social sciences

MDPI

Article

Privacy Threats and Protection Recommendations for the Use of Geosocial Network Data in Research

Ourania Kounadi [1,2,*], Bernd Resch [2,3] and Andreas Petutschnig [2]

1 Faculty of Geo-Information Science and Earth Observation (ITC), Department of Geo-information Processing, University of Twente, 7514 AE Enschede, The Netherlands
2 Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria; bernd.resch@sbg.ac.at or bresch@fas.harvard.edu (B.R.); andreas.petutschnig@sbg.ac.at (A.P.)
3 Center for Geographic Analysis, Harvard University, Cambridge, MA 02138, USA
* Correspondence: o.kounadi@utwente.nl; Tel.: +31-(0)53-4891-525

Received: 8 August 2018; Accepted: 9 October 2018; Published: 11 October 2018

Abstract: Inference attacks and protection measures are two sides of the same coin. Although the former aims to reveal information while the latter aims to hide it, they both increase awareness regarding the risks and threats from social media apps. On the one hand, inference attack studies explore the types of personal information that can be revealed and the methods used to extract it. An additional risk is that geosocial media data are collected massively for research purposes, and the processing or publication of these data may further compromise individual privacy. On the other hand, consistent and increasing research on location protection measures promises solutions that mitigate disclosure risks. In this paper, we examine recent research efforts on the spectrum of privacy issues related to geosocial network data and identify the contributions and limitations of these research efforts. Furthermore, we provide protection recommendations to researchers that share, anonymise, and store social media data or publish scientific results.

Keywords: privacy; geoprivacy; geosocial network data; location-based social networks; anonymisation

1. Geosocial Network Data in Research

In recent years, data from geosocial networks such as Twitter, Flickr, Instagram, Foursquare and others have become a comprehensively used basis for geospatial analysis in a number of application areas, including disaster management (Laituri and Kodrich 2008; Resch et al. 2018), public health and epidemiology (Santillana et al. 2015; Boulos et al. 2011), urban planning (Foth et al. 2011; Resch et al. 2016), traffic management (Pan et al. 2013; Steiger et al. 2016a), crime analysis (Malleson and Andresen 2015; Ristea et al. 2018; Kounadi et al. 2018), and others. While early research efforts focused on simple analysis using traditional methods (Girardin et al. 2008; Sagl et al. 2012), more recent research has developed more sophisticated approaches, including self-learning systems such as artificial neural networks (ANN) (Steiger et al. 2016b), machine learning semantic topic models (Hasan and Ukkusuri 2014; Kovacs-Gyori et al. 2018) or real-time analysis algorithms (Sakaki et al. 2010).

Resulting from the rapid development of social media analysis, data analysis methods have become more robust and results more reliable. In turn, geosocial networks are meanwhile acknowledged as a high-quality data source that supports the investigation of real-world problems and subsequent decision-making. This development has been fostered by the dramatically increasing availability of social media posts throughout the world, particularly in urban settings. Consequently, we have witnessed the emergence of far-reaching analysis efforts that investigate urban processes at a remarkably high spatial and temporal resolution.

On the downside, this leads to the pressing question of how to preserve the privacy of social media users, an issue which is becoming more and more serious as the spatial and temporal density of social media posts increases. This is because extracting user profiles and identifying single users can be done relatively easily by analysing accumulated social media posts, particularly when coupled with other data sources such as demographic data, statistical data or household data, which are increasingly available as open-source repositories (Steiger et al. 2015).

Potential infringements, which may arise when analysing geosocial media data and when publishing according results, include revealing a user's identity, building behaviour profiles of users, generating political profiles of users, putting users at physical risk (e.g., lateral thinkers, public figures, etc.), or making information permanently available on the Internet. Therefore, these infringements are particularly critical as fine-grained research outputs may only constitute a surrogate for more concrete and personal influences on users: A spatial accumulation of "negative emotions", "high traffic volumes", or "bad air quality" may have very direct consequences on a user's permanent stress level, their quality of life or even their life expectancy. Thus, more accurate, finer-grained or more complete information may in some cases not necessarily be desirable, as this would potentially allow for conclusions regarding the subjects' identity on a very small scale or, in extreme cases, even on the individual level (Resch 2013).

For the particular case of geosocial media, the responsibility for sharing data in "appropriate" semantic, content-wise, spatial and temporal detail cannot be shifted to the user because terms and conditions of the use of social networks are mostly articulated in convoluted and hardly understandable language. Posts in geosocial media are usually available through (public) Application Programming Interfaces (API), which enable data access without the users' awareness even though users knowingly and consensually share their data as agreed in the terms and conditions of a geosocial network application. The above-mentioned conditions necessitate a rigorous way of handling data from geosocial media to preserve the users' privacy. The challenge of this goal lies in the spatial nature of geosocial media data: The first of the 21 theses in the "Geoprivacy Manifesto" by Keßler and McKenzie (2018) says that "information about an individual's location is substantially different from other kinds of personally identifiable information". Individual time trajectories in space reveal activity spaces (e.g., locations of work, social clubs, day care, grocery, place of worship, etc.) that can, in turn, be processed to get insights on human behaviour and detect personal profiles (Armstrong et al. 2018). Therefore, the major factor that makes spatial information more challenging is the potential of inferences on identity that may be drawn.

This paper discusses privacy risks associated with research efforts using geosocial media data, identifies the limitations of existing studies regarding inference and protection, and proposes a set of geoprivacy-by-design recommendations with respect to sharing these data, anonymising them, publishing the resulting maps, modalities of data storage, and privacy-preserving measures. This is followed by a thorough discussion of the proposed recommendations, particularly in light of the recent General Data Protection Regulation (GDPR), and a set of future research directions in the area of geoprivacy. Furthermore, this paper addresses the use (storage, analysis, visualisation and sharing) of geosocial network data, but it shall not be understood as a guideline relevant to building location-based social networks.

2. Background on Inferences, Users, and Policies

Studies on location inference attacks examine the types of personal information that can be revealed from individual-level spatial trajectories and the accuracy of inferred information. Thus, privacy policies in research efforts involving LBSN data should take possible inference attacks into consideration (Section 2.1) and provide mechanisms that are in line with the users' preferences and attitudes towards privacy protection (Section 2.2).

2.1. Inference Attacks vs. Risk of Re-Identification

In this section, we review the literature on inference attacks and re-identification risk from spatial trajectory data. First, we should clarify the difference between inference attacks and re-identification. Inference is to draw conclusions based on observations and analytical results of the data. For instance, given a set of locations of a GPS user, the clusters of high point intensity can be computed. Then, the central point of the cluster with the highest density can be assumed as the home location of the user, especially if there are temporal signals during late night hours when people usually stay at home. In statistical terms, inference is accompanied by accuracy results. To calculate the accuracy of inference, true outcomes should be measurable. This means that, in the previous example, the true home location of the user is known and compared with the estimated one. However, this is not always the case for the studies on inference attacks from location data. Hence, in many cases, inference attacks show the potential of the kind of information that can be disclosed without necessarily validating the conclusions. On the other hand, re-identification involves a disclosure method as well as an accuracy assessment against the actual information.

