


**Please cite the Published Version**

Azzopardi, Leif, Briggs, Jo , Duheric, Melissa, Nash, Callum, Nicol, Emma, Moncur, Wendy and Schafer, Burkhard (2022) Are Taylor's posts risky? Evaluating cumulative revelations in online personal data. In: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 11 July 2022 - 15 July 2022, Madrid, Spain.

**DOI:** <https://doi.org/10.1145/3477495.3531659>

**Publisher:** Association for Computing Machinery (ACM)

**Version:** Accepted Version

**Downloaded from:** <https://e-space.mmu.ac.uk/632091/>

**Additional Information:** This is an Accepted Manuscript of a conference paper presented at SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

# Are Taylor's Posts Risky?

## Evaluating Cumulative Revelations in Online Personal Data

A persona-based tool for evaluating awareness of online risks and harms

### ABSTRACT

Searching for people online is a common search task that most of us have performed at some point or another. With so much information about people available online it is often amazing what one can find out about someone – especially when information taken from various places is pieced together to create a more detailed picture of the individual and used to make inferences about them (leading to *Cumulative Revelations*). As such the relevance of one piece of information is often conditional and dependent on the other pieces of information found – leading to interesting challenges in evaluating the “relevance” or in the case of searching personal profiles, posts, and related information about a person, the potential “risks” given these revelations. In this demonstration paper, we present a tool designed to explore how people assess and judge the relevance and the risks of *small, apparently innocuous pieces of information* associated with fictitious personas, such as “Taylor Addison” when searching and browsing online profiles. The demonstrator also acts as cyber-safety tool which aims to provide education and awareness to participants of the potential risks of cumulative revelations by working through different scenarios (where the relevance of pieces of information depends on the searcher and their “search” task).

### KEYWORDS

Information Revelation; Digital Identity; Data Self; Privacy; Cyber Safety, Training Tool

### ACM Reference Format:

. 2022. Are Taylor's Posts Risky? Evaluating Cumulative Revelations in Online Personal Data: A persona-based tool for evaluating awareness of online risks and harms. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

### 1 INTRODUCTION

In Information Retrieval, we often consider documents as discrete entities, both materially and in terms of their content as independent of each other [23]. As such, documents are often judged based on their relevance, significance and usefulness, individually, without taking account of other, even closely proximate documents. However, in practice, information from documents is typically combined and used together, including to make greater inferences. Not

only is the sum of all the information greater than its parts, through combination any apparent “gaps” can be filled, or inferred by the user of the information. This is particularly the case when looking for and retrieving information about personal information such as the information people post online. Most of us have gone online and conducted a search for someone, whether it be to find out about: a colleague – *what have they been working on?*, a friend – *what have they been up to lately?*, a date – *what are they interested in?*, or just someone we met online and are curious about – *are they a catfish?* However, from the sum of all this information, whether posted by us or about us – a digital footprint is generated that could be used maliciously by others, and result in harm, loss or detriment to the person.

So while the web, through social media, online networking sites, etc. presents an opportunity for people to build useful connections, construct personalized profiles, and so on – where they can express their personality, thoughts, feelings and other personal values (e.g., interests, opinions, livelihood, place of work, relationship status, sexual orientation, religion, etc. [11, 17, 22]) using these platforms also leaves people open and vulnerable to potential exploitation and harms. This is because the small pieces of information shared online across multiple networks and websites, individually may seem innocuous or harmless. However, over time they may reveal more cumulatively than the person intends as these traces can be linked together to make greater inferences about the individual. For example, Taylor may post messages that indicate that they live alone, while their jogging data posted online shows the routes and times that Taylor runs. Taken together, one may start to infer where Taylor lives, when and where Taylor goes, and that no one is waiting for them! Thus, these shared data can lead to revealing more about one's identity, habits, work/life patterns, personality, and so forth than a person intends – which may result in loss of privacy or worse. Clearly, such risks can have potentially negative and even disastrous consequences for the person (e.g., stalking [15], identity theft [1], financial loss [2], damage to reputation [6]), cyber-bullying [5], for their employer (e.g., by creating opportunities for cyber-crime, damage to corporate reputation, etc.), and even for national security (e.g. by revealing deployment details, security access, etc.) [10]. Each of these different risks represent different search scenarios that different actors could undertake. For example, an employer may screen potential employees by searching through social media accounts to see if they can amass a picture of the candidate and whether they have a track record of inappropriate behaviour, while a hacker may be more interested in collecting details that could be used to socially engineer access to the person's account or place of work. However, exploring and investigating such scenarios in practice is particularly challenging for a number of reasons (e.g. privacy issues over sharing the data, ethical issues over exposing individuals, curating profiles that contain sufficiently

