Ethics and privacy in online social networks

By Simona Bisiani and Jan Sodoge

Big Data: Social Processes and Ethical Issues

Fall Semester 2020

1 Introduction

Digitalization contributed to a non-linear increase in the size of digitally generated and stored data. Within this expansion, online social networks (OSN) record significant growth in size of data as well as membership numbers (Richterich, 2018). The data OSN generate is considered highly valuable from the perspectives of both businesses and academic research, as it reveals information on human behavior and societal functioning. Thus, the appeal of OSN data led to its utilization within various research fields. Within this essay OSN are defined following three central components delineated by Body and Ellison (2007): (i) user-level profiles ranging between a public or semi-public status, (ii) a network structure holding ties between user profiles, (iii) personal ties being visible to a user as well as further ties within the overall network. The public accessibility of ties between users represents a feature that differentiates OSN from other web platforms (e.g. forums). The same feature introduces significant effects on individual privacy, which we investigate within this essay. Privacy is traditionally defined following Westin (1970) as "the ability for people to determine for themselves when, how, and to what extent information about them is communicated to others" (Mondal et al., 2014: 6). Within the context of scientific research, the issue of privacy in OSN is becoming increasingly relevant as academic and non-academic research using OSN data is becoming more ubiquitous. Meanwhile, the collection of user-level data in OSN is increasing. As we will outline in this essay, OSN fundamentally challenge the idea of privacy by Westin (1970). Considering the idiosyncratic features of OSN compared to other digital platforms and online services, we will discuss particular challenges and specifics of OSN with respect to privacy. While the general usage of online-services introduces disclosure of personal information, there are significant differences that singularize OSN. Consider the example of a person using a credit card service: as a purchase takes place, personal information of the individual such as the purchased service or product, time, location, etc. are collected. This process takes place and can be considered in isolation from the actions of other customers using credit card services. Instead, OSN display complex interactions and mechanisms as these platforms are used by modes of social interaction between a large number of individuals connected within social network structures. Disentangling the actions of individuals from such larger network structures with respect to privacy imposes difficulties. Such challenges are explored and discussed within this essay, which structures as follows. First, we present three features of datafication that singularize OSN with regard to predictive analytics. Second, we highlight challenges on privacy in OSN based on these features. Third, we discuss the conceptualization of privacy in OSN based on current and alternative privacy paradigms and reflect on the idea of networked privacy as a promising alternative available in the existing literature.

2 Particular features of OSN for predictive analytics

The progression of datafication in its extent and capabilities, which captures the idea of storing information on the social world as machine-readable data, has been enabled by the assembly and retention of large datasets and the algorithmic innovations within predictive analytics (Richterich, 2018). Predictive analytics has established itself as a practice from the field of data engineering, and refers to the extraction of data and its consequent *mining*, seeking for patterns and information that is generated using statistical and mathematical techniques, such as community detection, dimensionality reduction and social network analysis (Mishra and Silakari, 2012). Holding large and increasing numbers of users who interact and disclose personal characteristics, OSN developed as a key area for datafication and predictive analytics (Agrawal et al., 2014). In this essay we focus on three dimensions of datafication and predictive analytics, specific to OSN, which impose consequences for individuals' privacy: social network features, user-generated textual content, and location-based information.

