

A FIELD  
GUIDE TO

# FAKE NEWS

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

IPR  
Institute for  
Policy Research

UNIVERSITY OF  
BATH

POLITECNICO  
MILANO 1863

CJS

UNIVERSITÄT  
SIEGEN

Unil  
UNIL | Université de Lausanne  
Institut des sciences sociales  
STSLab & LaDHUL

LISIS

SciencesPo  
MÉDIACLAS

Citizen Data Lab

TANTARAJA.DK

A FIELD GUIDE TO FAKE NEWS

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

This guide explores the use of digital methods to trace the production, circulation and reception of fake news online. It is a project of the Public Data Lab with support from First Draft. The Public Data Lab ([publicdatalab.org](http://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](http://firstdraftnews.com)) is dedicated to improving skills and standards in the reporting and sharing of information that emerges online.

This is a **SAMPLE** of A Field Guide to Fake News released to coincide with the International Journalism Festival in Perugia in April 2017. If you'd like to be notified when the full version is released, please visit: [fakenews.publicdatalab.org](http://fakenews.publicdatalab.org) For further information you can email [contact@publicdatalab.org](mailto:contact@publicdatalab.org)

© 2017 Public Data Lab.

The guide is released under the Creative Commons Attribution License ([creativecommons.org/licenses/by/3.0/](http://creativecommons.org/licenses/by/3.0/)). 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, 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  
Citizen Data Lab, Amsterdam University of Applied Sciences  
Density Design, Politecnico di Milano  
Digital Methods Initiative, University of Amsterdam  
Institut National de Recherche en Informatique et en Automatique (INRIA)  
Institute for Policy Research, University of Bath  
Laboratoire d'Étude des Sciences et des Techniques (STS-Lab), Université de Lausanne  
Laboratoire Interdisciplinaire Sciences Innovations Sociétés (LISIS), Université Paris-Est Marne-la-Vallée  
Médialab, Sciences Po, Paris  
Techno-Anthropology Lab, Aalborg University Copenhagen  
University of Siegen

Published by Public Data Lab.  
Sample released in April 2017.



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



PUBLIC  
DATA  
LAB

FIRST

DRAFT

DEN -

SITY

GN +

DIGITAL  
METHODS  
INITIATIVE

inria



# CONTENTS

## Contributors

### Co-investigators

Liliana Bounegru, University of Ghent, University of Groningen  
Jonathan Gray, University of Bath  
Tommaso Venturini, Univ Lyon, Inria, ENS de Lyon, CNRS, UCB  
Michele Mauri, Density Design, 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 Weltverde, University of Amsterdam  
Jonathan Gray, University of Bath  
Jorn Preuss, University of Siegen  
Liliana Bounegru, University of Ghent, University of Groningen  
Marc Tuters, University of Amsterdam  
Mathieu Jacomy, Sciences Po  
Natalia Sánchez, University of Amsterdam  
Nicolas Baya-Laffite, University of Lausanne  
Richard Rogers, University of Amsterdam  
Sabine Niederer, Hogeschool van Amsterdam  
Tommaso Venturini, Univ Lyon, Inria, ENS de Lyon, CNRS, UCB

### Analysts

Agata Troost, University of Utrecht  
Anders Grundtvig, Aalborg University  
Anna Keuchenius, University of Amsterdam  
Antonio Martella, DataActive  
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 Grosby Jorgensen, Aalborg University  
Emillie Victoria de Keulenaar, University of Amsterdam  
Joep Voorn, University of Amsterdam  
Jorn 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

### Research and Editorial Assistant

Daniela Demarchi, University of Amsterdam

## 7 Introduction

### CHAPTER 1

<b>15</b>	<b>MAPPING FAKE NEWS HOTSPOTS ON FACEBOOK</b>
18	1. What publics does fake news animate on facebook?
38	2. How may the trajectory of a fake news story be traced on facebook?
50	3. Do fact-checking initiatives reach the publics of fake news on facebook?

### CHAPTER 2

<b>59</b>	<b>TRACING THE CIRCULATION OF FAKE NEWS ON THE WEB</b>
62	1. Where do fake news originate? By what sites are they first retransmitted?
76	2. Which are the most visible sources related to a fake story? When and by whom are they mentioned?

### CHAPTER 3

<b>93</b>	<b>USING TRACKER SIGNATURES TO MAP THE TECHNO-COMMERCIAL UNDERPINNINGS OF FAKE NEWS SITES</b>
95	1. Do fake news sites use different kinds of trackers from mainstream media sites?
101	2. How can fake news and mainstream media sites be profiled based on their tracker usage?
107	3. How do tracker ecologies on fake news sites change over time?
113	4. Which other websites share the same tracker IDs as fake news sites?
119	5. Do trackers associated with hyper-partisan and misinformation sites vary across language spheres?

*This is a sample of A Field Guide to Fake News, which will contain further chapters on how to map and investigate fake news, online misinformation and other associated phenomena – including recipes on memes, bots, trolls, propaganda and fact-checking initiatives.*

# A FIELD GUIDE TO FAKE NEWS

## INTRODUCTION

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.<sup>[1]</sup> Others suggest that new online platforms accelerate and “supercharge” their circulation in a way which introduces hitherto unprecedented challenges and dynamics.<sup>[2]</sup> Proposed responses include new media literacy, educational and fact-checking initiatives; new laws, policies and fines for technology companies who fail to remove

[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, John Naughton, “Facebook and Twitter could pay the price for hate speech”, *The Guardian*: <https://www.theguardian.com/commentisfree/2017/mar/19/john-naughton-germany-fine-social-media-sites-facebook-twitter-hate-speech>

[4] See, for example, Full Fact, “The State of Automated Factchecking”, <https://fullfact.org/automated/> and <http://www.512tech.com/technology/austin-startup-wants-end-fake-news-and-fake-everything-else-the-internet/EcchWFgrl4PQmjPvm-kzycJ/>

offending content;<sup>[3]</sup> and a host of new startups and technical fixes – from authenticated content to automated fact-checking projects.<sup>[4]</sup>

Across these different kinds of responses, observers agree that fake news cannot be straightforwardly defined. “Fake news” comes in many different shades. This need not be taken as proof of the futility of investigating this phenomenon. On the contrary: accepting that there is no easy way to demarcate between “fake” from “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 phenomenon. 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 fake news in isolation to understanding the contexts in which it *circulates* online. This shift points to the limits of a “deficit model” approach – which might imply that fake news thrives 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.

While online and platform metrics often serve to take measure of engagement by means of what Theodore Porter calls “thin descriptions” – 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.

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

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.<sup>[5]</sup> 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 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”<sup>[6]</sup> – 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”.<sup>[7]</sup> 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.<sup>[8]</sup> 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

[6] 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*, Mahwah, NJ: Erlbaum, 1997. Available at: <http://polaris.gseis.ucla.edu/pagre/critical.html>

[7] Shannon Mattern, “Cloud and Field”, *Places Journal*: <https://placesjournal.org/article/cloud-and-field/>

[8] 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> and <http://www.makingandknowing.org/>

We hope that the recipes in this guide will enrich investigations of fake news 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.

In Edgar Allan Poe’s classic mystery story “The Purloined Letter”, 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 and Carolin Gerlitz characterise as “distributed accomplishment” or what Mike Ananny and Kate Crawford describe as “relational achievement”.<sup>[9]</sup> 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.

[9] 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.

[10] M. S. Abildgaard, A. Birkbak, L. Bounegru, J. Gray, M. Jacomy, A. K. Madsen, A. K. Munk (forthcoming) “Fake News: Seven Lines of Inquiry”.

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.<sup>[10]</sup> 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 and research.<sup>[11]</sup> We also draw upon our field’s interest

in public engagement and participation around digital technologies and data infrastructures.<sup>[12]</sup> 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.<sup>[13]</sup>

Two of us have a background in data journalism, having co-edited The Data Journalism Handbook and undertaken various initiatives in this field.<sup>[14]</sup> 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”,<sup>[15]</sup> 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

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

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

[13] 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.

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

[15] 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.

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 Density Design (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), the Techno-Anthropology Lab (Aalborg University Copenhagen, Denmark) and the Utrecht University – without whose energy, creativity and dedication this project would not have been possible.

Jonathan Gray, Liliana Bounegru, Tommaso Venturini  
London, March 2017

## *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”, August 2016, *The New York Times*: <https://www.nytimes.com/2016/08/28/magazine/inside-facebooks-totally-insane-unintentionally-gigantic-hyperpartisan-political-media-machine.html>

generating engagement<sup>[1]</sup>. 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?

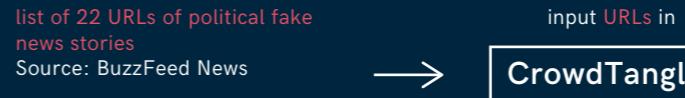
## BEFORE STARTING

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.

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 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

