**REGULAR ARTICLE**

**Open Access**

CrossMark

# Collective aspects of privacy in the Twitter social network

David Garcia[1,2]* , Mansi Goel[3], Amod Kant Agrawal[3] and Ponnurangam Kumaraguru[3]

*Correspondence: garcia@csh.ac.at
[1]Section for Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, Vienna, 1090, Austria
[2]Complexity Science Hub, Josefstaedterstrasse 39, Vienna, 1080, Austria
Full list of author information is available at the end of the article

**Abstract**

Preserving individual control over private information is one of the rising concerns in our digital society. Online social networks exist in application ecosystems that allow them to access data from other services, for example gathering contact lists through mobile phone applications. Such data access might allow social networking sites to create shadow profiles with information about non-users that has been inferred from information shared by the users of the social network. This possibility motivates the shadow profile hypothesis: the data shared by the users of an online service predicts personal information of non-users of the service. We test this hypothesis for the first time on Twitter, constructing a dataset of users that includes profile biographical text, location information, and bidirectional friendship links. We evaluate the predictability of the location of a user by using only information given by friends of the user that joined Twitter before the user did. This way, we audit the historical prediction power of Twitter data for users that had not joined Twitter yet. Our results indicate that information shared by users in Twitter can be predictive of the location of individuals outside Twitter. Furthermore, we observe that the quality of this prediction increases with the tendency of Twitter users to share their mobile phone contacts and is more accurate for individuals with more contacts inside Twitter. We further explore the predictability of biographical information of non-users, finding evidence in line with our results for locations. These findings illustrate that individuals are not in full control of their online privacy and that sharing personal data with a social networking site is a decision that is collectively mediated by the decisions of others.

**Keywords:** privacy; online social networks; location

## 1 Introduction

Since the leaks of the National Security Agency global surveillance by Edward Snowden [1], privacy in online activity has been one of the rising concerns for Internet users [2]. While these concerns date back nearly two decades [3] and have not led to wide use of privacy-enhancing technologies [4], the topic of privacy rights in online activity is higher than ever on political agenda and media attention. Sharing private information can be motivated by services or information received in exchange, for example in the case of sharing health information with a doctor. Nevertheless, this does not need to be the case in online social networks: A 2016 Pew Research Center survey [5] showed that more than 51% of respondents consider it *not acceptable* to share private information with an online

Springer

social network that shows personalized advertisement, in fear of third parties accessing such private data.

Social networking sites do not exist in isolation in cyberspace, they exist in the *Platform Society* [6] and can integrate information from other sites and applications. For several years, more than a billion users of the Facebook mobile applications [7] have given permissions to Facebook to read their phone contact lists.[a] This motivated the protest of the Europe-vs-Facebook advocacy group about Facebook building *shadow profiles* [8, 9]: hidden files on individuals with private information that has been inferred through the individual's friends inside the social network. These shadow profiles can potentially be built for non-users without an account in the social networking site and that did not agree to its privacy policy [10]. The possibility to build shadow profiles would pose an important concern with respect to privacy rights and informational self-determination [11].

Previous research has extensively evaluated privacy risks for users of social media, in general by evaluating the inference of personal attributes of users from their digital traces [12]. For example, Twitter data can be used to predict user locations [13, 14] as well as gender, age, and political orientation of users [15]. Publicly available information in location-based social networks, such as Foursquare, can also be used to predict the home location of users [16]. Facebook has also been shown to be extremely informative of user personal information, including sexual orientation [17], romantic partnerships [18], and a wealth of other private attributes that can be inferred from user "likes" [19, 20]. These inferences are possible thanks to the patterns of human interaction, such as the assortativity of personal attributes in social networks [15]. These patterns can be so strong that personal data can be inferred using only information given by the friends of a user and not the user itself. As an example, a recent study shows how Twitter user attributes can be inferred using only text produced by the social contacts of users [21].

While the above mentioned research has shed light on various properties and risks to privacy, it has generally been limited to the analysis of actual *users* of social networks. For this reason, their generalizability to the question of shadow profiles is limited. The scarcity of research on shadow profiles is due to the difficulty to generate inferences on attributes of non-users, but some exceptions provide a background for the topic. Horvát et al. [22] were the first ones to empirically evaluate if information about non-users could be inferred from online social networks. Combining samples of Facebook data with simulations of Facebook's growth, they could verify that friendship links between people outside Facebook could have been predicted with data contained in Facebook. In the same line of research, Sarigöl, Garcia, and Schweitzer [23] used the temporal sequence of users joining a shutdown social network, Frienster, to verify that the sexual orientation of non-users could have been predicted using data from Friendster. Furthermore, a recent article formulates and empirically tests the *shadow profile hypothesis* [10], i.e. that personal information of non-users can be predicted using information given by the users of an online service. More precisely, [10] shows that predictions of sexual orientation and marital status in Friendster improved with the amount of data shared by the users of the social network. That result illustrates that the decision of individuals not to share information is mediated by the decisions of others.

While the above articles provide evidence supporting the shadow profile hypothesis, they suffered certain limitations. First, the lack of precise data required the use of growth simulations or other heuristics [22]. Second, observational data of users did not contain

any information about which users share their contact lists, requiring certain assumptions in the analysis process [23]. And third, previous evaluations are either supported on small datasets or on data from shutdown social networks [10], leaving open the question of whether shadow profiles can be built in a current and large online social network. This article aims at overcoming those three limitations in a study of personal information in the Twitter social network, through the high-quality data provided by the Twitter API. As a result, we test shadow profile hypothesis in an active and large online social network, using the precise time sequence of Twitter users joining the network, and identifying which users share their contact lists as revealed in the metadata of their tweets.