2.1 Social network features

Various features that sociological and network-science research found in social networks of the physical world are proven to exist within OSN as well. One of these properties is homophily, which describes a pattern observed in various social contexts where those individuals in a social network who are connected by ties are on average more similar to one another than to those with whom no ties are shared (McPherson et al., 2001). For example, Mercken et al. (2012) show that there is homophily with respect to smoking behavior in friendship networks in schools among early adolescents. Research on OSN provides similar findings, where individuals are significantly more similar (based on voluntarily disclosed personal information) with their peers in OSN (Gayo-Avello et al., 2011). Using this finding for predictive analytics, a substantial strand of literature provides evidence on how network characteristics based on sociological theory like homophily, change the (in-)voluntary disclosure of personal attributes. A study by Mislove et al. (2010) provides a first example where the authors mine data from two OSN and use information about social network structures and attributes of peers for a particular individual (ego) with the aim of inferring a given ego characteristics. Following up on this study, further publicized research presented statistical models which are able to draw inferences from social ties to personal attributes of ego and share high levels of certainty in prediction (e.g. Gayo-Avello et al., 2011; Sarigol et al., 2014). While these studies focused on prediction based on friendship ties of a particular individual and available information on peers, research by Zhelva et al. (2009) showed that personal attributes can be inferred from group-level structures in OSN, too. The presumption of the study, that groups act as carriers of information that reflect individual-level characteristics, can be linked to sociological research on culture and group-identities by DiMaggio (1997). Zhelva et al. (2009) use group-level classification algorithms to make inferences on the attributes of individuals. Their presented statistical model is capable of discovering group-structures within friendship networks and respectively identifying groups an individual belongs to. As users can be identified as members of groups of individuals with similar attributes, the personal information disclosed by other group members can be used to infer attributes of individuals. Thereby, this approach transcends the previously discussed approach that instead focusses on the level of social ties to peers treated in isolation. Yet a third approach is reflected in the literature, and it highlights the potential of inferring personal characteristics based on diversity in the social ties in the OSN (instead of homophily). Following the argument by Backstrom et al. (2014), the ties of an individual reflect different relationships such as co-workers, childhood friends, geographically close individuals, etc. The ability to draw inferences based on diversity in social ties is exemplified in a study by Backstrom et al. (2014), where the authors try to use the information on the strengths of the social ties of an individual to infer the person's romantic partner. Compared to the prior studies cited, the authors develop a more nuanced understanding of ties by measuring tie strength and further tie-related characteristics. Based on tie characteristics and strengths, the authors present a predictive model that is capable to reliably predict romantic relationships between individuals. Overall, the three presented approaches highlight how social network structures in OSN enable inference of personal attributes with high levels of accuracy.

2.2 User-generated textual content

While social network structures operate as clues to the reconstruction of a person's private attributes, research has developed within further dimensions of datafication in OSN to serve similar investigative purposes. Language is one dimension, yet less explored within academia, that acts as a tool for OSN-based inferencing. Research in sociolinguistics shows that the way we speak, in the vocabulary we use, the topics we bring up, and even in our usage of punctuation, is not fully idiosyncratic, and that discourse-variation correlates with individual-level attributes (Labov, 1972; Coates, 1996; Macauley, 2005). Language serves the purpose of communicating to others just as it operates as a tool for status display, and proof of belonging and identification with particular groups. Also, language is homophilous, and we are likely to express shared identities with peers or social groups throughout common expressions or slang, may those be ethnicity, religion, gender or social class based or nationality bounded (Labov, 1972; Coates, 1996; Macauley, 2005). These findings bear substantial implications for inferring of personal characteristics: language is predictable, and thus can be used for prediction of personal characteristics. The availability of publicly available text uploaded by users in OSN creates powerful datasets for the development of classifier-algorithms and models that place individuals in groups by their linguistic choices and style. Such an approach is presented in an innovative study by Rao et al. (2010) where the authors use a large Twitter dataset and machine learning to train a classification algorithm and infer user characteristics such as gender, age, political orientation and religion. The results show that variation in linguistic choices across groups at the four different levels exists, and the researchers' models significantly outperform baseline models.

2.3 Location-based information

Often, when we think of OSN we do so within the context of digital spaces, where tangible geographical dimensions lose their traditional definitions and relevance. However, location is embedded within OSN in different ways, and thus can either be exploited to make inferences, or can be inferred. The form and degree to which geographic information features OSN varies. For example, some OSN might allow users to tag the location of a picture, or to specify the city of provenance or residence. Other OSN are more specifically defined as location-based social networks (LSBN) and use location as a central feature to their services. For instance, within the LSBN *Foursquare* users contribute by describing places they have visited. There is little doubt that geographic information related to an individual carries sensitive value (Pangburn, 2017). As Pangburn (2017) points out, from the home address to the precise location of a person in the present moment, unveiling such private details of an individual means accessing information that can be used to learn about a person's whereabouts or deliver more effective advertising, exploiting personalization of content and timeliness of advertisement delivery. Critically, the sensitivity of geographical personal information is recognized by companies too, reflected in Foursquare's homepage slogan: "With uncompromising accuracy, accessibility, scale, and respect for consumer privacy, Foursquare is the location platform the world trusts" (Foursquare, 2020). A study by Pontes et al. (2012) avails of data from Foursquare, Google+ and Twitter to infer an individual's home location by using publicly available geographic information uploaded by the users, as well as the users' friends disclosed home location. The developed model showed capable to estimate a person's home city with accuracy of 74%, and for a smaller subset of users the home residence was inferred within a six kilometers radius with accuracy of 60% for Foursquare and Twitter, and just around

