

A FIELD
GUIDE TO

"FAKE NEWS" AND OTHER INFORMATION DISORDERS

A COLLECTION OF RECIPES
FOR THOSE WHO LOVE TO COOK
WITH DIGITAL METHODS

COMPILED BY

Liliana Bounegru
Jonathan Gray
Tommaso Venturini
Michele Mauri



Public
Data
Lab

FIRST
DRAFT

DEN-
SITY
GEN+

digital
methods
initiative

inria



CJS

UNIVERSITÄT
SIEGEN

Unil
UNIL Université de Lausanne
Institut des Sciences sociales
STSLab & LaDHUL

LISIS

SciencesPo
MEDIALAB

Citizen Data Lab



A FIELD GUIDE TO "FAKE NEWS" AND OTHER INFORMATION DISORDERS

Compiled by Liliana Bounegru, Jonathan Gray, Tommaso Venturini and Michele Mauri.

This guide explores the use of digital methods to study false viral news, political memes, trolling practices and their social life online. It is a project of the **Public Data Lab** with support from **First Draft**.

The Public Data Lab (publicdatalab.org) is an interdisciplinary network seeking to facilitate research, democratic engagement and public debate around the future of the data society.

First Draft (firstdraftnews.com) is dedicated to improving skills and standards in the reporting and sharing of information that emerges online.

© 2017 Public Data Lab. Amsterdam.

The guide is released under the Creative Commons Attribution License (creativecommons.org/licenses/by/4.0/). This means you can freely share, adapt and draw on this work as long as you give credit, as per the terms of the license.

If you reproduce or draw on material from this guide, we'd be grateful if you could credit and link back as per the following statement:

"This article draws on *A Field Guide to "Fake News" and Other Information Disorders*, a collaboration of the Public Data Lab and First Draft. For further details see: <http://fakenews.publicdatalab.org>"

The copyright for the research that this guide draws upon remains with its respective contributors.

This project would not have been possible without the contributions of researchers from the following institutions:

Centre for Journalism Studies, University of Ghent (NL)
Citizen Data Lab, Amsterdam University of Applied Sciences (NL)
DensityDesign Lab, Politecnico di Milano (IT)
Digital Methods Initiative, Media Studies, University of Amsterdam (NL)
Govcom.org Foundation, Amsterdam (NL)
Institut National de Recherche en Informatique et en Automatique (INRIA) (FR)
King's College London (UK)
Laboratoire d'Étude des Sciences et des Techniques (STS-Lab), Université de Lausanne (CH)
Laboratoire Interdisciplinaire Sciences Innovations Sociétés (LISIS), Université Paris-Est Marne-la-Vallée (FR)
Médialab, Sciences Po, Paris (FR)
Techno-Anthropology Lab, Aalborg University Copenhagen (DK)
University of Siegen (DE)

Published by Public Data Lab.
Released in December 2017.

A FIELD
GUIDE TO

"FAKE NEWS" AND OTHER INFORMATION DISORDERS

A COLLECTION OF RECIPES
FOR THOSE WHO LOVE TO COOK
WITH DIGITAL METHODS



Public
Data
Lab

FIRST
DRAFT

DEN-
SITY
GEN+

digital
methods
initiative

Inria



KING'S
COLLEGE
LONDON



IPR
Institute for
Party Research



UNIVERSITY OF
BATH

UNIVERSITY OF
BATH

UNIVERSITY OF
BATH



CJS



UNIVERSITÄT
SIEGEN



UNIL | Université de Lausanne

Institut des sciences sociales
STSLab & LaDHUL



LISIS



Sciences Po
MEDIHLAB



Citizen Data Lab



TANTLAB.AU.UK

CONTENTS

| | |
|------------------|---|
| 7 | Introduction |
| 15 | Conventions used in the book |
| <i>CHAPTER 1</i> | |
| 17 | <i>MAPPING FAKE NEWS HOTSPOTS ON FACEBOOK</i> |
| 20 | 1. What publics does fake news animate on Facebook? |
| 40 | 2. How may the trajectory of a fake news story be traced on Facebook? |
| 52 | 3. Do fact-checking initiatives reach the publics of fake news on Facebook? |
| <i>CHAPTER 2</i> | |
| 61 | <i>TRACING THE CIRCULATION OF FAKE NEWS ON THE WEB</i> |
| 64 | 1. Where do fake news originate? By what sites are they first retransmitted? |
| 78 | 2. Which are the most visible sources related to a fake story? When and by whom are they mentioned? |
| <i>CHAPTER 3</i> | |
| 95 | <i>USING TRACKER SIGNATURES TO MAP THE TECHNO-COMMERCIAL UNDERPINNINGS OF FAKE NEWS SITES</i> |
| 97 | 1. Do fake news sites use different kinds of trackers from mainstream media sites? |
| 103 | 2. How can fake news and mainstream media sites be profiled based on their tracker usage? |
| 109 | 3. How do tracker ecologies on fake news sites change over time? |
| 115 | 4. Which other websites share the same tracker IDs as fake news sites? |
| 121 | 5. Do trackers associated with hyper-partisan, and misleading information sites vary across language spheres? |

CHAPTER 4

**127 *STUDYING POLITICAL MEMES
ON FACEBOOK***

- 130 1. How can meme spaces on Facebook be traced?
- 139 2. How do memes frame political and media events?
- 148 3. How may the content of memes be studied?

CHAPTER 5

**161 *MAPPING TROLL-LIKE PRACTICES
ON TWITTER***

- 165 1. How may we detect Twitter accounts which negatively target political representatives?
 - 172 2. How may we characterise the sources of troll-like activity?
 - 184 3. How may troll-like practices be characterised?
- 199 Conclusion
- 203 Glossaries
- 210 Contributors and acknowledgements

A FIELD GUIDE TO "FAKE NEWS" AND OTHER INFORMATION DISORDERS

INTRODUCTION

[1] See, for example, Robert Darnton, "The True History of Fake News", *The New York Review of Books*, February 2017, available at: <http://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/>

[2] See, for example, "Sky Views: Facebook's fake news threatens democracy", *Sky News*: <http://news.sky.com/story/sky-views-democracy-burns-as-facebook-lets-fake-news-thrive-10652711>

[3] See Edson Tandoc, Zheng Wei Lim & Richard Ling "Defining 'Fake News'", *Digital Journalism*, 8(1), August 2017. pp. 1-17. 2017. Available at: <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1360143>

[4] See Claire Wardle & Dera Khshan Hossein, *Information Disorder: Toward an interdisciplinary framework for research and policy-making (Report to the Council of Europe)*, 2017.

What is "fake news"? And what can be done about it? Depending on who you ask, fake news is said to represent a step-change in information warfare; an emerging form of cynical profiteering; an engine for energising "alt-right" and other digitally mediated grassroots political mobilisations around the world; a partisan battle cry for a new liberal "ministry of truth"; an unwanted byproduct of the online platforms which organise our digital societies; or a canary call signalling a collapse of consensus around established institutions and processes of knowledge production, heralding a new "post-truth" era in politics and public life.

According to some commentators fake news is just old wine in new bottles – and similar misinformation phenomena have existed for at least as long as the printing press and other communication technologies through which they circulate.^[1] Others suggest that new online platforms accelerate and "supercharge" their circulation in a way which introduces hitherto unprecedented challenges and dynamics.^[2] Others even claim that the term "fake news" should be avoided altogether because it is too vague^[3], politically dangerous^[4]; indistinguishable from past forms of disinformation^[5]; charged with an over-simplistic idea of truth as direct correspondence to reality^[6], and missing the

- [5] See Caroline Jack, What's Propaganda Got To Do With It? Available at: <https://points.datasociety.net/whats-propaganda-got-to-do-with-it-5b88d78c3282>
- [6] See Michael Lynch, Post-truth, alt-facts, and asymmetric controversies, 2017, First 100 Days, Available at: <http://first100days.stsprogram.org/2017/02/06/post-truth-alt-facts-and-asymmetric-controversies-part-i/>
- [7] See Henry Jenkins, Sam Ford, & Joshua Benjamin Green, Spreadable media. New York: New York University Press, 2017, Available at: <http://doi.org/10.1017/CBO9781107415324.004>
- [8] See, John Naughton, "Facebook and Twitter could pay the price for hate speech", The Guardian, March 2017: <https://www.theguardian.com/commentisfree/2017/mar/19/john-naughton-germany-fine-social-media-sites-facebook-twitter-hate-speech>
- [9] See, for example, Full Fact, "The State of Automated Factchecking", <https://fullfact.org/automated> and Lori Hawkins, "Austin startup wants to end fake news - and fake everything else - on the internet", 512Tech, February 2017: <http://www.512tech.com/technology/austin-startup-wants-end-fake-news-and-fake-everything-else-the-internet/EcchWFgrl4PQmjPvmkzycJ/>
- [10] Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford: Oxford University Press, 2009
- most important and dangerous features of the phenomenon it describes which is not deceptiveness, but "spreadability".^[7]
- Proposed responses include new media literacy, educational and fact-checking initiatives; new laws, policies and fines for technology companies who fail to remove offending content^[8]; and a host of new startups and technical fixes – from authenticated content to automated fact-checking projects.^[9]
- Across these different kinds of responses, observers agree that the term "fake news" is deceptive and that these problematic fabrications cannot be straightforwardly defined. And while we "follow the actors"^[10] and retain the main name by which these activities have been originally turned into an issue of public concern to indicate the controversy that prompted us to undertake this empirical investigation, we recognise that fabrications gathered under the label "fake news" come in many different shades. This need not be taken as proof of the futility of investigating this phenomenon. On the contrary: their different shades are what is at stake in our investigation and accepting that there is no easy way to demarcate between "fake" and "non-fake" across all cases opens interesting research opportunities. It is precisely because its forms and contents are designed to mimic those of mainstream media – and precisely because it travels through similar circuits – that fake news offers us the occasion to study not just the strategies and formats of fakeness, but the politics and composition of the media and information environments of the digital age more generally.
- This guide aims to enrich public debate and catalyse collective inquiry around this rapidly evolving and highly contested issue – by suggesting different ways in which it can be empirically studied, mapped and investigated online. Ultimately our hope is not just to provide better accounts of the issue of fake news and phenomena associated with it, but also to contribute to more substantive forms of public engagement around it. We hope this guide will contribute to facilitating broader public debate and involvement

around processes of reshaping platforms and policies, laws and infrastructures, technologies and standards that are implicated in the circulation of fake news and other fabrications. This includes remaining attentive to possible unintended consequences of these different responses, as well as other interests and concerns.

The guide explores the notion that fake news is not just another type of content that circulates online, but that it is precisely the *character* of this online circulation and reception that makes something into fake news. In this sense fake news may be considered not just in terms of the form or content of the message, but also in terms of the mediating infrastructures, platforms and participatory cultures which facilitate its circulation. In this sense, the significance of fake news *cannot be fully understood apart from its circulation online*. It is the register of this circulation that also enables us to trace how material that starts its life as niche satire can be repackaged as hyper-partisan clickbait to generate advertising money and then continue life as an illustration of dangerous political misinformation.

As a consequence this field guide encourages a shift from focusing on the formal *content* of fabrications in isolation to understanding the contexts in which they *circulate* online. This shift points to the limits of a “deficit model” approach – which might imply that fabrications thrive only because of a deficit of factual information. In the guide we suggest new ways of mapping and responding to fake news beyond identifying and fact-checking suspect claims – including “thicker” accounts of circulation as a way to develop a richer understanding of how fake news moves and mobilises people, more nuanced accounts of “fakeness” and responses which are better attuned to the phenomenon.

[11] See Theodore Porter,
“Thin Description: Surface
and Depth in Science and
Science Studies.” *Osiris*,
2012: [www.jstor.org/
stable/10.1086/667828](http://www.jstor.org/stable/10.1086/667828)

While online and platform metrics often serve to take measure of engagement by means of what Theodore Porter calls “thin descriptions”^[11] – i.e. aggregated quantities such as total likes, shares, posts – we suggest different ways of

exploring how different publics engage with and ascribe meaning to fake news and how this moves and mobilises different actors in the process. In doing so while we start our inquiry with fake news, we end up surfacing a wide range of grassroots political, media and participatory cultures online and the social and political issues around which they assemble. Some of these may challenge and prompt a rethinking of our ideas of the forms and formats of grassroots political action online.

We have adopted the metaphor of the “field guide” in the tradition of a number of recent guides which transpose the language and imagery of mapping places, flora and fauna onto the cloud, digital infrastructures and life online.^[12] However this metaphor stands in need of some qualification. Many classical natural historical “field guides” aspire to provide systematic taxonomies of natural phenomena by taking them out of their contexts in order to abstract and compare their features. By contrast with our guide we aim to do precisely the opposite – not to *decontextualise*, but to *recontextualise* fake news phenomena by suggesting ways to follow them “in the wild”: as they travel across the web, search engines, digital platforms, fact-checking initiatives and news websites.

We do not set out to provide a definitive single set of watercolour portraits, anatomical illustrations, cartographic charts, satellite imagery or infrastructural diagrams of the phenomenon in question – or even lists of characteristic features which may be used for the purposes of identification. Instead we illustrate a range of methods and procedures which readers may use in order to explore fake news phenomena online for themselves. As part of this process we wish to extend the repertoire of mapping practices which are publicly available to make sense of fake news online and in this sense the graphics that we provide can be understood as temporary placeholders to encourage further exploration.

We also draw attention to different ways of examining how

[12] For a recent overview see Shannon Mattern’s “Cloud and Field”, *Places Journal*, August 2016: <https://placesjournal.org/article/cloud-and-field/>

[13] See Philip E. Agre, "Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI", in Geof Bowker, Les Gasser, Leigh Star and Bill Turner (eds), "Bridging the Great Divide: Social Science, Technical Systems and Cooperative Work", NJ: Erlbaum, 1997. Available at: <http://polaris.gseis.ucla.edu/pagre/critical.html>

[14] Shannon Mattern, "Cloud and Field", Places Journal, August 2016: <https://placesjournal.org/article/cloud-and-field/>

[15] See, for example, the work of Pamela Smith and her colleagues on the "Making and Knowing" project at Columbia University: <http://recipes.hypotheses.org/7430>, February 2016, and <http://www.makingandknowing.org/>

things are *categorised* and *labelled* as fake news and the politics of these practices of classification. In this sense we hope to cultivate what has been called "critical technical practice"^[13] – which in this case would include reflection on the use of digital methods and digital data and how these not only serve to *designate* phenomena which can be straightforwardly and independently picked out, but how these very methods may also be involved in the process of *articulating* what fake news is. As Shannon Mattern puts it, in undertaking to investigate fake news online we should be aware of "the shadows cast by our presence as explorers in the field."^[14] And rather than producing maps for the sake of producing maps, we should consider what maps *do*, who and what they are *for* and the *effects* that they produce as social, cultural and political devices.

Insofar as we focus on providing procedures for inquiry rather than pictures of the phenomena, this guide may also be considered a kind of "recipe book." Recent research suggests that there is an interesting relation between the documentation of recipes and the emergence of procedural knowledge in the early modern period – such that practices of writing down processes for cooking and craft are entangled with the history of the emergence of scientific method.^[15] Over the past few decades the metaphors of the "recipe" and the "cookbook" have also become popular in relation to software programming. In our guide, we illustrate different approaches to mapping and investigating fake news online through a series of methodological "recipes." As with many cookery books, our aim is not just to support readers in following the specific recipes that we present, but rather to use these recipes to illustrate a certain approach to cooking – with the hope that readers are inspired to adapt, modify and venture beyond them. We also include a number of "serving suggestions" about how they may be put to work.

We hope that the recipes in this guide will enrich investigations of fake news and other fabrications in a way which has affinities with a common narrative approach

in mystery fiction – namely the scenario that in pursuit of solving an apparently simple crime, the plot thickens, the cast grows, the questions multiply and there are unexpected twists or changes of perspective. By following the production, circulation and responses to fake news online – we may end up being drawn into things that we do not set out to investigate: whether the media strategies of fake news publishers, propagandists, trolls or bots; the commercial and technical architectures of online content; the politics and dynamics of viral content; and how social life adapts, evolves and innovates in response to some of the world’s biggest online platforms and websites. In this sense, it will be clear that fake news involves more than a few rogue producers or state conspiracies – and raises important and difficult questions about the role of digital technologies in society and how we mutually shape and are shaped by them.

[16] Edgar Allan Poe and Jacob Schwartz, "The Purloined Letter", London: Ulysses bookshop, 1931.

[17] See, Noortje Marres, "The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived", *The Sociological Review*, June 2012; and Mike Ananny and Kate Crawford, "Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability", *New Media and Society*, December 2016.

[18] Liliana Bounegru, Mette Simonsen Abildgaard, Andreas Birkbak, Jonathan Gray, Mathieu Jacomy, Torben Elgaard Jensen, Anders Koed Madsen and Anders Kristian Munk, "Five Provocations about Fake News" (under review).

In Edgar Allan Poe's classic mystery story "The Purloined Letter"^[16], the prefect of police – "G" – and his colleagues search for a letter said to contain scandalous information behind wallpaper, under carpets, in the legs of furniture and in cushions, only to eventually find the letter "hiding in plain sight". In a similar vein, we may consider the algorithmically mediated circulation of fake news on digital platforms in terms of what Noortje Marres characterises as "distributed accomplishment" or what Mike Ananny and Kate Crawford describe as "relational achievement".^[17] This entails a shift from "seeing in" systems as a kind of looking "under the hood", to "seeing across" a diverse range elements which are implicated in the patterning of collective life online.

Many of the researchers who have contributed to the guide share a background in a field called Science and Technology Studies (STS). Some of the lines of inquiry pursued in the guide are informed by a forthcoming paper exploring what STS can bring to the study of fake news.^[18] The recipes are also informed by a "digital methods" research approach that has developed through an engagement with this field and which many of us have contributed to through our teaching

[19] See, for example, Richard Rogers, "Digital Methods", 2013, Cambridge, MA: MIT Press.

[20] See, for example, Noortje Marres, "Material Participation: Technology, the Environment and Everyday Publics", London: Palgrave Macmillan, 2012.

[21] See Tommaso Venturini, Anders Munk and Axel Meunier, "Data-Sprint: A Public Approach to Digital Research" in C. Lury, P. Clough, M. Michael, R. Fensham, S. Lammes, A. Last, & E. Uprichard (Eds.) "Interdisciplinary Research Methods", London: Routledge, 2017.

[22] Jonathan Gray, Liliana Bounegru and Lucy Chambers (Eds.) "The Data Journalism Handbook", Sebastopol, CA: O'Reilly Media, 2012, available at: <http://datajournalismhandbook.org/>

[23] Upon being invited to become Director of the Argentine National Library at a moment which coincided with the deterioration of his eyesight, Borges famously wrote: "No one should read self-pity or reproach / into this statement of the majesty / of God; who with such splendid irony / Granted me books and night at one touch". See J. L. Borges, "Poem of the Gifts" in "Selected Poems: Volume 2", London: Penguin Books, 2000, p. 95.

and research.^[19] We also draw upon our field's interest in public engagement and participation around digital technologies and data infrastructures.^[20] As such our focus is less on advancing particular legal or technical fixes, than on facilitating processes of public engagement and democratic deliberation – including provoking curiosity about different ways of seeing the issue and imagination about the different ways in which we might respond.

The material in this guide has been produced through a series of "data sprints" and research workshops in Amsterdam, Copenhagen and Milan, hosted by members of the Public Data Lab. The "data sprint" is a short form working format that has emerged at the intersection between Science and Technology Studies and New Media Studies, drawing on approaches associated with open-source software development, open data and civic hacking in order to convene a range of actors to collaborate around the co-production of data and research projects – including between fields of practice with different outlooks.

Two of us have a background in data journalism, having co-edited The Data Journalism Handbook and undertaken various initiatives in this field.^[22] This guide builds on a long-standing interest in supporting productive encounters between data journalists and digital researchers. While fake news seems like a remarkably ripe area for experimentation between these two fields, just as the writer Jorge Luis Borges lamented being granted "books and night at one touch"^[23] it is not without a sense of irony that we note that as public attention around this issue grows, fake news websites are beginning to vanish – leading to proposals for a "fake news archive" amongst our contributing researchers. Happily the approaches and analytical techniques in this guide may be used to inform collaborations between data journalists and digital researchers around the study of other contentious issues and controversies as they unfold on digital media, as well as of the mediating capacities of platforms, algorithms and infrastructures which shape life online.

The data sprint format has also helped us to catalyse new experimentation and empirical work in a comparatively short period of time – a distinct advantage given the pace of developments around fake news. For this we are immensely grateful to researchers, graduates and students at DensityDesign Lab (Politecnico di Milano, Italy), the Digital Methods Initiative (University of Amsterdam, Netherlands), the European Journalism Centre, the Laboratoire Interdisciplinaire Sciences Innovations Sociétés (Université Paris-Est, France), the médialab (Sciences Po, Paris, France), the Media of Cooperation research group (University of Siegen, Germany), the STS-Lab (University of Lausanne, Switzerland) and the Techno-Anthropology Lab (Aalborg University Copenhagen, Denmark) – without whose energy, creativity and dedication this project would not have been possible.

*Jonathan Gray (@jwyg), Liliana Bounegru (@bb_liliana),
Tommaso Venturini (@TommasoVenturin)
London, March 2017*

CONVENTIONS USED IN THE BOOK

In this book we use the  (eye) symbol to indicate visual results, the  (wrench) to point to the tools glossary and the  (arrow) to point to the concepts glossary. To avoid distracting our readers we only use the glossaries icons to mark the first occurrence per recipe of the term or tool explained in glossaries.

Furthermore, each recipe in our chapters is introduced by a diagram, or a method map, representing the key analytical steps taken to arrive at our results. In each method map, arrows represent actions and icons represents their results. You can see the steps in the method maps as possible ingredients for your own recipes.

Some recipes lead to multiple outcomes. When this is the case you will find at the beginning of the recipe a complete method map for the entire recipe (on a blue background), and the parts relevant to each individual step in the recipe highlighted on a white background at the beginning of each recipe alongside the description of the relevant step.

Below you can find a list of all the icons we use for the methods maps.



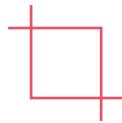
A **dataset** in the form of a table.



Any kind of **visualization**, such as a bubble graph or a network diagram. See the Concepts Glossary for the full list of visual models used in this guide.



A **list**. This could be, for example, a list of websites, or a list of Facebook pages.



A screenshot. Usually taken from a web browser with the aim of preserving a snapshot of a web page.



A corpus of images. A set of images captured with the same method.



User profile. It represents all the information related to a user in a social network. For example in Twitter the user profile contains the @name, the description, the profile picture.



Temporal information. Could be, for example, the creation time of a Facebook post.



Hashtag. Used in many social networks, for example in Twitter and Facebook.

ACTIONS

Sometimes, relevant actions have their own icons. Below you can find the full list of them.



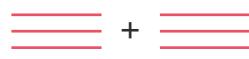
Automatic operations. Used to highlight when an action (e.g. dividing items into categories) is performed by a machine.



Manual operations. Used to highlight when an action (e.g. dividing items into categories) is performed by a human being.



Image comparison. It is used to highlight when the analyst must visually compare a corpus of images.



Union of lists. When two or more lists are merged into one.

Chapter 1

MAPPING FAKE NEWS HOTSPOTS ON FACEBOOK

What publics does fake news
animate on Facebook?

How may the trajectory of a
fake news story be traced on
Facebook?

Do fact-checking initiatives reach
the publics of fake news on
Facebook?

Introduction - This section provides a set of recipes for tracing the circulation of fake news on Facebook. The focus is on circulation because false and misleading knowledge claims are not born “fake news”. To become fake news they need to mobilise a large number of publics – including witnesses, allies, likes and shares, as well as opponents to contest, flag and debunk them. Facebook’s architecture poses challenges to the study of circulation of content due to the nature of its access and permissions system. Hence we focus on tracing the publics of fake news through its most publicly accessible entities: pages and groups, which may be considered to constitute already assembled publics.

Around the 2016 US presidential elections commentators have noted the emergence of a Facebook-native, hyper-partisan “political media machine” that was highly effective in gathering large numbers of → followers and

[1] See, John Herrman, "Inside Facebook's (Totally Insane, Unintentionally Gigantic, Hyperpartisan) Political-Media Machine", Agust 2016, *The New York Times*: <https://www.nytimes.com/2016/08/28/magazine/inside-facesbooks-totally-insane-unintentionally-gigantic-hyperpartisan-political-media-machine.html>

generating → engagement [1]. This fake news dissemination machine and responses to it, is what the recipes in this section enable to explore. The first two

recipes focus on mapping the publics that are energised by fake news on Facebook, as well as the trajectories through which fake news stories travel on Facebook. The third recipe provides an approach to address the effectiveness of fact-checking initiatives in reaching the publics of fake news on Facebook. Through these recipes we aim to gesture towards different ways of providing “thicker” accounts of circulation and engagement around fake news on social media beyond the “thin descriptions” of aggregated counts and metrics.



WHAT PUBLICS DOES FAKE NEWS ANIMATE ON FACEBOOK?

BEFORE STARTING

The starting point for this recipe is a list of fake news stories. There are different ways of obtaining these lists – including starting with existing lists as well as creating your own. To illustrate this recipe we use an already existing list of 22 fake news stories about various political issues pertaining to the 2016 presidential elections in the US that generated most → engagement on Facebook. These were identified by *BuzzFeed News*.

The recipe comprises of four steps. We start by identifying the themes that are exploited in our set of stories as well as the key political events which they editorialise (👁a). Next we identify the most prominent public Facebook pages and groups that share these stories (👁b). We also explore whether certain publics have preferred story themes (👁c) and profile the publics that are energised by fake news stories about the US elections (👁d).

START

list of 22 URLs of political fake news stories
Source: BuzzFeed News



input URLs in

output data



- > Fake news story URLs
- > Facebook pages and groups that share the URLs
- > Number of interactions per each page or group
- > Date of sharing of the story

identify time intervals with highest frequency of publication of fake news stories

identify key related events with



a

WHICH MEDIA AND POLITICAL EVENTS ARE SUCCESSFUL IN SETTING THE FAKE NEWS AGENDA?

import data in



b

WHICH FACEBOOK PAGES AND GROUPS PROMOTED THE HIGHEST NUMBER OF FAKE NEWS STORIES?

import data in



c

DO FACEBOOK PUBLICS HAVE PREFERRED STORY THEMES?

manually categorise Facebook pages



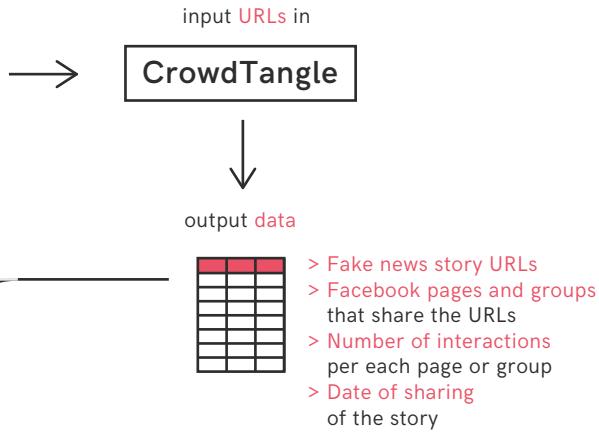
d

WHAT KINDS OF PUBLICS ARE ENERGISED BY FAKE NEWS?

START

list of 22 URLs of political fake news stories

Source: BuzzFeed News



identify time intervals with highest frequency of publication of fake news stories

identify key related events with

Google News



a

WHICH MEDIA AND POLITICAL EVENTS ARE SUCCESSFUL IN SETTING THE FAKE NEWS AGENDA?



**a. EXAMINE THE THEMES EXPLOITED
IN FAKE NEWS STORIES AND
IDENTIFY THE EVENTS WHICH THEY
EDITORIALISE**

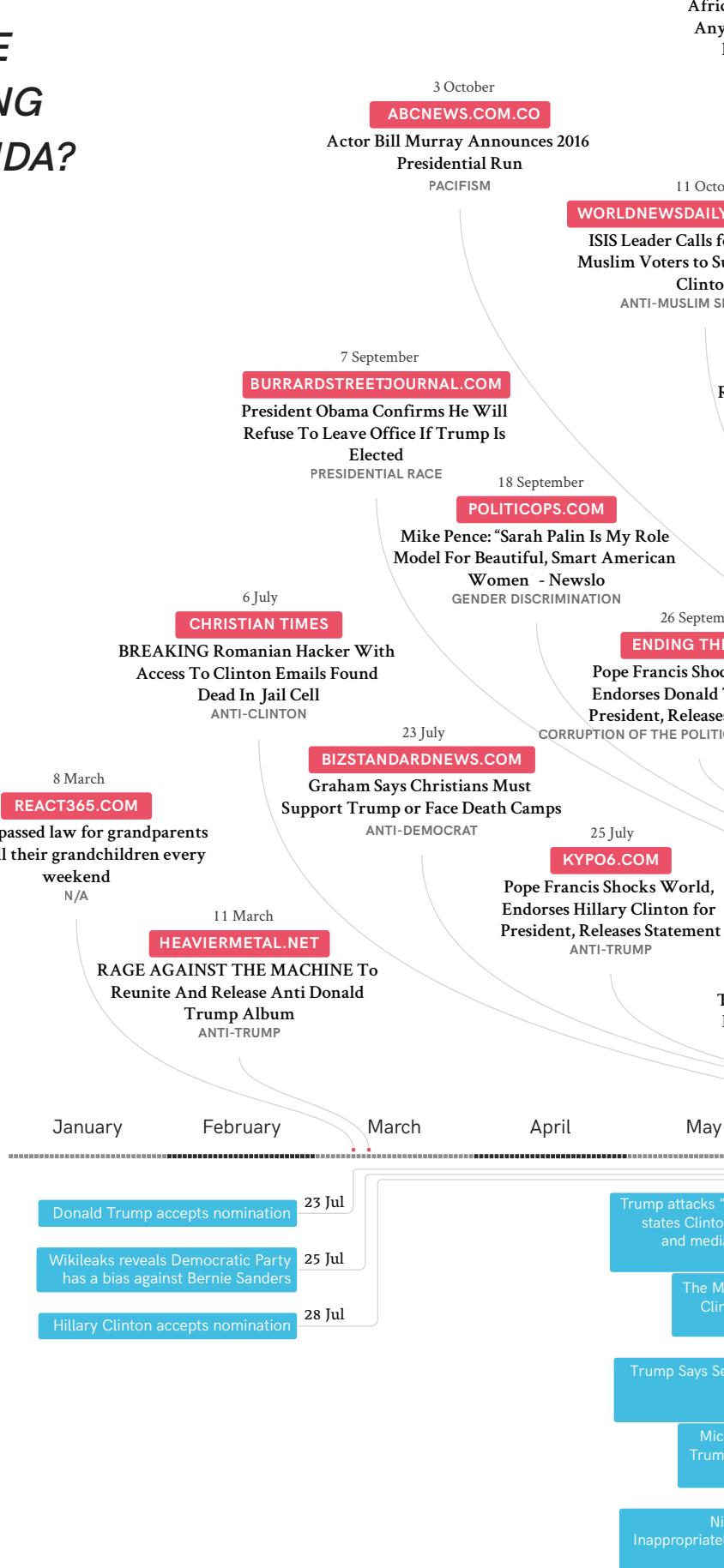
This analysis may be done by qualitatively examining the content of each article and identifying key political or media events related to the issues exploited in the articles, which occurred around the publication date of each story. This is done to enable a better understanding of the issues that animate the publics that circulate fake news.

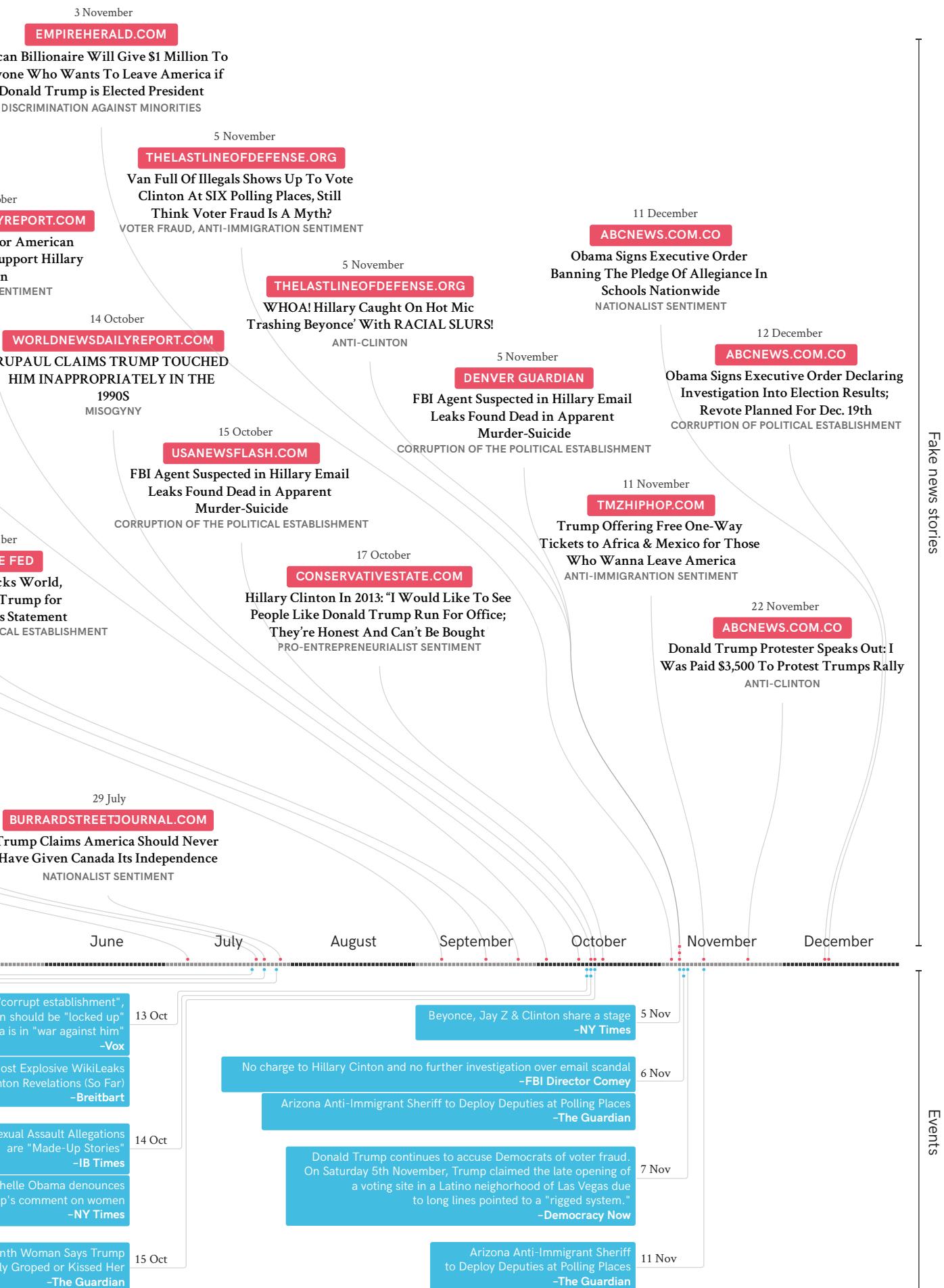
- ◊ If the content of the fake news article is no longer available on its original URL you may use the Internet Archive's  WayBack Machine to check whether an archived version of the URL is available.
- ◊ To identify key events occurring around the dates of publication of the stories which are related to the themes exploited in the stories you may use a news aggregator such as  Google News Search as well as news article archives.
- ◊ An annotated timeline of stories and relevant events occurring around the same dates might provide a starting point for reflection about the relationship between political and media events and fake news stories.

WHICH MEDIA AND POLITICAL EVENTS ARE SUCCESSFUL IN SETTING THE FAKE NEWS AGENDA?

Timeline of best performing fake news stories about the US elections on Facebook in 2016 and events they editorialise.

Successful fake news stories appear to exploit populist themes such as anti-establishment sentiment, nationalist and anti-immigration sentiment as well as perceived or projected weaknesses of political candidates such as misogyny and corruption. A number of events at the end of July, mid-October and early November are successful in setting the fake news "agenda".

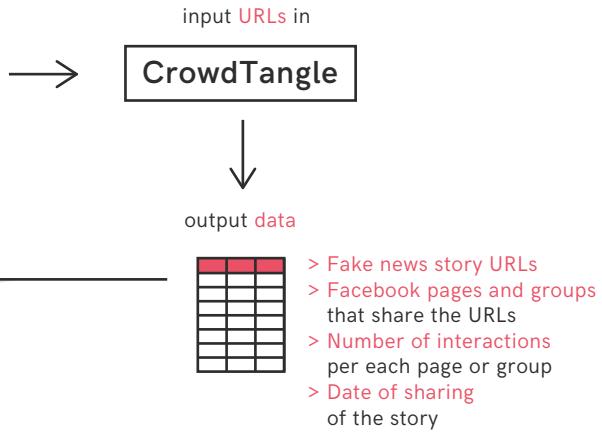




START

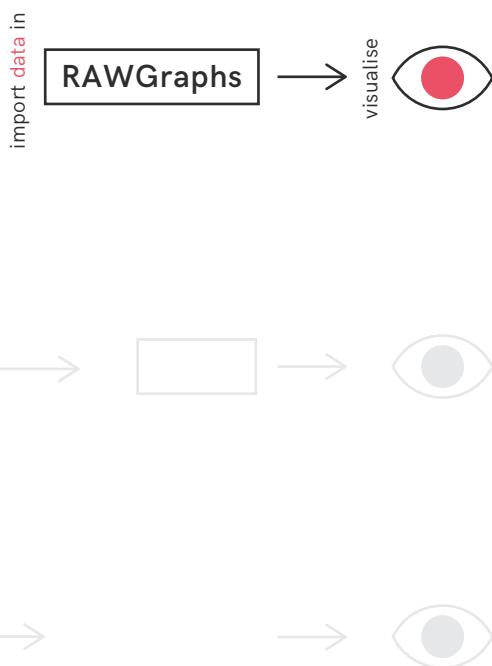
list of 22 URLs of political fake news stories

Source: BuzzFeed News



b

**WHICH FACEBOOK PAGES
AND GROUPS PROMOTED THE
HIGHEST NUMBER OF FAKE
NEWS STORIES?**



b. IDENTIFY THE FACEBOOK PAGES AND GROUPS THAT SHARE THESE STORIES

This may be done with a social media monitoring tool such as the browser extension of  CrowdTangle. The number of followers per page or group as well as the number of → interactions per posts should be recorded in a spreadsheet alongside the names of pages and groups that share fake news stories.

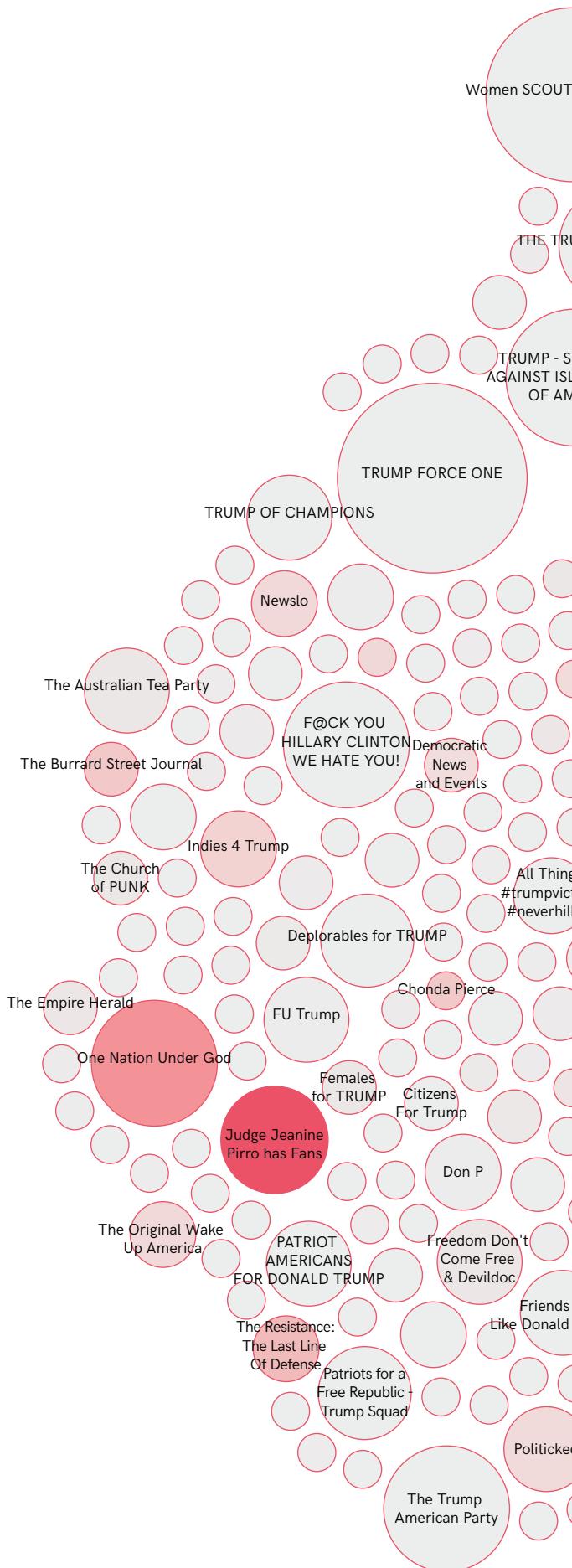
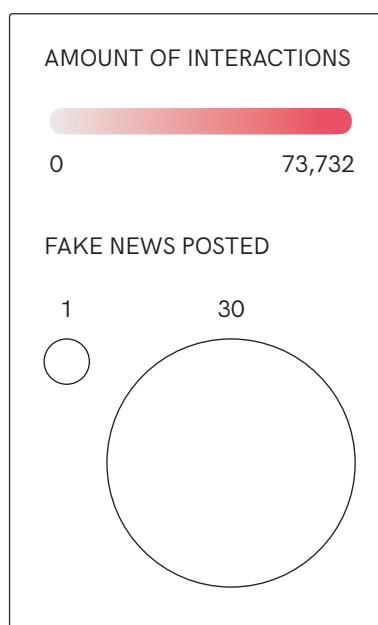
- ◊ Please note that a fake news story may be reposted on a number of different websites. For this reason a methodological decision needs to be taken from the outset as to whether only the pages and groups that share the original URL of the story will be recorded or whether all pages and groups that share all versions of the fake news story will be collected.
- ◊ You may want to take note of the pages or groups which shared the highest number of fake news stories as well as the total number of interactions generated by each group or page.
- ◊ If you use  CrowdTangle please note that for Facebook the tool returns the top 500 most popular public posts to verified pages as well as to pages with more than 125.000 fans.^[1]
- ◊ You may use a → circle packing visualisation to represent the pages and groups that share fake news items as well as the number of stories which they share and the number of interactions which they generate. You may use  RAWGraphs for this operation.

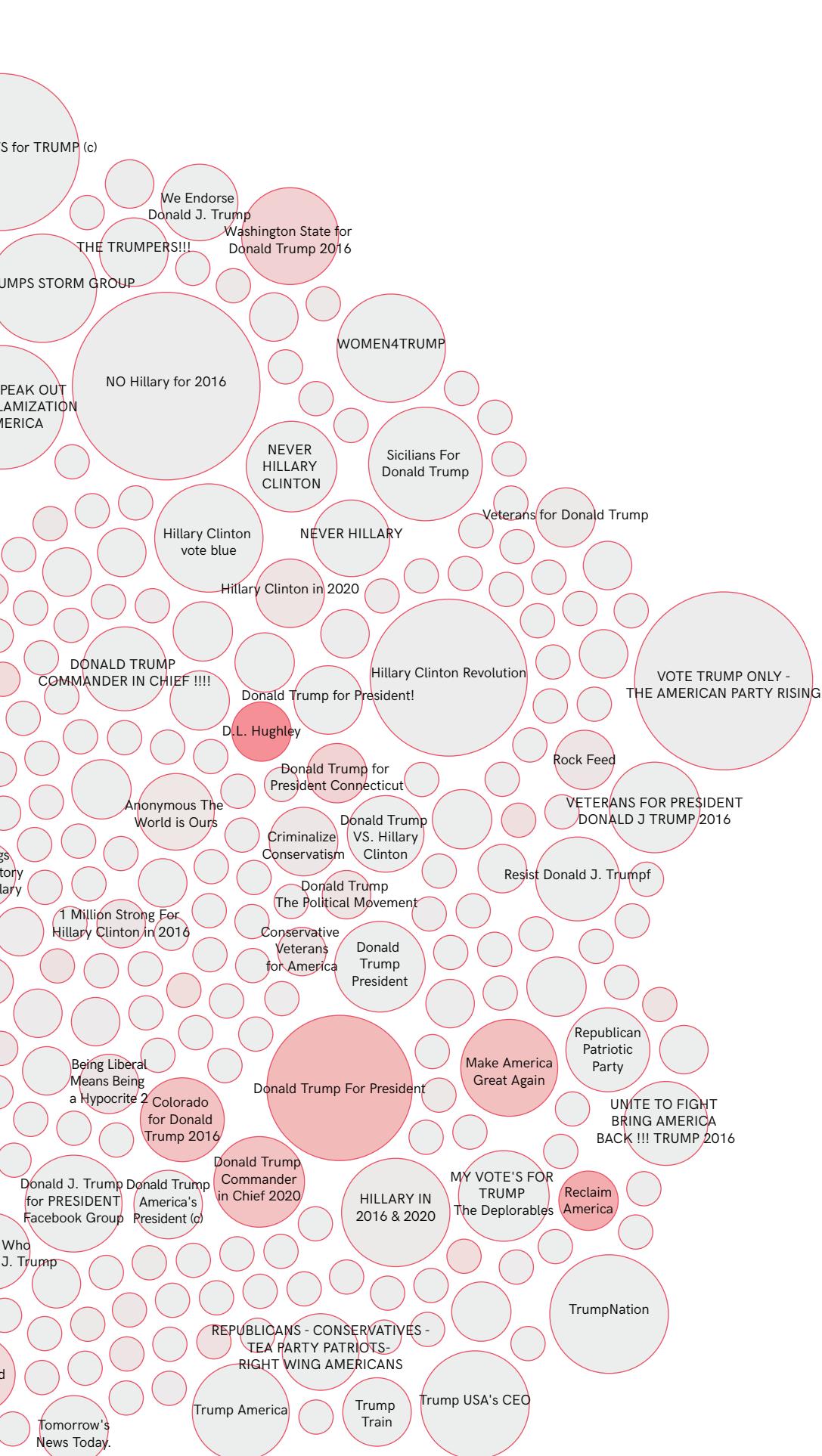
[1]

See, CrowdTangle's "Frequently Asked Questions", available at: <https://apps.crowdtangle.com/chrome-extension/faq>

WHICH FACEBOOK PAGES AND GROUPS PROMOTED THE HIGHEST NUMBER OF FAKE NEWS STORIES? WHICH ONES CREATED THE HIGHEST ENGAGEMENT?

Public Facebook pages and groups that share fake news items, sized according to the number of items they share and coloured according to their number of followers. Each page can share the same item more than once. The pages and groups that share the highest number of stories are primarily pro-Trump supporters and anti-Hillary groups. The page that generates the highest number of interactions with fake news stories is the fan page dedicated to republican TV commentator, Jeanine Pirro.

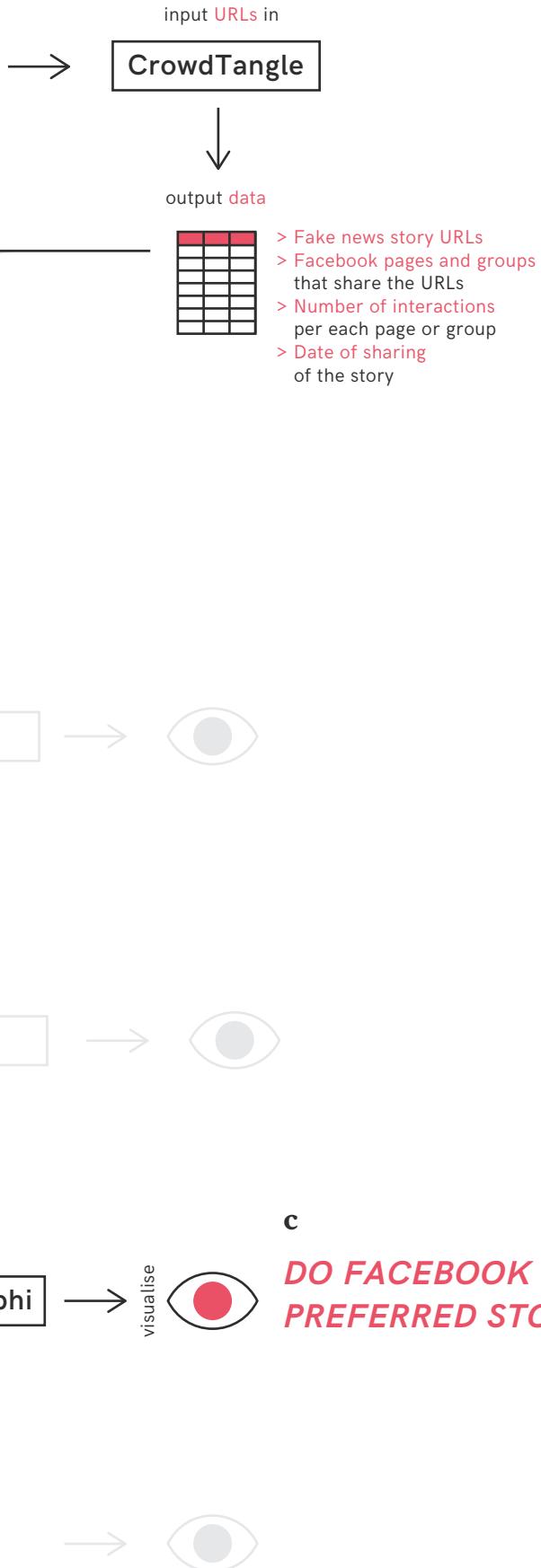




START

list of 22 URLs of political fake news stories

Source: BuzzFeed News



c. IDENTIFY WHETHER FACEBOOK PUBLICS HAVE PREFERRED STORY THEMES

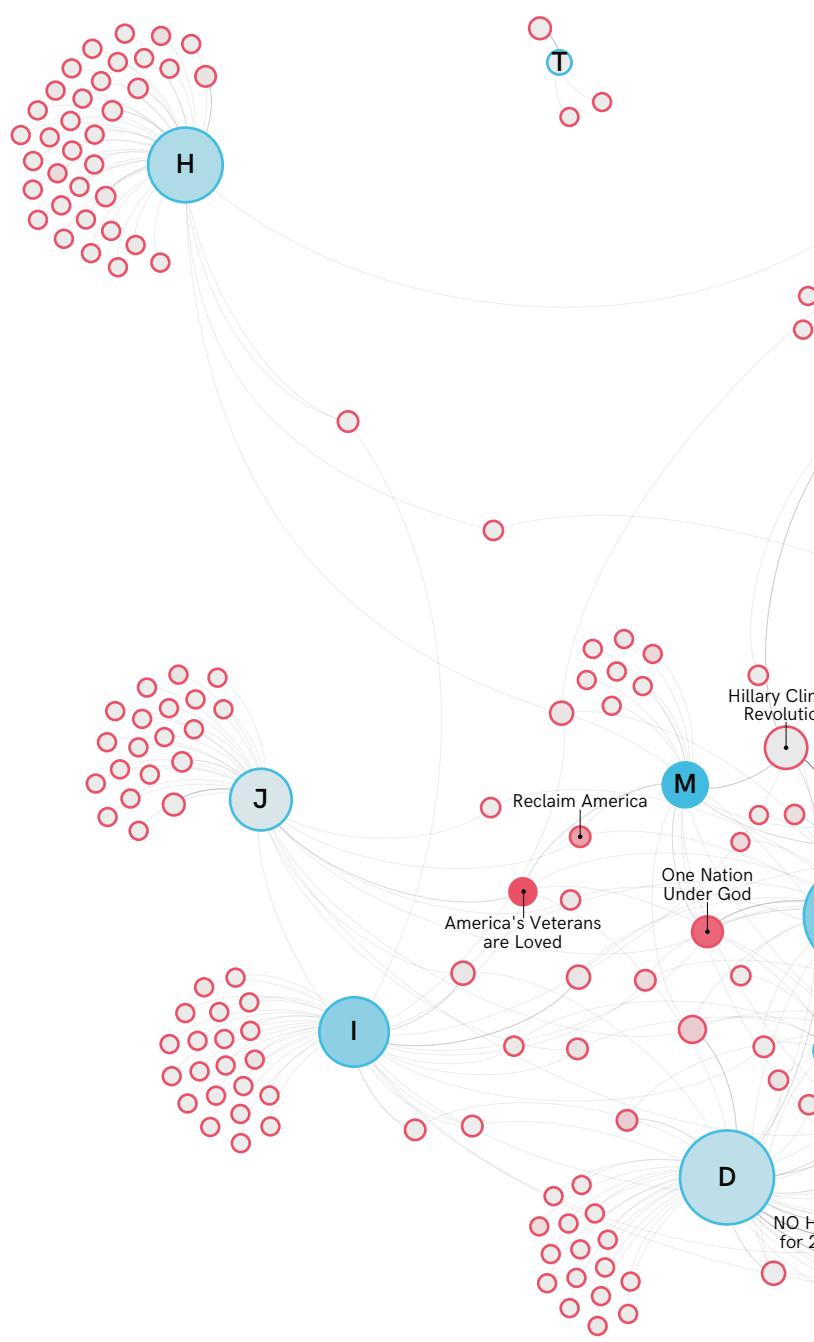
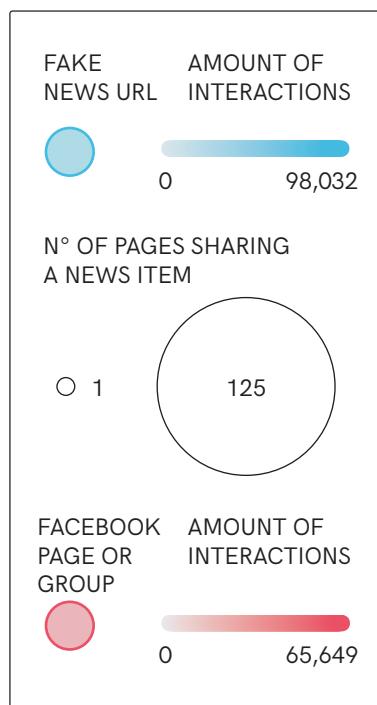
To explore whether particular story themes assemble publics and to qualitatively profile those publics based on the stories that animate them you may conduct a network analysis of public Facebook pages and groups connected by the stories which they share.