Taking this distinction into consideration, most of the studies have examined the potential of drawing conclusions about the private matters of individuals, and only few of these studies validated the degree to which such conclusions are accurate (Table 1, category: validation data). Types of re-identified information are, for example, the prediction of a social media user's next location in a georeferenced post (Preoţiuc-Pietro and Cohn 2013), the location of social media posts from a georeferenced dataset (Schulz et al. 2013), or the home address and identity of individuals that carry GPS receivers (Krumm 2007). Other studies evaluated their inference results questionably because the validation data that were used had significant limitations. Zang and Bolot (2011) aimed at detecting the home and work locations of cell phone users, but only had 12 subscribers to validate their results. Li et al. (2016) employed a significant number of participants in their experiments in order to reveal highly sensitive information from geosocial network media data and Wi-Fi traffic records such as age, gender, education, living place, and location patterns. However, the participants are only representative of a particular subgroup of the population, that is, people who work, study, or live on a university campus. One could argue that the sample's characteristics are considerably less variant than a representative sample of the general population and thus easier to predict. For example, more than half of the participants have a bachelor's degree, and there are only three types of education levels (i.e., bachelor, master, and PhD). In addition, Schulz et al. (2013) inferred the home location of Twitter users and then validated the estimated home location by using the last location of the user as ground truth. Similarly, Pontes et al. (2012) validated an estimated city of a user using as ground truth the information provided in the user's home city attribute. All approaches highlight the potential for re-identification but do not reveal the actual re-identified information.

The information that can be inferred or re-identified varies from barely sensitive, such as the country of an individual, to highly sensitive such as private locations, next locations or even the identity of a user. Most inference approaches are based on heuristics (Krumm 2007; Gambs et al. 2010; Zang and Bolot 2011; Pontes et al. 2012; De Montjoye et al. 2013; Schulz et al. 2013; Li and Goodchild 2013; Lampoltshammer et al. 2014), and on clustering and classification (Lampoltshammer et al. 2014; Preoţiuc-Pietro and Cohn 2013; Gambs et al. 2010; Li et al. 2016). Although data from public sources can be used as a ground truth information, there has been only one study in which they were used (Krumm 2007) (Table 1, category: inference approach).

Furthermore, some of these studies propose measures to protect subjects' anonymity in location trajectories, such as perturbation, aggregation (e.g., areal, point, or temporal), considering the desired level of privacy defined by user preferences, shortening the time collection period, and removing sensitive areas (i.e., spatial cloaking) (Table 1, category: countermeasures). Most of these measures deliver sufficiently large anonymous datasets and may work well for plain spatiotemporal trajectories. Nevertheless, in Section 7, we outline the limitations of the k-anonymity concept with respect to

geosocial network data, due to the diversity and variety of potential disclosed information, and we explain why alternative measures based on differential privacy or l-diversity are preferable.

Table 1. Literature review on re-identification and inference attacks from location data.

Inferred or Re-Identified Information		Location Data
• Identity: 1 • Home: 1, 2, 3, 6, 9 • Home, work: 3, 8 • Home, work, private locations: 3 • City: 4 • State: 4	• Country: 4 • Trace uniqueness: 5 • Location of post: 6 • Cluster: 7 • Next movement: 7 • Location pattern: 10 • Demographics: 10	• GPS: 1, 2 • Cell Phone data: 3, 5 • LBSN data: 4, 6, 9, 7, 8, 9, 10 • Wi-Fi traffic records: 10
Inference Approach	**Validation Data**	**Countermeasures**
• Heuristics: 1, 2, 3, 4, 5, 6, 8, 10 • Clustering/Classification: 2, 7, 9, 10 • Public data: 1 • Social applications data: 10 • Probabilistic models: 7, 10	• Yes: 1, 6, 7 • No: 2, 3, 5, 9, 8 • Non-significant sample size: 3 • Non-representative sample: 10 • Questionable accuracy: 4, 6	• Spatial cloaking: 1 • Perturbation: 1, 2 • Spatial aggregation: • 1, 3, 5 • Temporal aggregation: • 2, 5 • Shorten collection time: 3 • Based on user preferences: 10

Study (Reference, Study Area (If Stated), Data, and Subjects)
1. (Krumm 2007), 172 participants carrying GPS receivers on their vehicles
2. (Gambs et al. 2010), San Francisco—US, 90 taxi trails
3. (Zang and Bolot 2011), 50 states—US, 30 billion call records of 20 million cell phone users
4. (Pontes et al. 2012), Global, Foursquare public lists of 13 million users
5. (De Montjoye et al. 2013), Western country, 15 months of mobility data from phone interactions of 1.5 M people
6. (Schulz et al. 2013), Global, 1 Million georeferenced tweets
7. (Preoţiuc-Pietro and Cohn 2013), Globe, Foursquare check-ins of 9167 users
8. (Li and Goodchild 2013), Los Angeles—USA, georeferenced tweets of 5 heavy users
9. (Lampoltshammer et al. 2014), London—UK, georeferenced tweets about crime events
10. (Li et al. 2016), China, GPS trajectories and social media data of 30 participants in five apps and Wi-Fi traffic records of 22,843 users

2.2. Users' Privacy Preferences

Geosocial network data are provided voluntarily by the users who are also the data subjects. Their preferences on the protection of personal privacy in LBSN have been studied and conceptualised as opinions, attitudes, and behaviours. Beldad and Kusumadewi (2015) identified major determinants of sharing locations in LBSN. The first two are related to personal benefits, such as entertainment and impression management (i.e., controlling the impression they have on others), while the third one is trusting the competences of an application to protect personal privacy. A similar study on location information disclosure behaviour also confirms that privacy risks weaken the relationship between perceived benefits and intention to disclose personal information (Sun et al. 2015). In addition, both studies on location disclosure behaviour found that there are significant gender differences in the responses of the participants. Benisch et al. (2011) performed a survey on 27 participants and collected their location trajectories over three weeks. Then, the participants ranked and explained their disclosing criteria. One of the most significant findings was that the decision of users of whether or not to disclose their locations varies depending on the time of the day, the day of the week, and their exact

location. A second finding, which is important for policy implementation, is that users would prefer a more complex location- and time-based privacy set of rules over a simpler approach that restricts disclosed information to a particular group (i.e., friends or family).

On the other hand, people's opinions, attitudes, and behaviours regarding geoprivacy risks are connected to their geoprivacy awareness, which is not yet widely spread and well understood. Half or more of the participants in a location awareness study had *no idea* if: (a) their profiles in Twitter and Instagram are private or public; (b) they use the geolocation feature; and (c) they ever changed default privacy settings (Furini and Tamanini 2015). Another study asked participants to state their awareness of 14 types of inference attacks from geosocial network data (e.g., to infer home and work location, to know their friend network and weekly habits, etc.) (Alrayes and Abdelmoty 2014). More than one-third of the participants were not aware of possible attacks such as these related to other people being able to know what their personal activities are.

Users' preferences regarding the protection of their geosocial data are diverse and probably linked to their awareness about which data are available (on the Internet) and how they can be used. Thus, it is not advisable for researchers or institutions to construct privacy by design guidelines based on generalised preferences of the public who lack specialised knowledge on geoprivacy implications.