10% for Google+. While this research highlights that what a person chooses to upload, combined with his friends' choices on what to disclose, can be used to indirectly infer on something more private such as the home address, further research has reversed the mechanism and used available geographic information to infer individuals' social relations. In this context, Crandall et al. (2010) build an inference model of social ties based on geographic co-occurrences. Their research fundamentally answers the question: when two people appear in nearby locations a given number of times, what is the likelihood that they know each other? Using a dataset from Flickr, an OSN where users showcase photographic work which can be manually or automatically geotagged, their research demonstrates that it takes very few co-occurrences to infer the underlying social network structure of individuals. As pointed out in the paper, this type of inference highly intertwines with questions about *coincidences*. When two people *coincide* in time and space, what is the probability that this is merely a *coincidence* (Crandall et al., 2010)? And how does that probability change as the number of spatio-temporal co-occurrences increases? Fundamentally, Crandall et al. (2010) remind us that probabilistic models, combined with OSN, are a powerful tool to learn about the world, as they can intricately and yet accurately shine light on the resemblances, the patterns, and the commonships of human behavior.

3 Ethical considerations on privacy in OSN

As we learn how to use OSN to deepen our knowledge of social mechanisms through the types of research presented in the previous sections, a discussion within the literature emerges about defining and setting limits to the prediction of personal-level attributes with respect to ethical and privacy considerations. Based on the previously outlined inferential practices in OSN, this section discusses the challenges these pose to privacy. Here, we first consider implications based on individual privacy and the definition of privacy brought forward by Westin (1970). We explore the possibilities for ethical boundaries to be set within the context of the crucial challenge of circumnavigating the gist of public and academic discourse that OSN in themselves pose as a threat to individuals' privacy. In a second step, we discuss to what degree the idea of individual privacy in OSN can consequently continue to hold relevant. Based on the review of inferential practices in section 2, a discussion arises on the extent of consumer privacy and its preservation. A resulting debate can be exemplified for language based inferences. The study of how communication and language works in the online, given the constraints on length of messages and other features of OSN, can generate invaluable insights on how society comes together today. However, the task of inferring individual private characteristics carries with itself ethical concerns. Particularly, research using tools to bypass users' privacy choices and unmask users' characteristics for the purpose of "advertising, personalization, and recommendation" (Rao et al., 2010:44), requires careful considerations. In their paper, Rao et al. (2010) state as a justification for the study the interest of inferring attributes users have chosen to keep private, but how can stripping individuals of their protected identity be justified when it is deliberately undisclosed, particularly in the aim of delivering them targeted advertising? Ultimately, these are considerations that need to be addressed, especially as language is so publicly accessible in OSN, and communication in online platforms is likely to increase. The lack of ethical considerations by Rao et al. (2010), particularly the risk of compromising an individual's privacy, is concerning. Their research provides an example of predictive analytics where the inferential practice is done without (i) an ethical framework posing boundaries to the study and (ii) a clear indication of how subjects are protected and defended in their right to privacy. Considering more abstract challenges to individual privacy, multiple aspects need to be considered. We review literature and identify the following academic and public themes that we discuss in the following section.

