



NatCen

Social Research that works for society

Understanding political behaviour by linking survey & social media data





Acknowledgments

- Josh Smith & Christopher Lambin at Centre for the Analysis of Social Media (CASM)
- Martin Wood at NatCen Social Research

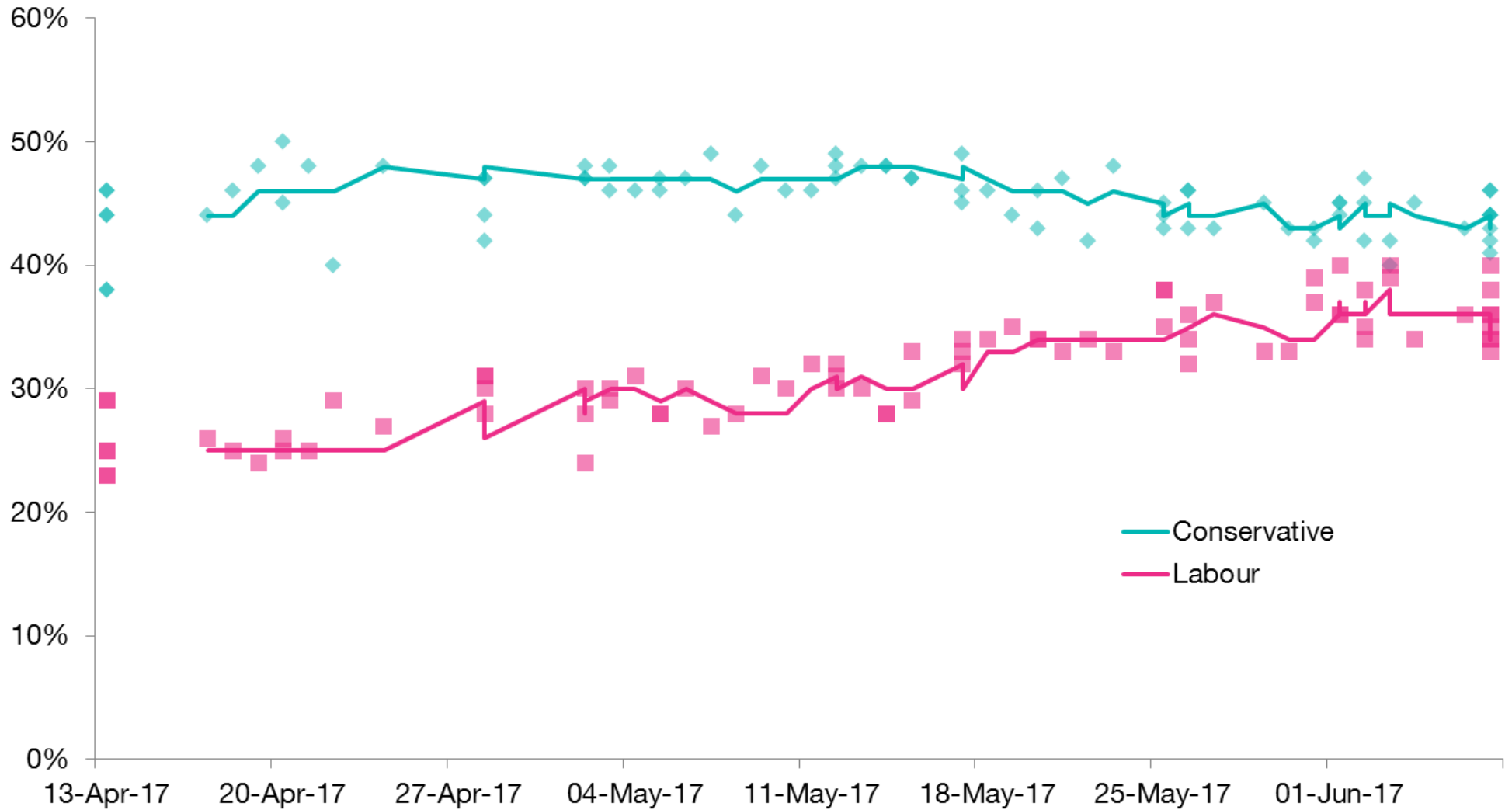