In the following, we present a dataset that we produced to evaluate the shadow profile hypothesis when predicting user location. We use that data to evaluate the shadow profile hypothesis and analyze how the quality of location prediction depends on the tendency of users to disclose information and on the number of friends that a non-user has in Twitter. We continue by testing the shadow profile hypothesis for simplified features of user biographical texts. This analysis not only has the potential to robustly test the shadow profile hypothesis in a current social network, but also explores possible inequalities in the accuracy of shadow profiling and the collective aspects of privacy decisions in our current digital society.
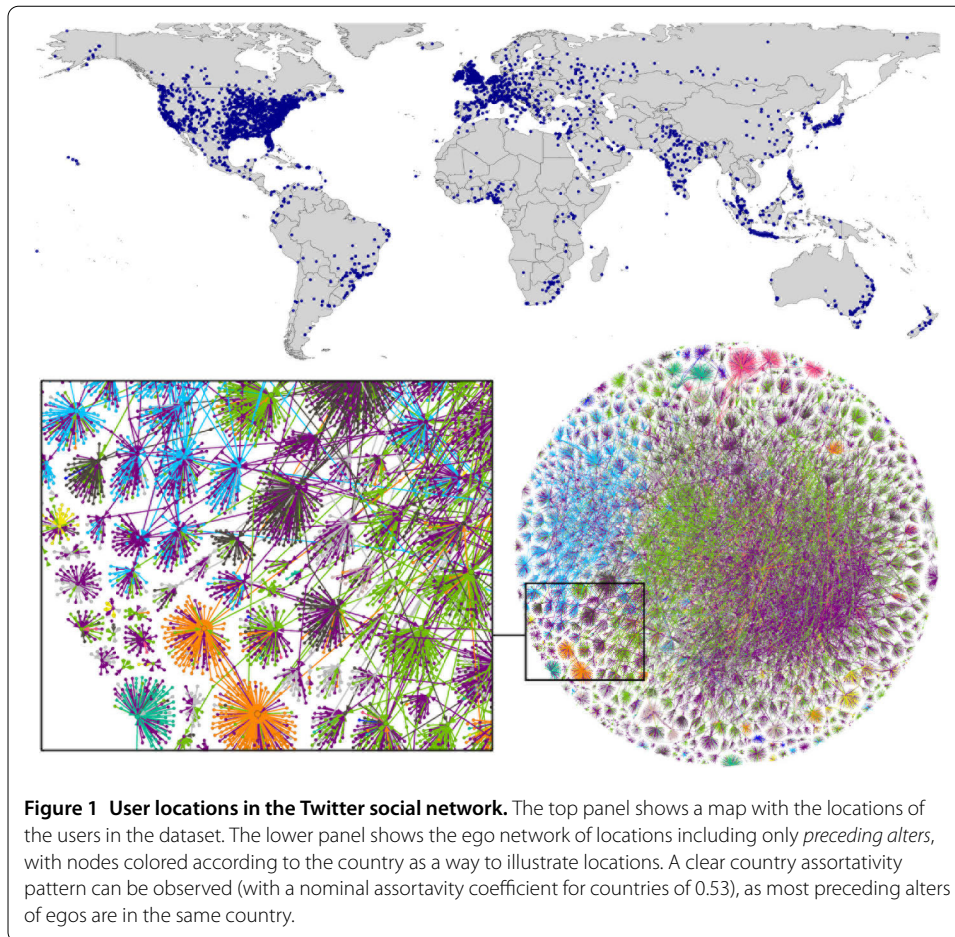
## 2 Data and methods

### 2.1 Ego network construction

We started our data collection by producing a set of ego users whose information will constitute the ground truth to evaluate predictions. To generate an initial unbiased random sample of users, we applied the Random Digit Search method [24, 25]: We generated random Twitter user ids in the range between 1 and 30 Billion, looked them up through the Twitter REST API,[b] and saved the basic user information of the valid sampled users. To avoid celebrities and spammers, we filtered out users with a ratio of followers to friends below 0.1 or above 10, as well as users with less than 50 friends or followers. To have a homogeneous sample for biographical data analysis, we included only users that have English as the language of the Twitter account. This process generated a set of 1,017 ego users, which are the starting point of a larger dataset including their social contacts and their activity in Twitter.