- ◊ Starting from the dataset extracted with  **CrowdTangle's** browser extension, you may create a network file where each time a Facebook group or page posts a fake news story a link is established between that page or group and that story.
- ◊ You may use  **Table2Net** to convert your CSV (comma-separated values) file into a network file and  **Gephi** to explore the network. A force-directed layout algorithm such as ForceAtlas2^[2] can help you visualise the outcomes.
- ◊ Identify which stories are most successful in energising publics as well as whether publics have preferred story themes.

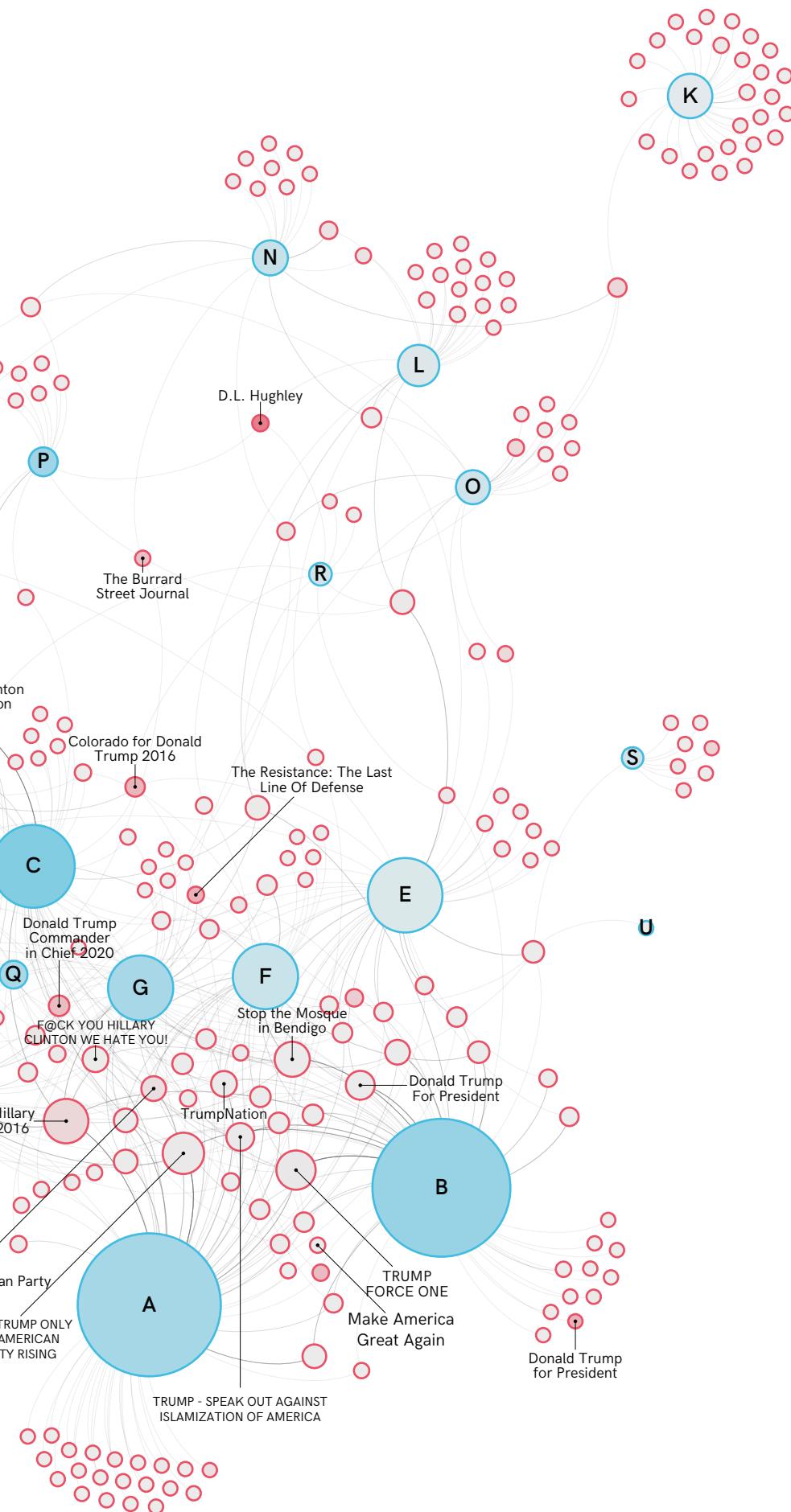
[2] See, Mathieu Jacomy, Tommaso Venturini, Sébastien Heymann and Mathieu Bastian, "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software", June 2014, *PLoS ONE*: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679>

DO FACEBOOK PUBLICS HAVE PREFERRED STORY THEMES?

Network of public Facebook pages and groups connected by the fake news stories which they share. Notable is the core of the network which consists of a series of pages and groups associated with Trump supporters which are animated by anti-Hillary stories.



VOTE THE PAR



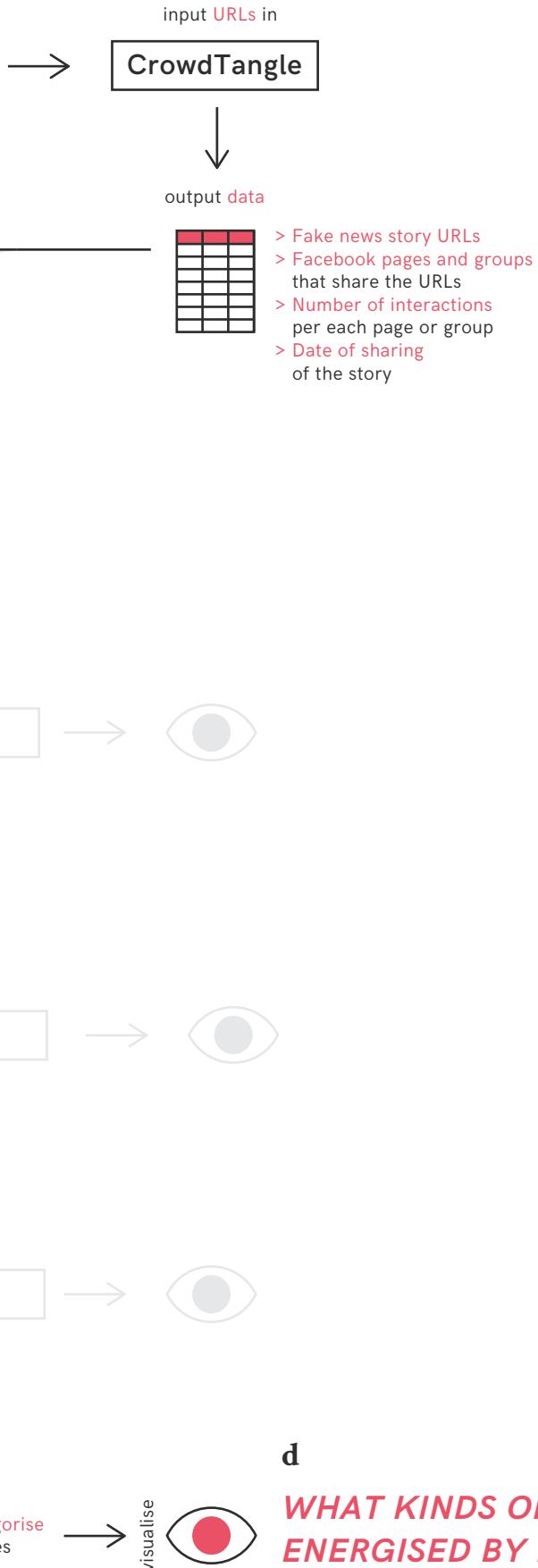
Fake News Headlines

- A FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide
- B Hillary Clinton In 2013: "I Would Like To See People Like Donald Trump Run For Office; They're Honest And Can't Be Bought"
- C ISIS Leader Calls for American Muslim Voters to Support Hillary Clinton
- D Donald Trump Protester Speaks Out: "I Was Paid \$3,500 To Protest Trump's Rally"
- E Obama Signs Executive Order Declaring Investigation Into Election Results; Revote Planned For Dec. 19th
- F WHOA! Hillary Caught On Hot Mic Trashing Beyonce' With RACIAL SLURS!
- G Van Full Of Illegals Shows Up To Vote Clinton At SIX Polling Places, Still Think Voter Fraud Is A Myth?
- H RAGE AGAINST THE MACHINE To Reunite And Release Anti Donald Trump Album
- I Obama Signs Executive Order Banning The Pledge Of Allegiance In Schools Nationwide
- J BREAKING Romanian Hacker With Access To Clinton Emails Found Dead In Jail Cell
- K Actor Bill Murray Announces 2016 Presidential Run
- L Pope Francis Shocks World, Endorses Hillary Clinton for President, Releases Statement
- M Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement
- N Trump Claims America Should Never Have Given Canada Its Independence
- O Mike Pence: "Sarah Palin Is My Role Model For Beautiful, Smart American Women" - Newslo
- P RUPAUL CLAIMS TRUMP TOUCHED HIM INAPPROPRIATELY IN THE 1990S
- Q President Obama Confirms He Will Refuse To Leave Office If Trump Is Elected
- R Graham Says Christians Must Support Trump or Face Death Camps
- S African Billionaire Will Give \$1 Million To Anyone Who Wants To Leave America if Donald Trump is Elected President
- T Trump Offering Free One-Way Tickets to Africa & Mexico for Those Who Wanna Leave America
- U Obama passed law for grandparents to get all their grandchildren every weekend

START

list of 22 URLs of political fake news stories

Source: BuzzFeed News



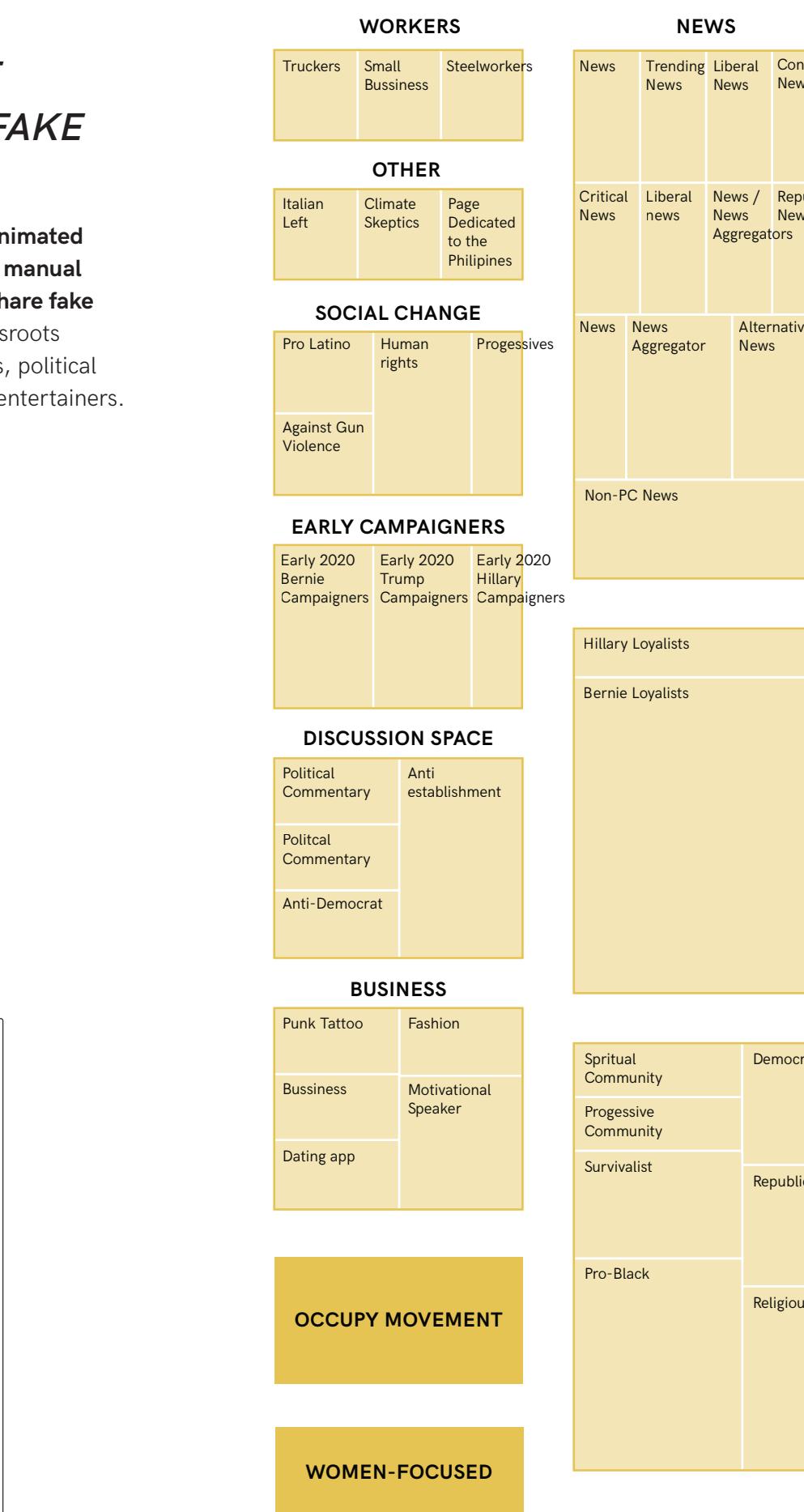
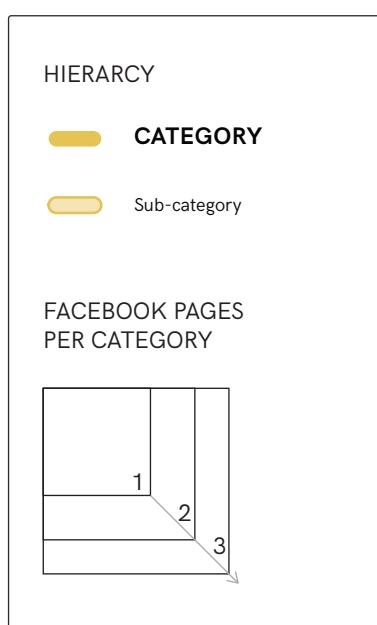
d. PROFILE THE PUBLICS ANIMATED BY FAKE NEWS

This may be done by conducting a qualitative analysis of all public Facebook pages that share fake news items based on self-descriptions available on their “About” pages.

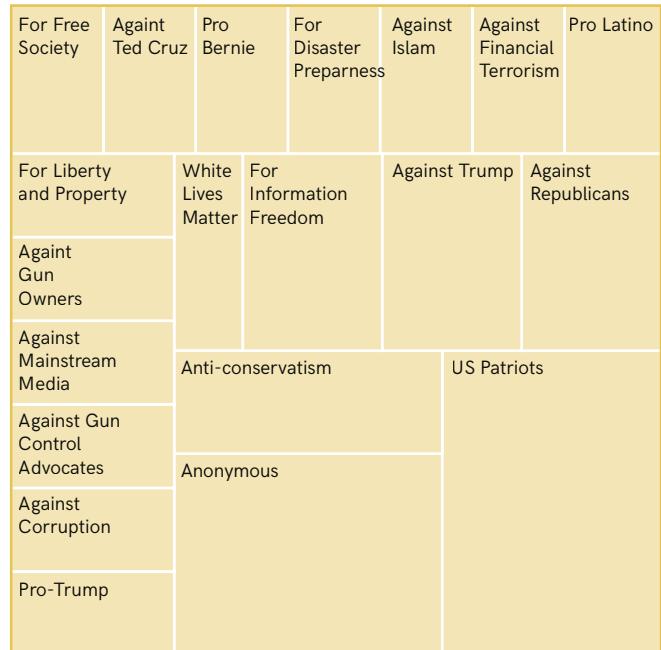
- ◊ You may take an → **emergent coding** approach to identify the themes that emerge from the description of pages. You may take note of a more generic category (e.g. “grass-roots activism”) as well as a more specific one (e.g. “anti-establishment”).
- ◊ Sum up the amount of followers across all pages belonging to the same category.
- ◊ A → **treemap** visualisation may be used to represent the weight and hierarchy of each category. You may use  **RAWGraphs** for this operation.

WHAT KINDS OF PUBLICS ARE ENERGISED BY FAKE NEWS?

Types of Facebook publics animated by fake news, according to a manual classification of pages that share fake news items. Notable are grassroots activists for a variety of issues, political candidate loyalists as well as entertainers.



GRASSROOTS ACTIVISTS/ CITIZEN VIGILANTES



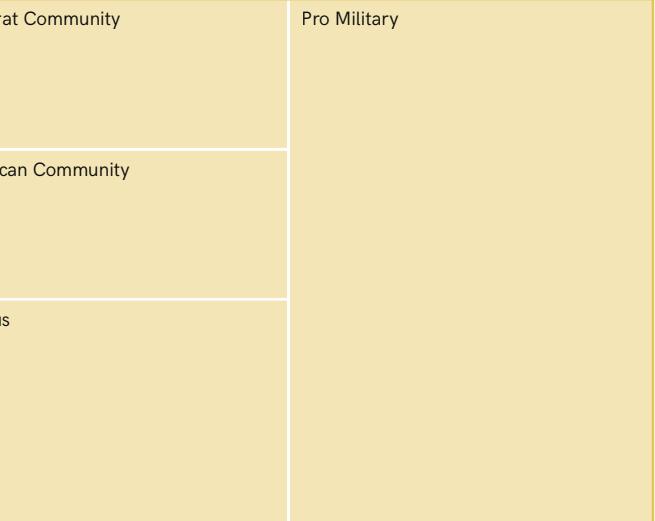
PUBLIC FIGURE



LOYALISTS

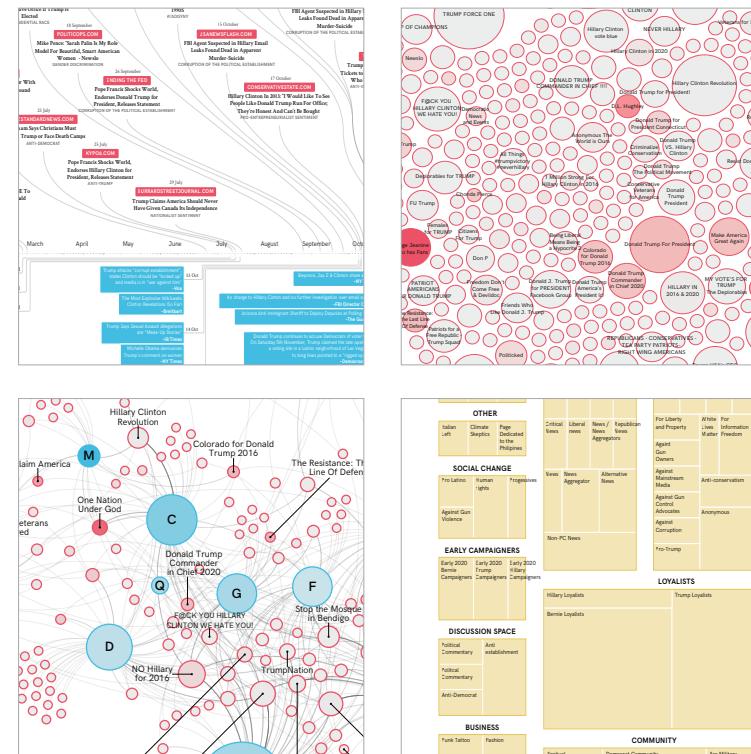
Trump Loyalists

COMMUNITY



MUSIC/ENTERTAINMENT





SERVING SUGGESTIONS

This recipe may be used to better understand the publics that are animated by fake news and the meaning making activities that they engage in around fake news, i.e. how they enroll fake news in the service of their own issue work. This approach may inform a thicker description of the impact of fake news that moves away from its viral character (the single engagement number or metric) to understanding who it mobilises and how.

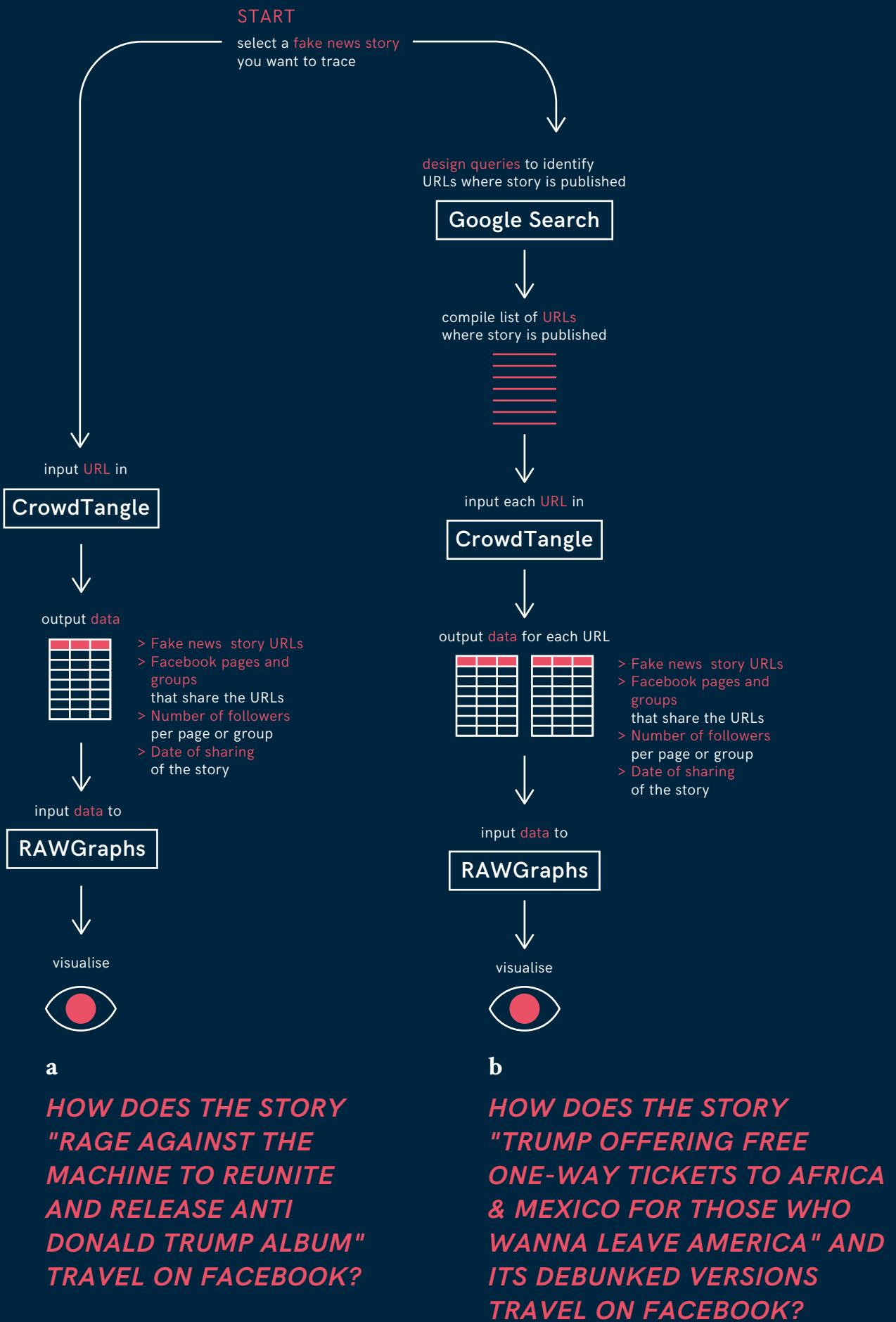


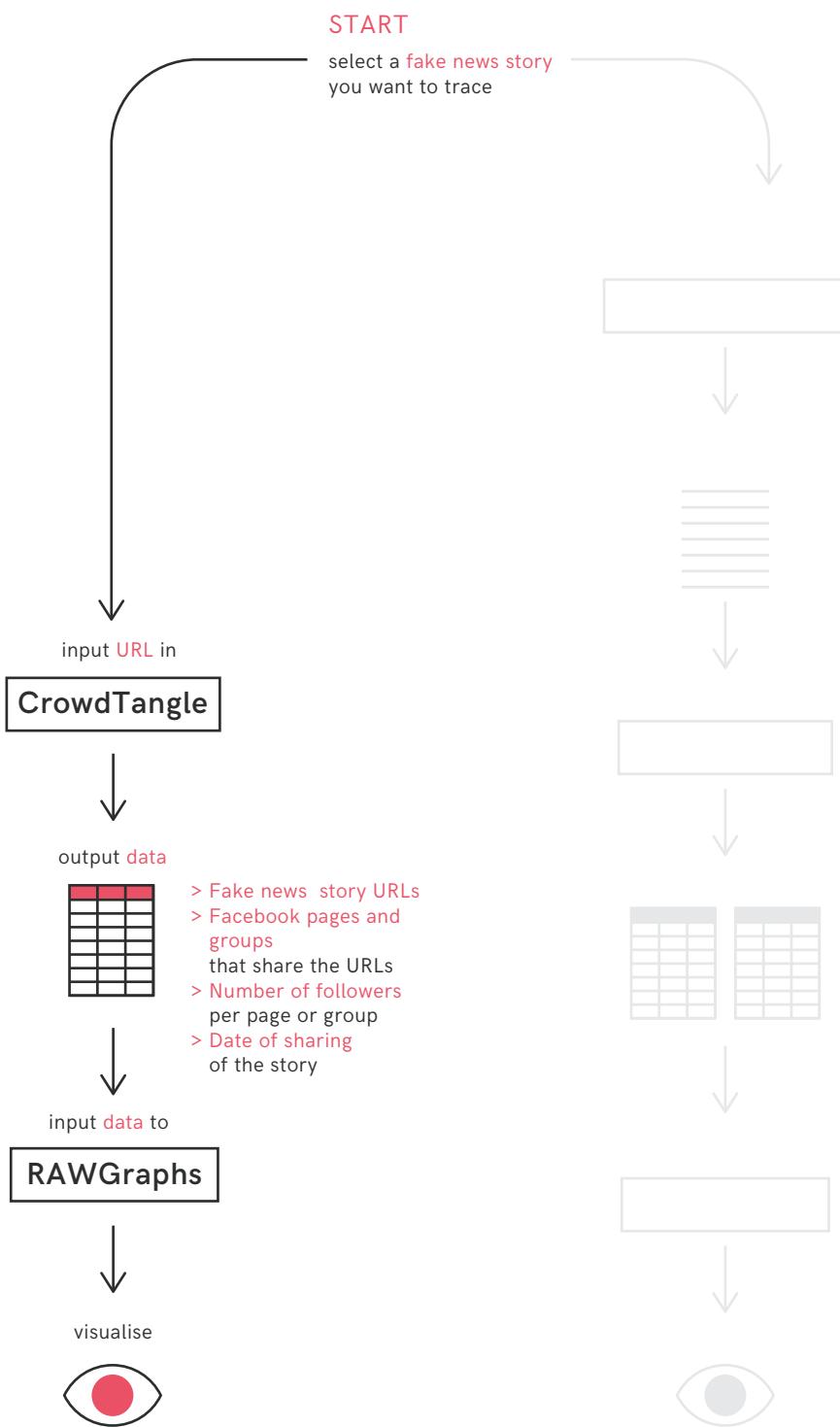
HOW MAY THE TRAJECTORY OF A FAKE NEWS STORY BE TRACED ON FACEBOOK?

BEFORE STARTING

For this recipe it is recommended that a fake news story is taken as a starting point and the URL or URLs on which it is published are identified. To illustrate this recipe we have selected as case studies two prominent stories about the 2016 US presidential elections, namely "Trump Offering Free One-Way Tickets to Africa & Mexico for Those Who Wanna Leave America," a story that exploits anti-immigrant sentiment and "Rage Against the Machine to Reunite and Release Anti Donald Trump Album," which exploits anti-Trump sentiment.

This recipe comes in two flavours. In step one you will learn to trace public Facebook pages and groups in which the original story URL is posted and plot them on a timeline (👁a). In step two this analysis will be extended to all URLs on which a story has been republished (👁b).





a

**HOW DOES THE STORY
"RAGE AGAINST THE
MACHINE TO REUNITE
AND RELEASE ANTI
DONALD TRUMP ALBUM"
TRAVEL ON FACEBOOK?**

a. IDENTIFY FACEBOOK PAGES AND GROUPS THAT SHARE A FAKE NEWS STORY VIA THE ORIGINAL URL

Public Facebook pages and groups that share a fake news story may be detected through a social media monitoring tool such as  CrowdTangle's browser extension.

- ◊ The names of these pages and groups, their  **followers'** count, the  **interactions** that they generate as well as the date of sharing of the story may be recorded in a spreadsheet per story URL.
- ◊ To explore the temporal dynamics of the circulation of the fake news story on Facebook, you may plot its trajectory across pages and groups on a timeline.   RAWGraphs can be used to create the base layer of the visualisation. Note which publics engage with the story as well as whether the moment of debunking of a story affects its circulation.^[1]
- ◊ To take the analysis one step further, a qualitative analysis of how fake news is enrolled by each of these pages to support issue work may be undertaken. This may be done by examining the context in which the stories are shared, i.e. whether they are shared uncritically or called out as fake as well as how they are framed in relation to the issues represented by the pages that share them. It is to be noted that such analysis might at times be difficult due to the fact that Facebook posts that share the most prominent fake news stories may be removed from the Facebook interface and API.

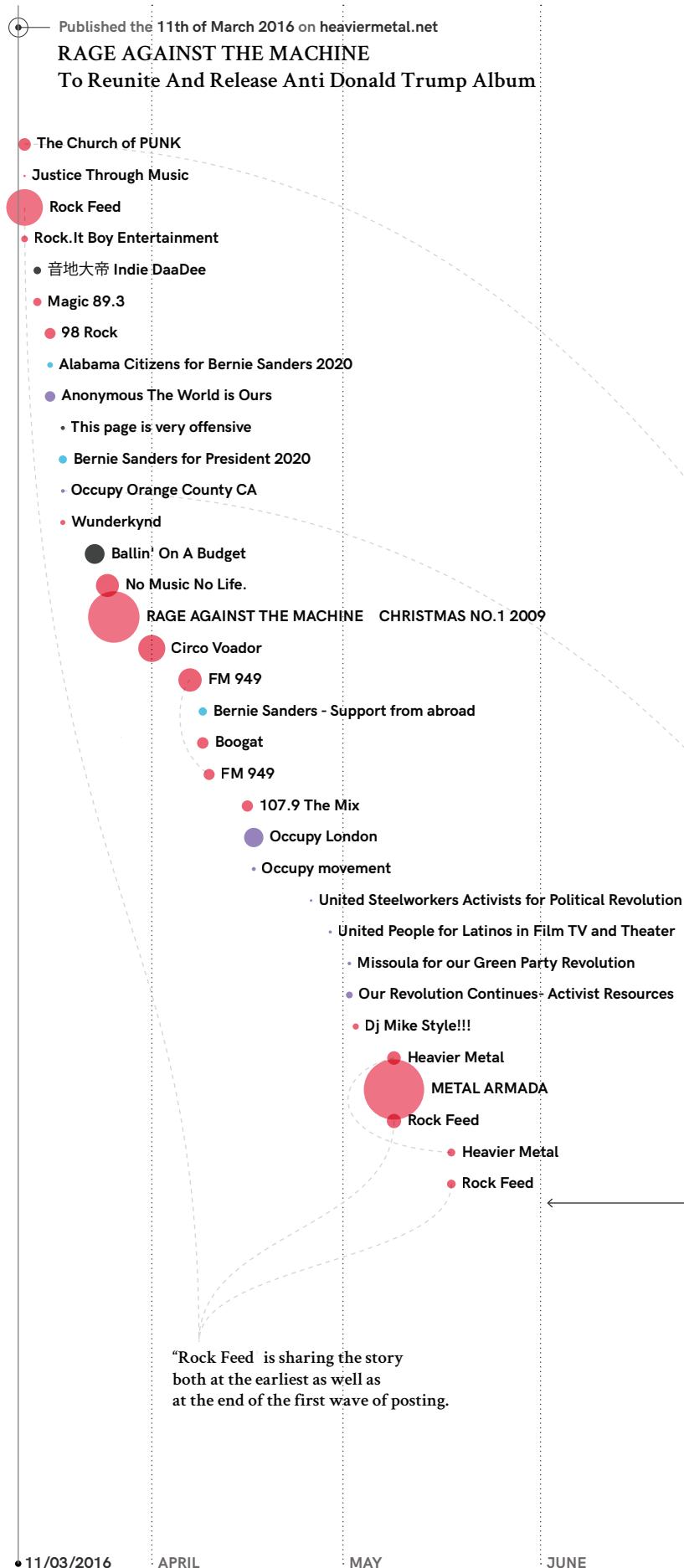
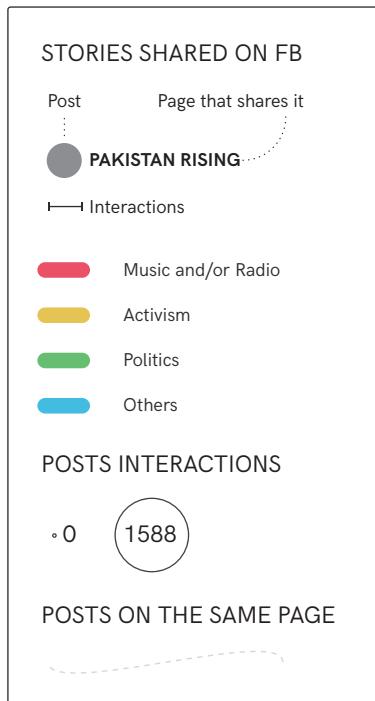
[1]

Please note that this analysis will not account for all instances of sharing of a fake news story URL on Facebook but only for the top 500 instances (per URL) of prominent sharing to public Facebook pages which CrowdTangle monitors. For more information see the note on CrowdTangle data on p.27.

HOW DOES THE STORY "RAGE AGAINST THE MACHINE TO REUNITE AND RELEASE ANTI DONALD TRUMP ALBUM" TRAVEL ON FACEBOOK?

Trajectory of "Rage Against the Machine to Reunite and Release Anti Donald Trump Album" story on Facebook pages and groups retrieved with CrowdTangle.

The story circulates best between March and June 2016 as satire amongst English language music and entertainment groups. It is revived in November after the US elections, when it is also picked up by Italian music and political pages.



Replace with "No prominent sharing of the story URL on public Facebook pages and groups from late May to early November according to CrowdTangle data"

The story reappears on Facebook on the 11th of November, shared by the page "Apartment Khunpa"

- Apartment Khunpa
- The Guitar Mag
- Banda Bassotti
- Marcello Belotti - Sinistra Italiana EU
- ROCK Mundial
- The Church of PUNK
- Radioactivo 98.5
- Los Angeles Punk Museum
- Veterans Against Rep. Ignorance
- FU Trump
- Occupy Orange County CA

JULY

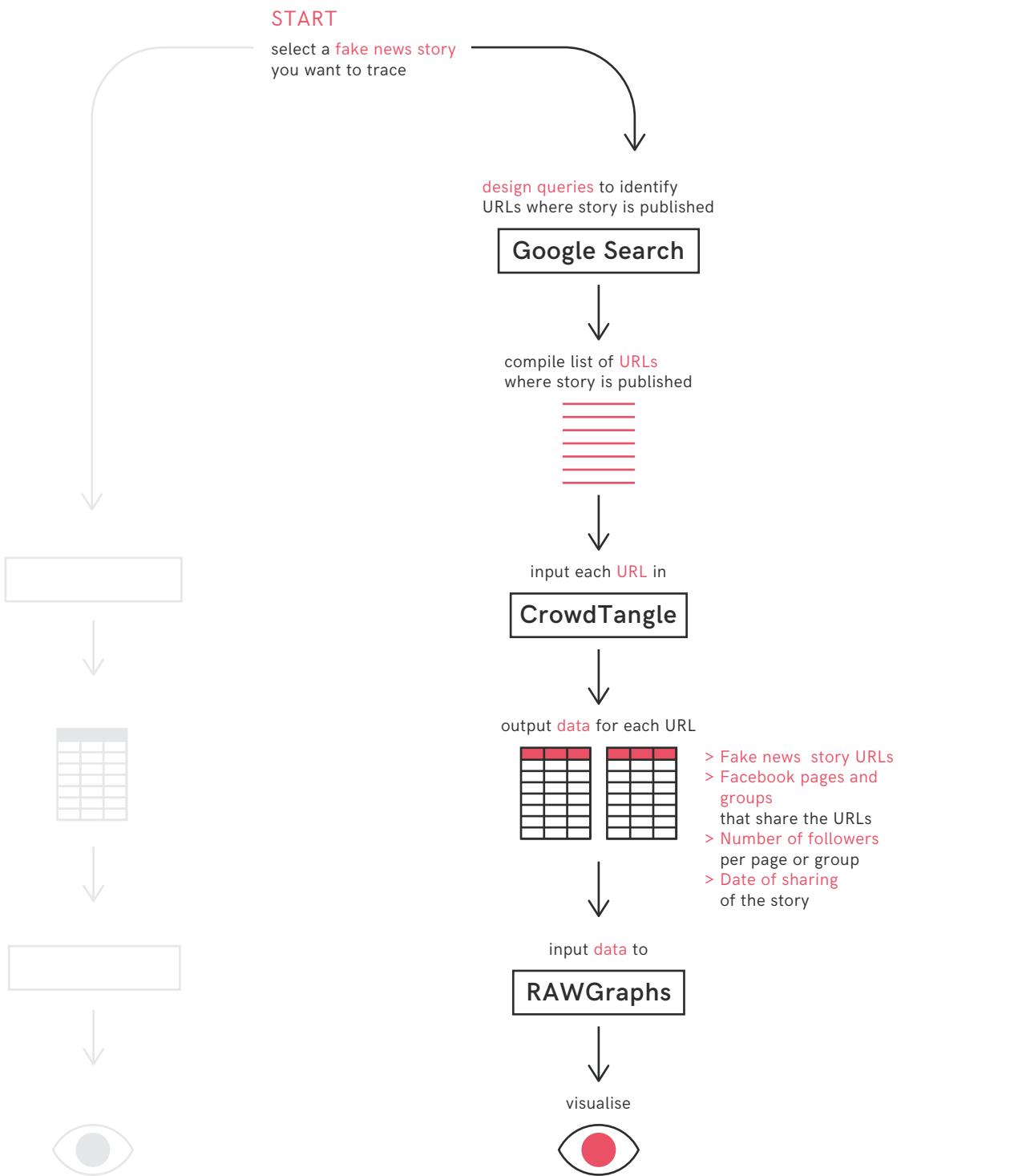
AUGUST

SEPTEMBER

OCTOBER

NOVEMBER

DECEMBER



b

HOW DOES THE STORY "TRUMP OFFERING FREE ONE-WAY TICKETS TO AFRICA & MEXICO FOR THOSE WHO WANNA LEAVE AMERICA" AND ITS DEBUNKED VERSIONS TRAVEL ON FACEBOOK?

b. IDENTIFY FACEBOOK PAGES AND GROUPS THAT SHARE ALL INSTANCES OF A FAKE NEWS STORY

As fake news stories may be republished by a number of sources, the previous analysis may be enriched by tracing the circulation not only of the original URL on which the chosen story is posted but all instances of story republication across a number of different sites.

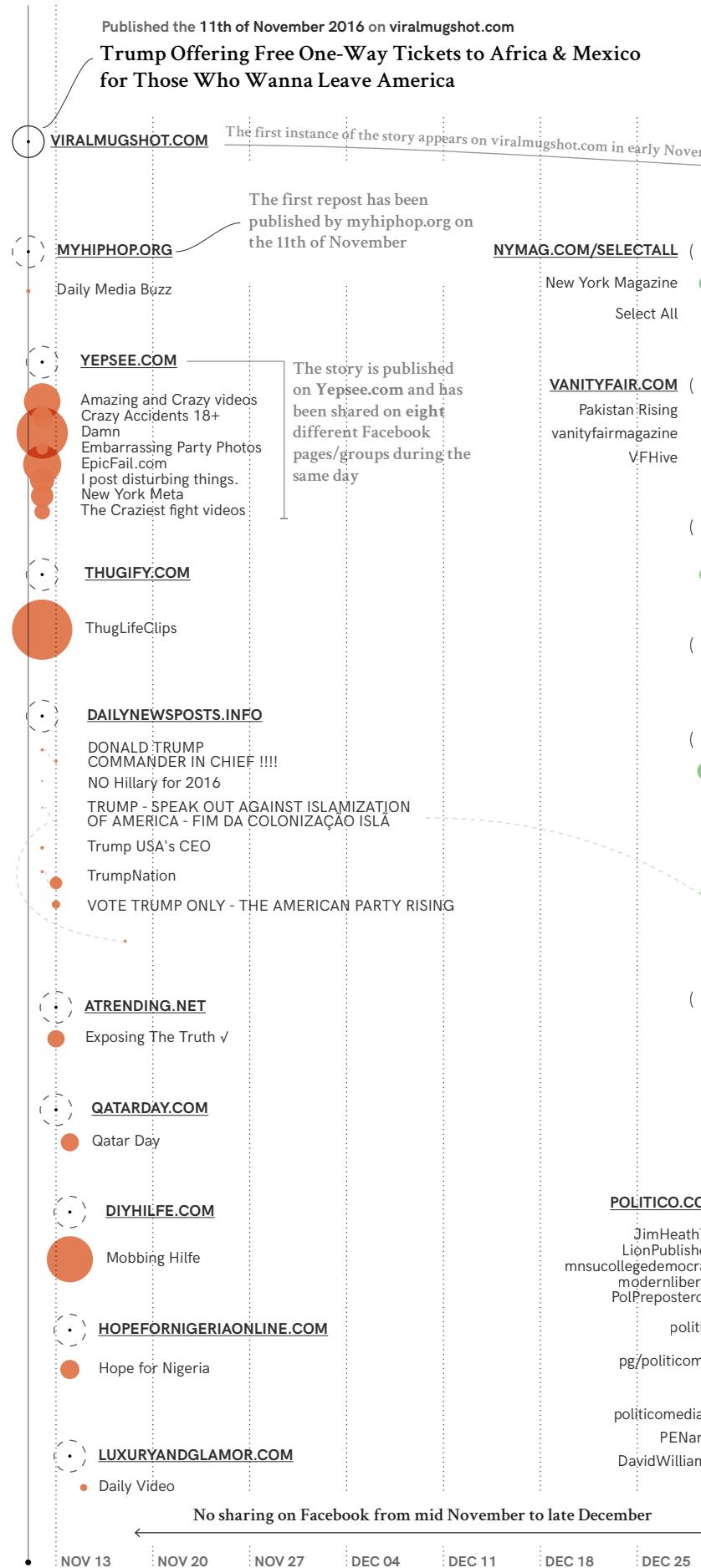
- ◊ To identify the websites which republish a story as well as those which debunk it, you may query the title of the fake news in a search engine of choice (e.g.  Google Web Search) using a research browser^[2] and extract the URLs corresponding to instances of republication and debunking of the story from the returned list of results.
- ◊ Query the resulting URLs in a social monitoring tool (such as  CrowdTangle) to get the list of Facebook groups and pages that prominently share the URLs corresponding to both the fake story and its debunked versions.^[3]
- ◊ You may plot these pages on a timeline to see whether different fake news sources spawn different story trajectories on Facebook and whether debunked versions are being acknowledged.

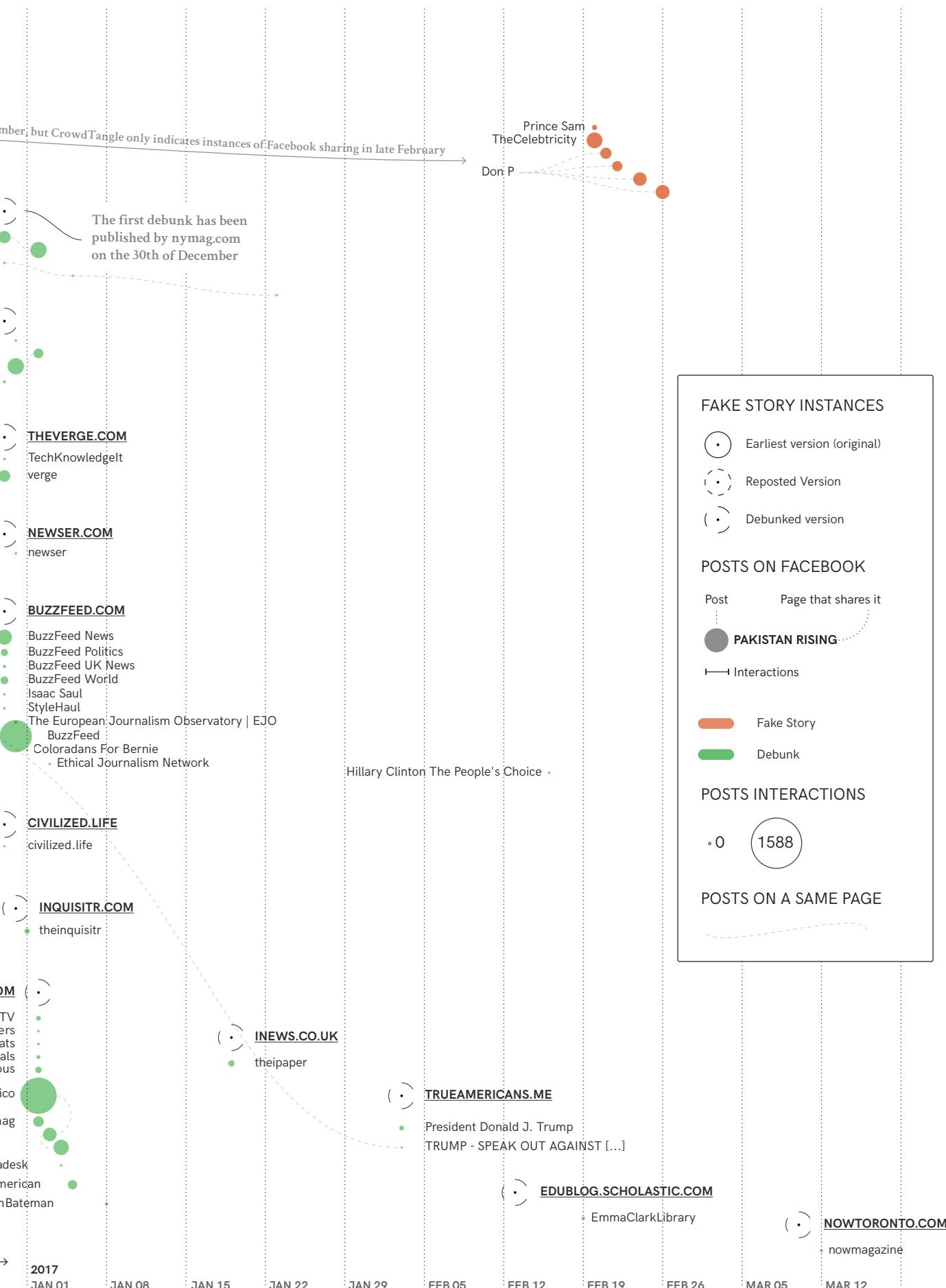
[2]
See instructions on how to set up a research browser in this video tutorial: <https://www.youtube.com/watch?v=bj65Xr9GkJM>

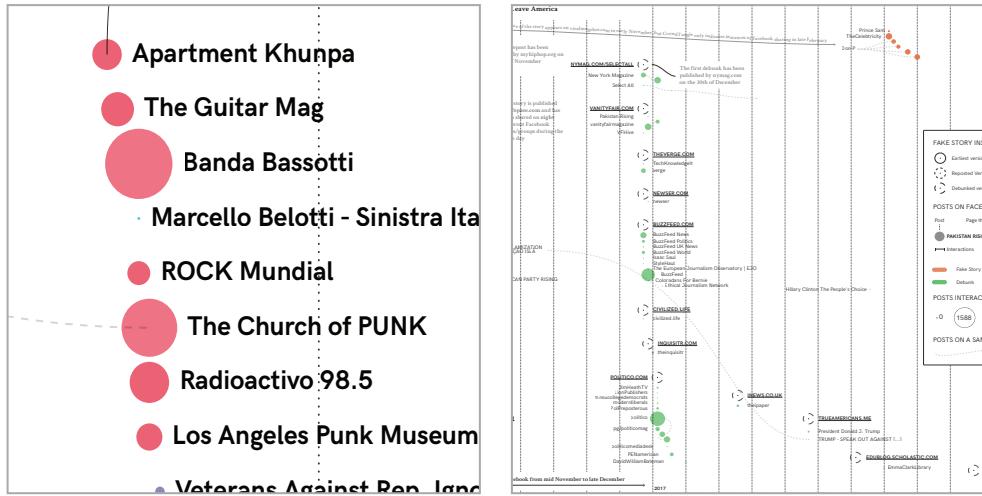
[3]
Please see the note on p.43 about the limitations of using CrowdTangle data.

HOW DOES THE STORY "TRUMP OFFERING FREE ONE-WAY TICKETS TO AFRICA & MEXICO FOR THOSE WHO WANNA LEAVE AMERICA" AND ITS DEBUNKED VERSIONS TRAVEL ON FACEBOOK?

Timeline of "Trump Offering Free One-Way Tickets to Africa & Mexico for Those Who Wanna Leave America" story and its sites of publication on the web and Facebook. The story is republished without critical context on multiple → **clickbait** sites in the week following its original publication. This gives the story multiple lives on Facebook. Its sharing on a fake news site animates political publics while its sharing on clickbait sites sees the story being recycled as clickbait by viral pages. The publics sparked into being by the fake news story and the debunked version thereof do not overlap.







CHAPTER 1 → RECIPE 2

SERVING SUGGESTIONS

This recipe may be used to understand the trajectory of a fake news story on Facebook, the different phases of its life cycle as well as key moments and intermediaries associated with its dissemination.



DO FACT-CHECKING INITIATIVES REACH THE PUBLICS OF FAKE NEWS ON FACEBOOK?

BEFORE STARTING

This recipe takes as a starting point a list of fake news stories. There are different ways of obtaining these lists – including starting with existing lists as well as creating your own. To illustrate this recipe we use an already existing list of 22 fake news stories about various political issues pertaining to the 2016 presidential elections in the US that generated most engagement on Facebook. These were identified by *BuzzFeed News*.

There are two steps to this recipe. The first is to identify URLs that circulate corrections or “debunking web pages” for each fake news story. The second is to explore how public Facebook pages engage with both fake news stories and their corresponding debunking web pages (☞b).