2.3. Privacy Policies in LBSN

Gambs et al. (2011) gave a comprehensive overview of the privacy policies implemented by four LBSN (Foursquare, Qype, La Ruche and Twitter). They identified the following eight privacy criteria and also checked whether the LBSN adhere to them:

Privacy criteria in LBSN:

1. Registration information: How much personal information from users is needed for registration?
2. Real identities versus pseudonyms: Are users allowed to use pseudonyms instead of their real name?
3. Information available to others (friends, public, and third parties): What personal information about users is disclosed to other parties operating on the LBSN?
4. Privacy settings: Do users have control over how their data is collected, used and disseminated?
5. Terms of use and privacy policy: Does the LBSN provide an explicit and easily understandable policy in which users are informed about how their data are used?
6. Policy of data retention in case of account deletion: Does the LSBN delete all data from a user after they delete their network account?
7. Mobility data collection and management: Are location data collected continuously or only when a user action requires location data access?
8. Security features: Does the LBSN implement reasonable IT security measures to prevent data theft?

The authors concluded that the platforms largely do not implement measures to fit the criteria and provide a list of practical recommendations for LBSN to use. Vicente et al. (2011) performed a similar study that examines a larger number of LBSN and outlines the features of the services that increase the re-identification risk. These features are the real-time publication time (occurs in 43% of the examined LBSNs of the study), the use of exact location (occurs in 62% of the LBSNs), and the ability to tag or check-in multiple users (occurs in 19% of the LBSNs). However, only 14% of the LBSNs use anonymous user identities, which is a feature that decreases the re-identification risk. Furthermore, some of their listed LBSN pose privacy issues for users, although they leave a threat formalisation to future studies and suggest spatial and temporal cloaking as a possible privacy protection measure. Further, they provided an outlook, in which they name user awareness of publishing location information as a factor in privacy protection.

3. Data Sharing

Researchers may share processed or unprocessed datasets for several reasons, for example, to allow research replicability, to establish synergies with research partners, or to publish in open data scientific journals. These datasets, typically, do not contain key identifiers (i.e., name, home address, etc.), but pseudonyms (i.e., username) that can be used to derive subsets for each subject. Inferential disclosure can be applied to the attributes (e.g., location) and reveal not only the identity but also further personal information about the subject.

Starting with the inferential disclosure of the subject's identity, an attacker may use the subject's space-time stamps to make a guess about their potential home address. In a study that used GPS trajectories of participants, the author re-engineered the real home addresses with a median distance error of 60.7 meters and the identity of a small fraction of the participants (Krumm 2007). Furthermore, LBSN data contain additional attributes that can lead to greater information disclosure compared to GPS data. Alrayes and Abdelmoty (2014) enlisted twelve types of disclosure that can be inferred from combining and analysing the spatial, temporal, and non-spatial semantics such as times spent away from home, activities during weekends, and time and location of meetings with friends, amongst others.

The simplest approach to mitigating such privacy threats is to remove pseudonyms and other information that can be used to derive subsets per subject prior to the release of a dataset. This can indeed be an effective solution, since many studies are interested in aggregated results, such as the spatial-temporal distribution of a topic of interest in a study area. If analysis by user or group of users is needed, pseudonyms may be stored, but the dataset has to be anonymised. Prior to the anonymisation, the data holder should consider the total number of observations by time intervals per user. Time information is critical in inferring personal information from geosocial network data. For example, locations derived after midnight and during weekdays can be used as a starting dataset to infer home addresses. The probability of an accurate inference is related to the entropy of the locations; in other words, the lower the entropy, the higher or more confident the inference is. More observations within a time interval may lower the entropy, resulting in easier detectability of a pattern. Ultimately, this depends on the temporal frequency of posts by users. A user with sporadic posts may be harder to identify compared to a frequently posting user. Thus, restricting the temporal frequency of the observations per user can be an anonymisation strategy to mitigate disclosure risk.

There are several methods for the anonymisation of LBSN data, which are critically discussed in Section 7. The primary criterion for the selection of an optimal method is that it protects the data sufficiently based on an anonymity measure, such as k-anonymity (Sweeney 2002), l-diversity (Machanavajjhala et al. 2007), or differential privacy (Dwork 2008). Another significant criterion is the spatial effect that a method imposes on the anonymised dataset. For instance, anonymised data produced by random perturbation approaches detect spatial clusters more accurately than data produced by aggregation approaches (Hampton et al. 2010; Kounadi and Leitner 2016). On the other hand, aggregation may be preferred if data are to be analysed or visualised in the same aggregation level as the anonymised data. The effect should also be calculated and communicated to future users. Kounadi and Resch (2018) proposed several measures that calculate the effect of the spatial error of the spatial analysis to be performed.

Recommendations:

1. Remove pseudonyms or other subject identification attributes.
2. Anonymise data if subject distinguishability is required.
3. Ensure anonymisation method provides sufficient protection based on an anonymity measure.
4. Select a method that minimises the spatial effect on the anonymised dataset based on their utility.
5. Calculate the spatial effect of anonymised data on certain types of analysis.
6. Communicate anonymity level and accuracy errors of anonymised data to future users.

4. Anonymised Data

Post-anonymisation practices may also involve disclosure risks. One example is when the data holder releases multiple versions of anonymised copies of original data. Research on a method that is based on a Gaussian perturbation shows that the more anonymised copies are released, the more accurately an attacker can infer the original locations (Cassa et al. 2008). The rationale is that multiple versions of the same datasets reveal hints regarding the anonymisation approach and facilitate the re-engineering process. This also implies that data holders should carefully consider how much and what type of information on the anonymisation method should be published.

In particular, Zimmerman and Pavlik (2008) suggested that metadata information on methods such as aggregation and perturbation may increase the disclosure risk. For example, if an attacker knows that data are perturbed based on a normal distribution, they will recognize that there are greater chances to re-identify an original location when looking in close proximity to the anonymised location. The disclosure risk, in this case, is higher to that of a perturbation based on a random distribution. Therefore, metadata information on the anonymisation method should be restricted unless the data holder is confident that releasing this information will not increase the disclosure risk. Fortunately, some methods guarantee privacy even when the attacker knows details about the data or the anonymisation process. These are discussed in Section 7.

Recommendations:

1. Do not release multiple versions of anonymised datasets.
2. Restrict metadata information on the anonymisation method.

5. Publication of Maps

Publication deliverables such as maps of confidential and private data may also lead to information disclosure. More specifically, research on confidential point data shows that locations depicted on maps in scientific publications can be re-engineered with considerable accuracy either from a digital or a printed map (Brownstein et al. 2006; Leitner et al. 2007). To the best of our knowledge, no similar research has been conducted regarding social media data. However, if a map distinguishes locations or trips by data subject, a similar re-engineering process can be used and, thus, the risk of disclosure remains.

Researchers must ensure that public cartographic visualisations do not compromise the privacy of the individuals involved in the dataset. A simple way to ensure privacy protection in maps is to lower the spatial or the temporal precision and present aggregate data (Graham 2012). In addition, the independent body Information Commissioner's Office (ICO) in the UK suggests to use heat maps (i.e., continuous or aggregated surfaces of densities) or explore alternatives of representing confidential information on maps (ICO 2012). Of course, if researchers wish to present detailed unprocessed information on maps, they should use the anonymised versions of their data. In line with Section 3, the spatial error of the map should be evaluated concerning the impact it may have on the readers when interpreting the map.