We collect the timeline of tweets of each ego user up to 3,200 tweets.[c] Based on those timelines, we identify alter users as the ones that have been mentioned at least four times by an ego user, following this way a set of friendship links that capture communication and not just followership or retweeting [26]. We use these links as an approximation to the underlying social network between Twitter users that is revealed when users share their contact lists through mobile phone apps or through importing tools. This way we generate a set of 68,447 alter users, collecting also their timeline of tweets and biographical information. As a result, we count with a total of 157,408,012 tweets in our dataset from both ego and alter users.
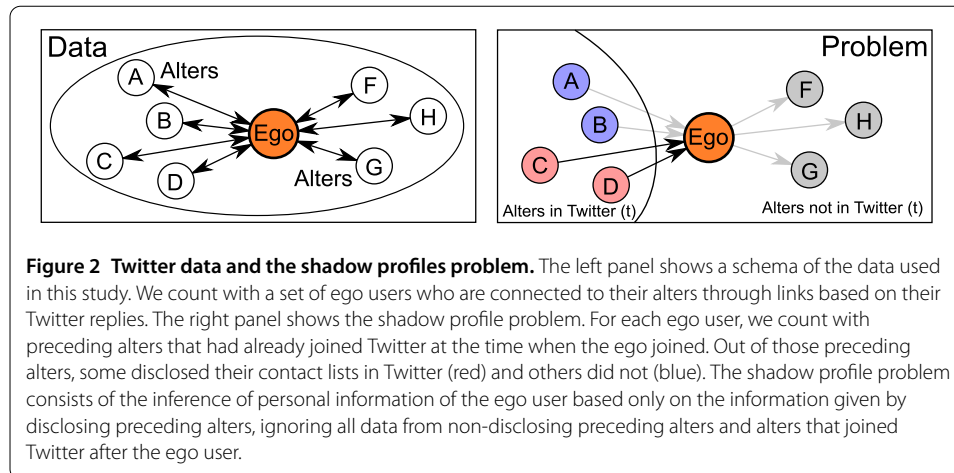
### 2.2 User analysis

We identify the location of users by combining geographical data of their tweets as well as their self-disclosed location and biographical text. Our dataset contains more than 5.6 Million geotagged tweets that contain a geographical location reference with precise coordinates. We process those coordinates with the Google Maps Geocoding API[d] to identify

**Figure 1 User locations in the Twitter social network.** The top panel shows a map with the locations of the users in the dataset. The lower panel shows the ego network of locations including only *preceding alters*, with nodes colored according to the country as a way to illustrate locations. A clear country assortativity pattern can be observed (with a nominal assortativity coefficient for countries of 0.53), as most preceding alters of egos are in the same country.

the municipality in which they are located, which we refer to as their city. For each user with at least one geotagged tweet, we label their location as the most frequent city where their tweets have been geotagged. For users without geotagged tweets, we process the location and biographical text of users through Google's Text Analysis API.[e] As a result, we located 630 ego users and 38,936 alters, taking this data as an approximation to the better location information that Twitter has access to. The top panel of Figure 1 shows the locations of users in the dataset, illustrating that users come from a wide variety of countries but are generally located in countries where Twitter adoption is high [27, 28]. The lower panel of Figure 1 shows the ego network using only preceding alters as explained below, with nodes colored according to the user country. A clear assortativity pattern can be observed, which is the foundation of the unsupervised predictors explained below.

We processed the biographical text provided by each user in the dataset by removing stop words that appear in the NLTK stopword list[f] and stemming its tokens with Porter's stemming algorithm [29]. For our analysis, we consider only those users which have 3 or more tokens in their biographical text after this step (49,576 alters and 676 ego users). Over these texts, we applied a 100-dimensional Doc2Vec [30] model trained on a separate corpus of 1.7 Million biographical texts generated in previous research [31]. Doc2Vec fits a language model that represents documents (in this case bios) as vectors such that semantically and linguistically similar documents are close in the representation space. We chose a dimensionality of 100 to follow previous applications of similar models [32]. To

**Figure 2 Twitter data and the shadow profiles problem.** The left panel shows a schema of the data used in this study. We count with a set of ego users who are connected to their alters through links based on their Twitter replies. The right panel shows the shadow profile problem. For each ego user, we count with preceding alters that had already joined Twitter at the time when the ego joined. Out of those preceding alters, some disclosed their contact lists in Twitter (red) and others did not (blue). The shadow profile problem consists of the inference of personal information of the ego user based only on the information given by disclosing preceding alters, ignoring all data from non-disclosing preceding alters and alters that joined Twitter after the ego user.

further reduce the dimensionality of the dataset, we applied Principal Component Analysis (PCA) and took the two most informative components as a quantification of the content in biographical texts. More details about the PCA are presented in the Additional file 1. As a result, we count with a 2-dimensional biographical vector for each user such that semantically similar texts will have similar orientations of their vector representations and dissimilar texts will be pointing in very different directions.

The Twitter API provides a source field for each tweet that identifies the way the tweet was produced. Among these sources we can find mobile phone applications for Android and iPhone, allowing us to identify wich users installed one of this applications and shared their contact lists with Twitter.[g][h] We mark as *disclosing alters* all the alters that produced at least one tweet with the source "Twitter for iPhone" or "Twitter for Android". This way we identify 54,658 disclosing users (934 ego users and 53,724 alters), which amount to more than 78% of the users in our dataset.

### 2.3 The shadow profile problem

We adapt the problem formulation of shadow profiles for Facebook [22] and Friendster [10] to the case of Twitter. The left panel of Figure 2 shows a schema of the ego-centered data we use: we count with the connections between ego and alter users and the location and biographical vectors for the users that shared that data on Twitter. The right panel of Figure 2 shows the problem of constructing a shadow profile for the ego user, in which only a historical subset of the data is used to evaluate if the information provided by users (alters) was predictive of non-users in the past (ego users).

