Towards reproducibility in online social network research

Luke Hutton, Tristan Henderson

Abstract—The challenge of conducting reproducible computational research is acknowledged across myriad disciplines from biology to computer science. In the latter, research leveraging online social networks (OSNs) must deal with a set of complex issues, such as ensuring data can be collected in an appropriate and reproducible manner. Making research reproducible is difficult, and researchers may need suitable incentives, and tools and systems, to do so.
In this paper we explore the state-of-the-art in OSN research reproducibility, and present an architecture to aid reproducibility. We characterise reproducible OSN research using three main themes: reporting of methods, availability of code, and sharing of research data. We survey 505 papers and assess the extent to which they achieve these reproducibility objectives. While systems-oriented papers are more likely to explain data-handling aspects of their methodology, social science papers are better at describing their participant-handling procedures. We then examine incentives to make research reproducible, by conducting a citation analysis of these papers. We find that sharing data is associated with increased citation count, while sharing method and code does not appear to be. Finally, we introduce our architecture which supports the conduct of reproducible OSN research, which we evaluate by replicating an existing research study.

Index Terms—Data sharing, online social networks, reproducibility, survey

✦

1 INTRODUCTION

The consideration of the scientific method in computational science has grown of late, as researchers increasingly recognise computation as a separate methodological branch of science in its own right [9]. Fundamental to the scientific method is the notion of *reproducibility*; that a researcher should be able to take an experiment or study and perform it again or build on it to create new works.

The challenges associated with the reproducibility of research have been widely studied, with solutions proposed across myriad domains. One area that has not been examined is Online Social Network (OSN) research. With research in OSNs straddling graph theory, human-computer interaction (HCI), networking, and social sciences, there may well be new challenges in reproducibility.

Enabling reproducibility requires support for conducting research through the entire workflow, from initial data collection through to processing, analysis, and publication of research artefacts such as papers, data, and source code. Thompson and Burnett [37] suggest that reproducibility consists of three elements:

- Supporting computationally intensive research, by sharing source code, tools, and workflows for executing these tools.
- Supporting structured analysis, by encoding the scripts which conduct analyses and produce components of publications such as tables and figures.
- Allowing the dissemination of research artefacts, such as papers, and raw data. Rather than treating papers as a static piece of text, they should include, or provide access to executable code, and other resources needed for replication.

We can think of these elements as broadly encapsulating three key themes: code, methods, and data, respectively.

There are particular challenges to conducting OSN research in a reproducible manner, some of which arise from the tension between social science and systems-oriented approaches which manifest in much OSN work. Trevisan and Reilly consider whether social media platforms ought to be considered public spaces [38], which has implications for how data are collected, processed, and shared with other researchers. Most major OSNs, such as Facebook and Twitter, provide fettered access to their data through application programming interfaces (APIs), the use of which is subject to a license agreement. These providers assert control over how the data that they host are used, and actively disallow large datasets of their content to be published.[1][2] This may impede one of the tenets of reproducible research, particularly when work concerns a specific corpus of content, such as Denef et al.'s examination of tweets during the 2011 London riots [8], rather than a random sample of content generated by a certain population. If OSN data cannot be directly shared, then it might be possible to instead repeat the experiment, but only if the sampling strategy of the original experiment can be replicated. This is challenging, however, when papers do not sufficiently disclose how their participants were recruited, and data were collected. In user studies where users interact with OSNs, the range of variables make it difficult to replicate the participant's experience, from the text used in prebriefing and consent documentation, through

- *School of Computer Science, University of St Andrews, St Andrews KY16 9SX, UK*
  E-mail: [lh49,tnhh]@st-andrews.ac.uk

1. Twitter Developers Terms of Service: https://dev.twitter.com/terms/api-terms
2. Facebook Platform Policy: https://developers.facebook.com/policy

to the implementation of user interface elements.

Where research is dependent on the use of APIs provided by OSNs, there are additional challenges to reproducibility. Code that evokes certain API endpoints is dependent on that API being online and its design remaining consistent, which may be an impractical expectation for actively developed services where new features are developed and retired over time. In 2013 alone, Facebook announced seven sets of "breaking" changes, where developers needed to amend their code if it used certain features, incorporating the change or withdrawal of 47 API endpoints.[3] As some of these changes concern the removal of features, some legacy code will not be usable even if actively maintained. This is a significant challenge to reproducing results dependent on live OSN data. More recently, Facebook has introduced an API versioning scheme which will go some way to improving this situation, but retired API versions will only receive support for one to two years[4], and such approaches are not common to all OSNs.

These challenges are encapsulated in the three themes identified above. If OSN researchers publish the *code* used to conduct their studies, explicitly outline their *methodology*, and share *data* where possible, then our hope is that the state of reproducibility in the field will improve.

In this paper, we make the following contributions:

- We conduct the first comprehensive study of 505 papers which collect or use OSN data, to assess the extent of reproducibility.
- We examine the common practices and challenges we see in recent OSN research, from which we propose a set of recommendations for the benefit of OSN researchers in all disciplines.
- We introduce a framework for conducting reproducible OSN studies, and demonstrate its effectiveness by reproducing one of the experiments from our survey.

2 RELATED WORK

Despite being a fundamental aspect of the scientific method, reproducibility in computational sciences has only recently been identified as a significant issue. Stodden has substantially contributed to the discourse, surveying researchers to understand attitudes towards reproducibility [35], and developing a set of recommendations concerning the storage, versioning, and publication of data [36]; however these do not address the domain-specific challenges of conducting OSN research that we explore in this paper.

Frameworks for improving the state of reproducibility are nascent, and often attempt to address one dimension – by supporting recomputation or archiving of code, encoding the methodology of an experiment as a workflow, and for supporting the sensitive management of data for reuse. Holistic solutions are not common, perhaps due to the complexity of addressing reproducibility, and the

domain-specific issues which emerge in different fields. For example, biologists have developed standards for the transmission of proteomics data, such as the protein sequence database UniProt [20], but such efforts are extremely domain-specific. Hurlin et al. propose RunMyCode.org as a tool to allow researchers to bundle artefacts such as data and code for others to access [17], while VisTrails is a system for encoding workflows, to allow the reconstruction of visualisations and plots from the original data [10]. We are not aware, however, of any solutions which aim to deal with the specific challenges of OSN research.

Researchers conducting OSN studies have found the APIs provided by OSNs to be a barrier to conducting their studies. Nichols and Kang's analysis of responses to tweeted questions was thwarted by their accounts being suspended by Twitter due to perceived spamming behaviour [27]. Morstatter et al. find that Twitter's unfettered "firehose" API, available only to high-profile commercial partners, provides a significantly different sample of tweets than the widely available "streaming" API [24]. These challenges restrict researchers' ability to replicate studies if they are not able to collect a similar distribution of content, depending on their license agreement with the OSN provider. For example, De Choudhury et al. leveraged their corporate Twitter access to collect depression-indicative content which others might not be able to recreate [7].

In many cases, access to the original data is not necessary. Unlike some more theoretical fields where reproducibility may concern the replication of results by seeding a simulation with data, or evaluating a statistical model, many OSN papers consist of user studies which use OSNs as a conduit for examining behaviour of a population. In such instances, replication of methods is key. For example, even subtle changes in the presentation of consent forms can have an impact on how people interact with an experiment [23], and the act itself of asking for consent may bias results [31]. Failure to encode such methodological details can make it difficult to accurately replicate studies and meaningfully compare results.