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference'17, July 2017, Washington, DC, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

rich relationships, etc.). To overcome these challenges, we have developed a bespoke test collection (albeit small by modern standards) of fictitious personas containing curated posts, profiles, web pages, blog posts, and so forth, for participants to inspect, explore and search in order to undertake such scenarios. To further enable the exploration process, we have developed a tool for participants to search and browse the collection for each scenario – where they can rate and annotate items, first individually and then collectively, in order to assess and evaluate participants’ ability to identify information that, taken together, could increase one’s online risk. Our demonstrator is broadly positioned towards raising awareness within the general public, but is also aimed at providing training in workplace operational security to employees and educating young adults in online risks.

## 1.1 Motivation and Background

Personal online cyber-safety presents many challenges to individuals, especially as their digital footprints span and encompass many different sites and platforms. While it has been found that people say that they understand the need to protect their privacy and security online, they often do not take the necessary steps to do so [7, 12]. In addition, as people become accustomed to searching and browsing other people’s information online, they are likely to underestimate how their own online sharing behaviors “give off” insights about them to others, or even feel like such practices are the norm [3, 4, 21]. Our research shows that even among those who profess to be digitally literate, participants struggled to recall what they had shared online, and found it difficult to envision potentially harmful future scenarios emanating from their digital footprints [19]. In other studies, even large-scale data violations e.g., Cambridge Analytica, which led to increased sensitivities around information sharing, did not significantly improve reported “digital hygiene” practices [21].

So how can people use social media and other online platforms, enjoying the benefits, while minimising their risk of negative or unintended consequences? One possible solution proposed in [10, 14] is the use of a personal informatics system, that enables people to examine and reflect upon the details that they are sharing online, to increase their awareness of the privacy risks. For example, the *DataSelfie*<sup>1</sup> project provided a browser plug-in for Facebook to show individuals what their interactions online might reveal about them. Another project called *WASP* [16] provided a prototype of a personal web archive and search system, integrating archiving, indexing, and reproduction technology into a single application. *DataMirror* [13] aimed to use people’s own social media data to let them inspect and explore different aspects of their social media posts – such as the sentiment conveyed, etc. However, such solutions are technically problematic – and require the consumption and ingestion of many different feeds across many different APIs, from which the data needs to be indexed and made sense of. Within the data, there may be many thousands of posts over many years, but only some of the possible types of risks – and so reflecting on one’s own practices, while potentially insightful – may not lead to the sufficient level of awareness regarding all types of risks. Moreover, these solutions pose a number of ethical

concerns regarding how to deal with and handle an individual’s personal data – making it a legal and ethical minefield to create test collections for researching such issues. For example, such collections could be used to automate finding security vulnerabilities in people’s online digital footprints and then be exploited by nefarious actors. In this work, we side step these issues by creating fictitious personas and bespoke curated collections to help study and explore people’s information seeking and sharing behaviours and practices – and, specifically how they rate and assess the risks stemming from combining disparate pieces of information which lead to cumulative revelations.