START

List of 22 URLs of political
fake news stories
Source: BuzzFeed News

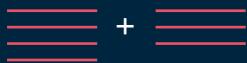
query
"Fake News Title 1" + Fake
"Fake News Title 2" + Fake
"..." in

Google Web Search

↓
Retain the top ranked URL
of a debunk per fake news story



merge all URLs in a single list



↓
input URLs in

CrowdTangle



output data



- > URLs of fake news story
or debunked version thereof
- > Facebook pages and groups
that share either the fake
news story or its correction
- > number of followers
per page or group

↓
input data to

RAWGraphs



visualise



**ARE DEBUNKING WEB
PAGES ACKNOWLEDGED
BY THE PUBLICS OF FAKE
NEWS?**

START

List of 22 URLs of political
fake news stories
Source: BuzzFeed News

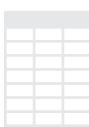
query
"Fake News Title 1" + Fake
"Fake News Title 2" + Fake
"..." in

Google Web Search

↓
Retain the top ranked URL
of a debunk per fake news story



↓
===== + =====



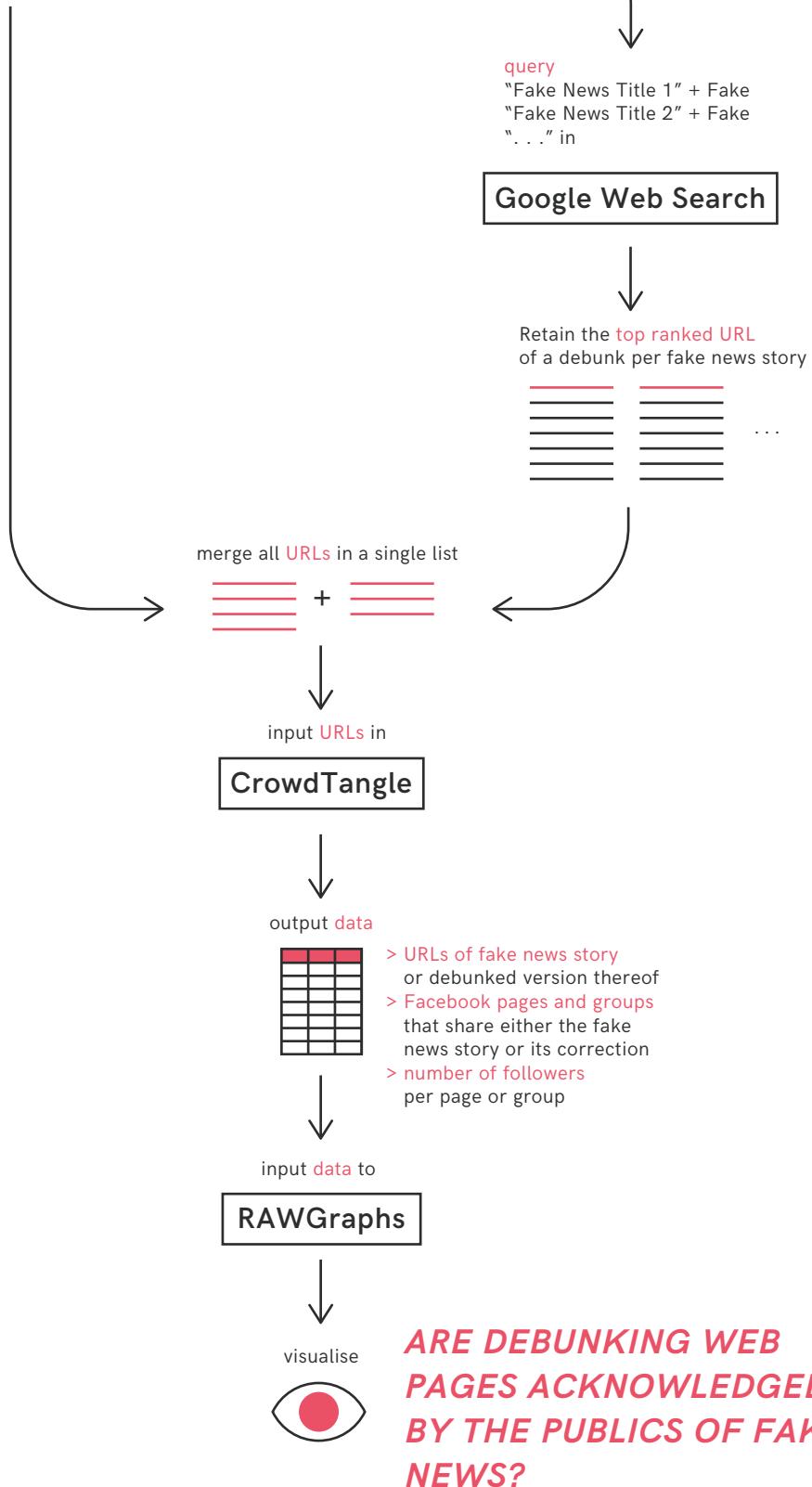
a. IDENTIFY WEB PAGES WHICH AIM TO DEBUNK FAKE NEWS STORIES

To identify prominent debunking web pages for a given fake news story you may use the  Google Web Search engine. In addition to this, you may also query fact-checking sites for keywords describing a fake news story.

- ◊ In order to find corrections of fake news articles queries need to be designed for each fake news item in your list. One strategy would be to use the title of the story in quotation marks followed by the word “fake” (e.g. “Trump Offering Free One-Way Tickets to Africa & Mexico for Those Who Wanna Leave America’ fake”).
- ◊ You may use the search engine ranking as an indication of salience of correction and select the highest ranked URLs corresponding to a corrected version of the fake news story in question.
- ◊ The result of this step is a list of URLs containing the most highly ranked debunking web pages per fake news story.

START

List of 22 URLs of political
fake news stories
Source: BuzzFeed News



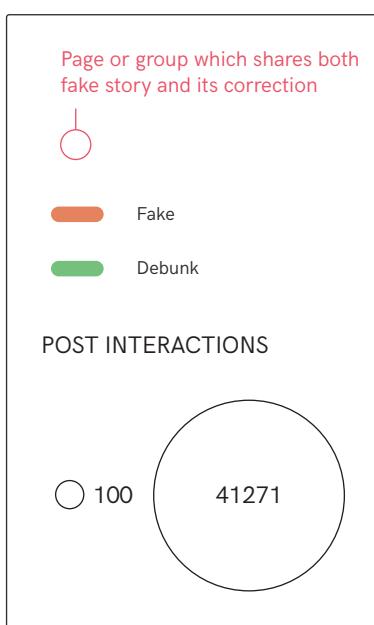
b. MAP THE OVERLAP BETWEEN THE PUBLICS OF FAKE NEWS STORIES AND WEB PAGES WHICH AIM TO DEBUNK THEM

Public Facebook pages and groups that prominently share both fake news stories as well as web pages which aim to debunk them may be detected through a social media monitoring tool such as  CrowdTangle's browser extension.

- ◊ To explore whether the debunking web pages are acknowledged by the publics which share the fake news stories, identify whether there is an overlap between the public Facebook pages and groups that share fake news stories and those debunking web pages issued in response.
- ◊ This may be illustrated by means of a → **circle packing** visualisation. You may use  RawGraphs for this operation.

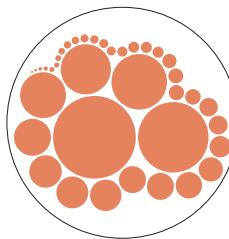
ARE DEBUNKING WEB PAGES ACKNOWLEDGED BY THE PUBLICS OF FAKE NEWS?

Fake news pages and debunking web pages have different publics on Facebook. Only six of the public pages that share fake news stories have acknowledged web pages which aim to debunk them in our CrowdTangle dataset. While Google looks to prioritise debunking web pages, on Facebook it is fake news stories that circulate better. While both progressive and conservative pages share fake news stories it is primarily progressive Facebook pages and those pertaining to journalists and fact-checking initiatives that share web pages which aim to debunk fake news stories.



Rage Against the Machine To Reunite And Release Anti Donald Trump Album

Published the 11th of March 2016 on heaviermetal.net



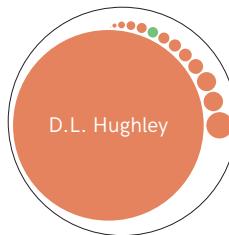
Actor Bill Murray Announces 2016 Presidential Run

Published the 3rd of October 2016 on www.abcnews.com.co



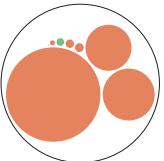
Rupaul claims Trump touched him inappropriately in the 1990s

Published the 14th of October 2016 on worldnewsdailyreport.com



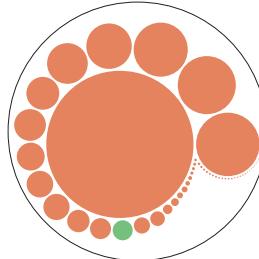
African Billionaire Will Give \$1 Million To Anyone Who Wants To Leave America if Donald Trump is Elected President

Published the 3rd of November 2016 on www.empireherald.com



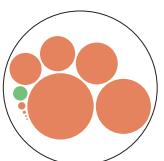
Hillary Clinton In 2013: "I Would Like To See People Like Donald Trump Run For Office; They're Honest And Can't Be Bought"

Published the 17th of October 2016 on conservativestate.com



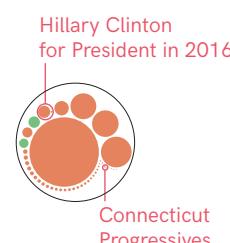
Graham Says Christians Must Support Trump or Face Death Camps

Published the 23th of July 2016 on www.bizstandardnews.com



Obama Signs Executive Order Declaring Investigation Into Election Results; Revote Planned For Dec. 19th

Published the 12th of December 2016 on www.abcnews.com.co



Pope Francis Shocks World, Endorses Hillary Clinton for President, Releases Statement

Published the 25th of July 2016 on www.kypo6.com



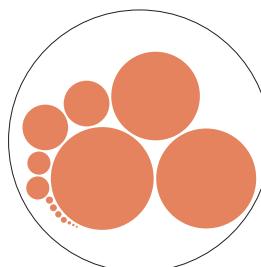
Mike Pence: "Sarah Palin Is My Role Model For Beautiful, Smart American Women"

Published the 18th of September 2016
on www.politicops.com



President Obama Confirms He Will Refuse To Leave Office If Trump Is Elected

Published the 7th of September 2016
on www.burrardstreetjournal.com



Trump Offering Free One-Way Tickets to Africa; Mexico for Those Who Wanna Leave America

Published the 11th of November 2016
on www.tmzhiphop.com



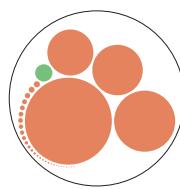
Obama passed law for grandparents to get all their grandchildren every weekend

Published the 8th of March 2016
on www.react365.com



WHOA! Hillary Caught On Hot Mic Trashing Beyonce' With RACIAL SLURS!

Published the 5th of November 2016
on www.thelastlineofdefense.org



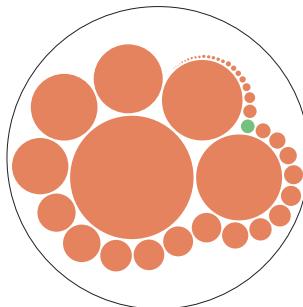
BREAKING Romanian Hacker With Access To Clinton Emails Found Dead In Jail Cell

Published the 6th of July 2016
on Christian Times



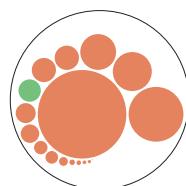
ISIS Leader Calls for American Muslim Voters to Support Hillary Clinton

Published the 11th of October 2016
on worldnewsdailyreport.com



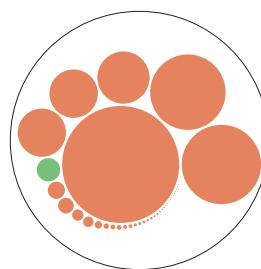
Trump Claims America Should Never Have Given Canada Its Independence

Published the 29th of July 2016
on www.burrardstreetjournal.com



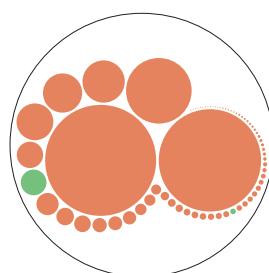
Van Full Of Illegals Shows Up To Vote Clinton At SIX Polling Places, Still Think Voter Fraud Is A Myth?

Published the 5th of November 2016
on www.thelastlineofdefense.org



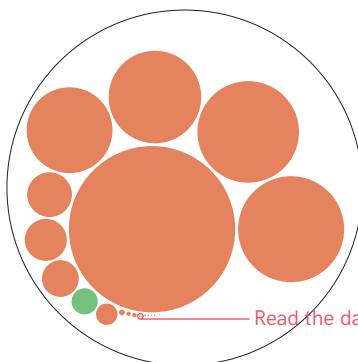
FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide

Published the 15th of October 2016
on www.usanewsflash.com



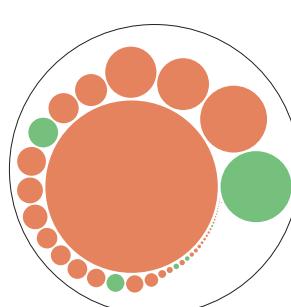
Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement

Published the 26th of September 2016
on [Ending The Fed](http://EndingTheFed.org)



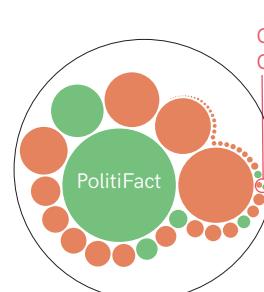
Obama Signs Executive Order Banning The Pledge Of Allegiance In Schools Nationwide

Published the 11th of December 2016
on www.abcnews.com.co

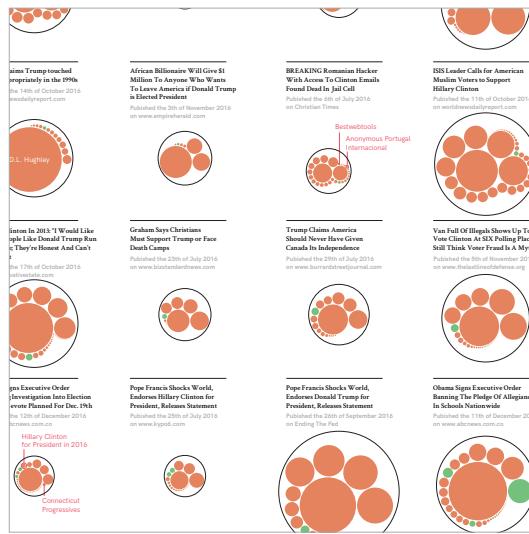


Donald Trump Protester Speaks Out: I Was Paid \$3,500 To Protest Trump's Rally

Published the 22th of November 2016
on www.abcnews.com.co



CHAPTER 1 → RECIPE 3



SERVING SUGGESTIONS

This recipe may be used as one way to assess the impact of attempts to debunk fake news by examining whether debunking responses to fake news are acknowledged on the platform that generates most engagement with fake news, Facebook, and by the particular publics which share and engage with fake news.

Chapter 2

TRACING THE CIRCULATION OF FAKE NEWS ON THE WEB

Where do fake news stories originate?

By which sites are they first disseminated?

Which are the most visible sources related
to a fake story? When and by whom are
they mentioned?

Introduction - Fake news are not just “false news”. They are interesting not so much because their content or form are different from that of “authentic news”, but because they travel as much as (and sometimes more than) mainstream news. If a blog claims that Pope Francis endorses Donald Trump, it's just a lie. If the story is picked up by dozens of other blogs, retransmitted by hundreds of websites, cross-posted over thousands of social media accounts and read by hundreds of thousands, then it becomes fake news.

The following recipes investigate the circulation of fake news for two reasons. Firstly, from a political point of view many have expressed disappointment that techniques and tactics commonly used to tackle fake news have not lived up to expectations. Fact-checking and debunking, in particular, often do not succeed in preventing the circulation of hoaxes and rumors. On the contrary: they can inadvertently

contribute to making them even more visible on the web. A better understanding of how fake news travels online can help to inform responses that are more attuned to the phenomena.

Secondly, from a methodological point of view, as there is no “ontological” difference between fake and authentic news, studying fake news circulation can help us understand more about how other kinds of news travel.

This recipe comes in two flavours. Firstly, we propose a manual variant which can be executed without the need for any particular tool or technical knowledge. This variant is easy to execute but also time consuming. It is based on a search for web pages referring to fake news stories and on the manual identification of the dates of their publication and the sources that they cite. Secondly, we propose a semi-automated version which allows this approach to be scaled to more pages, but demands more technical skills and may require more manual verification.



WHERE DO FAKE NEWS ORIGINATE? BY WHAT SITES ARE THEY FIRST RETRANSMITTED?

BEFORE STARTING

For this recipe you will need to choose a fake news story whose circulation you would like to trace. The more distinctive your story is, the easier it will be to follow its circulation. In the example, we focus on the “Pope Endorses Trump” story that was widely circulated around the US Elections.

START

open Chrome browser
in **incognito** mode



query

[Pope AND endors* AND (Trump OR Clinton)
AND NOT (hoax OR "fake news" OR lie OR debunk)]

Google Web Search



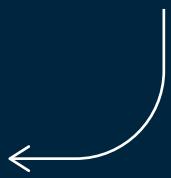
compile a list with
the top **resulting URLs**



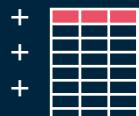
a

input **URLs** in a spreadsheet

- > The page **advocating** or **debunking** the fake news item
- > **Rank** in the search engine results
- > the **date of publication** of each page



record **all sources** cited in the occurrences
of your story, noting down:



- + if they are cited as **evidence** or **counter-evidence**
- + if they are cited through **hyperlink**, **textual reference** or **copying/pasting** of its content



Make sure you visit both the **results** of
your initial search and **all the sources**
cited by those results



extract **network** of instances
and references with

Table2Net



Gephi



b

HOW DO THE OCCURRENCES OF THE "POPE ENDORSES TRUMP" STORY CITE EACH OTHER?



Plot **URLs** on a **timeline** with

Graph Recipes



c

WHAT IS THE LIFE OF THE "POPE ENDORSES TRUMP" STORY ACCORDING TO PAGES IN SEARCH ENGINE RESULTS?

START

open Chrome browser
in **incognito** mode



query

[Pope AND endors* AND (Trump OR Clinton)
AND NOT (hoax OR "fake news" OR lie OR debunk)]

Google Web Search

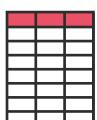


compile a list with
the top **resulting URLs**

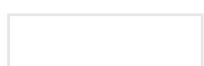


a

input **URLs** in a spreadsheet



- > The page **advocating** or **debunking** the fake news item
- > **Rank** in the search engine results
- > the **date of publication** of each page



a. IDENTIFY AND ANALYSE THE OCCURRENCES OF YOUR STORY IN SEARCH ENGINE RESULTS

Successful fake news stories always appear on several web pages. In the first step of this recipe, you will identify and collect information about these occurrences.

- ◊ Identify the occurrences of your story by querying one or more search engines (we used  Google Web Search). Since fake news stories evolve while circulating, consider keywords that may capture different variants of the story.
- ◊ Rely on search engines to rank results by relevance and concentrate on the first results (under the working assumption that they are the ones that circulated the most).
- ◊ To avoid "filter bubbles" and personalised results, consider using a dedicated research browser.^[1]
- ◊ Be aware that search engines may give more visibility to debunkers than to original sources – and be careful not to overestimate the circulation of debunkers based on this. Also bear in mind that with this approach you see the phenomena “through the eyes” of the search engine that you selected – which will become a part of your story or research.
- ◊ Record all relevant metadata for each relevant search result. You can collect as many variables as you like, but make sure to characterise how the page refers to the story (e.g. as a reliable news source, or as a problematic claim to be debunked), as well as some indicator of its “visibility” (e.g. the rank in the search results).

[1]

See instructions on how to set up a research browser in this video tutorial: <https://www.youtube.com/watch?v=bj65X-r9GkJM>

START

open Chrome browser
in **incognito** mode



query

[Pope AND endors* AND (Trump OR Clinton)
AND NOT (hoax OR "fake news" OR lie OR debunk)]

Google Web Search

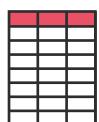


compile a list with
the top resulting URLs



a

input URLs in a spreadsheet



- > The page **advocating** or **debunking** the fake news item
- > **Rank** in the search engine results
- > the **date of publication** of each page



record **all sources** cited in the occurrences
of your story, noting down:



- + if they are cited as **evidence** or **counter-evidence**
- + if they are cited through **hyperlink**, **textual reference** or **copying/pasting** of its content



Make sure you visit both the **results** of
your initial search and **all the sources**
cited by those results



extract **network** of instances
and references with

Table2Net

→

Gephi

→

visualise



b

**HOW DO THE
OCCURRENCES OF THE
"POPE ENDORSES
TRUMP" STORY CITE
EACH OTHER?**



b. EXTRACT THE NETWORK OF REFERENCES AROUND YOUR STORIES

Fake news is supported or opposed through a network of references: websites that share rumours cite other pages to support their claims, while debunking initiatives flag toxic websites or refer to sources denying them.

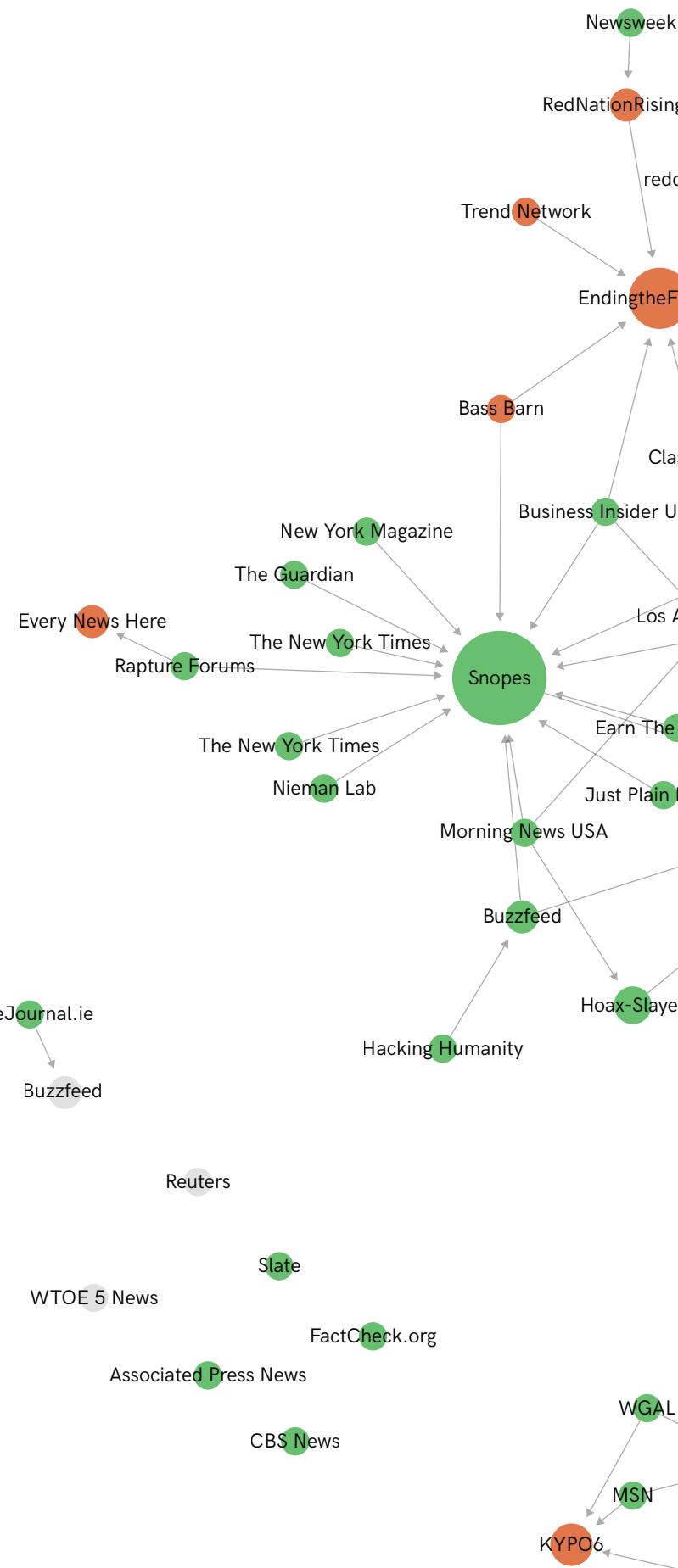
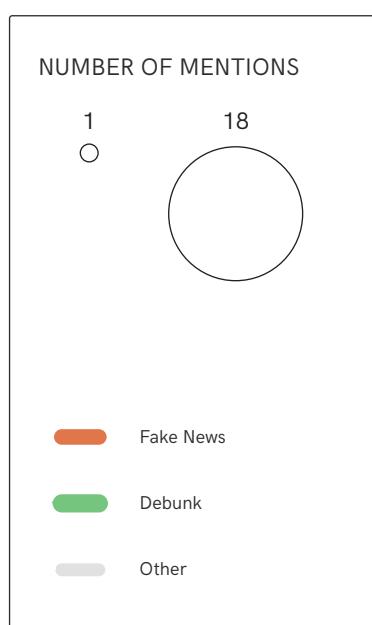
In this step, we will trace the network of references of a specific fake news story.

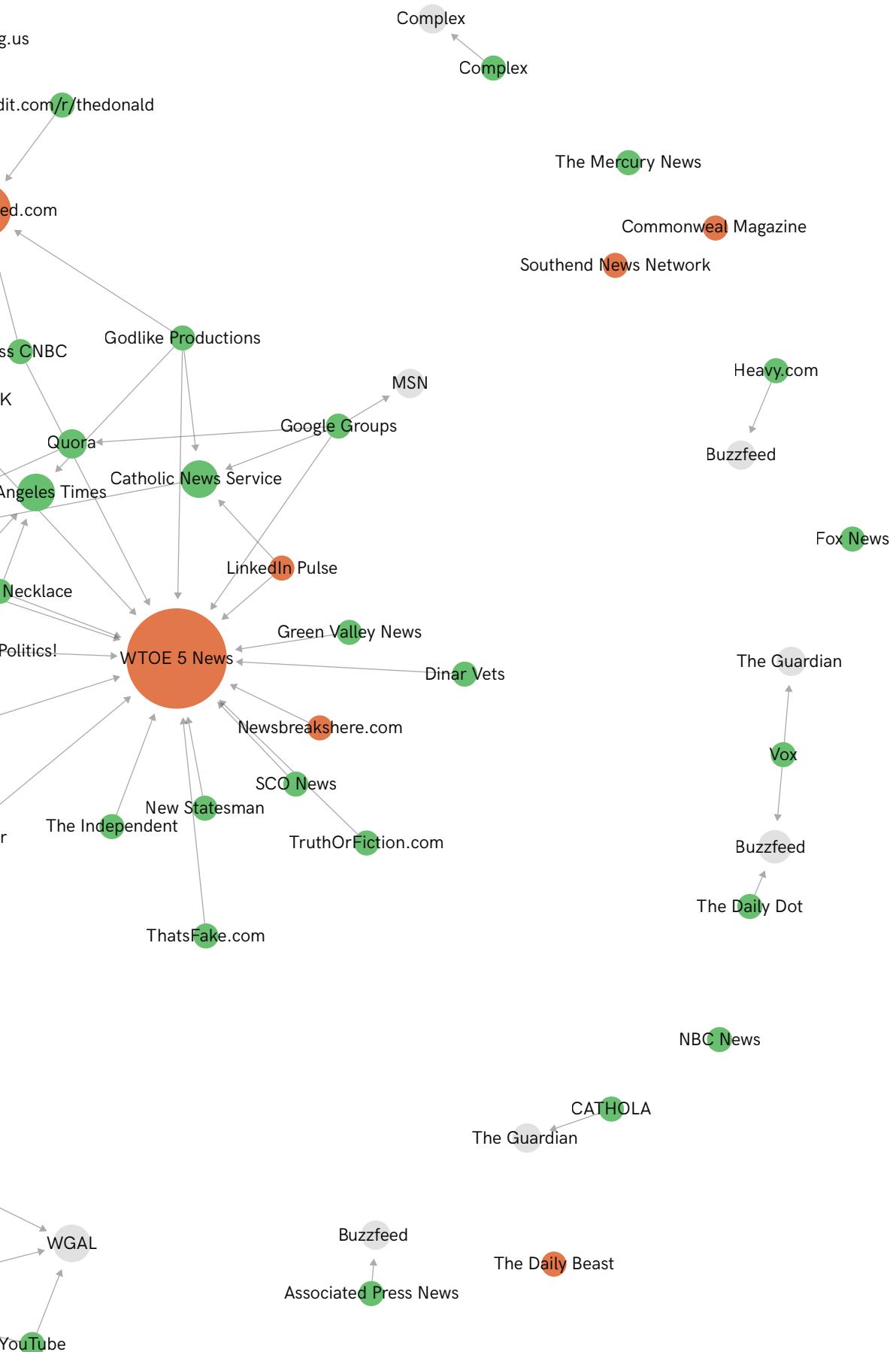
- ◊ Record all sources cited in the occurrences of your story. For each, note down if the source is cited as evidence or counter-evidence and if it is cited through a hyperlink (e.g. <http://snopes.com>), a textual reference (e.g. “the website WTOE 5 News”) or copying and pasting its content.
- ◊ Make sure you visit not only all the results of your initial search, but also all the sources cited by those results.
- ◊ Extract the networks of occurrences and references (you can use  Table2Net).
- ◊ Visualise the network (using  Gephi, for instance), applying a force-directed layout; sizing the nodes according to the number of citations they receive; and colouring the nodes according to how they report the story (advocating or debunking).

HOW DO THE OCCURRENCES OF THE "POPE ENDORSES TRUMP" STORY CITE EACH OTHER?

Network of cross-references between the pages mentioning the "Pope Endorses Trump" story.

In the image the nodes represent the different pages on which the fake story appears. The comparison of the colour of the nodes (which indicates whether the page affirm or debunks the story), their size (which indicates the number of citations received by the page) and their number reveal the great visibility of the debunking and neutral pages compared to websites that spread the fake story as authentic.





START

open Chrome browser
in **incognito** mode



query

[Pope AND endors* AND (Trump OR Clinton)
AND NOT (hoax OR "fake news" OR lie OR debunk)]

Google Web Search

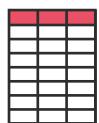


compile a list with
the top resulting URLs



a

input URLs in a spreadsheet



- > The page **advocating** or **debunking** the fake news item
- > **Rank** in the search engine results
- > the **date of publication** of each page



record **all sources** cited in the occurrences
of your story, noting down:



- + if they are cited as **evidence** or **counter-evidence**
- + if they are cited through **hyperlink**, **textual reference** or **copying/pasting** of its content



Make sure you visit both the **results** of
your initial search and **all the sources**
cited by those results



extract **network** of instances
and references with

Table2Net



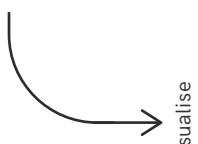
import **data** in

Gephi



Plot **URLs** on a timeline with

Graph Recipes



c

**WHAT IS THE LIFE OF
THE "POPE ENDORSES
TRUMP" STORY
ACCORDING TO PAGES
IN SEARCH ENGINE
RESULTS?**

c. VISUALIZE THE FAKE NEWS INSTANCES OVER TIME

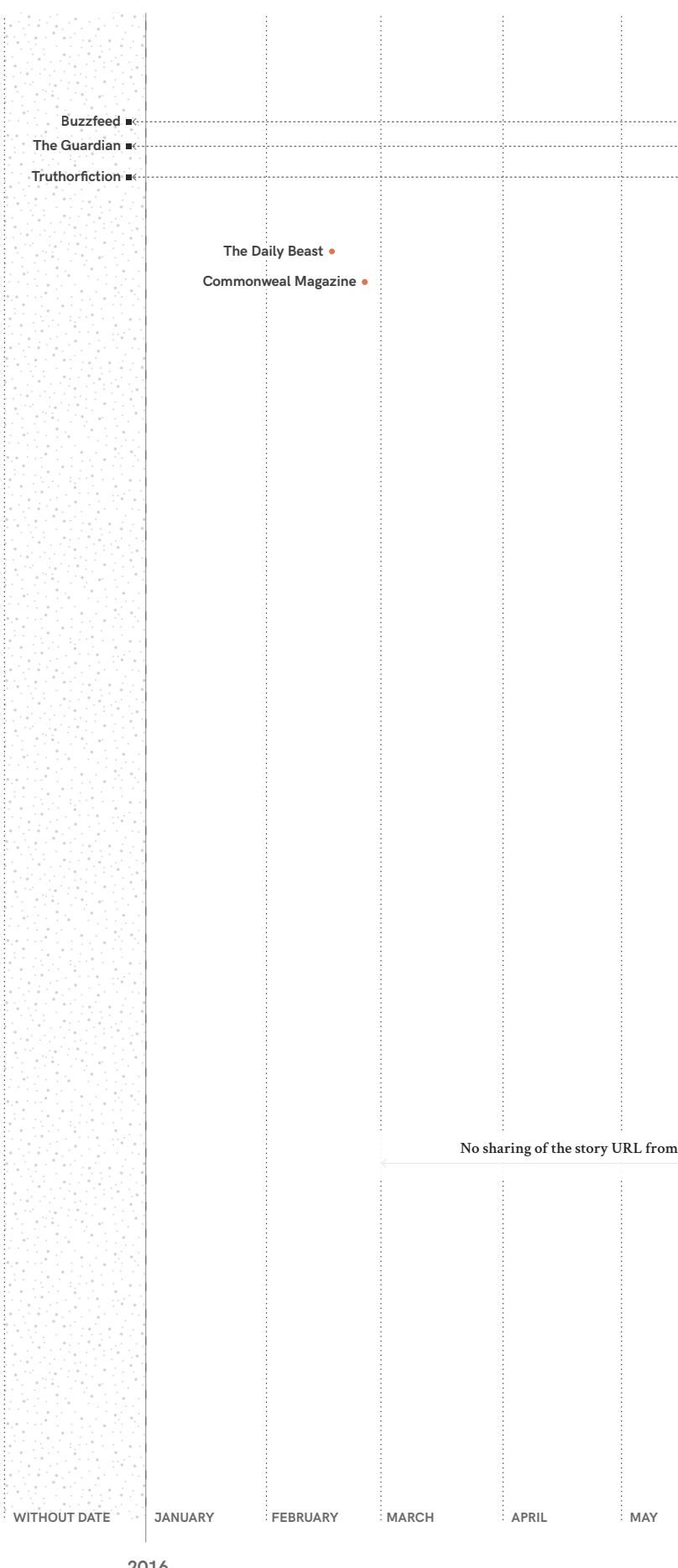
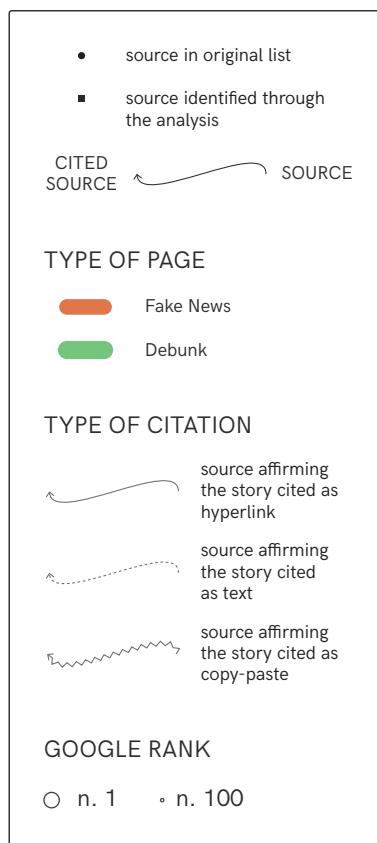
The network extracted in the previous step can help you understanding not only who cited whom, but also how and in which direction your fake news travelled. To reveal the circulation use the dates that you collected in the first step of this recipe.

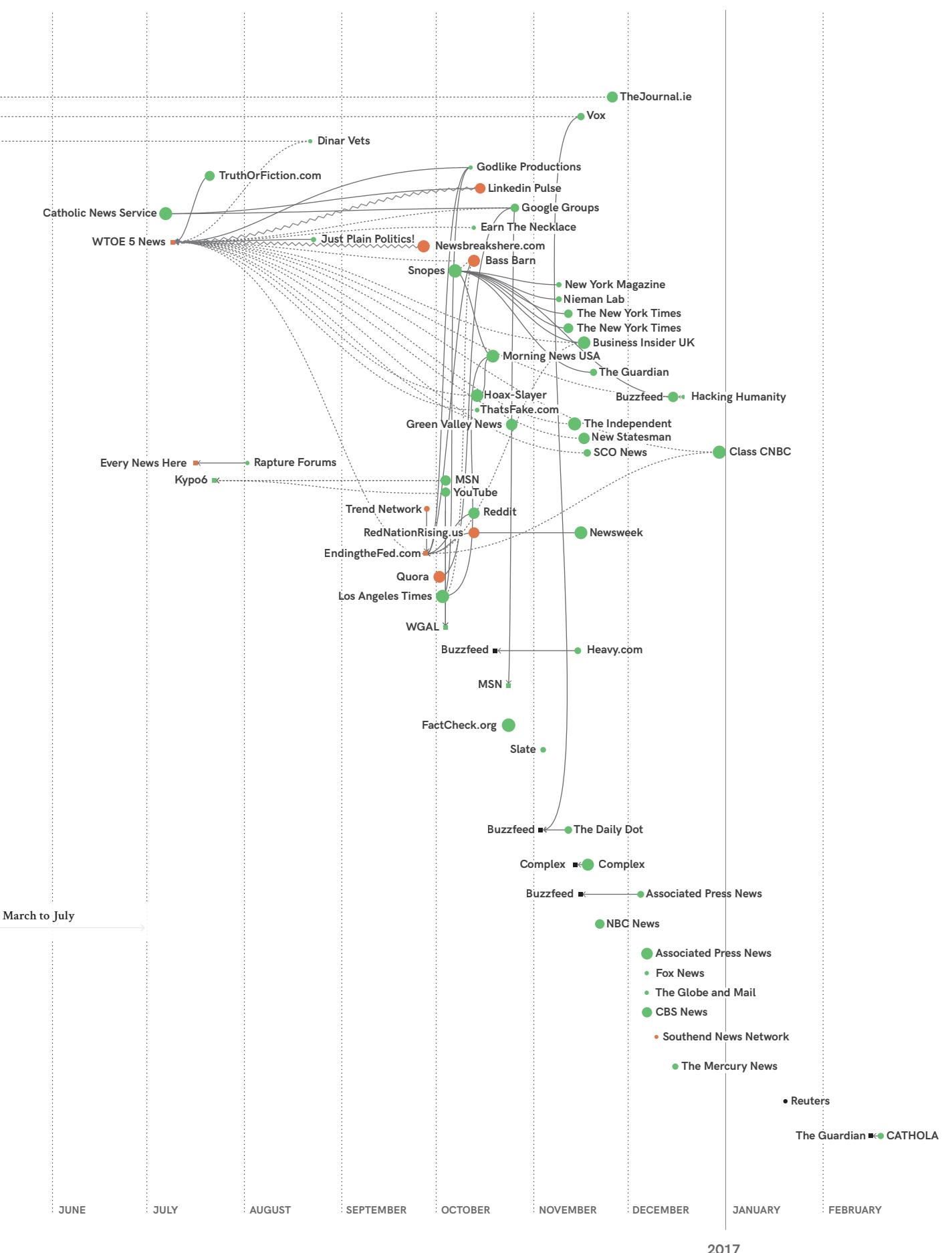
- ◊ Arrange the network of instances extracted in the previous step chronologically. You can use different visual styles to represent the different kind of citations (to do so, we used a custom script of  Graph Recipes tools).

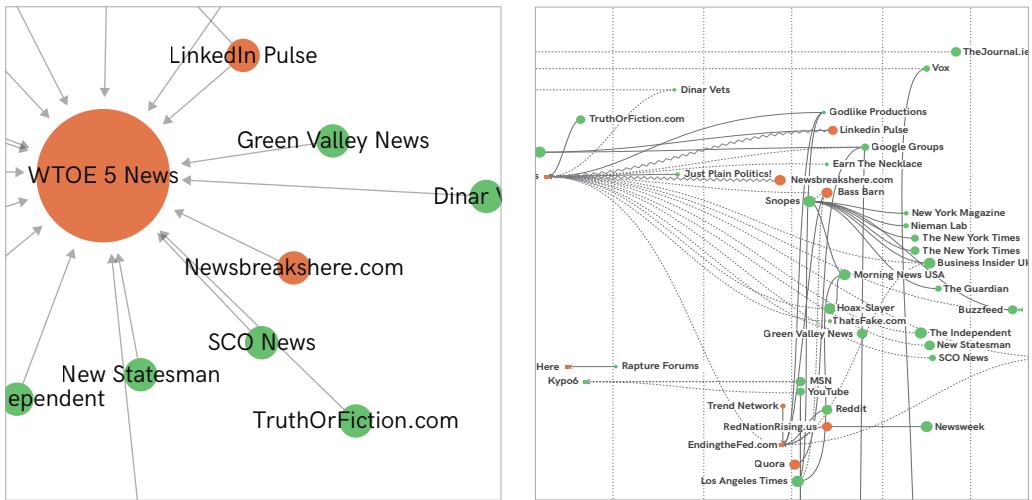
WHAT IS THE LIFE OF THE "POPE ENDORSES TRUMP" STORY ACCORDING TO PAGES IN SEARCH ENGINE RESULTS?

Chronological network of the cross-references between the pages mentioning the "Pope Endorses

Trump" story. In this image, the colour of the nodes indicates whether the page affirms or debunks the story and the type of line indicates how different pages cite each other. The high presence of dotted lines going from green to orange or gray nodes shows how debunking initiatives tend to mention original sources but not link to them. This technique is used to flag fake websites without increasing their online visibility by explicitly linking to them.







SERVING SUGGESTIONS

This recipe may be used to repurpose data obtained through search engine results in order to identify and follow the different occurrences of a given fake news story as it is cited and referenced by different online sources, as well as to retrace its circulation over time. You will also be able to see when the debunking activities took place, who promoted them and what effect this had on the circulation of fake news stories and web pages.



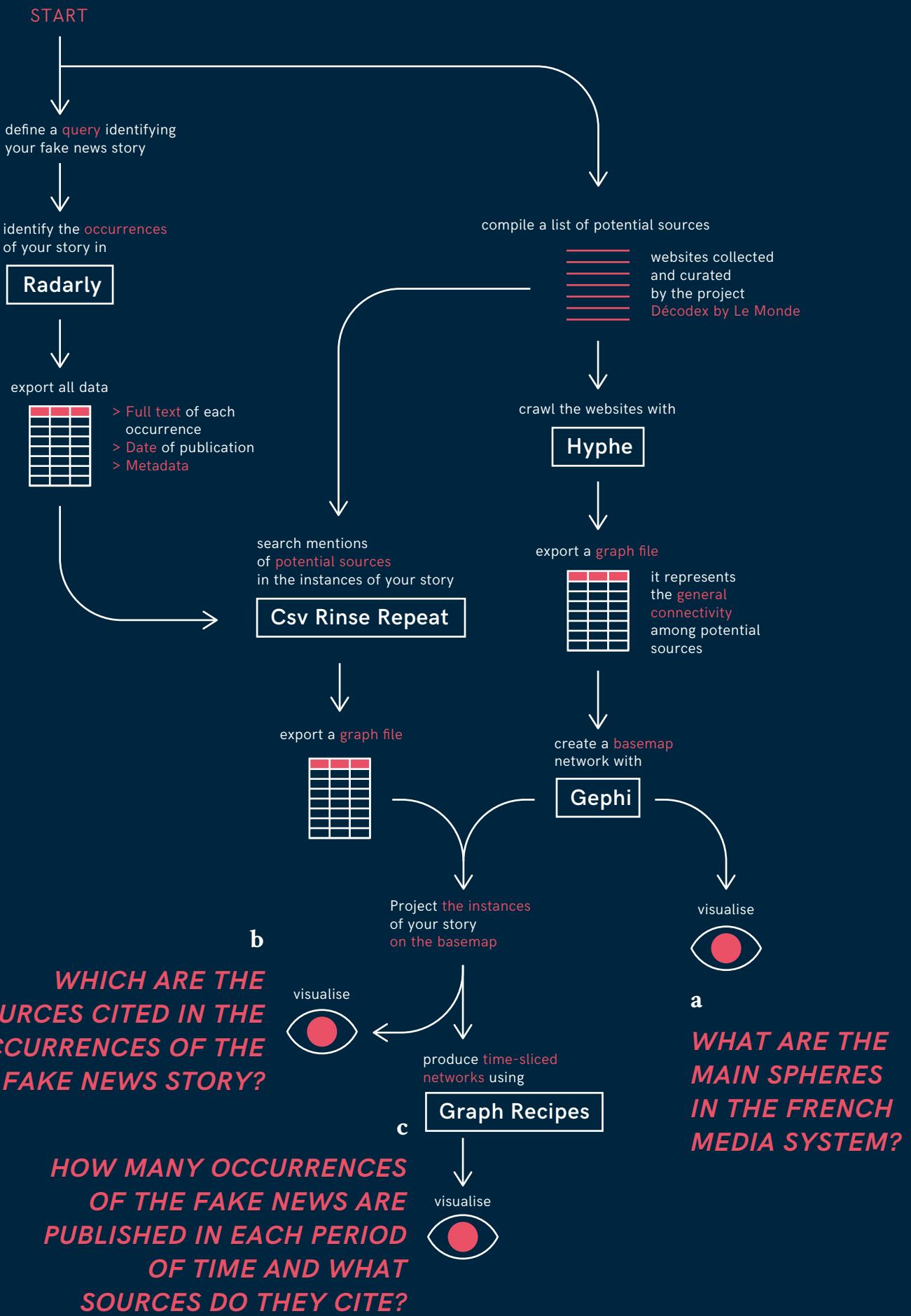
WHICH ARE THE MOST VISIBLE SOURCES RELATED TO A FAKE STORY? WHEN AND BY WHOM ARE THEY MENTIONED?

BEFORE STARTING

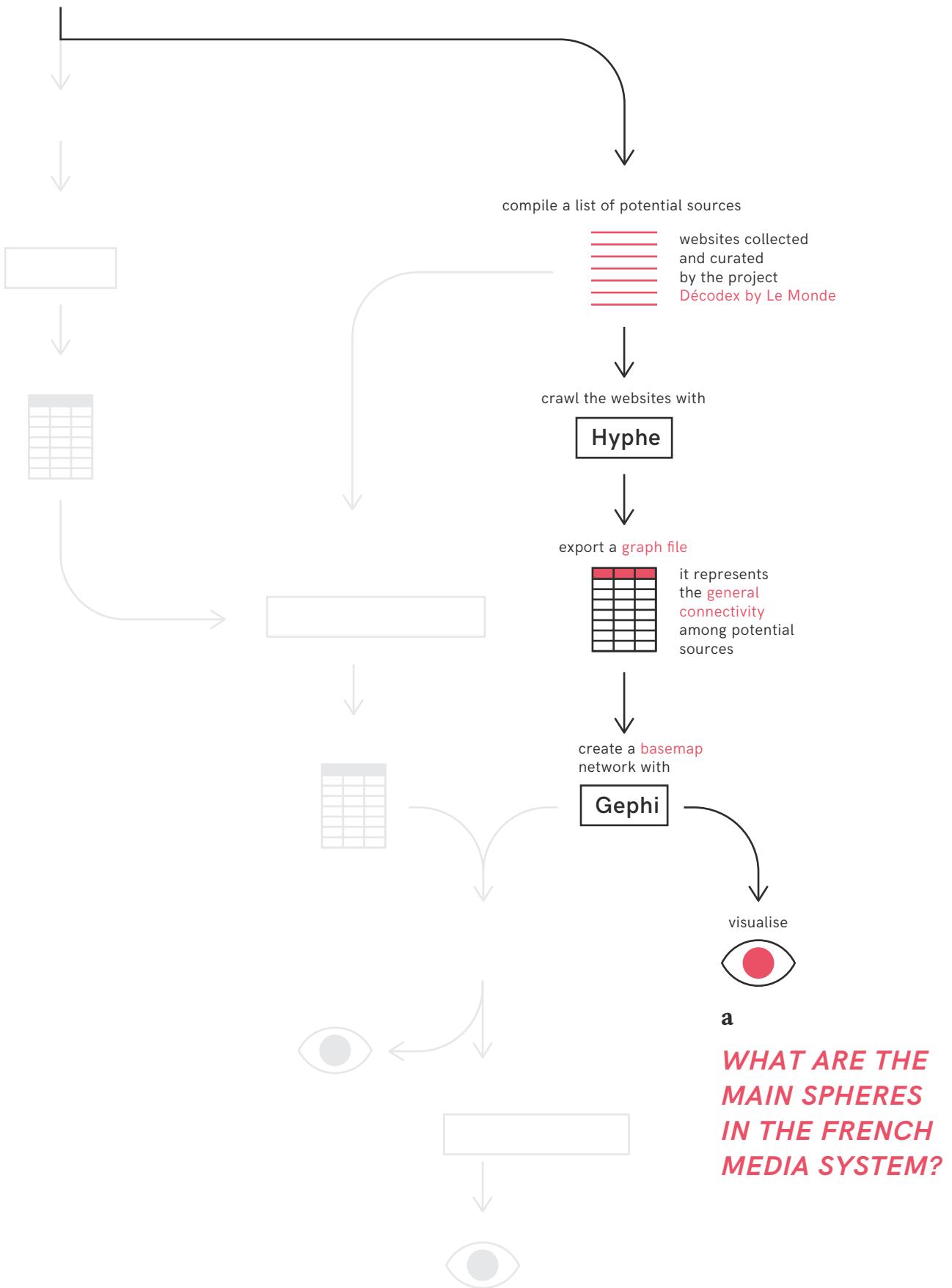
This recipe enables a scaling up of the approach presented in the previous recipe, but requires a bit more technical knowledge, as well as some bigger datasets. In particular, you will need to have access to:

- ◊ A web archive (we used  Radarly by Linkfluence).
- ◊ A list of all the possible web sources in which your chosen fake news story may have appeared (we used the list curated by  Le Monde Décodex).

To illustrate this recipe, we focus on a false story that circulated during the 2017 French presidential election and referred to the presumed homosexuality of Emmanuel Macron.



START



a. DEFINE A BASE MAP OF NEWS PROVIDERS

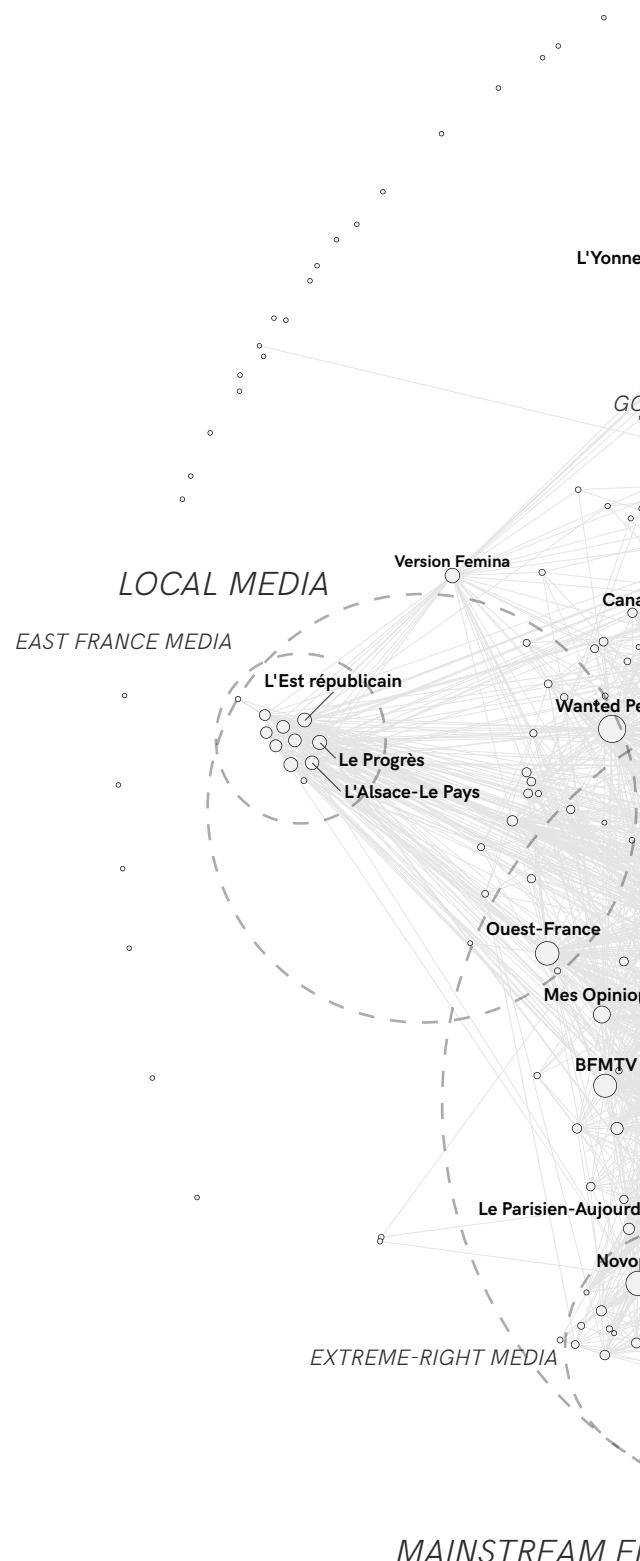
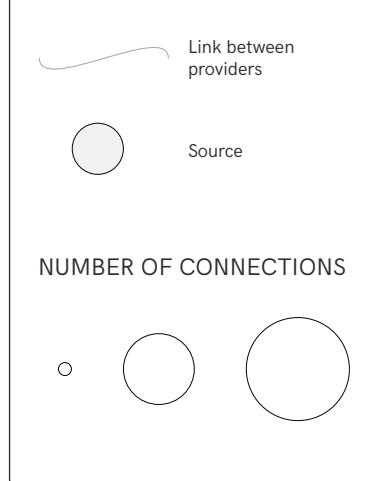
Identify (or compile) a list of all the possible Web sources in which you think your fake news item might have appeared (try to be as exhaustive as possible). You can use one of the many lists of fake news websites maintained by debunking initiatives and combine it with a list of mainstream media outlets.

