

NatCen

Social Research that works for society

Linking Survey & Social Media Data

A Twitter Case Study

25/06/1019



Contents

1. Background
2. Case Study – 2017 General Election
3. Practicalities & ethics
4. Looking forward

Background



1

What do social media data offer?

'The 6 Vs'

Volume

Variety

Velocity

Veracity

Virtue

Value

Why link social media data and survey data?

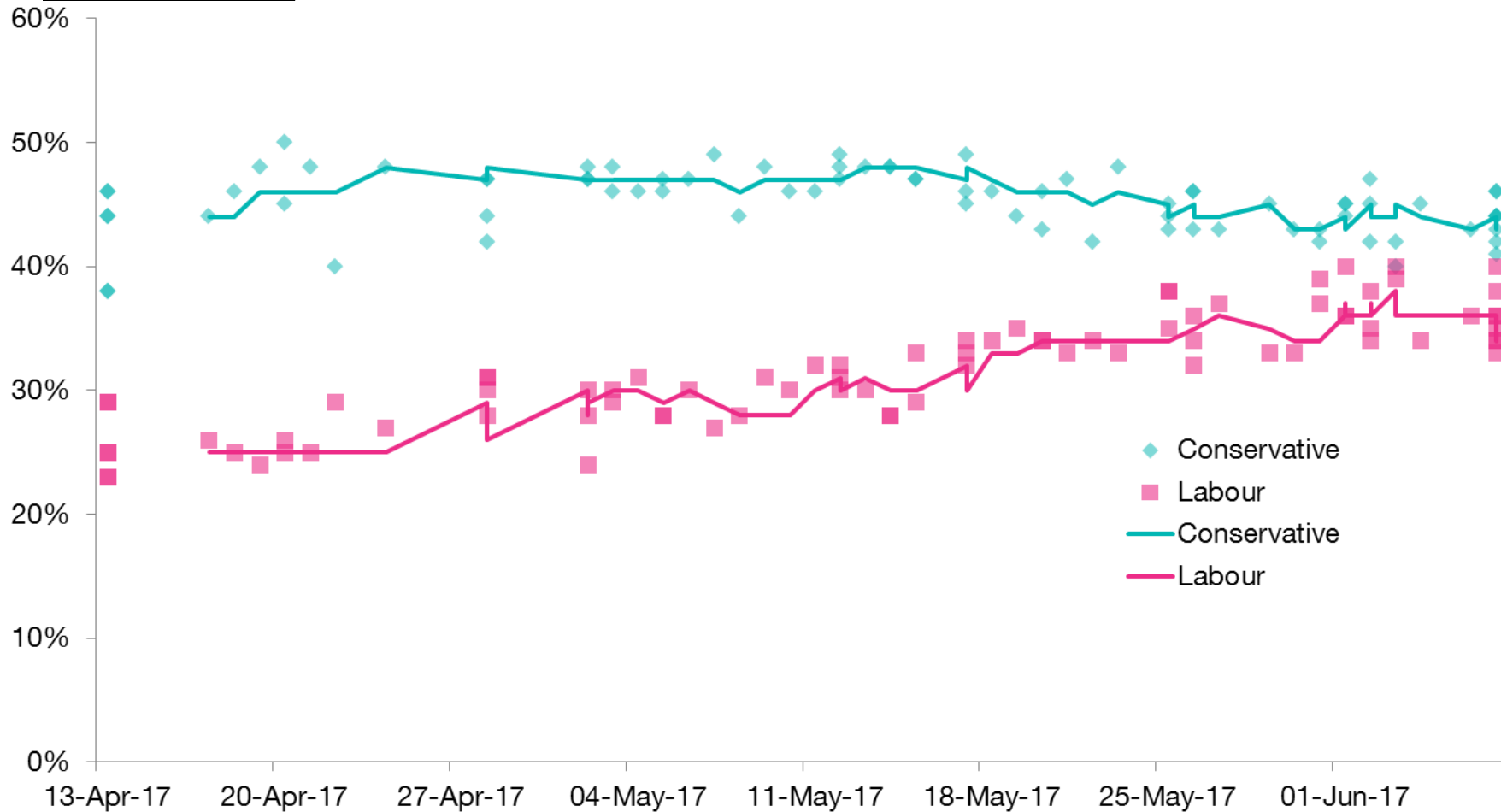
- Continued & increasing interest in using social media data for social research
- But they continue to have their drawbacks
- Linking survey & social media data attempts to address some of these by:
 - Collecting informed consent
 - Putting sample in context of the population
 - Understanding whose data you are analysing
 - Validating machine-based classifications