## 2 DEMONSTRATOR

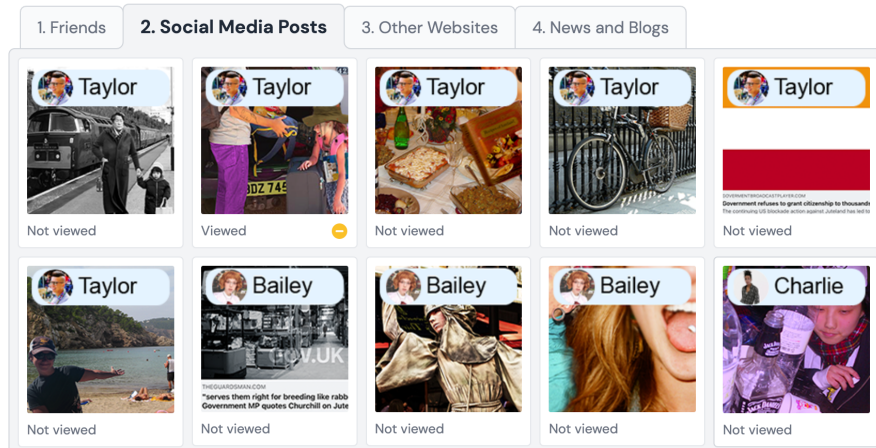
The design and development of the scenarios and tool build on outcomes from our previous work [18–20]. We performed a data narrative inquiry into people’s awareness of risks stemming from their everyday online information sharing practices in their personal and working lives. During online interviews, participants described their practices and also conceptualised them in ideographic form. We found that participants adopted incomplete risk models when assessing the dangers arising from individual pieces of shared information rather than accounting for their connection or summation together [18, 19]. When challenged to foresee possible future risks stemming from hostile actors they found it difficult to envision any such scenarios. Some participants described their online selves as boring – seemingly rationalising that their online information was of no interest or value to others. Taken together, this suggested that it would be difficult for people to scrutinize their own profiles for particular risks – which may or may not contain items of information relevant to particular scenarios. Prior to conducting the research, we thought that we could help participants to gain greater awareness of the risks by having them reflect upon their own posts, profiles, pages, etc. via the *Data Mirror* [13]. However, in light of their responses, coupled with the technical and privacy challenges already outlined, rather than having participants work through their own posts, we felt that it could be more powerful to develop a bespoke set of personas where different scenarios presented different potential risks for participants to find – so that we could avoid any ethical or privacy issues in dealing with personally identifiable data as well as ensuring that the different risks were present to explore.

### 2.1 Scenarios

To date, we have developed two personas with associated collections of posts, pages, blogs, news articles, etc. by the person of interest, their friends or other entities. The two fictitious personas created were, “Taylor Addison” and “Alex Smith”. While each sub collection currently consists of approx. 50 items, the possible combination of dependent judgements is approx. 2500 (as each post can potentially be combined with other posts to make greater inferences, e.g.  $50 \times 49$ ). So while the number of items in the collection is small, the number of possible combinations is large. For each persona, we then developed different scenarios (which can be regarded as topics). For Taylor Addison, for instance, we developed the following scenarios:

- **Hacked / Identity Theft:** Taylor has seen some usual activity on one of their accounts. They think that their account

<sup>1</sup><http://dataselfie.it/>



**Figure 1:** Search result page containing different verticals of (1) friends and followers, (2) social media posts (3) other websites, and (4) news and blog sites.

might have been hacked. Taylor asks you for help. You point out that certain information can be used by hackers to gain access to accounts, such as one’s date of birth, etc.

- **Unwanted Attention:** Taylor feels rather paranoid as if someone has been looking over their shoulder. They think that someone might be following them. You wonder if someone can trace Taylor’s movements from their online posts and the platforms that they use.
- **Lost Employment Opportunity:** Taylor tells you that they just received a call saying they didn’t get a job. Taylor thinks the recruiter was not open about the reason they were turned down given how perfect they were for the role. You think that perhaps Taylor’s online personality might have been a bit too much for them employer.
- **Political Victimisation:** Taylor feels persecuted for having strong beliefs about people’s rights to live in a free and democratic world. Taylor thinks that trolls have been targeting them because of Taylor’s friends. You think it might not just be what Taylor’s friends do or say, but also Taylor’s online activity.

In each of the different scenarios the participant is challenged to search through the profiles, posts and pages to identify items that might be relevant to the specific scenario. The scenarios created employ ambiguity and game play to provoke curiosity and encourage exploration across the individual pieces of a persona’s online information.