Privacy-related decision-making

Knowledge that predictions can be made, with high level of accuracy, revealing your sensitive and undisclosed information, can act as a reinforcing mechanism to the privacy paradox (Kokolakis, 2017). The privacy paradox refers to the inconsistency between users' attitudes towards privacy and their behaviour online, which has found enormous support in the literature (see Kokolakis, 2017 for a review). Lack of control over one's private information due to OSN inferential practices can contribute to the decisions individuals make in relation to their online privacy. Stemming from behavioural economics, the literature on privacy-related decision-making shows that people are limited by incomplete information, time and resources, when deciding upon their personal information disclosure (Acquisti, 2004; Kokolakis, 2017). Acquisti's (2014) formal model of the privacy paradox shows that individuals act following an *immediate gratification* heuristic: the benefit of disclosing personal information in the present has higher perceived value than the future privacy risks. It remains untested, to the best of our knowledge, how OSN and inferential practices affect the perception of future privacy risks. We hypothesize that the complexity of the inferential practices and the possibility to bypass an individual's active choice to hide or reveal personal information, might hold some weight within the privacy decision function of a person, though it is at present unknown whether it is the case, and in which way it might do so.

Informed Consent and Transparency

Questions of privacy and ethics are inevitably questions of informed consent, as Richterich (2018) hints. Informed consent has been a critical pillar of the moral framework guiding research ever since the Nuremberg Code of 1947. In Big Data ethics, a central challenge is in the definition itself of informed consent. What type of practices are subjects informed about? Multiple reported cases such as the Target store prediction of a customer pregnancy status (Duhigg, 2012) shine light on the fact that predictive analytics generates controversies when defining whether informed consent is given or not. Is publicly accessible information ethically acceptable to use for research? And does the same apply when it is used to create new information about an individual and his connections? Ultimately, informed consent highly intertwines with questions about transparency, and the subjects of an inferential study knowing that their data is and can be used in such a way. OSN companies, in particular, often do not disclose the practices they carry out to generate advertising revenue or improve their platforms. These practices to a significant degree are achieved by means of increasing personalization of the platform, obtained through a deeper understanding of human behaviour. Lack of transparency is thus a central challenge to ethical debates around OSN and predictive analytics. Within this context, recent years have seen the emergence of a debate around the ethics of a particular practice of OSN companies, in what has become known as *shadow profiles*. *Shadow profiles* describe a practice of OSN providers which uses predictive models to collect personal information a user does not disclose. Crucially, these are maintained and build privately by OSN providers aside of user agreement, permission or terms of service. A reported case of Facebook in 2013 showed such practices exist (Blue, 2013). Following the report, Facebook collected data of phone numbers from mobile phone books of their users. Thereby, Facebook was able to infer phone numbers of individuals who did not directly disclose these but were stored in peers' phone books. More generalizing evidence is provided in a study by Garcia (2017) who shows the capabilities to construct shadow profiles of users. Here, the number of individuals within the analyzed social network as well as the openness to disclose attributes among users show significant impact on the abilities to predict shadow profile data.

Sample Bias and Discrimination

OSN, a source of so-called Big Data, provide data that appeals thanks to its renowned characteristics: velocity, volume and variety (Laney, 2001). The societal hype for Big Data, however, might diverge attention from its flaws. As Boyd (2012) points out, OSN do not directly offer what can be considered an equivalent to a research's well-considered, methodically designed sample. Rather, the socio-demographic characteristics of the digital population within an OSN are non-representative, pointing rather to a sub-set of the general population (Boyd, 2012). Additionally, as Boyd (2012) mention, not all accounts are used in the same way: some people have more than one account, some accounts are held by more than one person, and there are bots acting as individuals. This ultimately serves as a reminder that the predictions we make are merely built on, and for, a the type of individuals that populate OSN specifically, and thus they are far from being useful statements applicable to the society at large. Moreover, insights such as the classification of individuals using language into different genders, political groups and religion, can be inappropriately used by information holders. An example of this is provided by the controversial study carried out by AI researchers Kosinsky and Wang (2017), which used the profile pictures of members of an online dating community to predict sexual orientation. The findings were quite remarkable, as the authors were able to predict with 91% accuracy whether a male member defined himself as heterosexual or gay, while for women the figure sat at 83%. The implication of such study are substiantial: classifying individuals by their biological and facial traits can lead to discrimination by actors opposing homosexuality, may those be an employer or a government.