CrowdTangle



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

Google News



→ visualise

a

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

import data in

RAWGraphs



→ visualise

b

**WHICH FACEBOOK PAGES AND GROUPS SHARE FAKE NEWS?**

import data in

Gephi



→ visualise

c

**DO FACEBOOK PUBLICS HAVE PREFERRED STORY THEMES?**

manually categorise facebook pages



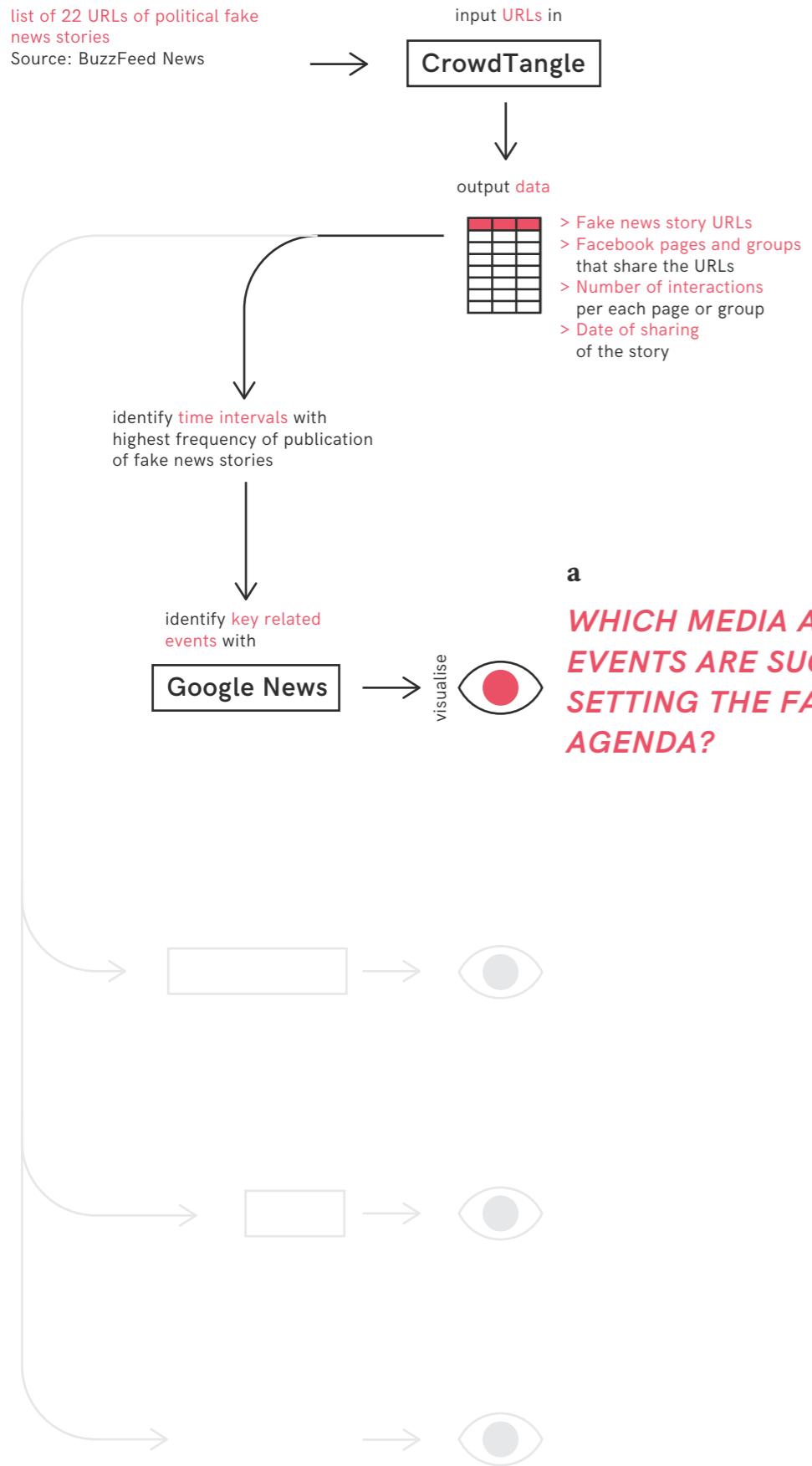
→ visualise

d

**WHAT KINDS OF PUBLICS ARE ENERGISED BY FAKE NEWS?**

## START

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



## CHAPTER 1 → RECIPE 1

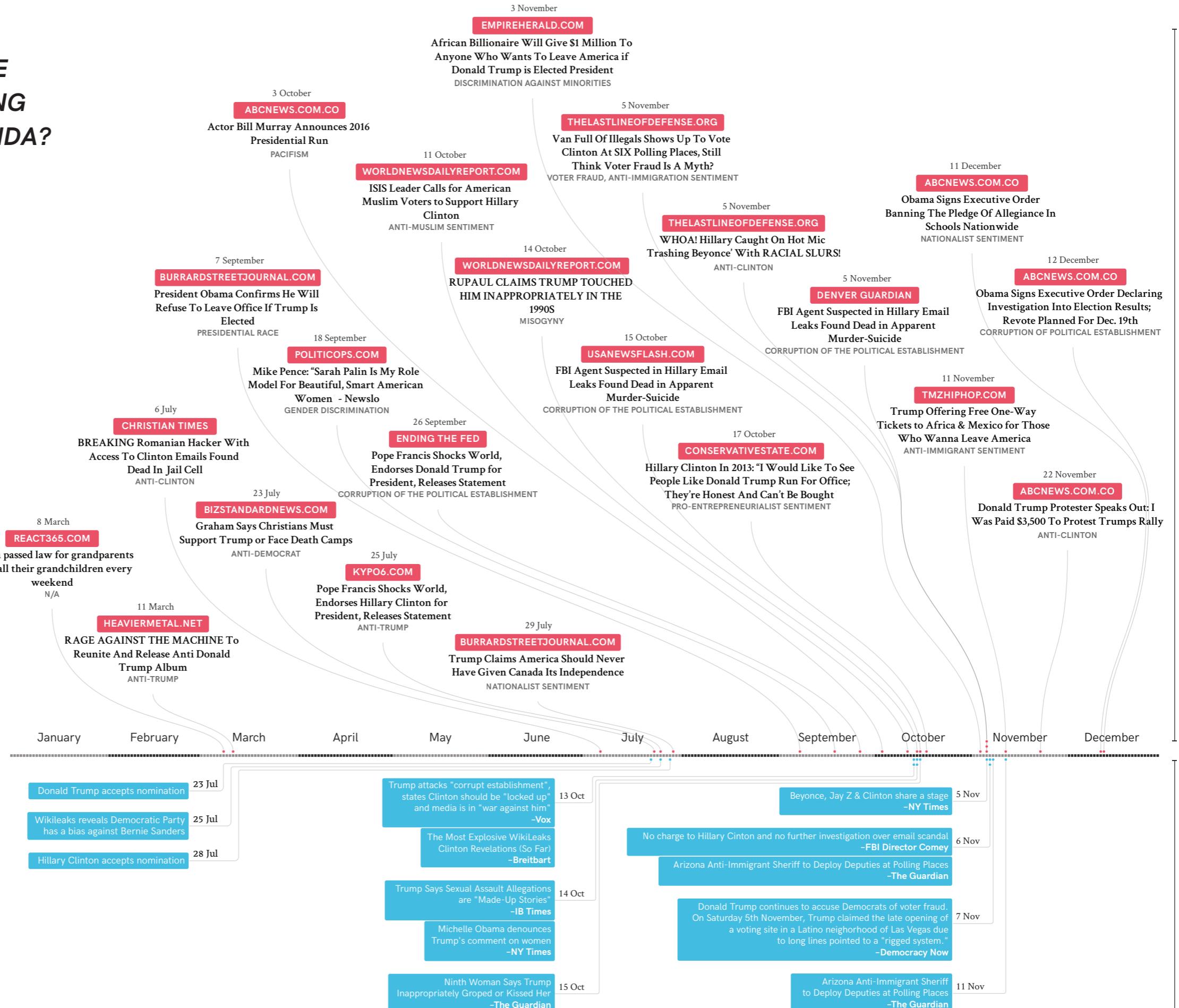
### a. EXAMINE THE THEMES EXPLOITED IN FAKE NEWS STORIES AND IDENTIFY THE EVENTS WHICH THEY EDITORIALISE

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.

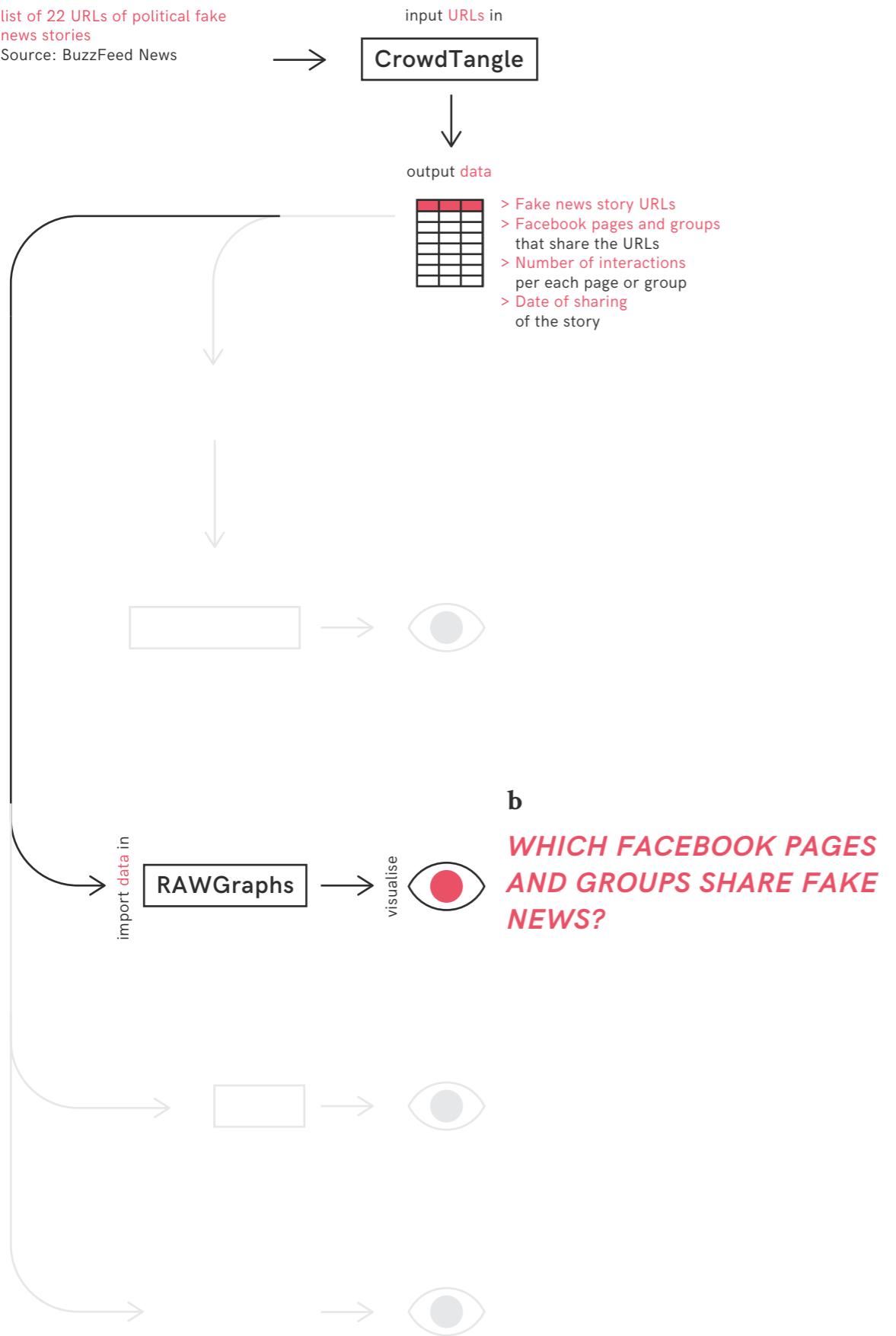
# WHICH MEDIA AND POLITICAL EVENTS ARE SUCCESSFUL IN SETTING THE FAKE NEWS AGENDA?

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

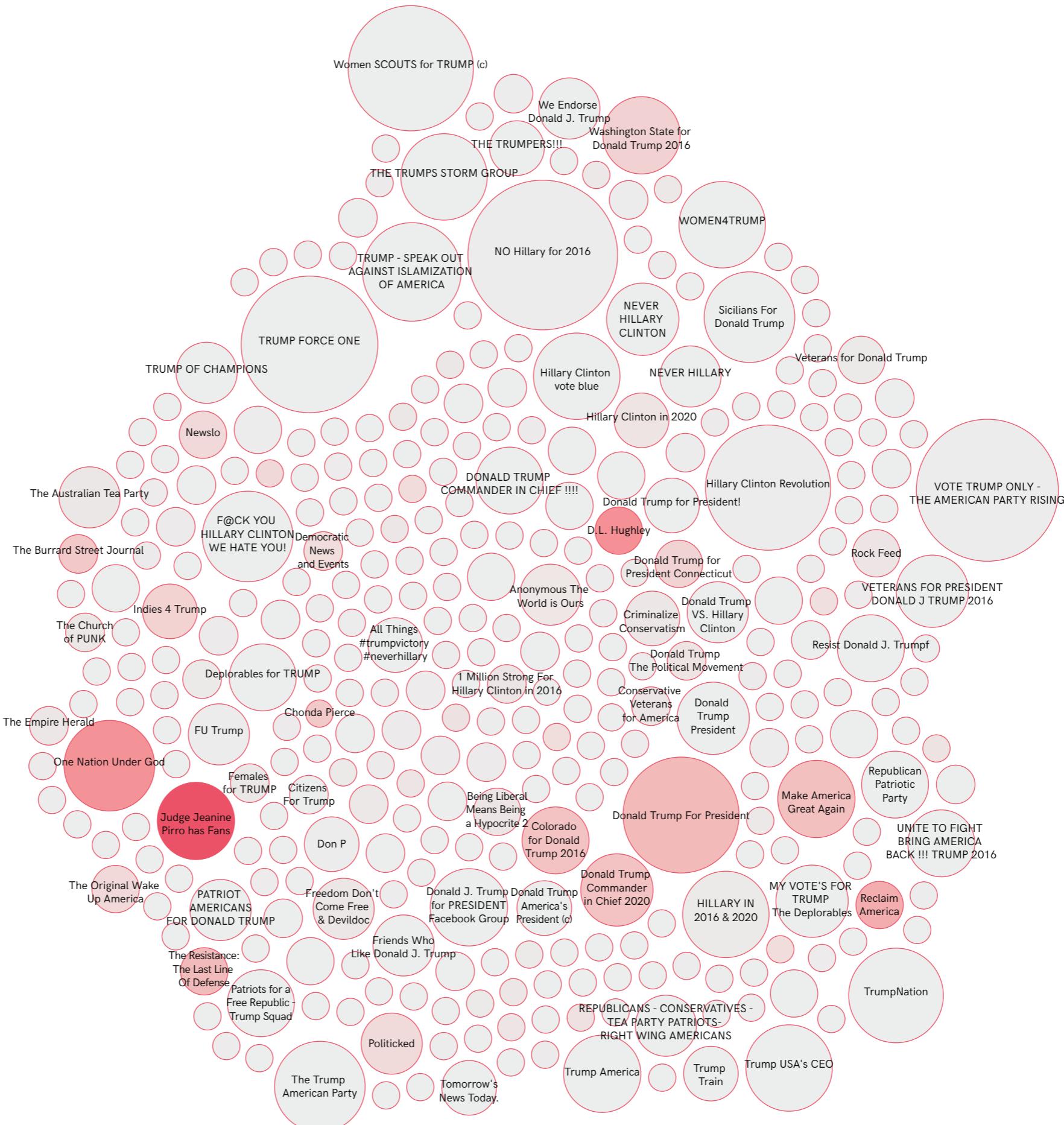
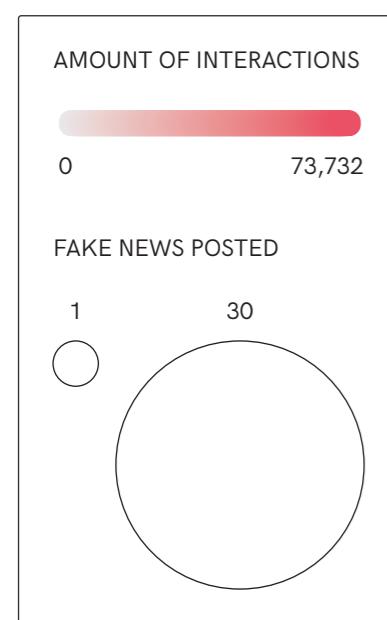
This may be done with a social media monitoring tool such as the browser extension of → CrowdTangle. The number of followers per page or group as well as the number of interactions per posts should be recorded in a spreadsheet alongside the names of pages and groups that share fake news stories.

- ◊ 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 important posts from verified pages as well as from pages with more than 750.000 followers. [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>

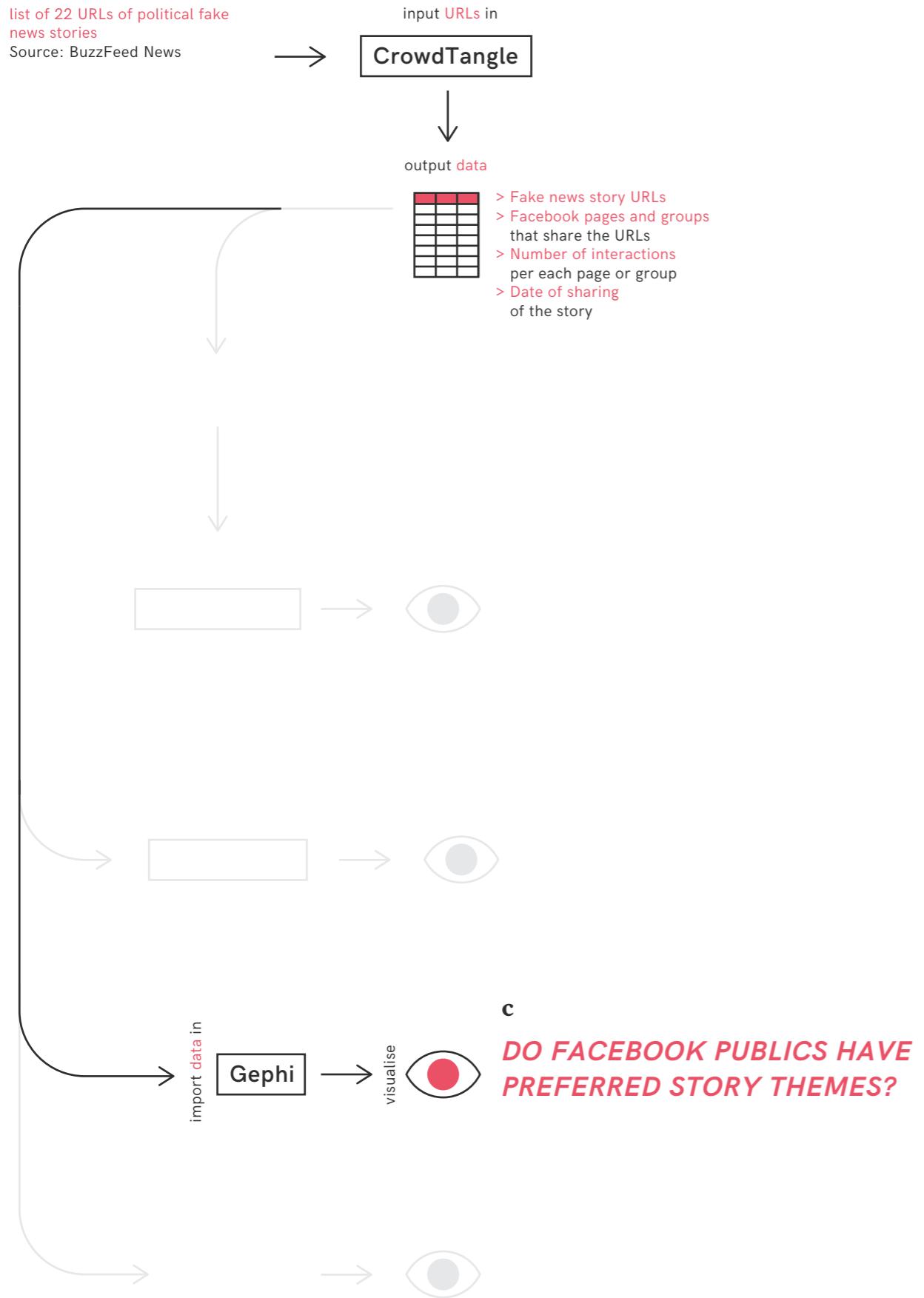
# WHICH FACEBOOK PAGES AND GROUPS PROMOTED THE HIGHEST NUMBER OF FAKE NEWS? WHICH ONES CREATED THE HIGHEST ENGAGEMENT?