The difficulty of adequately anonymising sensitive OSN data is another challenge. Anonymisation has a temporal quality - what might be sufficiently obfuscated today may be deanonymised tomorrow. Narayanan and Shmatikov demonstrate how many apparently anonymised datasets simply replace names with random identifiers, rather than obfuscating uniquely identifying attributes, permitting reidentification [26]. Dawson surveyed 112 articles to show participants quoted from public web sources could trivially be reidentified [6]. Concerns about sufficiently protecting the privacy of participants after their data have been released, while maintaining their utility in further studies, is a constant tension for OSN research.

Other surveys of the OSN literature have been conducted. boyd and Ellison's survey of OSNs provides a de facto definition of such applications, and identifies early work exploring the behavioural and graph theoretic perspectives on social network structures [3]. Mullarkey developed a typology of OSN papers based on a smaller sample of pa-

---

3. Facebook Platform Roadmap: https://developers.facebook.com/roadmap/completed-changes

4. Facebook Platform Upgrade Guide: https://developers.facebook.com/docs/apps/upgrading

pers, to illustrate biases in the nature of OSN research [25]. Wilson et al. [40] look at 412 papers that use a single OSN, Facebook, while Caers et al. [5] find 3,068 papers in a broader search, but neither focuses on the reproducibility of work as we do. Golders and Macy conduct a wide-ranging survey of OSN research in sociology [14], and outline privacy as a research challenge, but not ethics, and discuss methodology but in the context of training sociologists in methods for collecting OSN data. Alim surveys OSN researchers about ethics concerns, and finds that 25% of respondents sought ethics approval for their studies [2]. This is higher than the proportion that we find reported on ethics approval, although not all authors might report this, and indeed there might be an element of selection bias, since researchers more interested in ethics might have responded to this particular survey.

3 WHAT IS REPRODUCIBLE OSN RESEARCH?

To examine the state of reproducibility in the field, we examine 901 papers from 26 venues, published between 2011 and 2013. A range of venues were included to gain a diverse range of perspectives, including top-tier HCI conferences, network science workshops, and social science journals. We first collected all papers which satisfied the search terms shown in Table 1. For each paper, we then assessed whether the paper involved the handling of OSN data. If a paper's methodology concerned the collection or publication of data intended for an OSN, whether already established (such as Facebook or Twitter), or developed as a testbed for academic study, it was included. This was the case whether the authors directly processed the data themselves, or a previously crawled dataset was utilised.

Of the 901 papers examined, 505 met this criteria and were then tested against the ten criteria we devised for assessing reproducibility.

Field	Keywords
Abstract contains any of	Facebook
	Twitter
	Foursquare
	LinkedIn
	Friendster
	Weibo
	Flickr
	LiveJournal
	MySpace
	"Online social network"
	"Social network site"
	"Social networking site"
	SNS
	OSN
Publication date between	01-01-2011
	31-12-2013

TABLE 1
The semantics of the search term used to identify papers in the study (the exact syntax for expressing the search varied from source to source).

Venue type	Venue	Total	Relevant
Computer science	IEEE T Mobile Comput	2	1
	Comput Netw	7	4
	Commun ACM	46	2
	Comput Commun	9	5
	IEEE Pervas Comput	2	1
Security & Privacy	NDSS	2	1
	SOUPS	9	2
	S&P	3	1
	CCS	13	1
	WPES	4	2
Information systems	COSN	14	12
	EuroSys SNS	13	7
	WOSN	9	7
	WebSci	40	33
	ICWSM	200	177
	ASONAM	155	120
	HotSocial	9	6
Human-centered computing	CHI	82	39
	CSCW	73	45
	Pervasive	9	0
	UbiComp	23	9
Multidisciplinary	Nature	10	3
	P Natl A Sci USA	7	4
	Science	7	2
Communication	J Comput-Mediat Comm	18	3
Psychology	Comput Hum Behav	129	13
Anthropology	Soc Networks	6	5
Total		**901**	**505**

TABLE 2
Breakdown of the surveyed papers by venue. The "total" column indicates how many papers matched our search term, while the "relevant" column indicates how many used OSN data, meriting further study.

azines were grouped by field, using the publication's top category as listed by Thomson Reuters[5], while conferences were grouped by the best-fitting top-level category in the ACM Computing Classification System.[6] A summary of the venues, their classifications, and the number of papers examined is shown in Table 2. Finally, for each paper included in the survey, we conducted a citation analysis by querying Google Scholar to receive a citation count for each paper on July 8th 2014. While Google Scholar may not provide an exhaustive count of all citations, it allows us to study the relative performance of the papers we examine, in a similar fashion to other studies [29].

3.1 Explanation of criteria

Each of the 505 papers was tested against the following set of criteria. These align with the three aspects of re-producibility outlined earlier. For each criterion, a paper is assigned a binary flag to indicate satisfaction. Note that this was determined by our reading of the papers, and not the result of an automated content analysis process.

To better understand trends across the literature, we categorised venues in one of two ways. Journals and mag-

3.1.1 Methods

- **Source OSN**: User behaviour is not identical across social network sites, so replications are dependent on knowing where data were collected, either to collect data from a similar population, or to show differences between OSNs. Thus we note whether the paper explicitly identify the OSN(s) from which data were collected or published to. If the authors note that data were collected from an OSN aggregation service such as FriendFeed[7] without clarifying which underlying OSNs were accessed, this criterion is not met.

- **Sampling strategy**: Just as the choice of underlying OSN may indicate biases in the resulting data, the way participants in the research were chosen is an important consideration. When conducting user studies, it is important to know whether the authors were investigating a certain population, or whether they intend their findings to be generally applicable to a wider population, as this has implications for how participants are recruited for replications. Similarly, large-scale crawling exercises may be biased if, for example, user IDs are collected in increments from an arbitrary starting point. To satisfy this criterion, the paper must explain how participants were recruited, either explaining the sampling technique, or offering a breakdown of the participants' demographics. If the study used an existing dataset, the authors must explain how the underlying data were collected.

- **Length of study**: OSNs exhibit a number of temporal effects. As the functionality of services evolve, the way they are used changes [16], and people's online behaviours change as they age [34]. Accordingly, in order to replicate OSN studies, it is important to know the length of time over which data were collected, as this can affect user behaviour [19], and ideally at what time data were collected. To satisfy this criterion, the period of data collection must be identified.

- **Number of participants**: As the number of participants will affect the number of results, and the effect size of analyses, it is important to disclose how many were collected. To satisfy this criterion, the number of participants, or users whose data were crawled, must be identified. In user studies, if participants were in one of many experimental conditions, the distribution of participants among these conditions must be disclosed.

- **Data processing**: Understanding how data are handled throughout an experiment is an important detail, from both reproducibility and ethical perspectives. Knowing precisely which attributes of sensitive OSN data were collected is important to both replicate the study, and ensure data collection is proportionate to requirements, especially as OSN APIs make it trivial to collect significant amounts of information. In addition, knowledge of how data were sanitised is important, particularly when releasing data which relates to sensitive OSN content. For example, have identifying characteristics been anonymised or aggregated, and how? To satisfy this criterion, the paper must have answered at least one of the following questions: Is the data handling strategy identified? Are the attributes of collected data enumerated? Were the data sanitised? How were they stored? Who had access to the data?