Adding the 'who' to social media data



- 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



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

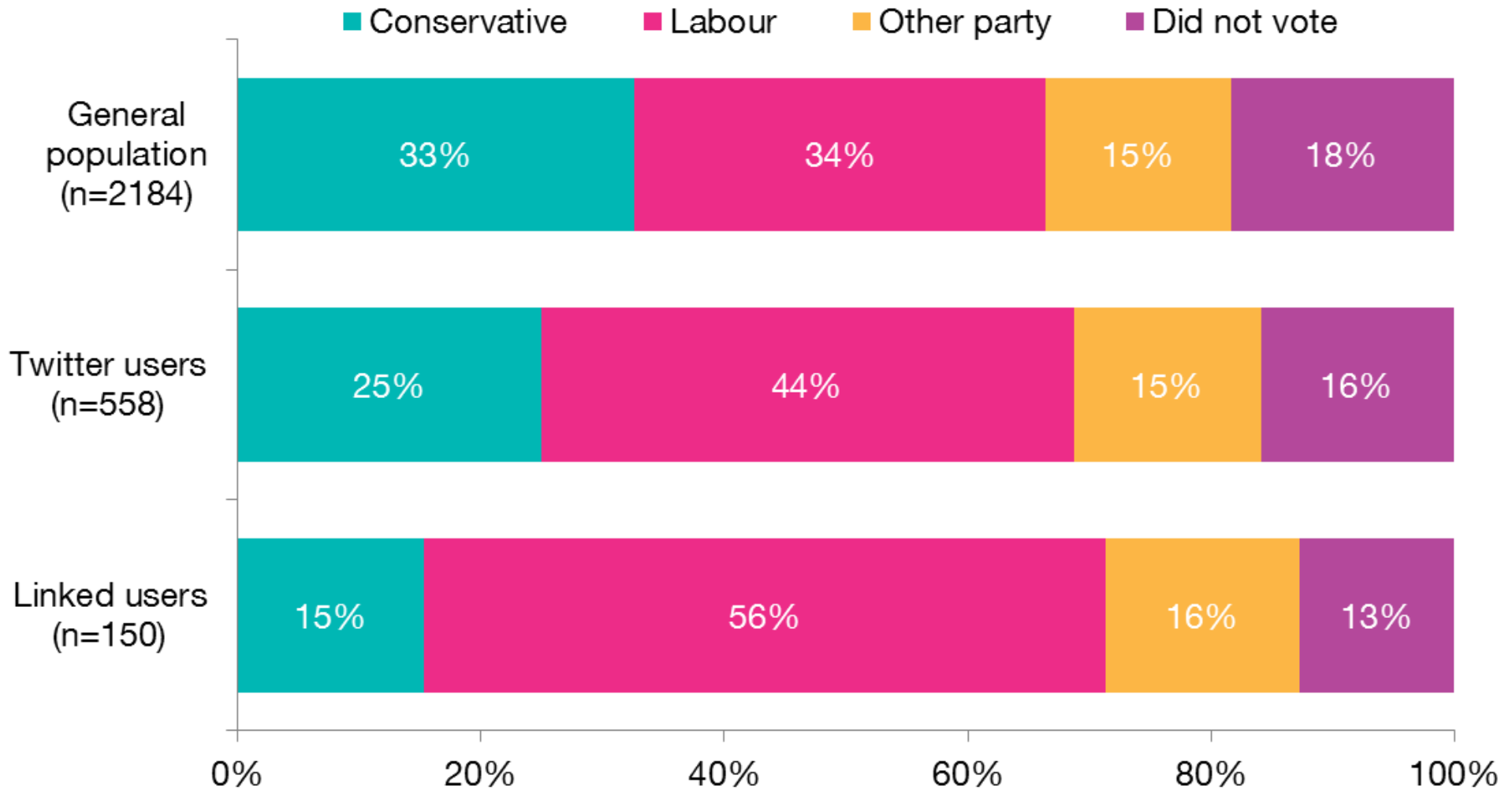


Data collection

- 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?