Recommendations:

1. Ensure privacy protection of public cartographic visualisations.
2. Reduce the spatial and/or temporal resolution of public maps.
3. Consider the use of heat maps or other types of cartographic visualisations.
4. Use anonymised data if it is necessary to publish detailed maps (i.e., locations or trajectories distinguished by subject).
5. Assess the spatial error and its impact when anonymised data are used in maps.

6. Data Storing

Boulos et al. (2009) described *data security* as the "missing ring" in privacy-preserving discussions that have predominately neglected risks such as data theft, data loss, or data disclosure to non-authorised parties. The authors highlighted several security measures (e.g., building security, cable locks, cryptography, access authentication, etc.) and suggested a "*purpose-built*" combination of measures that depend on the type, sensitivity, value, and risk of data. An expert, who acts as a designated privacy manager and whose knowledge extends beyond location-related disclosure risks, should oversee data storing and processing tasks. If unauthorised persons can physically access the storage devices, sensitive data on them should be encrypted to avoid theft. In case data are to be stored or processed on machines provided by third parties within a cloud computing environment, the entire workflow from sending the data to receiving results has to be subject to encryption, which must not be compromised at any stage (e.g., by the use of client-side encryption). Chen and Zhao (2012) gave a more detailed overview of cloud computing security architectures and data security issues. On top of these measures, it is of course also important to adhere to well-known security routines that help prevent data theft. Examples of such measures are locking computers when not needed, not writing down passwords, using strong passwords, and not reusing passwords.

Recommendations:

1. Assign a privacy manager or security expert to oversee data storage and processing tasks.
2. Apply all necessary security measures and best practices throughout the entire workflow.
3. If storing or processing data on third-party machines, ensure that security standards are upheld throughout the entire workflow.

7. Privacy Concepts and Protection Methods

An approach to protect the locations of LBSN data is to prevent them from being released to unauthorised parties. This can be achieved by allowing the user to set up their own privacy preferences for location disclosure or by transmitting data in an encrypted form. For example, the data can be encrypted when shared with untrusted third-party servers, and then decrypted by the users that the data is intended for (e.g., friends) (Puttaswamy and Zhao 2010). In addition, encryptions can be transferred to the hands of users who may apply policies on who may access their private data based on their attributes (i.e., attribute-based encryption) (Baden et al. 2009). Another possibility is that the users decide and adapt the granularity of their shared locations, while probabilistic encryption ensures that their data and preferences remain private (Hu et al. 2017). A third approach is to divide the released information between social network servers and location-based servers (Wei et al. 2012) or to further split them into multiple location servers to prohibit access to users' social network topology based on their friend sets (Li et al. 2017).

Although encryption, adaptive privacy preferences, and location servers are straightforward privacy protection approaches, they prohibit or limit the use of LBSN data for secondary purposes, such as research studies, which are the scope of this paper. On the other hand, location transformation promises privacy protection while data are shared openly, and data can, thus, be extracted and used for research purposes. Armstrong et al. (1999) were the first scholars to anonymise data by transforming their locations and established the term "geographical masking" for the protection of discrete spatial datasets. Later approaches applied geographical masking with the privacy measure of k-anonymity, which ensures that a data subject cannot be distinguished amongst k-1 other subjects (Sweeney 2002), (Cassa et al. 2006; Hampton et al. 2010). This concept is best applicable to confidential datasets (e.g., health and crime information) such as locations of domestic violence events where each location can be a direct link to a building or a household. In practice, spatially anonymised regions are defined based on the dataset and underlying population, and then data are either displaced (Kounadi and Leitner 2016) or aggregated within these regions (Croft et al. 2017). Furthermore, spatiotemporal versions

of k-anonymity, commonly known as cloaking, have been applied to location-based services data by degrading the location and/or time information that is sent to a server to ensure that queries contain at least k-1 users (Gruteser and Grunwald 2003; Mokbel et al. 2006; Kalnis et al. 2007). Regarding geosocial networks, Freni et al. (2010) proposed a technique that combines generalisation, spatial cloaking, and temporal cloaking to protect two types of privacy concerns. The first concern is the uncontrolled disclosure of a user's location at specific times and the second concern is the uncontrolled disclosure of the absence of a user at a location at specific times (e.g., a user is not at work or home).

Shokri et al. (2011), in their work on location privacy of mobile users, developed a quantifier (metric) for location privacy that is based on the incorrectness of the adversary in their inference attack (i.e., the higher the number of incorrect inferences, the higher the privacy level is achieved). The authors analysed the localisation of users over time from trajectories protected via k-anonymity but found that the desired anonymity level was in some cases over or underestimated. Another limitation of k-anonymity is that it cannot prevent disclosure from a homogeneity attack (i.e., knowing a person who is in the database) and a background knowledge attack (i.e., knowing a person who is in the database and additional information on the distribution of the sensitive attribute/attributes). Unfortunately, there are many types of datasets, including LBSN data, which may suffer from these attacks. For example, an attacker may know a user or groups of users that have accounts in a geosocial network as well as other background information on the type of inference he/she is about to make. Privacy concepts such as l-diversity (ensures that an equivalent class has at least l "well-represented" values for the sensitive attributes) (Machanavajjhala et al. 2007) or differential privacy (ensures that the presence or absence of a subject in the data does not alter the probability of the properties of a query answer) (Dwork 2006) are able to protect against these two types of attacks. L-diversity results in protected datasets and differential privacy yields answers to aggregate queries. Although both approaches were formulated in the context of statistical databases, they show great potential for protecting data from geosocial networks and spatial data in general. Nevertheless, we should stress that l-diversity data are still vulnerable to composition attacks (i.e., an attacker uses independent anonymised releases about overlapping populations to compromise privacy), but a differential privacy based approach may satisfy such conditions (Ganta et al. 2008).

One of the first attempts to prevent these attacks for location data is the work by (Cormode et al. 2012) who adapted spatial indexing methods such as quadtrees and kd-trees to provide spatial decompositions that are differentially private. The decompositions allow queries to know how many individuals (or other point objects in question) fall within a given region. However, considering the complexity of geosocial network data, this is only one of the many possible queries that entail private information.

Another possible query is to identify locations of interest near other locations. For instance, social media applications use the users' personal trajectories (captured vie check-ins) to give them suggestions about which places to visit. An attacker may use the recommended locations to make individual inferences such as the user's actual trajectory. Zhang et al. (2014) proposed sanitation approaches that allow recommendations queries without revealing the user's trajectory. In a similar way, the LocBorg approach retains online personas by suggesting users add posts that are similar to their topics of interest but have fake locations (Zakhary et al. 2017). This approach might be useful for a-spatial studies, in which sensitive attributes and personal profiles are important, but location accuracy is of no interest.

Another approach based on differential privacy for individual-level location data is the notion of "geoindistinguishability" (Chatzikokolakis et al. 2015), which allows users to be protected within an adaptive radius of r, for which the desired privacy (l-*privacy*) increases with the distance. The advantage of geoindistinguishability compared to the previous two approaches is that it is applicable to queries related to a single user (location at specific time) rather than providing aggregate information about several users.