- **Consent**: The issue of obtaining informed consent when conducting online research is contentious [33], [39]. Depending on its nature, OSN research may constitute human subjects research, in which case data-handling practices should be subject to the participants' informed consent. Understanding whether consent was sought is important for replications, as the process may have implications on the results [23]. To satisfy this criterion, the authors must note whether the human subjects of the data collection provided consent to participate. The authors do not need to have sought consent to satisfy this criterion, but the issue must have been considered in the text.

- **Participant briefing**: As with the acquisition of consent, the briefing and debriefing experience is an important ethical consideration when conducting human subjects research. These procedures ought to be explained in the text such that other studies can replicate the procedures for the most consistent participant experience. To satisfy this criterion, the paper must disclose whether participants were briefed and debriefed to bookend their participation in the study.

- **IRB/Ethics**: Alongside disclosure of consent and briefing procedures, studies should disclose whether the procedures of an experiment were approved by an Institutional Review Board (IRB), ethics committee, or equivalent. The need for such approval is dependent on what certain institutions or jurisdictions deem to be human subjects research, but disclosure can support replications, as IRB oversight may affect the ultimate data collection protocol of an experiment. To satisfy this criterion, the authors must note whether such bodies have approved the practices of the study.

3.1.2 Data

- **Data shared**: The studies we examine may concern first-hand collection of data, perhaps by crawling an OSN, or conducting a user study to examine behaviour in an OSN. Alternatively, studies may use existing datasets, either provided through arrangement through a third-party, or by using a public dataset. Data sharing is acknowledged as an important aspect of reproducibility, but for all OSN research it is not essential, particularly where the data collection practices are sufficiently explained to allow other researchers to collect their own data. Nonetheless, we consider for each paper whether the data are shared with the research community, or if the authors explicitly entertain requests for access to the data. Where an existing dataset is used, the authors must explicitly cite it.

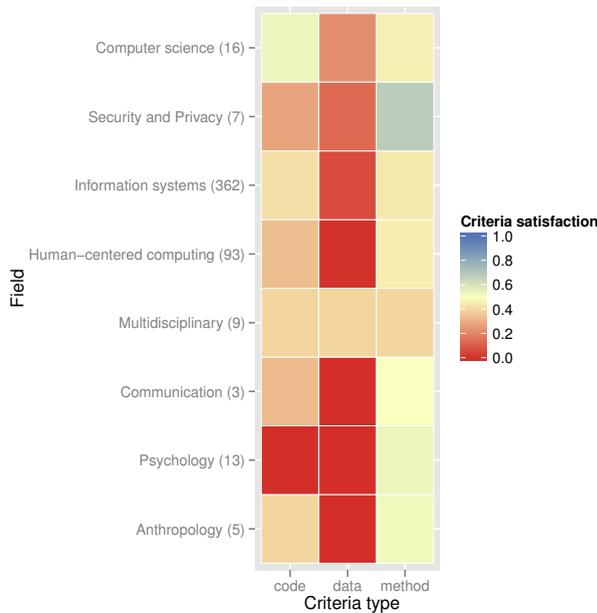7. FriendFeed: http://friendfeed.com

Fig. 1. Heatmap showing how different fields achieve our three criteria types. Data-sharing is particularly poor across most disciplines, while reporting of methodologies is generally stronger.
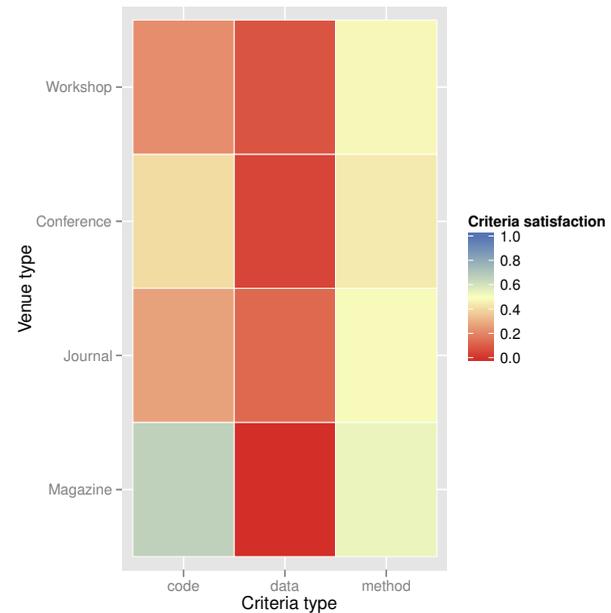

Fig. 2. Heatmap showing how well each type of venue achieve our three criteria types. Data-sharing and methodology reporting are similar, however conferences and magazines are better at sharing code.

3.1.3 Code

- **Protocol**: Another pillar of reproducibility concerns access to software artefacts necessary for collecting data, conducting analysis, and generating outputs such as plots. If a study concerns a bespoke visualisation, or the development of a new OSN or alternative OSN interface, these should be accessible openly, and ideally the source should be available for others to use. To satisfy this criterion we check whether authors who develop their own software make this available to other researchers, and whether statistical analyses are explained in such a way that they can be replicated.

4 STATE OF THE ART

Our survey highlights differences in how well papers in different venues achieve reproducibility. Fig. 1 shows a high-level summary of how different fields satisfy the three criteria types we introduced in Section 3.

4.1 Few OSN researchers share their data

The most striking finding is that few papers share their data at all, with only 6.1% of papers in our survey doing so. Unsurprisingly, this is closely associated with the data- sharing policies of different venues. Multidisciplinary journals such as Nature and Science mandate authors to include data such that reviewers and other researchers can replicate results[8] [9],

and accordingly are a notable exception to this trend, with 40% of papers sharing their data. We are not aware of any conferences in our survey which mandate data necessary for replication must be shared, although conferences such as SOUPS do allow authors to include appendices which support replication.[10] Similarly, ICWSM operates a data sharing initiative to encourage the sharing of datasets[11], which may explain why 35.4% of the papers which shared data came from this venue. We note that papers at some information systems venues, such as EuroSys SNS and COSN, are moderately better at their data sharing practices, with authors at both sharing data twice as often as the venue average. This appears to be a side-effect of many papers using crawled social graphs, rather than datasets of content, such as tweets, which are licensed under terms which prohibit redistribution. As shown in Fig. 2, papers in venues of all types are quite poor at routinely sharing their data. Journals fare better with 13.9% of papers sharing their data; however a chi-square test of independence does not suggest this is a significantly greater effect than other venue types ($\chi^2 = 4.38$, df = 2, $p = 0.11$).

Recommendation 1: Researchers should endeavour to share their datasets where possible with the community. Providers of OSNs should develop ways to allow researchers to share data collected from their services, and to mitigate the inequalities between institutions with different degrees of access to OSN data, such as those without Twitter Firehose access. This is echoed in the final "rule" for

---

8. Nature data policy: http://www.nature.com/authors/policies/availability.html

9. Science data policy: http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml

10. SOUPS 2014 CFP: http://cups.cs.cmu.edu/soups/2014/cfp.html

11. ICWSM Data Sharing Initiative: http://icwsm.org/2015/datasets/datasets

reproducible computational research proposed by Sandve et al. [32], which argues that "all input data, scripts, versions, parameters, and intermediate results should be made publicly and easily accessible", and by the Yale Law School Roundtable on reproducible research [41]: "When publishing computational results, including statistical analyses and simulation, provide links to the source-code (or script) version and the data used to generate the results to the extent that hosting space permits"; however neither set of recommendations acknowledges the challenge of redistributing sublicensed datasets.

