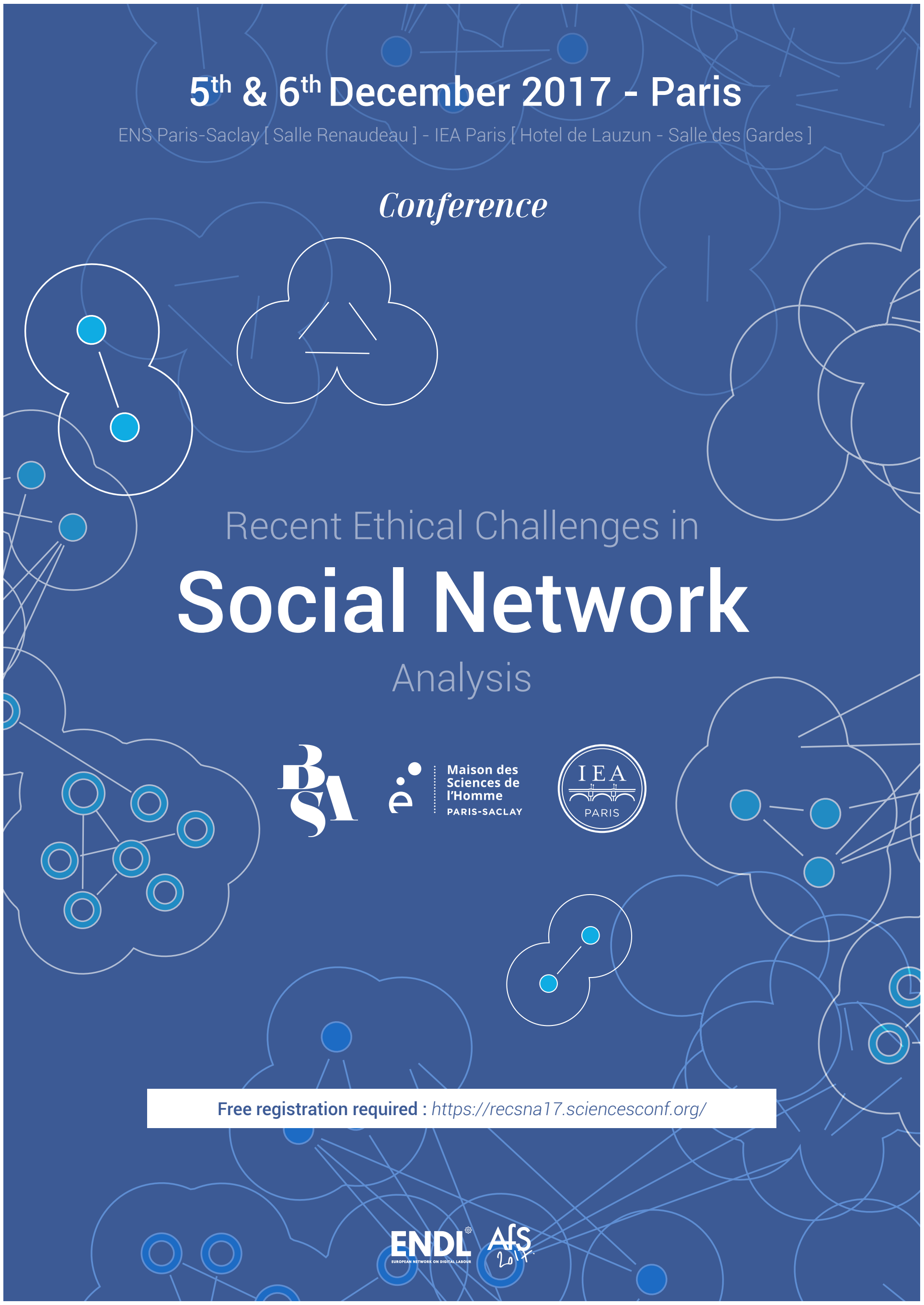

5th & 6th December 2017 - Paris

ENS Paris-Saclay [Salle Renaudeau] - IEA Paris [Hotel de Lauzun - Salle des Gardes]