Furthermore, a typical use of social media data in research is to identify and examine spatial clusters of features of interest. Wang et al. (2016) proposed a method that provides differentially private results in areas with high concentrations of privacy-preserved tweets. The outcome can be used to identify correlations between users and events without identifying the exact locations. Moving away from differentially private methods but still looking into research applications, the location history of users can be used to predict their next locations. Xue et al. (2017) developed a destination prediction model that explores the check-in service of geosocial networks and a privacy protection method against such attacks.

8. Discussion and Future Research Directions

The way in which LSBN data are analysed and published in a responsible manner may not only be a technical question but a legal one as well. Depending on which country data are from and published in, different legal restrictions that may go beyond the recommendations given in this paper may apply. Examples of such legal frameworks can be found accompanying the many open data portals that some governments operate to share their data.

8.1. The EU Open Data Portal and the General Data Protection Guideline (GDPR)

The EU Open Data Portal (European Parliament 2011) is used to unify all open data portals of the EU member states. As a legal framework, they adhere to the Regulation 45/2001 on processing of personal data by the EU institutions (European Parliament 2001), which applies by proxy to the EU member states. However, taking a closer look at how it is implemented in the respective member states reveals that, even with a universal privacy protection law, differences may occur in this respect. This is shown by Custers et al. (2017), who compared how different EU governments and their citizens enforce and allocate resources for data protection and engage in debate about the topic.

As of 25 May 2018, the General Data Protection Regulation (GDPR) (European Parliament 2016) has been in effect, which affects LBSN operating within the European Union or the European Economic Area. Its goal is to empower users by enforcing transparency and constraints for the storage and processing of personal data, thus forcing LBSN and other organisations to design their data storage facilities in a way that precludes misuse. The GDPR requires that personal data are stored according to principles such as privacy by design and by default, minimising data storage time, informing individuals about how their data are processed, and purpose limitation. Concretely, it regulates data processing with respect to the following aspects:

- **Lawful basis for processing**: if no user consent for data processing has been provided, there needs to be a legal basis for analysing data, such as public interest, contractual obligations or to protect the interest of the subject
- **Responsibility and accountability:** responsibility and the liability of the data controller to implement effective data and privacy protection measures
- **Data protection by design and by default**: high level of privacy by default, including encryption, and rules for the analysis of data
- **Pseudonymisation**: replacing bits of information with random information (e.g., replacing names with random names) to avoid re-identification
- **Right of access**: a subject's right to access their personal data
- **Right to erasure**: a subject may request the erasure of all their personal data
- **Records of processing activities**: documentation of the data processing steps, including their purpose, the categories of used personal data, the projected time limits for erasure, or a general description of taken security measures
- **Data protection officer**: a data protection manager has to be assigned in every institution
- **Data breaches**: the data controller is legally obliged to notify the supervisory authority about any data breach

- **Sanctions**: warnings, audits or fines can be issued
- **Business to business (B2B) marketing**: allowed, provided consent or legitimate interest is given

Importantly, for research campaigns, the GDPR does not apply in the following circumstances:

- Lawful interception, national security, military, police, justice
- ***Statistical and scientific analysis***
- Deceased persons are subject to national legislation
- There is a dedicated law on employer-employee relationships
- Processing of personal data by a natural person in the course of a purely personal or household activity

8.2. The Challenges of Diverging National and Supra-national Legislation

The legal constraints for storing and processing personal data and the right to privacy differ widely between countries. Noorda and Hanloser (2011) provided an overview of selected national legislations from across the world and point out some of their incompatibilities. They also gave examples of cases in which such incompatibilities allowed privacy violations to be committed with impunity, thus pointing out the impactful consequences of such unclear legal situations. Custers et al. (2017) showed that even on a smaller scale, within the EU, such incompatibilities exist.

The most severe limiting factor in this regard is the varying interpretation of "privacy" in different parts of the world. For instance, privacy can be traded as an economic good by its owner in the USA, whereas it is protected by law in the European Union. An ideal, but unlikely case would be that supra-national legislation bodies and initiatives set up appropriate world-wide regulations (Resch et al. 2012). As shown in Figure 1, legislation and governments play highly different roles in these two environments.

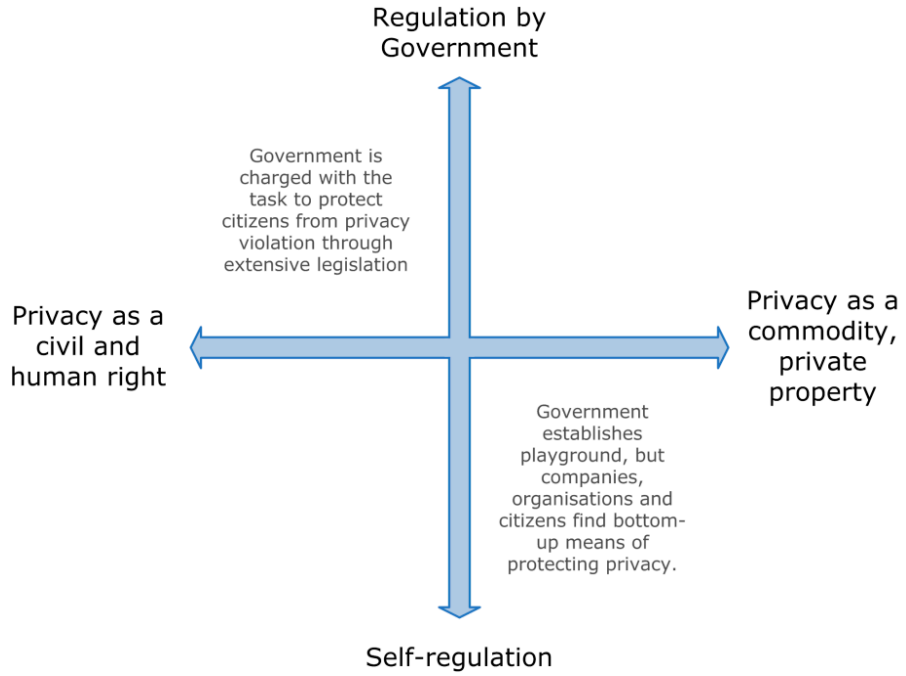

Figure 1. Different Understandings of Privacy and the Roles of Governments (Resch et al. 2012).

This makes it impossible to draw up one universally applicable and legally binding set of rules for data storage. As a consequence, researchers must not only respect the data storage and processing conditions set forth by data providers and best practice guidelines but also by their national jurisdiction and the jurisdiction of their data subjects.

8.3. Future Research Directions

Section 2.1 outlines the limitations of existing studies on ***inference attacks***. They mainly arise from the lack of true and measurable actual data. When inferences are being made about private information of individuals, the ideal means of validating them is by confirming with the tested individuals. However, in reality, it is virtually impossible to get these individuals to report on their private matters (e.g., where do you live? Where do you go on weekends?). Furthermore, if private aspects are reported, they are oftentimes prone to a number of biases such as the cooperative principle (respondents may alter their statements when answering questions repeatedly), retrospective biases that may be caused through delayed responding (e.g., inaccurate recall, recency effects, false memories), or the fact that some respondents may stick with what they answered earlier in order to appear consistent and not contradict themselves (Bluemke et al. 2017).

A major incentive of inference attack studies is to raise awareness on the negative implications of regular geosocial media practices of people (e.g., geotagging posts). If such studies are ethically responsibly conducted and performed by reliable research campaigns, people would potentially become motivated to participate for the benefit of the society.