4.2 Social scientists rarely share code for experiments and analyses

We find that code-sharing practices are generally better, which includes the distribution of theorems or algorithms which support replication, but notably no venue types, except for multidisciplinary journals, include a majority of papers who satisfy this.

In this analysis, *Computers in Human Behavior* was notable in that none of the papers we examined shared code. CHB's simultaneous computational and social science angle attracts authors from diverse disciplines and may go some way to explaining this. Of the 13 papers that we examined, first authors are affiliated with computer science, communications, political science, management, humanities, psychology, and law faculties. For many such fields, there may be no expectation that quantitative methods are shared to allow replication. As multidisciplinary efforts like this gain traction, it is important that the strengths of social sciences – such as experience with qualitative methods – feed into computer science, just as traditional CS strengths – such as an emphasis on sharing code – are accepted by the wider computational social sciences community.

Of note, code-sharing rates increase dramatically between publication types. As shown in Fig. 2, protocols are shared in approximately a quarter of workshop and journal papers, while 41.4% of conference papers satisfy this. in Paek and Hsu's work to create phrase sets for text entry experiments from large corpora, the researchers made the phrase sets and code available, and included detailed algorithmic details within the paper [28]. As noted earlier, we attribute this trend towards sharing to more stringent requirements for supplementary materials in such publications. As workshops are often used for work in progress, it may be that researchers are reticent to share unfinished code. We would hope to see this change, however, to help engage the community in the development and re- use of software even in an unfinished state.

Recommendation 2: CS and social sciences need to merge their strengths to bring domain knowledge from both perspectives. Leveraging experience of human subjects research from social scientists can improve the reporting and ethical conduct of such studies in a computer science context, while a background in computational research can encourage others to share source code and details of analyses to support reproducibility. This exchange of knowledge can both improve the state of both fields individually and in collaborative efforts.

4.3 Reporting of core experimental parameters is strong

The focus of this paper is on the state of reproducible methodologies in OSN papers. Reporting of the methodological attributes appears strong across all papers; however, the breakdown of these criteria in Fig. 3 shows a more complex dichotomous story. The first four criteria illustrate the extent to which studies report the core aspects of their data collection practices, critical to reproduce any such studies, including the source OSNs, how participants and their data were sampled, for how long data were collected, and the number of participants. Generally, papers are very good at reporting this information, with some notable exceptions. Just as studies which used existing datasets are inherently better at sharing the data they use, they tend to be worse at reporting the provenance of their datasets, such as the composition of the dataset's participant pool. These are crucial details which are required to replicate such studies, particularly if the original dataset is not to be used – such as aiming to replicate the findings of a user study with a different population.

Recommendation 3: Even when studies use existing datasets, researchers must explain core methodological details to support replication. Such sharing can minimise the duplication of effort which currently prevails when researchers attempt to build on the findings of those before them, as well as supporting direct replications.

4.4 Participant-handling and ethical considerations are not discussed

The final four criteria in our methodology breakdown concern data processing and participant ethics, two critical aspects of reproducibility, where consistently most papers do not report core methodological concerns: did participants give consent? Were procedures approved by an IRB? How were the collected data handled? Again we see a divide in approaches between systems papers, and social sciences. Quantitative work, for example, is better at reporting how their data were handled, such as anonymisation practices, and which attributes of datasets were stored. As Fig. 3 shows, the seven *Security & Privacy* papers we consider are better at reporting these concerns. We attribute this to a culture of reporting these details at SOUPS, while WPES allows appendices with supplementary information to be provided. Conversely, the social science background of many CHB papers is highlighted in the marginal improvement in reporting of ethical concerns, shown in the *Psychology* group. We were surprised to find that HCI papers were not particularly strong in this regard. Indeed, such reporting is so uncommon that attention should be drawn to positive cases, such as Johnson et al.'s description of their recruitment material and consent procedures [21], and Ali et al.'s reporting of their study's participant briefing process [1]. Simply reporting the existence of briefing
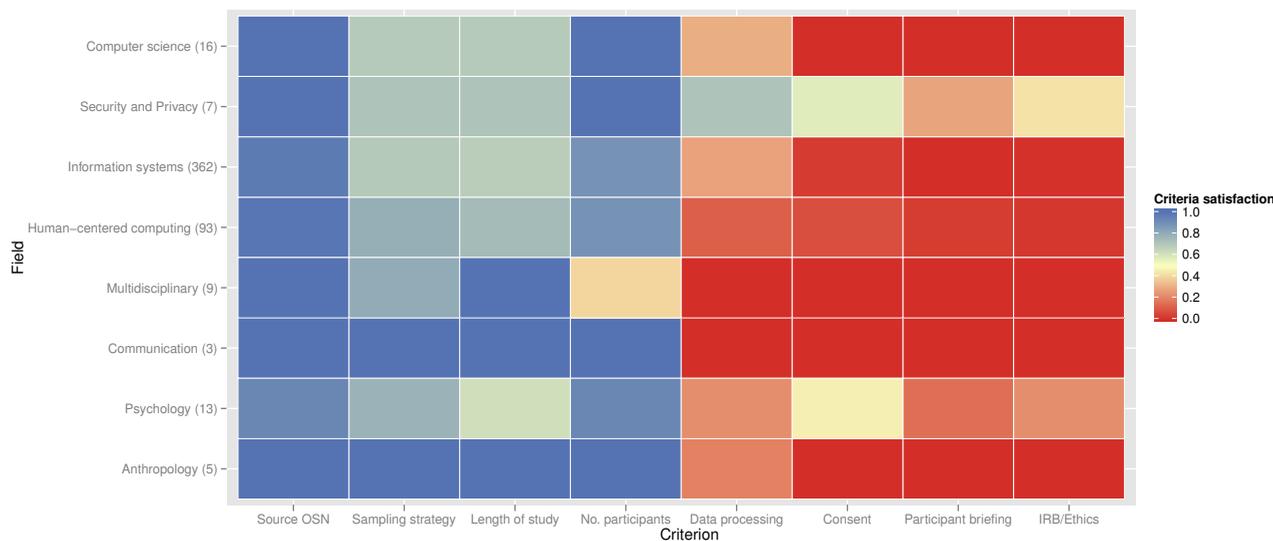
Fig. 3. Breakdown of the eight criteria we assess for "methods". Generally, papers successfully report descriptive attributes of their study, but often, participant handling and data processing are not sufficiently explained.

and consent procedures generally does little to support replication. Our concern with the lack of robust description of such methods, is that as previous work shows the briefing experience can affect people's disclosure behaviours in OSN experiments [23], it is important that researchers can replicate these procedures when conducting user studies using OSN data.

Recommendation 4: Briefing procedures, IRB protocols, and other auxiliary materials should be made available. Beyond recreating the experiment itself, this will ensure ethical standards can be preserved, and that the requirements of a study can be communicated to other ethics boards when replicating studies.