Conference

Recent Ethical Challenges in
Social Network
Analysis

BSA

Maison des Sciences de l'Homme
PARIS-SACLAY

IEA
PARIS

Free registration required : *https://recsna17.sciencesconf.org/*

ENDL
EUROPEAN NETWORK ON DIGITAL LABOUR

AFS 2017

Contents

Day 1: 5 December 2017, Salle Renaudeau, ENS Paris-Saclay

Breakfast, 8:30 - 9:00

Welcome, 9:00 - 9:10

Coffee break, 10:40 - 11:00

Lunch, 12:30 - 13:30

Break, 15:00 - 15:20

Day 2: 6 December 2017, Salle des Gardes, IEA de Paris

Breakfast, 8:30 - 9:00

Break, 11:00 - 11:15

1 Ethical standards, 9:10 - 10:40

Chair: Michael Schönhuth, Universität Trier – Trier, Germany

1.1 Relational vs individual rights? Applying a relational autonomy lens to evaluating social network analysis research ethics applications (Mandy Lee, Trinity College Dublin), 9:10 - 9:30

In this paper, I would outline a new ethical framework for assessing research ethical approval for studies involving social network analysis, particularly those in healthcare. Traditionally, research ethical governance in healthcare has been derived primarily from bioethics guidelines which emphasizes individual autonomy and informed consent for use of personal data. The duty of care from the investigator concerns primarily giving enough information about the study, including risks and benefits assessed on an individual basis, to enable informed consent and thence voluntary participation to take place. However, this framework regarding individual risks and benefits and individual voluntary autonomy rests on assuming that the research being conducted follows the traditional atomised model of variable-based research, whereby only the attributes of individual participants are of interest, and no interrelationships are/can be assumed to exist amongst the study population. Such a bioethical frame towards research ethics has long been criticised as unsuitable to social research, especially those derived from qualitative traditions of inquiry that are predicated upon exploring reciprocal and long-standing relations amongst community members. This problem is exacerbated for SNA research which has been increasingly applied in the health sector, not only because SNA takes as its central premise the relationships amongst actors as paramount to understanding social phenomena, but the quantitative measures and visualisations make traditional anonymisation procedures problematic to execute. I submit that we need to overhaul our guiding ethical framework for evaluating such relationally-focused research, by proposing the concept of relational autonomy as existing independently and additionally to individual autonomy, and how this could be translated to a research ethical framework theoretically – in terms of guiding principles for research ethical governance; and practically – in terms of information to be provided for ethics committees and target participants, and the protocols to be followed to enable relational informed consent to take place.

1.2 The presentation of the networked self: Ethical challenges in social network analysis (Alessio D'Angelo, Middlesex University & Louise Ryan, University of Sheffield), 9:30 - 9:50

For decades, social network analysis (SNA) has experienced a growth in volume and subject areas, accompanied by the development of technological tools and increasing availability of data. Whilst opening new pathways for sociological investigation, this also raises specific ethical challenges, an aspect which has been often overlooked in academic discussions.

This paper aims to bring to the fore some major issues in SNA research and to explore them in light of broader developments within contemporary society. In particular, we recognise that, with the rise and omnipresence of social media, most people tend to have very strong ideas about what social networks are and on how to present their own 'networked self'. This compares to less than two decades ago, when 'social network' was a concept that had to be explained carefully to most research participants.

In regards to this, drawing on Goffman's concept of 'presentation of self', we explore ethical issues around how people present themselves when disclosing data about their own personal or professional networks. This is also relates to the challenges of network mapping with people who know each other, how this can influence responses, affect those involved and impact on data reliability. Finally, we address the new shape of old challenges of confidentiality and anonymity for network researchers.

The paper builds on a range of research projects undertaken by the authors in different settings (including research with EU migrants; secondary school students; and BME organisations) and with different methods (qualitative, quantitative and mixed-methods SNA), to ensure its theoretical contributions are well grounded in empirical data.

1.3 Challenges in ethics application and ethics research design for the study of social networks of healthcare providers (Rosica Pachilova, University College London), 9:50 - 10:10

This paper presents challenges encountered in the application for ethics approvals and the research design of a study that investigates communication networks amongst healthcare providers in a hospital. Of particular interest is the 'chicken-egg problem': where the researcher should start from and how to test research methodologies prior to obtaining ethics approval. In addition, the practicalities of the study design in busy clinical wards will be discussed.