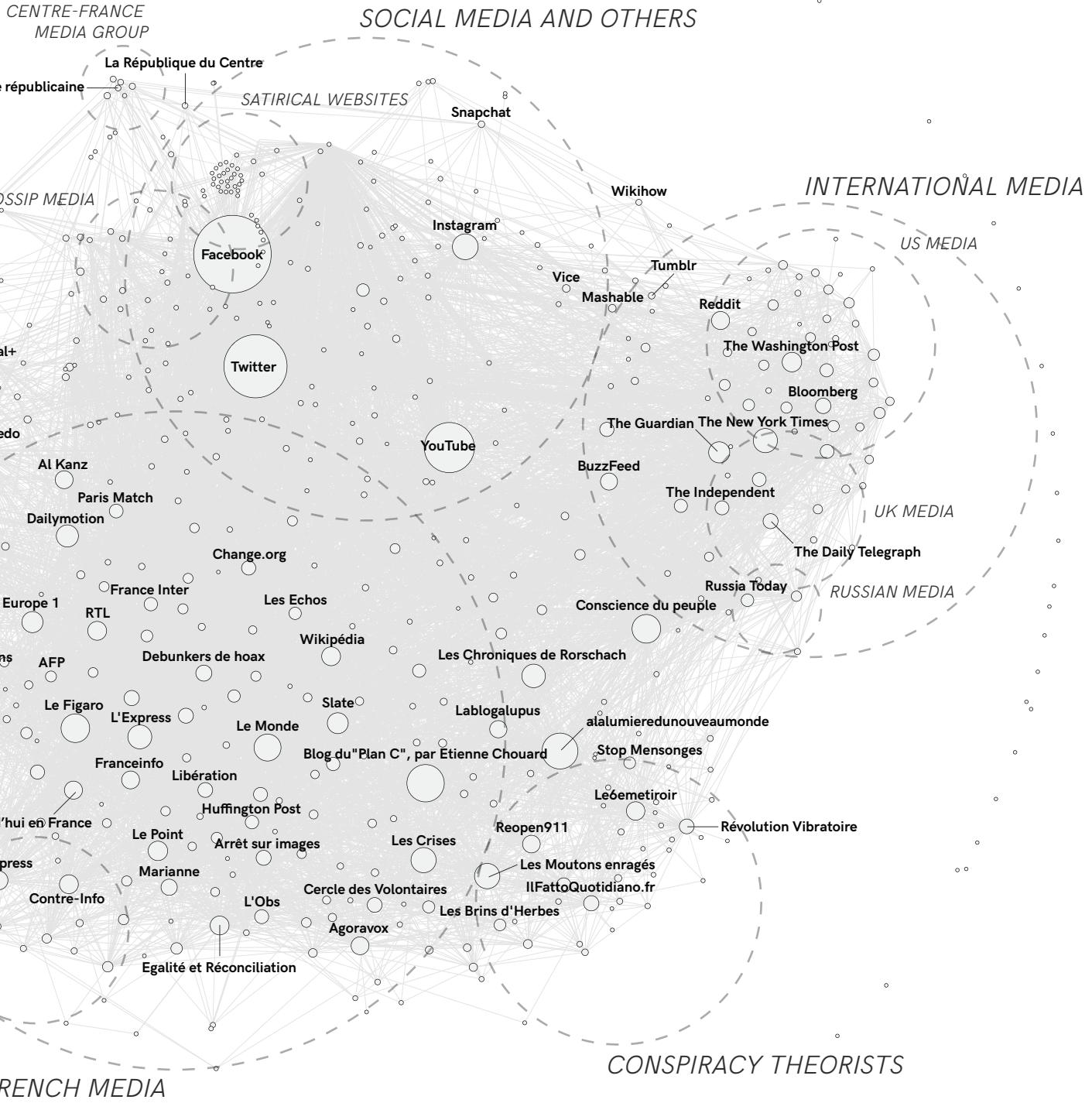
- ◊ Identify how the sources in your list are associated with each other through → **web crawling** and hyperlink analysis. We used  **Hyphe** for this.
- ◊ Visualise the resulting network and apply a force-directed layout algorithm to identify clusters of sources. You can use  **Gephi** for this task.
- ◊ Manually highlight and name the clusters.

WHAT ARE THE MAIN SPHERES IN THE FRENCH MEDIA SYSTEM?

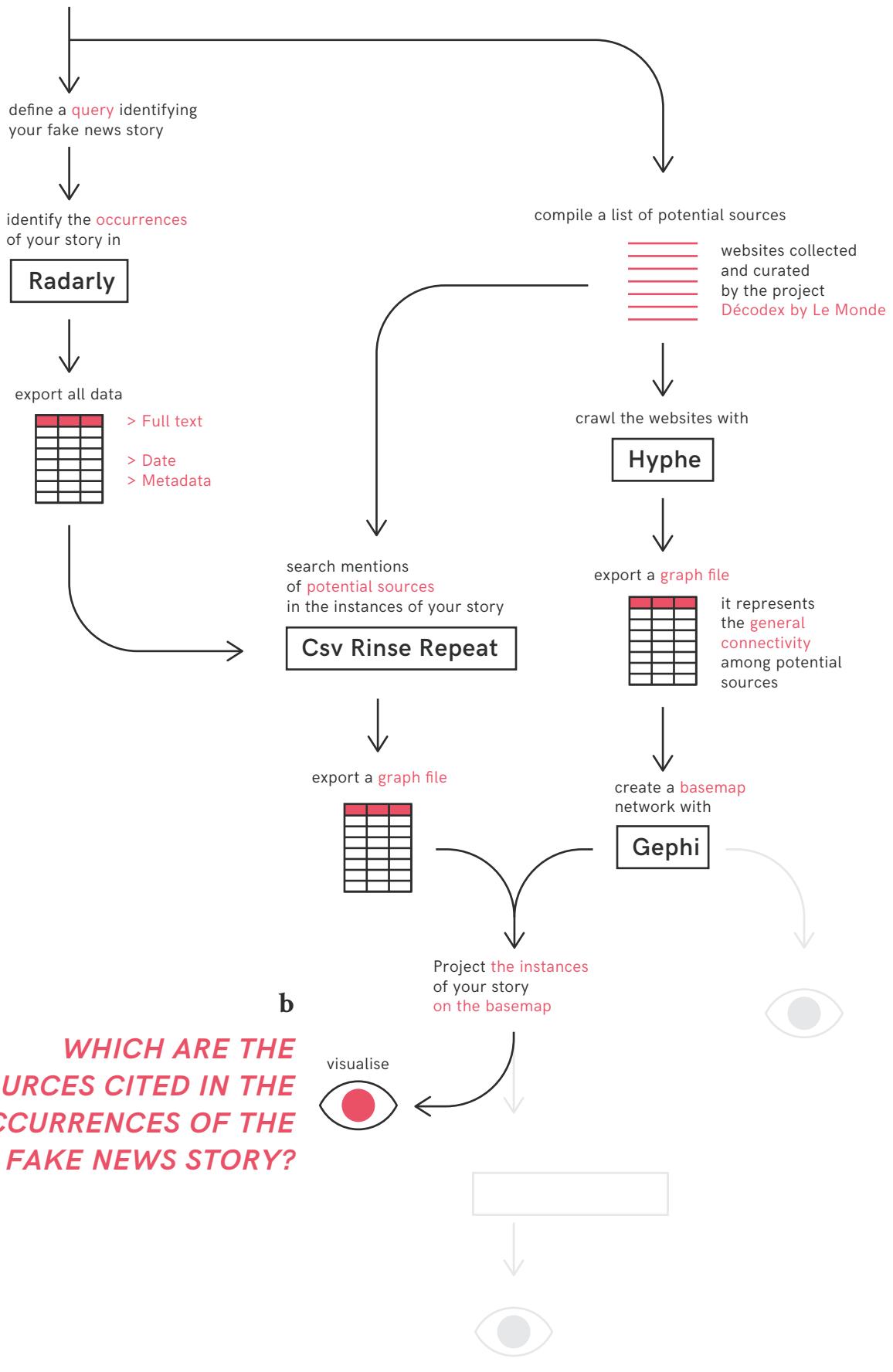
Network analysis of the media sources

active in French public debate. The image shows the news sources listed by the Décodex project by *Le Monde* and the hyperlinks connecting them. A force-directed layout has been applied to reveal the main clusters of websites and their respective associations and positions.





START



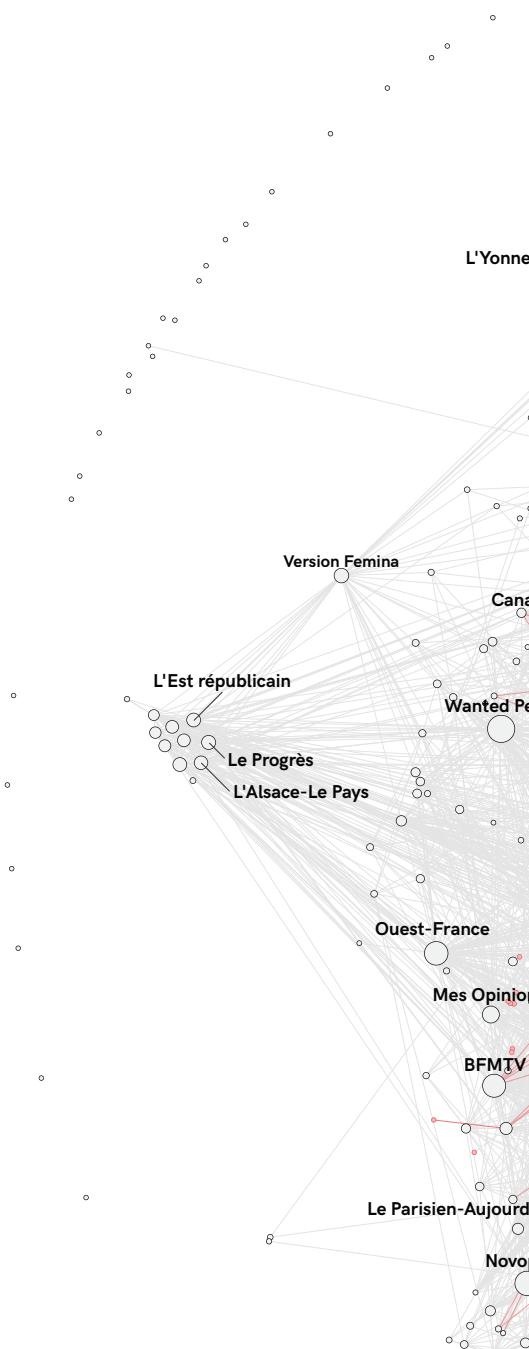
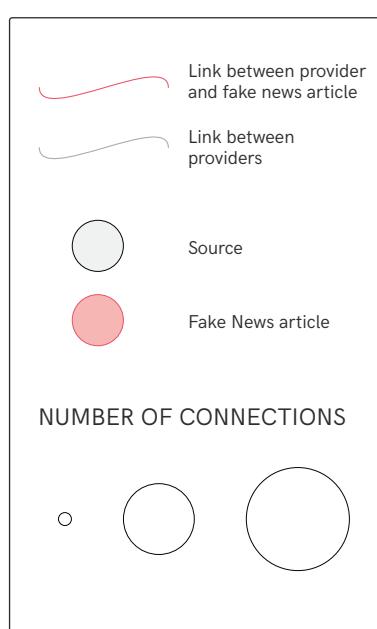
b. HIGHLIGHT THE OCCURRENCES OF YOUR STORY ON THE BASE MAP

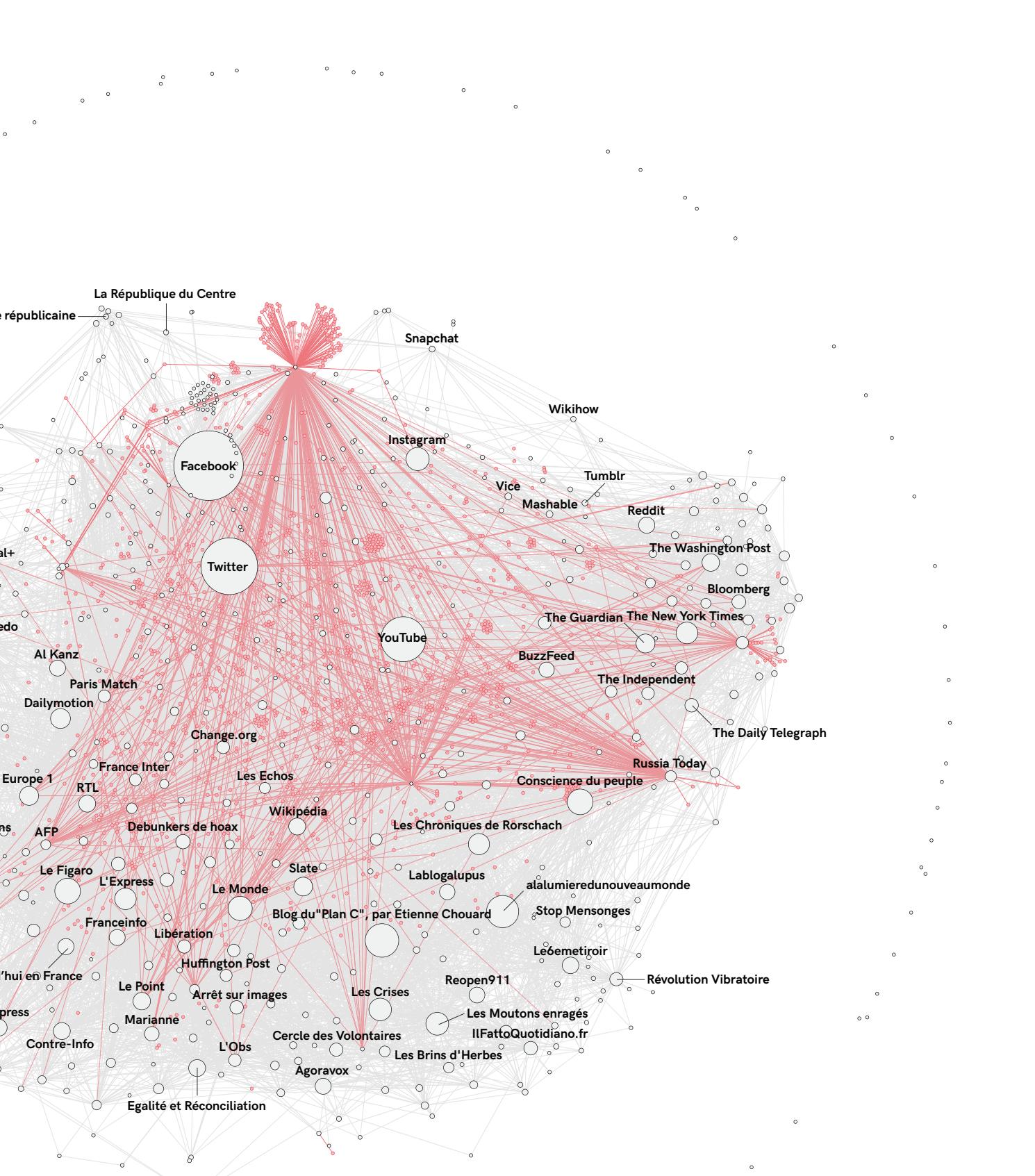
In this step we will explore how fake news stories are associated with different sources. This is interesting, as while a fake news story might – for example – start out its life as a piece of satire, as it travels it can become more prominently associated with non-satirical sources. Here we will identify which of the sources in the base map of the French media system are mentioned in the pages in which your fake news story occurs.

- ◊ Create a query that identifies the fake news story that you want to trace. Use keywords specifically associated with your story and the stop-words to exclude "false positives".
- ◊ Identify the occurrences of your story, running your query on the archive that you have chosen to use. For each of the results, collect the full text and the date of publication.
- ◊ Detect, in the occurrences of the story, mentions of the sources of your base map. Search for URLs as well as for the names of your sources (e.g. sputniknews.com, Sputnik). In our example we used a custom script for  CSV Rinse Repeat.
- ◊ Project the occurrences of your story onto your base map, by connecting each of them to the sources that they mention. While keeping the source-nodes fixed, apply a force directed spatialisation algorithm (you can do this using  Gephi) to move the nodes representing the fake story occurrences closer to clusters of the base map that they cite the most.

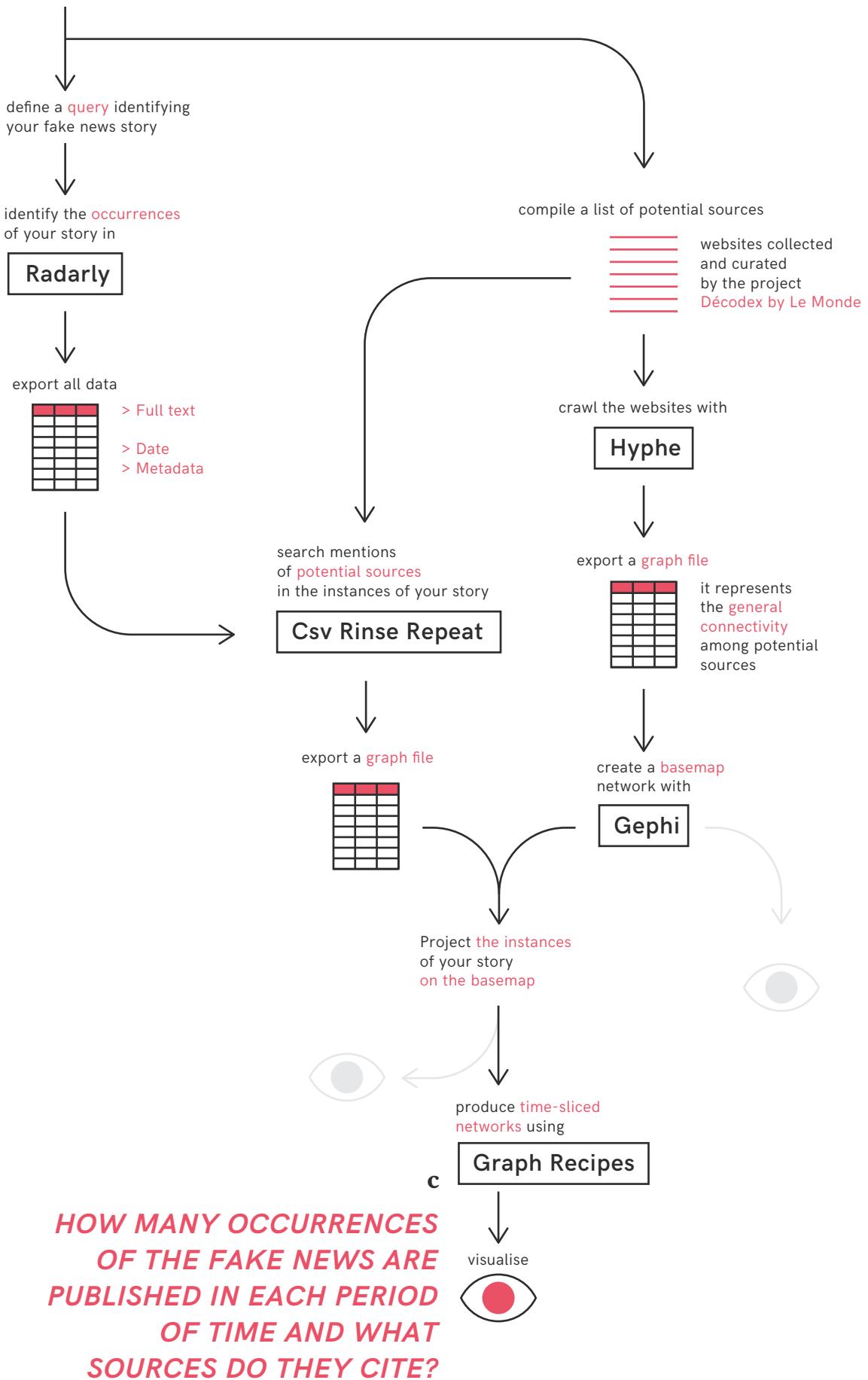
WHICH ARE THE SOURCES CITED IN THE OCCURRENCES OF THE FAKE NEWS STORY?

Projection of the fake news occurrences on the network of media sources. In this image, the occurrences of the fake news story are positioned on the base map displayed by the previous network according to the sources they cite. The tendency to refer to social media is visible as well as the relevance of *Russia Today* and *Sputnik International* in this particular story.





START



c. VISUALISE THE SPREAD OF YOUR FAKE STORY ON THE BASE MAP

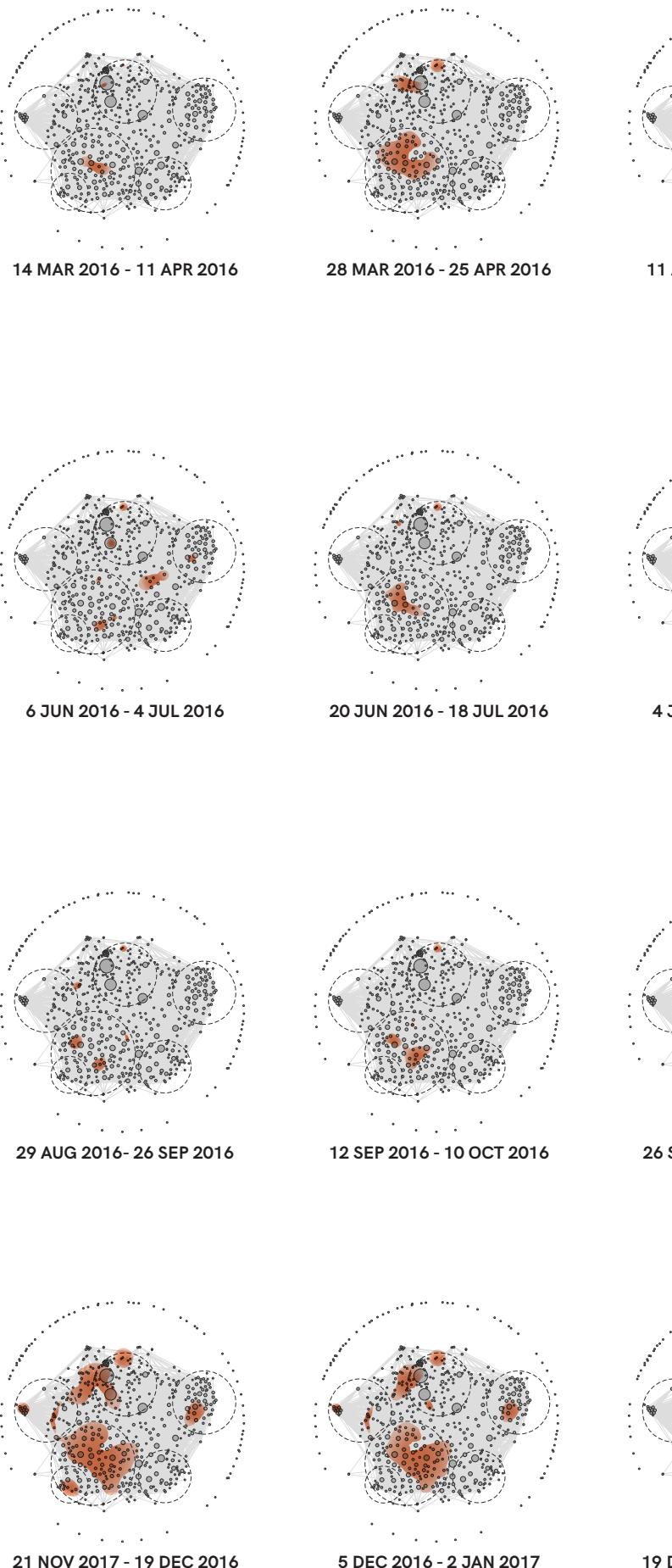
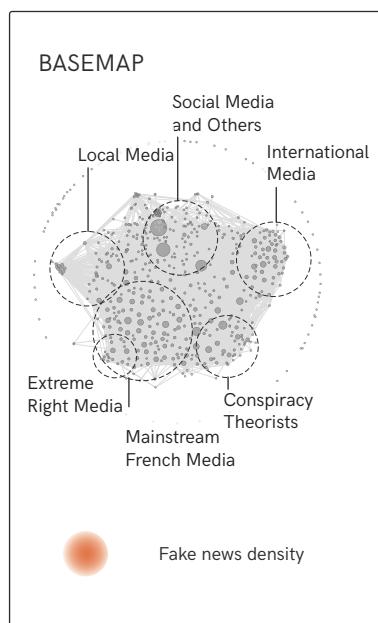
In this step, you will reveal how the reference patterns identified in the previous step evolve over time.

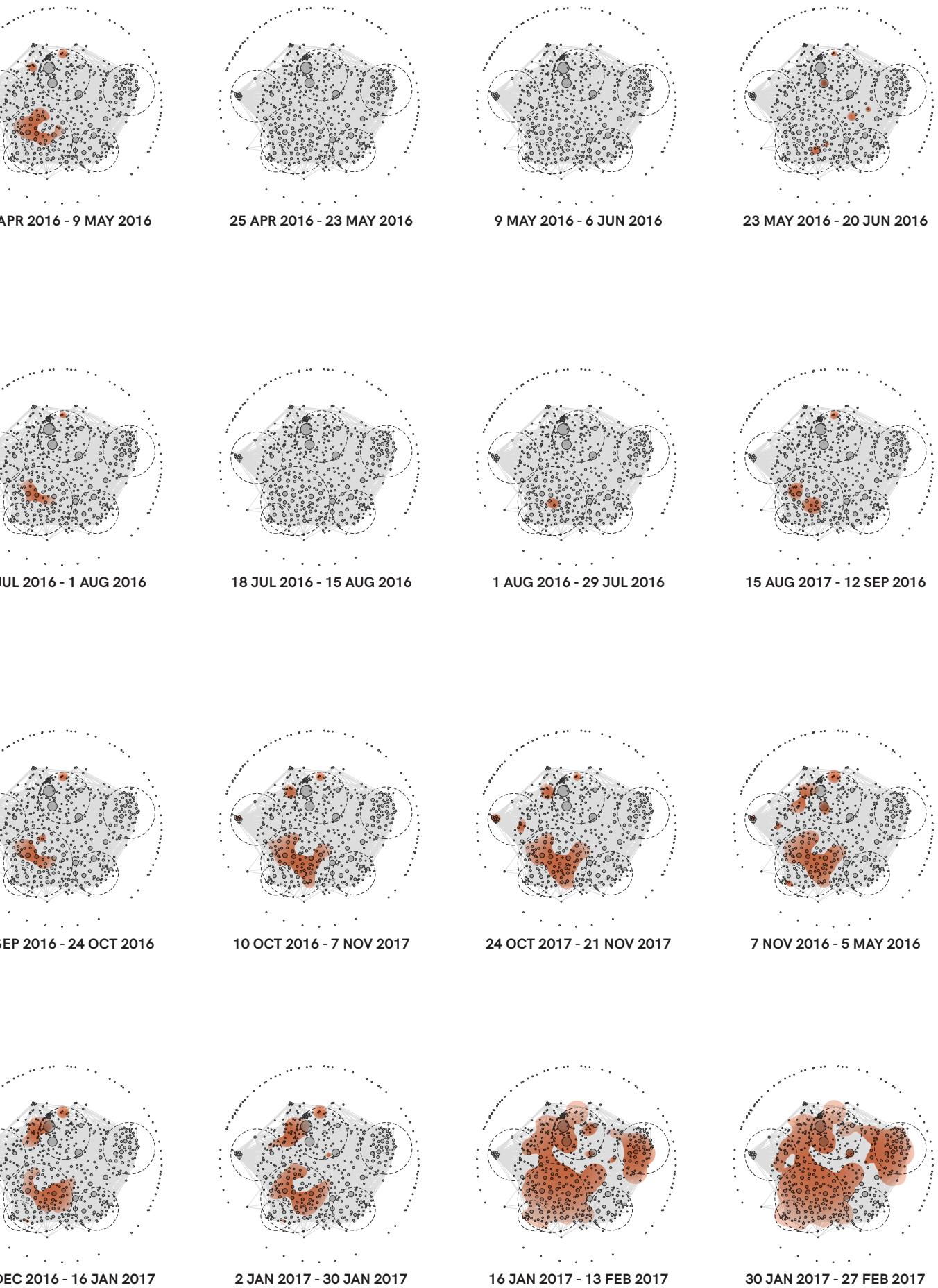
- ◊ Slice your network of occurrences and references by month, by week or by day, according to the speed of circulation of your story. In this example we grouped news by month and then zoomed in on a four day window to explore the most important period of circulation.
- ◊ While keeping the source base map stable, visualise the different temporal slices of fake news story occurrences.
- ◊ In order to make the changes and patterns more legible, you can represent the fake story occurrences not as single nodes, but through a density heatmap (the example has been produced using a  Graph Recipes).

HOW MANY OCCURRENCES OF THE FAKE NEWS STORY ARE PUBLISHED IN EACH PERIOD AND WHAT SOURCES DO THEY CITE?

Temporal evolution of the fake news story in the whole observed period.

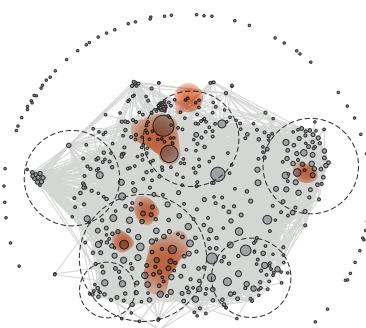
In this image, the occurrences of the fake news story are divided in slices of 4 weeks (with an overlap of two weeks) and represented as a density heat map rather than as individual points. Though mentions of the story have been present for more than one year, its circulation appears to spike up in February 2017, when a new strand of the fake story is published by the Russian website *Sputnik International*.



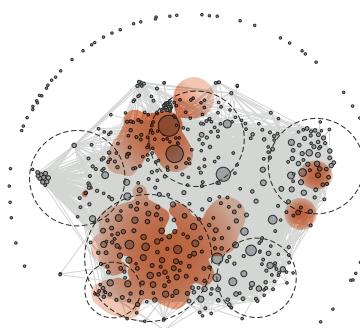


HOW MANY OCCURRENCES OF THE FAKE NEWS STORY ARE PUBLISHED IN FEBRUARY 2017 AND WHAT SOURCES DO THEY CITE?

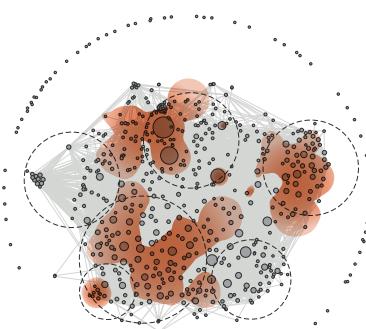
Temporal evolution of the fake news story in February 2017. This image represents a 'temporal zoom' of the previous one. Here the occurrences of the fake news story are broken up in slices of 4 days (with an overlap of two days).



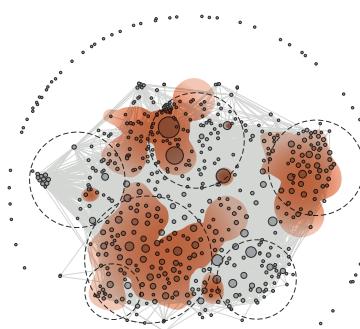
1 FEB 2017 - 5 FEB 2017



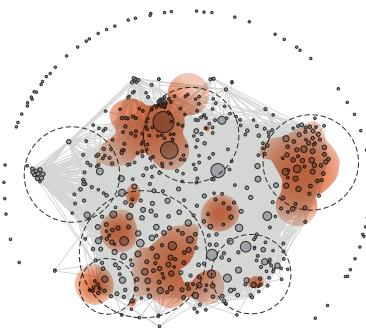
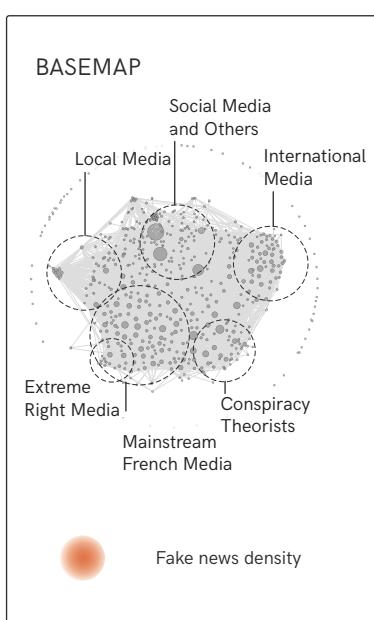
3 FEB 2017 - 7 FEB 2017



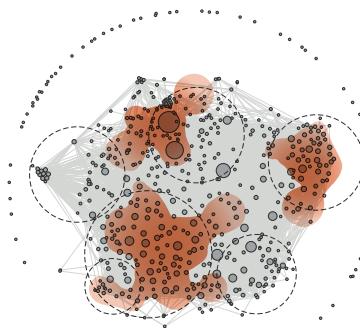
11 FEB 2017 - 15 FEB 2017



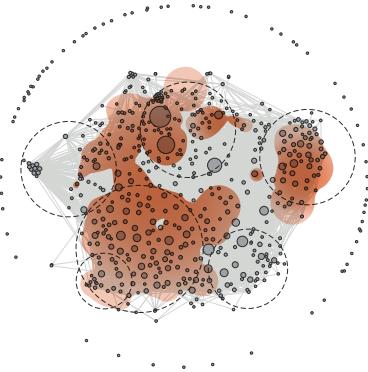
13 FEB 2017 - 17 FEB 2017



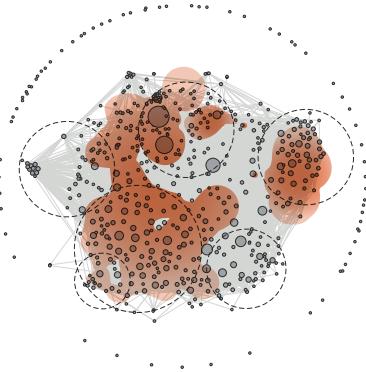
21 FEB 2017 - 25 FEB 2017



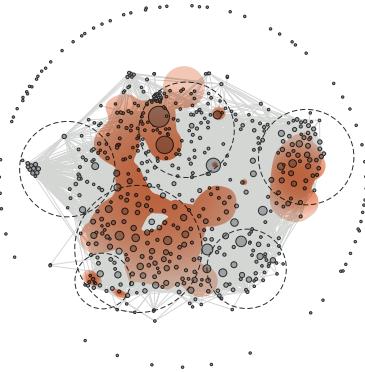
23 FEB 2017 - 27 FEB 2017



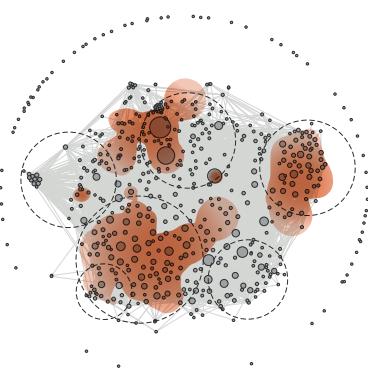
5 FEB 2017 - 9 FEB 2017



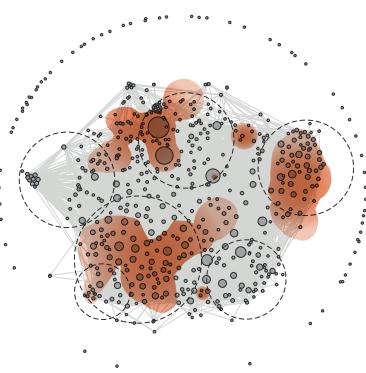
7 FEB 2017 - 11 FEB 2017



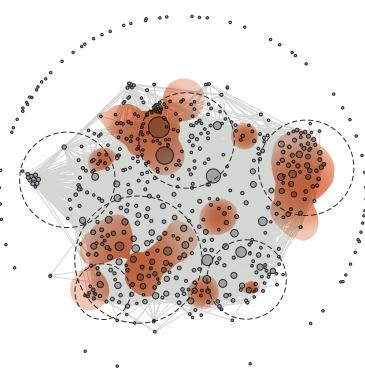
9 FEB 2017 - 13 FEB 2017



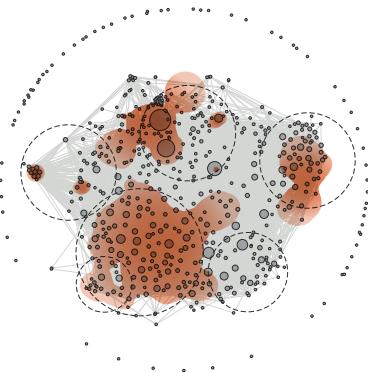
15 FEB 2017 - 19 FEB 2017



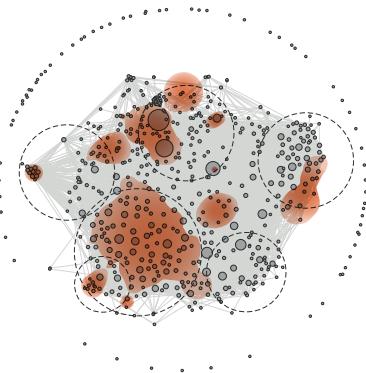
17 FEB 2017 - 21 FEB 2017



19 FEB 2017 - 23 FEB 2017

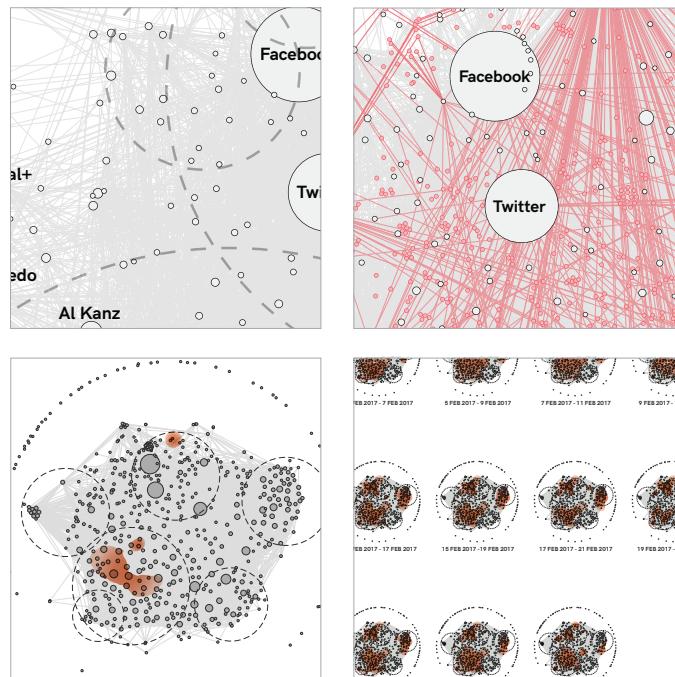


25 FEB 2017 - 1 MAR 2017



27 FEB 2017 - 3 MAR 2017

CHAPTER 2 → RECIPE 2



SERVING SUGGESTIONS

This recipe may be used to identify which websites reference a fake news story most often in different spheres. These are not necessarily the original sources of the fake news, but are often the most influential media outlets that contribute to its circulation (whether as a rumour or as a debunked story).

Chapter 3

USING TRACKER SIGNATURES TO MAP THE TECHNO- COMMERCIAL UNDERPINNINGS OF FAKE NEWS SITES

Do fake news sites use different kinds of trackers from mainstream media sites?

How can fake news and mainstream media sites be profiled based on their tracker usage?

How do tracker ecologies on fake news sites change over time?

Which other websites share the same tracker IDs as fake news sites?

Do trackers associated with hyper-partisan, and misleading information sites vary across language spheres?

Introduction - Over the past few decades many responses to misinformation and disinformation have focused on mapping and debunking claims made or repeated by politicians, journalists or other public figures. What are the prospects of mapping fake news online not just by looking at the circulation of claims, but by examining the technical infrastructures of the websites through which these claims are published?

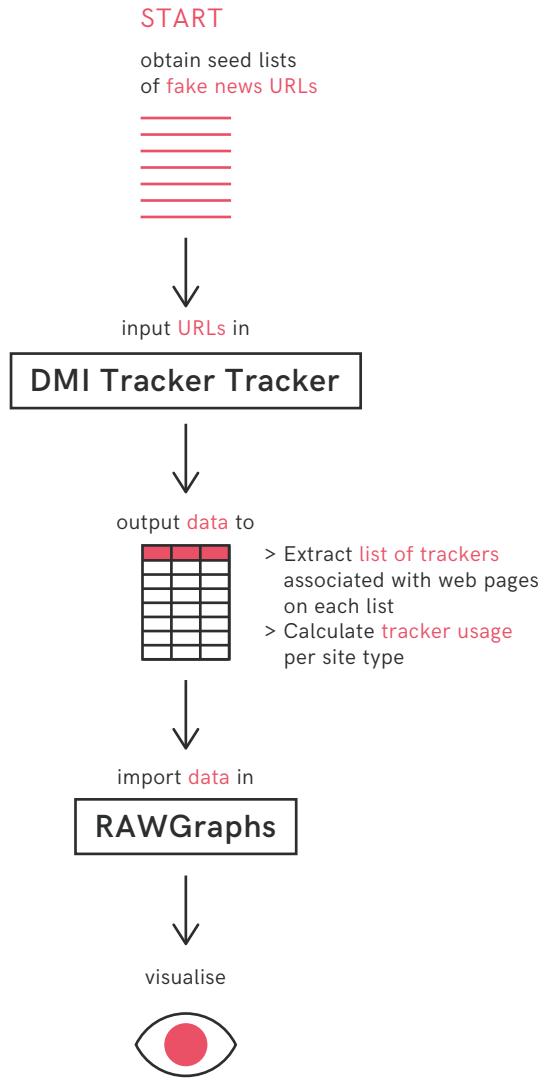
Many websites use “trackers” – small bits of embedded code – in order to monitor engagement, including visitor numbers, visitor behaviour and the effectiveness of ads. In this section we look at how data about web trackers can be repurposed in order to investigate the technical and commercial underpinnings of websites associated with fake news and other misleading information phenomena.



DO FAKE NEWS SITES USE DIFFERENT KINDS OF TRACKERS FROM MAINSTREAM MEDIA SITES?

BEFORE STARTING

For this recipe you will need two lists of URLs: one list of fake news URLs and one list of mainstream media URLs. How these lists are obtained is a crucial part of the research process. You can either draw on existing lists, or create your own (e.g. by compiling a selection, triangulating from other sources, or obtaining from different platforms or media sources). The starting point that you choose will affect how to read and what you can do with the results. To illustrate this recipe, we start with a selection of fake news pages obtained from a list created by *BuzzFeed News* (ordered by most engaged with content according to the  **BuzzSumo** tool), as well as a list of mainstream media web pages obtained by triangulating lists from *BuzzFeed News* and *Alexa*.



***DO MAINSTREAM
MEDIA AND FAKE NEWS
WEBSITES SHARE THE SAME
TRACKER ECOLOGY?***

CALCULATE TRACKER USAGE PER SITE TYPE

From the → **source code** of web pages it is often possible to see which third-party tracking services are used.

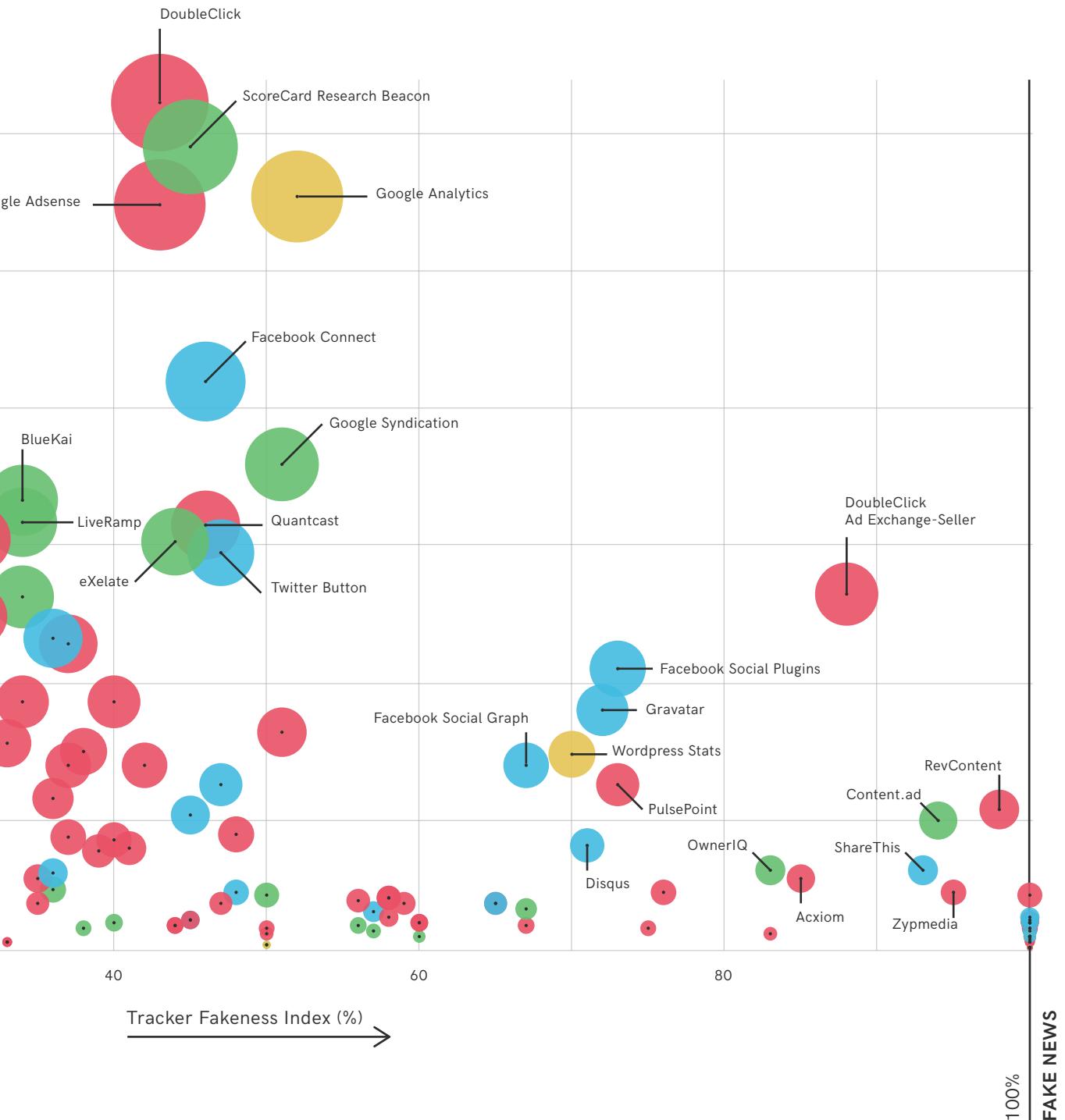
- ◊ Collect data about trackers associated with the web pages on each list. You may use the  **DMI Tracker** tool to collect this information.
- ◊ Count the usage of each tracker in fake news websites and in mainstream news websites.
- ◊ You may use a → **scatter plot** to visualise the resulting data. Each circle represents one tracker coloured by category. On the horizontal axis, you can show, for example, the distribution of trackers usage by mainstream media and fake news websites. On the vertical axis, you can indicate the overall usage of the tracker. We used the  **RAWGraphs** tool to generate this visualisation.

DO MAINSTREAM MEDIA AND FAKE NEWS WEBSITES SHARE THE SAME TRACKER ECOLOGIES?

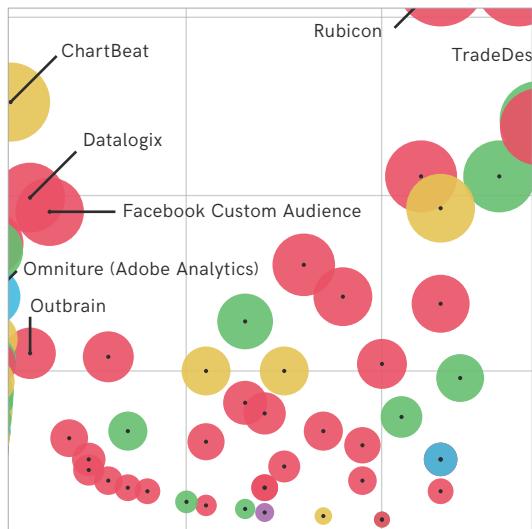
Scatterplot representing tracker usage on a series of fake news and mainstream media sites.

While fake news sites and mainstream media sites share popular tracker services such as Google Adsense, DoubleClick and Google Analytics, mainstream media appears more mature and sophisticated in its use of trackers in terms of the number and diversity of trackers that it uses.





CHAPTER 3 → RECIPE 1



SERVING SUGGESTIONS

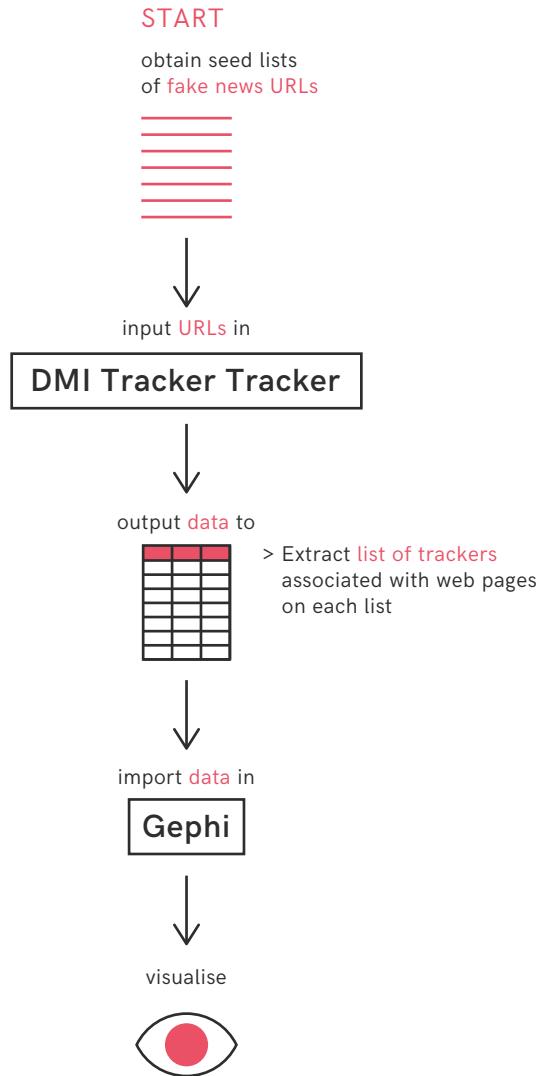
This recipe can be used to profile the tracking practices associated with different kinds of websites – including which trackers are either mainly or exclusively associated with fake news websites and what these trackers do – as well as identifying most commonly used trackers. It can also be used for exploring the “long tail” of smaller and more specialised trackers.



HOW CAN FAKE NEWS AND MAINSTREAM MEDIA SITES BE PROFILED BASED ON THEIR TRACKER USAGE?

BEFORE STARTING

For this recipe you will need two lists of URLs: one list of fake news URLs and one list of mainstream media URLs. How these lists are obtained is a crucial part of the research process. You can either draw on existing lists, or create your own (e.g. by compiling a selection, triangulating from other sources, or obtaining from different platforms or media sources). The starting point that you choose will affect how to read and what you can do with the results. To illustrate this recipe, we start with a selection of fake news pages obtained from a list created by *BuzzFeed News* (ordered by most engaged with content according to the  **BuzzSumo** tool), as well as a list of mainstream media web pages obtained by triangulating lists from *BuzzFeed News* and *Alexa*.