In this section, we have looked at how well the state of the art addresses a number of facets of reproducibility in OSN research. We find that venues from more technical backgrounds differ in their reporting from the social sciences, and identify four best practices which combine the strengths of both to improve the state of the art. The results of the survey can be viewed in full at a CiteULike group[12], which provides links to all publications we considered, and the ability to search for papers based on which criteria were satisfied. Next, we discuss the challenges in encouraging researchers to make the effort to share their protocols to support replications, and how a culture can be developed to incentivise such efforts with direct benefits to the researcher.

5 ENCOURAGING REPRODUCIBILITY IN OSN RE-SEARCH

In recent years, bibliometrics have increasingly been used to measure the impact of research, with implications for

funding and career advancement [4]. It follows that researchers will then consider such metrics when making decisions about how to conduct their work. Recent interest in data-sharing has led to an increasing number of venues encouraging, and in many cases mandating, the sharing of raw data with published papers. Our own results highlght this change in culture, with Fig. 2 showing that data-sharing has increased among prestigious journals such as *Nature* and *Science*.

An important factor in motivating such data sharing has been the incentives it can provide to researchers, specifically encouraging other researchers to cite their data, increasing the visibility of their work. Piwowar et al. show that cancer clinical trial papers which make their data publicly available account for a 69% increase in citations [29]. Our results support this finding to an extent, with papers that share their data receiving 21.6% more citations. We believe this increase is smaller due to a less embedded culture of data-sharing in social network research compared to many biological fields. In addition, as our survey only examines papers published between 2011 and 2013, many papers have not yet had an opportunity to be cited.

We investigate whether nascent efforts to encourage broader disclosure of methodological details has so far led to a similar increase in citations. Fig. 4 shows a similar distribution, where most papers are rarely cited, followed by a long tail of highly-cited papers. Sharing data is associated with the greatest change in distribution, with more highly-cited papers among those who share their data. This trend is not repeated for the other metrics we examine to a significant extent. We attempt to fit a linear model to the distribution to determine whether an increase in citations can be attributed to sharing these details. Due to its long-tailed distribution, we apply Yeo-Johnson transformations to coerce a normal distribution for analysis. A regression suggests only data sharing is significantly associated with
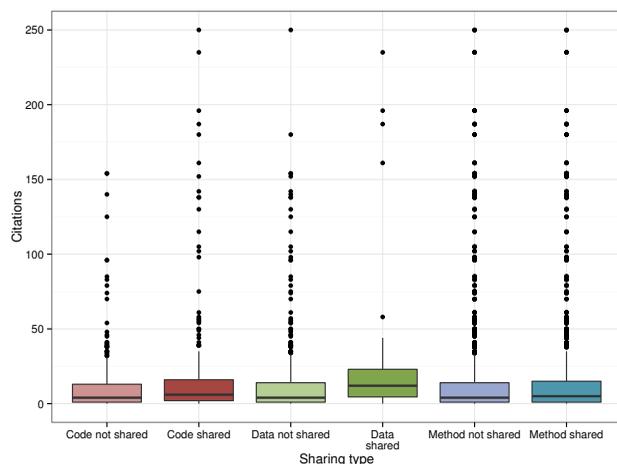
12. http://www.citeulike.org/group/19063

Fig. 4. Boxplot showing citation rates for papers which do and do not share their data, code, and methods. Papers which share their data are more likely to be higher cited. Sharing other details also leads to improved citations but to a lesser extent.

an increase in citations, however with such a poor fit ($R^2 = 0.003$), this model does not adequately explain the effect on citation rates. No such effect can be found for papers which share code and methods.

To understand why this is the case, we can look to the culture shift that occurred with the data-sharing movement. As discussed earlier, various initiatives aim to convince researchers of the merits of sharing their data with the hope that taking the time to prepare data for wider sharing would both benefit the field, while delivering a benefit to the researcher of increased citations as others use these data to generate more publications, creating a compelling symbiotic relationship, with incentives for all parties. No such movement has yet motivated increased disclore of experimental protocols to support replications, because the case has not yet been made that doing so is in a researcher's personal interest, beyond improving the state of the field. Papers which simply reproduce a previous experiment are unlikely to be published, so researchers may wonder what the merit is of disclosing such detailed protocols.

This motivates efforts to incentivise researchers to make better efforts to share the protcols necessary to replicate experiments. Other fields, such as biology, have an embedded culture of sharing protocols and workflows such that other researchers can re-use and adapt these protocols in their own experiments. One such initiative is Goble et al.'s myExperiment system, for encoding and sharing scientific workflows on a service which adopts concepts from existing OSNs [13]. Since its launch in 2007, there have been attempts to quantify the impact such services have had on the scientific method in fields where it has been adopted, particularly bioinformatics. Procter at al. conducted interviews with researchers using myExperiment in 2009, finding that building a social network based on shared workflows as well as other traditional scholarly outputs was

an attractive incentive for adopting such systems, as well as the hope of building social capital, with workflow-sharing perceived as a reputation-building exercise. Researchers found limitations with reproducing workflows "off-the-shelf", however, due to poor annotation and documentation of many workflows [30].

Where a culture of workflow-sharing is nascent, we can expect the level of curation to be low, although one can expect this to improve in time as de facto standards for annotation emerge. Fridsma et al. introduce the BRIDG Project, which defines consistent semantics in clinical trial protocols and datasets to aid the sharing and re- use of such artefacts [11]. Just as such tools are gaining traction in bioinformatics and other fields, we believe such approaches can translate well to the sharing of protocols for handling OSN data. We propose a similar culture shift has to occur to convince researchers of the benefit of sharing methodological details of OSN experiments. One such way to encourage this is through the provision of tools and proposed standards which encourage interoperable experimental workflows.

In the next section, we discuss our work towards developing such tools to support reproducible OSN research. We replicate one experiment from our survey that achieves these best practices, to show how such tools can be applied.

6 REPRODUCING OSN EXPERIMENTAL WORK-FLOWS

Our analysis of the state of the art in reproducibility in OSN research shows mixed progress towards the three pillars of data sharing, code re-use, and methodology capture. While the first two are generally not well-achieved in our survey, there are increasing efforts in this space, such as the data-sharing repository FigShare[13], and Gent's work towards a recomputation framework which allows legacy experimental code to be executed in the future [12], but the applicability of such methods to human subjects research is unclear. There has been relatively little attention paid to capturing the methodology of experiments, particularly those concerning OSNs. The low rate of sharing such methodological details may be attributable to its difficulty. With most venues making little effort to encourage researchers to disclose such details, and no standards for communicating protocols or workflows in OSN research, it is perhaps not surprising that authors are unwilling to take the time to disclose such details without being confident that the community is willing and able to make use of them. To aid this, we have developed tools to simplify the sharing of OSN experimental workflows.

6.1 PRISONER: An architecture for reproducing OSN research workflows

PRISONER (Privacy-Respecting Infrastructure for Social Online Network Experimental Research) is a framework which aims to support the execution of reproducible and

13. FigShare: http://figshare.com/

privacy-respecting experiments which use OSN data. PRIS-ONER abstracts experimental artefacts, such as questionnaires or OSN user studies, from the OSN on which they depend, and from the data-handling practices of an experiment. This allows the same experiment to be conducted on different OSNs with minimal effort, and encapsulates various methodological concerns, such as data collection, processing, and participant consent, as a worklow which can be shared and replicated. The framework is described in full in [18].

