Digital Wildfires, hyper connectivity and the responsible governance of social media

Helena Webb, June 2018

Digital Wildfires = *the rapid viral spread of information that is either intentionally or unintentionally misleading or provocative with serious consequences* (World Economic Forum report 2013 "Digital Wildfires in a hyperconnected world")

UNIVERSITY OF OXFORD

Digital wildfires as a challenge to the governance of social media

"***Establishing reasonable limits to legal freedoms of online speech is difficult because social media is a recent phenomenon,*** *and digital social norms are not yet well established. The question raises thorny issues of the extent to which it would be possible to impose limits on the ability to maintain online anonymity, without seriously compromising* **the usefulness of the Internet as a tool for whistle-blowers and political dissidents in repressive regimes**.*"*

WEF, 2013: 'Digital Wildfires in a hyperconnected world'

UNIVERSITY OF OXFORD

"Digital Wildfire: (mis)information flows, propagation and responsible governance" project
November 2014-November 2016

Marina Jirotka
Rob Procter
Bernd Stahl
William Housley
Omer Rana
Adam Edwards
Matt Williams
Pete Burnap
Helena Webb

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYdd

UNIVERSITY OF OXFORD

DE MONTFORT UNIVERSITY LEICESTER

THE UNIVERSITY OF WARWICK

http://www.digitalwildfire.org/

UNIVERSITY OF OXFORD

RESEARCH AIM
To build an empirically grounded methodology for the study and advancement of the responsible governance of social media

RESEARCH QUESTION
Does the risk of digital wildfires necessitate new forms of social media governance?

UNIVERSITY OF OXFORD

WP1**:** Scoping ethical and regulatory issues in relation to digital wildfires. 4 key governance mechanisms: legal, social media platform, institutional and user self-governance.

WP2: Examination of social media content. *Qualitative (interaction analysis of social media data) and quantitative computational analysis. Inspection of self-governance in action*

WP3: Policy Delphi. Seeking *the informed opinion of individuals on the governance of social media* (social media platforms, users, institutions, lawyers)

WP4: Governance practices observed. *Fieldwork at social media platforms, schools, police, policy institutions, civil rights groups etc.*

Key project findings

1) Scale and breadth of the 'problem'
2) The complexities and limitations of current governance
3) The potential value of counter speech and user self-governance
4) The value of education and engagement

1. Scale and breadth of the 'problem'

"Numerous community groups raise concerns about social media and about material available on social media …it's a massively increasing area." *(Staff member at Ministry of Justice)*

"People sometimes, I don't think, do the research because they see a tweet, similar to what Google may be I think sometimes …someone will tweet it and that tweet becomes the truth". *(Council worker)*

"[Sexting] is certainly happening, yes... We're constantly talking about the fact that you lose control as soon as you send it.." *(Secondary school teacher)*

"70% of our resources go on online cases. 70% of our cases go on chasing the nonsense day in, day out. I have temp staff and it costs £400,000 a year." *(Staff member at anti hate crime organisation)*

2. The complexities and limitations of current governance

UNIVERSITY OF OXFORD

- Legal governance

- Social media platform governance mechanisms

- Institutional social media policies

- User self-governance

Crime

Two jailed for Twitter abuse of feminist campaigner

Isabella Sorley receives 12-week sentence and John Nimmo eight weeks for 'extreme threats' against Caroline Criado-Perez

facebook

Facebook Terms and Policies

Nottingham University lecturer forced to apologise after branding students 'idiots' and 'semi-literate' on Facebook

- As a result of the comments, students' essays will have to be re-marked
- On January 14, Anthony Fisher also made light of a student's panic attacks
- In a bizarre lecture on Tuesday, Fisher apologised to his students
- The sociolinguistics lecturer also lashed out at Nottingham University's HR department

By DAILY MAIL REPORTER

PUBLISHED: 17:57, 7 February 2014 | UPDATED: 11:45, 7 February 2014

UNIVERSITY OF OXFORD

"But that's also because from my experience when we've reported something to the Police sometimes they take it seriously and sometimes they don't. I think they struggle as much as I do to know is it against the law or not to be honest. No, I think it's a very grey area, very difficult." *(Staff member at equality organisation)*

"We would report it to Twitter to try to take [racist posts] down and again, the majority of the cases don't get taken down…It is a branding and a numbers game for Twitter." (*Staff member at anti hate crime organisation*)

"If you attempt to police every single transmission on social media you're setting yourself up for quite a large task, … there are millions of communications every hour, let alone every day, so, realistically, you would have an entirely separate service and police service and prosecution service devoted to that one task, if that was what you wanted to do to vet every tweet, every message" *(Staff member at CPS)*

UNIVERSITY OF OXFORD

- Legal governance
- Social media platform governance mechanisms
- Institutional social media policies

- Disruption
- User self governance

limitations of application

tend to focus on individual posts/users rather than the multiple users spreading content

tend to be retrospective—deal with posts/ users after harm has been caused

Have some capacity to be prospective and limit the spread of harmful content in real time

---

3. Counter speech and user self-governance

UNIVERSITY OF OXFORD

"Social media is a place where everyone is entitled to their opinions but everyone is responsible for that opinion as well. Therefore we should all be able to self-regulate and be responsible for what we say online and be able to defend, question or even apologise for words said" *(Delphi panel survey respondent)*

"Some people on social media are 'professional users'. They have the skills to —[intervene in inflammatory discussions] in an effective way and also the armour to protect themselves from harm." *(Staff member at anti hate crime organisation)*

"The best response to bad speech is more speech...Online communities have a right to ostracize, call out, criticize bad actors. Such mechanisms alone are probably not going to stomp out every digital wildfire, but neither have we managed to stomp out every wildfire in real life." *(Delphi panel survey respondent)*

Analysis of 'cyber hate' on Twitter

Computational analysis of Twitter threads

collation of conversational 'threads' on Twitter	development of classifier tools to identify rumour/provocative posts and counter speech
Interaction feature identification - characteristics of posts and responses	disagreement with hate speech tends to extend thread length; however as the number of users disagreeing rises, the length of the thread goes down.
possibility of responses as counter speech	**Multiple voices of disagreement work to quell thread length.**
panel annotation task	Disagreement can therefore function as a form of counter speech – could it be promoted through further mechanisms?

4. The value of education and engagement

UNIVERSITY OF OXFORD

"Work with the communities and see how the communities can help without the police being involved, or the criminal justice system, but then we need to work with the communities to get those outcomes" *(Staff member at equality organisation)*

"We would run counter-speech projects online. We would run projects that actually would provide engagement for community groups to come together".
(Staff member at anti hate crime organisation)

"One of the best ways to address 'responsible' behaviour on social media is through education and social media literacy."
(Staff member at online platform)

UNIVERSITY OF OXFORD

#TakeCareOfYourDigitalSelf

Youth panel competitions

Teaching and learning materials tes.com

Further projects

- Impact Acceleration Project and British Academy Small grant project – co-construction and delivery of e-safety teaching materials

- PHEME – verification of online content
  https://www.pheme.eu/

- UnBias – user experience of algorithm driven internet platforms
  https://unbias.wp.horizon.ac.uk/

Digital Wildfires, hyper connectivity and the responsible governance of social media

helena.webb@cs.ox.ac.uk