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



## CHAPTER 1 → RECIPE 1

### 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 Facebook pages and groups connected by the stories which they share.

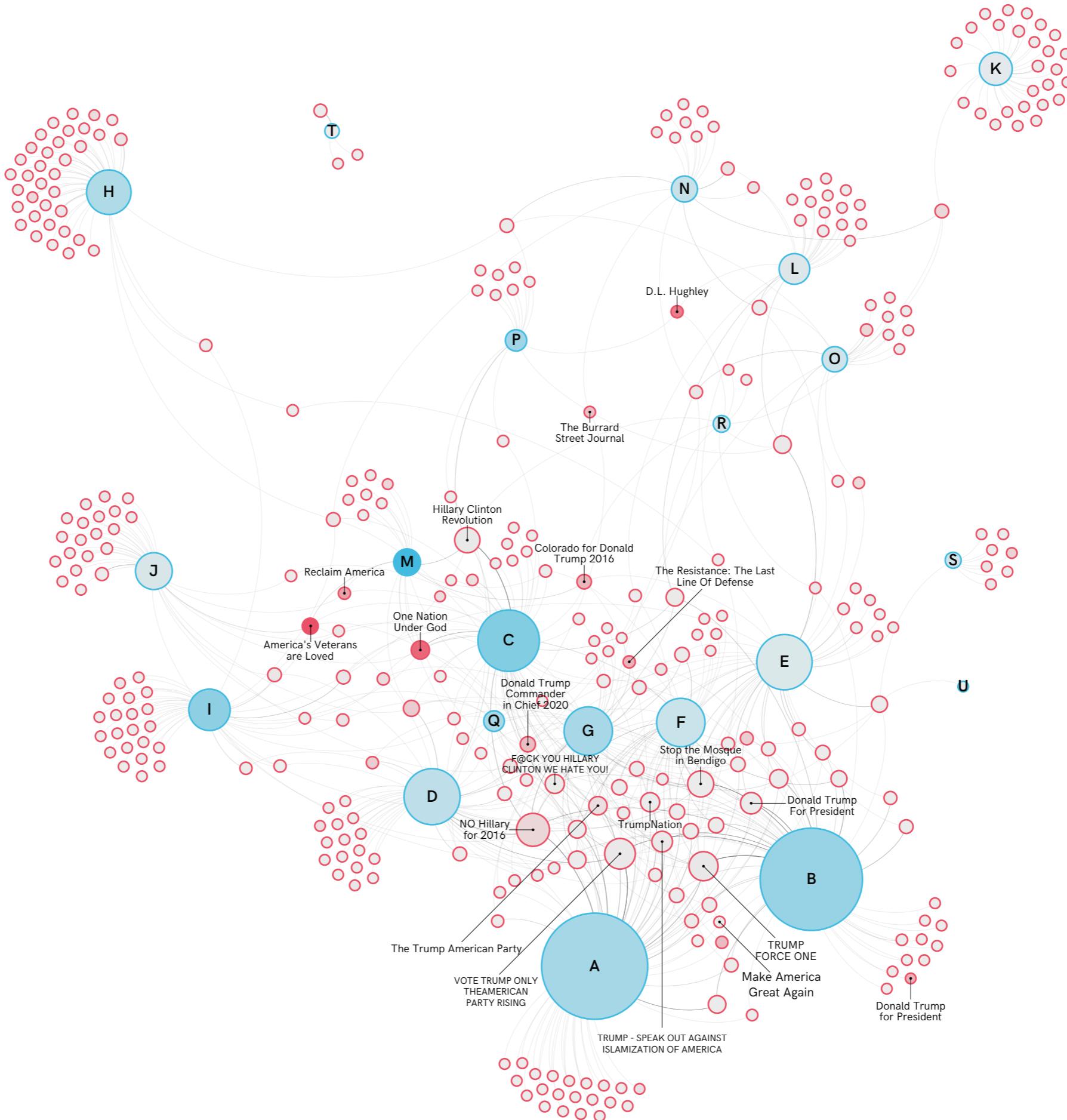
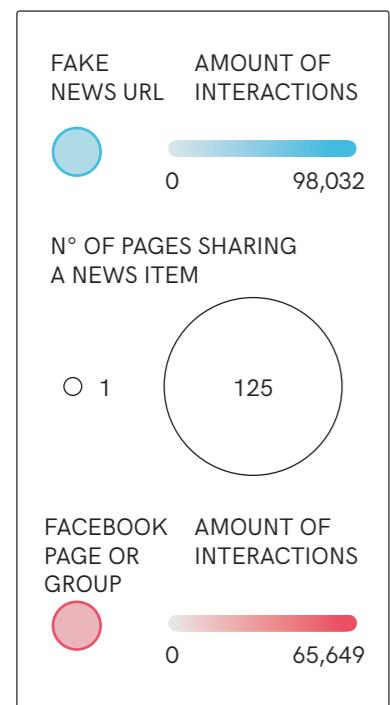
- ◊ Starting from the dataset extracted with → CrowdTangle's browser extension, you may create a network file where each time a Facebook group or page posts a fake news story a link is established between that page or group and that story.
- ◊ You may use → Table2Net to convert your CSV (comma-separated values) file into a network file and → Gephi to explore the network. A force-directed layout algorithm such as ForceAtlas2<sup>[2]</sup> 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>

### c DO FACEBOOK PUBLICS HAVE PREFERRED STORY THEMES?

# DO FACEBOOK PUBLICS HAVE PREFERRED STORY THEMES?

**Network of 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.



## 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



## CHAPTER 1 → RECIPE 1

### d. PROFILE THE PUBLICS ANIMATED BY FAKE NEWS

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

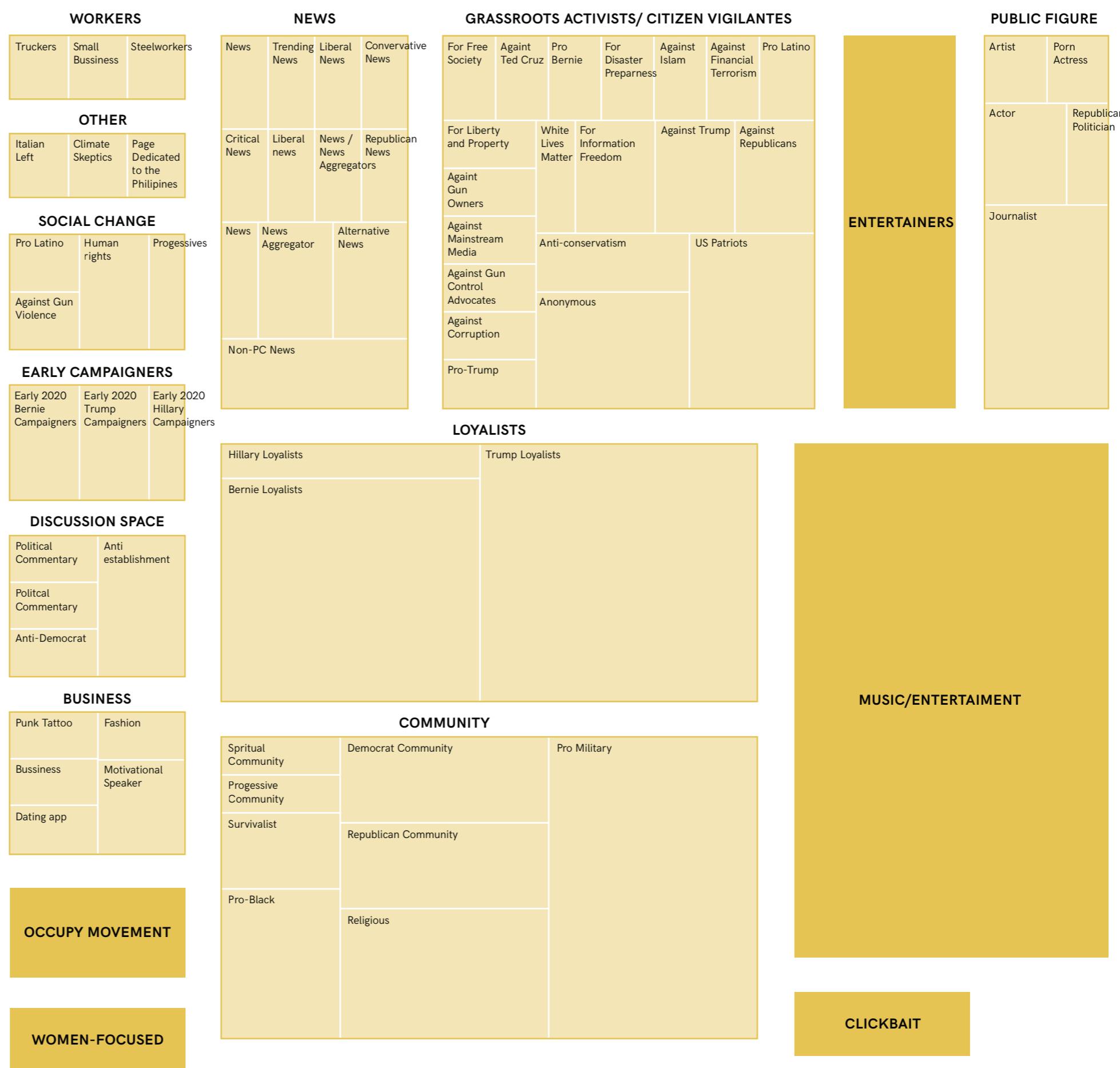
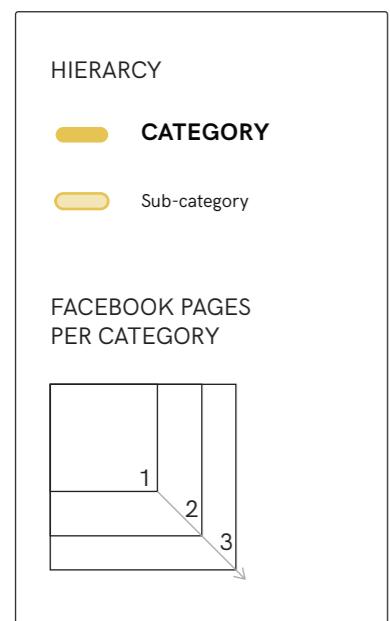
- ◊ You may take an “emergent coding” approach to identify the themes that emerge from the description of pages. You may take note of a more generic category (e.g. “grass-roots activism”) as well as a more specific one (e.g. “anti-establishment”).
- ◊ Sum up the amount of followers across all pages belonging 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.

**d**

**WHAT KINDS OF PUBLICS ARE ENERGISED BY FAKE NEWS?**

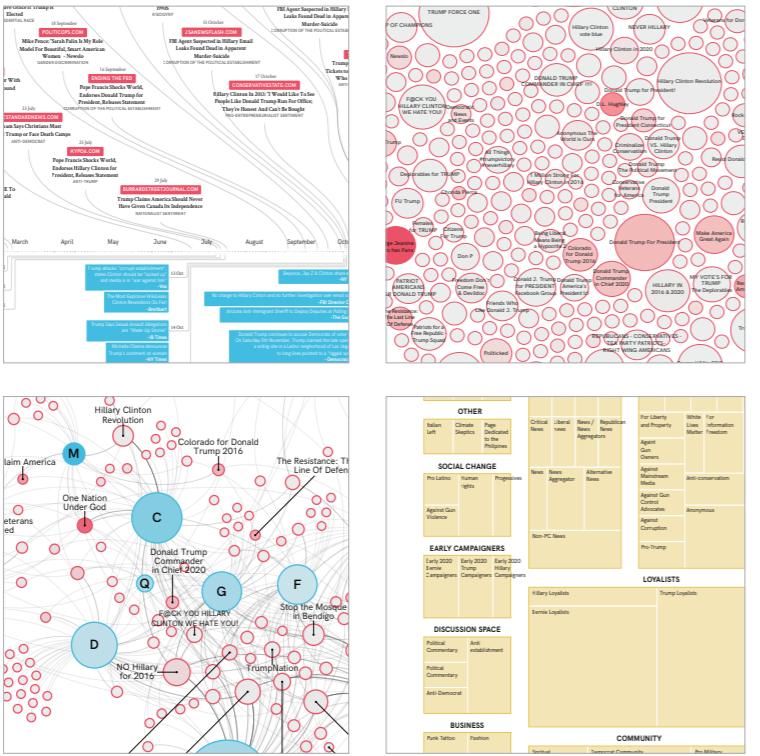
# WHAT KINDS OF PUBLICS ARE ENERGISED BY FAKE NEWS?