With PRISONER, researchers define how their OSN experiment collects, processes, and stores data by writing a *privacy policy*, enumerating the OSNs from which data are used, the types of data to be collected, and how those data should be sanitised throughout the life of the experiment. Instead of directly accessing the APIs of OSNs, experimental applications make requests to PRISONER, which provides a consistent interface to the implementation of different OSNs, while validating requests and sanitising responses to respect the researcher's policy. This approach has some key benefits: experiments can be targeted at different OSNs with minimal adjustment, and policies can easily be shared with other researchers to ensure the same data collection practices are used in their replications.

PRISONER supports the reproducibility of OSN experiments by addressing many of the challenges discussed in our survey. To illustrate this, we take one of the papers identified in our survey [22], and reproduce its data collection procedures using PRISONER. We choose this paper as it is the only study in our analysis to meet all ten criteria, suggesting it should be possible to fully recreate its procedures.

Our chosen paper [22] studies attitudes towards information-sharing with third-party Facebook applications, by evaluating how well participants understand the data-handling practices of applications, and the differences between features operated by Facebook and applications provided by third-parties. The authors built a Facebook application to deliver a survey about privacy attitudes, which masqueraded as a personality quiz to encourage participation. Participants believed their responses would be used to classify them as one of a number of personality types. In reality, the application measured a participant's level of engagement with Facebook based on how many profile attributes they disclose (such as age, gender, and work history), and how many status updates they shared. This was used to provide a classification "for entertainment value" to the participant, while providing a quantitative measure of how much information they disclose on Facebook. To achieve this, the researchers collected significant amounts of information from a participant's profile using the Facebook API. The authors "collected data about each respondent's profile (but no actual profile data) in order to compute measures of how much information people were sharing on Facebook. For most fields we computed a simple binary score (1 if the field contained data, 0 if blank) or a count if available (such as the total number of status updates and the number of status updates in the past 30 days)". This

suggests that at no stage were any sensitive data stored, but in order to compute these measures, requests for the data had to be made. In this instance, the authors make good faith efforts to protect the privacy of their participants, but in replications, such details are easily overlooked, and could easily lead to inappropriate quantities of information being stored.

This study is ideal to model using PRISONER, as it relies on the collection of large quantities of data, while demonstrating a clear workflow that dictates how data should be sanitised and aggregated through the duration of the experiment. To recreate this workflow with PRISONER, we create a *privacy policy* which encodes the requirements we have discussed, in terms of which OSNs are accessed, which data types we require, and how they should be sanitised. We then write an exemplar web-based application which supplies this policy to the PRISONER web service, then makes requests to the PRISONER API whenever Facebook data are required.

To illustrate how this works in practice, we take each of the criteria in the *methods* category of our survey, and explain how we apply PRISONER to achieve that aspect of reproducibility. The privacy policy which we created for this example can be accessed online [15].

- **Source OSN** - To replicate this study, we need to collect data from Facebook. PRISONER allows researchers to request generic *social objects* which exist on various OSNs, such as *people* or *notes* (which can resolve to Facebook status updates, or tweets, for example). As this experiment only uses a single OSN, however, we make this explicit in the experiment's privacy policy. We create policies for each type of object our experiment needs to retrieve. Our policy for *Facebook:User* only allows us to retrieve profile information from that OSN as it is explicitly namespaced. This policy can be shared with others to ensure any further data collection comes from the same service, while a policy for a generic *Person* object could allow a replication to use data from any compatible OSN.

- **Length of study** - This study requires us to collect all of a participant's status updates in order to determine how many have been posted. PRISONER allows privacy policies to include temporal constraints on the data collected. This ensures that when the policy is reused, data from evolving sources, such as Facebook status updates, are only accessible from the same time period, or over the same duration. This study requires that a user's entire history of status updates is collected, so that the total number can be counted, so we did not provide an explicit time limit in this instance.

- **Data processing** - This study outlines some crucial data sanitisation requirements which must be preserved to both replicate the conditions of the study and preserve participant privacy. As described earlier, we do not need to collect the content of profile attributes or status updates, but rather a count of how many are accessible. When manually evoking the Facebook

API to do this, it would be necessary to collect the sensitive data then manually sanitise it. While achieving the desired result, this is not ideal, due to the possibility that data may be inappropriately stored in an unsanitised form, especially when using third-party bindings which may implement their own clientside caching behaviours. This may risk participant privacy. By encoding these data-handling requirements in a declarative manner in the experiment's privacy policy, researchers do not need to be concerned with such implementation details. PRISONER includes *transformation* primitives which support such declarations by providing a range of common sanitisation techniques. To ensure we do not inadvertently collect too much information, we only request the *id* attributes from status updates, as shown in the *attributes* collection in the policy. On all other requests for sensitive attributes, such as work history or gender, we use *reduce* transformations whenever we retrieve data. The *bit* attribute immediately sanitises the response from the Facebook API to only return 1 if the attribute is present, or 0 if it is not, before the data are made available to the experimental application. As well as only collecting the number of profile attributes, the study requires that "respondents' Facebook user IDs were hashed for anonymization purposes". The transformation policy for the *User* object shows we hash the user ID using SHA-224 after retrieving it. Note that while this technique is commonly used to provide a degree of obfuscation, it is not impervious to attack. PRISONER does not provide any guarantees about the anonymity afforded by use of such techniques, and we are looking into incorporating other approaches such as differential privacy into the framework in later work.

- **Consent** - The authors note that "a consent statement appeared on the first page of the survey.", but this is not sufficient to replicate the study, as language used to obtain consent can impact the results of OSN research [23]. As all attributes collected from OSNs are encoded in an experiment's policy, PRISONER can generate participant consent forms that explain which OSNs data are collected from, which attributes are collected, and how data are processed through the life of the experiment. This information is provided in a consistent, human-readable format which ensures a participant's informed consent is tied to the exact procedures of the experiment. When PRISONER workflows are replicated, the consent language is consistent.

- **IRB/Ethics** - The authors explain that "our design was reviewed and approved by our university's IRB." While it is encouraging to see this confirmed, the tendency to not routinely share IRB protocols presents some challenges to reproducibility, particularly where the actual procedures of an experiment have drifted from the previously agreed protocol, so-called "ethical drift". While it is beyond the scope of PRISONER to resolve these challenges, allowing researchers to share a testable specification of the data-handling
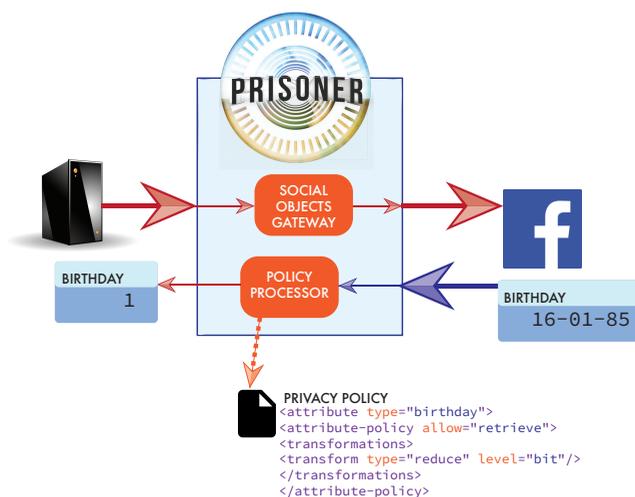

Fig. 5. Illustration of how a request for data is handled by PRISONER. An experimental application makes a request to PRISONER's social objects gateway for an object, which is delegated to the appropriate social network. The object is returned and handled by PRISONER's policy processor, which invokes the privacy policy for the experiment to ensure the data are suitably sanitised. In this example, the participant's birthday has been reduced to a bit indicating its presence before being returned to the application.

requirements of a study with their IRB when making an application, rather than a speculative protocol, constitutes an improvement on the state of the art.