The empirical study investigates how the layout of hospital wards influences communication networks amongst healthcare professionals and how this affects the quality of care provided to patients. Several different methods - social network surveys, ego network interviews and sociometric badges, were considered to collect social network data in a hospital in the UK. Each method is described and discussed in terms of ethics considerations and difficulties to recruit participants and obtain informed consent. The ethics approval documents and process are reviewed and the journey of obtaining ethics approval for the study is described. What works well on paper does not necessary works in practice and difficulties in collecting data after the ethics approval of the project are discussed.

Recommendations on improving the ethics application process are considered in the conclusion. It is suggested that a community with experience in REC approval similar to SOCNET would be a good starting point to advise researchers prior to applying for ethics approval. In addition, the research ethics committee should be able to advise prior to the official submission of an application. Currently medical RECs lack understanding of social network methods and more diversity in the committee would be required to fill this gap.

1.4 Discussion, 10:10 - 10:40

2 Impact and audiences, 11:00 - 12:30

Chair: TBC

2.1 Whose results are these anyway? Negotiating diverging agendas after SNA research (Paola Tubaro, Centre National de la Recherche Scientifique Paris-Saclay), 11:00 - 11:20

What ethical issues may arise in the process of 'returning' results to the community where a social network analysis (SNA) research was undertaken? Standard ethical review processes provide little guidance as they tend to focus on the first stages of a study – before rather than after the data collection. The field methods literature does hint at the importance of rendering something to study participants in return for data access, but often neglects their broader social environment. Through a reflective account of two empirical social network studies, the proposed paper engages with issues arising from results-sharing in regard to: 1/ study participants; 2/ researchers themselves; 3/ policymakers, the press and the wider public.

First, discussion is on how potential effects on participants ' well-being are magnified in SNA insofar as digital technologies make connectedness increasingly visible – and attach values and status to it. By the same token, there is a high likelihood that participants (or their social network contacts) may appropriate researchers' concepts for their own purposes. The paper leverages Chardel' s (2013) notion of dual vulnerability – of participants and researchers alike – to apprehend the emergence of such issues from the relational design of a social networks study, and to identify means to mitigate them.

Finally, the paper addresses results-sharing with a broad range of stakeholders including journalists and policymakers. Despite growing public interest in SNA, researchers may struggle to contrast misunderstandings as to the nature of social networks – often conflated with social media in common parlance – and prejudices against a presumed negative effect that they (or rather, social media) may have. Again, researchers' vulnerability mirrors that of the community under study. By way of conclusion, it is discussed how a researcher-driven ethical framework – based on shared hands-on experience and a reflective attitude – may moderate these effects.

2.2 That's you here on the map : introducing results to network participants (Clément Renaud, Ecole Polytechnique Fédérale de Lausanne), 11:20 - 11:40

By producing representations of networks, researchers are often seeking to unfold things that have been left unseen. The process of visualization is used to cast new lights and reveal critical informations about relationships in these networks. The "bird-eye view" and its relative distance to the studied object can sometimes leave outside the scope of the researcher how unpleasant and unwelcome these representations can be.

In this communication, we will describe the challenges faced on the field during the confrontation of networks participants with observations made from data analysis. Our examples rely on extensive fieldwork and use of social network analysis in China. We will see how the publications of visualizations based on social media network analysis have - unwillingly - contribute to expose political activists by revealing specific network structures and practices. We will then focus on solo or group interviews presenting visualizations as supports for comments and discussions. Interviews constitute a powerful tools for social scientists to confront the results and implications of their - even web-based - research. Showing maps and images usually generate much interest, especially about oneself and others positions. While these on-site interviews are precious to validate observations, they may also endanger other stakeholders represented on the visualizations.