**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.



## 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 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).

## START

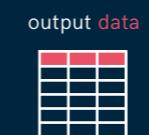
select a **fake news story**  
you want to trace

design queries to identify  
URLs where story is published

**Google Search**

compile list of **URLs**  
where story is published

input **URL** in  
**CrowdTangle**



output **data**  

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

input **data** to  
**RAWGraphs**



**a**

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

input each **URL** in  
**CrowdTangle**



output **data** for each **URL**

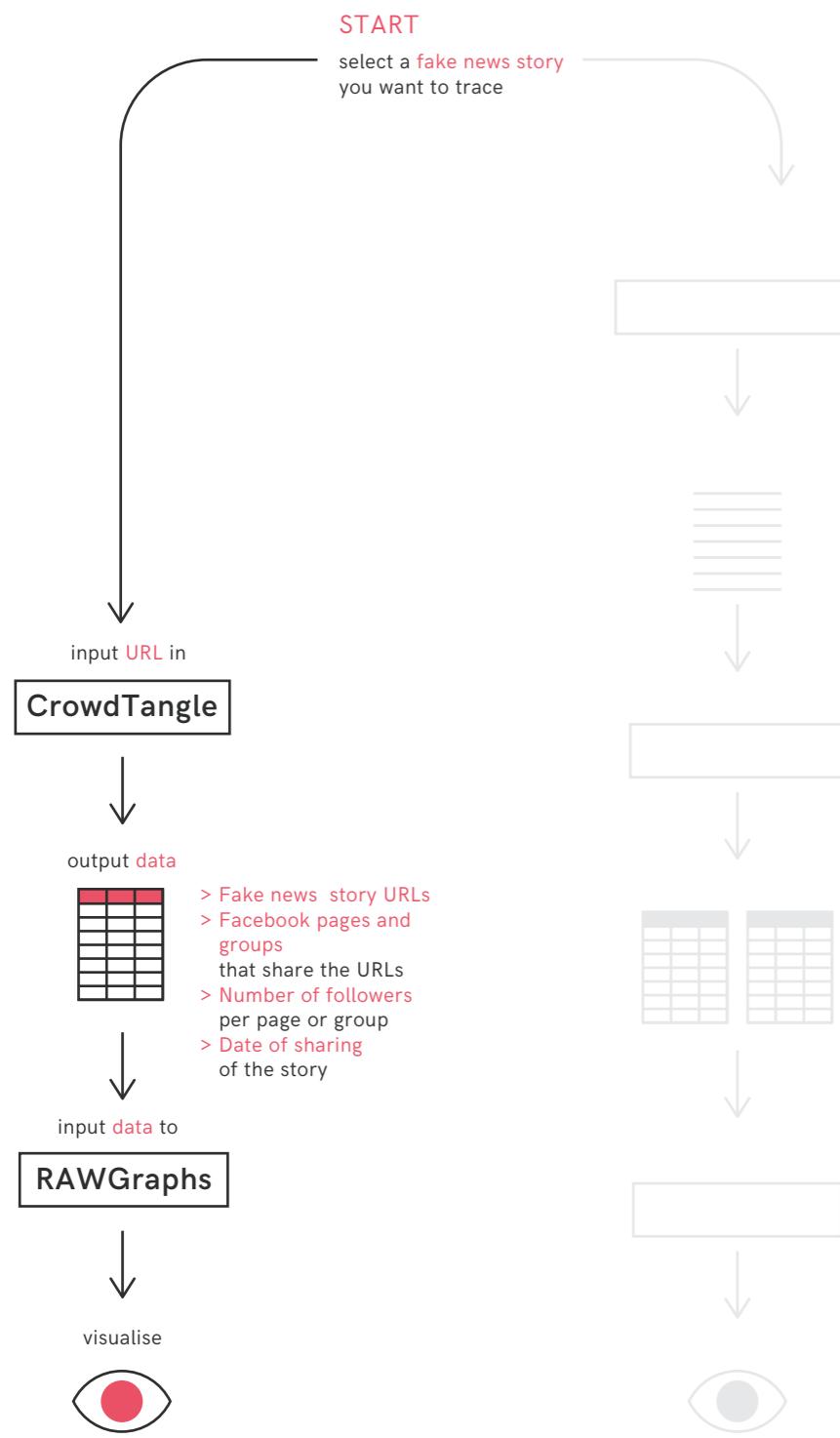
- > Fake news story URLs
- > Facebook pages and groups that share the URLs
- > Number of followers per page or group
- > Date of sharing of the story

input **data** to  
**RAWGraphs**



**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?**



## CHAPTER 1 → RECIPE 2

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

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.

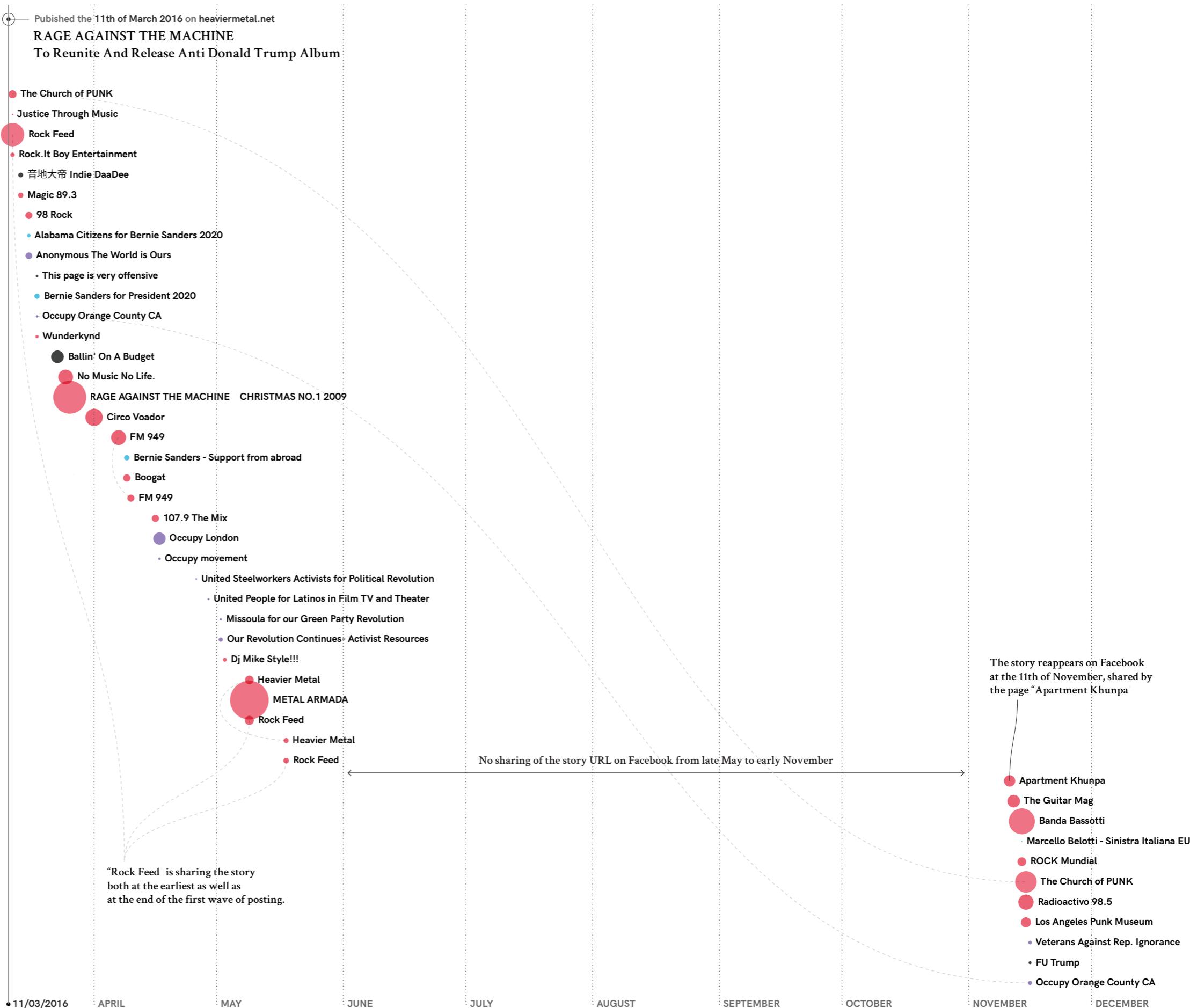
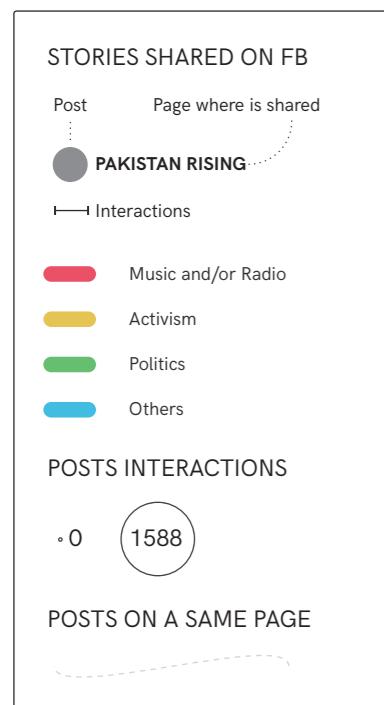
- ◊ The names of these pages and groups, their followers' count, the interactions that they generate as well as the date of sharing of the story may be recorded in a spreadsheet per story URL.
- ◊ To explore the temporal dynamics of the circulation of the fake news story on Facebook, you may plot its 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 a story affects its circulation.
- ◊ 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 for the most prominent fake news stories Facebook posts that share them are often no longer available via the Facebook interface and API.

