'Digital Wildfires': a challenge to the governance of social media?

Conference Paper · July 2015

9 authors, including:

Helena Webb
University of Oxford
53 PUBLICATIONS   **158** CITATIONS

SEE PROFILE

Rob N Procter
The University of Warwick
327 PUBLICATIONS   **5,295** CITATIONS

SEE PROFILE

Marina Jirotka
University of Oxford
99 PUBLICATIONS   **1,445** CITATIONS

SEE PROFILE

Bernd Carsten Stahl
De Montfort University
233 PUBLICATIONS   **1,964** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Alan Turing Institute View project

Project   Dependability Interdisciplinary Research Collaboration (DIRC) View project

'Digital Wildfires': a challenge to the governance of social media?

Helena Webb
Marina Jirotka
University of Oxford, Department of
Computer Science
Oxford, United Kingdom
helena.webb@cs.ox.ac.uk
marina.jirotka.cs.ox.ac.uk

Bernd Carsten Stahl
De Montfort University, Department of
Informatics
Leicester, United Kingdom
bstahl@dmu.ac.uk

William Housley
Adam Edwards
Matthew Williams
Cardiff University, School of Social
Sciences
Cardiff, United Kingdom
housleyw@cardiff.ac.uk
edwardsa2@cardiff.ac.uk
williamsm7@cardiff.ac.uk

Rob Procter
University of Warwick, Department of
Computer Science
Coventry, United Kingdom
rob.procter@warwick.ac.uk

Omer Rana
Pete Burnap
Cardiff University, School of Computer
Science and Informatics
Cardiff, United Kingdom
ranaof@cardiff.ac.uk
p.burnap@cs.cardiff.ac.uk

ABSTRACT
The increasing popularity of social media platforms such as Facebook, Twitter, Instagram and Tumblr has been accompanied by concerns over the growing prevalence of 'harmful' online interactions. The term 'digital wildfire' has been coined to characterise the capacity for provocative content on social media to propagate rapidly and cause offline harm. The apparent risks posed by digital wildfires create questions over the suitable governance of digital social spaces. This paper provides an overview of some preliminary findings of an ongoing research project that seeks to build an empirically grounded methodology for the study and advancement of the responsible governance of social media.

Categories and Subject Descriptors
K.4 [**Computers and Society**]: Public and Policy Issues – *abuse and crime involving computers, ethics, regulation.*

General Terms
Management, Human Factors, Legal Aspects.

Keywords
Social media, governance, responsible research and innovation.

1. INTRODUCTION
The increasing popularity of social media platforms such as Facebook, Twitter, Instagram and Tumblr has been accompanied by concerns over the growing prevalence of 'harmful' online interactions. Since social media users can share and forward

content instantly to multiple others, behaviours such as trolling, harassment and abuse, the spread of rumour etc. can also propagate rapidly, with the potential to cause offline harm. The apparent risks posed by social media communications are characterised in the 2013 World Economic Forum (WEF) report "Digital wildfires in a hyperconnected world" [1]. The report argues that the modern status of hyperconnectivity sets up the global risk of 'digital wildfires', in which provocative content of different kinds spreads rapidly and on a massive scale to 'create havoc in the real world'. For instance, misleading content can cause harm to the reputation of individuals, organisations and markets before there is a chance to correct it. Content containing hate speech or incitement to violence etc. can also cause considerable harm to individuals and generate social tension. The WEF report calls for a 'global digital ethos' to regulate digital social spaces and prevent, govern and manage digital wildfire scenarios.

This paper forms part of a wider project that explores the issues raised in the WEF report, concerning the spread of provocative content on social media, its offline impacts and the governance of digital social spaces. Here we describe some preliminary project work which has scoped existing relevant governance frameworks in the UK context.

2. THE GOVERNANCE OF DIGITAL SOCIAL SPACES
This section provides an overview of a scoping exercise conducted to explore mechanisms relevant to the regulation of digital social spaces and potential digital wildfire scenarios. In the UK context the key mechanisms are:

2.1 Legal governance
In England and Wales existing civil and criminal legal codes can be applied to pursue and penalise individuals who have posted certain kinds of provocative content. One specific case reported in the WEF report is the Sally Bercow defamation trial [2]. In a typical digital wildfire scenario, content is posted and spread by multiple

users. However only a small number of posts are likely to be reported to the police and even fewer will be pursued in the courts.

2.2 Social media governance

Social media platforms typically require users to agree to Terms of Use in order to post content and can apply penalties when these are breached. Automated processes may be used to block certain forms of illegal content but more often the platforms rely on user reports to deal with content retrospectively. Some platforms – for instance Twitter [3] – have become stricter over what kinds of posts are allowable in the face of public criticism. However they tend to be run on principles that support freedom of expression and are reluctant to (pre)monitor content.