Based on these reflexions, we will propose different solutions to conduct such interviews and avoid exposure for other participants by preparing visualizations - focusing, splitting and anonymizing networks, hiding and grouping nodes, etc. This discussion will be the occasion to reassert the importance of field investgation as a complementary tool to validate data findings - and as a reminder of the unintentional impact of most data analysis. We will insist on the ethical responsability of the researcher in the production and diffusion of representations that can easily be misread and misused.

2.3 The impact of network research on postgraduate students – a case study (Neylson Crepalde, Minas Gerais Federal University & Methodist Technological University Izabela Hendrix), 11:40 - 12:00

In 2015, the author conducted a network study with 47 postgraduate students of a Brazilian university aiming to discuss the social capital inequalities among them1. That study showed how postgraduate students build ties with each other increasing their productivity and also that tie formation is most conditioned by the recognition of methodological skills and participation in the same research groups. In that process, the author faced two problems: (1) getting a good response rate and (2) presenting the results to the community. These problems bring to discussion two ethical issues related to network science: first, the impact on the interviewee of having to decide between his colleagues (what of his colleagues he would indicate for a job or not, who he considers to have methodological or theoretical abilities). Second, the impact of making him aware of his position within the network structure. Even if the data is anonymized, a very central node or a very peripheral one would easily recognize himself in the presentation of the results. Considering that research in postgraduate programs is considered one of the most stressful activities nowadays, if, say, a peripheral person was aware of his condition, would he decide to take charge of his own career and move on or would he fall into depression and anxiety and probably quit the program? The main goal of this paper is to discuss these ethical issues and present some strategies to induce a positive impact both in the results reception and on students self-esteem.

2.4 Discussion, 12:00 - 12:30

3 Online research and social media, 13:30 - 15:20

3.1 Ethical issues for online social media data in SNA (Nina MacFarlene, Edinburgh Napier University), 13:30 - 13:50

Although analysis of SNA online data is a fairly new phenomenon, there has been a long-standing debate about the ethical side of it. The main focus of this paper is how to treat an online forum, which is the source of SNA data, – as a public or as a private space. This can usually be established through the fact if the registration is required or not (Eysenbach and Till, 2001). Thus, if the information is available from a web-space that is accessible without registration or a request to join, it is usually considered that the users of the forum are aware that the information they post is publicly available, and in some cases they even want the information to reach a wider audience (Vayreda and Antaki, 2009; Horne and Wiggins, 2009; Tabor and Milfont, 2011).

The reading audience on such forums is a lot bigger than the posting community. As the registration is only required for posting, it is possible that the users only do it when they decide to post, which is supported by the fact that there are sometimes 10 times more viewers than registered users on the forum.

If SNA data corpus is collected from publicly accessible online forums, which do not require registration to view, however they do require registration to post – would it be considered a public or private space and could the data be used without overt permission of the authors? Who owns online social media data – the author, the website or the public? The paper is looking at the issues of social media data ownership and the ethical ways of collecting, analysing and publishing it, to ensure maximum protection of the authors identity.

3.2 The ethics ecosystem: Personal ethics, standard ethical practice and regulating actors governing social media research data (Gabby Samuel, Gemma Derrick & Thed Van Leeuwen, Lancaster University), 13:50 - 14:10

This paper examines the consequences of a culture of ”personal ethics” when using new methodologies, such as the use of social networking sites (SNS) as a source of data for research.More broadly, this paper explores the practices of a number of actors and researchers within the ”Ethics Ecosystem” which governs the regulation and promotion of ethically responsible research behaviour when using SNS data for research. The ethical use of this data is currently in dispute, since even though it is seemingly publicly available, concerns relating to privacy, vulnerability, potential harm and consent blur the lines of responsible ethical research behaviour. The findings point to the dominance of a personal, bottom-up, researcher-led,'ethical barometer' for making decisions regarding the permissibility of using SNS data. Given the ethical complexities of using SNS data in research, this bottom-up, researcher-led approach is viewed as best practice. However, we show that the use of different barometers by different researchers can lead to wide disparities in ethical practice - disparities which are compounded by the lack of firm guidelines for responsible SNS use. This has widespread consequences on the development of shared norms and understandings at all levels, and by all actors within the ethics ecosystem,and risks inconsistencies in their approaches to ethical decision-making. This paper argues that this governance of ethical behaviour by individual researchers perpetuates a negative cycle of academic practice that is dependent on subjective judgments by researchers themselves, rather than governed by more formalised academic institutions such as the research ethics committee and funding council guidelines.