Case Study – Survey & Twitter data



2

GE2017 – Understanding the rise of Labour



Survey data can tell us a lot about who...

“Labour was relatively successful at winning the support of those who did not vote in 2015”

“Labour’s advance was strongest amongst those who were keenest on staying in the EU and those who were least concerned about immigration”

... but social data can provide more 'texture'



- What were people talking about in the run-up to the general election?
- Where were people getting their information from?
- How were people talking about Corbyn/May; Labour/Conservatives?

Enhancing our understanding of the election on Twitter



- Restrict Twitter sample only include the population of interest (GB, 18+)
- Cover the ‘Twitter population’, not the ‘Tweeting population’
- Put findings in context: how the Twitter population differs from the general population
- Match on characteristics to understand how behaviour varies between voter groups – topics discussed, content shared, networks interact with

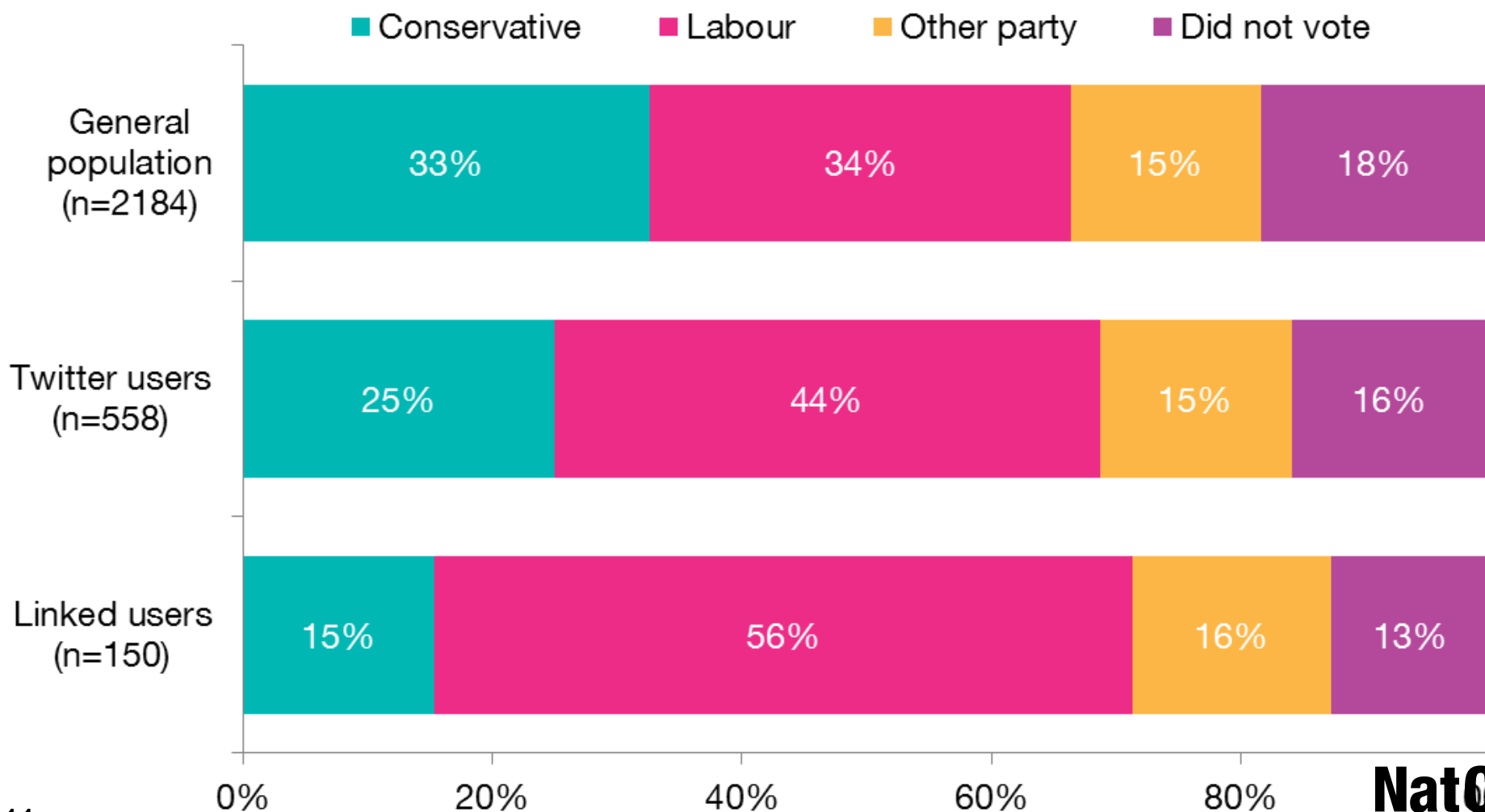


Data collection

- Survey data collected using NatCen Probability Panel in July 2017 (n = 2184)
- Range of questions:
 - Voting behaviour and social, economic, & political attitudes
 - Consent to survey data to Twitter account
- Twitter handles for 150 who agreed to linkage passed to CASM to collect data using Method52
- 7,555 Tweets sent between 17th April & 14th June



Putting Twitter data in context



What were people talking about?

Parties or politicians

“@jeremycorbyn doesn’t avoid voters or hire a crowd, he reads out questions from real people #PMQs”

Policy

“Labour will build affordable homes and crack down on rogue private landlords”

Pro-Labour or Anti-Conservative?