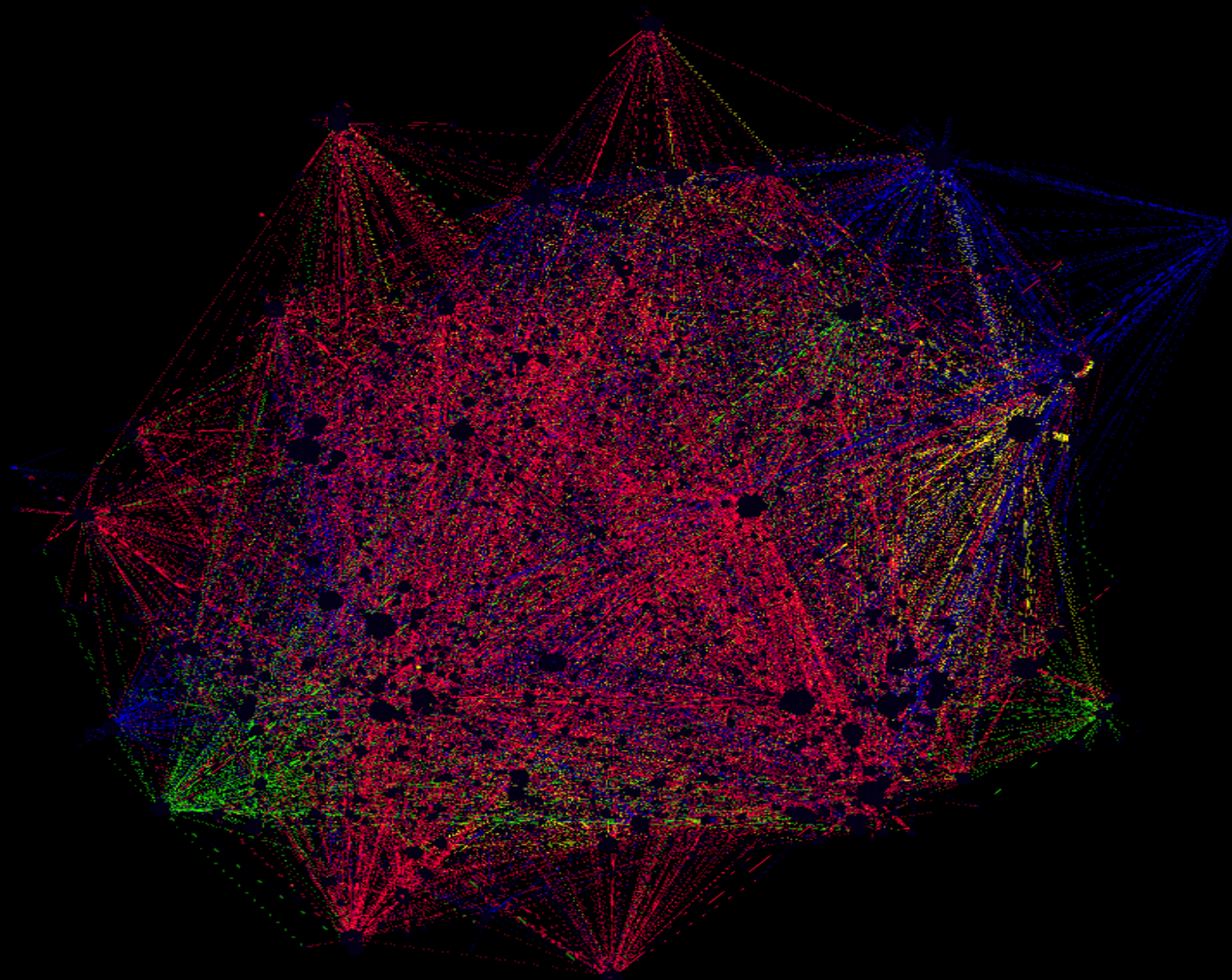
Pro-Labour

“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?





Discussion points

- How can we boost sample sizes?
- Tweets or people?
- Quantitative or qualitative analysis?
- Should we weight results?
- Using survey data for validation?
- How can we archive this data for others to use?