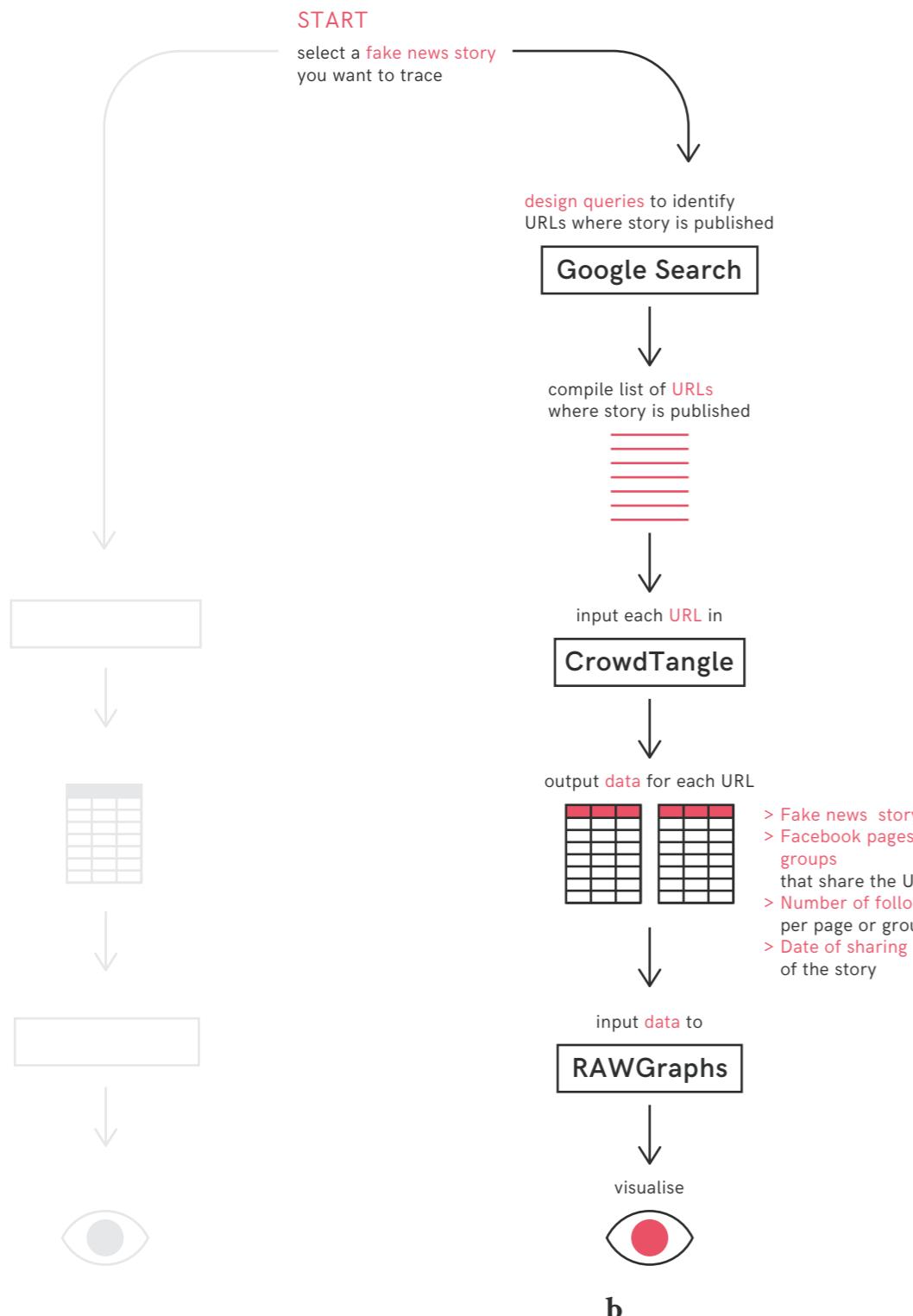
**a**

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

# **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.** 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.





**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?**

CHAPTER 1 → RECIPE 2

### 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.

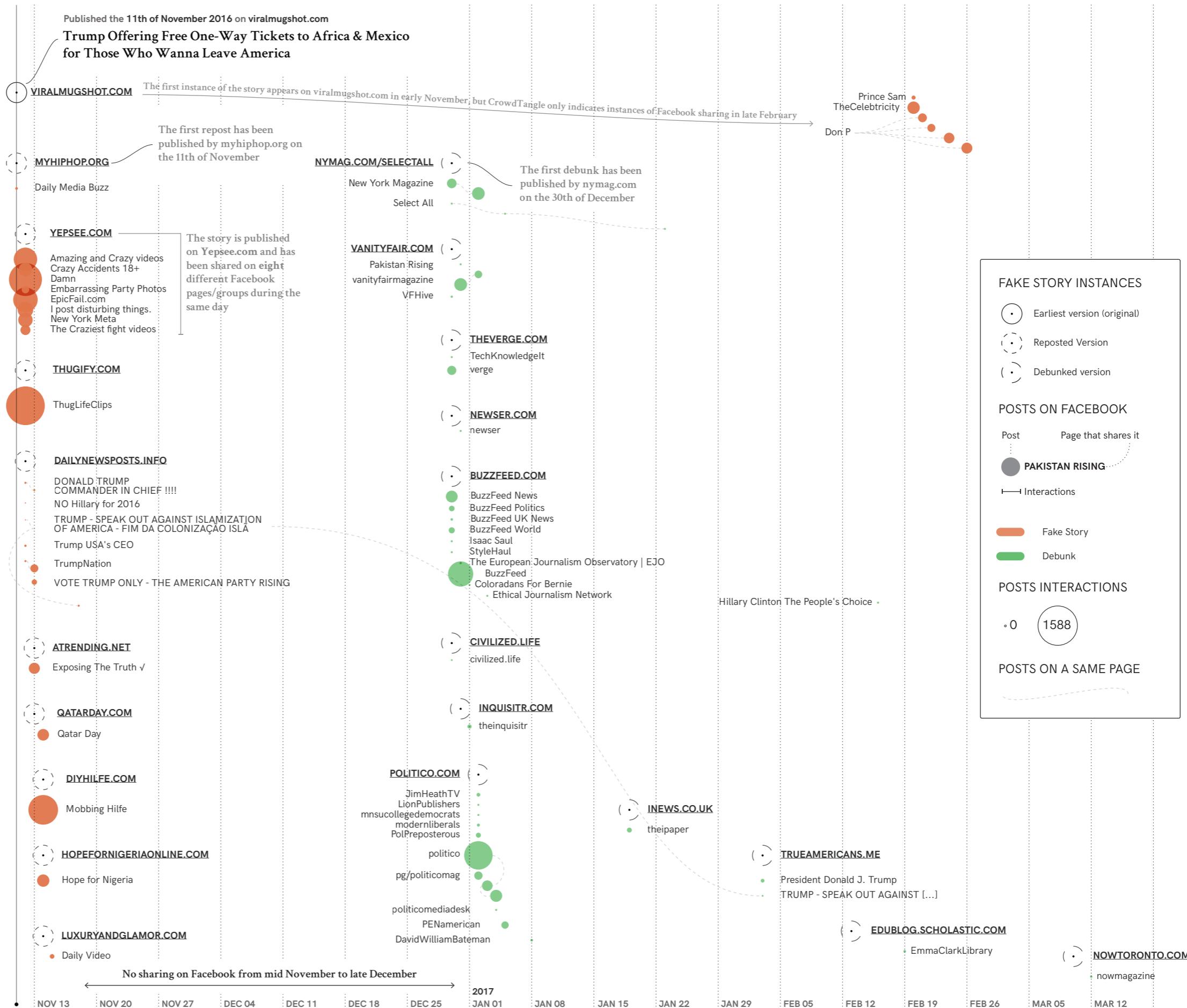
- ◊ To identify the websites which republish a story as well as those which debunk it, you may query the title of the fake news in a search engine of choice (e.g. → **Google Web Search**) using a research browser [1] and extract the URLs corresponding to instances of republication and debunking of the 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 share the URLs corresponding to both the fake story and its debunked versions.
- ◊ 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.

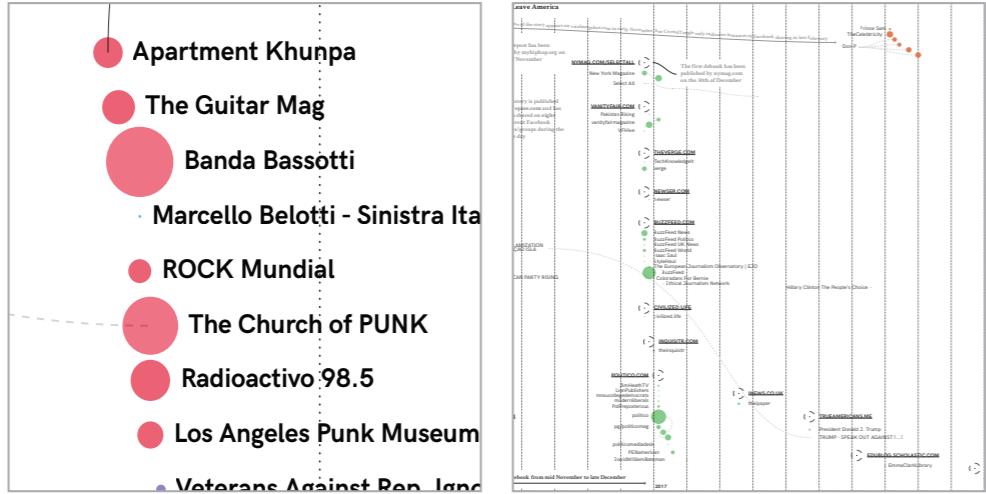
[1]  
See instructions on how to set up a research browser in this video tutorial: <https://www.youtube.com/watch?v=bj65Xr9GkJM>

**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.





## 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 Facebook groups 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  
≡ + ≡

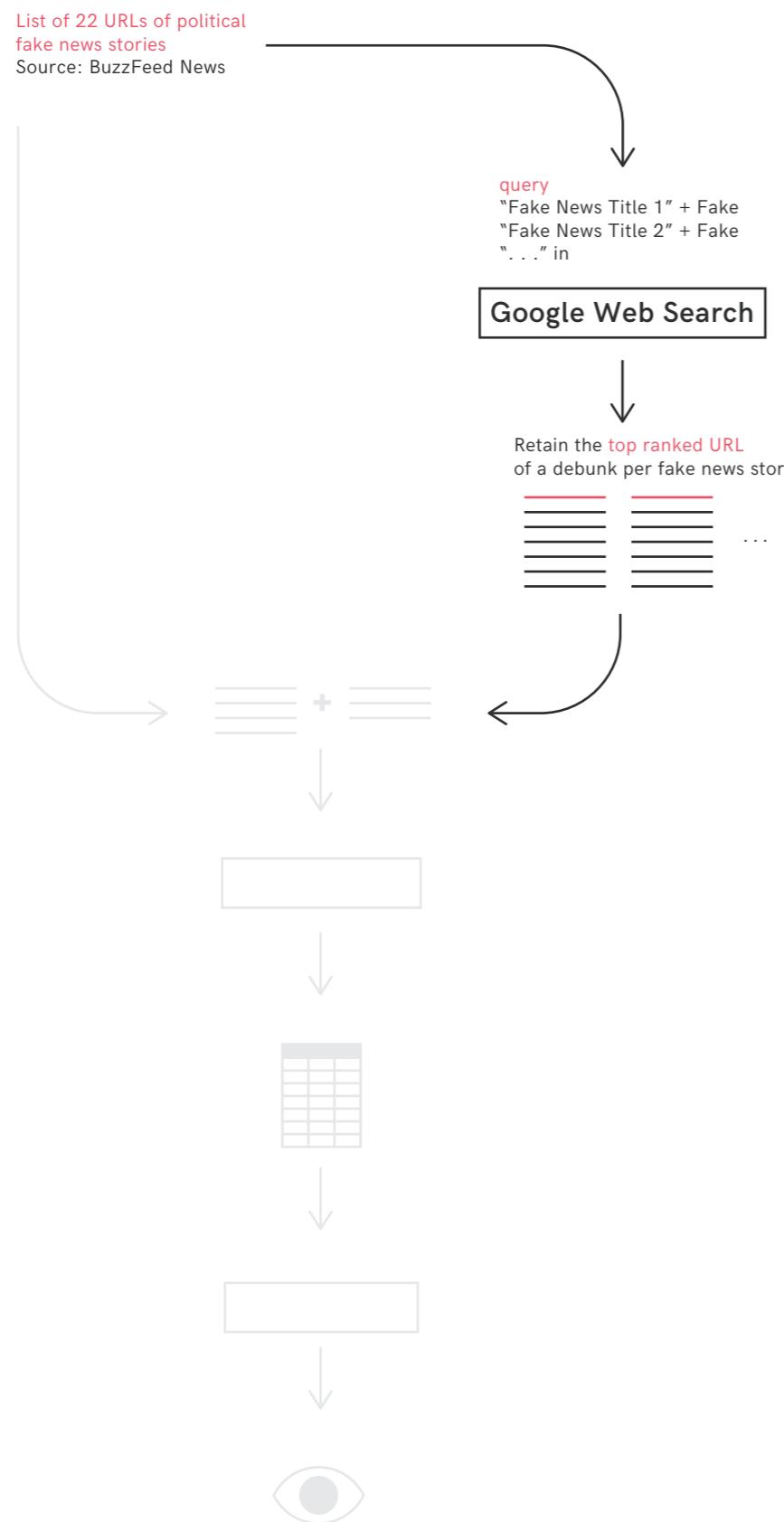
↓  
input URLs in  
**CrowdTangle**

↓  
output data  
grid  
> 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  
eye icon  
**ARE DEBUNKED VERSIONS  
OF FAKE NEWS STORIES  
ACKNOWLEDGED BY THE  
PUBLICS OF FAKE NEWS?**

## START



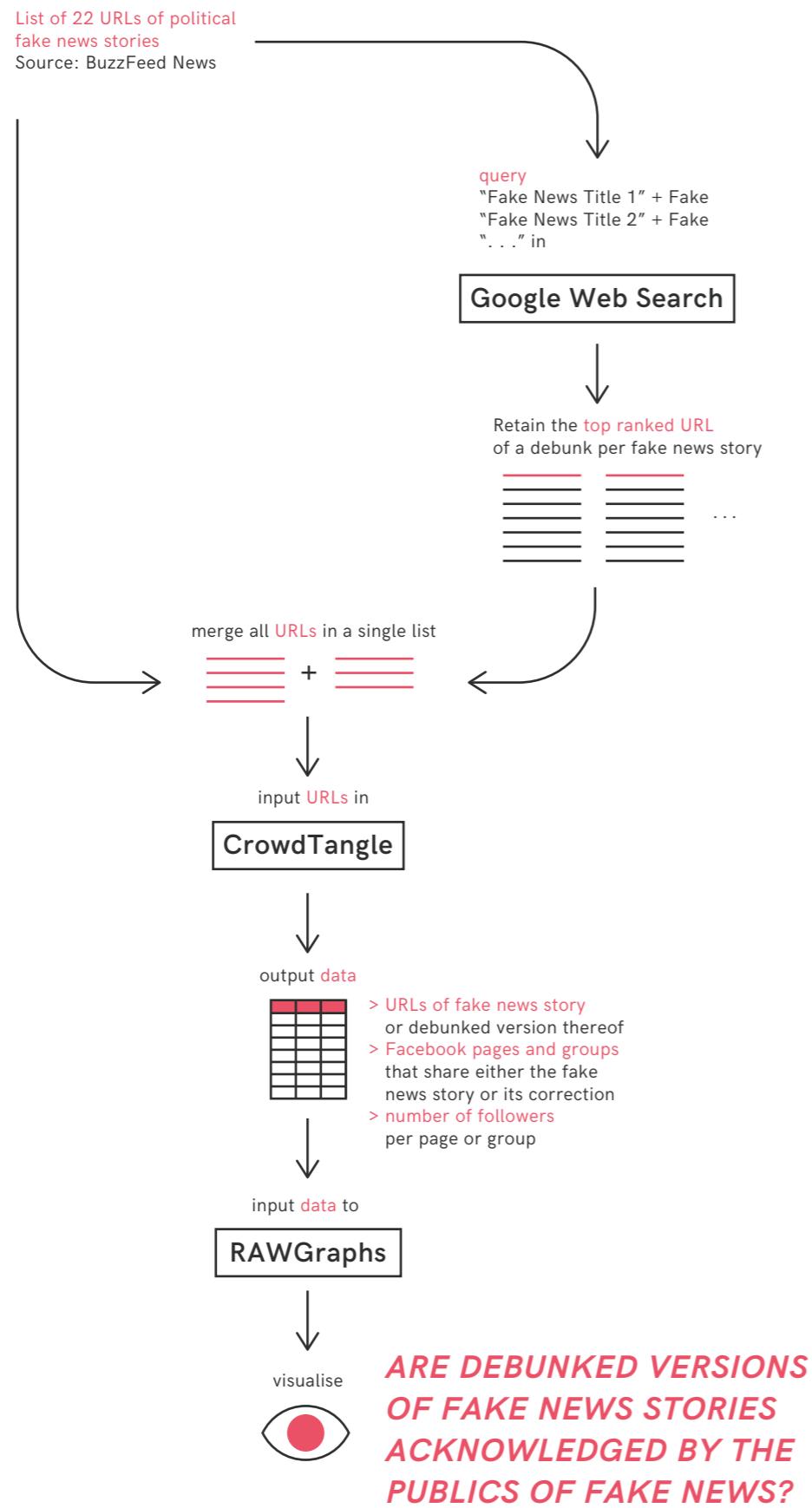
CHAPTER 1 → RECIPE 3

### 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. Alternatively, 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