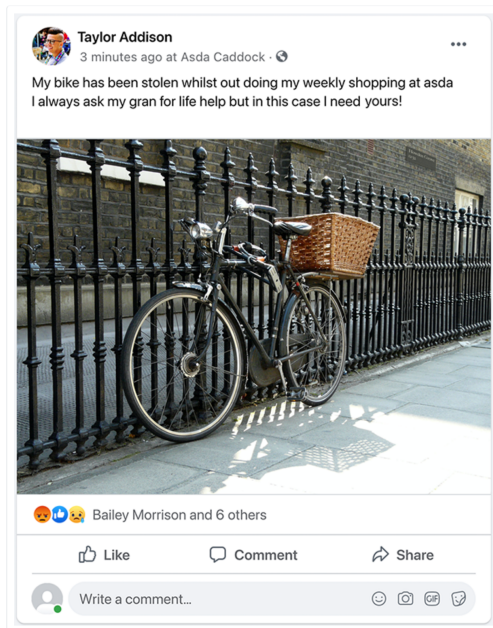
## 2.2 Interface

Below, we describe the main pages of the demonstrator. **Search Result Page** Participants are presented with a search result page containing several verticals (see Figure 1) – breaking the different posts, pages, profiles, and sites into different groups: friends and followers, social media posts from various fabricated platforms (e.g. “Friendbook”, “Tweeta”, “InLinked”, etc.), other sites (e.g. government websites with open data, open fitness app data, etc.) and news (e.g. articles from online newspapers, blogs, etc.).

**Result Annotation Page** When the participant clicks on a result, they are taken to a page to inspect and assess the item in question. Participants are then asked to rate the item – in terms of its relevance to the scenario – specifically, how concerning the item is w.r.t the scenario, then, whether the item is concerning for some other reason. For both questions, the participants can rate the item as either: (0) no, (1) possibly or (2) yes. Rather than using binary relevance, we included a middle ground where participants that were unsure could flag up whether they were concerned but not convinced. This is because the nature of relevance in these scenarios is conditional – a post may only become relevant – in light of other information that has been found. Participants could re-visit results, and thus subsequently revise their decision in light of new information. Participants were also asked to provide a free text description of the concern that they had regarding the item.

**Post Scenario Annotation Page** After the participant completes the scenario, for those posts that they marked as concerning, a subsequent rating page is presented. On this page participants are asked to grade the relevance of the items taken together – and thus provide their assessment of the risk of items individually and cumulatively. Participants are also asked to provide a free text description of their concern regarding the information that they found.

**Relevance and Risk Assessments** The tool allows us to capture the order of inspection of items, the amount of time spent reviewing items, how many times participants visit and revisit items, along with the participant’s relevance and risk assessments (including changes to their decisions). While our demonstrator focuses on scenarios specifically for risks associated with online cyber-safety, it could also be used to capture assessment for other scenarios where relevance is conditional and dependent. For our scenarios, we are particularly interested in understanding how participants explore and rate items during interactive search tasks – and, specifically how well they are able to identify relevant (risky) items. After the scenario is completed participants are provided with a debriefing page explaining the relationships between posts according to our gold set of judgements.



Do you think this post might reveal information that could be used to hack Taylor's account or steal Taylor's identity?

No  Possibly  Yes

Do you think that this post is concerning for other reasons?

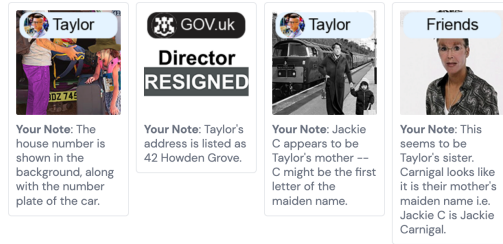
No  Possibly  Yes

If you think this post is of concern, please explain below.

**Figure 2: Page Annotation.** Participant's flag and note the concerns with the post – which they can revise during the course of the task.

### 3 SUMMARY AND FUTURE WORK

People's online profiles, posts and pages, whether constructed by themselves, friends or others leave a digital footprint that can be searched and explored – this motivates people-based search tasks – an area that is largely under investigated in Information Retrieval – except perhaps in the context of Expert Search [8] and Celebrity Profiling [9]. Dealing with and processing personally identifiable data, however, is fraught with challenges – while many of the potential risks may not even manifest in one's individual collection of profiles, pages, posts, etc. As such, how to best create and build collections to examine, explore and investigate how people search for people – and thus how well people can identify potential risks to their security, safety and well-being, online and offline, is largely unknown. Our demonstrator not only enables us to study such questions, but also provides a novel and engaging tool to educate people in online cyber-safety – by raising awareness of these potential risks, consequences and harms through game play and, to some extent, gamification of the search and annotation process. As previously mentioned, raising awareness and understanding of



Given these posts that you found, please tell us why you think they may risk giving away information about Taylor's identity to hackers and thieves.

