt
0	156				
1	19	10	8	0	1
2	29	15	11	2	1
3	33	16	9	8	1
4	18	12	3	2	0
5	15	8	1	4	2
6	12	9	1	2	0
7	2	2	0	0	0
8	4	4	0	0	0
9	3	3	0	0	0
10	2	2	0	0	0
11	4	3	0	0	1
12	1	1	0	0	0
13	0	0	0	0	0
14	1	1	0	0	0
15	1	1	0	0	0
total	300	87	33	18	6

Figure 10: Statistics on the 300 profiles shown

6.3 Analysis

The first finding is that with an account with 100 interests cleverly chosen and an account localized at the good place, it is really easy to find a Facebook account linked to a Tinder account.

We obtained for every experiment less than 50% with at least one common interest. There must be various explanation to this. First of all, there is a large portion of bots and other fake account on Tinder (see [6]). Those fakes users are often not linked to a Facebook account. Some other users did not link their Facebook account to their Tinder account, some did not give the permission to Tinder to see their interests (but we assume few people do this). Finally, many users does not have common in-

terests with our account (which have only 100 interests).

Among the users with common interests, we were not able to find at least 30% of users. There is some possible explanation : Sometimes there were too many results and if the profile picture on Facebook is not the one of the Tinder account it is really hard to find the target. Sometimes Tinder photos or Facebook photos are too unclear to create a real link between the two. And in other cases, people secure the access to their interests.

We assume that the number of people who secure the access to their interest is low, because this setting does not appear in general settings of our Facebook account so people have to search deeply to find it.

With several pages, we queried to Facebook users who liked this page and we did the ratio between the number of results we obtain and the number of likes actually shown on the page. We obtain a number of results between 70% and 80% of the total amount of likes on the Facebook page.

We did another experiment to estimate more precisely access control policy of Facebook users : We did the same query on the friends of one user with (1) this user account, (2) an account of a friend of this user which have not other friends and (3) with an account without any friends. What we see is that between 70% and 75% of Facebook users let their studies,jobs and current city public (See Figure 11 to 13).

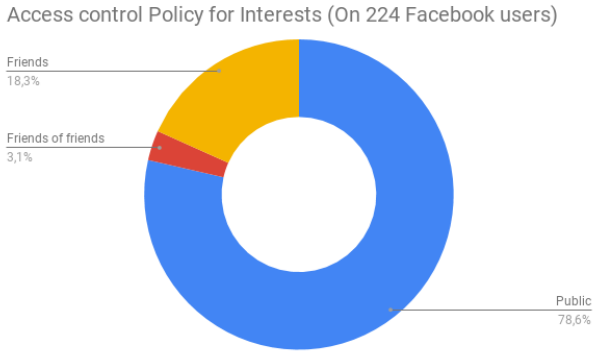

Figure 11: Access control policies for interests

7 Facebook and Tinder protections

It is not so easy to create a Facebook account (you need a valid e-mail address, sometimes you have to send an original photo) but we think that a clever script can overpass that.

Same thing to create a Tinder account, you need a phone number, but you can use the same number several

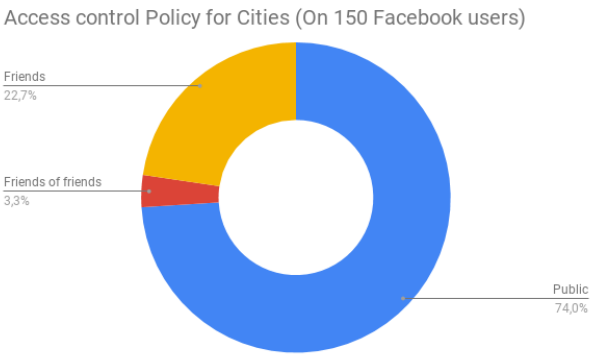

Figure 12: Access control policies for cities

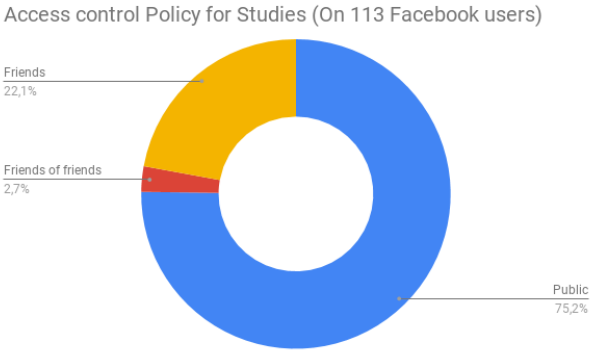

Figure 13: Access control policies for studies

times. Moreover, there exists a lot of ”disposable phone number” on the net which are not detected by Tinder.

Facebook limit the number of search query per day and per account. But there is many way to overpass it : By doing an union of query, with the mobile version of the site (m.facebook.com) or with a lot of accounts.

The number of like par day and per account is also limited by Facebook. But since you do only 100 likes with each account it is not a problem.

8 How to protect yourself from those attacks

The best thing to do is to not link your Facebook account to your Tinder account. In any case, you will need to give a phone number so it does not simplify the registration.

If you really want to use Facebook, don’t allow Tinder to see your interests, and set restrictive access controls on your Facebook account for pages liked, residence city, jobs and studies.

References

[1] A Tool to find your match on Tinder.

[2] C. STENSON, A. BALCELLS, M. C. Burning up privacy on Tinder.

[3] CHOO, M. C.-K. R. Tinder me softly – how safe are you really on tinder?

[4] JODY FARNDEN, BEN MARTINI, K. K. R. C. Privacy Risks in Mobile Dating Apps.

[5] T. DELEMAZURE, P. BOURHIS, R. R. W. R. Documentation Facebook Query Language 2.

[6] TAHORA H. NAZER, FRED MORSTATTER, G. T. H. L. A Close Look at Tinder Bots.

[7] VEYTSMAN, M. How I was able to track the location of any Tinder user. .