CHAPTER 1 → RECIPE 3

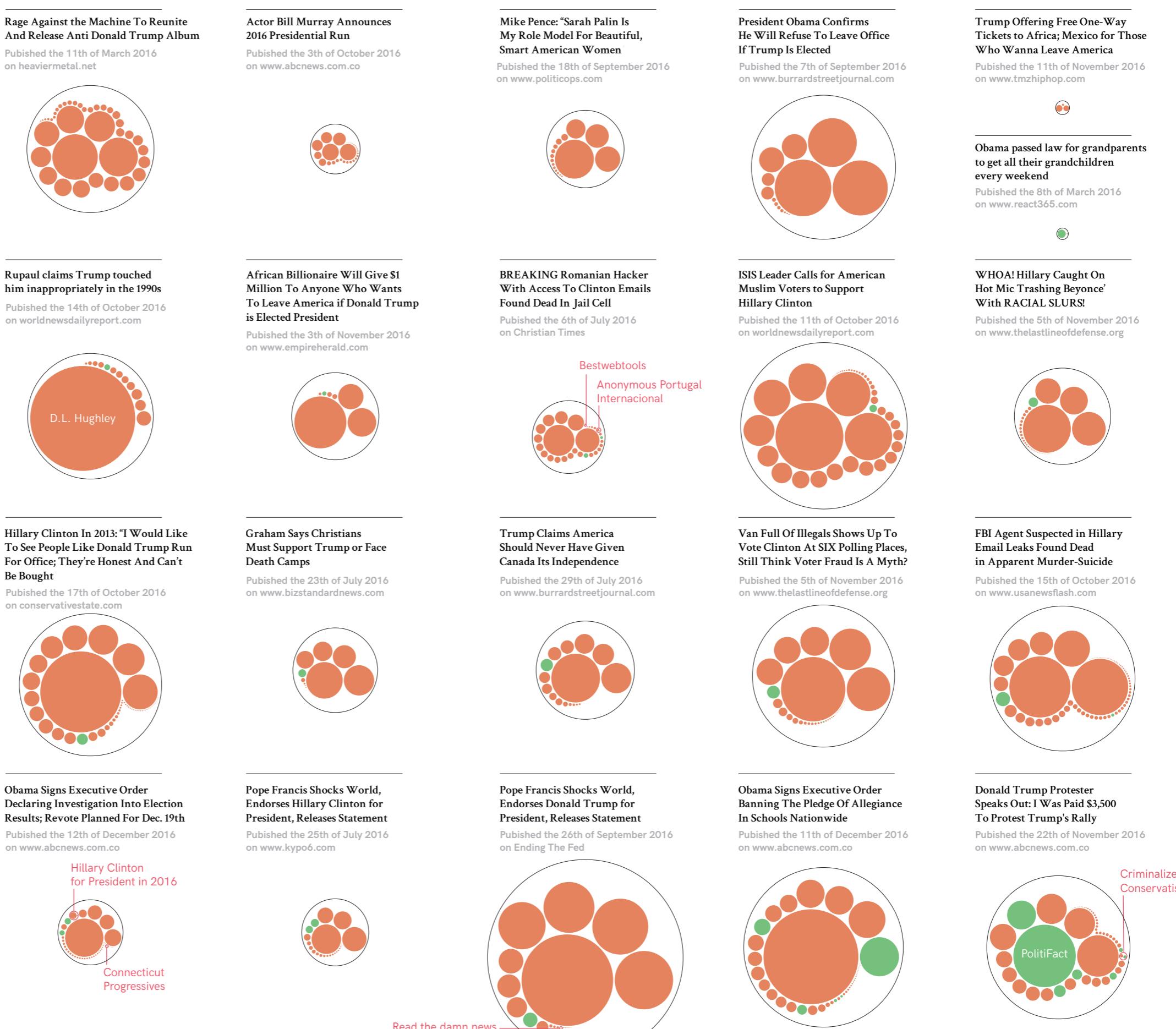
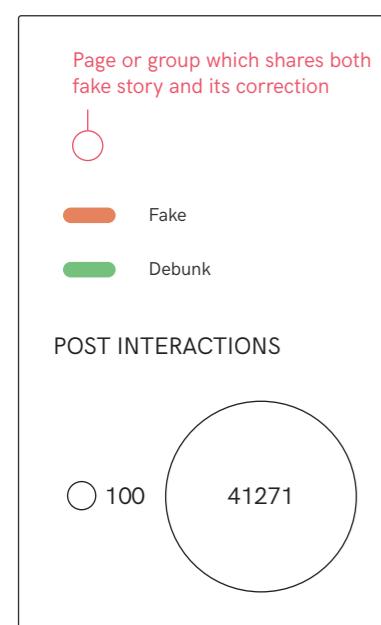
### b. MAP THE OVERLAP BETWEEN THE PUBLICS OF FAKE NEWS STORIES AND WEB PAGES WHICH AIM TO DEBUNK THEM

Facebook pages and groups that 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 Facebook pages and groups that share fake news stories and those debunking web pages issued in responses.
- ◊ This may be illustrated by means of a “circle packing” visualisation. You may use → RawGraph for this operation.

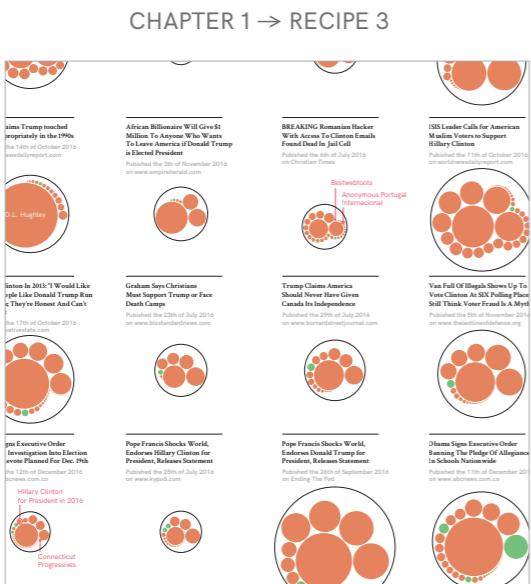
# ARE DEBUNKING WEB PAGES ACKNOWLEDGED BY THE PUBLICS OF FAKE NEWS?

**Fake news pages and debunking web pages have different publics on Facebook.** Only six of the pages that share fake news stories have acknowledged web pages which aim to debunk them. 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.



## Chapter #2

# TRACING THE CIRCULATION OF FAKE NEWS ON THE WEB



### 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.

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

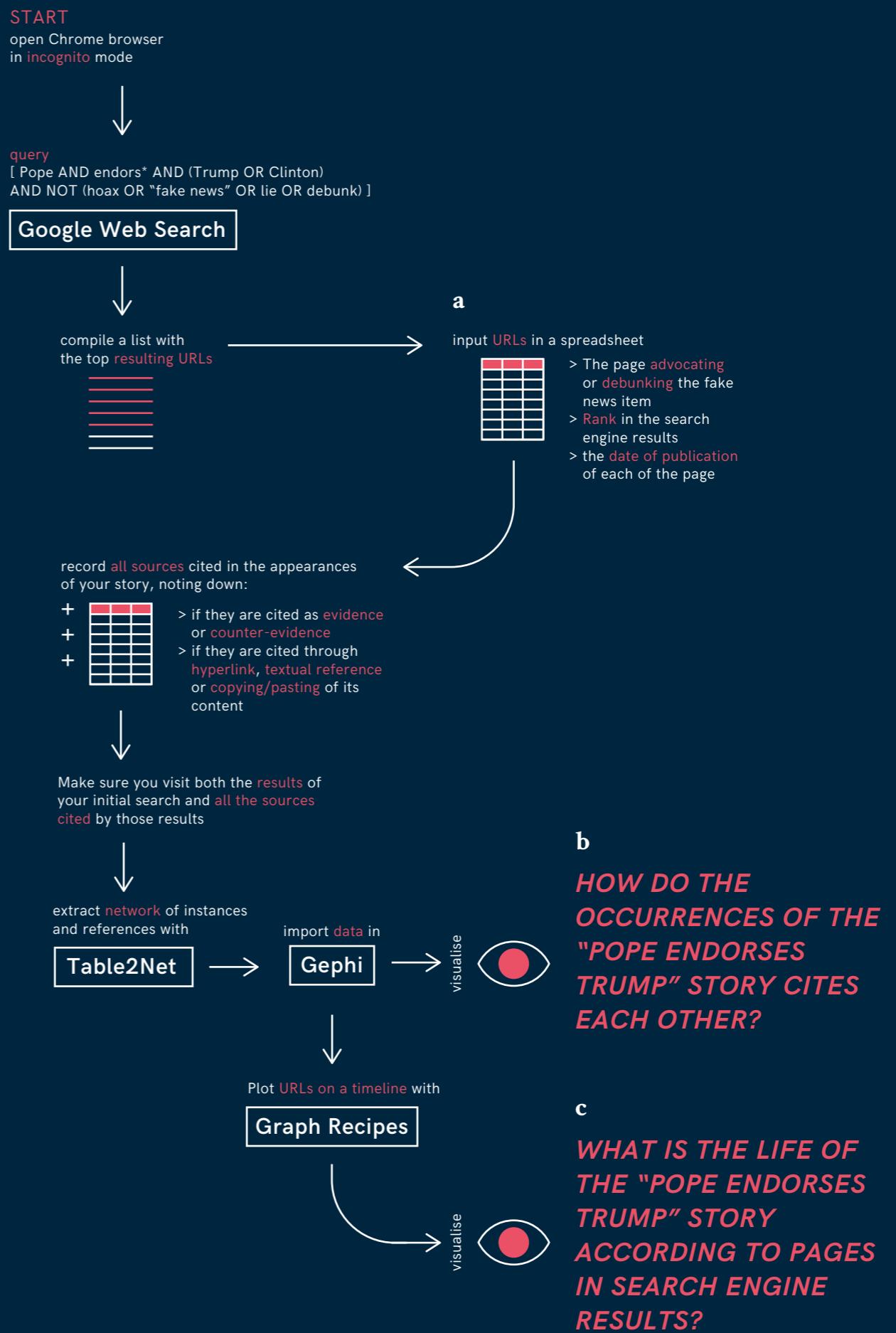
contribute to making them even more visible on the web. A better understanding of how fake news travels online can help to inform responses that are more attuned to the phenomena. From a methodological point of view, as there is no “ontological” difference between fake and authentic news, studying fake news circulation can help us understand more about how other kinds of news travels.

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.



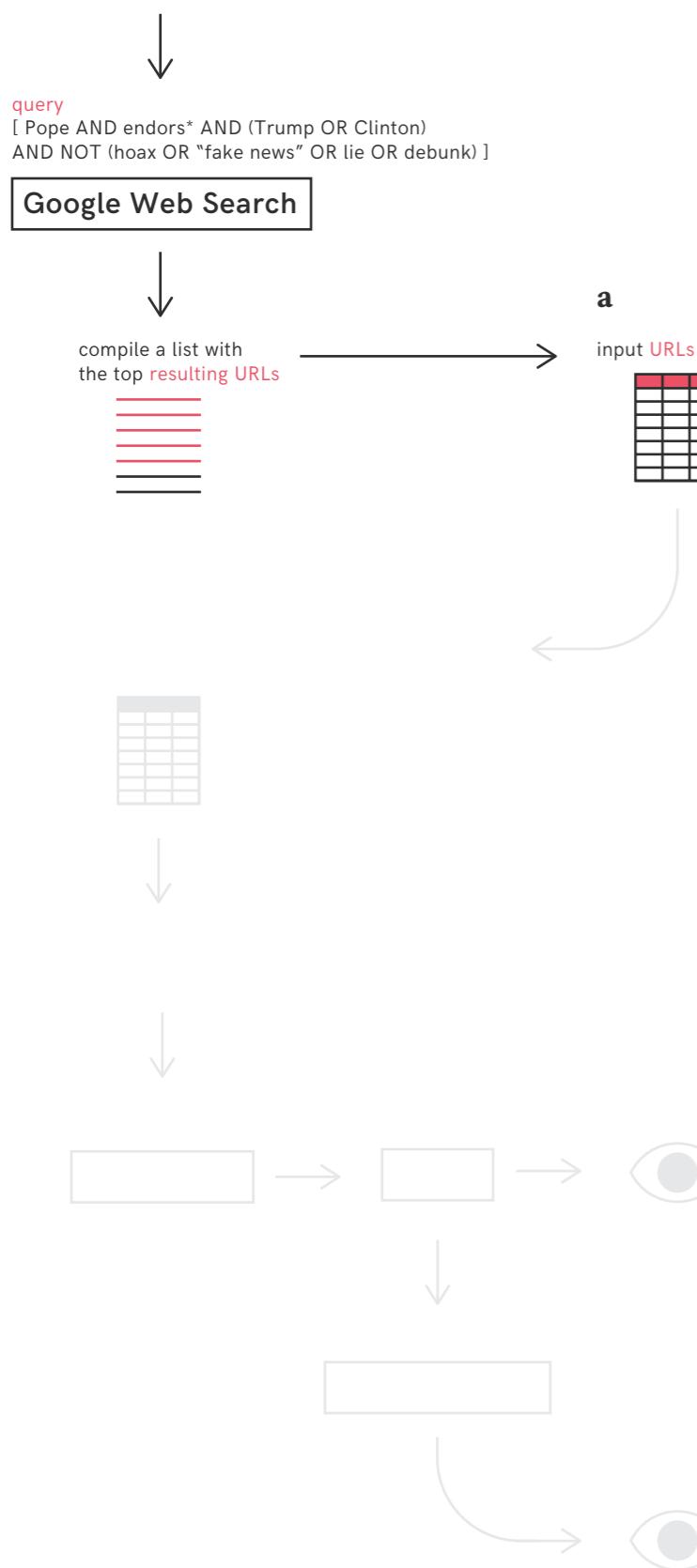
## **BEFORE STARTING**

For this recipe you will need to choose a fake news story whose circulation you'd 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



## CHAPTER 2 → RECIPE 1

### 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.<sup>[1]</sup>
- ◊ 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=bj65Xr9GkJM>

## 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) ]

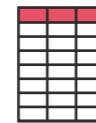
### Google Web Search



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

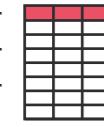
**a**

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 of the page

record **all sources** cited in the appearances  
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