- **Participant briefing** - The authors explain some of their briefing procedures, particularly "Our university's name and seal were featured prominently on every page of the survey and on the app's home page on Facebook.", which may have a priming effect and is important to be able to replicate. While researchers are responsible for conducting their own participant briefing, PRISONER provides a consistent "bookending" experience, including the presentation of consent forms, which explain the procedures of the experiment. This, when augmented by other cosmetic details, such as those outlined by the researchers in this study, provides a degree of consistency between replications.

We do not replicate the entire experiment in [22], but rather recreate its data collection requirements which we demonstrate in a simple example which attempts to retrieve a plethora of information about participants.

As discussed earlier, this experiment requires access to a participant's Facebook profile to determine the presence of certain attributes. Fig. 5 illustrates for one such attribute, how data are handled by the framework. As shown, at the beginning of the experiment, the participant provides the PRISONER gateway with access to their Facebook account, binding this to their PRISONER session, and ensuring any requests for profile data are made via the PRISONER proxy. When the experimental application requests these data, PRISONER's policy processor consults the applica-

tion's privacy policy for an appropriate "retrieve" clause to determine whether the application can access the attribute, and if any sanitisation should occur. In the example shown, the experimental application needs to determine whether the participant discloses their birthday. Thus, the policy processor sanitises the attribute before making it available to the application, and the sensitive attributes are discarded.

Having produced a policy file, it can be distributed to other researchers who can subsequently replicate the workflow. Even if researchers do not have access to the original code for the experiment, they can build an application against the same policy to make requests for data. They bootstrap an instance of PRISONER by providing a URL to the policy to generate all consent forms, briefing materials, and gain access to the OSN authentication flow and sanitisation API without writing any code. In addition, if a researcher wished to run the same experiment using, e.g., Twitter as the source OSN, simply replacing any reference to "Facebook" with "Twitter" will provide this without any further modification. We are still developing the PRISONER tools with the aim of releasing them to the community in the near future. We have made available the privacy policy and source code for this replication to show how these components are written [15].

In this example, we have shown how an experiment can be managed by PRISONER in a reproducible and ethical manner. Even if we were to conduct this experiment and not share our application's source code, other researchers can replicate the experiment in their environment of their choice, but re-use our experiment's workflow to ensure data are collected under the same conditions. It is important to note, however, that workflow sharing alone is not sufficient to guarantee the accuracy of replications, particularly as this may not consider all possible corner cases which could affect the result of a replication. As we have discussed, a wider culture change is needed to achieve a higher degree of reproducibility.

7 FURTHER WORK

Our results illustrate some of the challenges in conducting reproducible OSN research, and we have demonstrated our work towards tackling some of the challenges in capturing and replicating the methodology of OSN experiments with our architecture, PRISONER. In further work, we will develop the architecture further. Currently, the architecture handles Facebook, Twitter, and Last.fm, but we intend to include support for other popular OSNs. We will also be extending our sanitisation tools to include state of the art techniques for supporting anonymous data disclosure, such as differential privacy. We will be developing tools to help researchers define their PRISONER policies without having to manually write XML, as well as developing features to abstract experiments further from the implementation of APIs by mapping older experimental code to newer versions of the underlying APIs, solving the challenge of not being able to reuse legacy code which uses OSN APIs. PRISONER is still in development and will be

publicly released to allow other researchers to benefit and contribute, and to use it to enable reproducibility in other OSN research.[14]

8 CONCLUSIONS

In this paper we have conducted a comprehensive survey of the recent OSN literature to assess to what extent research in the field supports reproducibility.

We find that publications across a range of venues rarely share their data or support recomputability of results. As there is other work which strives to improve this, we focus on the challenge of capturing the methodology in OSN experiments. Our analysis of the state of the art shows that while systems-oriented papers are often better at reporting some of the fundamental attributes of their methodology, they often do not consider the provenance of their data, such as how the data were sampled, or how participants were handled. Conversely, social science papers are often better at explaining their participant-handling procedures, such as whether informed consent was obtained, but may not explain some structural details of their dataset. Our findings motivate a set of four recommendations for OSN researchers, which combine the strengths we find across the disciplines. We have built an architecture which aims to support these recommendations, which we demonstrate by recreating one of the experiments from the survey. We hope that this will be of use to the community in encouraging a shift towards reproducibility in OSN research.

9 ACKNOWLEDGEMENTS

REFERENCES

[1] A. E. Ali, S. N. A. van Sas, and F. Nack. Photographer Paths: Sequence Alignment of Geotagged Photos for Exploration-based Route Planning. In *Proc. CSCW '13*, pages 985–994, San Antonio, TX, USA, 2013. doi:10.1145/2441776.2441888.

[2] S. Alim. An initial exploration of ethical research practices regarding automated data extraction from online social media user profiles. *First Monday*, 19(7), 06 July 2014. doi:10.5210/fm.v19i7.5382.

[3] d. boyd and N. B. Ellison. Social Network Sites: Definition, History, and Scholarship. *J Comput-Mediat Comm*, 13(1):210–230, Oct. 2007. doi:10.1111/j.1083-6101.2007.00393.x.

[4] L. Butler. Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework. *Ethics in Science and Environmental politics*, 8(1):83–92, 2008. doi:10.3354/esep00077.

[5] R. Caers, T. De Feyter, M. De Couck, T. Stough, C. Vigna, and C. Du Bois. Facebook: A literature review. *New Media Soc*, 15(6):982–1002, Sept. 2013. doi:10.1177/1461444813488061.

[6] P. Dawson. Our anonymous online research participants are not always anonymous: Is this a problem? *Br J Educ Technol*, 45(3):428–437, May 2014. doi:10.1111/bjet.12144.

[7] M. De Choudhury, S. Counts, and E. Horvitz. Social Media As a Measurement Tool of Depression in Populations. In *Proc. WebSci '13*, pages 47–56, Paris, France, 2013. doi:10.1145/2464464.2464480.

[8] S. Denef, P. S. Bayerl, and N. A. Kaptein. Social Media and the Police: Tweeting Practices of British Police Forces During the August 2011 Riots. In *Proc. CHI '13*, pages 3471–3480, Paris, France, 2013. doi:10.1145/2470654.2466477.

14. http://prisoner.cs.st-andrews.ac.uk

[9] D. L. Donoho, A. Maleki, I. U. Rahman, M. Shahram, and V. Stodden. Reproducible research in computational harmonic analysis. *Comput Sci Eng*, 11(1):8–18, Jan. 2009. doi:10.1109/mcse.2009.15.

[10] J. Freire, D. Koop, F. Chirigati, and C. T. Silva. Reproducibility Using VisTrails. In V. Stodden, F. Leisch, and R. D. Peng, editors, *Implementing Reproducible Research*. Chapman and Hall/CRC, 2014. Online at http://osf.io/c3kv6/.

[11] D. B. Fridsma, J. Evans, S. Hastak, and C. N. Mead. The BRIDG project: a technical report. *JAMA-J Am Med Assoc*, 15(2):130–137, 2008. doi:10.1197/jamia.m2556.