3.1 Propositions for privacy in OSN

Our review of the research shining light on OSN's embedded threats to privacy points to two main intuitions. First, there is a growing awareness of what can be learned about individuals by the means of datafication and predictive analytics. Second, we are simultaneously gaining insights as to what it takes to learn about certain individual-level attributes. Zheleva and Getoor (2009) discuss one way in which these insights can be used to protect individuals from what seems to be *unprotectable* in the online. Individuals can be educated in relation to OSN and inference, and made aware of what type of insights attributes, activities and associations generate, as demonstrated by research. Such practice is discussed to empower individuals and enable them to make better choices in their privacy personalization settings of different OSN. Zheleva and Getoor (2009) exemplify this idea for group-based inference: knowing that groups' homogeneity makes the prediction of personal attributes highly accurate, individuals could use this information to reflect on their group characteristics, and if concerned with their privacy, try to diversify their group properties. Reflecting on the authors' proposal, we perceive difficulties in its practical implementation: groups evolve overtime both structurally and internally, meaning that individuals change and add information within their personal profiles, altering fundamentally the predictable outcomes for any given individual in the network. Also, the extent to which individuals can educate themselves about inference is limited, as new tools and new insights are also dynamically generated. Furthermore, understanding of the given models and underlying statistics is limited, which can hinder the learning and empowering process. Finally, the insights we are aware of are limited to those publicly available. Knowledge generated internally to the OSN provider remains often undisclosed, yet they are arguably the ones carrying the most value, given the quality of the data in their possession. Nonetheless, educating individuals gives them the opportunity to be aware of the burden carried by seemingly distant and unrelated choices, and the weight brought further by their digital social circles. From a structural perspective, OSN, in themselves, carry risks to individual privacy. Developing an OSN which protects users from inferential practices would thus imply to strip an OSN of its key characteristics: a social platform where individuals generate information about themselves and the world as they go along, where they are linked to each other, and can see those linkages (Body and Ellison, 2007). This said, there are at times easily identifiable improvements within the design of an OSN, that would help individuals protect their identity. For example, Zheleva and Getoor (2009) mention the possibility for individuals to be able to mask their affiliations to public groups, and make that visible only to friends, a feature that at that time of writing was not available on neither Facebook nor Flickr.

While the prior propositions are situated at the level of individual privacy, a stream of literature has emerged discussing the appropriateness of the concept of individual privacy within the context of digital platforms. The origin of this discourse by scholars (see Garcia, 2017) relates to the findings discussed within section 2, that stress that it is not entirely up to oneself to shape personal privacy in social networks. Following a study by Bargow et al. (2019) which uses Twitter data to predict individual-level behavior, the predictions inferred from social network ties of an individual had a higher prediction accuracy than those inferred from user-level data. Consequently, the decision of disclosing personal information (and thus privacy in OSN) is not governed entirely by individual agency but shifts towards a collective level. Critically, this shift has significant implications for privacy and policy (Sarigol, 2014), as recognized in the literature with the call for *networked* *privacy* by scholars like Garcia (2019). Contemporary models of privacy in OSN are categorized as *access control* models within the literature (Mondal et al., 2014). Here, to achieve privacy in the online realm users are able to restrict other's access personal information based on specific persons, groups or roles and by individually deciding to disclose or not disclose personal information. Mondal et al. (2014) argue that this is an outdated form and needs to be replaced. The argument is based on different historic events (introduction of Facebook News Feed, data aggregation Spokeo, Google Street View), where these did not breach privacy according to the concept of access control. Scholars following up on the work of Westin stress that the functions of privacy do not take place in isolation i.e. on an individual-level with no interactions to other members of societal structures. Instead, relational features of privacy are highlighted (Cohen, 2012), where privacy is not a binary state but instead contextual within networks. Thus, Bannerman (2019) argues that the conceptualization of privacy by Westin is increasingly outdated in OSN as the networked context and structure of these cannot be ignored and isolation is just not possible. As previously stated, we found in the literature calls for an alternative paradigm, conceptualized as *networked privacy.* Marwick and Boyd (2014) discuss this paradigm to be a consequence of findings we previously reviewed. They highlight the importance of individual agency in governing one's own privacy while they stress social network structures impose hardship on maintaining agency as a result of the difficulties for an individual to take account of the social context and influence and behavior of peers. Thus, to achieve privacy following this paradigm, privacy needs to equip individuals with the ability of controlling the information that arises from the social network structure and flows within it. On a level of implementation, this is described to imply equipping individuals with knowledge and authority to "shaping the context in which information is being interpreted" (Marwick and Boyd, 2014:1063). Consequently, a shift in the paradigm of privacy can be accounted as a solution providing privacy while maintaining the funcionality of OSN.