In the shadow profile problem, all alters that joined Twitter later are excluded in the prediction of ego data, since the information of these alters was not available to Twitter before the ego created an account. Out of the alters that joined Twitter before the ego user, only a subset of them were *disclosing alters*, i.e. they shared their contact lists with Twitter, for example through the mobile phone app. The shadow profile problem in Twitter is to generate predictions of the location and biographical vectors of ego users based only on the information of their disclosing alters who joined Twitter before. Therefore, we go over the history of Twitter and evaluate predictions only based on data that was available before the ego user joined.

We study the conditions that drive the quality of shadow profiles in two analysis scenarios. First, we perform an *empirical shadow profile analysis* by applying the predictors

explained below over the set of disclosing alters found through the source of their tweets. We evaluate the predictions against the ground truth of ego users and compare against a Null Model to test the shadow profile hypothesis. Furthermore, we analyze the relationship between the quality of predictions and the number of disclosing alters of each ego user with the purpose to evaluate if profiles are more accurate for users with more friends who already joined Twitter.

Second, we perform a *disclosure tendency analysis* to study how the tendency of users to share their contact lists can affect the quality of shadow profiles. In this analysis scenario, instead of using the set of disclosing alters, we randomly sample subsets of all alters that joined Twitter before the ego user. We define the disclosure parameter $\rho$ as the probability that an alter shares its contact lists with Twitter, to analyze how the quality of shadow profiles depends on disclosure tendencies. For each value of $\rho$ between 0.1 and 0.9 in increments of 0.1, we generate 1000 samples and generate predictions based on that subset of the data. In addition, we record the number of alters sampled this way to evaluate if any relationship between prediction quality and number of friends also appears in this sampling scenario.

## 2.4 Unsupervised predictors and evaluation

We apply two unsupervised predictors for location and biographical vectors to evaluate the shadow profile hypothesis on Twitter. To predict the location of ego users, we take the locations of all disclosing alters and identify the most frequent city among alters, i.e. the modal predictor. We use this location as the unsupervised prediction of location to be compared against the ground truth of the location of the ego user. We evaluate the quality of the prediction by measuring the Haversine distance in Km between the predicted point and the ground truth. We predict the biographical vector of each alter as the average vector of its disclosing alters. We evaluate this prediction through the cosine similarity of predicted and ground truth vectors. Therefore, a high similarity will mean a high accuracy of the predictor.
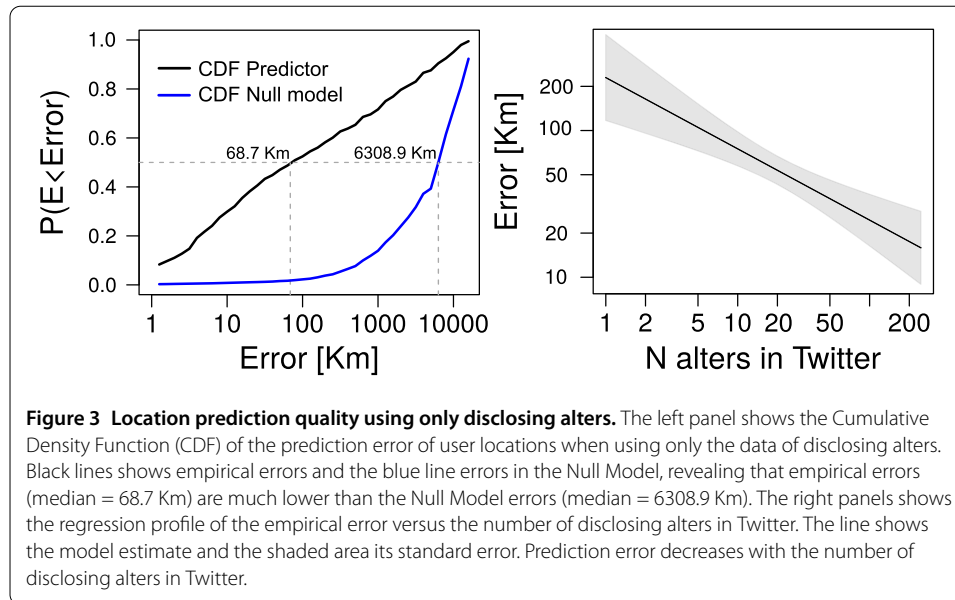
We compare both predictors against a Null Model that takes a uniformly random sample of all users to construct a prediction. For each prediction of the model, we generate 100 Null Model predictions by sampling the same number of users from the whole dataset. By comparing the Null Model with the shadow profile predictions we ensure that our results are not an artefact of limited data samples or uneven distribution of locations and biographical data.

## 3 Results

## 3.1 Predicting locations

We first evaluate the predictive power of the data shared by *disclosing alters* in the empirical shadow profile analysis exercise explained above. Disclosing alters are friends of an ego user who joined Twitter before that user and who shared their contact lists with Twitter through a mobile app. Given this sample of users, we construct a historical shadow profile for the location of each ego user based only on the data provided by its disclosing alters, as explained in the Data and Methods section.