***HOW DO FAKE NEWS SITES
AND MAINSTREAM MEDIA
CLUSTER ACCORDING THEIR
TRACKER USAGE?***

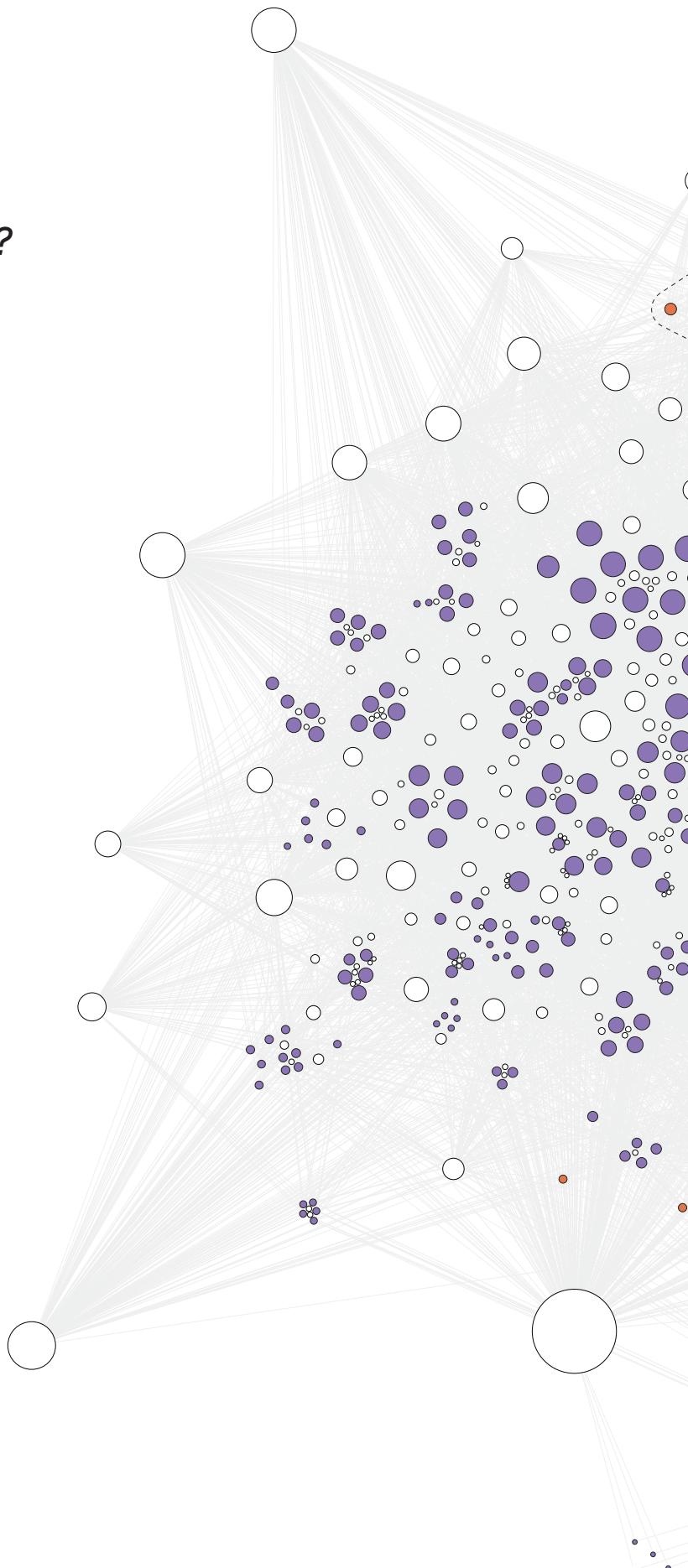
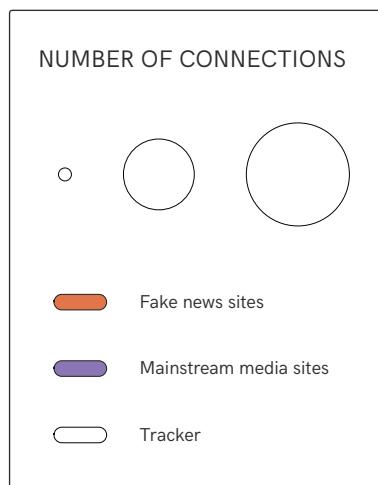
GRAPH RELATIONS BETWEEN PAGES AND TRACKERS

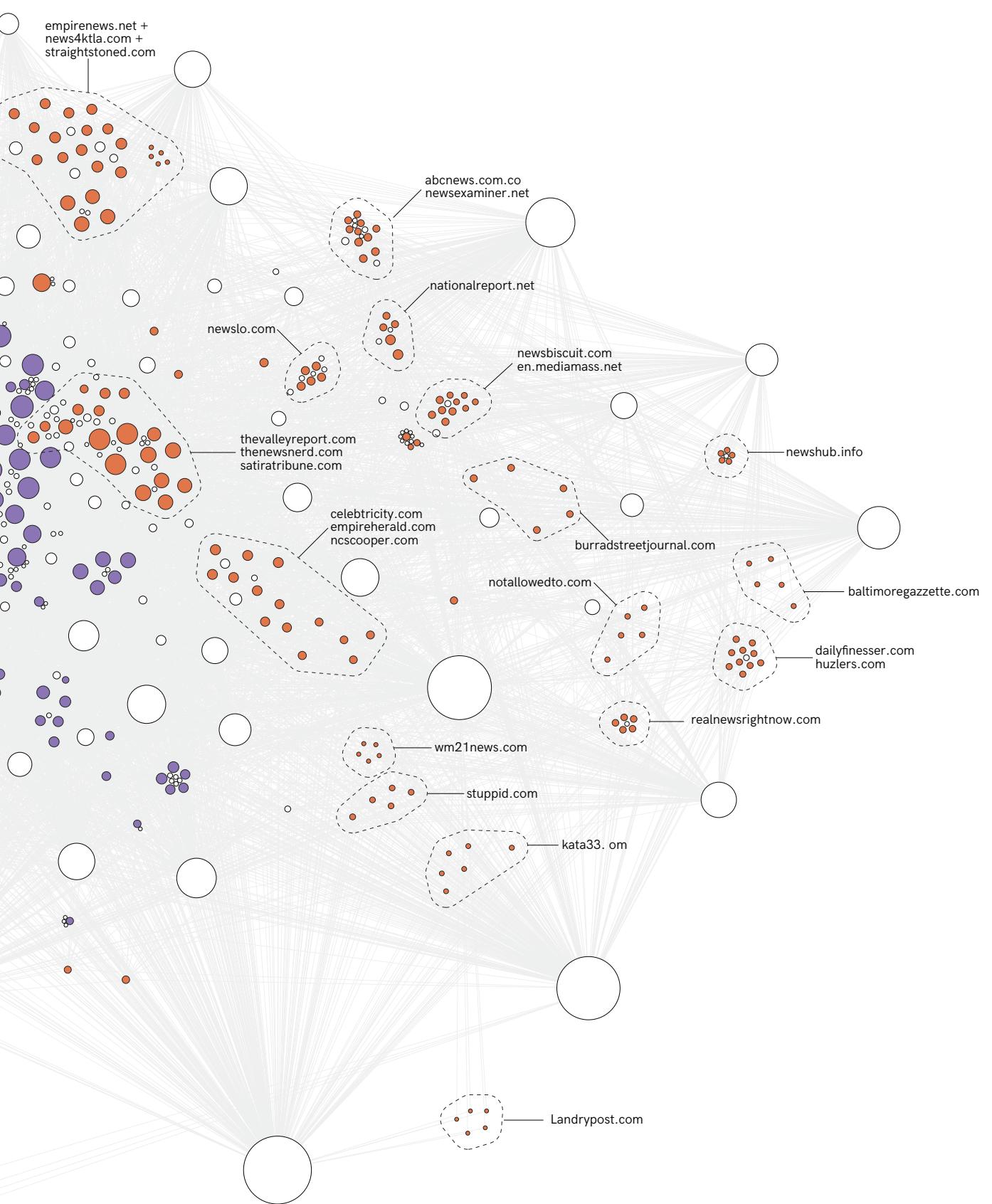
In order to explore how different URLs share the same patterns of tracker usage we can create a → **network graph** to highlight associations between web pages to their corresponding trackers.

- ◊ Extract lists of trackers associated with the initial lists of fake news and mainstream media pages. You may use the → **DMI Tracker Tracker** tool to collect this information.
- ◊ Create a network in order to show the tracker usage patterns of the different web pages. We used → **Gephi** in order to visually explore the network using a → **force directed network layout** to help read the data.
- ◊ You can annotate the network graph in order to highlight the clusters of URLs (e.g. fake news clusters, or mainstream media clusters).

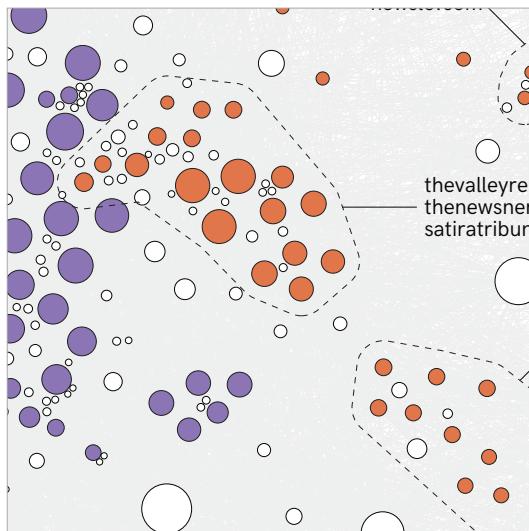
HOW DO FAKE NEWS SITES AND MAINSTREAM MEDIA CLUSTER ACCORDING THEIR TRACKER USAGE?

Bipartite network of trackers and websites that use them. Shared tracker signatures may be used to explore tracker practices or strategies amongst a set of websites or to detect fake news “media groups.”





CHAPTER 3 → RECIPE 2



SERVING SUGGESTIONS

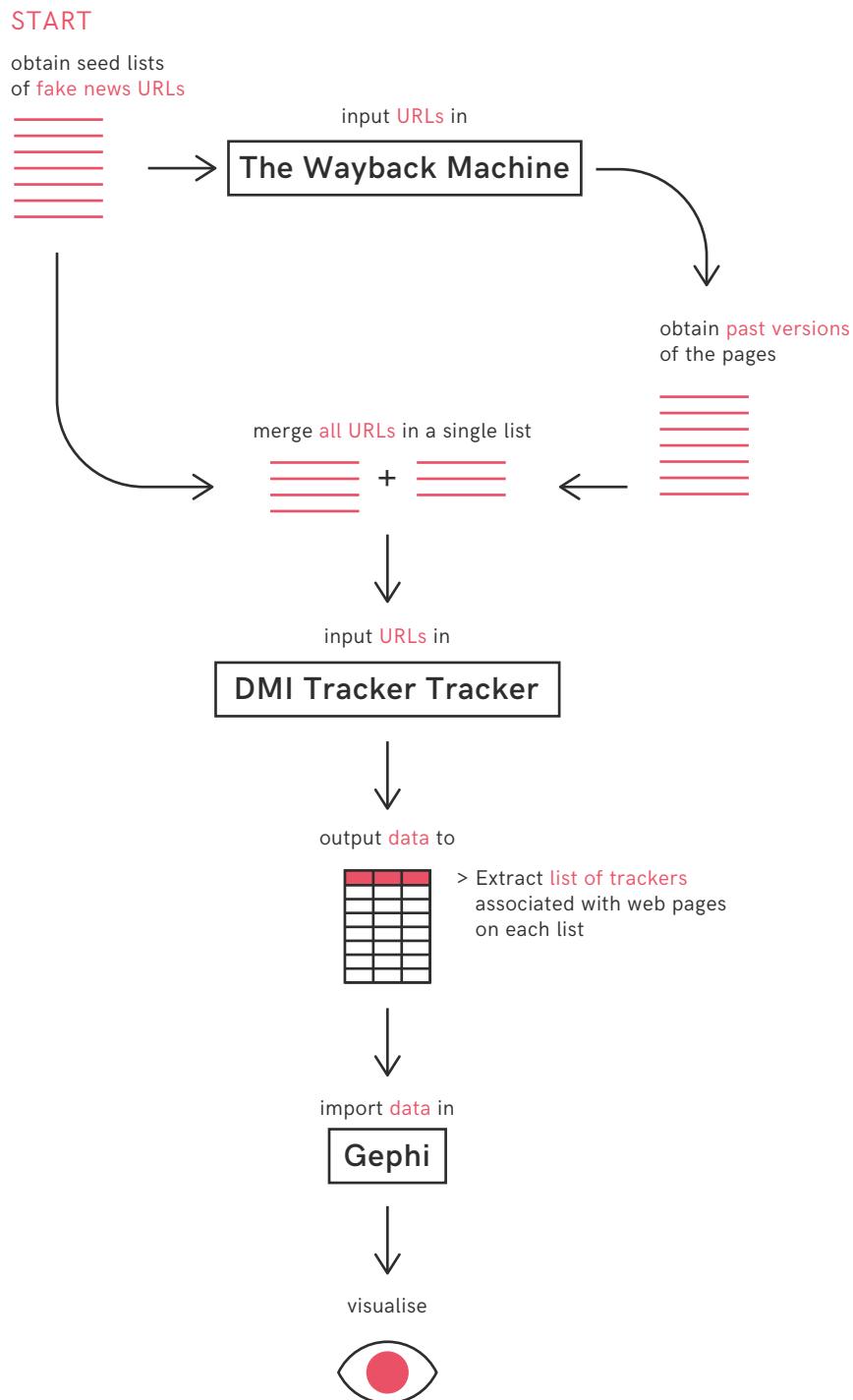
This recipe can be used to explore how a set of web pages can be grouped based on their tracker signatures. This provides a complementary picture to lists or metrics (e.g. of most and least used trackers across the pages) by facilitating exploration of relations between trackers and websites. For example it could be used as a starting point to identify potential fake news “media groups” for further investigation, or to explore the different web tracking practices, styles and footprints of fake news web pages – including comparisons between pages associated with different regions, issues or sources.



HOW DO TRACKER ECOLOGIES ON FAKE NEWS SITES CHANGE OVER TIME?

BEFORE STARTING

For this recipe you will need the → source code of the same web page (or set of web pages) at two different moments in time. You can obtain saved copies of the same page over time (e.g. through manually or automatically saving the source code yourself) or you can use public web archiving projects such as the Internet Archive's  [The Wayback Machine](#).



**HOW DO FAKE NEWS SITES
ADAPT THEIR TRACKER USAGE
IN RESPONSE TO BLACKLISTING
FROM MAJOR AD NETWORKS?**

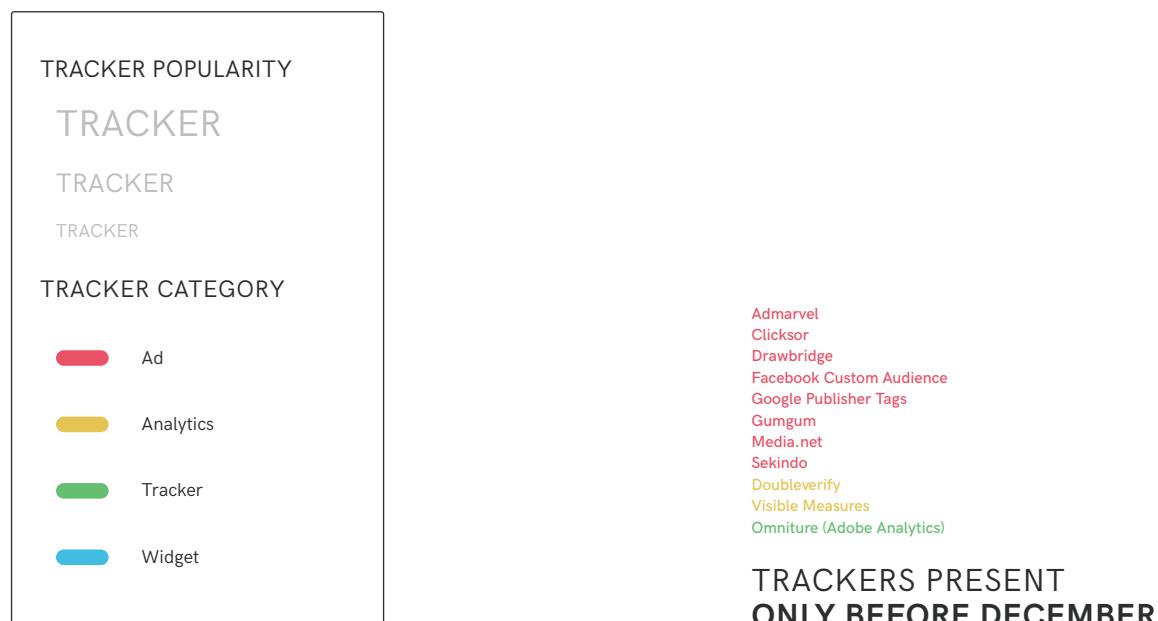
GRAPH RELATIONS BETWEEN PAGES AND TRACKERS

This recipe can be used to identify which trackers were being used by a given web page at different moments in time. It might be useful to chart changes in tracking practices – for example by examining the impact and responses to events like Google and Facebook’s bans of fake news providers from their ads programs in November 2016.

- ◊ Obtain archived copies of a webpage. You may use  **The Wayback Machine** to see how a given page changed over time.
- ◊ Identify associated trackers with the current and previous version of the page. You may use the  **DMI Tracker Tracker** tool to collect such information.
- ◊ Identify the trackers which are only present on the first date, the ones that are only present on the second date and the ones that are shared across both dates.
- ◊ You can group trackers into three lists, colouring them accordingly.

HOW DO FAKE NEWS SITES ADAPT THEIR TRACKER USAGE IN RESPONSE TO BLACKLISTING FROM MAJOR AD NETWORKS?

Tracker ecologies on fake news sites before and after blacklisting from major ad networks. While ad networks from which fake news sites have been blacklisted remain in the source code of these sites and hence are present in the graphic even after the moment of blacklisting, the visualisation also illustrates new ad networks that fake news sites have moved to. A manual review of ad services used to serve ads on the website interface may help to further refine this analysis and identify false positives (i.e. tracker services that are no longer in use but whose code remains embedded in these sites).



DoubleClick
Google Analytics
Google Adsense
DoubleClick Ad Exchange-seller
Google Syndication
Scorecard Research Beacon
Gravatar
Wordpress Stats
Facebook Connect
Exelate
Disqus
Brightroll
Adobe Audience Manager
Bluekai
Amazon Associates
Appnexus
Bidswitch
Criteo
Mediamath
Openx
Pubmatic
Tradedesk
Facebook Social Plugins
Taboola
Twitter Button
Adtech
Advertising.com
Index Exchange (Formerly Casale Media)
Pulsepoint
Quantcast
Rubicon
Spoutable
Stickyads
Tapad
Teads
Turn Inc.
Yahoo Ad Exchange
Tubemogul
Krux Digital
Liveramp
Accloudimages
Acuity Ads
Adscale
Eyeview
Revcontent
Smart Adserver
Sovrn (Formerly Lijit Networks)
Twitter Advertising
Zypmedia
At Internet
Twitter Analytics
Aggregate Knowledge
Lotame
Owneriq
Rocket Fuel
Videology
Addthis
Facebook Social Graph
Lockerz Share
Pinterest
Sharethis
Twitter Badge
Typekit By Adobe

Adform
Infectious Media
Yahoo Ad Manager Plus
Adap.tv
Adroll
Bidswitch
Crimtan
Datalogix
Dataxu
Digitant
Distillery
Getintent
Improve Digital
Infolinks
Internet Billboard
Smaato
Smartclip
Spotchange
Switch Concepts
Yieldlab
Kxcdn
Beeswax
Bidtheatre
Chango
Dotomi
Kixer
Mythings
Netmining
Pagefair
Radiumone
Sumome
Tumblr Dashboard

TRACKERS
ALWAYS PRESENT

TRACKERS PRESENT
ONLY AFTER JANUARY

Google Analytics
Google Adsense
Doubleclick AdSense
Google Syndication
Scorecard Research
Gravatar

SERVING SUGGESTIONS

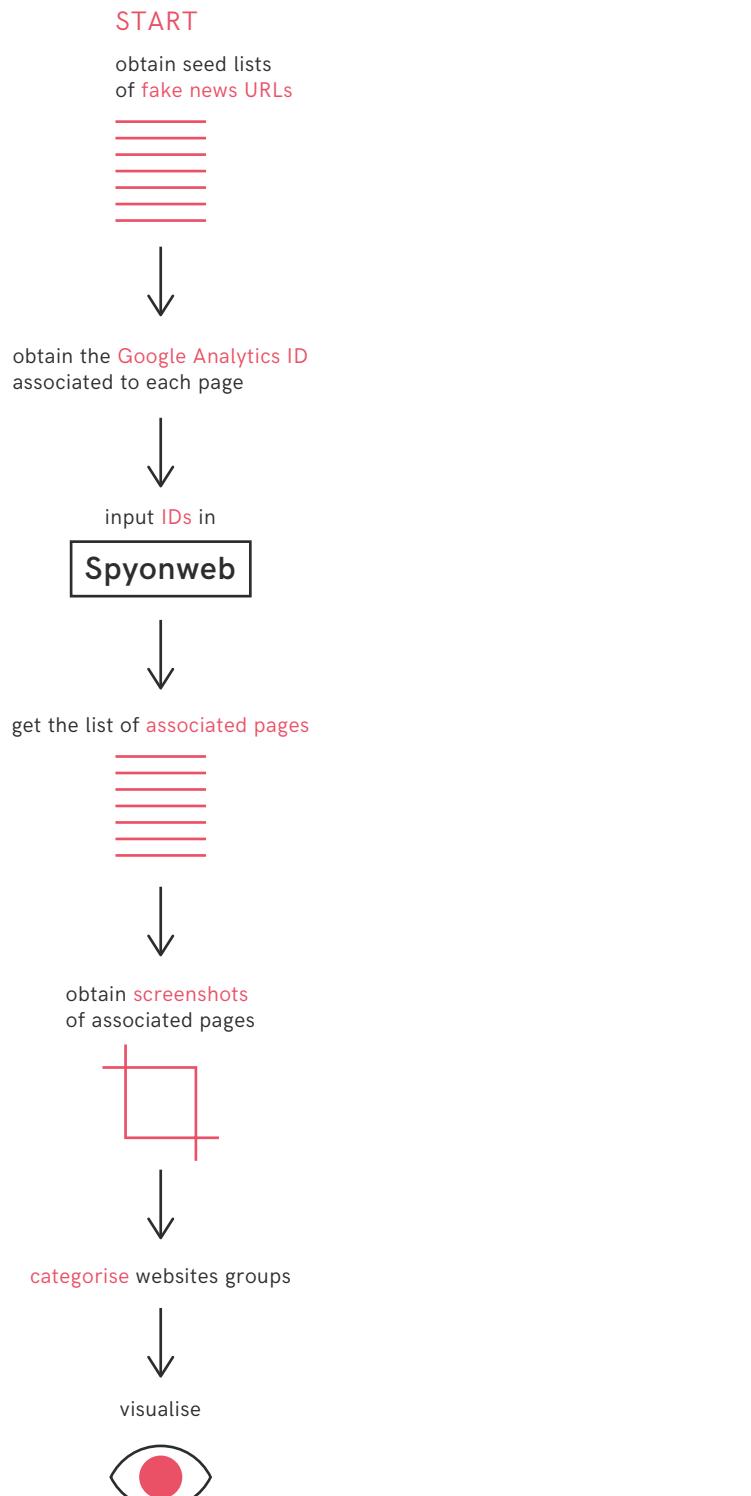
Given debates and proposals about stopping the ad revenue of fake news, this recipe may be used to understand how fake news websites are adapting to the measures taken by trackers services, technology companies and advertisers – as well as how effective these measures are. For example, it can show which trackers have been dropped, which remain and which are added at different moments in time.



WHICH OTHER WEBSITES SHARE THE SAME TRACKER IDS AS FAKE NEWS SITES?

BEFORE STARTING

Before you start you will need to compile or identify seed lists of fake news and other misleading information websites. We illustrate this recipe by examining which websites use the same → Google Analytics IDs as a list of websites from the EU Disinformation Review.



**WHAT MEDIA GROUP STRATEGIES
CAN BE DETECTED THROUGH
SHARED GOOGLE ANALYTICS IDS?**

IDENTIFY WEBSITES WHICH SHARE TRACKER IDS WITH A SEED LIST OF PAGES OR SITES

This recipe can be used to identify which other websites share the same tracker IDs as web pages on a given list.

- ◊ Extract the Google Analytics ID for each URL in your starting list. You can do this manually (e.g. by looking in the → **source code** for a string in the form “UA-xxxxxxx”) or automatically through → **web scraping** or other tools (in this example we wrote a custom script in order to extract this information from the metadata of the website).
- ◊ Obtain a list of pages associated with the same ID. We used the → **API** of  **Spyonweb.com** to get this information.
- ◊ Take a screenshot of each web page. We used a script to automate the process of obtaining screenshots, in order to visually compare the different websites to identify different kinds of media groups.
- ◊ Place together screenshots of pages with the same ID to spot differences and similarities between websites across and within groups.

WHAT MEDIA GROUP STRATEGIES CAN BE DETECTED THROUGH SHARED GOOGLE ANALYTICS IDS?

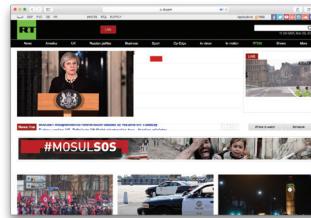
A selection of websites which share the same Google Analytics IDs, based on seed list from EU Disinformation Review.

Review. This illustrates the diversity of online settings where claims labelled as Russian disinformation are shared – from large media groups such as *Russia Today*, to themed clusters (e.g. military or mysticism), and geographical clusters (e.g. Canadian). One can also identify distinctive visual styles and possible shared → CMS features amongst different websites in these clusters, which may be used as the basis for further investigations into the media, publication and communication strategies of websites associated with misleading information online.

rt.com

19 disinformation stories

UA-5773642



xryshaygh.com

1 disinformation story

UA-4839940

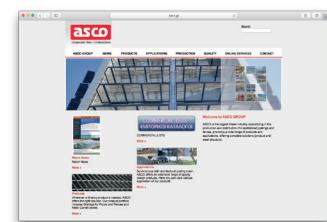


Media group
(Russia Today)

Lone
Webmaster



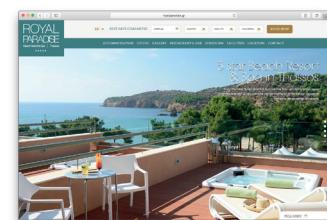
assange.rt.com



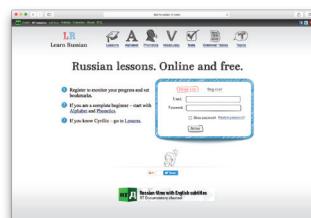
asco.gr



doc.rt.com



royalparadise.gr



learnrussian.rt.com



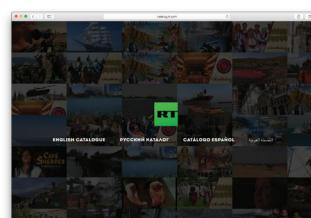
discoverthassos.com



russian.rt.com



thassosinn.gr



catalog.rt.com

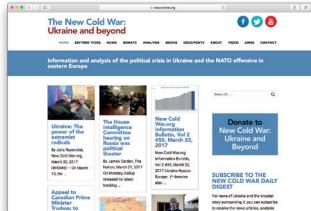
1 disinformation story

UA-841082



7 disinformation stories

UA-15942468



3 disinformation stories

UA-3004323



Themed network (Navy, Airforce)



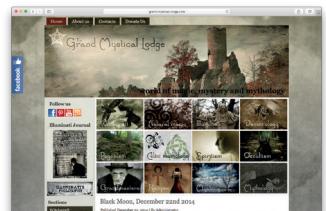
armedforcesjournal.com

Canadian socialism and unions



oakvillendp.ca

Themed network (Mysticism, Liberland)



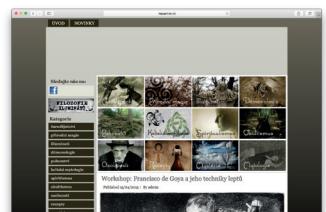
grand-mystical-lodge.com



sightlinemediagroup.com



socialiststudies.com



malachim.cz



militarytimes.com



ndpsocialists.ca



liberlandpress.com



airforcetimes.com



ccu-csc.ca



illuminati-journal.com

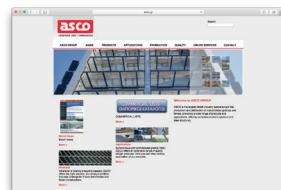


marinecorpstimes.com



oldm.cz

CHAPTER 3 → RECIPE 4



asco.gr



royalparadise.gr

SERVING SUGGESTIONS

This recipe may be used in the service of expanding a group of fake news web pages – in order to derive lists of other websites which share the same tracker IDs.

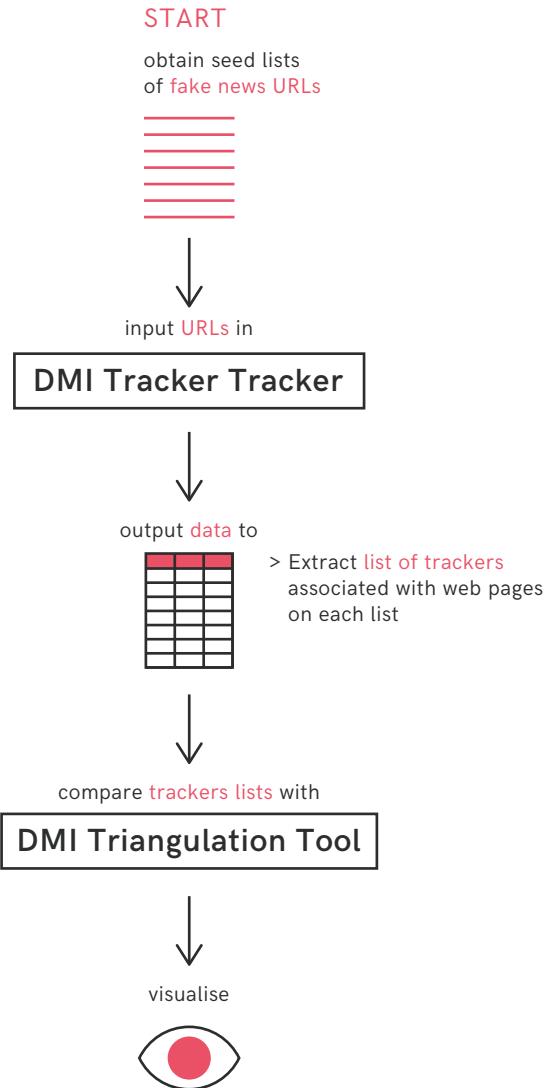
It may also be used to provide context to the digital strategies and “media groupings” of fake news providers



DO TRACKERS ASSOCIATED WITH HYPER- PARTISAN AND MISLEADING INFORMATION SITES VARY ACROSS LANGUAGE SPHERES?

BEFORE STARTING

For this recipe, you will need lists of fake news, hyper-partisan or misleading information sites in different language spheres in order to compare their trackers and tracking practices. We illustrate this recipe with reference to hyper-partisan, fake news and misleading information sites in Dutch, English and German language spheres.



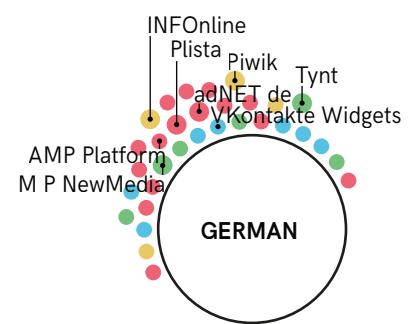
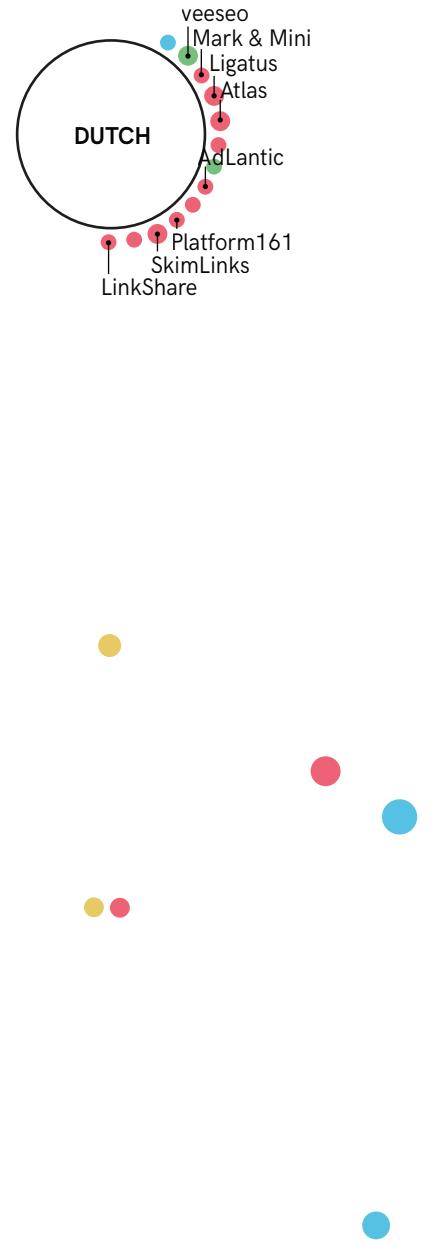
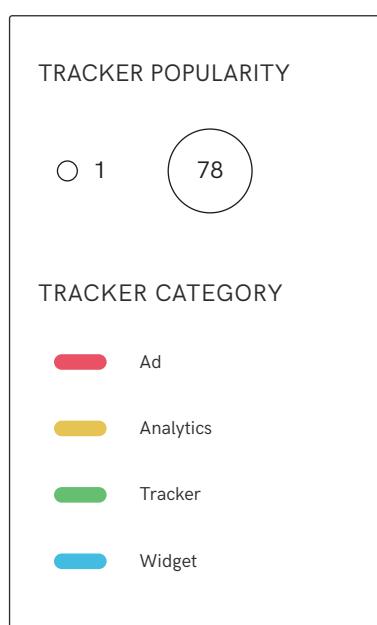
DO MISINFORMATION AND HYPER-PARTISAN WEBSITES IN DIFFERENT LANGUAGE SPHERES HAVE DISTINCT TRACKER ECOLOGIES?

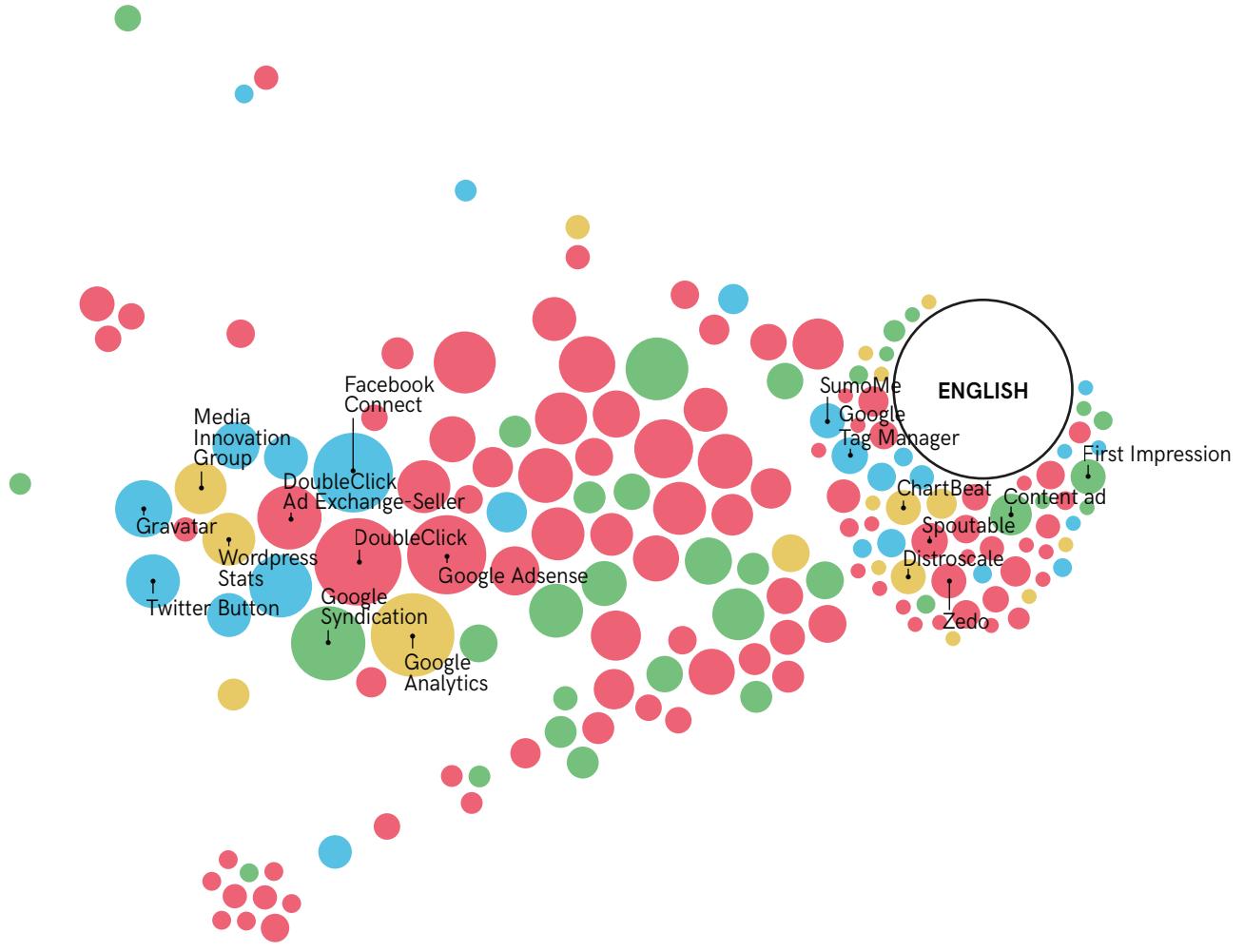
IDENTIFY TRACKERS PER LANGUAGE SPHERE

- ◊ Extract trackers associated with lists of the web pages for each language sphere. We did this using the  DMI Tracker tool.
- ◊ Identify the trackers which are shared across and which are unique to different languages spheres within the dataset. We did this using the  DMI Triangulation tool.
- ◊ You can illustrate the results using the visual metaphor of magnets. Each of the three languages are represented on the corner of a triangle. The trackers are distributed in the triangle according to their usage: if a tracker is used by all three languages it will appear in the middle, if it is used by two languages the tracker will be placed on the edge between the two and so on.

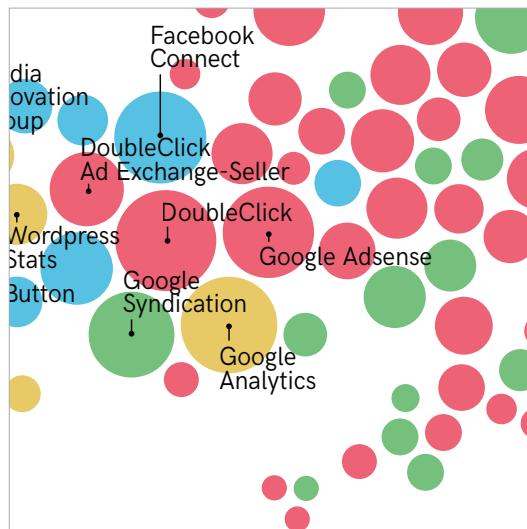
DO MISLEADING INFORMATION AND HYPER-PARTISAN WEBSITES IN DIFFERENT LANGUAGE SPHERES HAVE DISTINCT TRACKER ECOLOGIES?

Visualisation of tracker ecologies associated with hyper-partisan or misleading information sites across three language spheres. While popular ad and widget services such as DoubleClick, Google Adsense and Facebook Connect are shared across language spheres, unique services per language sphere may also be detected. For example, trackers associated with the Russian-language focused Mail.ru Group are only found in the set of websites associated with the German language sphere.





CHAPTER 3 → RECIPE 5



SERVING SUGGESTIONS

This recipe can be used to identify trackers for further investigation – including language sphere specific and cross-language trackers. It may help to provide lines of inquiry for looking into what is distinctive about the commercial and technical underpinnings of fake news in different language spheres.

Chapter 4

STUDYING POLITICAL MEMES ON FACEBOOK

How can meme spaces on
Facebook be traced?

How may the content of memes
be studied?

How do memes frame political
and media events?

Introduction - So far the recipes in this guide have focused on media artefacts which mimic the news genre. But successful hyper-partisan content, misinformation, disinformation and propaganda do not always look and feel like news pages with the familiar combination of headlines, pictures and text that we see in sites like the *BBC*, *CNN* and countless other outlets. In fact images, and particularly image-based memes, circulate just as well (if not better) in social media ecosystems.

According to a piece for the *Columbia Journalism Review*, the media format which generated most engagement on *Breitbart's* Facebook page is the image-meme [1]. Hence, it is essential to consider not just how fake news pages but also other kinds of viral content genres such as memes participate in political agenda setting, political processes and political culture.

This section provides a set of approaches to investigate political memes. We focus on memes that take political topics, actors and events as their object. The case study used to illustrate these approaches is alt-right and pro-Trump memetic activity on Facebook around the 2016 US presidential election. We shall use the term "memetic activity" in this section to designate the multiple ways in which users act around memes online, including circulating, imitating and transforming them. The first recipe focuses on how to identify and map meme spaces on Facebook. The second recipe explores ways to investigate how Facebook users engage with political events through memetic activity. The

[1] See, Nausicaa Renner, "Memes Trump articles on Breitbart's Facebook page", *Columbia Journalism Review*: 2017. Available at: http://www.cjr.org/tow_center/memes-trump-articles-on-breitbart-facebook-page.php.

third and final recipe provides a series of approaches for analysing the content of memes.



HOW CAN MEME SPACES ON FACEBOOK BE TRACED?

BEFORE STARTING

To investigate who engages with memes around a political issue of interest on Facebook, you should start by identifying a page with significant following and memetic activity around your topic of interest. As an example, we selected the *Disdainus Maximus* page, which is very active in pro-Trump and alt-right activity. We traced the network of connections around this page and explored its topology.

START

Choose a Facebook page
with memetic activity



trace the network of connections with

Netvizz



extract page like network (depth 2)



import data in

Gephi



Quantitative Analysis

- InDegree
- OutDegree
- Betweenness Centrality
- Netvizz Fan Count Metrics



Qualitative Analysis

- Prominent clusters identification
- Cluster thematic annotation
- Exploration of content and self-description of pages



visualise



a

**WHAT ISSUES ANIMATE
THE INTER-LIKED FACEBOOK
PAGE NETWORK SEEDED
BY A PRO-TRUMP POLITICAL
MEME REPOSITORY?**

START

Choose a Facebook page
with **memetic activity**



trace the **network of connections** with



extract page **like network** (depth 2)



import **data** in



Quantitative Analysis

- InDegree
- OutDegree
- Betweenness Centrality
- Netviz Fan Count Metrics



Qualitative Analysis

- Prominent clusters identification
- Cluster thematic annotation
- Exploration of content and self-description of pages



a

**WHAT ISSUES ANIMATE
THE INTER-LIKED FACEBOOK
PAGE NETWORK SEEDED
BY A PRO-TRUMP POLITICAL
MEME REPOSITORY?**

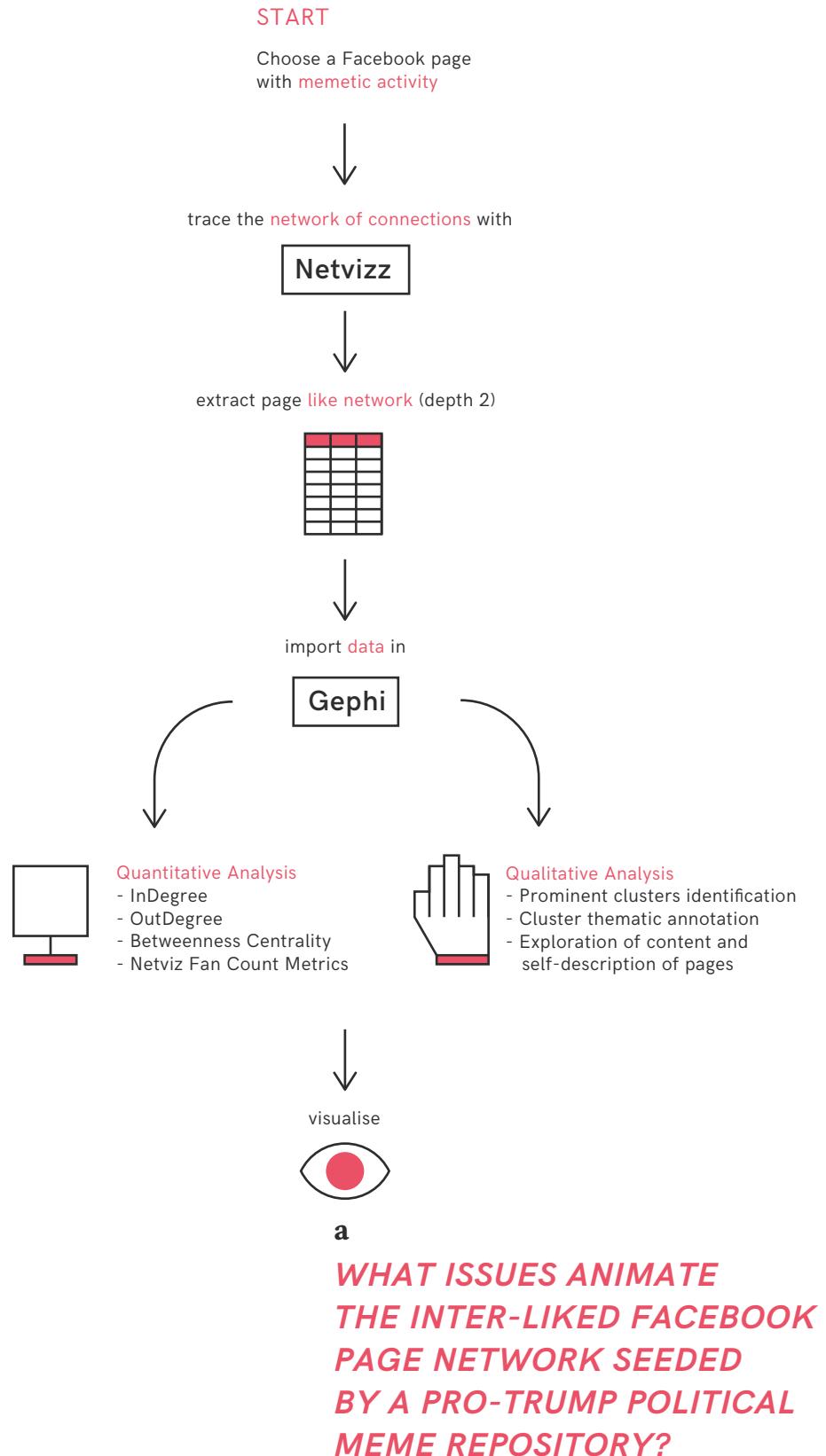
IDENTIFY THE NETWORK OF TIES AROUND A CHOSEN FACEBOOK PAGE

To trace the network of affinities around a Facebook page we followed the “likes” from our page to other pages (to be distinguished from the "likes" received from users).

- ◊ A Facebook crawler may be used to extract the “likes” network around a page. We used  Netvizz’s “page like network” module to “create a network of pages connected through the likes between them.”
- ◊ We set the crawler to a depth of two to extract the pages liked by our seed page and those liked by them. We thus obtained a directed network file where nodes are pages and edges represent acts of liking.
- ◊ You may use a network analysis tool such as  Gephi to examine the graph of interconnected Facebook pages. A force-directed layout algorithm (such as ForceAtlas2) can help you visualise the shape of the network and explore the interconnected space of memetic activity.

[1]

See Netvizz, Facebook application, version: 1.3, 2017, at: <https://apps.facebook.com/netvizz/>



PROFILE THE ISSUES AND THEMES THAT ANIMATE THE MEMETIC SPACE

This step consists of qualitative and quantitative analysis of the composition and arrangement of the network of Facebook pages obtained in the previous step. The configuration of the → **network graph** may be analysed:

- Quantitatively, by:

- ◊ Identifying which pages are most popular in the network by using a graph metric such as indegree, i.e. the count of the likes received from other pages in the network.
- ◊ Identifying the pages most active in liking other pages by using a graph metric such as outdegree, i.e. the count of likes given to other pages in the network.
- ◊ Identifying which pages bridge or connect different clusters in the network by using a graph metric such as betweenness centrality.
- ◊ Identifying which pages are most popular among Facebook users by using the Facebook's "fan count" metric.

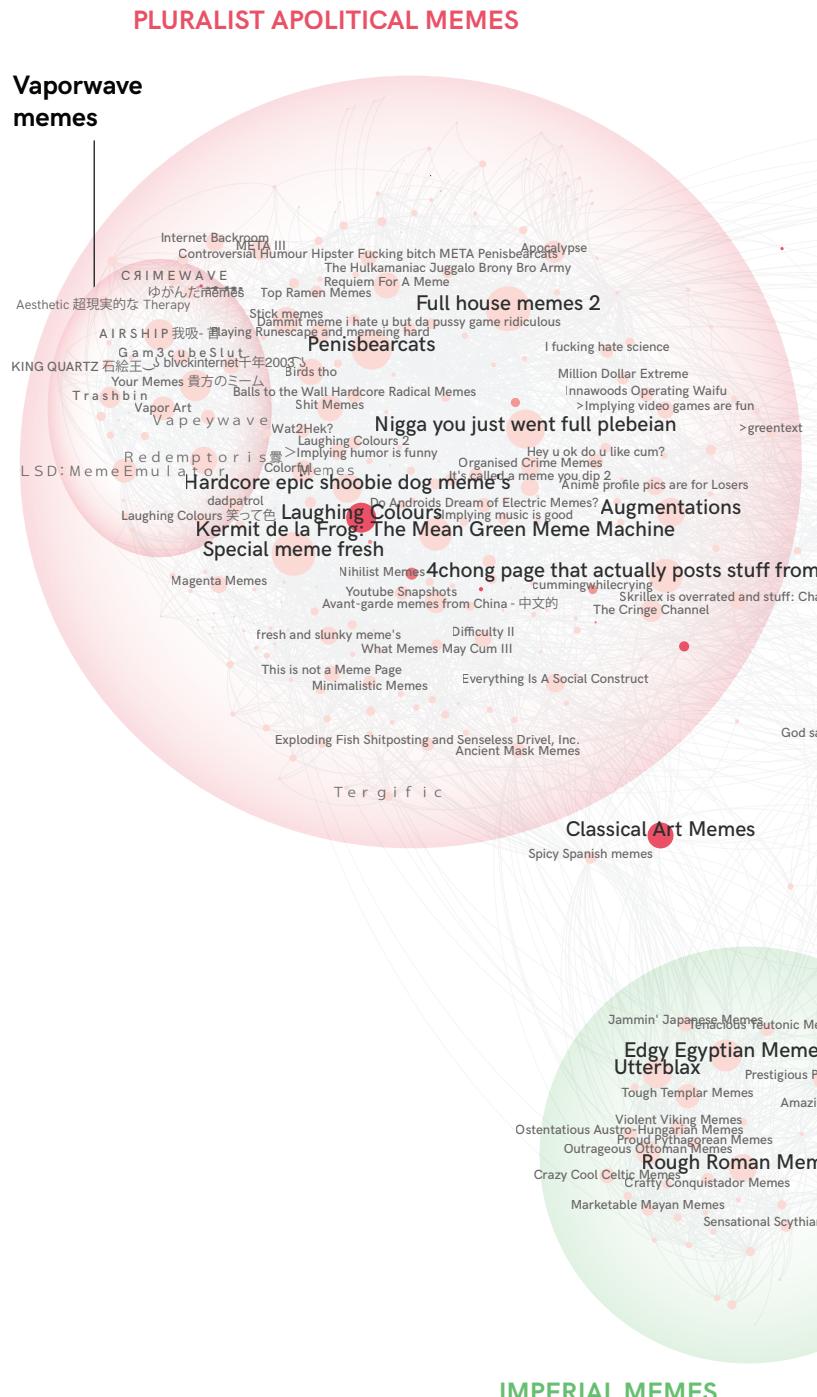
- Qualitatively, by:

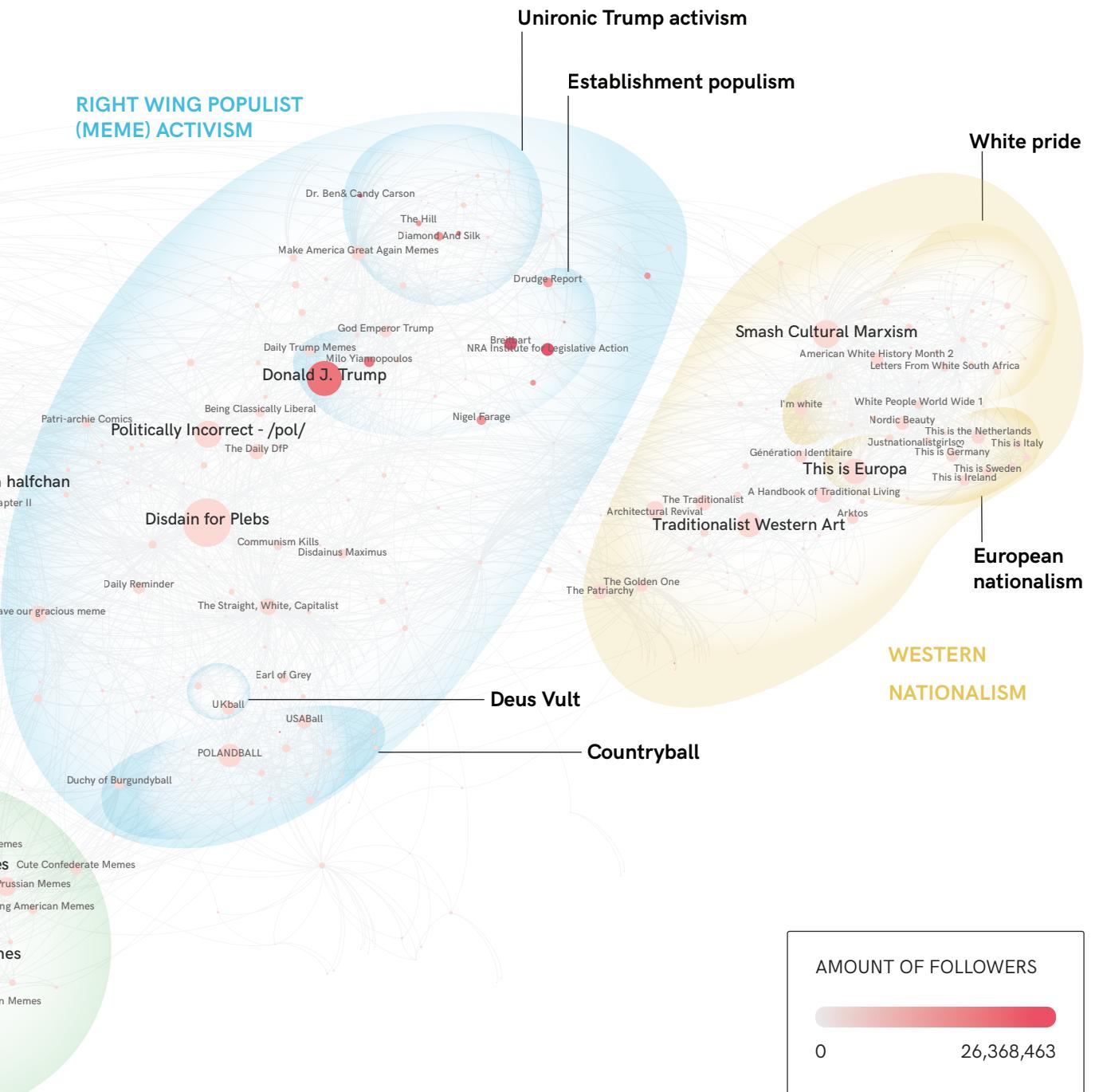
- ◊ Identifying prominent clusters by visually exploring the shape and density of nodes groupings in the graph.
- ◊ Examining the content shared by the pages in the network as well as their titles and self-descriptions to identify shared issues of concern within each cluster.