4 Conclusion

In this essay, we bring together research on inferential practices within OSN and provide an overview of the different applications for predictive analytics made possible by the datafication of OSN. Multiple aspects of OSN can be used to *mine for meaning* and information about individuals. We review how social network features, language and location-based information act as the ingredients to the prediction of individual-level characteristics. Critically, the nature and inherent features of OSN can serve as a tool to unveil information about an individual which he has chosen not to disclose. It is the availability of information of social networks and the characteristics of nodes within those that allow to generate predictions on hidden information for any particular individual in the network. We discuss how such opportunity to bypass the readily and publicly available information about an individual and infer additional knowledge about a person, poses threats to an individual's privacy, and raises ethical questions in regards to transparency, informed consent, and algorithmic bias. Finally, we discuss networked privacy as a new emerging privacy paradigm that accounts for the shared responsibility towards privacy in OSN. We identify, in what Marwick and Boyd (2014) call *networked privacy*, the key to protect individuals' privacy in the complex and often undiscernible inferential practices carried out within OSN.

References

[1] Acquisti, Alessandro. ”Privacy in electronic commerce and the economics of immediate gratification.” In Proceedings of the 5th ACM conference on Electronic commerce, pp. 21-29. 2004.

[2] Agrawal, Divyakant, Ceren Budak, Amr El Abbadi, Theodore Georgiou, and Xifeng Yan. ”Big data in online social networks: user interaction analysis to model user behavior in social networks.” In International Workshop on Databases in Networked Information Systems, pp. 1-16. Springer, Cham, 2014.

[3] Backstrom, Lars, and Jon Kleinberg. ”Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook.” In Proceedings of the 17th ACM conference on Computer supported cooperative work social computing, pp. 831-841. 2014.

[4] Bannerman, Sara. ”Relational privacy and the networked governance of the self.” Information, Communication Society 22, no. 14 (2019): 2187-2202.

[5] Bagrow, James P., Xipei Liu, and Lewis Mitchell. ”Information flow reveals prediction limits in online social activity.” Nature Human Behaviour 3, no. 2 (2019): 122-128.

[6] Blue, Violet. ”Anger mounts after Facebooks shadow profiles leak in bug”. 2013. https://www.zdnet.com/article/anger-mounts-after-facebooks-shadow-profiles-leak-in-bug/

[7] Boyd, Danah, and Kate Crawford. ”Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon.” Information, communication society 15, no. 5 (2012): 662-679.

[8] Boyd, Danah M., and Nicole B. Ellison. ”Social network sites: Definition, history, and scholarship.” Journal of computer-mediated communication 13, no. 1 (2007): 210-230.

[9] Coates, Jennifer. ”Women talk: Conversation between women friends.” (1996): 265-268.

[10] Cohen, Julie E. Configuring the networked self: Law, code, and the play of everyday practice. Yale University Press, 2012.

[11] Crandall, David J., Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg. ”Inferring social ties from geographic coincidences.” Proceedings of the National Academy of Sciences 107, no. 52 (2010): 22436-22441.

[12] Debatin, Bernhard, Jennette P. Lovejoy, Ann-Kathrin Horn, and Brittany N. Hughes. ”Facebook and online privacy: Attitudes, behaviors, and unintended consequences.” Journal of computer-mediated communication 15, no. 1 (2009): 83-108.

[13] DiMaggio, Paul. ”Culture and cognition.” Annual review of sociology 23.1 (1997): 263-287.