The left panel of Figure 3 shows the Cumulative Density Functions (CDF) of the error in Km of the shadow profile predictor and of the Null Model. The shadow profile model clearly outperforms the Null Model: the median error of the shadow profile predictor is
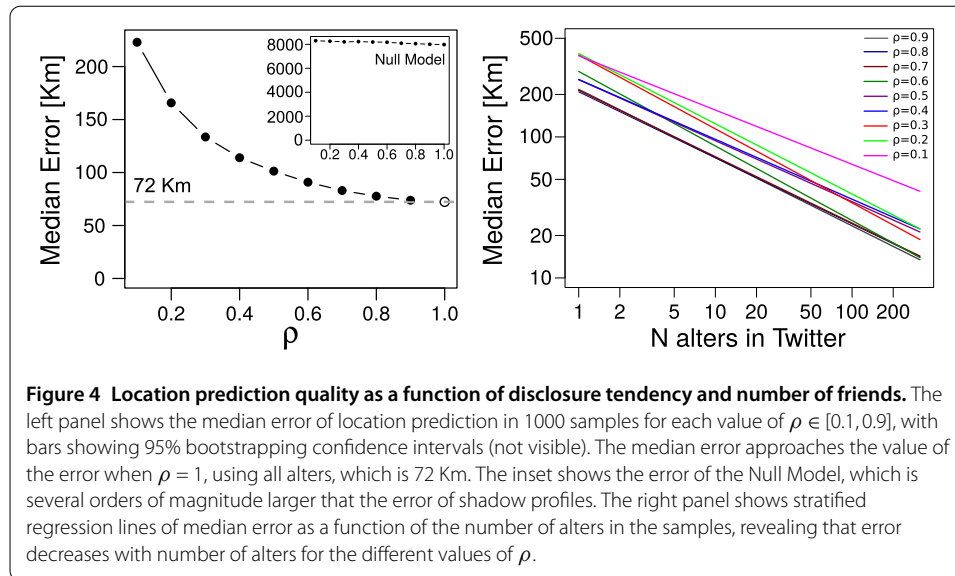
**Figure 3 Location prediction quality using only disclosing alters.** The left panel shows the Cumulative Density Function (CDF) of the prediction error of user locations when using only the data of disclosing alters. Black lines shows empirical errors and the blue line errors in the Null Model, revealing that empirical errors (median = 68.7 Km) are much lower than the Null Model errors (median = 6308.9 Km). The right panels shows the regression profile of the empirical error versus the number of disclosing alters in Twitter. The line shows the model estimate and the shaded area its standard error. Prediction error decreases with the number of disclosing alters in Twitter.

68.7 Km while the median error of the Null Model is several orders of magnitude larger (6308.9 Km). Furthermore, when comparing the names of cities as a binary prediction, we find an accuracy of 32%, while the Null Model has an accuracy near zero. These results lend support to the shadow profile hypothesis for location in Twitter, as the information of users is predictive of the location of non-users.

We analyze how this predictive power varies across users, testing a relationship between the shadow profile error and the number of disclosing alters in Twitter. The right panel of Figure 3 shows a regression profile of the logarithm of prediction error versus the logarithm of the number of disclosing alters of each user. There is a significant negative association between both variables ($\beta = -0.486$, $p < 0.05$, more details in the SI Table 1), which is also confirmed in when computing Spearman correlation ($\sigma = -0.155$, 95%CI = $[-0.232, -0.076]$). This confirms that the error of the shadow profile predictor for location in Twitter decreases monotonically with the number of friends of a user who already are in Twitter and share their contact lists.

## 3.2 Disclosure tendency analysis of locations

We analyze the dependence of the quality of shadow profiles for location as a function of the disclosure of contact lists, sampling alters as disclosing users for increasing values of the disclosure parameter $\rho$. The left panel of Figure 4 shows the median prediction error averaged over 1000 user samples for each value of $\rho$, with an inset showing the equivalent for the Null Model. It is clear that the same observation as above holds here: the error of the shadow profile prediction is much lower than the error of the Null Model, even for low values of $\rho$. Median errors decrease monotonically with $\rho$, which is confirmed by the Spearman correlation coefficient ($\sigma = -0.06$, 95%CI = $[-0.086, -0.037]$). This supports the hypothesis that, as more users share information in the social network, the predictor accuracy increases.

We analyzed the relationship between prediction error and the number of alters in Twitter in this analysis scenario. A linear regression model of log-transformed variables over all samples shows a negative association ($\beta = -0.48$, $p < 0.001$, more details in SI Table 2),
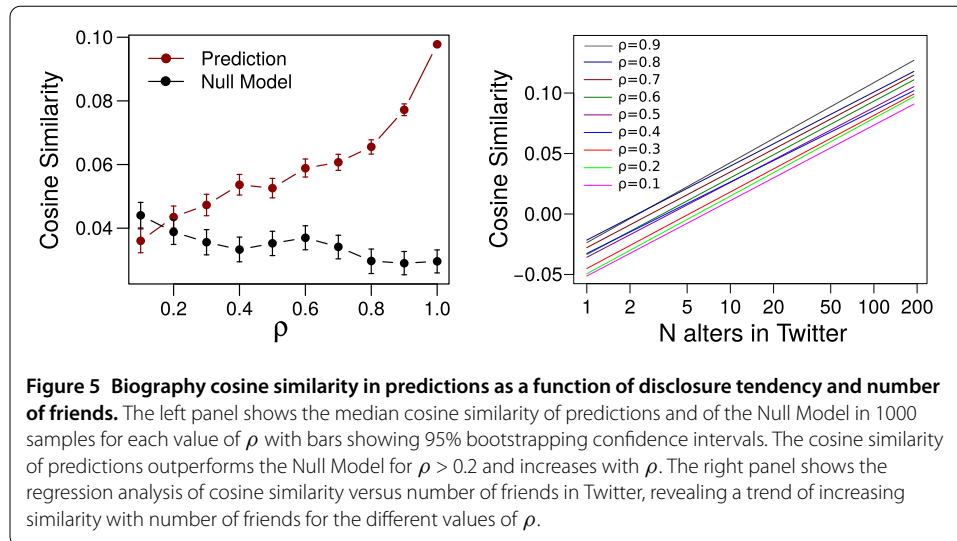
**Figure 4 Location prediction quality as a function of disclosure tendency and number of friends.** The left panel shows the median error of location prediction in 1000 samples for each value of $\rho \in [0.1, 0.9]$, with bars showing 95% bootstrapping confidence intervals (not visible). The median error approaches the value of the error when $\rho = 1$, using all alters, which is 72 Km. The inset shows the error of the Null Model, which is several orders of magnitude larger that the error of shadow profiles. The right panel shows stratified regression lines of median error as a function of the number of alters in the samples, revealing that error decreases with number of alters for the different values of $\rho$.