In addition, studies on inference attacks typically use the spatial information (i.e., coordinates and trajectories) to make inferences. However, McKenzie et al. (2016) demonstrated that the protection of private location information should not be exclusively handled from a spatial perspective. Place-based information co-exists in the ***semantic signatures*** of geosocial footprints such as the spatial, temporal, and thematic inductiveness of posts. Furthermore, another feature of LBSN that has not been discussed in this paper is ***image data*** (geocoded or not). In fact, the link between image data and information disclosure remains a significant research gap in the literature of geoprivacy. Image data generally belong to the LBSN features that may increase the re-identification risk. For example, similarity and clustering algorithms can match an image "A" (or set of images) to another image "B" (or set of images) (Kawakubo and Yanai 2011; Lv et al. 2004; Chen et al. 2005). If image "A" belongs to a fully anonymous account of an LBSN user that has location information data (e.g., geotagged messages and geotagged pictures), and image "B" belongs to another identifiable account (or any source of information linked to individuals), then one can draw conclusions about the involved individuals or even infer that they are the same person. As a result, information from the anonymous account can be disclosed.

Most importantly, the field of computer vision deals, amongst other things, with image location recognition algorithms (Arase et al. 2009; Zhang and Kosecka 2006; Hays and Efros 2008; Li et al. 2010). Thus, images of anonymous accounts can be directly processed to identify location patterns of the users. However, this has not yet been studied in the context of geoprivacy and spatial re-identification risk.

Moreover, there is a need to match and harmonise scientific knowledge (e.g., protection methods and privacy by design guidelines), and the legal aspects of location privacy (e.g., how should privacy be protected in Europe based on the GDPR?) with the use of technological tools. One such tool is a ***spatial decision support system*** (SDSS). SDSS can be specified for the application domain of geoprivacy in order to help and guide "data holders", researchers (or principal investigators in larger research campaigns) when anonymising their data. This system can have the form of a graphical user interface (GUI) to allow users to interact with the program and make informative decisions. As we explained earlier in the paper, decisions on anonymisation or protection of LBSN data involve certain standard principles but also depend on, or could be adapted based on, future analyses (e.g., regression, classification, point pattern analysis, clustering, etc.), as well as the type of release (e.g., an aggregated table or a detailed point distribution map).

Finally, an essential aspect of future research efforts in the area of geoprivacy are the unclear consequences that the GDPR may pose. Although the GDPR is in place and has to be followed, it is not entirely clear which measures researchers need to take to comply with the regulation with respect to data acquisition, storage, processing, visualisation or sharing. This problem is rooted in the ambiguous and non-exhaustive formulations of the GDPR. Consequently, detailed interpretations of the GDPR may only be possible after a number of jurisdictional cases, which may potentially

compromise current research practices and put severe limits on operational procedures of research involving personal geodata.

Author Contributions: Conceptualization, Review, and Drafting: All authors; Inferences and Users, Recommendations, Anonymization, and Future Directions: O.K. Funding Acquisition, Introduction, Discussion, and Future Directions: B.R.; Privacy Polices, Data Storing, and Discussion: A.P.

Funding: We would like to express our gratitude to the Austrian Science Fund (FWF) for supporting the project "Urban Emotions" (reference number I-3022) and the project "The Scales and Structures of Intra-Urban Spaces" (reference number P 29135-N29), as well as to the Austrian Research Promotion Agency (FFG) for funding the project "HUMAN+" (reference number 865697).

Conflicts of Interest: The authors declare no conflict of interest.

References

Alrayes, Fatma, and Alia Abdelmoty. 2014. No place to hide: A study of privacy concerns due to location sharing on geo-social networks. *International Journal On Advances in Security* 7: 62–75.

Arase, Yuki, Xing Xie, Manni Duan, Takahiro Hara, and Shojiro Nishio. 2009. A game based approach to assign geographical relevance to web images. Paper presented at the 18th International Conference on World Wide Web, Madrid, Spain, April 20–24.

Armstrong, Marc P., Gerard Rushton, and Dale. L. Zimmerman. 1999. Geographically masking health data to preserve confidentiality. *Statistics in Medicine* 18: 497–525. [CrossRef]

Armstrong, Marc P., Ming-Hsiang Tsou, and Dara E. Seidl. 2018. Geoprivacy. In *Comprehensive Geographic Information Systems*. Amsterdam: Elsevier, pp. 415–30.

Baden, Randy, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. 2009. Persona: An online social network with user-defined privacy. In *ACM SIGCOMM Computer Communication Review*. Barcelona: ACM, vol. 39, pp. 135–46.

Beldad, Ardion, and Margareta Citra Kusumadewi. 2015. Here's my location, for your information: The impact of trust, benefits, and social influence on location sharing application use among Indonesian university students. *Computers in Human Behavior* 49: 102–10. [CrossRef]

Benisch, Michael, Patric G. Kelley, Norman Sadeh, and Lorrie F. Cranor. 2011. Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing* 15: 679–94. [CrossRef]

Bluemke, Matthias, Bernd Resch, Clemens Lechner, René Westerholt, and Jan-Philipp Kolb. 2017. Integrating Geographic Information into Survey Research: Current Applications, Challenges and Future Avenues. *Survey Research Methods* 11: 307–27. [CrossRef]

Boulos, Maged N. Kamel, Andrew J. Curtis, and Philip AbdelMalik. 2009. Musings on privacy issues in health research involving disaggregate geographic data about individuals. *International Journal of Health Geographics* 8: 46. [CrossRef] [PubMed]

Boulos, Maged N. Kamel, Bernd Resch, David N. Crowley, John G. Breslin, Gunho Sohn, Russ Burtner, William A. Pike, Eduardo Jezierski, and Kuo-Yu Slayer Chuang. 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples. *International Journal of Health Geographics* 10: 67. [CrossRef] [PubMed]

Brownstein, John S., Christopher A. Cassa, Isaac S. Kohane, and Keneth D. Mandl. 2006. An unsupervised classification method for inferring original case locations from low-resolution disease maps. *International Journal of Health Geographics* 5: 56. [CrossRef] [PubMed]

Cassa, Christopher A., Schaun J. Grannis, Overhage J. Marc, and Keneth D. Mandl. 2006. A context-sensitive approach to anonymizing spatial surveillance data: Impact on outbreak detection. *Journal of the American Medical Informatics Association* 13: 160–65. [CrossRef] [PubMed]

Cassa, Christopher A., Shannon C. Wieland, and Keneth D. Mandl. 2008. Re-identification of home addresses from spatial locations anonymized by Gaussian skew. *International Journal of Health Geographics* 7: 45. [CrossRef] [PubMed]

Chatzikokolakis, Konstantinos, Catuscia Palamidessi, and Marco Stronati. 2015. Geo-indistinguishability: A principled approach to location privacy. International Conference on Distributed Computing and Internet Technology. In *International Conference on Distributed Computing and Internet Technology*. Berlin: Springer, pp. 49–72.

Chen, Deyan, and Hong Zhao. 2012. Data security and privacy protection issues in cloud computing. Paper presented at 2012 International Conference on Computer Science and Electronics Engineering, Hangzhou, China, March 23–25; pp. 647–51.