[14] Duhigg, Chris. ”How Companies Learn Your Secrets”. New York Times. (2012). https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

[15] Erdős, Dóra, Rainer Gemulla, and Evimaria Terzi. ”Reconstructing graphs from neighborhood data.” ACM Transactions on Knowledge Discovery from Data (TKDD) 8.4 (2014): 1-22.

[16] Garcia, David. ”Leaking privacy and shadow profiles in online social networks.” Science advances 3, no. 8 (2017): e1701172.

[17] Gayo Avello, Daniel. "All liaisons are dangerous when all your friends are known to us." In Proceedings of the 22nd ACM conference on Hypertext and hypermedia, pp. 171-180. 2011.

[18] Horvát, Emöke-Ágnes, Michael Hanselmann, Fred A. Hamprecht, and Katharina A. Zweig. "One plus one makes three (for social networks)." PloS one 7, no. 4 (2012): e34740.

[19] Kim, Myunghwan, and Jure Leskovec. "The network completion problem: Inferring missing nodes and edges in networks." In Proceedings of the 2011 SIAM International Conference on Data Mining, pp. 47-58. Society for Industrial and Applied Mathematics, 2011.

[20] Kokolakis, Spyros. "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon." Computers and security 64 (2017): 122-134.

[21] Labov, William. Language in the inner city: Studies in the Black English vernacular. No. 3. University of Pennsylvania Press, 1972.

[22] Laney, Doug. "3D data management: Controlling data volume, velocity and variety." META group research note 6, no. 70 (2001): 1.

[23] R. K. Macaulay. Talk that counts: Age, Gender, and Social Class Differences in Discourse. Oxford University Press, 2005.

[24] Marwick, Alice E., and Danah Boyd. "Networked privacy: How teenagers negotiate context in social media." New media & society 16.7 (2014): 1051-1067.

[25] Mercken, Liesbeth, Christian Steglich, Philip Sinclair, Jo Holliday, and Laurence Moore. ”A longitudinal social network analysis of peer influence, peer selection, and smoking behavior among adolescents in British schools.” Health Psychology 31, no. 4 (2012): 450.

[26] Mishra, Nishchol, and Sanjay Silakari. ”Predictive analytics: A survey, trends, applications, oppurtunities & challenges.” International Journal of Computer Science and Information Technologies 3, no. 3 (2012): 4434-4438.

[27] Mislove, Alan, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. ”You are who you know: inferring user profiles in online social networks.” In Proceedings of the third ACM international conference on Web search and data mining, pp. 251-260. 2010.

[28] Mondal, Mainack, Peter Druschel, Krishna P. Gummadi, and Alan Mislove. ”Beyond access control: Managing online privacy via exposure.” In Proceedings of the Workshop on Useable Security, pp. 1-6. 2014.

[29] Pangburn, DJ. Even This Data Guru Is Creeped Out By What Anonymous Location Data Reveals About Us. (2017). https://www.fastcompany.com/3068846/how-your-location-data-identifies-you-gilad-lotan-privacy

[30] McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. ”Birds of a feather: Homophily in social networks.” Annual review of sociology 27.1 (2001): 415-444.

[31] Pontes, Tatiana, Gabriel Magno, Marisa Vasconcelos, Aditi Gupta, Jussara Almeida, Ponnurangam Kumaraguru, and Virgilio Almeida. ”Beware of what you share: Inferring home location in social networks.” In 2012 IEEE 12th International Conference on Data Mining Workshops, pp. 571-578. IEEE, 2012.

[32] Rao, Delip, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. "Classifying latent user attributes in twitter." In Proceedings of the 2nd international workshop on Search and mining user-generated contents, pp. 37-44. 2010.

[33] Richterich, Annika. The big data agenda: Data ethics and critical data studies. University of Westminster Press, 2018.

[34] Sarigol, Emre, David Garcia, and Frank Schweitzer. "Online privacy as a collective phenomenon." In Proceedings of the second ACM conference on Online social networks, pp. 95-106. 2014.

[35] Wang, Yilun, and Michal Kosinski. "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images." Journal of personality and social psychology 114, no. 2 (2018): 246.

[36] Zheleva, Elena, and Lise Getoor. "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. 2009.