How risky would you rate these posts, on average?

Not Risky      Extremely Risky

Taking these posts together, how risky would rate these posts, overall?

Not Risky      Extremely Risky

**Figure 3: Post Scenario Annotation.** Participants then rate the posts collectively on a graded scale in terms of risk.

these issues this is particularly important at both an individual and societal level.

We hope to receive feedback on how to develop and expand the collection to other scenarios and personas, in order to build further insights into how people search personal information, but also in terms of understanding the conditional and dependent nature of relevance when embedded in such contexts. Perhaps insights from this work could be used more generally for future test collection development to studying relevance more deeply. With respect to our primary goal of understanding and educating how well people can identify risks in people's digital footprints, we plan to run a number of subsequent studies evaluating people's search behaviours and their ability to connect information together – and, moreover whether engaging with and undertaking in such scenarios leads to greater awareness and longer term changes to people's information behaviours and practices that improve their cyber-safety.

The demonstrator focusing on Taylor's posts is available at:

- <https://bit.ly/sigir-demo-2022-submission-risky-relevance>

and the proposed study is available at:

- <https://bit.ly/study-on-risks-of-sharing-and-posting-online>.

### ACKNOWLEDGMENTS

*Cumulative Revelations of Personal Data* This project is supported by the UKRI's EPSRC under Grant Numbers: EP/R033889/1, EP/R033889/2, EP/R033897/1 EP/R033854/1, EP/R033870/1. The personas in this work are fictional. Any similarity to actual persons, living or dead, or actual events, is purely coincidental.