To increase readability of the network map you may annotate it with the resulting thematic classification of the clusters.

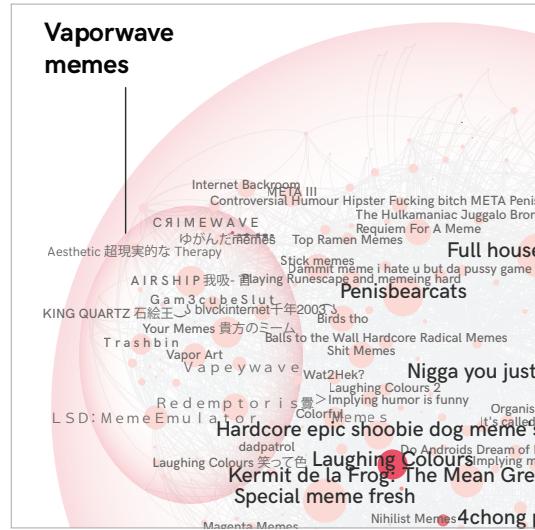
WHAT ISSUES ANIMATE THE INTER-LIKED FACEBOOK PAGE NETWORK SEEDED BY A PRO-TRUMP POLITICAL MEME REPOSITORY?

Network of inter-linked Facebook pages seeded from the pro-Trump meme repository *Disdainus Maximus*. The network comprises 751 pages featuring memetic activity. Pages are connected through ties representing likes among them. The network is spatialised with a force-directed layout algorithm. Another algorithm, modularity, is used to identify clusters. Nodes are sized according to the number of likes received and coloured according to their follower count. Pro-Trump memes are prominent within the Facebook meme culture, as shown by the fact that even the pages which are not politically oriented share ties with pages circulating pro-Trump memes (as well as with pages dedicated to various forms of nationalism and populism). The presence of Donald Trump's official page at the center of the network may indicate that his image plays an energising role in the alt-right meme space.





CHAPTER 4 → RECIPE 1



SERVING SUGGESTIONS

This recipe may be used to identify political meme repositories on Facebook and the relations between them as well as the themes that lend themselves to memefication.

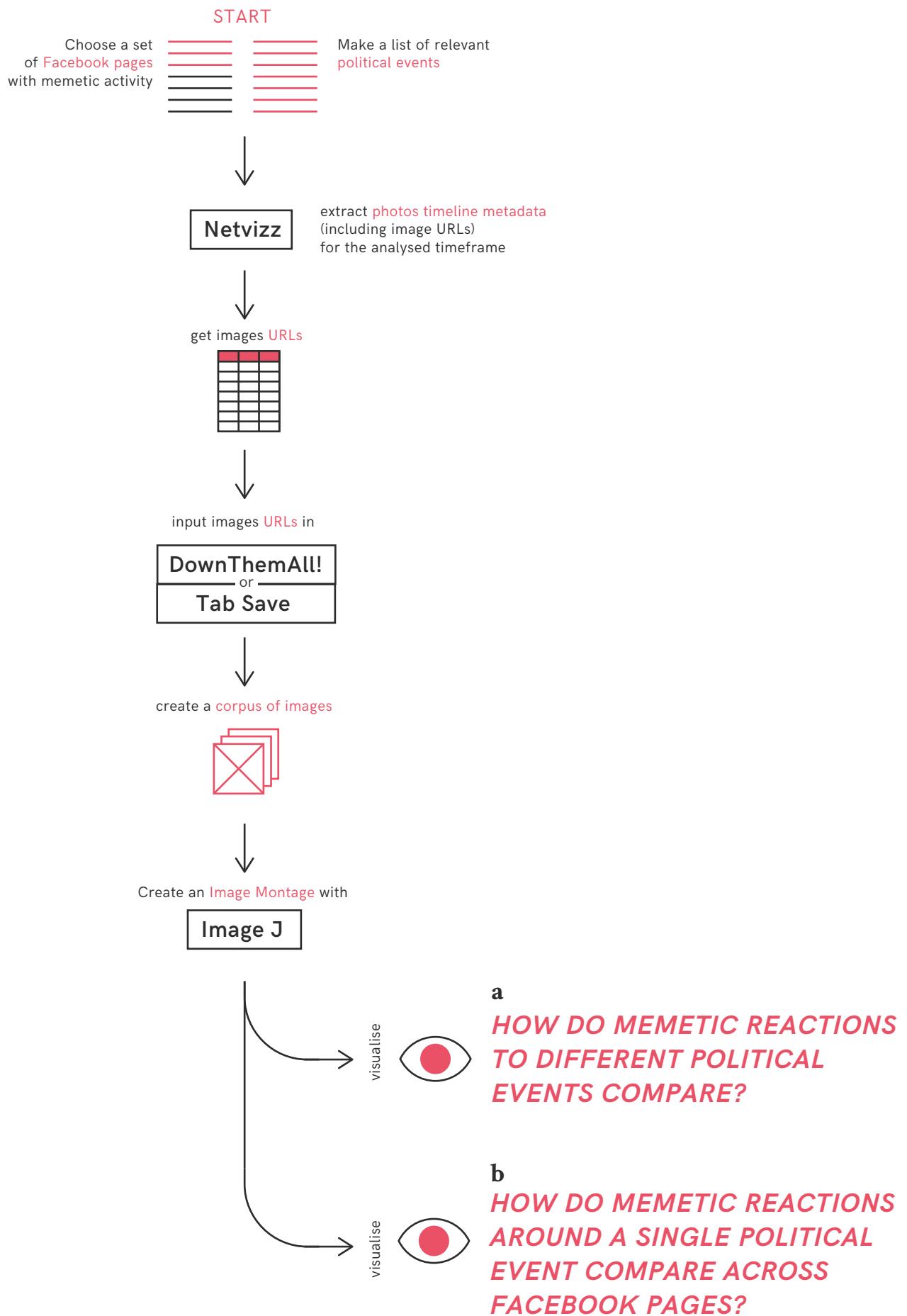


HOW DO MEMES FRAME POLITICAL AND MEDIA EVENTS?

BEFORE STARTING

To identify how memes frame political or media events the first step is to identify the events to be examined. To illustrate this recipe we used the following key events associated with the US elections:

| Date | Description |
|-------------------|---|
| March 3, 2016 | The eleventh Republican debate |
| June 3, 2016 | Trump's Mexican judge remark |
| September 9, 2016 | Clinton's 'basket of deplorables' comment |
| October 8, 2016 | Trump's taped comments about women |
| October 29, 2016 | Podesta emails |
| January 27, 2017 | Trump announces first travel ban |



IDENTIFY MEMES RELATED TO EVENTS

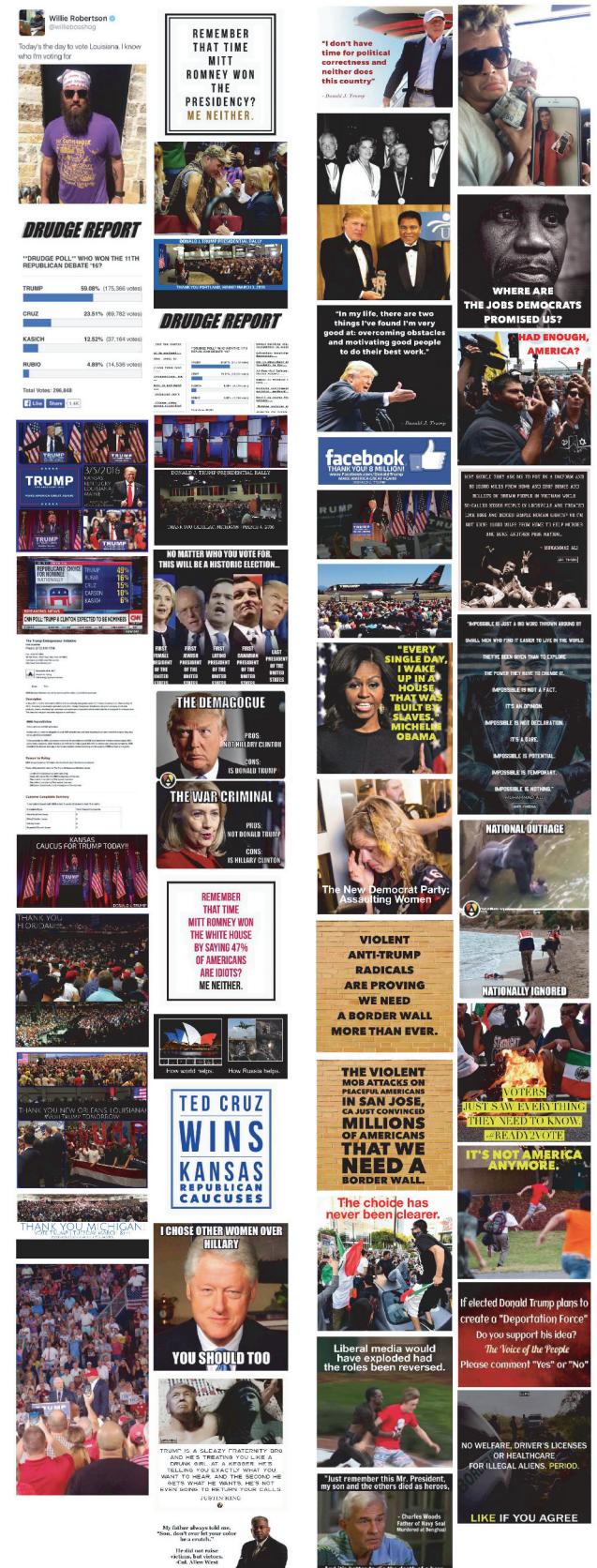
To illustrate this recipe we use a sub-selection of 46 pages from the corpus identified in recipe 4.1. The selection was done based on a set of qualitative and quantitative criteria, including their → **engagement** counts and the thematic cluster that they belong to.

- ◊ Define a timeframe for meme selection.
For this analysis, we selected three days starting with the date of the event under examination.
- ◊ Extract a list of memes posted to your corpus of pages in the selected timeframe. You may use a tool such as  **Netvizz** to collect a list of images and related metadata.
- ◊ Download the images for each URL. You may use a browser extension such as  **Tab Save** for Chrome or  **DownThemAll!** for Firefox.
- ◊ Visually juxtapose the images grouping them by event to enable comparison of memetic reactions across events. You may use an image montage tool such as  **ImageJ** (4.2.a).
- ◊ You can also explore how different pages react to a single event to enable comparison across pages (4.2.b).

HOW DO MEMETIC REACTIONS TO DIFFERENT POLITICAL EVENTS COMPARE?

Image montage of memetic activity on selected Facebook pages around six key events in the 2016 US presidential campaign. Memetic activity around different events may be compared in terms of scale and framing.

Eleventh Republican debate; blood coming out of her... wherever
3rd of March 2016



Trump's Mexican Judge remark

3rd of June 2016

Basket of Deplorables comment

11th of September 2016



HOW DO MEMETIC REACTIONS AROUND A SINGLE POLITICAL EVENTS COMPARE ACROSS FACEBOOK PAGES?

Image montage of a particular interpretative frame prominent in memetic activity around Trump's taped comments about women across a set of pro-Trump Facebook pages.

The visualisation shows how a number of pages pointed to a particular image of Bill Clinton confronted with alleged victims of sexual harassment from his own past in response to the "groping tape" event. Notably, this interpretative frame appears consistent with the frame proposed by *Breitbart*.



FIRST A 01:55 (GMT)



B 02:21



C 02:28



I 04:30



J 05:10

thealexjoneschannel



K 05:11

GET



P 11:28



Q 14:39

wakeupandreclaimamerica

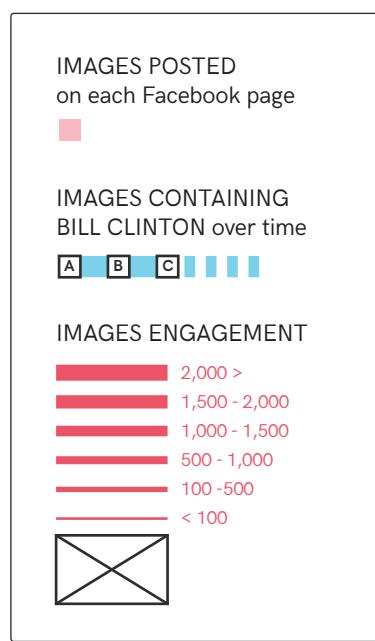
GET

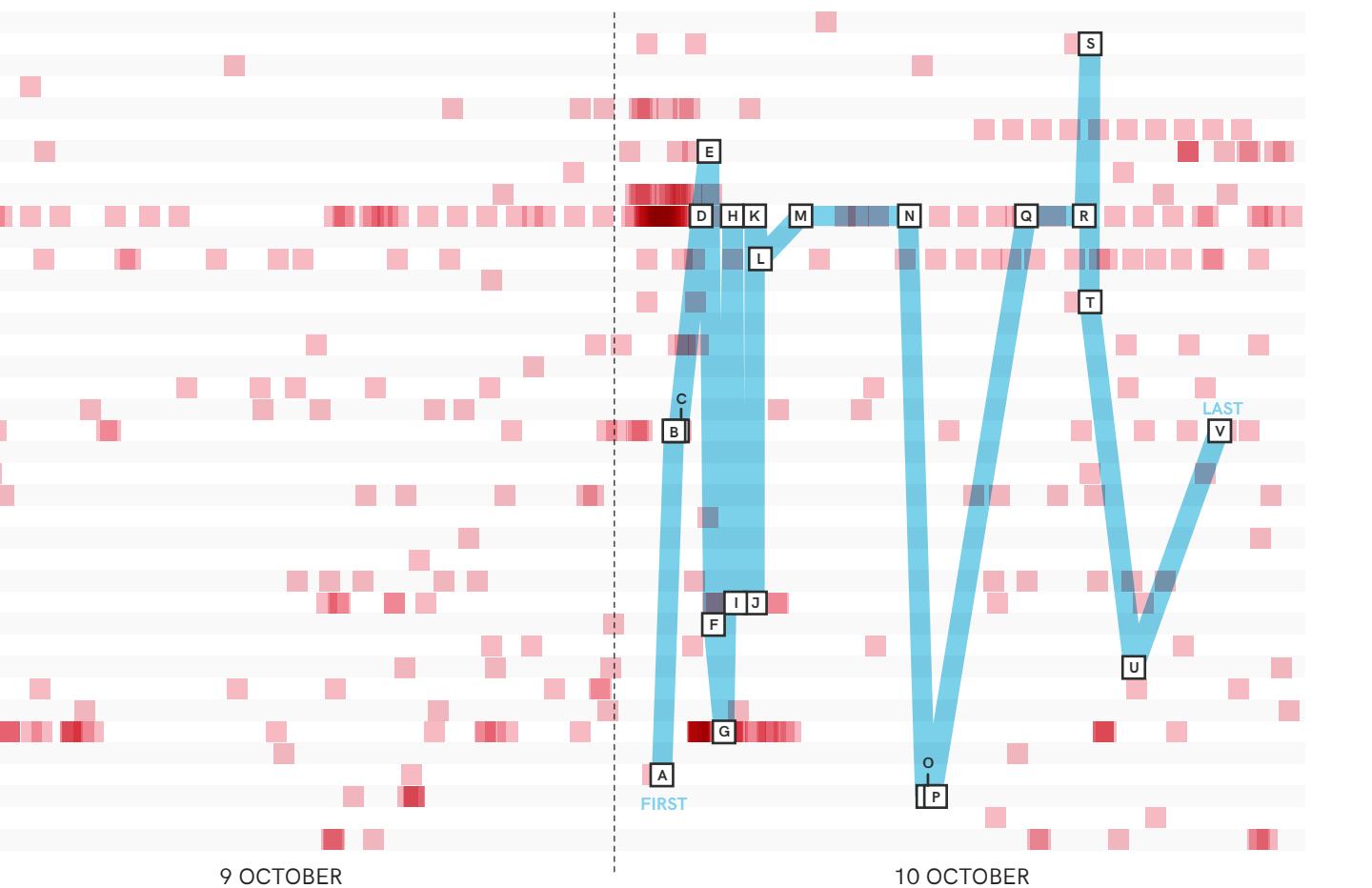


R 16:40

wakeupandreclaimamerica

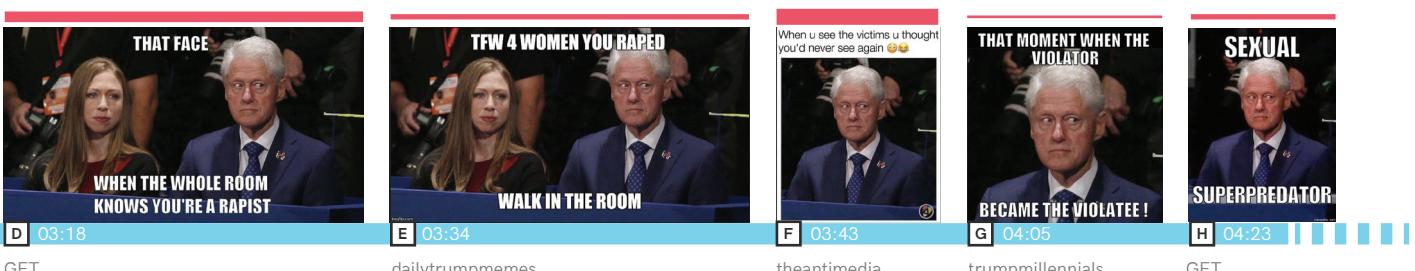
GET





9 OCTOBER

10 OCTOBER



GET

dailytrumpmemes

theantimedia

trumpmillennials

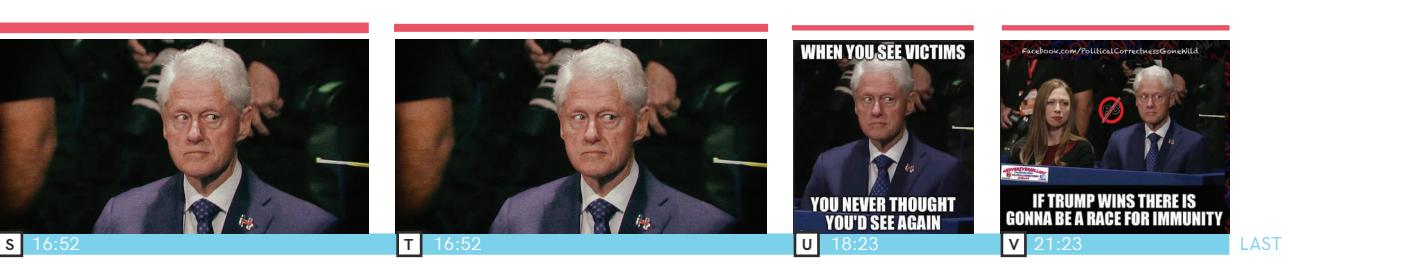
GET



guardiansoffreedom.com.au

GET

wakeupandreclaimamerica

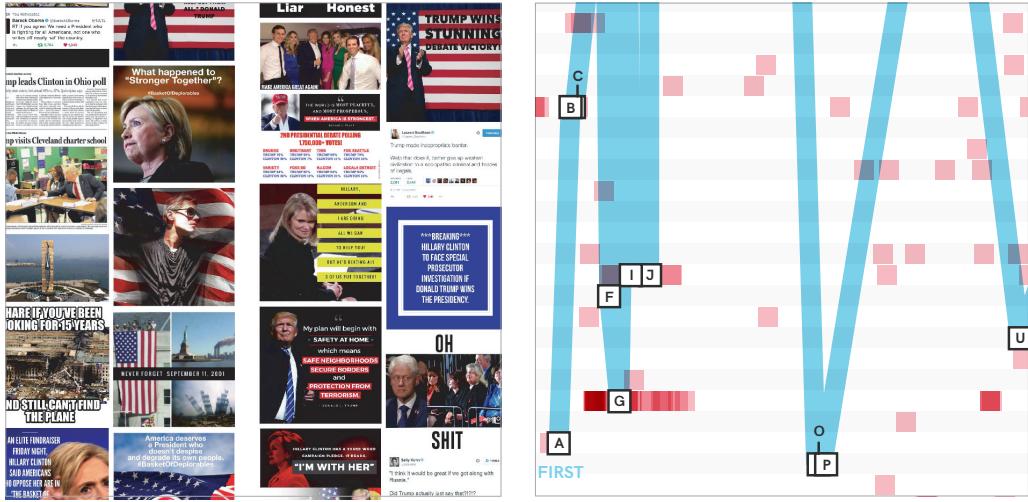


alexanderemerickjones

infowars

thedailypfp

politicalcorrectnessgonewild



SERVING SUGGESTIONS

This recipe may be used to explore participatory production of visual culture around political events and how it contributes to agenda-setting efforts.



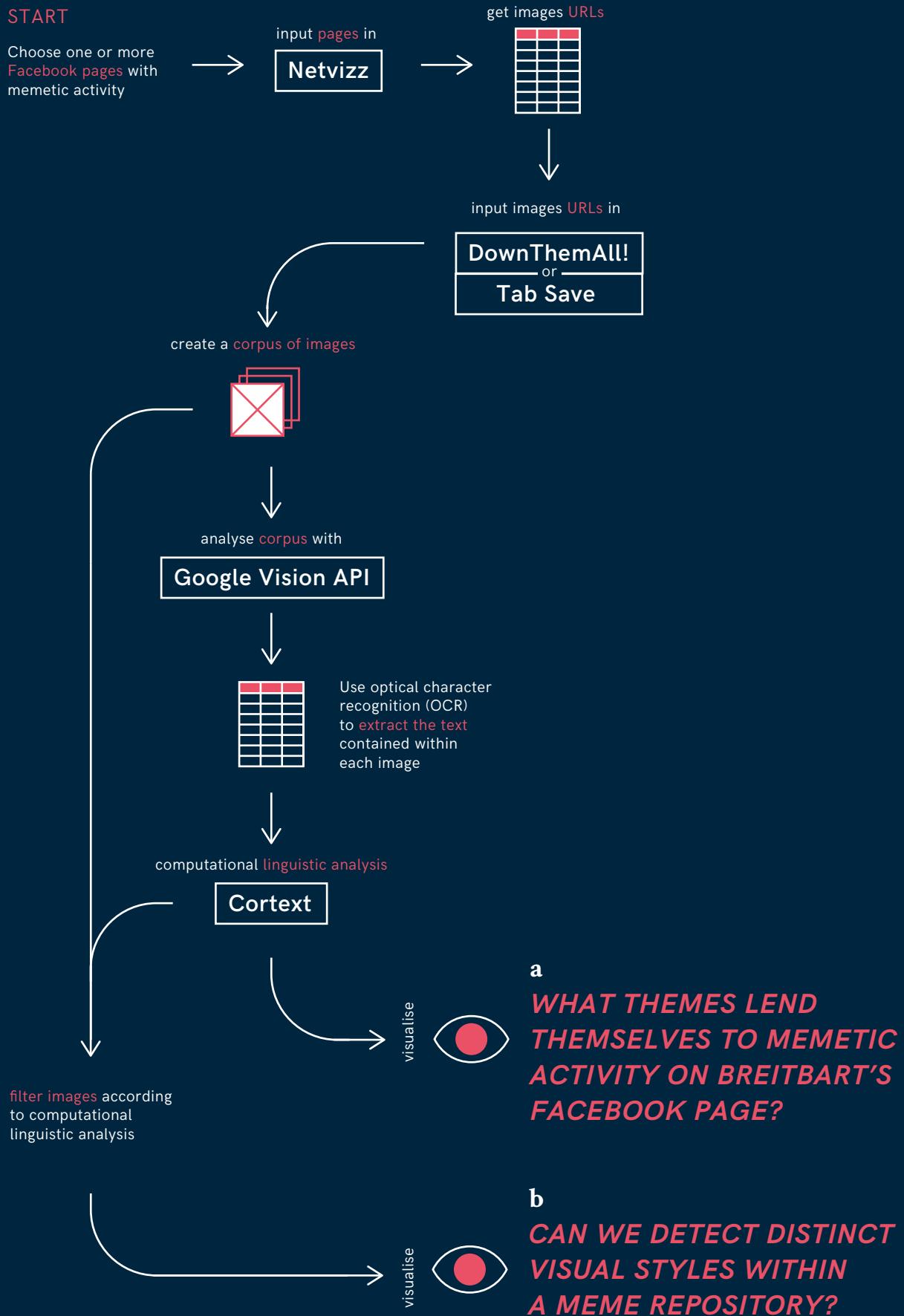
HOW MAY THE CONTENT OF MEMES BE STUDIED?

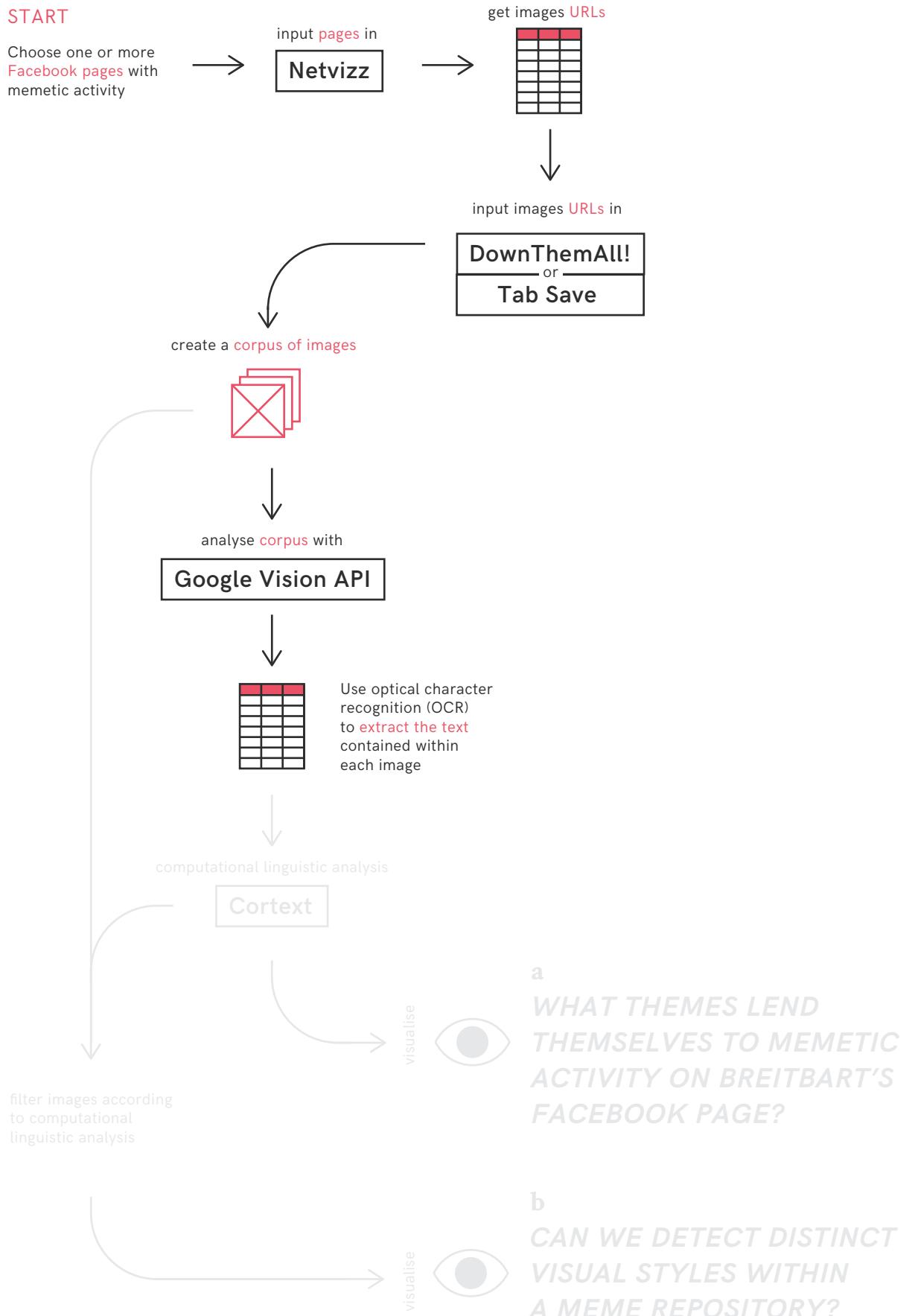
BEFORE STARTING

The recipe illustrates techniques to support content analysis of texts and images contained in memes as well as the detection of genres or styles of memetic activity.

As examples, we use two Facebook pages which feature pro-Trump memes: *Breitbart* and *God Emperor Trump*. Even if the images posted to these pages do not comply with classic meme formats (e.g. image macros), they exhibit memetic features such as virality, user-driven remixing, imitation and intertextuality. *Breitbart* has been selected due to its central role in animating the alt-right culture^[1]. The *God Emperor Trump* page is one of the most popular pro-Trump, alt-right meme pages with over 245,000 likes and → followers as of the time of writing. In *Breitbart* only the page administrator can post images, while in *God Emperor Trump* users may submit their productions to the administrator for posting.

[1] Y., Faris, R., Roberts, H., & Zuckerman, E. (2017, March 3). Study: Breitbart-led right-wing media ecosystem altered broader media agenda. Retrieved March 8, 2017, from <http://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>





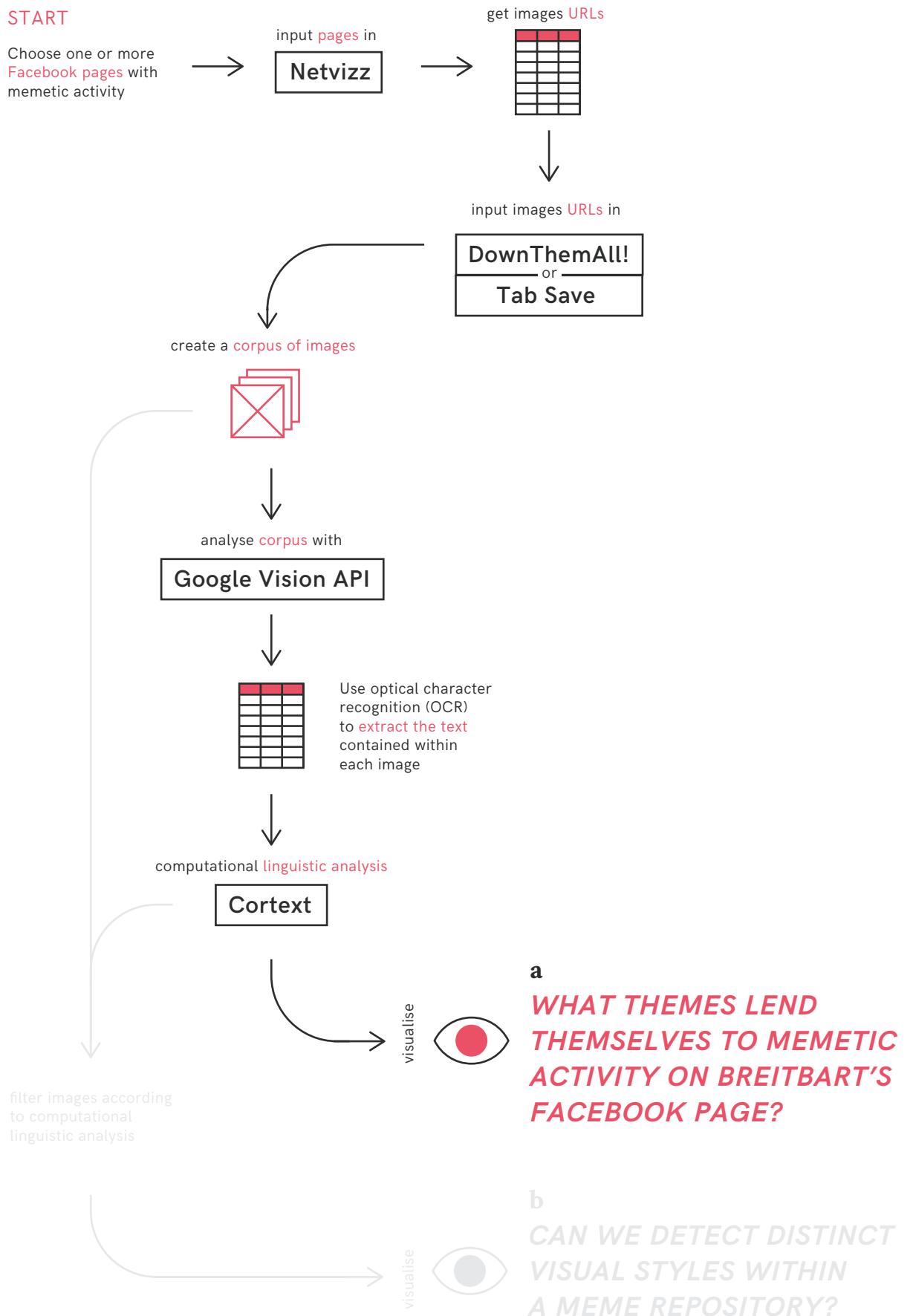
CREATE A CORPUS OF IMAGES POSTED TO THE CHOSEN FACEBOOK PAGE AND THEIR ASSOCIATED METADATA

To create a corpus choose a timeframe of interest (for instance the days around a particular political or media event - see recipe 4.2) or use all images posted to a Facebook page.

- ◊ The Facebook → API enables the extraction of metadata associated with images posted to a page and available via the “Photos” tab.
- ◊ Metadata capture may be done with a data extraction tool such as ↗ Netvizz, using the “page timeline images” module.
- ◊ The outcome of running ↗ Netvizz’s “page timeline images module” is a tab-separated file containing metadata associated with each image, including its creation date, its URL and its “likes”, reactions and comment count.
- ◊ A browser extension such as ↗ Tab Save for Chrome or ↗ DownThemAll! for Firefox may be used to download the images.
- ◊ To extract the text contained within each image you may run the images through some optical character recognition (OCR) software. For this recipe, we used ↗ Google's Vision API and a script feeding the list of image URLs to the Vision API^[2].
- ◊ You may also use a piece of image analysis software to generate additional metadata for your corpus of images through the detection and labelling of entities, objects and attributes.

[2]

See memespector script written by Bernhard Rieder, University of Amsterdam, at <https://github.com/bernorieder/memespector>



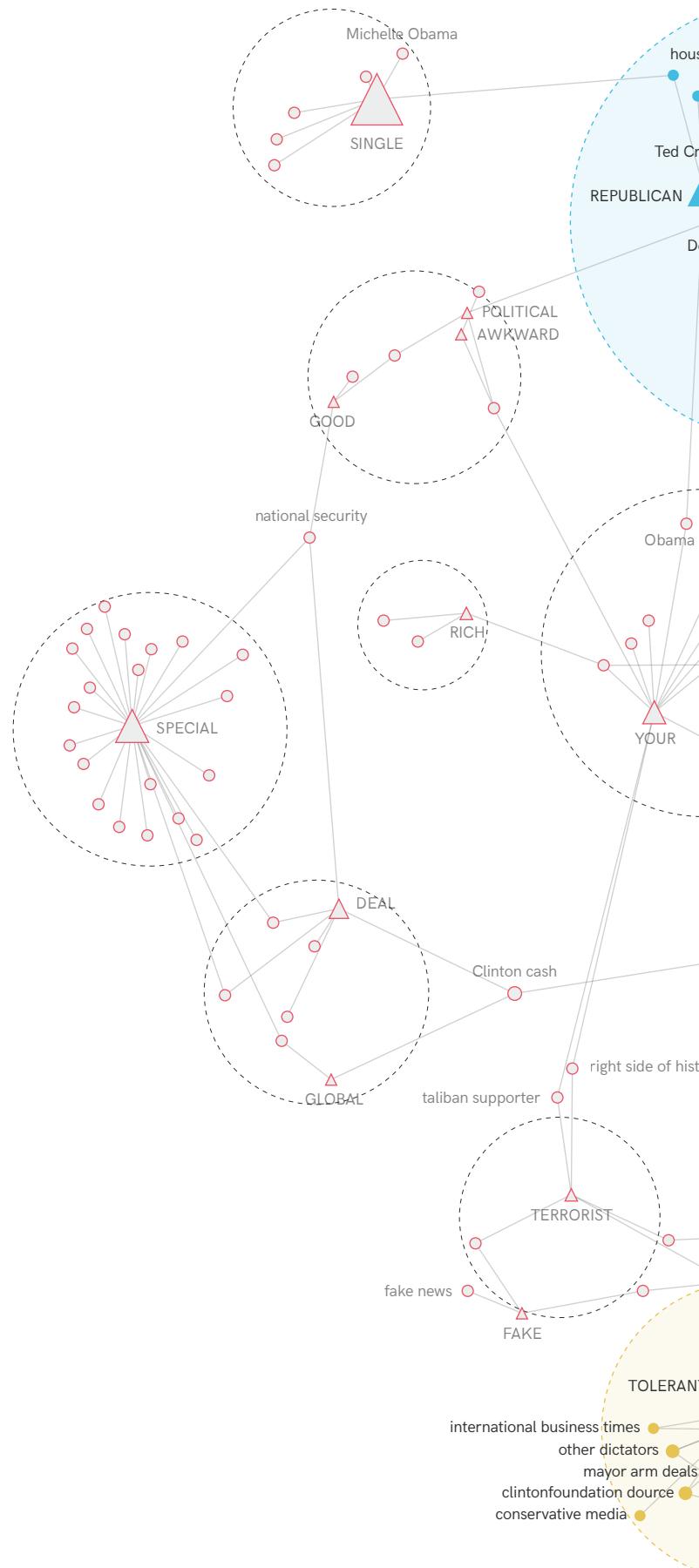
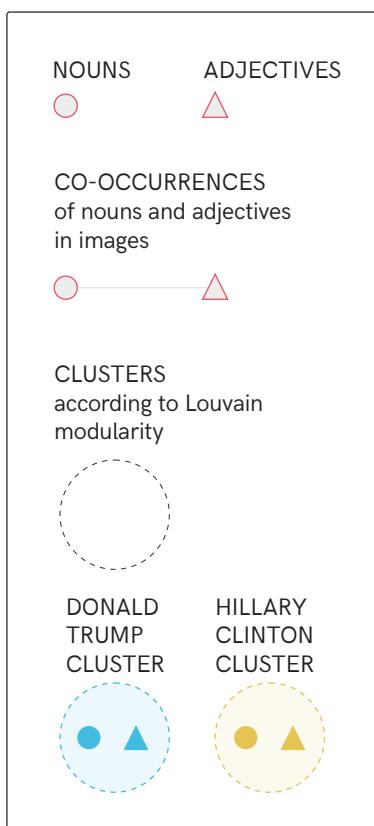
EXAMINE THEMES EXPLOITED IN THE CORPUS OF MEMES WITH TEXT ANALYSIS (EXPERIMENTAL)

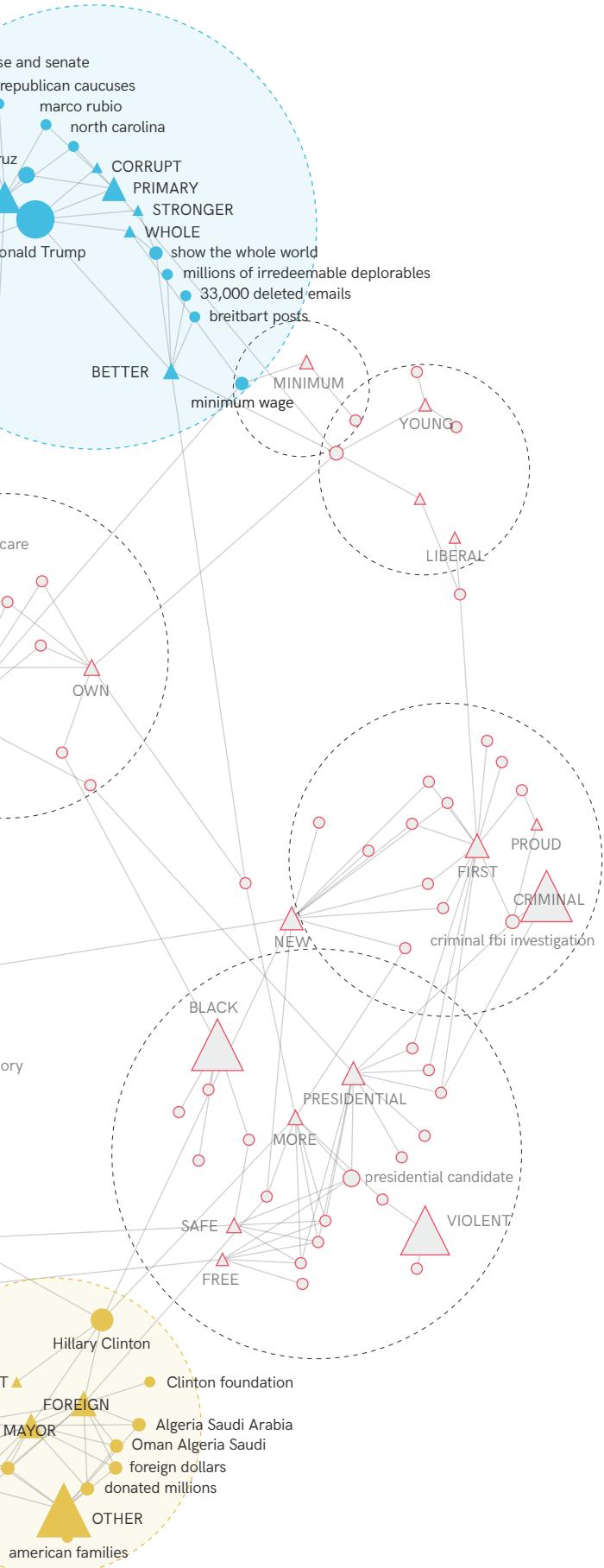
To examine the issues that trigger memetic activity you may analyse the text extracted from the images through manual qualitative analysis or semi-automated semantic analysis. The results of this analysis are affected by the quality of the OCR. A computational linguistics tool (e.g.  CorText) can be used, but the analyst's judgement remains crucial to evaluate the relevance of the extracted terms and to set the parameters of the tool (what algorithms to use, what types of words to keep, how frequently should they occur, etc.).

- ◊ Lexical analysis may help you identify the most relevant terms used in your memes.
- ◊ You may also run queries on the OCR outputs to explore the resonance of particular issues.
- ◊ Analysis of term co-occurrences across the corpus of memes allows you to explore the main themes within the corpus and their relationships. The network of term co-occurrence may be visually explored in order to identify prominent thematic clusters and the relationships between them and identify how key political issues are articulated through associations of terms in memes.

WHAT THEMES LEND THEMSELVES TO MEMETIC ACTIVITY ON BREITBART'S FACEBOOK PAGE?

Network of nouns and adjectives co-occurring in images posted in 2016 on Breitbart's Facebook page (two words are connected if they are present in the same image). Colors identify clusters according to Louvain modularity. The two most prominent clusters are centered around Donald Trump (top, blue) and Hillary Clinton (bottom, yellow). One may examine the terms present in the Hillary Clinton cluster and in clusters in its proximity in terms of framing and agenda setting.





most specific images for the **Hillary Clinton** cluster



HILLARY: YOU CAN'T



LEAD THOSE YOU HATE.



#BASKETOFDEPLORABLES



THE FIX IS IN!
BOMBSHELL
HILLARY CLINTON'S
FBI TESTIMONY
WASN'T
UNDER OATH
OR EVEN RECORDED.

FACT: HILLARY CLINTON APPROVED MAJOR ARMS DEALS TO QATAR, OMAN, ALGERIA, SAUDI ARABIA AND OTHER DICTATORS AFTER THEY DONATED MILLIONS OF FOREIGN DOLLARS TO THE CLINTON FOUNDATION.
Source: International Business Times

most specific images for the **Donald Trump** cluster

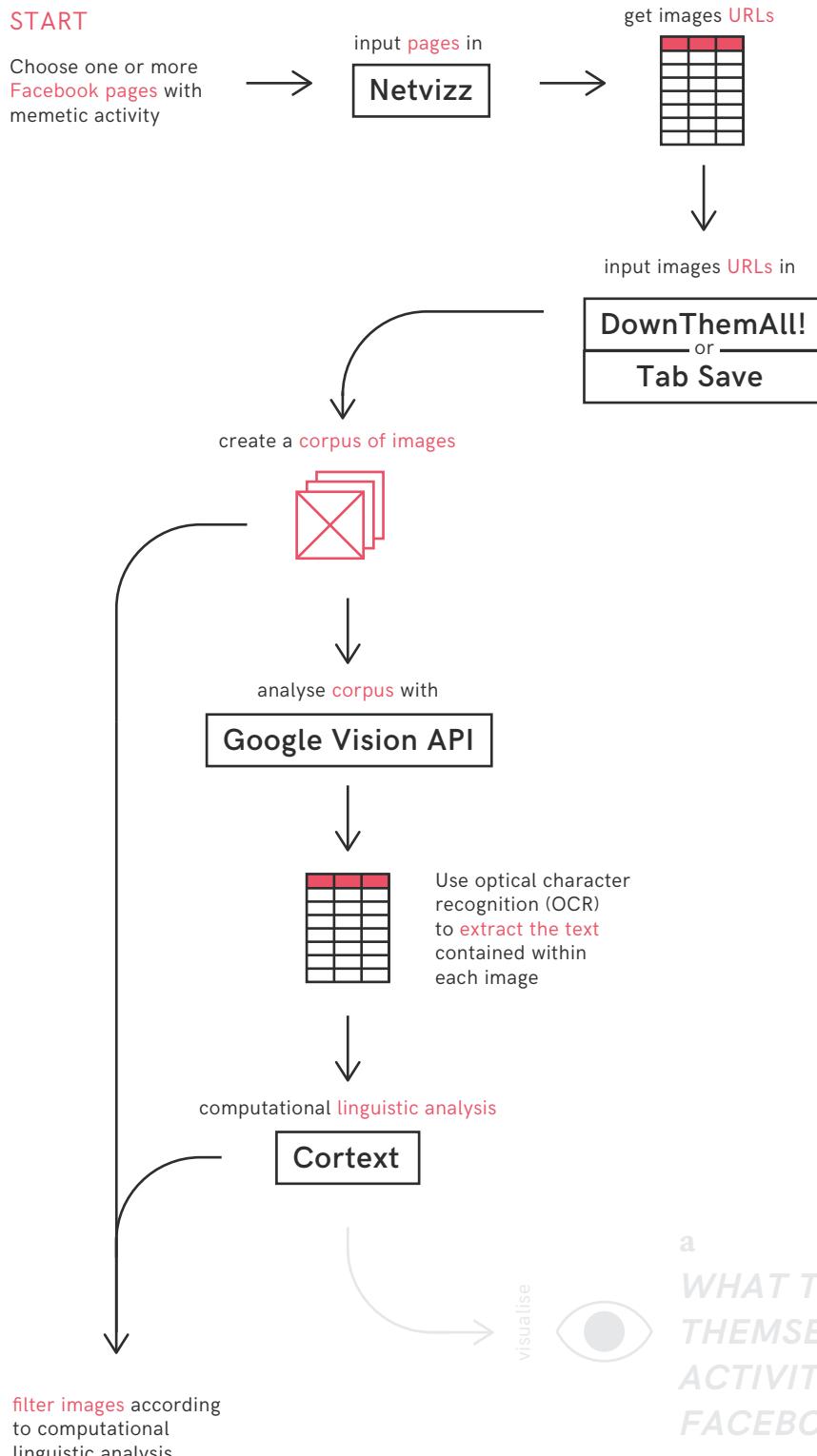


Donald Trump

Wins Maryland
Republican Primary



Marco Rubio
Wins Minnesota
Republican
CAUCUSES



a
WHAT THEMES LEND THEMSELVES TO MEMETIC ACTIVITY ON BREITBART'S FACEBOOK PAGE?

b
CAN WE DETECT DISTINCT VISUAL STYLES WITHIN A MEME REPOSITORY?