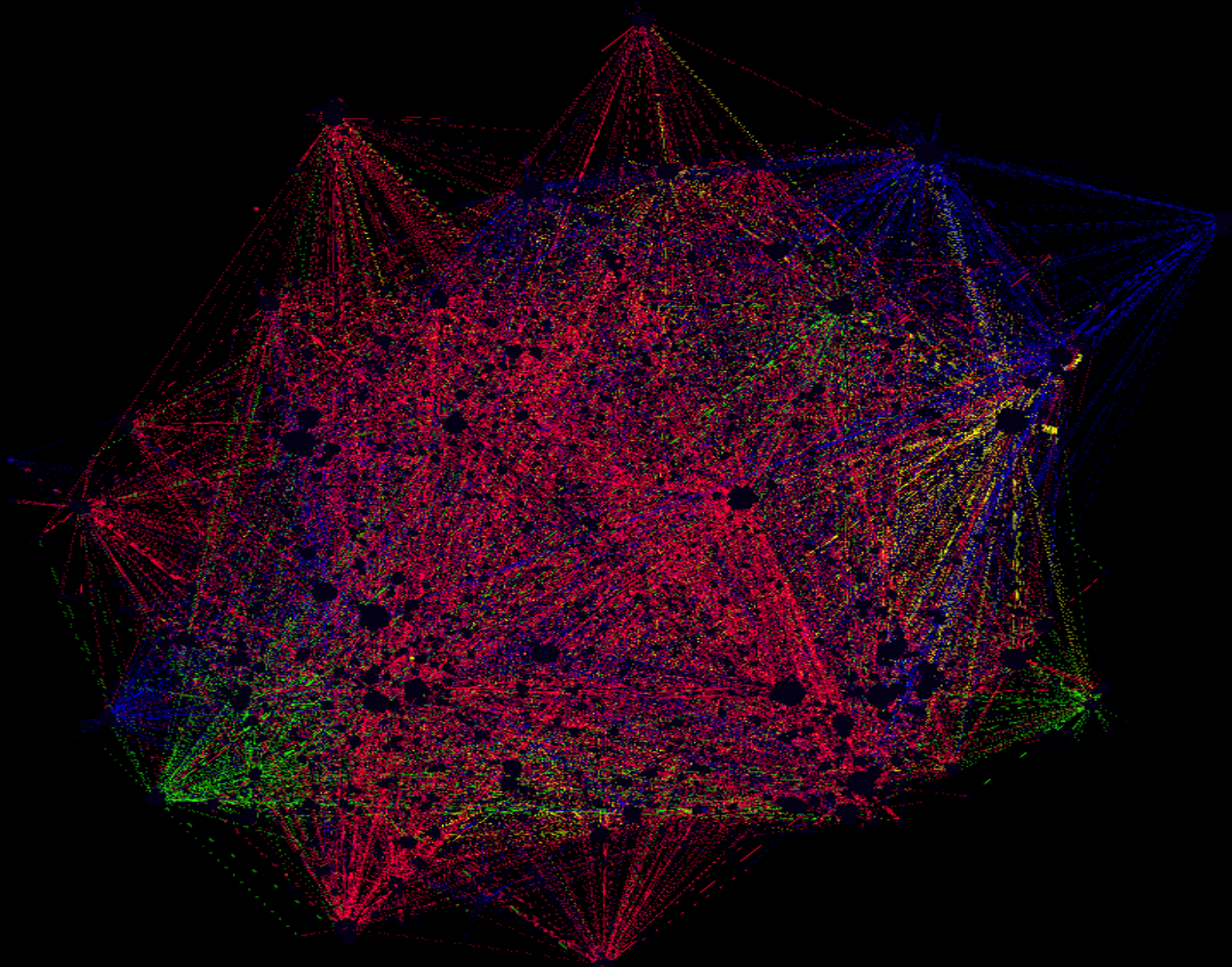
Pro-Labour

“Labour: minimum wage, school cuts reversed, universal child care. Tories: kill the foxes!”

Anti-Conservative

“Trying to humanise May just shows us she’s personally awful as well as politically vile”

Do different groups interact?



Ethics & practicalities

3

Consent

- Need to ask consent to access & link Twitter data
 - We have the opportunity
 - Linking data creates additional risk
- Balance level of detail:
 - Enough detail that consent is informed
 - Not so much that people are overwhelmed & don't read/understand it

Consent – Question wording

As social media plays an increasing role in society, we would like to know who uses Twitter, and how people use it. We are also interested in being able to add people's, and specifically your, answers to this survey to publicly available information from your Twitter account such as your profile information, tweets in the past and in future, and information about how you use your account.

Your Twitter information will be treated as confidential and given the same protections as your interview data. Your Twitter username, and any information that would allow you to be identified, will not be published without your explicit permission.

Consent – Helplinks

What information will you collect from my Twitter account?

What will the information be used for?

Who will be able to access the information?

What will you do to keep my information safe?

What if I change my mind?

Consent rates

	Web	CAT/F2F	Overall
BSA 2015	-	37%	37%
NatCen Panel	26%	34%	27%
USoc IP10	24%	41%	31%

Maintaining security

- Cannot rely on 'anonymisation' to maintain data security
- But there are other approaches to help minimise risks:
 - Systematic processing
 - Data reduction
 - Controlled access
 - Data deletion



Looking forward

- Linking is feasible, but key challenge of low consent rates
 - Understand why & how to address
- Next step of demonstrating value in applied setting
 - Practical applications for understanding online & offline society
- Infrastructure for archiving and sharing Social Media data
 - Secure access
- Expansion to other platforms; is Twitter 'the future'?
 - The 'digital eco-system'
 - New & changing digital data forms

Contact

Curtis Jessop

Research Director

curtis.jessop@natcen.ac.uk

NatCen

Social Research that works for society