[12] I. P. Gent. The recomputation manifesto, 12 Apr. 2013. Online at http://arxiv.org/abs/1304.3674.

[13] C. A. Goble, J. Bhagat, S. Aleksejevs, D. Cruickshank, D. Michaelides, D. Newman, M. Borkum, S. Bechhofer, M. Roos, P. Li, and D. De Roure. myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Res*, 38(suppl 2):W677–W682, 1 July 2010. doi:10.1093/nar/gkq429.

[14] S. A. Golder and M. W. Macy. Digital footprints: Opportunities and challenges for online social research. *Annu Rev Sociol*, 40(1):129–152, July 2014. doi:10.1146/annurev-soc-071913-043145.

[15] T. Henderson and L. Hutton. Data for the paper "Towards reproducibility in online social network research", Aug. 2014. doi:10.6084/m9.figshare.1153740.

[16] C. M. Hoadley, H. Xu, J. J. Lee, and M. B. Rosson. Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry. *Electron Commer R A*, 9(1):50–60, 10 Jan. 2010. doi:10.1016/j.elerap.2009.05.001.

[17] C. Hurlin, C. Pérignon, and V. Stodden. RunMyCode.org: A Research-Reproducibility Tool for Computational Sciences. In V. Stodden, F. Leisch, and R. D. Peng, editors, *Implementing Reproducible Research*. Chapman and Hall/CRC, 2014. Online at http://osf.io/39eq2/.

[18] L. Hutton and T. Henderson. An architecture for ethical and privacy-sensitive social network experiments. *SIGMETRICS Perform. Eval. Rev.*, 40(4):90–95, Apr. 2013. doi:10.1145/2479942.2479954.

[19] G. Iachello and J. Hong. End-User Privacy in Human-Computer Interaction. *Foundations and Trends in Human-Computer Interaction*, 1(1):1–137, 2007. doi:10.1561/1100000004.

[20] E. Jain, A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. Suzek, M. Martin, P. McGarvey, and E. Gasteiger. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, 10(1):136+, 2009. doi:10.1186/1471-2105-10-136.

[21] M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and Privacy: It's Complicated. In *Proc. SOUPS '12*, Washington, DC, USA, July 2012. doi:10.1145/2335356.2335369.

[22] J. King, A. Lampinen, and A. Smolen. Privacy: Is there an app for that? In *Proc. SOUPS '11*, Pittsburgh, PA, USA, 2011. doi:10.1145/2078827.2078843.

[23] S. McNeilly, L. Hutton, and T. Henderson. Understanding ethical concerns in social media privacy studies. In *Proc. ACM CSCW Workshop on Measuring Networked Social Privacy: Qualitative & Quantitative Approaches*, San Antonio, TX, USA, Feb. 2013. Online at http://tristan.host.cs.st-andrews.ac.uk/pubs/mnsp2013.pdf.

[24] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *Proc. ICWSM '13*, 2013. Online at http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6071.

[25] M. T. Mullarkey. Socially immature organizations: A typology of social networking systems [SNS] with organizations as users [OAU]. In *Proc. CSCW '12*, pages 281–292, Seattle, WA, USA, 2012. doi:10.1145/2141512.2141604.

[26] A. Narayanan and V. Shmatikov. De-anonymizing Social Networks. In *Proc. 30th IEEE Symposium on Security and Privacy*, pages 173–187, Oakland, CA, USA, May 2009. doi:10.1109/sp.2009.22.

[27] J. Nichols and J. H. Kang. Asking questions of targeted strangers on social networks. In *Proc. CSCW '12*, pages 999–1002, Seattle, WA, USA, 2012. doi:10.1145/2145204.2145352.

[28] T. Paek and B. J. Hsu. Sampling Representative Phrase Sets for Text Entry Experiments: A Procedure and Public Resource. In *Proc. CHI '11*, pages 2477–2480, Vancouver, BC, Canada, 2011. doi:10.1145/1978942.1979304.

[29] H. A. Piwowar, R. S. Day, and D. B. Fridsma. Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3):e308+, 21 Mar. 2007. doi:10.1371/journal.pone.0000308.

[30] R. Procter, M. Poschen, W. Lin Y, C. Goble, and D. De Roure. Issues for the Sharing and Re-Use of Scientific Workflows. In *Proc.*

5th International Conference on e-Social Science*, 2009. Online at http://www.escholar.manchester.ac.uk/uk-ac-man-scw:117546.

[31] M. A. Rothstein and A. B. Shoben. Does Consent Bias Research? *The American Journal of Bioethics*, 13(4):27–37, 2013. doi:10.1080/15265161.2013.767955.

[32] G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig. Ten simple rules for reproducible computational research. *PLoS Comput Biol*, 9(10):e1003285+, 24 Oct. 2013. doi:10.1371/journal.pcbi.1003285.

[33] L. Solberg. Data mining on Facebook: A free space for researchers or an IRB nightmare? *University of Illinois Journal of Law, Technology & Policy*, 2010(2), 2010. Online at http://www.jltp.uiuc.edu/works/Solberg.htm.

[34] C. Steinfield, N. B. Ellison, and C. Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *J Appl Dev Psychol*, 29(6):434–445, Nov. 2008. doi:10.1016/j.appdev.2008.07.002.

[35] V. Stodden. The Scientific Method in Practice: Reproducibility in the Computational Sciences. Technical Report 4773-10, MIT Sloan School of Management, 09 Feb. 2010. doi:10.2139/ssrn.1550193.

[36] V. Stodden and S. Miguez. Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. *Journal of Open Research Software*, 2(1):21+, 09 July 2014. doi:10.5334/jors.ay.

[37] P. A. Thompson and A. Burnett. Reproducible Research. *CORE Issues in Professional and Research Ethics*, 1(6), 2012. Online at http://nationalethicscenter.org/content/article/175.

[38] F. Trevisan and P. Reilly. Ethical dilemmas in researching sensitive issues online: lessons from the study of British disability dissent networks. *Information, Communication & Society*, 17(9):1–16, 2014. doi:10.1080/1369118x.2014.889188.

[39] J. G. Warrell and M. Jacobsen. Internet research ethics and the policy gap for ethical practice in online research settings. *Canadian Journal of Higher Education*, 44(1):22–37, 2014. Online at http://ojs.library.ubc.ca/index.php/cjhe/article/view/2594.

[40] R. E. Wilson, S. D. Gosling, and L. T. Graham. A review of Facebook research in the social sciences. *Perspec Psychol Sci*, 7(3):203–220, May 2012. doi:10.1177/1745691612442904.

[41] Yale Law School Roundtable on Data and Code Sharing. Reproducible research. *Comput Sci Eng*, 12(5):8–13, Sept. 2010. doi:10.1109/mcse.2010.113.

Luke Hutton is a Research Fellow in Computer Science at the University of St Andrews. He holds a BSc degree in Applied Computing from the University of Dundee, and a PhD in Computer Science from the University of St Andrews. His research interests include privacy, ethics, and reproducibility of social network research.

Tristan Henderson is a Senior Lecturer in Computer Science at the University of St Andrews. He holds an MA in Economics from Cambridge University, and an MSc and PhD in Computer Science from University College London. His research interests include privacy, network measurement and wireless networks, including the running of the CRAWDAD wireless network data archive (http://crawdad.org/).