3.3 Informed consent, disclosure of anonymity, and trust: Ethical challenges in a case study of the networks of inter-urban itineraries of Russian tourists in Spain (Deniza Alieva & Isidro Maya Jariego, Universidad de Sevilla), 14:10pm - 14:30

We conducted an online survey to 354 Russian tourists that had visited Andalusia (Southern Spain), to describe the network of itineraries between cities. Participants were contacted through two big internet forums where people share their tourist experiences. In this study case, we examine the ethical challenges of obtaining informed consent, getting participants to reveal their pseudonyms (or disclose anonymity) and to generate sufficient confidence in the interviewees to obtain a valid and reliable report about their trip in Spain. First, although the internet forums contained information about the tourist itineraries and the list of places they had visited, we decided not to use such information without the consent of the users of the forum. Instead, we conducted an online survey using the contact information available about users. However, this was a second difficulty, since the participants of the forum received a request that meant revealing their identity: that is, in order to participate in the survey they had to communicate their email and provide some personal information. In fact, only 4.4 percent of the people we contacted declined to participate, and 14.7 percent of those who agreed to participate did not finally complete the questionnaire. In any case, once they agreed to be contacted by mail they showed less privacy concerns than when they shared information in the forums. Third, the application of online questionnaires makes it more difficult to develop a personal rapport so that respondents share information openly and confidently. To address these three challenges, we deployed a research strategy that facilitated a climate of trust with the participants. The survey was coordinated by a Russian-speaking researcher, who reduced perceived cultural distance with the target population. This researcher participated in the forum from a few months before starting the survey, which facilitated the establishment of personal links. The questionnaires were completed in Russian. After the first interviews, the other participants were selected following a snowball procedure, from the network of contacts in the forums. It was also useful to propose flexible time margins to return the survey, which in the case of the Russian population seems to facilitate answering questionnaires and generally increases participation rates. Finally, we obtained a completion rate 85.3 per cent and a response rate of 81.6 per cent (Table 2). This is well above what is usual in internet surveys, and is even above the data that are normally obtained in face-to-face interviews.

3.4 Discussion, 14:30 - 15:00

4 Economic and political risks and threats, 15:20 - 16:50

Chair: Catherine Comet, Université de Lille

4.1 Social network data in reinventions of collective responsibility (Emmanuel Lazega, CSO - Institut d'Etudes Politiques de Paris) 15:20 - 15:40

It has become routine for many consulting firms serving large multinational corporations to buy data from private online social network platforms and to use this data for recruitment (and presumably for redundancies) of personnel. Based on interviews, we know for example that, in the analyses of this data, consulting firms currently differentiate between 'personal' data and 'intimate' data. They consider using 'personal' data to be appropriate, while using 'intimate' information to be inappropriate –the difference and boundary between the two, however, being considered a matter of interpretation. Regulators (see the forthcoming European GDRP) are vigilant about the use of personal data, but they face the classic problem of not being organized to enforce the law efficiently and impose new rules to well known foreign private social network hegemons that are beyond their jurisdiction and protected by their respective governments. This paper focuses on the implications of such redefinitions of boundaries by reflecting on the extent to which social network databases can be used by powerful public and private actors to redefine the notion of collective responsibility for social control purposes (in family life, business life, public health-related issues, etc.), i.e. to reinvent collective responsibility by extending it in practice to newly selected alters in anyone's personal networks.

4.2 The Trump-SCL debacle: a case study in social network analysis abuse and digital labor exploitation (Antonio A. Casilli, Telecom ParisTech & Rochelle Laplante), 15:40 - 16:00

In 2014, Strategic Communication Laboratories (SCL), a shady Cambridge-based research company turned to Amazon Mechanical Turk to recruit participants for a network survey. MTurk is a marketplace for micro-work. 185,000 "Turkers" were recruited to fill in a questionnaire and install a small app that downloaded their Facebook friends' data without their knowledge or consent.. This not only raises several privacy concerns, but also a set of questions as to the labor rights of the MTurk micro-workers.