consistent with the results using only disclosing alters of the previous section. Furthermore, this result holds across different levels of disclosure tendencies, as captured by the values of $\rho$. The right panel of Figure 4 shows regression profiles of the median error of the predictor versus the logarithm of the number of alters in Twitter for the nine values of $\rho$. All profiles have a negative trend, i.e. more alters are associated with lower error, being consistent with a model that includes an interaction term between $\rho$ and the logarithm of the number of alters ($\beta = -0.35$, $p < 0.001$, more details in SI Table 2). In particular, the interaction term is negative and significant ($-0.24$, $p < 0.01$), showing that the negative association between prediction error and number of friends becomes stronger for larger values of $\rho$.

### 3.3 Biographic vector prediction

We study the quality of unsupervised predictions of biographical vectors in the empirical shadow profile analysis, i.e using only information from disclosing alters. In this scenario, the cosine similarity between the predicted vector and the ground truth is not higher than the Null Model when compared over the whole set of ego users (Wilcoxon test $p$-value = 0.92, difference between medians = 0.005). On the contrary, when analyzing only ego users with an number of alters above the average, the cosine similarity of the prediction is significantly higher than the Null Model by a sizeable amount (Wilcoxon test $p$-value < 0.05, difference between medians = 0.20). This indicates that the data of users in Twitter is predictive of the biographical features of users outside Twitter, but only for egos with enough friends who already joined Twitter and shared their contact lists. This conclusion is further supported by the Spearman correlation coefficient between the cosine similarity of the prediction and the number of disclosing alters ($\sigma = 0.08$, 95%CI = [0.001, 0.15]).

We performed the disclosure tendency analysis over the predictions of biographical vectors to understand how their performance depends on the values of $\rho$. The left panel of Figure 5 shows the median cosine similarity of the predictor over 1000 samples for each value of $\rho$, comparing the result against the Null Model. Cosine similarities outperform the Null Model from relatively low values of $\rho$ and increase with it. This observation is confirmed when calculating the Spearman correlation coefficient ($\sigma = 0.028$, 95%CI = [0.003, 0.052]),

**Figure 5 Biography cosine similarity in predictions as a function of disclosure tendency and number of friends.** The left panel shows the median cosine similarity of predictions and of the Null Model in 1000 samples for each value of $\rho$ with bars showing 95% bootstrapping confidence intervals. The cosine similarity of predictions outperforms the Null Model for $\rho > 0.2$ and increases with $\rho$. The right panel shows the regression analysis of cosine similarity versus number of friends in Twitter, revealing a trend of increasing similarity with number of friends for the different values of $\rho$.

indicating that higher disclosure tendencies lead to higher accuracy in the estimation of non-user biographies.

The positive relationship between cosine similarity and disclosing alters is present for changing values of $\rho$. The right panel of Figure 5 shows regression profiles for the nine values of $\rho$, displaying a positive trend in all of them. The Spearman correlation coefficient between cosine similarity and the number of disclosing alters significant and positive ($\sigma = 0.04$, 95%CI = $[0.017, 0.068]$). Furthermore, it is consistent with the fit of a regression model of cosine similarity as a function of the logarithm of the number of alters ($\beta = 0.028$, $p < 0.001$, more details in SI Table 3), and is robust to the addition of a interaction term with $\rho$ ($\beta = 0.021$, $p < 0.05$, more details in SI Table 3). It is worth noting that the interaction term is not significant, i.e. we do not have evidence that the relationship between cosine similarity and the number of friends becomes stronger for higher values of $\rho$. Overall, these results support the shadow profile hypothesis for biographical data in Twitter, evidencing that features of the description of non-users can be predicted with data from users, but once that data is abundant enough.

## 4 Discussion

Our work shows that the data shared by Twitter users is predictive of personal information of individuals that are not users. We produced a dataset of the ego network of more than 1000 users, retrieving their timelines and timelines of their alters for a total of more than 150 Million tweets. Detecting users that use a mobile phone app, we could identify which users share their contact lists, and thus we provide the first empirical test of the shadow profile hypothesis on a dataset of a current social network. We found that the data shared by those users is informative in the prediction of location and approximates the biographical text of individuals that had not joined Twitter. This served as a historical audit to evaluate the shadow profile hypothesis, as Twitter had enough data to infer personal attributes of people that did not have an account at that time. Studying various disclosure tendencies in random samples of users, we found that the quality of those inferences improves with the tendency to disclose information of Twitter users. Furthermore, we analyzed the heterogeneity in the quality of these inferences and found that users with more friends with a Twitter account are subject to have more accurate shadow profiles.