Chen, Yixin, James Ze Wang, and Robert Krovetz. 2005. CLUE: Cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing* 14: 1187–201. [CrossRef] [PubMed]

Cormode, Graham, Cecilia Procopiuc, Divesh Srivastava, Entong Shen, and Ting Yu. 2012. Differentially private spatial decompositions. Paper presented at 2012 IEEE 28th International Conference on Data Engineering, Arlington, VA, USA, April 1–5; pp. 20–31.

Croft, William Lee, Wei Shi, Jörg-Rüdiger Sack, and Jean-Pierre Corriveau. 2017. Comparison of approaches of geographic partitioning for data anonymization. *Journal of Geographical Systems* 19: 211–48. [CrossRef]

Custers, Bart, Francien Dechesne, Alan M. Sears, Tommaso Tani, and Simone van der Hof. 2017. A comparison of data protection legislation and policies across the EU. *Computer Law & Security Review* 34: 234–43.

De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports* 3: 1376. [CrossRef] [PubMed]

Dwork, Cynthia. 2006. Differential Privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II*. Edited by Michele Bugliesi, Bart Preneel, Vladimiro Sassone and Ingo Wegener. Berlin and Heidelberg: Springer, pp. 1–12.

Dwork, Cynthia. 2008. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation. TAMC 2008. Lecture Notes in Computer Science*. Berlin and Heidelberg: Springer, vol. 4978.

European Parliament. 2001. Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the Protection of Individuals with regard to the Processing of Personal Data by the Community Institutions and Bodies and on the Free Movement of such Data. Available online: https://publications.europa.eu/en/publication-detail/-/publication/0177e751-7cb7-404b-98d8-79a564ddc629/language-en (accessed on 11 October 2018).

European Parliament. 2011. Commission Decision of 12 December 2011 on the reuse of Commission documents. *Official Journal of the European Union* 54: L 330.

European Parliament. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). *Official Journal of the European Union* 59: L 119.

Foth, Marcus, Laura Forlano, Christine Satchell, and Martin Gibbs. 2011. *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement*. Cambridge: MIT Press.

Freni, Dario, Carmen Ruiz Vicente, Sergio Mascetti, Claudio Bettini, and Christian S. Jensen. 2010. Preserving location and absence privacy in geo-social networks. Paper presented at the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, October 26–30; pp. 309–18.

Furini, Marco, and Valentina Tamanini. 2015. Location privacy and public metadata in social media platforms: Attitudes, behaviors and opinions. *Multimedia Tools and Applications* 74: 9795–825. [CrossRef]

Gambs, Sebastian, Marc-Oliver Killijian, and Miguel N. del Prado Cortez. 2010. Show me how you move and I will tell you who you are. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*. New York: ACM, pp. 34–41.

Gambs, Sébastien, Olivier Heen, and Christophe Potin. 2011. A comparative privacy analysis of geosocial networks. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*. New York: ACM, pp. 33–40.

Ganta, Srivatsava Ranjit, Shiva Prasad Kasiviswanathan, and Adam Smith. 2008. Composition attacks and auxiliary information in data privacy. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, pp. 265–73.

Girardin, Fabien, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. 2008. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing* 7: 36–43. [CrossRef]

Graham, Christopher. 2012. *Anonymisation: Managing Data Protection Risk Code of Practice*. Wilmslow and Cheshire: Information Commissioner's Office.

Gruteser, Marco, and Dirk Grunwald. 2003. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*. New York: ACM, pp. 31–42.

Hampton, Khrisen H., Molly K. Fitch, William B. Allshouse, Irene A. Doherty, Dionne C. Gesink, Peter A. Leone, Marc L. Serre, and William C. Miller. 2010. Mapping Health Data: Improved Privacy Protection With Donut Method Geomasking. *American Journal of Epidemiology* 172: 1062–69. [CrossRef] [PubMed]

Hasan, Samiul, and Satish V. Ukkusuri. 2014. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies* 44: 363–81. [CrossRef]

Hays, James, and Alexei A. Efros. 2008. IM2GPS: Estimating geographic information from a single image. Paper presented at CVPR 2008. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, June 23–28; pp. 1–8.

Hu, Peizhao, Sherman SM Chow, and Asma Alou. 2017. Geosocial query with User-Controlled Privacy. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. New York: ACM, pp. 163–72.

ICO. 2012. *Crime-Mapping and Geo-Spatial Crime Data: Privacy and Transparency Principles*. Wilmslow: Information Commissioner's Office.

Kalnis, Panos, Gabriel Ghinita, Kyriakos Mouratidis, and Dimitris Papadias. 2007. Preventing location-based identity inference in anonymous spatial queries. *IEEE Transactions on Knowledge and Data Engineering* 19: 1719–33. [CrossRef]

Kawakubo, Hidetoshi, and Keiji Yanai. 2011. Geovisualrank: A ranking method of geotagged imagesconsidering visual similarity and geo-location proximity. In *Proceedings of the 20th International Conference Companion on World Wide Web*. New York: ACM, pp. 69–70.

Keßler, Carsten, and Grant McKenzie. 2018. A geoprivacy manifesto. *Transactions in GIS* 22: 3–19. [CrossRef]

Kounadi, Ourania, and Michael Leitner. 2016. Adaptive areal elimination (AAE): A transparent way of disclosing protected spatial datasets. *Computers, Environment and Urban Systems* 57: 59–67. [CrossRef]

Kounadi, Ourania, and Bernd Resch. 2018. A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics* 13: 203–22. [CrossRef] [PubMed]

Kounadi, Ourania, Alina Ristea, Michael Leitner, and Chad Langford. 2018. Population at risk: Using areal interpolation and Twitter messages to create population models for burglaries and robberies. *Cartography and Geographic Information Science* 45: 205–20. [CrossRef] [PubMed]

Kovacs-Gyori, Anna, Alina Ristea, Clemens Havas, Bernd Resch, and Pablo Cabrera-Barona. 2018. London2012: Towards Citizen-Contributed Urban Planning Through Sentiment Analysis of Twitter Data. *Urban Planning* 3: 75–100. [CrossRef]

Krumm, John. 2007. Inference attacks on location tracks. In *Pervasive Computing*. Edited by Anthony LaMarca, Marc Langheinrich and Khai N. Truong. Berlin and Heidelberg: Springer, pp. 127–43.

Laituri, Melinda, and Kris Kodrich. 2008. On line disaster response community: People as sensors of high magnitude disasters using internet GIS. *Sensors* 8: 3037–55. [CrossRef] [PubMed]

Lampoltshammer, Thomas J., Ourania Kounadi, Izabela Sitko, and Bartosz Hawelka. 2014. Sensing the public's reaction to crime news using the 'Links Correspondence Method'. *Applied Geography* 52: 57–66. [CrossRef] [PubMed]

Leitner, Michael, Jacqueline W. Mills, and Andrew Curtis. 2007. Can Novices to Geospatial Technology Compromise Spatial Confidentially? *Kartographische Nachrichten('Cartographic News')* 57: 78–84.

Li, Linna, and Michael F. Goodchild. 2013. Is privacy still an issue in the era of big data?—Location disclosure in spatial footprints. Paper presented at 21st International Conference on Geoinformatics (GEOINFORMATICS), Kaifeng, June 20–22.