## REFERENCES

- [1] Alessandro Acquisti and Ralph Gross. 2009. Predicting social security numbers from public data. *Proceedings of the National academy of sciences* 106, 27 (2009), 10975–10980.
- [2] Angeliki Aktypi, Jason RC Nurse, and Michael Goldsmith. 2017. Unwinding Ariadne’s identity thread: Privacy risks with fitness trackers and online social networks. In *Proceedings of the 2017 on Multimedia Privacy and Security*. 1–11.
- [3] DM Boyd. 2008. *Taken out of context: American teen sociality in networked publics*. Ph.D. Dissertation. <https://search.proquest.com/openview/9cc930ef134daf46c17434d2992e8251/1?pq-origsite=gscholar&cbl=18750>
- [4] Danah Boyd. 2009. Why youth (heart) social network sites: The role of networked publics in teenage social life. *papers.ssrn.com* (2009). [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1518924](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1518924)
- [5] Xiongfei Cao, Ali Nawaz Khan, Ahsan Ali, and Naseer Abbas Khan. 2020. Consequences of cyberbullying and social overload while using SNSs: A study of users’ discontinuous usage behavior in SNSs. *Information Systems Frontiers* 22, 6 (2020), 1343–1356.
- [6] Hongliang Chen, Christopher E. Beaudoin, and Traci Hong. 2016. Protecting oneself online: The effects of negative privacy experiences on privacy protective behaviors. *Journalism and Mass Communication Quarterly* 93, 2 (2016), 409–429.
- [7] Lynne M. Coventry, Debora Jeske, John M. Blythe, James Turland, and Pam Briggs. 2016. Personality and Social Framing in Privacy Decision-Making: A Study on Cookie Acceptance. *Frontiers in Psychology* 7 (2016). <https://doi.org/10.3389/fpsyg.2016.01341>
- [8] Nick Craswell, Arjen P De Vries, and Ian Soboroff. 2005. Overview of the TREC 2005 Enterprise Track. In *Trec*, Vol. 5. 1–7.
- [9] Walter Daelemans, Mike Kestemont, Enrique Manjavacas, Martin Potthast, Francisco Rangel, Paolo Rosso, Günther Specht, Efstathios Stamatatos, Benno Stein, Michael Tschuggnall, et al. 2019. Overview of PAN 2019: bots and gender profiling, celebrity profiling, cross-domain authorship attribution and style change detection. In *International conference of the cross-language evaluation forum for european languages*. Springer, 402–416.
- [10] Judson C Dressler, Christopher Bronk, and Daniel S Wallach. 2015. Exploiting military OpSec through open-source vulnerabilities. In *MILCOM 2015-2015 IEEE Military Communications Conference*. IEEE, 450–458.
- [11] Oliver L. Haimson, Jed R. Brubaker, Lynn Dombrowski, and Gillian R. Hayes. 2016. Digital footprints and changing networks during online identity transitions. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 2895–2907.
- [12] Cormac Herley. 2009. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop (NSPW ’09)*. Association for Computing Machinery, New York, NY, USA, 133–144. <https://doi.org/10.1145/1719030.1719050>
- [13] Amal Htait, Leif Azzopardi, Emma Nicol, and Wendy Moncur. 2020. *DataMirror: Reflecting on One’s Data Self: A Tool for Social Media Users to Explore Their Digital Footprints*. Association for Computing Machinery, New York, NY, USA, 2125–2128. <https://doi.org/10.1145/3397271.3401398>
- [14] Danesh Irani, Steve Webb, Kang Li, and Calton Pu. 2011. Modeling unintended personal-information leakage from multiple online social networks. *IEEE Internet Computing* 15, 3 (2011), 13–19.
- [15] Puneet Kaur, Amandeep Dhir, Anushree Tandon, Ebtesam A. Alzeiby, and Abeer Ahmed Abohassan. 2021. A systematic literature review on cyberstalking. An analysis of past achievements and future promises. *Technological Forecasting and Social Change* 163 (2021), 120426. <https://doi.org/10.1016/j.techfore.2020.120426>
- [16] Johannes Kiesel, Arjen P de Vries, Matthias Hagen, Benno Stein, and Martin Potthast. 2018. WASP: web archiving and search personalized. (2018).
- [17] Hanna Krasnova, Sarah Spiekermann, Ksenia Koroleva, and Thomas Hildebrand. 2010. Online social networks: Why we disclose. *Journal of Information Technology* 25, 2 (jun 2010), 109–125.
- [18] Callum Nash, Daniel Carey, Emma Nicol, Amal Htait, Burkhard Schafer, Jo Briggs, Wendy Moncur, and Leif Azzopardi. 2022. Making Sense of Trifles: Data Narratives and Cumulative Data Disclosure. In *Proceedings of RISE22 INTERNATIONALES RECHTSINFORMATIK SYMPOSIUM*. 8p.
- [19] Emma Nicol, Jo Briggs, Wendy Moncur, Amal Htait, Daniel Carey, Leif Azzopardi, and Burkhard Schafer. 2022. Revealing Cumulative Risks in Online Personal Information: A Data Narrative Study. *PACM HCI* (2022).
- [20] Emma Nicol, Amal Htait, Leif Azzopardi, and Wendy Moncur. 2021. Towards identifying, understanding and controlling cumulative revelations in social media. *Proceedings of the Association for Information Science and Technology* 58, 1 (13 Oct. 2021), 798–800. <https://doi.org/10.1002/pr2.566> 84th Annual Meeting of the Association for Information Science and Technology (ASISampT) ; Conference date: 29-10-2021 Through 03-11-2021.
- [21] OfCom. 2019. *Ofcom Adults’ Media use and attitudes report*. Technical Report. [https://www.ofcom.org.uk/\\_data/assets/pdf\\_file/0021/149124/adults-media-use-and-attitudes-report.pdf](https://www.ofcom.org.uk/_data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf)
- [22] Ning Xia, Han Hee Song, Yong Liao, Marios Iliofotou, Antonio Nucci, Zhi-Li Zhang, and Aleksandar Kuzmanovic. 2013. *Mosaic: Quantifying Privacy Leakage in Mobile Networks*. 564 pages.
- [23] ChengXiang Zhai, William W Cohen, and John Lafferty. 2015. Beyond independence: methods and evaluation metrics for subtopic retrieval. In *Acmsigir forum*, Vol. 49. ACM New York, NY, USA, 2–9.