While our results show that shadow profiles are a possibility, we must note that we have not found empirical evidence of their existence. Our results only show that they can be constructed, which is sufficient to put in question the control that individuals have on their personal information online. Testing if these kind of profiles are actually being built is an open question that does not subtract importance from the fact that, without oversight or collective control mechanisms, individuals have little power to ensure that they are not being profiled without their knowledge or consent.

The prediction methods we applied in our study are simple, unsupervised strategies that take straightforward averages as predictors. The error level for shadow profiles of location (68.7 Km) is comparable to error levels using full information, which are typically between 57.2 Km and 28.3 Km [14]. More advanced supervised methods are likely to improve predictions, but developing techniques to infer data of non-users carries important ethical issues that need to be addressed before such research is performed. We tested the shadow profile hypothesis by showing that inferences are informative, and for the case of location these inferences greatly outperform the Null Model. Further research can examine if it is desirable to develop more accurate methods, and whether that needs to be done with user consent.

Our work suffers a series of limitations that need to be taken into account when generalizing. First, we performed a historical audit using future data as ground truth. While this can test the shadow profile hypothesis, we can only fully understand the risks it conveys when producing predictions of people that have never been users of an online service. This could be done combining user contact information, which is often proprietary, with factual data from non-users, which needs to be voluntarily provided by non-users for research purposes. Second, we have relied on a heuristic to infer friendships based on the intensity of interaction in Twitter. While social network arguments support this assumption [26, 33], future research should aim at accessing friendship lists or name generators that do not depend on online interaction. Third, we used a model for user biographical texts that does not allow a straightforward interpretation of what biographical qualities are being predicted. While this allows us to address the shadow profile hypothesis, larger user samples can quantify individual demographic markers in user biographical texts [21]. And finally, our analysis is based on a sample of users that might not be demographically representative. This means that, while we cannot conclude everyone can have a shadow profile, our results show that *someone* can have it. The evidence of this possibility is already a challenge to the current guarantees of the right to privacy, but generalizing these results to larger populations has the potential to reveal larger issues and risks for whole societies.

This article adds two new dimensions to the shadow profile question: location and biographical data. This adds up to the analysis of friendship links [22], sexual orientation [23], and marital status [10], but still there are many more private attributes that could be subject to inclusion in a shadow profile. For example political views, religious beliefs, and use of substances are private attributes that can pose important issues if inferred and that can be subject of future research. Furthermore, we have added Twitter as another case to previous research on samples form Facebook [22] and Friendster [10, 23]. Further research should try with other current social networks to avoid the Twitter model organism bias [34] and to ensure that our knowledge applies to the online society and not only a handful of social networking sites.

The implications of our results are clear: individuals do not have full control over their privacy and the decision not to share information with an online service is mediated by the decisions of other people. This means that we cannot conceive online privacy as a purely individual phenomenon that can be reduced to the decisions of a person. To ensure the right to privacy and of informational self-determination as democratic values, we need new legal and data management frameworks that empower users beyond their individual agency, taking into account the evidence of *collective aspects of online privacy*.

While Terms of Service and privacy policies are exclusive contracts between a user and the owners of an online service or social network, our results show that there are clear *data externalities* that affect other people. When creating an account and sharing information, we inadvertently share information about others [35], effectively affecting their privacy. The analogy that data is the new oil and not the new gold fits well this situation [36]: data does not have just intrinsic value, but also can generate costs and harm to people that do not directly benefit from it.

## Additional material

Additional file 1: Additional information about statistical analyses. (pdf)

**Abbreviations**
NLTK, Natural Language Toolkit; CI, Confidence Interval; REST API, Representational State Transfer Application Programming Interface.

**Availability of data and materials**
All data used for this study was publicly available in Twitter. User id lists to reproduce the results of the article are available under request.

**Ethics approval and consent to participate**
All data analyzed in this study was publicly available in Twitter. Only explicitly public information was used as ground truth (biographical text, locations) and no private information of any user was predicted as part of this research. This observational study was based only on public archival data and carried no intervention and no interaction with any individual. Thus, this research carries minimum risk and is exempt of the need of consent or of an institutional review.

**Competing interests**
The authors declare that they have no competing interests.

**Consent for publication**
Not applicable.

**Authors' contributions**
DG designed the research question, MG and AA retrieved and processed the data, DG and MG analyzed data and performed statistical analysis, DG and MG wrote the final manuscript, PK provided guidance over the project and feedback on the manuscript. All authors read and approved the final manuscript.

**Author details**
[1] Section for Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, Vienna, 1090, Austria. [2] Complexity Science Hub, Josefstaedterstrasse 39, Vienna, 1080, Austria. [3] IIIT Delhi, Okhla Industrial Estate, Phase III, New Delhi, India.