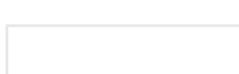
**Table2Net**

→ **Gephi**

import **data** in

**b**

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



CHAPTER 2 → RECIPE 1

## 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.

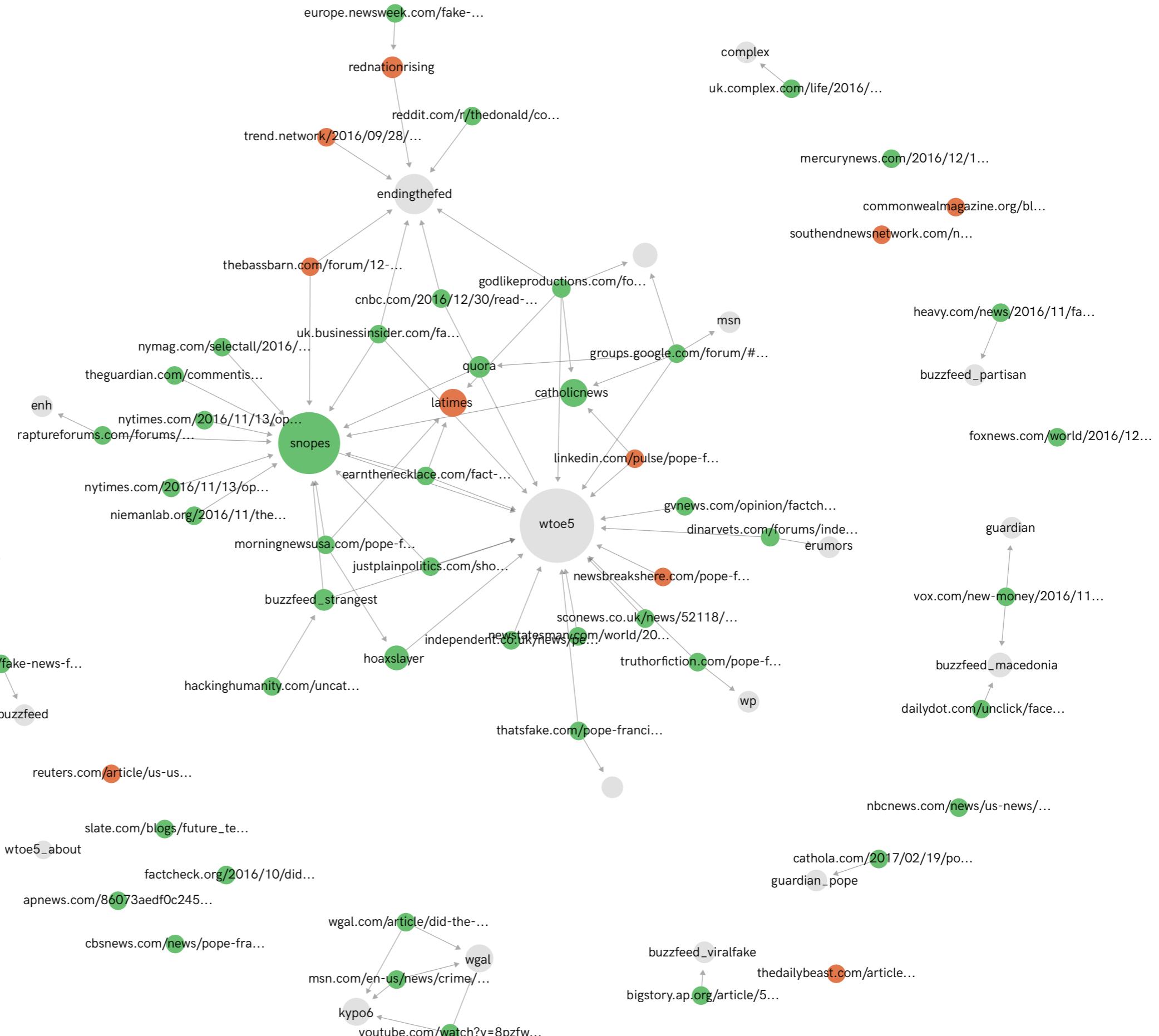
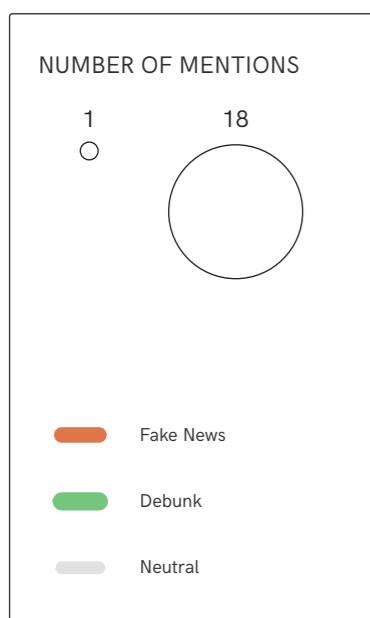
In this step, we will trace the network of references of a specific fake news story.

- ◊ 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 CITES 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 cites or debunks the story) and their size (which indicates the number of citations received by the page) reveals the greater 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 of the page

record **all sources** cited in the appearances  
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**

visualise

**b**

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

Plot **URLs** on a timeline with

**Graph Recipes**

visualise

**c**

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

CHAPTER 2 → RECIPE 1

### c. VISUALIZE THE FAKE NEWS INSTANCES OVER TIME

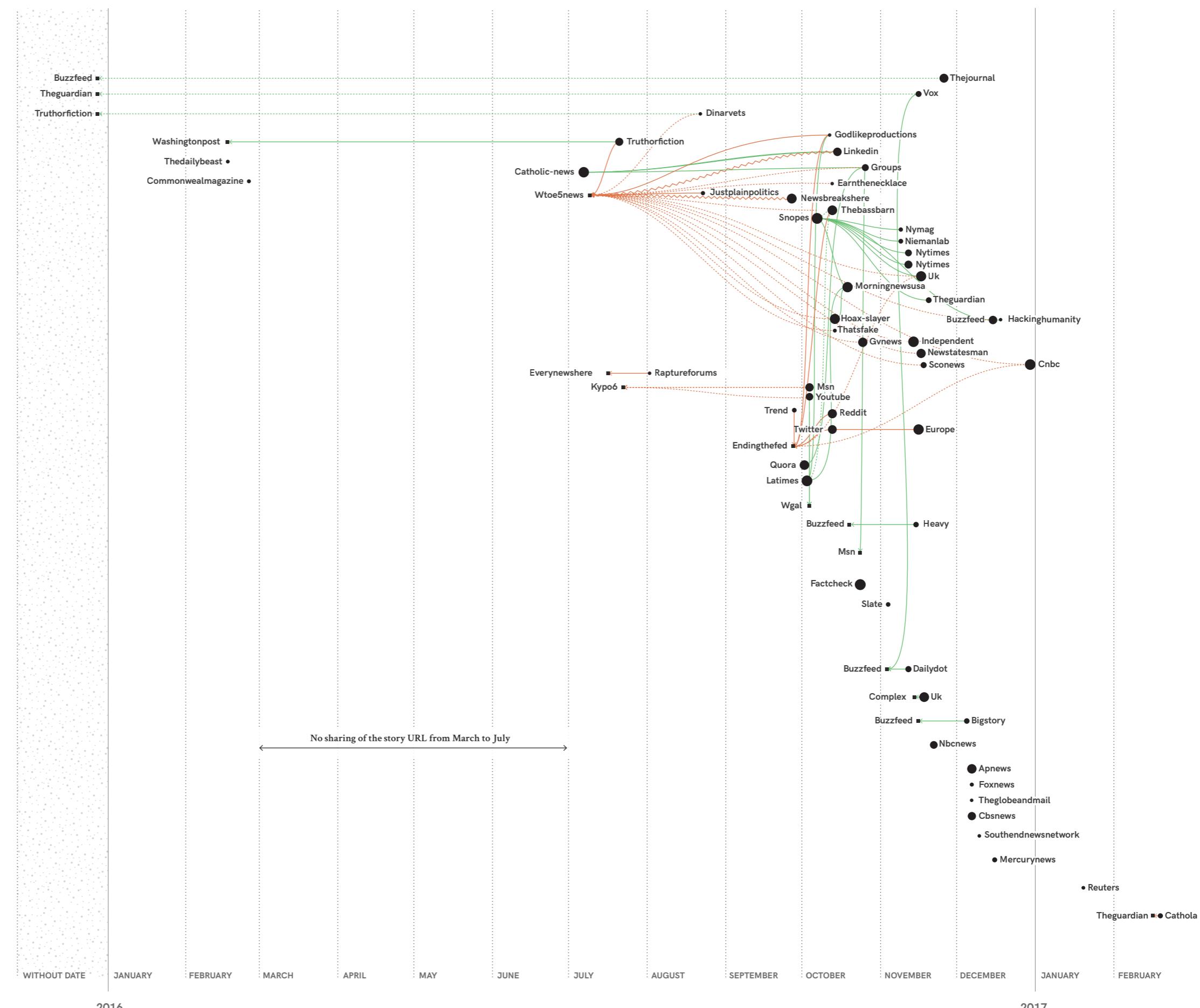
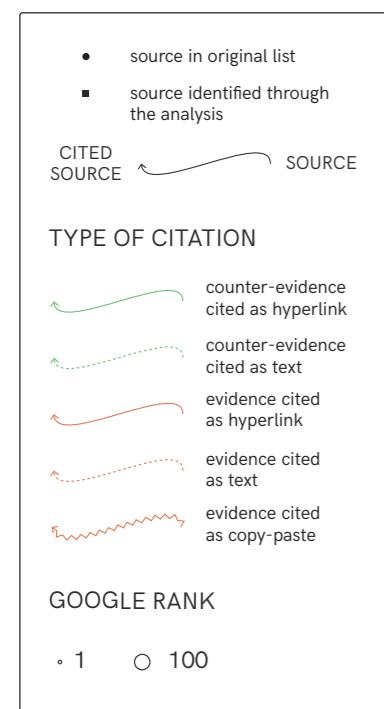
The network extracted in the previous step can help you understanding not only who cited whom, but also how and in which direction your fake news travelled. To reveal the circulation use the dates that you collected in the first step of this recipe.

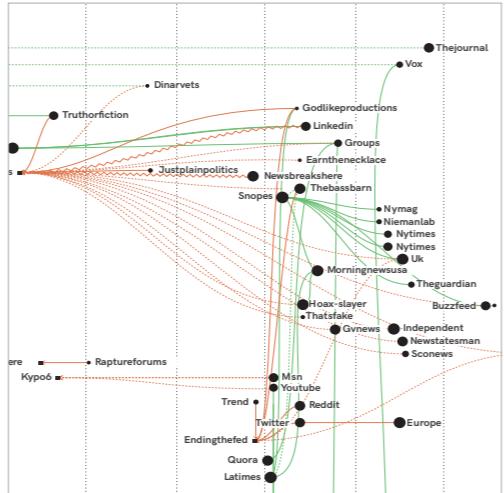
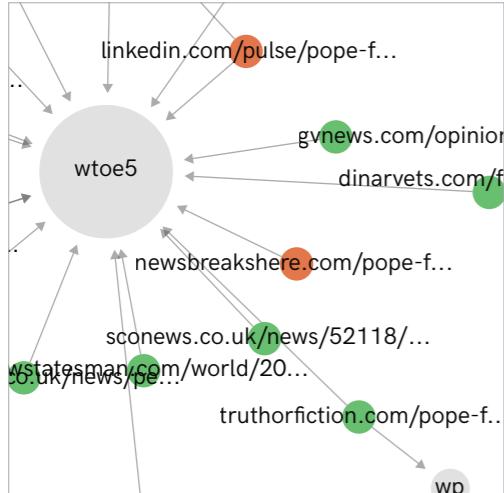
- ◊ Arrange the network of instances extracted in the previous step chronologically. You can use different visual styles to represent the different kind of citations (to do so, we used a custom script of → **Graph Recipes** tools).

# **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"

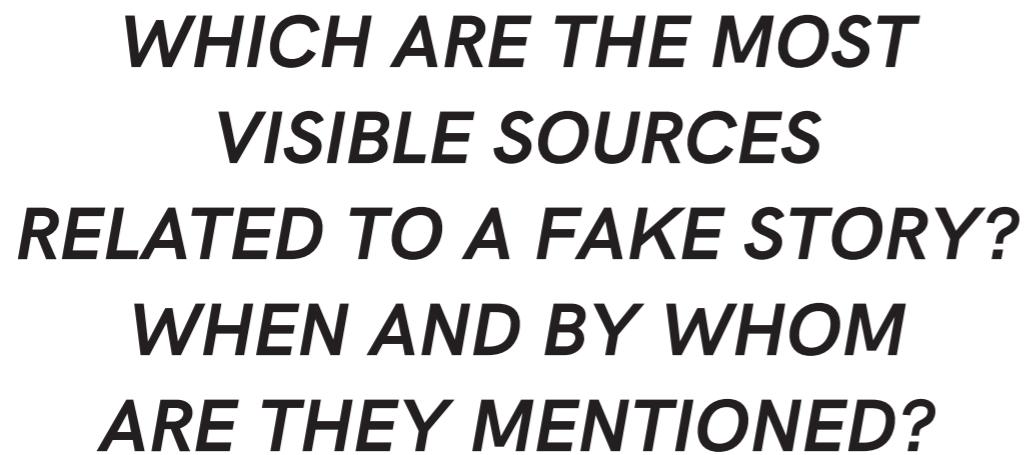
**Trump” story.** In this image it is not the nodes, but their connections that are highlighted: their colour indicates whether a source is cited as evidence or counter-evidence of the story and the style the type of citation. The high presence of orange dotted lines (counter-evidence textual citations) 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

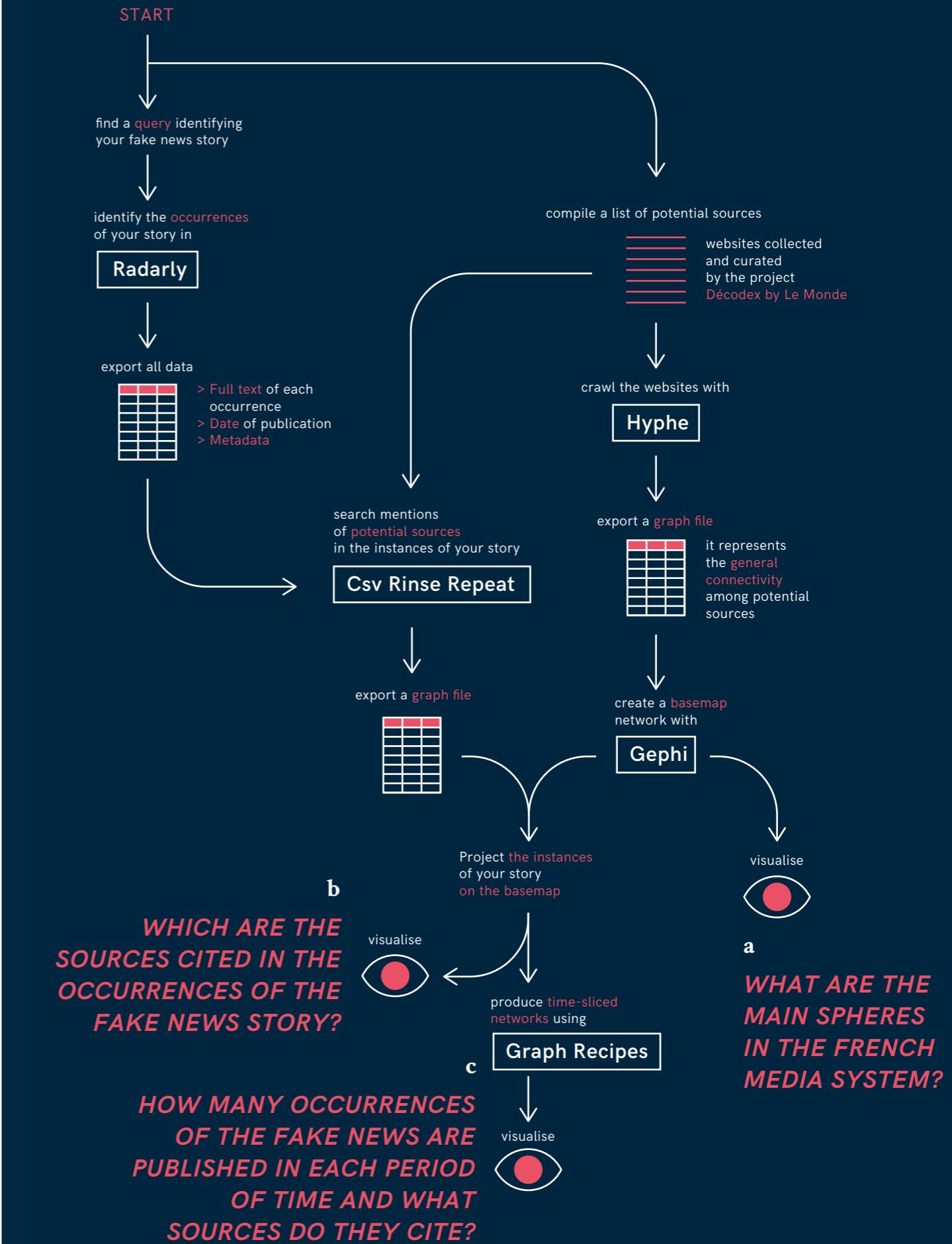
This recipe may be used to repurpose data obtained through search engine results in order to identify and follow the different occurrences of a given fake news story as it is cited and referenced by different online sources, as well as to retrace its circulation over time. You will also be able to see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