Despite efforts to create awareness via discussion forums, best practices, and software solutions, Turkers seldom know who they are working for and what is the purpose of their tasks. Their vulnerability mirrors that of their unaware Facebook friends, who were de facto made into uninformed, non-contractual, unpaid "data labor providers".

SCL's initial collection eventually evolved into a 30 million profiles database, which was put to good use by its American branch: Cambridge Analytica. Eventually the data landed in the lap of the 2016 US presidential Republican candidate. His campaign used the data to target voters in swing states with crowdturfed political messages. Crowdturfing consists in recruiting contingent meme and "fake news" producers while as well as crowds of micro-workers to share, click, follow, flood social media, and influence the public perception.

This case study addresses questions as to the ethics of network research. It also provides insight into a sea-change in data collection: the increasing reliance on respondents' underpaid or unpaid digital labor to replenish increasingly big databases. Freelancers, micro-workers and free data providers are recruited through online platforms which act as labor markets. Thus, ethical concerns about the use and abuse of network data become inextricable from questions about the effectiveness of democratic deliberation in the wake of the rise of digital labor platforms.

4.3 How ethical is to exploit open data extracted through 'ethical hacking' ? The case of the Panama Papers (Antoine Vion, Aix-Marseille Université), 16:00pm - 16:20

The structuring of the Panama papers open dataset proceeds from an operation known as "ethical hacking" of the database of assembly files of screens companies managed by Mossack Fonseca, a specialized firm in Panama. While the international consortium of investigative journalists (ICIJ) is a well-known and recognized organization, its means of access to the transaction files of a law firm specializing in tax evasion are naturally surrounded by a halo of mystery. How and by whom has this operation been carried out and protected? Technical assumptions have been made on how to extract data: SQL injection in the company's payment system, according to a hacker, access to mailboxes by exploiting flaws in the WordPress Revolution Slider plug-in according to experts from Wordfence, etc. (Paganini 2016).

In any case, the full archive contains more than 11.5 million files, weighing 2.6 terabytes of data relating to offshore company mounts used by a very diverse clientele (state or international organizations, corporations, liberal professions, money laundering professionals, etc.). The structure and the volume of this archive are not commensurate with the previous operations, since the Panama Papers represent ten times more data than the Offshore Leaks, and one thousand five hundred times more than Wikileaks.

The ICIJ website posts graphs of relationships in a hotel suite in the Seychelles.

The focus of this paper will be to discuss the methodological and ethical issues raised by this kind of research experience. My main purpose is to address three challenges:

- Should sociologists leave aside such kind of data while journalists exploit them to manage scoops ? Is it worth producing some structural knowledge from hacked data ? Is it ethical ? Is it legal ?
- Is it ethical to reconfigure such data with the help of semantic tools ? What are the risks related to this practice ?
- How should sociologists report their results ? Beyond anonymity, do some results call for inquiries of other kind, such as judicial inquiries ? Should sociologists become inofmers ? How far ?

4.4 Discussion, 16:20 - 16:50

5 Closure of the day, 16:50 - 17:10

6 Keynote 1, 9:00 - 10:00

Chair: Elise Panalva-Icher, Université Paris Dauphine

6.1 José Luis Molina, Universitat Autònoma de Barcelona: 'Hyper-Ethics: A Critical Account', 9:00 - 9:40

Are (social) scientists behaving really that 'badly'? Do they need to be supervised by a growing body of Ethics Committees and its far reaching regulations? Such questions are just rhetorical because the response is as simple as the law must be fulfilled. Instead, the right question seems to be: How did we arrive at this situation? The response is at least threefold: First, some of the stricter regulations deriving from bioethics are applied to social science and humanities as if the type of participation and its risks are even minimally comparable ('informed consent', the right to opt out, etc.); Second, there are objective interests in increasing bureaucracy, namely, professional careers, scarce jobs, and exemption of liabilities; Third, academics, who have been taught to behave ethically and in a disciplined manner, tend to accept without hesitation all kinds of bureaucratic and auditing requests as a further warranty of legitimization. These factors lead to what I have dubbed 'HyperEthics', i.e., hypertrophy of regulatory bodies supervising research-related behaviors. The label is especially ironic in a context where social media sites and telecommunication companies are systematically owning and collecting the majority of data on social interactions nowadays, but are not subject to the same regulations. Often, 'we' (social scientists in a wide sense) want to behave ethically, irrespective of the domains of research, i.e., digital and/or offline. This ethos can only be fulfilled taking the best from regulations, research ethics literature, and IT solutions for privacy, and developing a self-regulative culture, as we have always done. If we do not care about the core values of science, i.e., to advance knowledge and technology for the wellbeing of human kind, who will? Markets? States? I am afraid not. Scientia gratia scientiae.