2.3 Institutional social media policies

As social media platforms have gained popularity, organisations of all kinds have begun to enforce policies regarding their appropriate use. Social media policies are now commonplace in workplaces, schools and law courts and allow a range of penalties for misuse.

2.4 User self-regulation

Social media users can regulate their own and others' online behaviours. In addition to reporting posts, users can counter provocative content through actions such as: challenging the accuracy and/or appropriateness of posts; correcting false information; urging posters to remove their own content; mocking or shaming posts to minimise their value; ignoring provocative posts and posters to slow the spread of content.

Existing governance mechanisms for social media are often retrospective. Legal, social media and institutional mechanisms tend to deal with content after it has spread and had an impact. Beyond the deterrent effect of penalties they do little to prevent content propagating. They also deal with individual users rather than the multiple users who may be involved in a digital wildfire. By contrast user regulation has a real time element and may have the capacity to prevent or limit the spread of provocative content.

If we accept digital wildfires as a global risk factor, we might also ask whether it is necessary to introduce new governance mechanisms to overcome the limitations of existing ones. Proactive mechanisms could attempt to slow the spread of posts – for instance through the creation of a waiting time to forward rapidly spreading content. Alternatively self-governance could be further supported through the provision of visible esteem for users who intervene to counter inappropriate content, or the provision of a 'lie' button to allow users to indicate that the content of a post is not creditworthy.

3. JUSTIFICATION OF GOVERNANCE

Insights from computer ethics and responsible research and innovation [4] illustrate the importance of ethical justifications for the governance of digital social spaces. Calls for further governance relevant to digital wildfires are based on the assumption that the massive spread of provocative content is harmful and that limiting it is desirable. However it is possible some governance mechanisms – such as 'shaming' or blocking freedom of speech - may produce more harm than the content itself. Provocative content can sometimes be truthful and desirable, or open to interpretation. These issues are in part normative but can also be informed by empirical insights – as discussed next.

4. DISCUSSION: EXPLORING DIGITAL WILDFIRES

This paper reports on a scoping exercise to explore governance mechanisms in relation to 'digital wildfires'- the massive spread of provocative content on social media with serious offline impacts. In the UK context legal, social media and institutional mechanisms have the capacity to deal with some forms of provocative content on an individual and retrospective basis, whilst user self-regulation may be able to limit the real time propagation of posts. It may be possible for further governance mechanisms to halt or slow digital wildfires but these raise important ethical issues. This scoping of governance mechanisms poses questions that require empirical investigation. How do existing social media governance mechanisms operate in real time digital wildfire scenarios? What kinds of harm do digital wildfires inflict and do they create any benefits? Are the harms serious enough to warrant new mechanisms that may limit freedom of speech?

The "Digital Wildfire: (Mis)information flows, propagation and responsible governance" project [5] conducted by the authors of this paper takes up these research questions. This is an interdisciplinary study with the overall aim to build an empirically grounded methodology for the study and advancement of the responsible governance of social media. The study draws on the WEF report to explore the possible existence of digital wildfires, their offline impacts and the regulatory challenges they present. We combine approaches from computer science and the social sciences to examine information flows on social media and the occurrence of self-regulatory behavior such as counter speech. We also observe governance practices in action and solicit the informed opinion of relevant experts on the appropriate regulation of digital social spaces. This empirical work will be combined with the scoping work overviewed above to produce an ethical security map. This will be a practical tool to help various users navigate through social media policy and aid decision making.

5. REFERENCES

[1] World Economic Forum, 2013 *Digital Wildfires in a hyperconnected world*. Global Risks Report. World Economic Forum (Feb. 2013), DOI= http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/

[2] Lord McAlpine of West Green v Sally Bercow 2013 EWHC 1342 (QB). DOI= https://www.judiciary.gov.uk/judgments/mcalpine-bercow-judgment-24052013/

[3] Doshi, S. 2015. *Policy and product updates aimed at combating abuse*. Blog.twitter.com. DOI= https://blog.twitter.com/2015/policy-and-product-updates-aimed-at-combating-abuse

[4] For example see Carsten Stahl, B., Eden, G., Jirotka, M. and Coeckelbergh, M.. 2014. From Computer Ethics to Responsible Research and Innovation in ICT: The transition of reference discourses informing ethics-related research in information systems. *Information & Management*. 51, 6 (Sep 2014) 810-818. DOI= http://www.sciencedirect.com/science/article/pii/S037872061400007X

[5] For further information see: http://www.digitalwildfire.org/