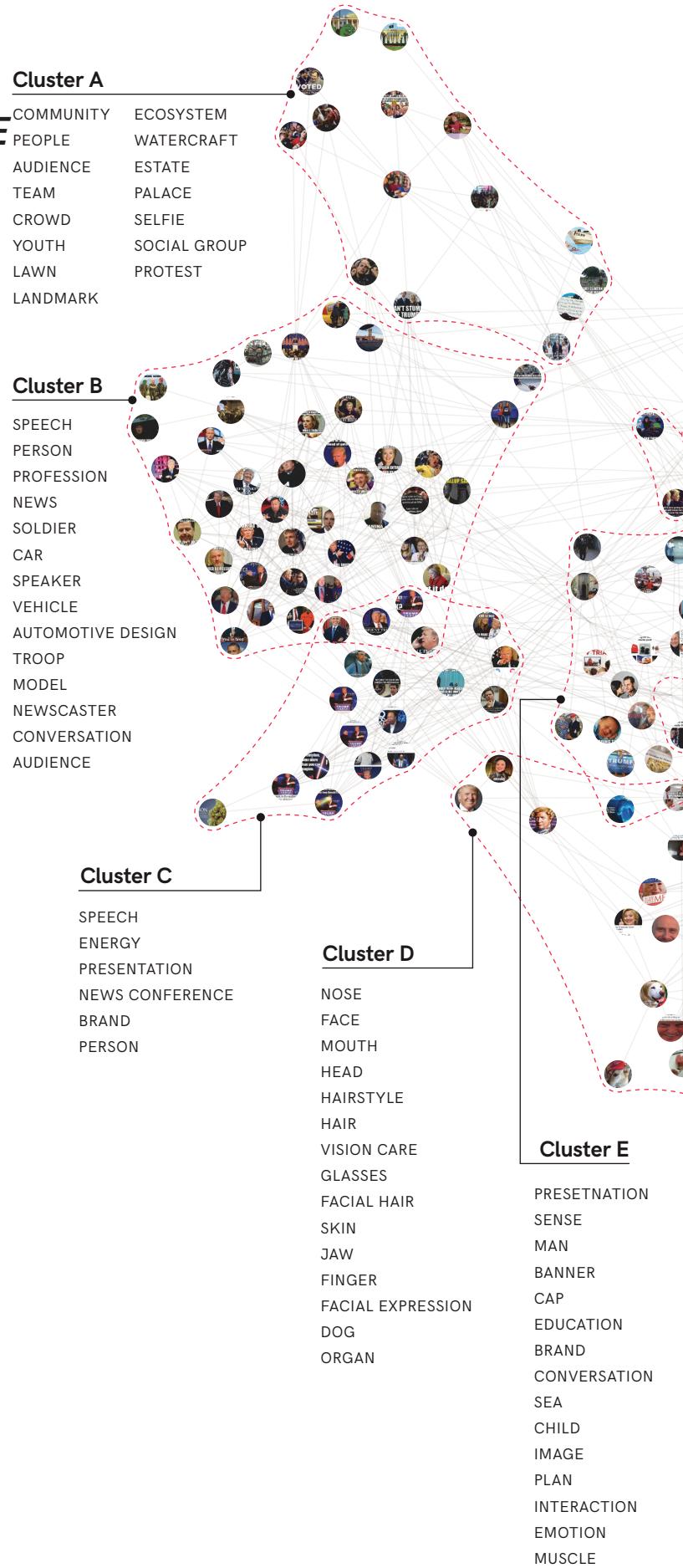
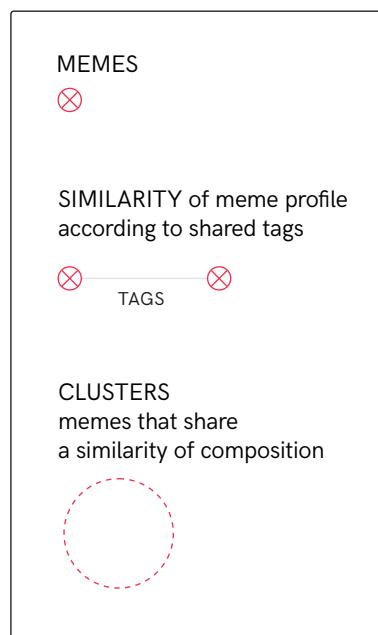
EXAMINE VISUAL STYLES OF MEMES WITH IMAGE ANALYSIS

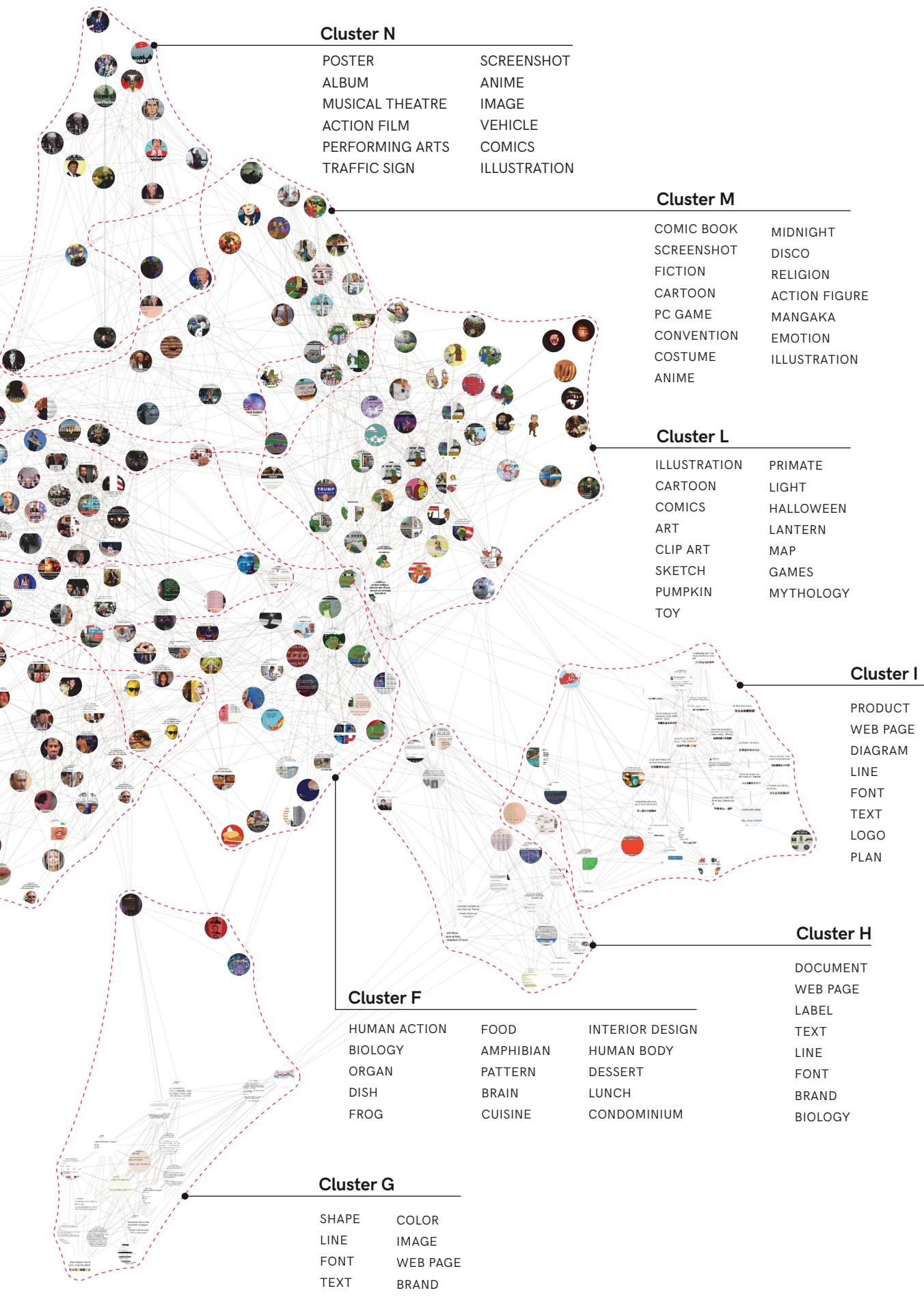
To explore the visual styles deployed in a meme repository, the outputs of the image analysis software described in step 4.2.a may be used, namely the labels or tags generated to describe the entities and attributes detected in our image corpus. We illustrate this analysis on the *God Emperor Trump* page.

- ◊ Images are analysed with Google Vision to extract tags describing the images and textual content.
- ◊ We used  CorText to examine associations between images based on shared labels.
- ◊ The configuration of the → **network graph** may be visually explored in order to identify and describe visual styles per cluster.
- ◊ This operation may be repeated with different pages to compare their different style.

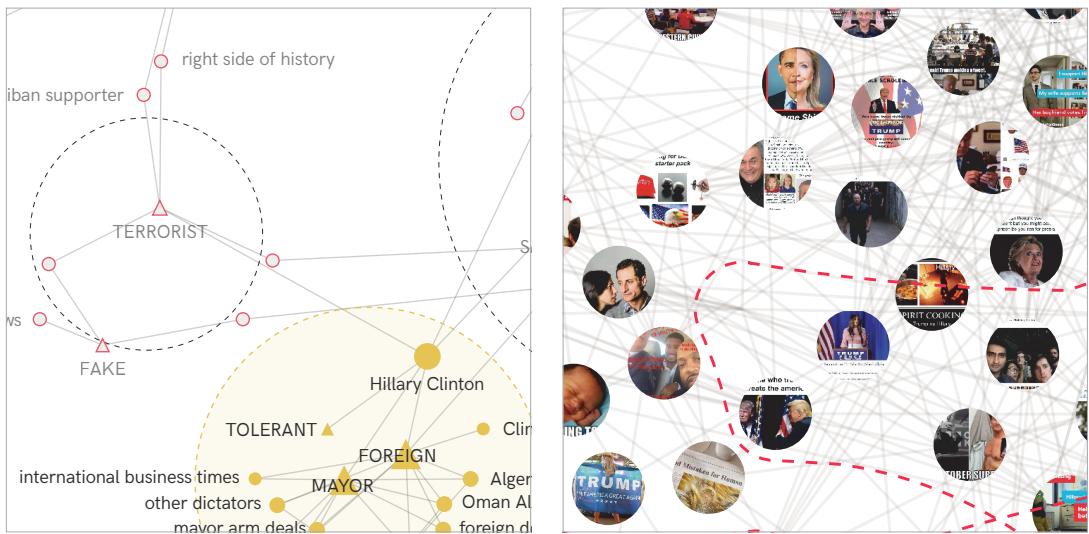
CAN WE DETECT DISTINCT VISUAL STYLES WITHIN A MEME REPOSITORY?

Network of images in the *God Emperor Trump* corpus, linked by similarity of meme profile, according to shared tags. The fuzzy boundaries between clusters show that memes share a similarity of composition, which may be associated with the memetic logic. Some genres or styles of memetic activity may be detected. A screenshot-based meme cluster may be noticed as well as a comics and cartoons cluster, including Pepe the Frog.





CHAPTER 4 → RECIPE 3



SERVING SUGGESTIONS

This recipe may be used to explore how memes frame political issues, events and personalities, in what may be considered a form of “participatory propaganda”.^[3]

[3] See, Alicia Wanless, "Have You Fallen Down a Participatory Propaganda Rabbit Hole?", *Politics Means Politics*, at: <https://politicsmeanspolitics.com/have-you-fallen-down-a-participatory-propaganda-rabbit-hole-6f71c83f04fa>

Chapter 5

MAPPING TROLL-LIKE PRACTICES ON TWITTER

How may we detect Twitter accounts which negatively target political representatives?

How may we characterise the sources of troll-like activity?

How may troll-like practices be characterised?

Introduction - Tactics such as trolling and the use of bots and “sock-puppet” accounts have been linked to the spread of political disinformation and propaganda online [1]. In the lead up to the 2017 general elections for the Dutch parliament, journalists pointed to the use of sock puppets (i.e. false online identities assumed to deceive and influence opinion) by some political parties to amplify their messages online and to attack their political opponents on social media [2].

In this recipe set we provide some methods that can be employed to detect and profile misleading information practices by taking troll-like behaviour around the 2017 Dutch election campaign as a case study. We focus on political troll-like practices, a term which we use in the narrower sense of attacks addressed at political representatives.

We focus on three aspects of political trolling: the sources of troll-like activity, the characteristics of these practices and their targets.

[1] See, for example, Alice Marwick and Rebecca Lewis, "Media Manipulation And Disinformation Online", *Data Society*, May 2017: <https://datasociety.net/output/media-manipulation-and-disinfo-online/>

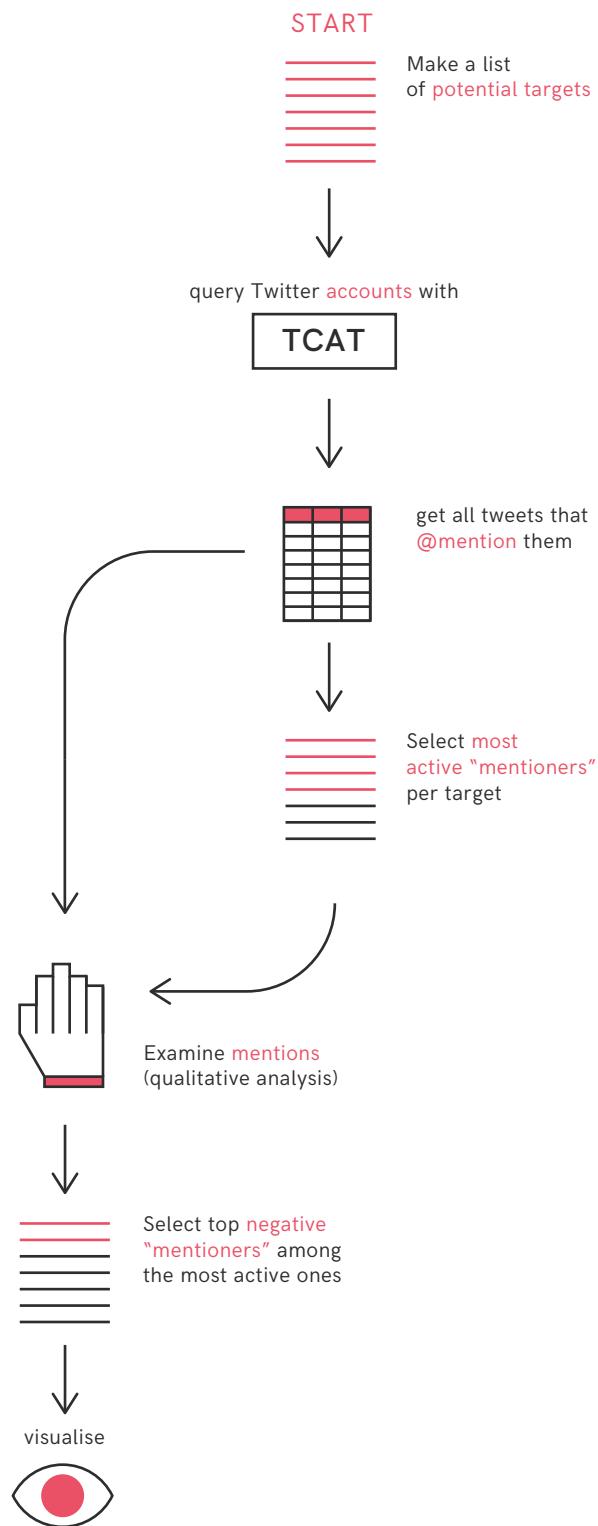
[2] See, Andreas Kouwenhoven and Hugo Logtenberg, "Hoe Denk met 'trollen' politieke tegenstanders monddood probeert te maken", *NRC*, 2017. Available at: <https://www.nrc.nl/nieuws/2017/02/10/de-trollen-van-denk-6641045-a1545547>



HOW MAY WE DETECT TWITTER ACCOUNTS WHICH NEGATIVELY TARGET POLITICAL REPRESENTATIVES?

BEFORE STARTING

To detect political troll-like activities and their sources, you should start from compiling a list of potential targets, for instance a set of Twitter accounts associated with political representatives. We used a set of 28 Twitter accounts associated with the leaders of parties participating in the 2017 Dutch general elections.



a

**HOW ARE ATTACKS DISTRIBUTED
ACROSS THE POLITICAL SPECTRUM
IN THE NETHERLANDS?**

IDENTIFY WHO NEGATIVELY TARGETS POLITICAL LEADERS ON TWITTER

To identify who negatively mentions a political leader on Twitter start from identifying all references of a political leader account on Twitter. Following Twitter practices, you can operationalise references with "mentions" defined by the "@" sign.

- ◊ Capture all @mentions of a set of accounts.
We used a Twitter data extraction and analysis toolset called  TCAT.
- ◊ We capture all tweets mentioning at least one of the 28 political leader Twitter accounts between 8 February and 8 March 2017, the month before the Dutch general elections. The result is a collection of 519,245 tweets which we use in all the recipes in this set.
- ◊ To identify the most active “mentioners”, find all the accounts mentioning one of the target accounts more than a given threshold. In our example, we retained the accounts mentioning one of the political leaders more than 100 times in our dataset.
- ◊ Examine mentions by most active “mentioners” identified in the previous step to qualify the nature of each of their references (i.e. whether it is in support of the political leader or negatively targeting them).
- ◊ Retain only the users who consistently negatively target one or more political representatives (in our case we narrowed down our list to 25 users).

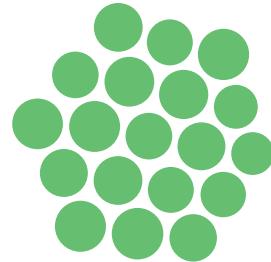
HOW ARE ATTACKS DISTRIBUTED ACROSS THE POLITICAL SPECTRUM IN THE NETHERLANDS?

Visualisation of Dutch political leaders who are targets of positive and negative mentions on Twitter (from users who mention them more than 100 times in a one-month period). Red circles represent users launching attacks and green circles represent users making favourable mentions. The size of the circle represents the total number of users mentioning a party leader. The asymmetry of the distribution of targets of troll-like behaviour across the political spectrum is notable as left-wing politicians are most often targeted by negative mentions.

@sybrandbuma (CDA)



@thierrybaudet (Forum voor Democratie)



MENTIONS



Each dot represents one mention

MENTION TYPE



Attack

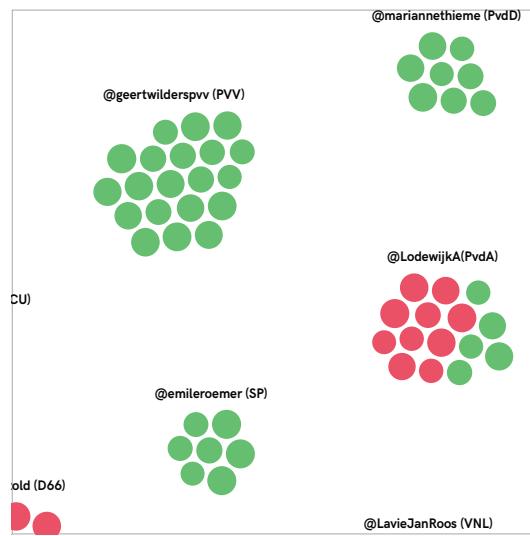


Not Attack

@JacquesMonasch (Nieuwe Wegen)







SERVING SUGGESTIONS

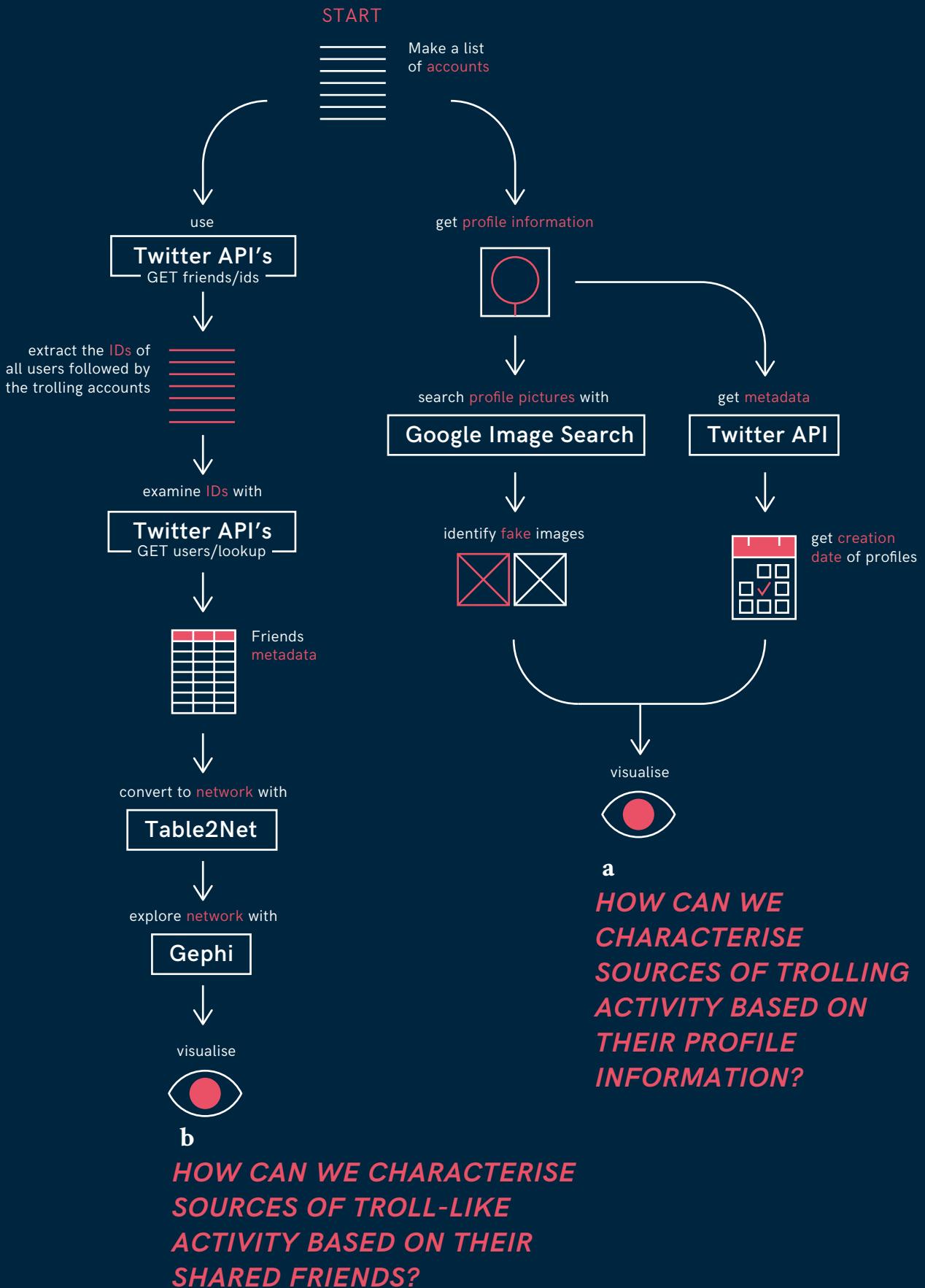
This recipe can be used to identify sources of personal attacks on Twitter and can be extended beyond the context of political trolling.

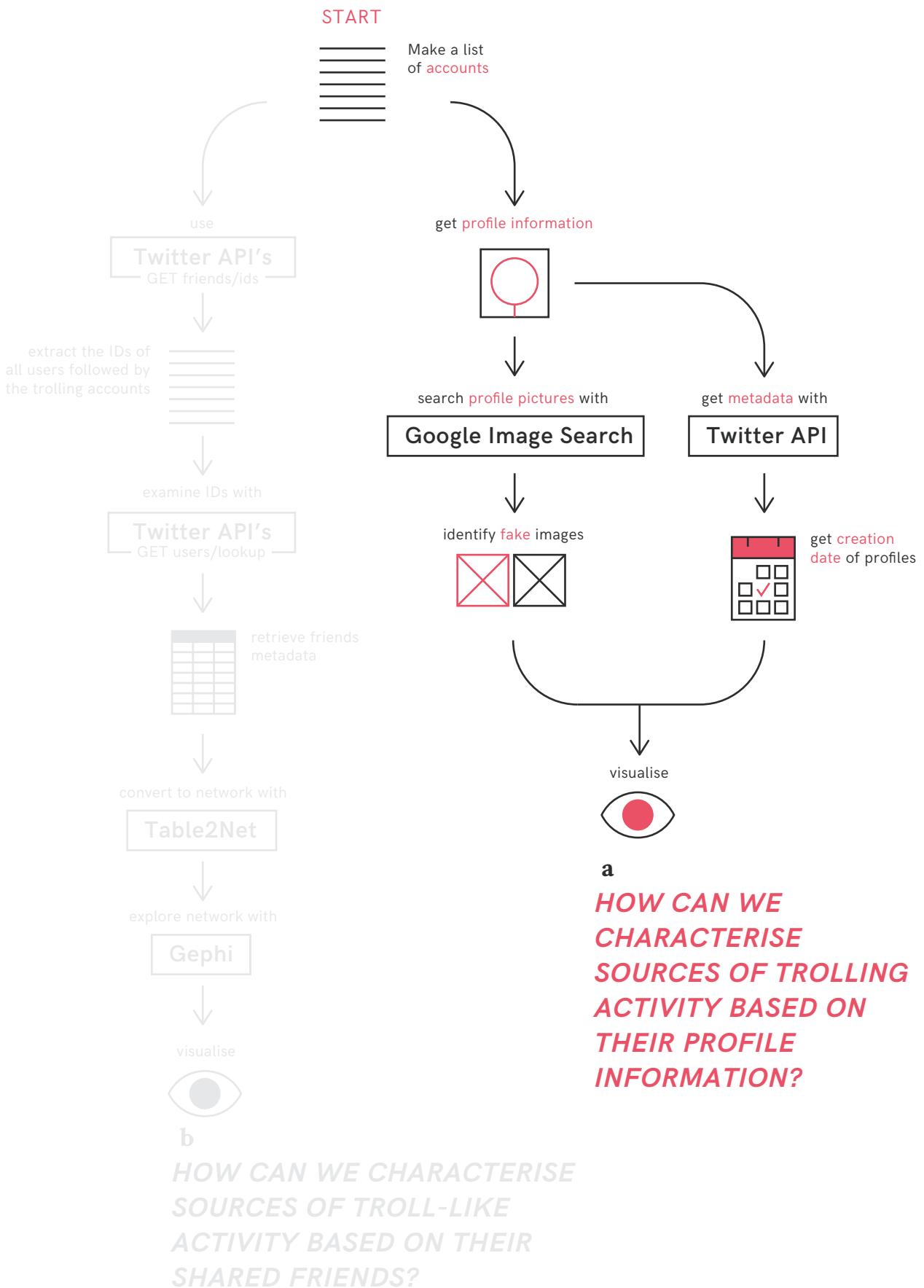


HOW MAY WE CHARACTERISE THE SOURCES OF TROLL-LIKE ACTIVITY?

BEFORE STARTING

For this recipe we take as a starting point the 25 accounts that mention at least one political leader at least 100 times identified in the previous recipe (we discarded one account because it was no longer active).





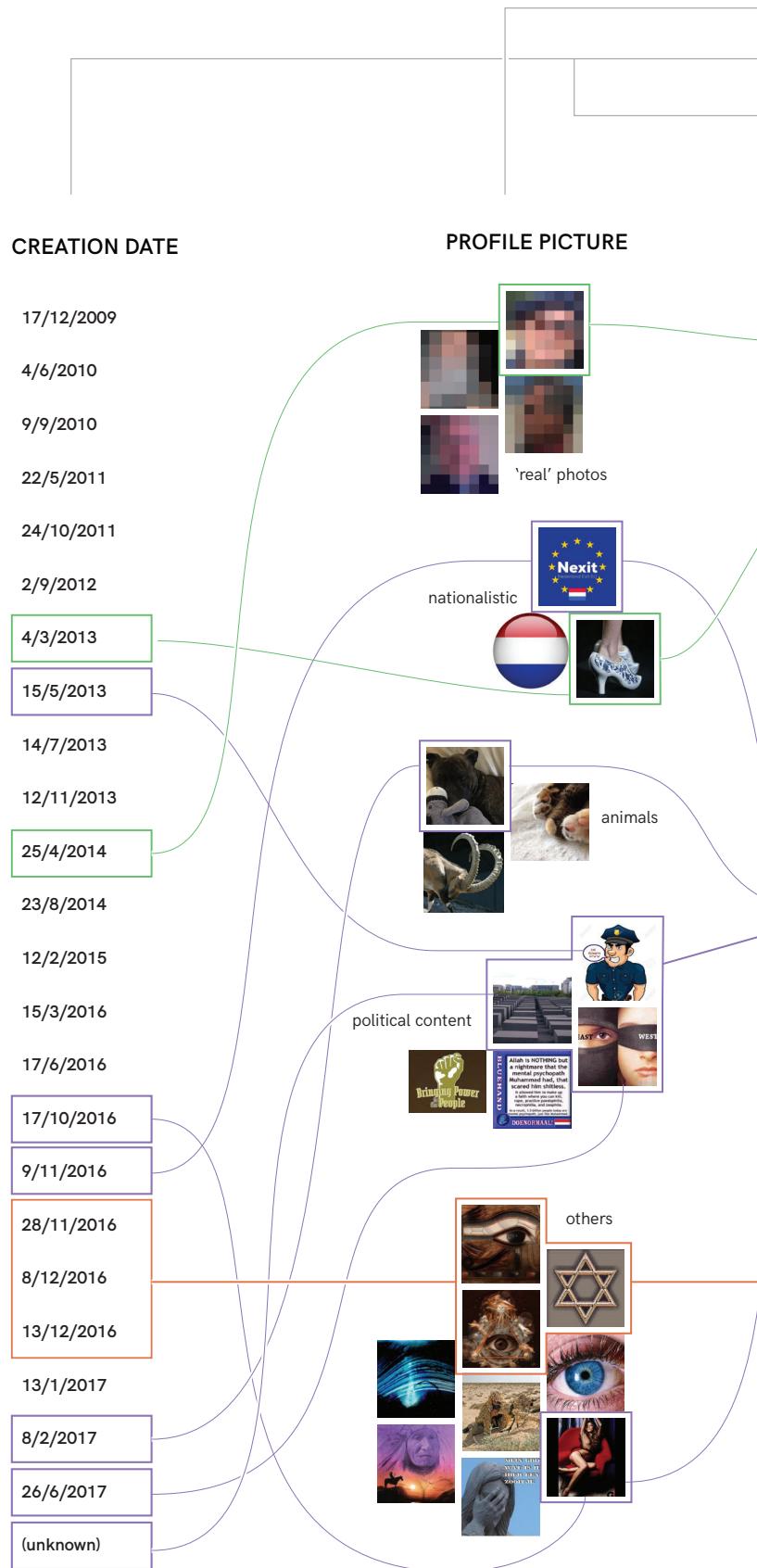
a. CHARACTERISE SOURCES OF TROLL-LIKE ACTIVITY THROUGH THEIR PROFILE INFORMATION

One way in which you can characterise the sources of troll-like activity is by examining their profile information.

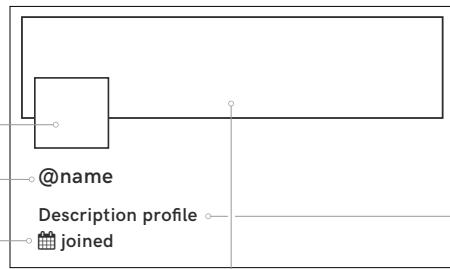
- ◊ Visit each of the accounts and collect their profile information from the Twitter interface (description, profile picture and banner).
- ◊ Analyse this information in order to identify political issues, hashtags mentioned and affiliations.
- ◊ Take note of the presence or absence of profile images and upload any images identified to  Google Image Search to detect whether any of the accounts use fake profile images.
- ◊ Using the Twitter → API, extract the creation date of the account and examine whether several accounts in your corpus have been created around the same date.

HOW CAN WE CHARACTERISE SOURCES OF TROLLING ACTIVITY BASED ON THEIR PROFILE INFORMATION?

Clustering of 24 accounts engaging in troll-like activity around the Dutch elections. The profile information is clustered according to similarities. Three users have very similar profiles and are created in a short amount of time: this helps us to identify them as 'sock-puppet' account created for trolling activities. Other six promote the same anti-islam agenda, but without being fake accounts.



ELEMENTS FOR CHARACTERIZATION ANALYSIS



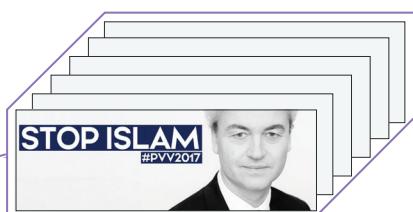
BANNER



BIO DESCRIPTION

TROLL
OR NOT TROLL?

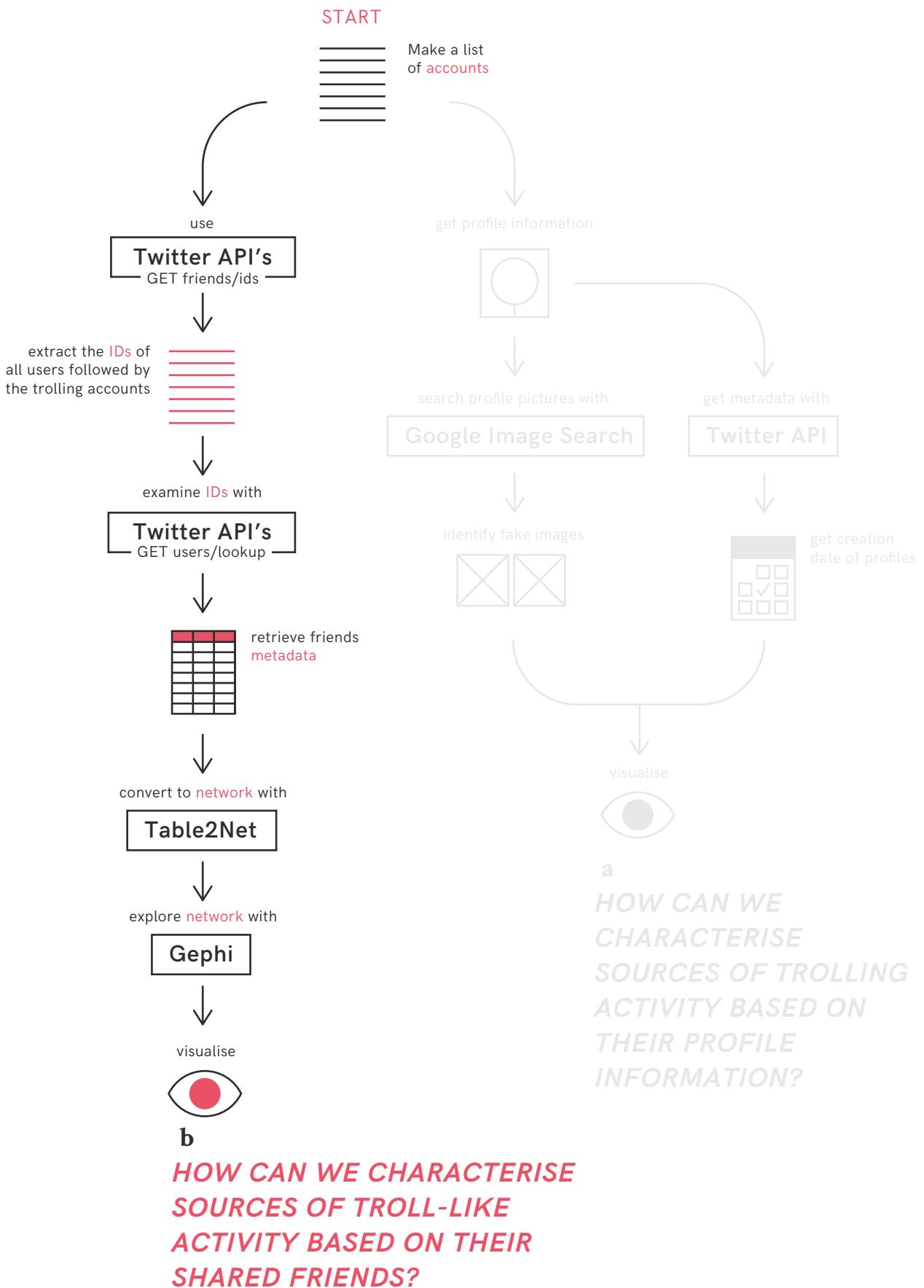
- @D***er
BOER, Economist, Father of six, Doctor/MBA Patriotically correct. Romantically conservative, islamophobe, aspirant lid PVV, lid FvD ?
- @M***te
(no bio description) X



- @H***an
(no bio description) ?
- @l***dy
#de-islamize #stoppolicor #stopterror#voteMarine #voteWilders#minderminderminder #lesslessless#maroccons ?
- @M***00
Ben fascist noch racist, maar realist. Links is het nieuwe rechts en blank het nieuwe zwart. Stop Brussels EU omvolking. Stop massa migratie, Stop Rutte & Timmermans ?
- @A***rt
Een kern van (mijn) waarheid overgotten met sarcasme, sadisme en galgenhumor. Neem niets persoonlijk op, maar neem het wel op! ?
- @g***91
Twitter Censors Everything, their Board members are Islamic. MOVE to Gab.ai for true FREE SPEECH FOR ALL <http://gab.ai/ger2519> #PVV #MEGA #MHGA #BanIslam ?
- @Y***NL
(account suspended) ?



- @j***33
PVV Geert Wilders Nexit, Trump Trooper MAGA, Love Le Pen Vive la France MEGA, Anti Islam, De-Islamization Forever ✓
- @j***33
Geert Wilders PVV MEGA Nexit De-Islamization worldwide, Love Le Pen Superwoman, Trump MAGA, Close borders, Anti islam, Anti EU, Jail Obama, Jail Hillary... ✓
- @k***33
Geert Wilders Nexit, Vive Le Pen, Trump MAGA, AfD Germany, Brexit, Pro Israël, ☽□□□□□□□□□□, Anti EU, De-Islamization... ✓



b. CHARACTERISE SOURCES OF TROLLING ACTIVITY THROUGH THEIR FRIENDS

Another way in which sources of troll-like activity may be characterised is by examining who they follow on Twitter, also known as their friends.

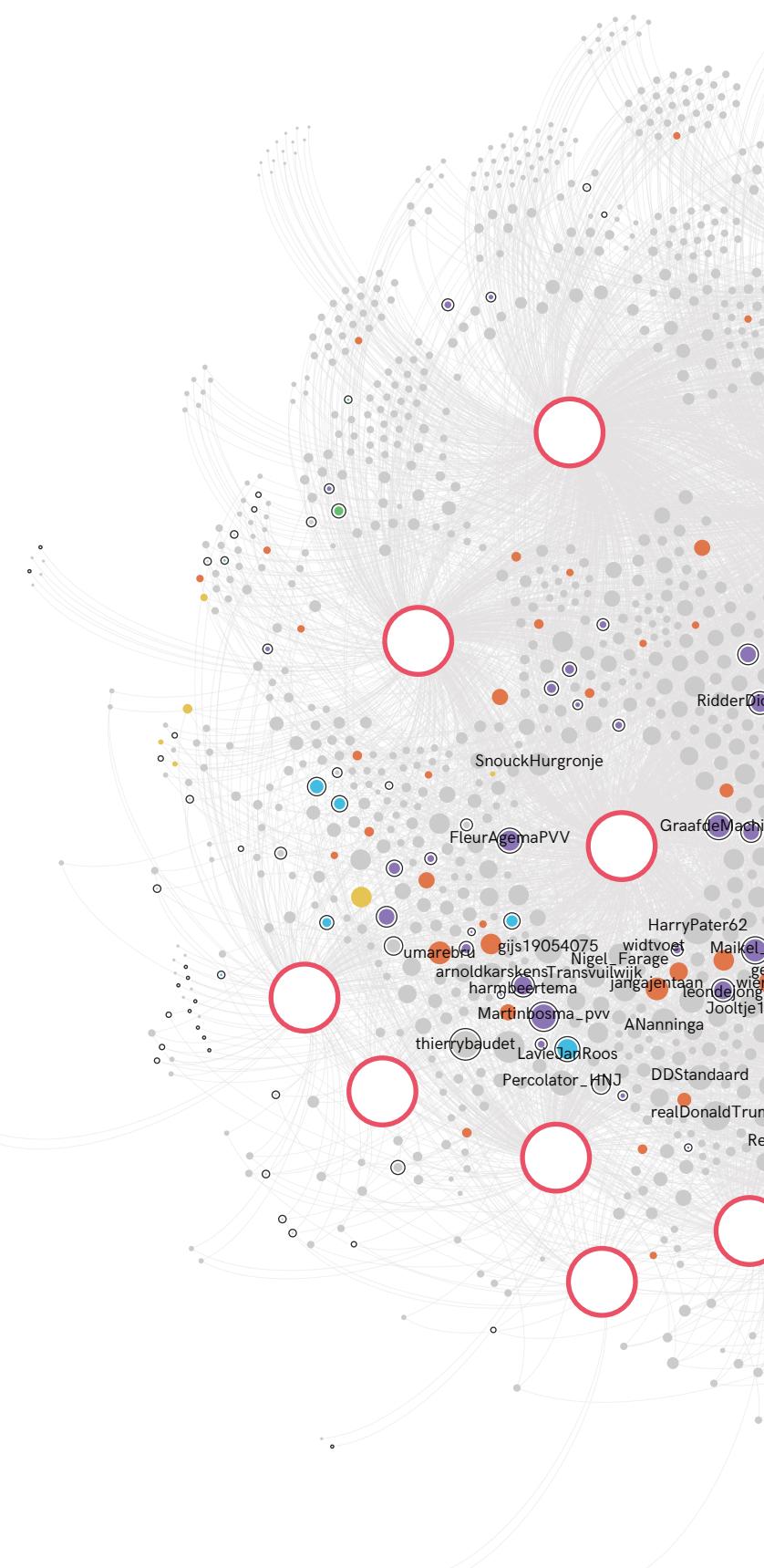
- ◊ Use the "GET friends/ids" function of the Twitter API to extract the IDs of all users followed by the trolling-accounts you have identified.^[1]
- ◊ Use the "GET users/lookup" function of the Twitter API to retrieve the profile information for each friend.^[2]
- ◊ Use  **Table2Net** to convert the table containing the studied accounts and their friends in a network of Twitter accounts connected by the “friendship” relationships between them.
- ◊ You may use a network analysis and visualisation tool such as  **Gephi** to explore the shared friends or followees between the accounts engaging in troll-like practices.

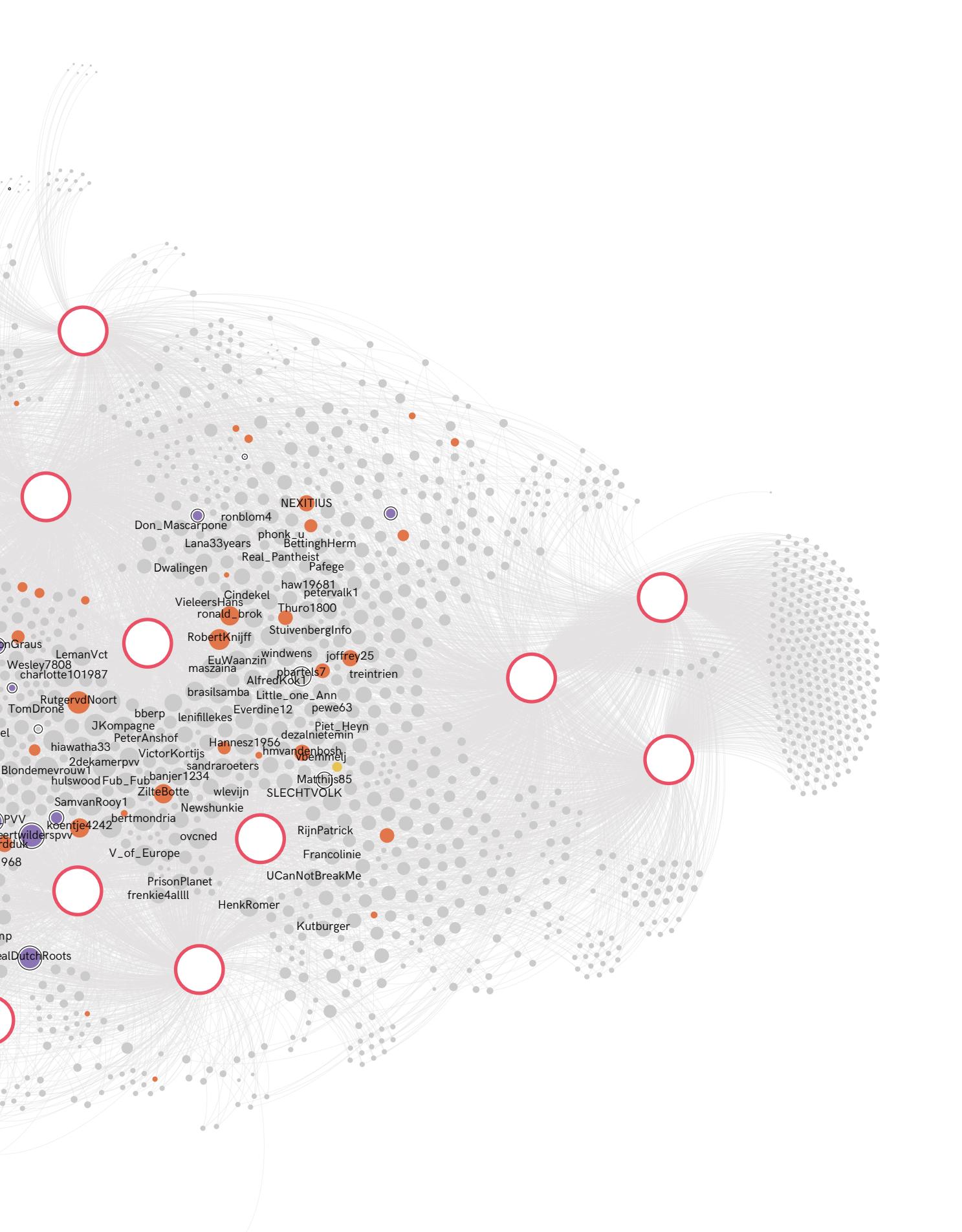
[1] See *Twitter Developer Platform*, API reference: GET friends/ids, 2017, at: <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-friends-ids>

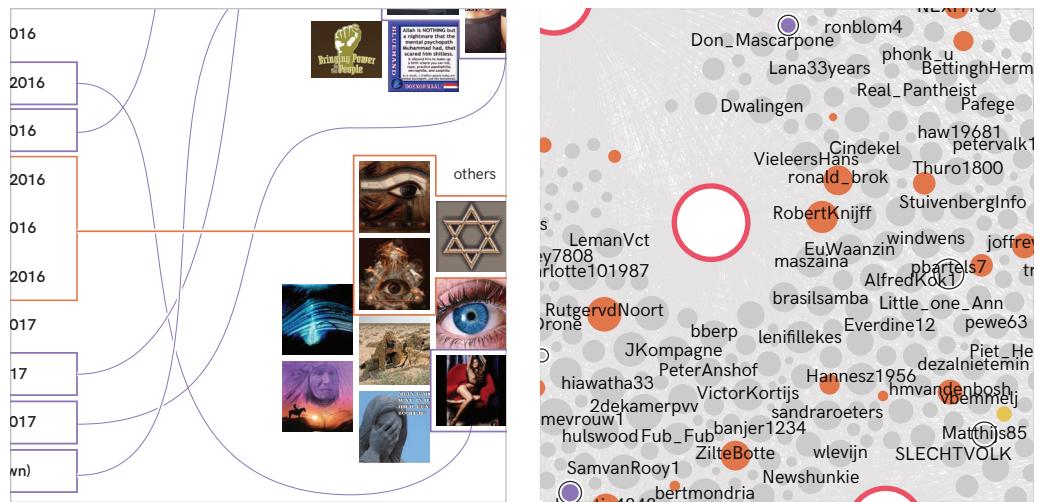
[2] See *Twitter Developer Platform*, API reference: GET users/lookup, 2017, at: <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-users-lookup>

HOW CAN WE CHARACTERISE SOURCES OF TROLL- LIKE ACTIVITY BASED ON THEIR SHARED FRIENDS?

Visualisation of shared friends or followees of 24 accounts engaging in trolling activity around the Dutch elections. The density of connections in the network shows that accounts share multiple followees. Multiple accounts which may be described as right-leaning based on their profile information (description, picture and banner) are present at the core of the network thus confirming earlier findings pertaining to the right/left asymmetry of sources and targets of attacks.







SERVING SUGGESTIONS

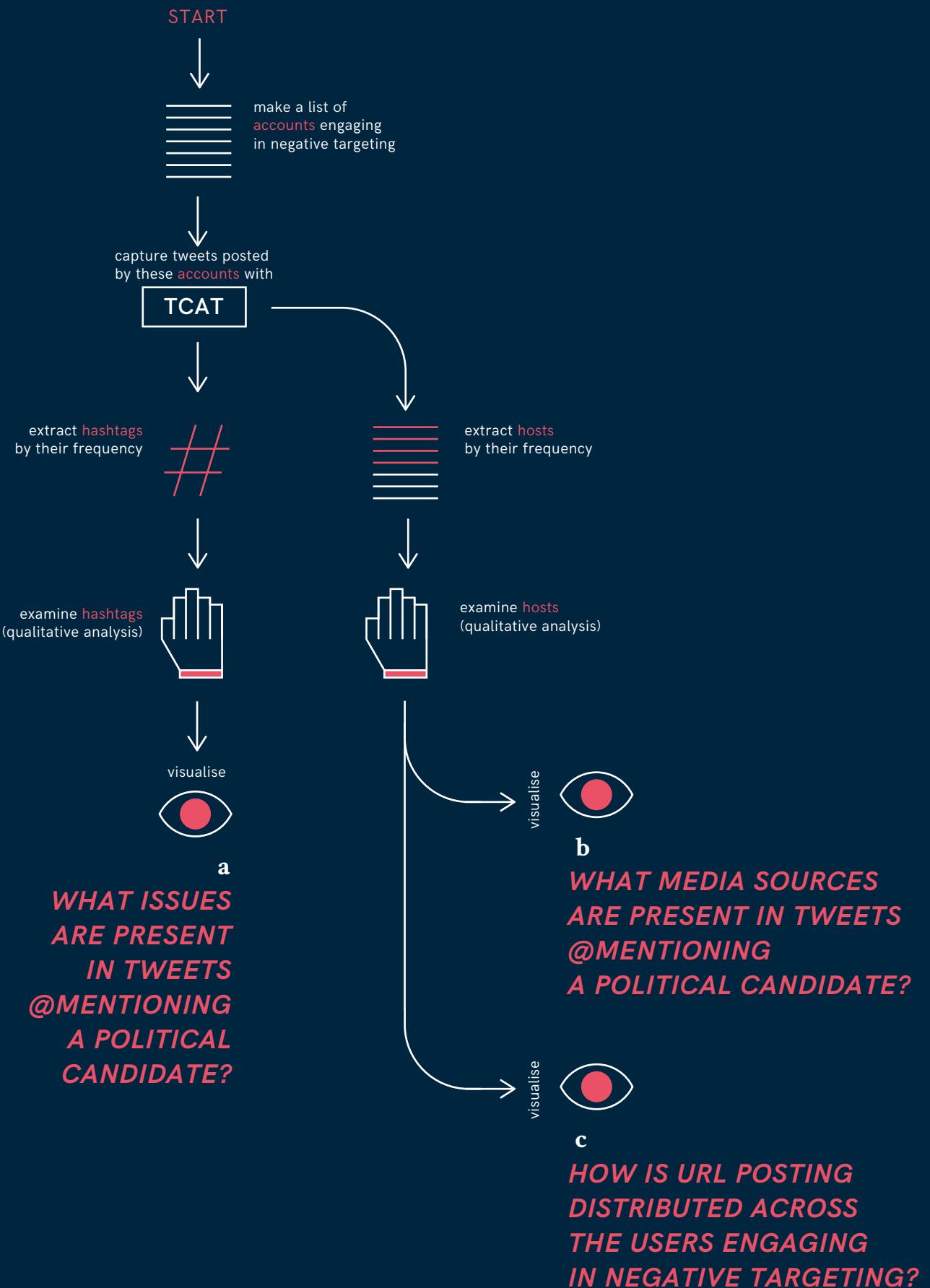
This recipe may be used to profile trolling accounts in the context of other political campaigns.

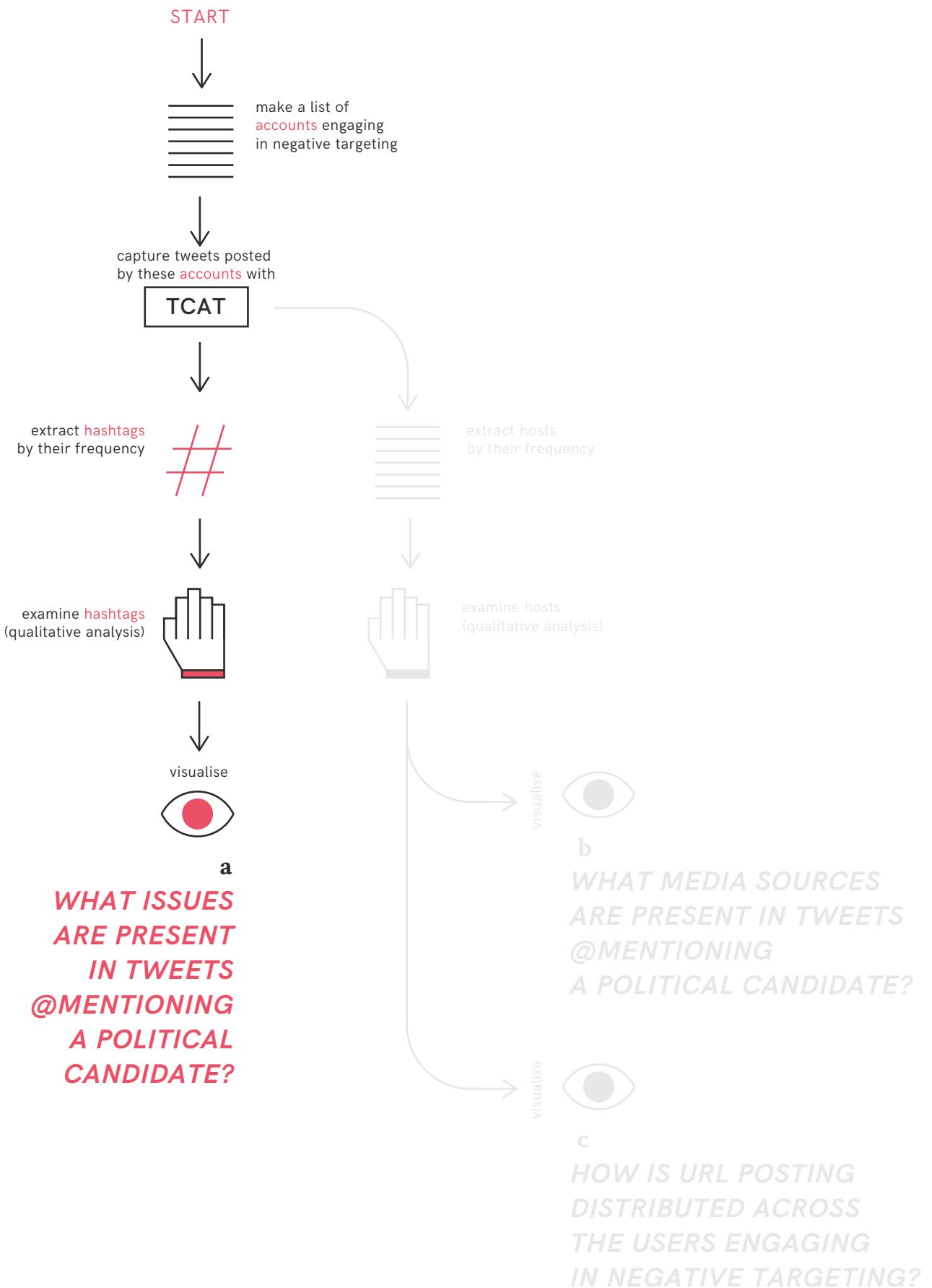


HOW MAY TROLL-LIKE PRACTICES BE CHARACTERISED?