6.2 Discussion, 9:40 - 10:00

7 Keynote 2, 10:00 - 11:00

Chair: Michel Grossetti, CNRS, Toulouse

7.1 Bernie Hogan, Oxford Internet Institute: 'Indirect ties, indirect consent? An ethical dilemma made worse online', 10:00 - 10:40

When collecting network data, we not only capture data on an individual, but on their relationships to other individuals. In many studies, the other individuals have not given informed consent to the researcher. This is the case in self-reported personal network studies. If informed consent is a precondition for the processing of personal information then virtually all self-reported network studies are in violation of standards of ethical research. One flimsy argument for these studies is that we are only getting details from 'the respondentâĂŹs point of view'. The network analysis of digital trace data undermines this argument as the data was not solely from recall of respondents but from actions of the respondent and their alters.

Perhaps informed consent, on its own, is not the right approach for evaluating the legitimacy of social network studies? Drawing upon work in other domains of research I explore the issue of indirect or implied consent. We

rarely think of indirect consent as being appropriate for personal information. Nevertheless, we routinely perform qualitative interviews where respondents denote names of others, albeit not in as systematic a fashion as in network studies. Much of our behaviour online would not make sense if we had to guard our identity from our established relationships.
In this paper I contend that we refer to a principle of harm minimization first and foremost. Instead of establishing deontological principles of indirect consent, we look to phenomenological and pragmatic approaches towards what names and identities mean in the context of establishing relationships. A pragmatic approach, while always provisional, helps to establish a basis for understanding how individuals make sense of their own needs to identify others. From this own need to identify others we can establish some boundaries for indirect consent while still making sense of social relationships in a way that minimizes harm and respects individual research subjects.

7.2 Discussion, 10:40 - 11:00

8 Panel 1: Researchers' perspectives, 11:15 - 12:00

Chair: Guillaume Favre, Université de Toulouse Jean Jaurès

8.1 Insight from Day 1, 11:15 - 11:45

This panel summarizes and presents key reflections and insights from paper presentations and discussions on Day 1. Panelists are:

- Mandy Lee, University College Dublin
- Louise Ryan, University of Sheffield

8.2 Discussion with the audience, 11:45 - 12:00

9 Panel 2: Policy perspectives, 12:00 - 13:00

Chair: Antonio A. Casilli, Telecom ParisTech

9.1 Panel on policy and regulation issues, 12:00 - 12:45

This panel brings the discussion to the policy and regulation levels. It bridges the viewpoints and experiences of researchers with broader issues such as orientations in public research policies; the new European GDRP (General Data Protection Regulation); the power of private digital companies and their, sometimes conflictual, relationships with public authorities; reconciling the drive for technological innovation with lawfulness and deontology; data access for (public) research; integrity of research and the establishment of codes of conduct. Panelists are:

- Christine Balagué, Professor at Institut Mines Telecom, former Vice-President of the French Digital National Council (Conseil National du Numérique, CNNum) and currently member of CERNA (Comité d'Ethique de la Recherche sur le Numérique d'Allistène), an ethics committee for digital research;
- Geoffrey Delcroix, Innovation & Foresight Project Manager at CNIL (Commission Nationale de l'Informatique et des Libertés), the French data protection authority;
- Rémy Mosseri, Research Professor at CNRS and Université Pierre & Marie Curie in Paris, and member of COMETS, the research ethics committee of CNRS.

9.2 Discussion with the audience, 12:45 - 13:00