Li, Yunpeng, Noah Snavely, and Daniel P. Huttenlocher. 2010. Location recognition using prioritized feature matching. In *European Conference on Computer Vision*. Berlin and Heidelberg: Springer, pp. 791–804.

Li, Huaxin, Haojin Zhu, Suguo Du, Xiaohui Liang, and Xuemin Shen. 2016. Privacy leakage of location sharing in mobile social networks: Attacks and defense. *IEEE Transactions on Dependable and Secure Computing* 15: 646–60. [CrossRef]

Li, Jin, Hongyang Yan, Zheli Liu, Xiaofeng Chen, Xinyi Huang, and Duncan S. Wong. 2017. Location-sharing systems with enhanced privacy in mobile online social networks. *IEEE Systems Journal* 11: 439–48. [CrossRef]

Lv, Qin, Moses Charikar, and Kai Li. 2004. Image similarity search with compact data structures. In *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*. New York: ACM, pp. 208–17.

Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1: 3. [CrossRef]

Malleson, Nick, and Martin A. Andresen. 2015. The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science* 42: 112–21. [CrossRef]

McKenzie, Grant, Krzysztof Janowicz, and Dara Seidl. 2016. Geo-privacy beyond coordinates. In *Geospatial Data in a Changing World*. Berlin and Heidelberg: Springer, pp. 157–75.

Mokbel, Mohamed F., Chi-Yin Chow, and Aref Walid G. 2006. The new Casper: Query processing for location services without compromising privacy. Paper presented at 32nd International Conference On Very Large Data Bases, Seoul, Korea, September 12–15; pp. 763–74.

Noorda, Catrien W, and Stefan Hanloser. 2011. *E-Discovery and Data Privacy: A Practical Guide*. Alphen aan den Rijn: Kluwer Law International.

Pan, Bei, Yu Zheng, David Wilkie, and Cyrus Shahabi. 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York: ACM, pp. 344–53.

Pontes, Tatiana, Marisa Vasconcelos, Jussara Almeida, Ponnurangam Kumaraguru, and Virgilio Almeida. 2012. We know where you live: Privacy characterization of foursquare behavior. In *Proceedings of the 2012 ACM Conference On Ubiquitous Computing*. New York: ACM, pp. 898–905.

Preoţiuc-Pietro, Daniel, and Trevor Cohn. 2013. Mining user behaviours: A study of check-in patterns in location based social networks. In *Proceedings of the 5th Annual ACM Web Science Conference*. New York: ACM, pp. 306–15.

Puttaswamy, Krishna P. N., and Ben Y. Zhao. 2010. Preserving privacy in location-based mobile social applications. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*. New York: ACM, pp. 1–6.

Resch, B. 2013. People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In *Progress in Location-Based Services*. Berlin and Heidelberg: Springer, pp. 391–406.

Resch, Bernd, Anja Summa, Peter Zeile, and Michael Strube. 2016. Citizen-centric urban planning through extracting emotion information from Twitter in an interdisciplinary space-time-linguistics algorithm. *Urban Planning* 1: 114–27. [CrossRef]

Resch, Bernd, Florian Usländer, and Clemens Havas. 2018. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science* 45: 362–76. [CrossRef]

Resch, Bernd, Alexander Zipf, Euro Beinat, and Marc Boher. 2012. Towards the Live City–Paving the Way to Real-time Urbanism. *International Journal on Advances in Intelligent Systems* 5: 470–82.

Ristea, Alina, Justin Kurland, Bernd Resch, Michael Leitner, and Chad Langford. 2018. Estimating the spatial distribution of crime events around a football stadium from georeferenced tweets. *ISPRS International Journal of Geo-Information* 7: 43. [CrossRef]

Sagl, Günther, Bernd Resch, Bartosz Hawelka, and Euro Beinat. 2012. From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In *Proceedings of the GI-Forum*. Berlin: Herbert Wichmann Verlag, pp. 54–63.

Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*. New York: ACM, pp. 851–60.

Santillana, Mauricio, André T. Nguyen, Mark Dredze, Michael J. Paul, Elaine O. Nsoesie, and John S. Brownstein. 2015. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Computational Biology* 11: e1004513. [CrossRef] [PubMed]

Schulz, Axel, Aristotelis Hadjakos, Heiko Paulheim, Johannes Nachtwey, and Max Mühlhäuser. 2013. A Multi-Indicator Approach for Geolocalization of Tweets. Paper presented at Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, USA, July 8–11; pp. 573–82.

Shokri, Reza, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. 2011. Quantifying location privacy. Paper presented at 2011 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, May 22–25; pp. 247–62.

Steiger, Enrico, Bernd Resch, João Porto de Albuquerque, and Alexander Zipf. 2016a. Mining and correlating traffic events from human sensor observations with official transport data using self-organizing-maps. *Transportation Research Part C: Emerging Technologies* 73: 91–104. [CrossRef]

Steiger, Enrico, Bernd Resch, and Alexander Zipf. 2016b. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science* 30: 1694–716. [CrossRef]

Steiger, Enrico, René Westerholt, Bernd Resch, and Alexander Zipf. 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems* 54: 255–65. [CrossRef]

Sun, Yongqiang, Nan Wang, Xiao-Liang Shen, and Jacky Xi Zhang. 2015. Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior* 52: 278–92. [CrossRef]

Sweeney, Latanya. 2002. K-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10: 557–70.

Vicente, Carmen Ruiz, Dario Freni, Claudio Bettini, and Christian S. Jensen. 2011. Location-related privacy in geo-social networks. *IEEE Internet Computing* 15: 20–27. [CrossRef]

Wang, Shuo, Richard Sinnott, and Surya Nepal. 2016. Protecting the location privacy of mobile social media users. In *2016 IEEE International Conference on Big Data*. Piscataway: IEEE, pp. 1143–50.

Wei, Wei, Fengyuan Xu, and Qun Li. 2012. Mobishare: Flexible privacy-preserving location sharing in mobile online social networks. Paper presented at 2012 Proceedings IEEE INFOCOM, Orlando, FL, March 25–30; pp. 2616–20.

Xue, Di, Li-Fa Wu, Hua-Bo Li, Zheng Hong, and Zhen-Ji Zhou. 2017. A novel destination prediction attack and corresponding location privacy protection method in geo-social networks. *International Journal of Distributed Sensor Networks* 13. [CrossRef]

Zakhary, Victor, Cetin Sahin, Theodore Georgiou, and Amr El Abbadi. 2017. Locborg: Hiding social media user location while maintaining online persona. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York: ACM, p. 12.

Zang, Hui, and Jean Bolot. 2011. Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*. New York: ACM, pp. 145–56.

Zhang, Jia Dong, Gabriel Ghinita, and Chi Yin Chow. 2014. Differentially private location recommendations in geosocial networks. Paper presented at 2014 IEEE 15th International Conference on Mobile Data Management, Brisbane, QLD, Australia, July 14–18; vol. 1, pp. 59–68.

Zhang, Wei, and Jana Kosecka. 2006. Image Based Localization in Urban Environments. Paper presented at Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), Chapel Hill, NC, USA, June 14–16; vol. 6, pp. 33–40.

Zimmerman, Dale L., and Claire Pavlik. 2008. Quantifying the effects of mask metadata disclosure and multiple releases on the confidentiality of geographically masked health data. *Geographical Analysis* 40: 52–76. [CrossRef]