BEFORE STARTING

For this recipe, take as a starting point the accounts identified in the previous recipes and the tweets posted from these accounts in the timeframe of the study (see recipe 5.1).





INVESTIGATE THE HASHTAGS THAT SOURCES OF TROLL-LIKE ACTIVITY USE

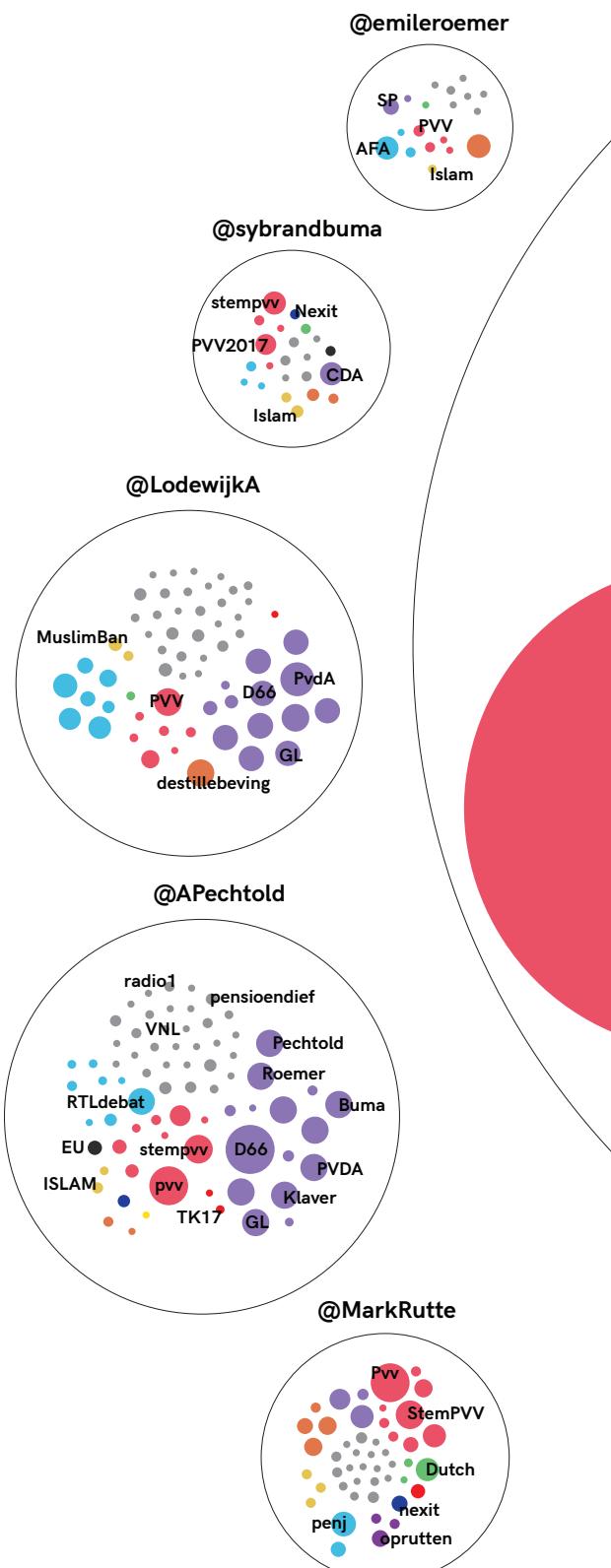
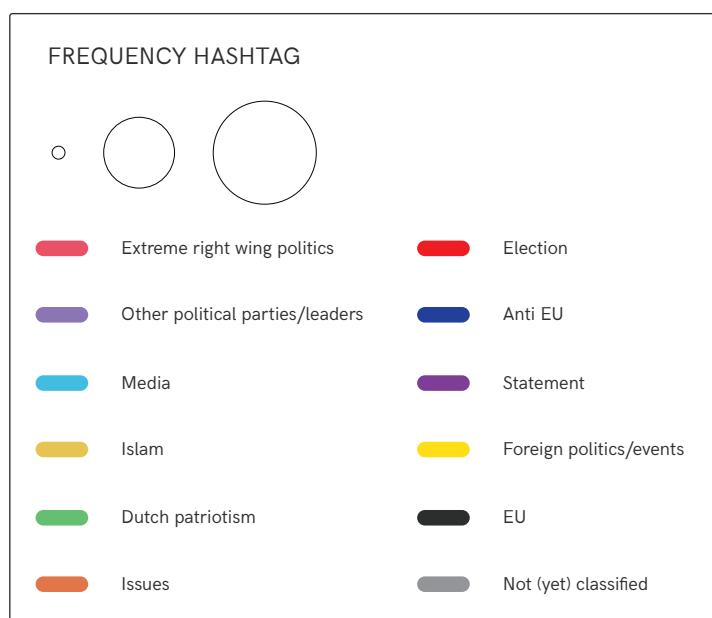
To identify what issues are associated with troll-like activity, examine the hashtags that are used in tweets that mention a politician posted by the users that frequently engage in negative targeting.

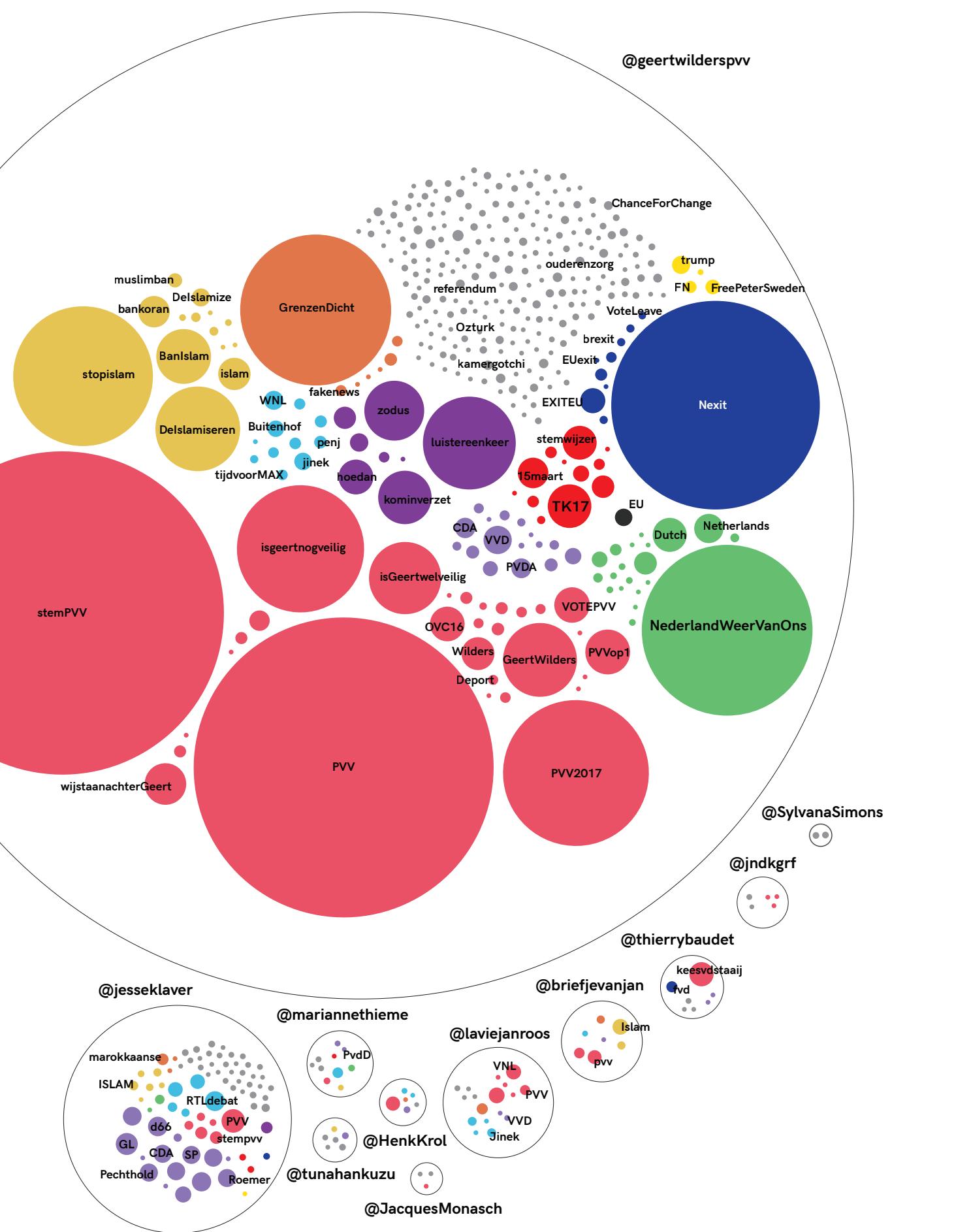
- ◊ Rank hashtags by their frequency using the “hashtag frequency” feature of  TCAT.
- ◊ Manually analyse the most frequently used hashtags to identify issues that animate activities of users engaged in trolling practices.

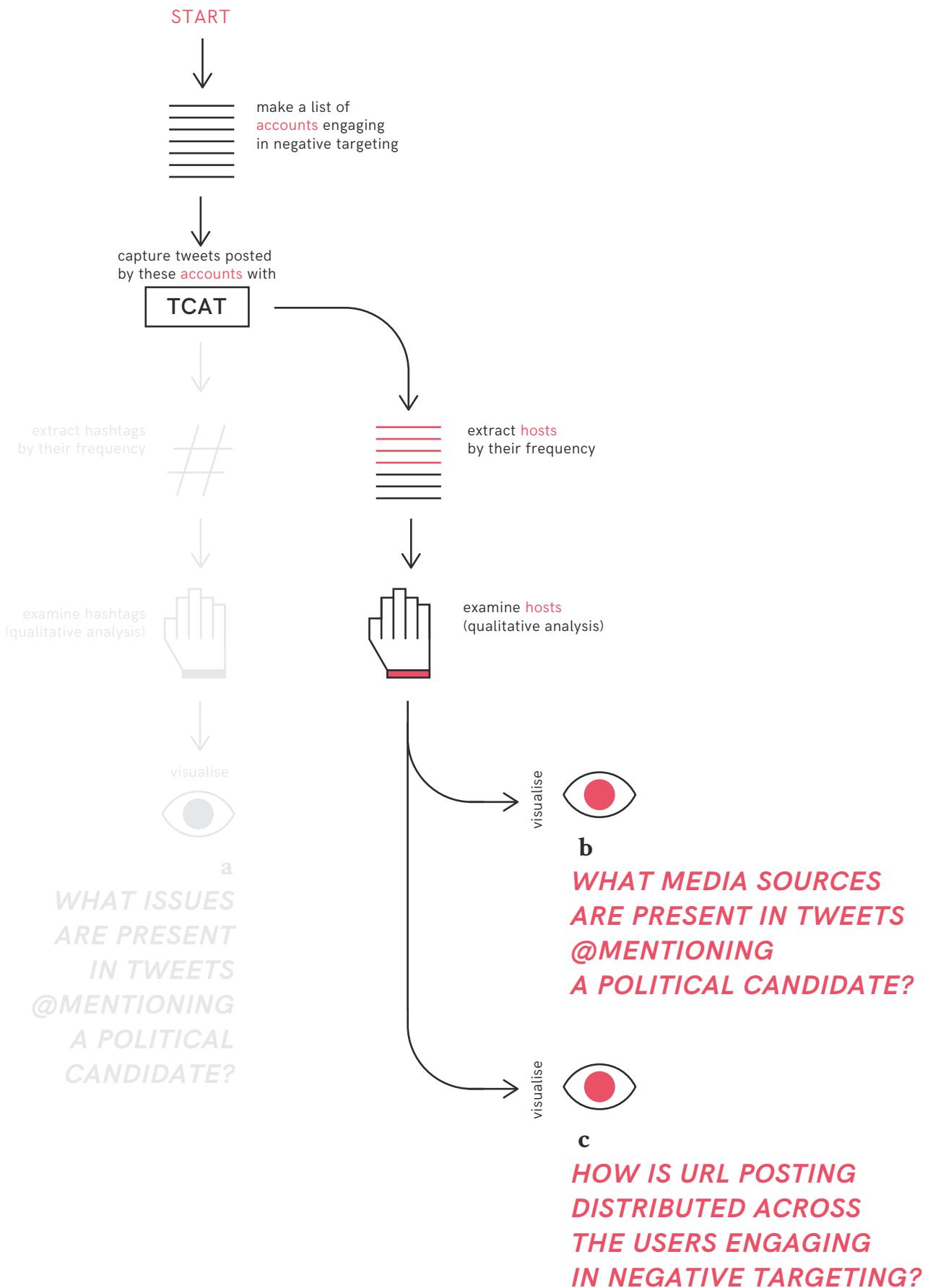
WHAT ISSUES ARE PRESENT IN TWEETS @MENTIONING A POLITICAL CANDIDATE?

→ Bubble graph of issues expressed through hashtags in tweets @mentioning candidates in the 2017 Dutch elections posted by the set of 24 accounts

engaging in troll-like activity. Issues are coloured by type, sized by frequency of occurrences and grouped according to the candidate mentioned in the tweet which contains them. Most tweets with hashtags mention the right-wing populist candidate Geert Wilders. Most prominent are issues related to PVV's political message ("Nexit", "StopIslam" and "BanIslam") as well as those pertaining to expressions of Dutch patriotism. Generally speaking, we can conclude that right-wing politicians receive mainly support from "troll-like users," while other politicians are the targets of attacks (as discussed in recipe 5.1).







INVESTIGATE THE MEDIA SOURCES SHARED BY THE ACCOUNTS ENGAGED IN NEGATIVE TARGETING

To identify the content shared in tweets posted by the users engaged in troll-like activities, examine the URLs inserted in their tweets.

- ◊ Use the “hosts frequency” feature of  TCAT to extract the media sources ranked by frequency of occurrence.
- ◊ Manually analyse the most frequently used media sources to determine their profile.
- ◊ Analysis of URL sharing behaviour across the set of users may be used as a means to detect troll-like activity.

WHAT MEDIA SOURCES ARE PRESENT IN TWEETS @MENTIONING A POLITICAL CANDIDATE?

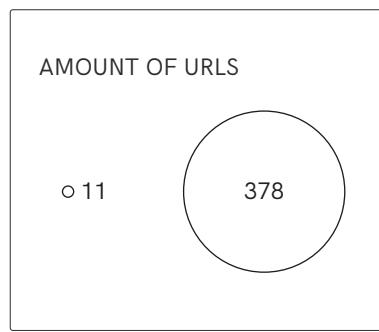
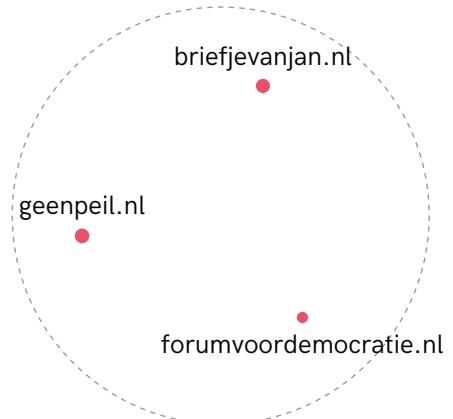
Venn diagram of most resonant media sources in tweets @mentioning candidates in the 2017 Dutch elections posted by the set of 24 accounts engaging in trolling activity. The most tweeted source is the Dutch alt-right blog *fenixx.org* followed by the anti-islam site *Jihad Watch* and the right-wing think tank *Gatestone Institute*.

OTHER MEDIA



libelletv.nl

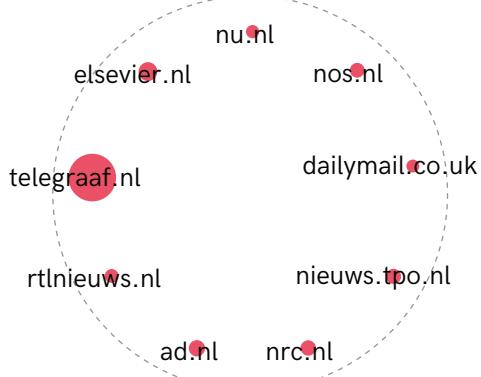
POLITICAL PARTY



PUBLIC BROADCASTER



NEWS



ANTI-ISLAM

PRO-ISRAEL

worldisraelnews.com

dagelijksestandaard.nl

kudtkoekiewet.nl

tpo.nl

fenixx.org

joostniemoller.nl

jihadwatch.org

clarionproject.org

gatestoneinstitute.org

unitedwithisrael.org

ejbron.wordpress.com

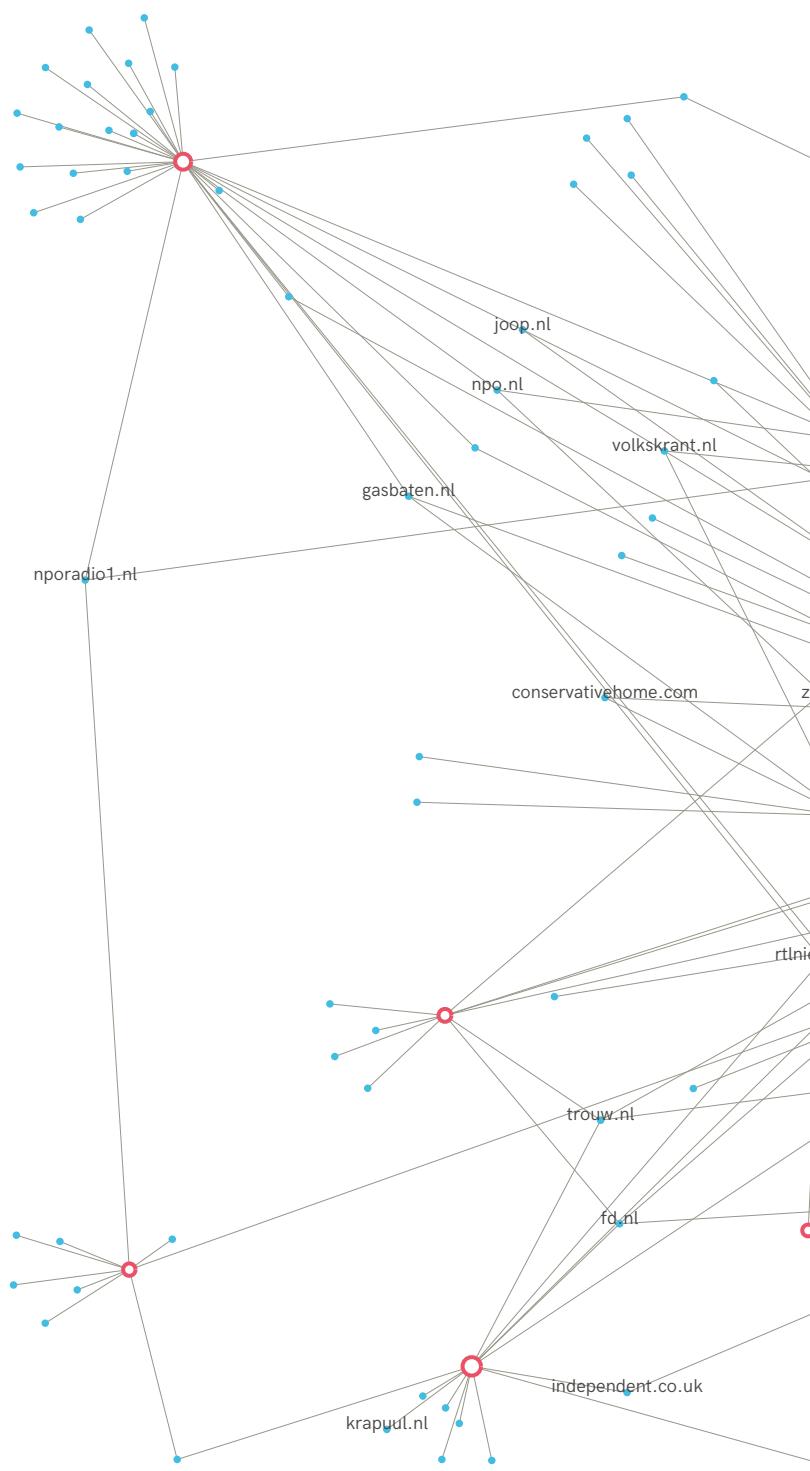
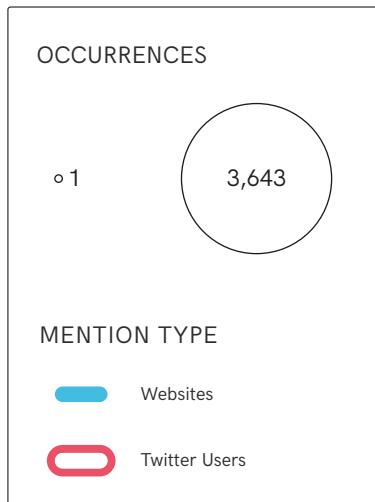
ALT-RIGHT

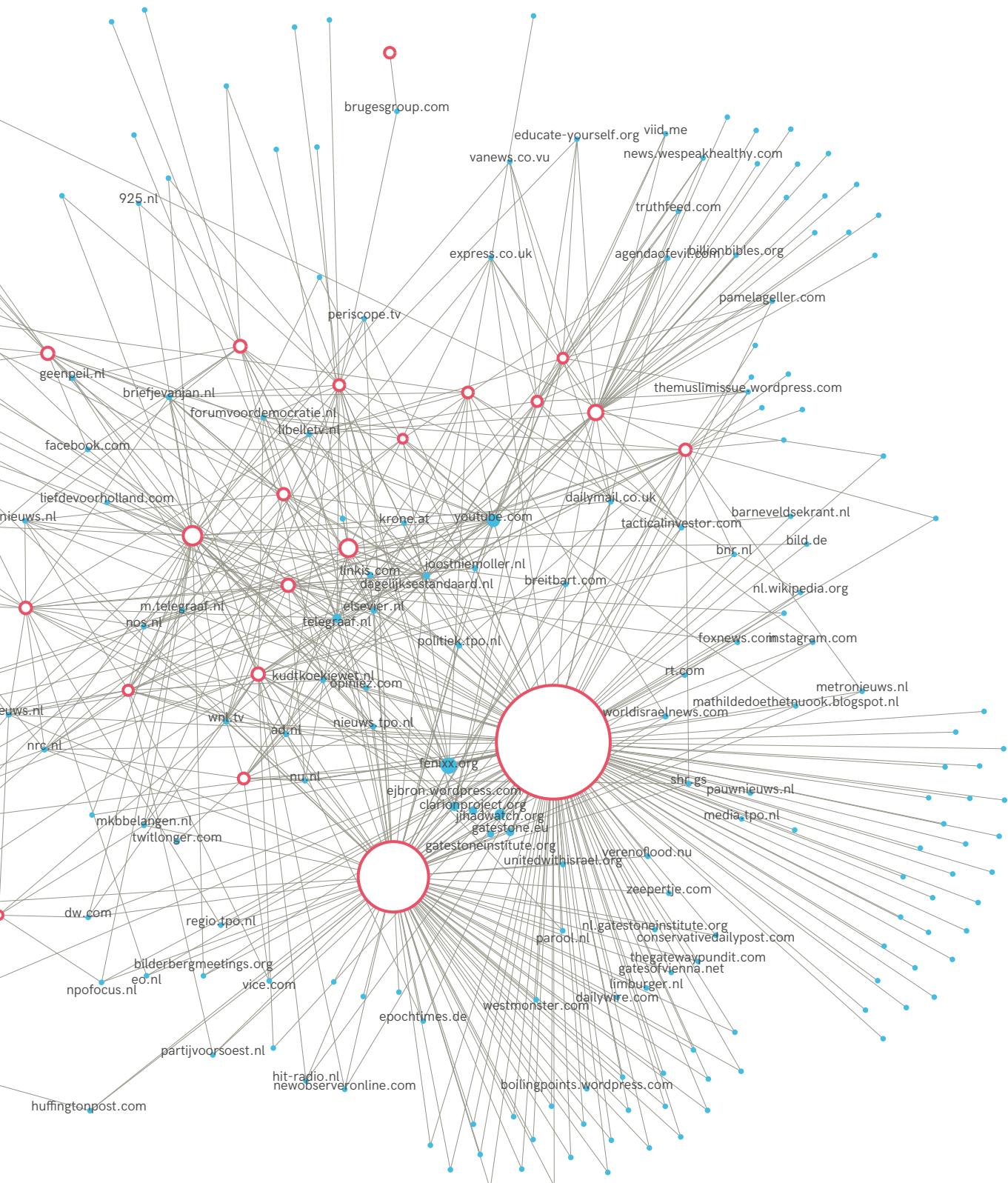


DUTCH-RIGHT

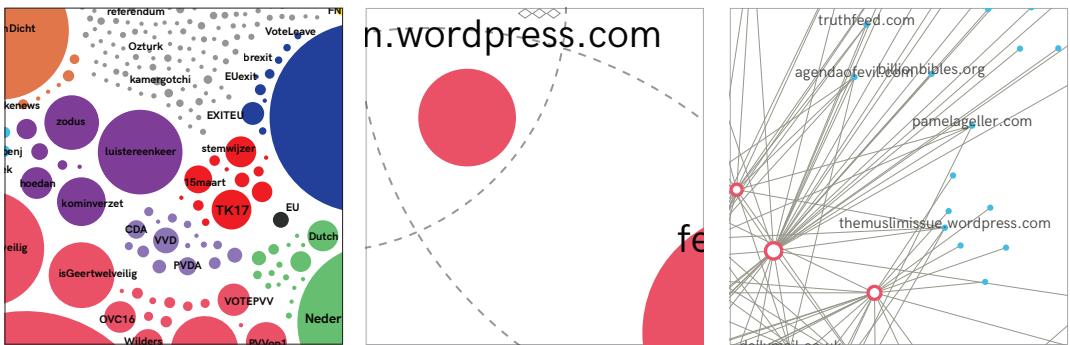
HOW IS URL POSTING DISTRIBUTED ACROSS THE USERS ENGAGING IN NEGATIVE TARGETING?

→ Network graph of distribution of URLs shared across the 24 users engaging in negative targeting of politicians in the month before the Dutch elections. The graph shows two users to be responsible for the majority of URLs posted in the studied timeframe.





CHAPTER 5 → RECIPE 3



SERVING SUGGESTIONS

This recipe can be used to profile the issues and media sources associated with social media accounts engaging in troll-like behaviour – building up a richer picture of these activities and their context. If you are exploring personal accounts you should make sure to consider the potential legal and ethical implications of your inquiry, and ensure that public-interest arguments are weighed against privacy considerations. You may consider whether to focus on networks and relationships between a number of accounts, and whether to remove or redact personally identifying information.

Conclusion

Glossaries

Contributors and
acknowledgements

CONCLUSION

Fake news, it can safely be said, is not a neglected issue. Every day it seems as though a new newspaper article, blog post, research report or project is released on the subject, and more and more academic articles are produced to reconnect public debates to scholarly literature. Over the course of this project we have been in touch with media organisations, journalists, civil society groups, public institutions, companies, researchers and students from around the world eager to understand, investigate, address and study to this issue

Amidst this intense public debate and mediatisation, concerns have been raised about the term “fake news”, and suggestions have been made to retire it. As we mention in the introduction, amongst other things, fake news has been said to be vague, politically dangerous (as it is appropriated as a tactical term by various parties), and indistinguishable from previous forms of propaganda, disinformation and misinformation. While in this guide we have not abandoned the term we have sought to address these legitimate concerns in a different way.

Over the course of the pages above, readers will not have failed to notice, we move away from a focus on defining

and identifying fake news based on its content. While such interest is certainly justified, we believe that attempts to classify and demarcate the terrain of associated phenomena should be grounded in empirical investigation of not just the *features* but also of the *social lives* of a variety of cases. We hope that such work will contribute to the development a more granular analytical, conceptual and theoretical vocabulary to describe the constellation of phenomena associated with the term.

What is to be done about fake news? As we mention in the introduction, if there is one single thing that we hope to achieve with this *Field Guide*, it is to broaden the emphasis of research, journalism and public debate to include a more substantive focus on the social lives of news and the digital *environments* in which they move. We hope that this work makes at least some modest contribution to the rather grander task of inspiring, mobilising and assembling publics who are capable of not only of studying and *interpreting* these environments but also *changing* them.

A year after we started work on this project around the time of the 2016 US presidential elections, the issue of fake news has “gone global”. The cast of characters has multiplied from an initial narrative focusing on grassroots hyper partisan propagandists, opportunistic Macedonian teenagers and Russian political operators targeting the United States, to include actors as diverse as Google and Facebook, the European Commission, the Chinese Communist Party, the Italian Five Star movement, UK’s intelligence agency MI6, Wikipedia, Web Inventor Tim Berners-Lee, election bots, messaging apps, nuclear threats, tech startups, security firms and “dark” money in numerous countries around the world.

And the question “what is to be done about fake news?” has broadened out into a series of questions not only about online misleading information, but also about online platforms and the broader digital cultures, practices and technologies associated with them. What started as a matter

of identifying and “weeding out” offending articles and deviant users has unfolded into a much bigger series of questions and debates about the organisation of public life online, and the attendant infrastructures and institutions through which information, knowledge and culture is created, vetted, shared, used and made meaningful.

As with all controversies, there are different ways of diagnosing, defining and scoping the problem, as well as different solutions and conceptions of how responsibility (and blame) should be apportioned to corporations, markets, states, politicians, policy-makers, media organisations, educational institutions, civil society groups and others. And as with many crises, there are a range of actors lined up with different agendas but sharing the same sentiment of “not letting a good crisis go to waste”.

While there are many pressures for a quick response to the issue of fake news, we hope the approaches that we explore in this guide encourage readers not to be *over-hasty* in appraising the situation, diagnosing the problem and in proposing fixes. Through the series of recipes in this guide we hope to provide some pointers about how to spend time with the phenomenon. In particular we hope the guide will inform and support research, investigations and public debate around one aspect of this broader set of concerns: the mediating capacities and cultures of online platforms and the web. While the set of recipes that we have provided focuses on following fake news and other fabrications online, many of them can easily be repurposed to examine many other aspects of knowledge politics, issues and controversies, and the online spaces and digital infrastructures upon which they play out.

In repurposing digital traces to study knowledge politics, we also advocate a shift from the examination and evaluation of claims *in themselves* and *in isolation*, to looking more closely at the various *networks in which they are embedded*. We thus propose a shift from the *atomistic* study of fake news artefacts

(apart from their contexts of circulation), to looking at their networked and distributed character, the social and cultural practices of meaning-making that emerge around them, as well as the media systems which underpin their circulation. In other words, we urge investigators to consider items which are classified as fake news not only in terms of truth, falsity, and the extent to which they accurately depict states of affairs in the world, but also in terms of how they are shared, amongst whom, what they depend on, and the many varieties of value and significance that may be attributed to them by different publics.

Why might we want to make such a shift? Firstly, a richer picture of social and cultural processes of making meaning around digital content might help to open up different kinds of questions. Is a particular group sharing something because it considers it is literally *true*, or because they think it is funny, germane, ridiculous, intolerable or resonant with other beliefs and backgrounds? Stemming the flow of a particular piece of content may have negligible (or even counterproductive) effects in addressing the beliefs, practices and concerns of groups which share it. Fact-checking corrections risk falling on fallow ground if they miss the point or punchline, which requires knowledge of the background against which their claims becomes poignant, salient or amusing. This is not to suggest that we should flatten the difference between subjective salience and objective accuracy, but that both depend on a shared background of social institutions and cultural practices which should not be taken for granted.

Secondly, a shift from atomistic to networked investigations of fake news may enable us to learn more about the specific ways in which social institutions and culture practices are enabled, constrained and organised through digital platforms and infrastructures. While we adopt the metaphor of the field guide from natural histories, the online spaces that we study should not be understood as natural *ecosystems*, but rather as manufactured *landscapes* where social and cultural

[1] Here we are inspired by work on “technological landscapes” in the study of science and technology, see, for example, Richard Rogers, *Technological Landscapes*, London: Royal College of Art, 1999.

life unfolds in tandem with specific technological devices and algorithms.^[1] At the same time the web and online platforms cannot determine how they are used, and so we must look “across” them to understand not only their techno-political “shape” but also how online life unfolds around them. In studying the social life of fake news and other fabrications we can explore *both* the capacities of online platforms and infrastructures and the social practices of their users.

[2] Sabine Niederer, *Networked Content Analysis: The Case of Climate Change*, unpublished manuscript, 2016.

Through this *Field Guide* we have sought to make clear how the issue of fake news may foreground central aspects of our digital environments and thus provides a good opportunity to study their dynamics. And that these dynamics can and should be empirically investigated. To illustrate different aspects of these dynamics we attended to the networked character of fake news and to its technicity, that is the way in which fake news is formatted, ordered, metrified, datafied and thus co-produced with digital platforms.^[2] Thus, chapters 1 and 2 discuss the publics and modes of circulation afforded by these platforms. Chapter 3 investigates the tracking networks in which online content is embedded and through which its readers are rendered into data. Chapter 4 analyses the media artefacts that circulate well online, namely image-based memes, and chapter 5 explores how platform features may be mobilised in the service of attacks directed at political representatives.

However, while empirical approaches to studying the social life of fake news and other fabrications online are necessary, they also brings a number of challenges. While we emphasize the need for studying how fake news circulates, the current configurations of digital platforms, for good (and less good) reasons, do not always allow this. As a consequence, all the recipes described in this book are meant to study the *public* circulation of fake news. Our study of Facebook provides a perfect illustration of such challenges. The API of the social network allows scholars to retrieve the contents of public pages, but prevents them from accessing the information exchanged through personal accounts (although some of

this information may be accessed via public pages). Online platforms have their own ways of organising the boundaries between public and private. And in the case of fake news, we also have to deal with the consequences of technological fixes to the phenomenon on possibilities to study it. In the case of Facebook, measures to remove problematic posts from the platform mean that researchers are unable to examine how users engaged with these items. How platforms, regulators, policy-makers, users and others negotiate these unfolding questions of the configuration of these emerging spaces of publicity and privacy, their attendant mechanisms of accountability, remains to be seen.

Beyond questions about how digital landscapes are studied, organised and reshaped, we hope this guide may also serve as inspiration for how digital methods may be used to study and intervene around data politics in the contemporary moment. Who will have the capacities to shape how data is created and used? How can data be used to not only to close debate but to enrich it? How can different kinds of data help us to pursue objectivity not just through a single picture, but through a plurality of different perspectives? How does the configuration of digital infrastructures shape what is hearable, sayable, seeable and doable with data? Who and what will stand to benefit from the data society? *The Field Guide to Fake News* is the first of an ongoing series of activities and experiments with the Public Data Lab through which we hope to continue to explore these themes.

*Jonathan Gray (@jwyg), Liliana Bounegru (@bb_liliana),
Tommaso Venturini (@TommasoVenturin)
Paris, November 2017*

TOOLS GLOSSARY

In this section we provide brief descriptions and links to various tools that are referenced throughout the field guide. These descriptions are intended to be sufficiently informative to enable readers to follow the text. It should be noted that it is very important in any research project or investigation to develop an appreciation of precisely how they work and what they do (and what they do not do). Hence we advise you to refer to the documentation and more detailed descriptions on the websites listed below before using them in your own project.

- ❖ **BuzzSumo:** a social analytics service which enables users to explore the most “engaged” content relative to a given topic or domain. You can filter the results by language, country, word count and content type (article, infographic, interviews, videos). (<http://buzzsumo.com/>)
- ❖ **CorText:** an online application used to analyse textual data. It allows users to create various types of statistical and network visualisations. (<http://www.cortext.net/>)
- ❖ **CrowdTangle:** digital tool that allows users to track how content spreads through the web and follow the performance of posts and accounts on Facebook, Twitter, YouTube, Instagram and Vine.
(<http://www.crowdtangle.com>)
- ❖ **CSV Rinse Repeat:** a JavaScript based tool to clean and

structure a csv files, including filtering, clustering, parsing, merging and matching regular expressions.
(<http://tools.medialab.sciences-po.fr/csv-rinse-repeat/>)

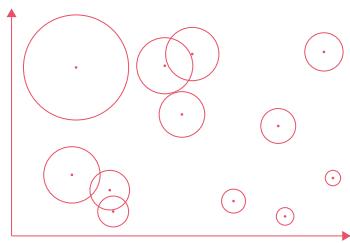
- ☞ **DMI Tracker Tracker:** a web-based tool which uses data from the Ghostery project to detect a set of over 900 “fingerprints” of analytics tools, widgets, social plugins, and other trackers in a given set of URLs.
(<https://tools.digitalmethods.net/beta/trackerTracker/>)
- ☞ **DMI Triangulation Tool:** identifies common items in two or more lists.
(<https://tools.digitalmethods.net/beta/triangulate/>)
- ☞ **DownThemAll!:** a Mozilla Firefox extension that allows you to collect all the links and images contained in a web page. (<https://addons.mozilla.org/en-US/firefox/addon/downthemall/>)
- ☞ **Gephi:** network analysis and visualization software. Gephi is particularly helpful for finding patterns, trends and highlights in large datasets. (<http://gephi.org>)
- ☞ **Google Image Search:** a search service provided by Google, which allows users to retrieve images related to a keyword or a query. (<http://images.google.com>)
- ☞ **Google News Search:** a news aggregator from Google which provides results on news articles, sorting them by date and time of publication. (<https://news.google.com/>)
- ☞ **Google's Vision API:** an image analysis tool which allows you to categorise pictures, detect objects or individual faces, as well as to extract textual content.
(<https://cloud.google.com/vision/>)
- ☞ **Google Web Search:** a search engine which provides results based on “Page Rank” (see concept dictionary).

- ☞ **Graph Recipes:** an online Javascript tool that allows you to generate static images and compute statistical metrics about networks. A number of default scripts are offered by the tool, but others can be added by the user.
[\(http://tools.medialab.sciences-po.fr/graph-recipes\)](http://tools.medialab.sciences-po.fr/graph-recipes)
- ☞ **Hyphe:** a semi-automated web crawler allowing users to identify and follow the hyperlinks on a series of webpages, to define and categorize a corpus of websites and to generate networks of web-entities and their connections. (<http://hyphe.medialab.sciences-po.fr/>)
- ☞ **Image J:** an open source, Java-based program used to edit, calibrate, process, measure and analyse visual data. (<https://imagej.nih.gov/ij/>)
- ☞ **Le Monde Décodex:** tool that helps users check the source of information circulating online and identify rumours or distortions.
[\(http://www.lemonde.fr/verification/\)](http://www.lemonde.fr/verification/)
- ☞ **Netvizz app:** a Facebook application that extracts a variety of data from different sections of the platform, including groups, fan pages and search function.
[\(https://apps.facebook.com/netvizz/\)](https://apps.facebook.com/netvizz/)
- ☞ **Radarly:** a commercial tool to monitor social media, which allows you to track what is being said about particular topics, people or events online.
[\(http://linkfluence.com/en/products/radarly/\)](http://linkfluence.com/en/products/radarly/)
- ☞ **RAWGraphs:** allows you to create vector-based visualizations of your dataset. Based on the svg format, RAWGraphs is highly customizable and visualizations can be imported in and edited with vector graphics applications for further refinements, or directly embedded into web pages. (<http://rawgraphs.io>)
- ☞ **Spyonweb:** allows you to identify websites associated

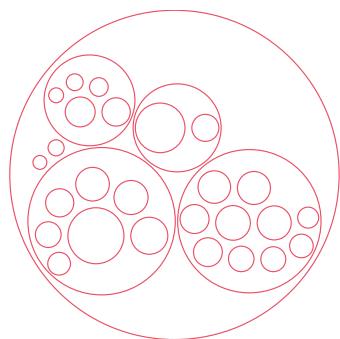
with the same IDs by querying the WHOIS protocol of registered users or assignees of an Internet resource. (<http://spyonweb.com>)

- ☞ **Tab Save:** a Google Chrome extension that allows you to collect and save files such as PDFs, images or list of URLs available on a web page.
(<https://chrome.google.com/webstore/detail/tab-save/lkngoeaeclaebmpkgapchgjdbaekacki?hl=en>)
- ☞ **Table2Net:** allows you to transform tables (.csv) into networks (.gexf).
(<http://tools.medialab.sciences-po.fr/table2net>)
- ☞ **TCAT:** (Twitter Capture and Analysis Toolset) a tool that allows you to retrieve and collect data from Twitter. The datasets can be collected based on keyword, user or hashtag queries.
(<https://wiki.digitalmethods.net/Dmi/ToolDmiTcat>)
- ☞ **The Wayback Machine:** an initiative by the Internet Archive, which archives versions of websites at regular intervals. (<https://archive.org/web/>)

CONCEPTS GLOSSARY

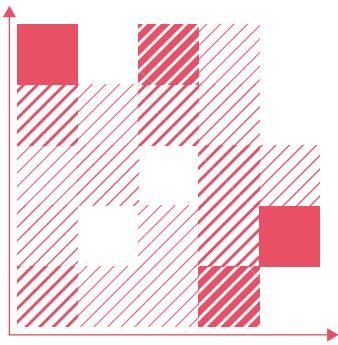


Bubble Graph

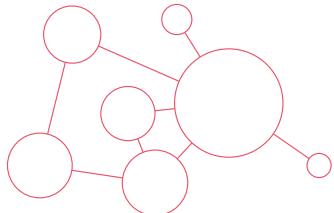


Circle packing

- **API (*Application Programming Interface*)**: a set of clearly defined methods of communication that allows two pieces of software to communicate with each. In Web research APIs are often used to extract data from public or private datasets (typically those collected by social platforms), without having direct access to the database that contains them.
- **Bubble Graph**: a type of scatterplot (see the definition in this glossary) in which the size of the projected point is proportional to a third variable.
- **Circle packing**: a type of data visualisation used to visualize hierarchically structured data. Each cluster/group is represented by a circle. The circle is then packed with smaller circles representing sub-groups. The size of the circle can represent different quantitative properties.
- **Click-bait**: online content with the main purpose of attracting attention and encouraging users to click on a link to a particular page.
- **CMS (*Content Management System*)**: is a software application used to create and manage digital content and websites in particular.
- **Emergent coding**: a technique to classify items through categories that are not presupposed before the observation, but are iteratively defined in the exploration process. The purpose of this type of coding is to remain as close as possible to the categories used by the studied actors themselves instead of fitting data into pre-established categories.



Heatmap



Network Graph

- **Facebook page or group followers:** number of users who have liked a Facebook page or joined a group.
- **Force-directed network layout:** a graph drawing algorithm used to spatialize items inside a network and help make sense of the data. The force-directed layout uses repulsive forces between the nodes while applying attractive forces between adjacent nodes.

→ **Google Analytics ID:** an identification assigned by Google Analytics (Google's service to tracks and reports website traffic) to identify a user account.

→ **Heatmap:** a type of data visualisation in which the variation of values present in a table or matrix are represented by a gradient of colors.

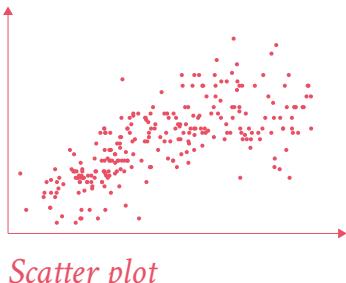
→ **Interactions/engagement:** the total number of likes, shares and comments on a Facebook post (source: <http://www.crowdtangle.com/resources/glossary>).

→ **Network analysis and visualisation:** the process of investigating the connections/relationships between individuals, webpages, accounts or any other group of entities. Using visualization tools such as Gephi it is possible to characterize associative phenomena in terms of nodes (individual actors, people, items) and the ties, or links, that connect them.

→ **Network Graph:** a type of data visualisation used to highlight the relationship between entities, where nodes (or points) represent the entities and lines (or arc, or edges) represent the relationship between them.

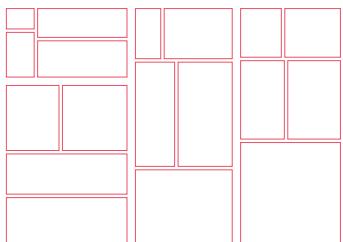
→ **PageRank:** the algorithm used by Google Search to rank the results of its queries. While there are several factors that influence the position of a website on a query (combined in ways that are not publicly known), the basis of the ranking is the recursive count to the references

pointing to a website (how many pages point to a page and how many pages point to those pages).



Scatter plot

- **Scatter plot:** a type of data visualisation in which points are positioned in a Cartesian diagram according to the value that they have on two different variables (corresponding to the axes of the diagram). This type of diagram is most often used to reveal a correlation between the two variables it represents.
- **Source code:** a set of instructions written in programming languages, such as HTML or JavaScript, defining how software should function or a document be displayed.
- **Subscriber Count on CrowdTangle Output Spreadsheet:** the number of subscribers the account had when the post was published - in contrast to the subscriber count found on the account, which represents the current number of subscribers an account has (source: <https://github.com/CrowdTangle/API/wiki/Post#statistics>).



Treemap

- **Treemap:** a type of data visualisation used to represent a hierarchical categorisation through nested rectangles. Each category is associated to a rectangle, whose size is proportional to the importance or weight of the group and which is then filled with smaller rectangles representing sub-groups.
- **Web crawling:** the process of extracting the network of hyperlinks connecting an ensemble of websites or webpages. Crawling is generally performed by automatic or semi-automatic tools called 'spiders' capable to identify and follow all the hyperlinks present on a set of HTML pages.
- **Web scraping:** a method for extracting structured information or content from a website (and saving it in a tabular format).

CONTRIBUTORS AND ACKNOWLEDGEMENTS

Contributors

Co-investigators

Liliana Bounegru, University of Groningen, University of Ghent
Jonathan Gray, King's College London
Tommaso Venturini, Univ Lyon, Inria, ENS de Lyon, CNRS, UCB
Michele Mauri, Density Design, Politecnico di Milano

Research and Editorial Assistant

Daniela Demarchi, University of Amsterdam

Editorial Design

Ángeles Briones, Politecnico di Milano
Carlo De Gaetano, Politecnico di Milano

DensityDesign Researchers

Agata Brilli, Politecnico di Milano
Ángeles Briones, Politecnico di Milano
Carlo De Gaetano, Politecnico di Milano
Gabriele Colombo, Politecnico di Milano
Mariasilvia Poltronieri, Politecnico di Milano
Michele Invernizzi, Politecnico di Milano
Michele Mauri, Politecnico di Milano
Paolo Ciuccarelli, Politecnico di Milano
Tommaso Elli, Politecnico di Milano

Facilitators

Alex Gekker, University of Amsterdam
Anders Munk, Aalborg University
Bilel Benbouzid, University of Paris-Est Marne-la-Vallée
Erik Borra, University of Amsterdam
Esther Weltevreden, University of Amsterdam
Jonathan Gray, King's College London
Jorn Preuss, University of Siegen
Liliana Bounegru, University of Ghent, University of Groningen
Marc Tuters, University of Amsterdam
Mathieu Jacomy, Sciences Po
Natalia Sánchez-Querubín, University of Amsterdam
Nicolas Baya-Laffite, University of Lausanne
Paolo Ciuccarelli, Politecnico di Milano
Richard Rogers, University of Amsterdam
Sabine Niederer, Hogeschool van Amsterdam
Tommaso Venturini, Univ Lyon, Inria, ENS de Lyon, CNRS, UCB

Analysts

Anders Grundtvig, Aalborg University
Anna Keuchenius, University of Amsterdam
Antonio Martella, University of Amsterdam
Asbjørn Fleinert Mathiasen, Aalborg University
Asger Gehrt Olesen, Aalborg University
Carlo Santagiustina, University of Amsterdam
Charlotte Leclercq, University of Amsterdam
Daniel Bach, Aalborg University
Daniela Demarchi, University of Amsterdam
Ecesu Erol, University of Amsterdam
Emil Jørgensen, Aalborg University
Joep Voorn, University of Amsterdam
Jörn Preuss, University of Siegen
Kaspar Beelen, University of Amsterdam
Katerina Gladkova, University of Amsterdam
Lieke Kersten, University of Amsterdam
Lisanne Blomberg, University of Amsterdam
Manon van Hoek, University of Amsterdam
Maria Hayat, University of Amsterdam
Marlene Scherf, University of Amsterdam
Michel Blonk, University of Amsterdam
Mintsje de Witte, University of Amsterdam
Mischa Benjamin Szpir, Aalborg University
Pieter Vliegenthart, University of Amsterdam
Rina Tsubaki, European Journalism Centre
Ronja Ingeborg Lofstad, Aalborg University
Sal Hagen, University of Amsterdam
Stefani Mans, University of Amsterdam
Stefanie Voortman, University of Amsterdam
Talia Castellanos Usigli, University of Amsterdam
Zoë Versteegen, University of Amsterdam

Contributors by chapter

chapter 1

MAPPING FAKE NEWS HOTSPOTS ON FACEBOOK

- Facilitators: Erik Borra, University of Amsterdam
Liliana Bounegru, University of Groningen,
University of Ghent
Jonathan Gray, King's College London
Natalia Sánchez-Querubín, University of
Amsterdam
Esther Weltevreden, University of Amsterdam
- Members: Lisanne Blomberg, University of Amsterdam
Talía Castellanos, University of Amsterdam
Tommaso Elli, DensityDesign Research Lab
Mintje de Witte, University of Amsterdam
Stefanie Voortman, University of Amsterdam
Stefani Mans, University of Amsterdam
Antonio Martella, University of Amsterdam
Rina Tsubaki, European Journalism Centre
Manon van Hoek, University of Amsterdam
Zoë Versteegen, University of Amsterdam
Joep Voorn, University of Amsterdam

chapter 2

TRACING THE CIRCULATION OF FAKE NEWS ON THE WEB

- Facilitators: Mathieu Jacomy, Sciences Po
Anders Grundtvig, Aalborg University
Tommaso Venturini, Univ Lyon, Inria, ENS de
Lyon, CNRS, UCB
- Members: Agata Brilli, DensityDesign Research Lab
Daniela Demarchi, University of Amsterdam
Ronja Lofstad, Aalborg University
Anders Kristian Munk, Aalborg University

chapter 3

USING TRACKER SIGNATURES TO MAP THE TECHNO-COMMERCIAL UNDERPINNINGS OF FAKE NEWS SITES

- Facilitator: Liliana Bounegru, University of Groningen,
University of Ghent
Jonathan Gray, King's College London
Richard Rogers, University of Amsterdam
- Members: Michele Invernizzi, DensityDesign Research Lab
Mischa Szpir, Aalborg University

chapter 4

HOW TO STUDY POLITICAL MEMES ON FACEBOOK

- Facilitators: Nicolas Baya-Laffite, University of Lausanne
Bilel Benbouzid, University of Paris-Est
Marne-la-Vallée
Marc Tuters, University of Amsterdam
- Members: Daniel Bach, Aalborg University
Carlo De Gaetano, DensityDesign Research Lab
Sal Hagen, University of Amsterdam
Emil Jørgensen, Aalborg University

chapter 5

MAPPING TROLL-LIKE PRACTICES ON TWITTER

- Facilitators: Erik Borra, University of Amsterdam
Sabine Niederer, Hogeschool van Amsterdam
Jörn Preuß, University of Siegen
Esther Weltevreden, University of Amsterdam
- Members: Ángeles Briones, DensityDesign Research Lab
Michel Blonk, University of Amsterdam
Lieke Kersten, University of Amsterdam
Carlo Santagiustina, University of Amsterdam
Marlene Scherf, University of Amsterdam
Pieter Vliegenthart, University of Amsterdam

We would also like to register our gratitude to the following people who provided invaluable input, feedback, advice and support at various stages: Claire Wardle and Jenni Sargent at First Draft; Ida Eklund-Lindwall at East Stratcom Task Force; Jayson Harsin, The American University of Paris (AUP); Craig Silverman and Lam Thuy Vo at BuzzFeed News; friends and colleagues at Le Monde, NRC, the New York Times and other organisations with whom we corresponded with about the guide. We also benefited from discussions with participants at public talks, workshops and events including the International Journalism Festival 2017 in Perugia; "Fake News, Algorithmic Accountability and the Role of Data Journalism in the Post-Truth Era" at the Centre for Research in the Arts, Social Sciences and Humanities (CRASSH), University of Cambridge; "Data Publics" at Lancaster University; "Politics, Fake News and the Post-Truth Era" at the University of Bath; "Les fausses nouvelles : le nouveau visage d'un vieux problème" at the Montréal University; "Data Storytelling: Engaging Visual Narratives" at Data for Culture conference, Katowice Miasto Ogrodów; the "Social Life of Fake News Online" at King's College London and "Social Media and Democracy: New Challenges for Political Communication Research" at Lund University and the University of Copenhagen.

<http://fakenews.publicdatalab.org/>