## Endnotes

a  https://play.google.com/store/apps/details?id=com.facebook.katana
b  https://dev.twitter.com/rest/reference/get/users/lookup
c  https://dev.twitter.com/rest/reference/get/statuses/user_timeline
d  https://developers.google.com/maps/documentation/geocoding/intro
e  https://cloud.google.com/natural-language/
f  http://www.nltk.org/nltk_data/
g  https://itunes.apple.com/in/app/twitter/id333903271?mt=8
h  https://play.google.com/store/apps/details?id=com.twitter.android

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Greenwald G, MacAskill E (2013) NSA prism program taps in to user data of apple, Google and others. Guardian 7(6):1-43
2. Rainie L, Maniam S (2016) Americans feel the tensions between privacy and security concerns. Pew Research Center Fact Tank. http://pewrsr.ch/1mMSuYp
3. Lessig L (1999) Code and other laws of cyberspace. Basic Books
4. Preibusch S (2015) Privacy behaviors after snowden. Commun ACM 58(5):48-55
5. Rainie L, Duggan M Privacy and information sharing. Pew Research Center, January 14, 2016
6. van Dijck J, Poell T (2015) Social media and the transformation of public space. Soc Media Soc 1(2):2056305115622482
7. Smith C (2016) Facebook Mobile Stats (November 2016) http://expandedramblings.com/index.php/facebook-mobile-app-statistics
8. Knibbs K (2013) What's a facebook shadow profile and why should you care? Digital Trends
9. Blue V (2013) Anger mounts after Facebook's 'shadow profiles' leak in bug. ZDNet
10. Garcia D (2017) Leaking privacy and shadow profiles in online social networks. Science Advances 3(8). https://doi.org/10.1126/sciadv.1701172
11. Rouvroy A, Poullet Y (2009) The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy. In: Reinventing Data Protection?, pp 45-76
12. Zheleva E, Getoor L (2009) To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th international conference on world wide web, pp 531-540
13. Jurgens D (2013) That's what friends are for: inferring location in online social media platforms based on social relationships. In: Proceedings of the 7th international AAAI conference on weblogs and social media (ICWSM), pp 273-282
14. Jurgens D, Finethy T, McCorriston J, Xu YT, Ruths D (2015) Geolocation prediction in Twitter using social networks: a critical analysis and review of current practice. In: Proceedings of the 9th international AAAI conference on weblogs and social media (ICWSM)
15. Zamal FA, Liu W, Ruths D (2012) Homophily and latent attribute inference: inferring latent attributes of Twitter users from neighbors. In: Proceedings of the 6th international AAAI conference on weblogs and social media (ICWSM)
16. Pontes T, Vasconcelos M, Almeida J, Kumaraguru P, Almeida V (2012) We know where you live: Privacy characterization of foursquare behavior. In: 4th International Workshop on Location-Based Social Networks (LBSN 2012)
17. Jernigan C, Mistree BFT (2009) Gaydar: Facebook friendships expose sexual orientation. First Monday 14(10)
18. Backstrom L, Kleinberg JM (2014) Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on Facebook. In: CSCW, pp 831-841
19. Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci USA 110(15):5802-5805
20. Youyou W, Kosinski M, Stillwell D (2015) Computer-based personality judgments are more accurate than those made by humans. Proc Natl Acad Sci 112(4):1036-1040
21. Jurgens D, Tsvetkov Y, Jurafsky D (2017) In: Ciampaglia GL, Mashhadi A, Yasseri T (eds) Writer profiling without the writer's text. Springer, Cham, pp 537-558
22. Horvát EÁ, Hanselmann M, Hamprecht FA, Zweig KA (2012) One plus one makes three (for social networks). PLoS ONE 7(4)
23. Sarigol E, Garcia D, Schweitzer F (2014) Online privacy as a collective phenomenon. In: Proceedings of the second ACM conference on Online social networks (COSN), pp 95-106
24. Zhu J., Mo Q, Wang F, Lu H (2011) A random digit search (rds) method for sampling of blogs and other user-generated content. Soc Sci Comput Rev 29(3):327-339
25. Liang H, Fu K-w (2015) Testing propositions derived from Twitter studies: generalization and replication in computational social science. PLoS ONE 10(8):1-14 https://doi.org/10.1371/journal.pone.0134270
26. Gonçalves B, Perra N, Vespignani A (2011) Modeling users' activity on Twitter networks: validation of Dunbar's number. PLoS ONE 6(8):22656
27. Leetaru K, Wang S, Cao G, Padmanabhan A, Shook E (2013) Mapping the global Twitter heartbeat: The geography of twitter. First Monday 18(5)
28. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A (2013) The Twitter of babel: mapping world languages through microblogging platforms. PLoS ONE 8(4):1-9 https://doi.org/10.1371/journal.pone.0061981

29. Porter MF (1980) An algorithm for suffix stripping. Program 14(3):130-137
30. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning (ICML-14), pp 1188-1196
31. Garcia D, Mavrodiev P, Casati D, Schweitzer F (2017) Understanding popularity, reputation, and social influence in the Twitter society. Policy Internet 9(3):343-364. https://doi.org/10.1002/poi3.151
32. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111-3119
33. Hill RA, Dunbar RI (2003) Social network size in humans. Hum Nat 14(1):53-72
34. Tufekci Z (2014) Big questions for social media big data: representativeness, validity and other methodological pitfalls. In: Eighth international AAAI conference on weblogs and social media (ICWSM)
35. Boyd D (2012) Networked privacy. Surveill Soc 10:348-350
36. Schep T Social Cooling. Accessed November 20, 2017 (archive in http://archive.is/DTIYb). https://www.socialcooling.com/