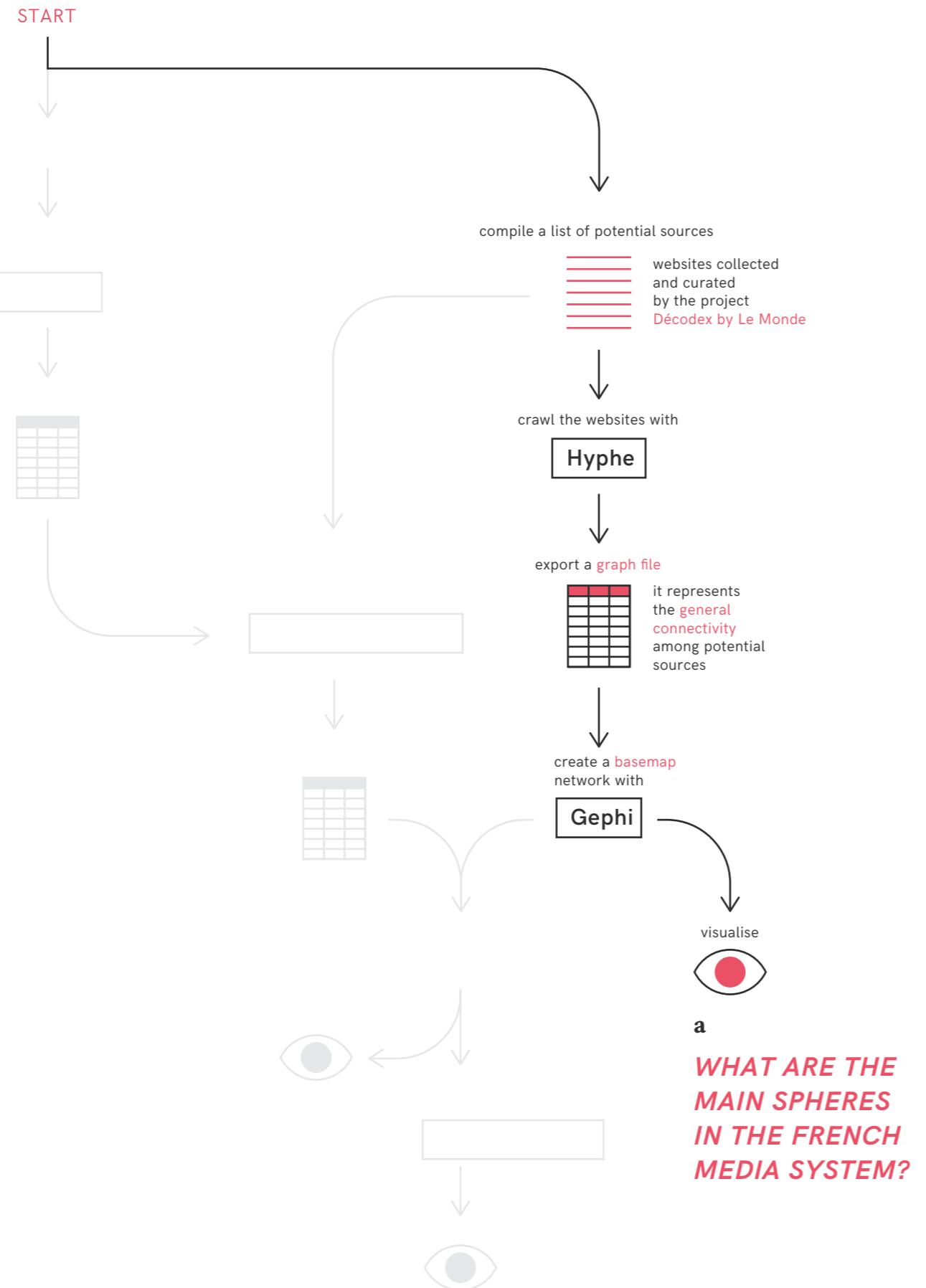


## **BEFORE STARTING**

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**).





CHAPTER 2 → RECIPE 2

### a DEFINE A BASE MAP OF NEWS PROVIDERS

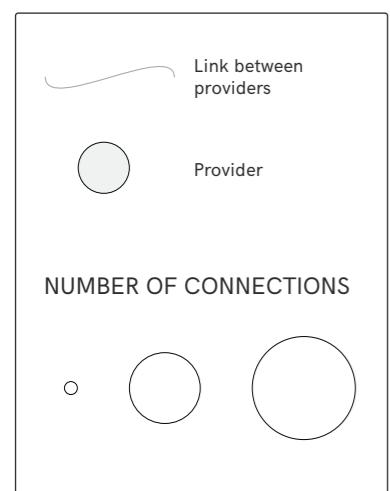
Identify (or compile) a list of all the possible Web sources in which you think your fake news item might have appeared (try to be as exhaustive as possible). You can use one of the many lists of fake news websites maintained by debunking initiatives and combine it with a list of mainstream media outlets.

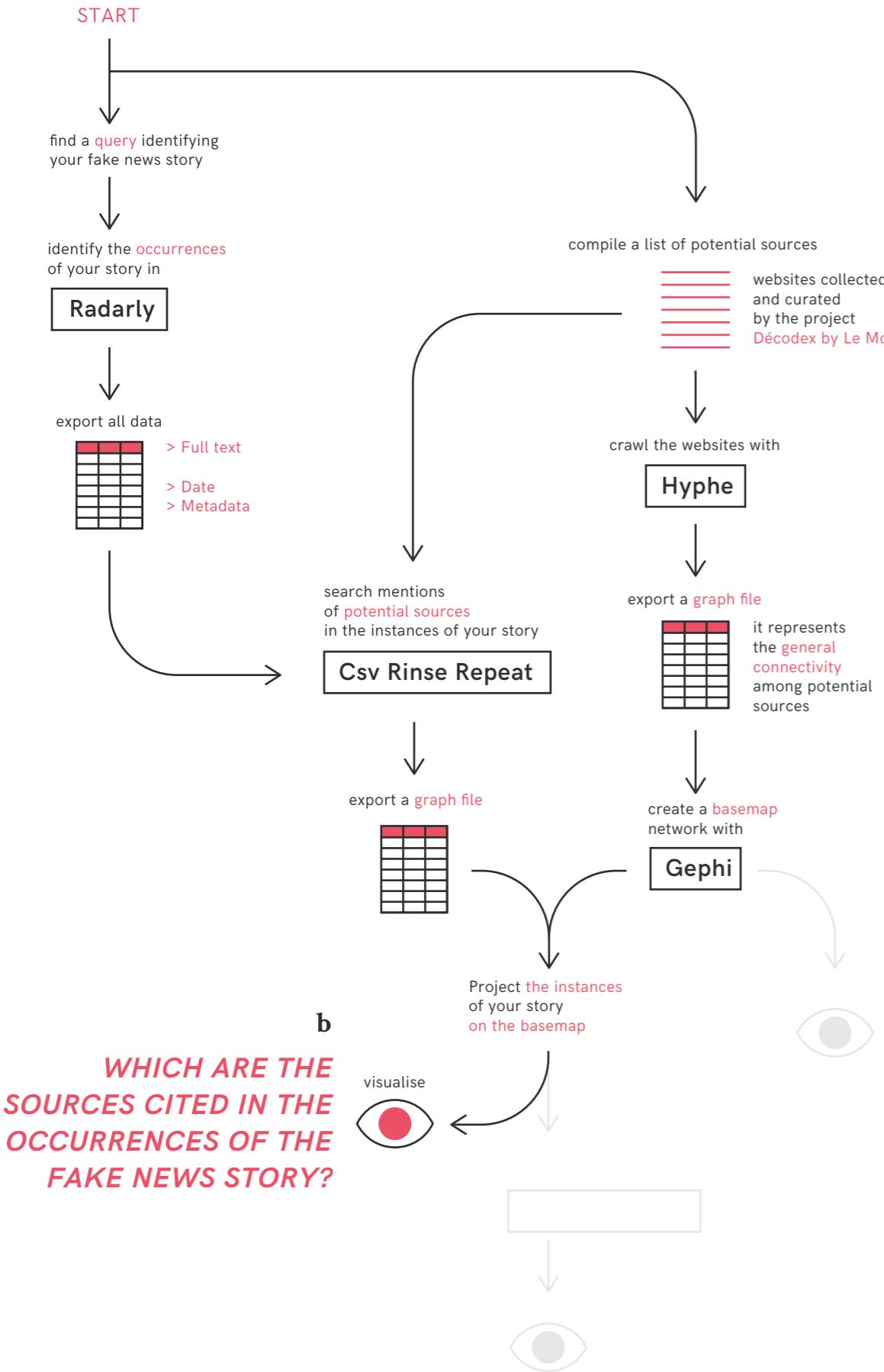
- ◊ Identify how the sources in your list are associated with each other through web crawling and hyperlink analysis. We used → **Hyphe** for this.
- ◊ Visualise the resulting network and apply a force-directed layout algorithm to identify clusters of sources. You can use → **Gephi** for this task.
-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





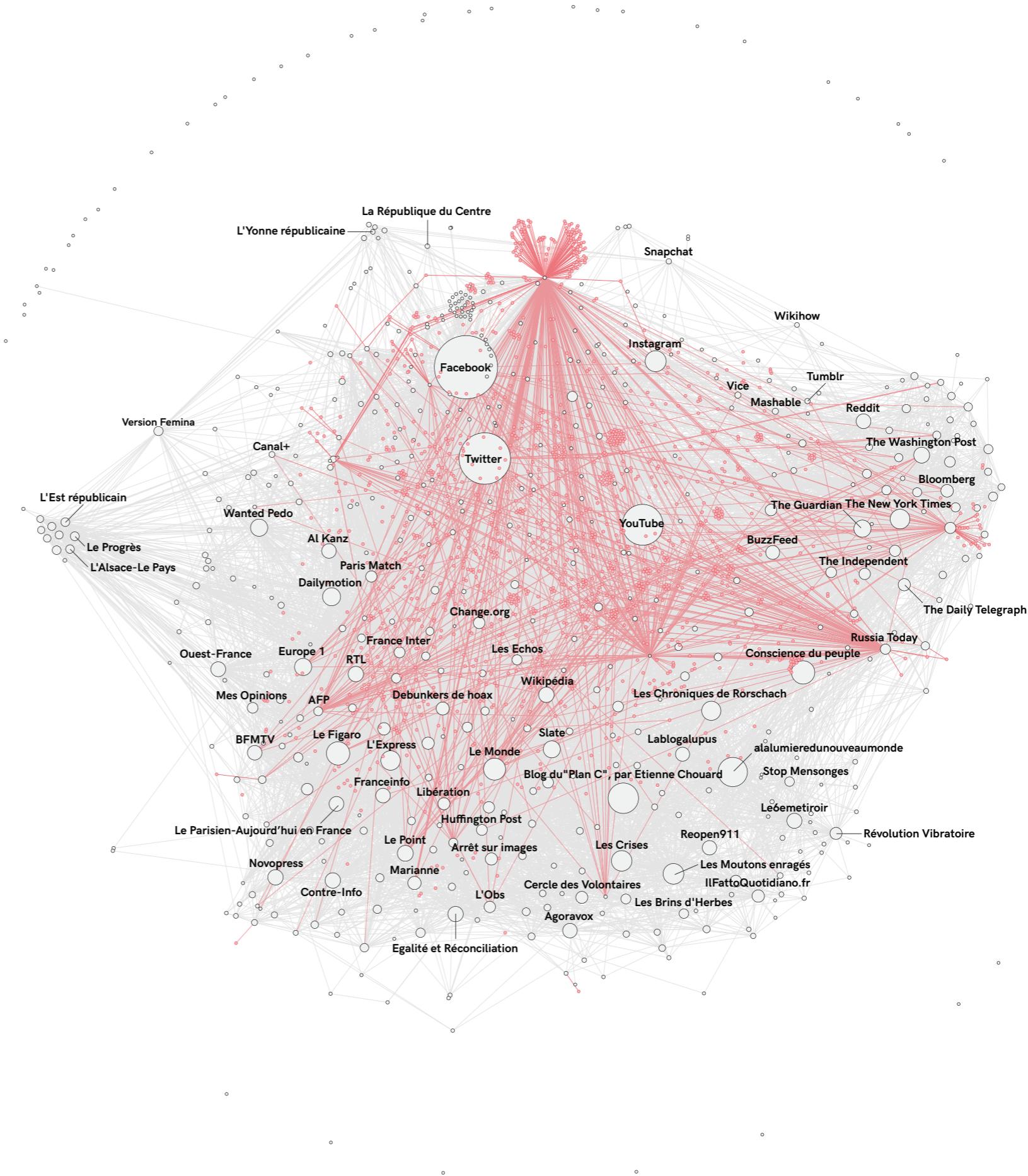
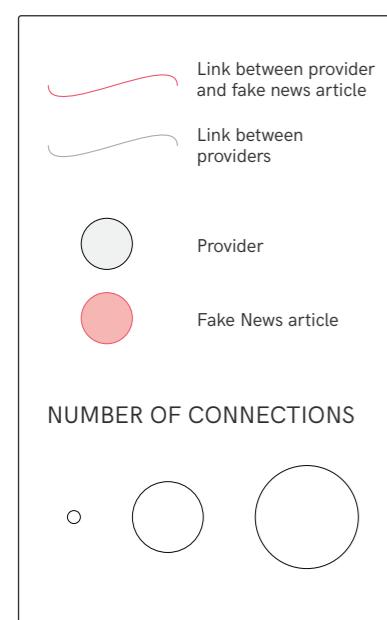
## b. HIGHLIGHT THE OCCURRENCES OF YOUR STORY ON THE BASE MAP

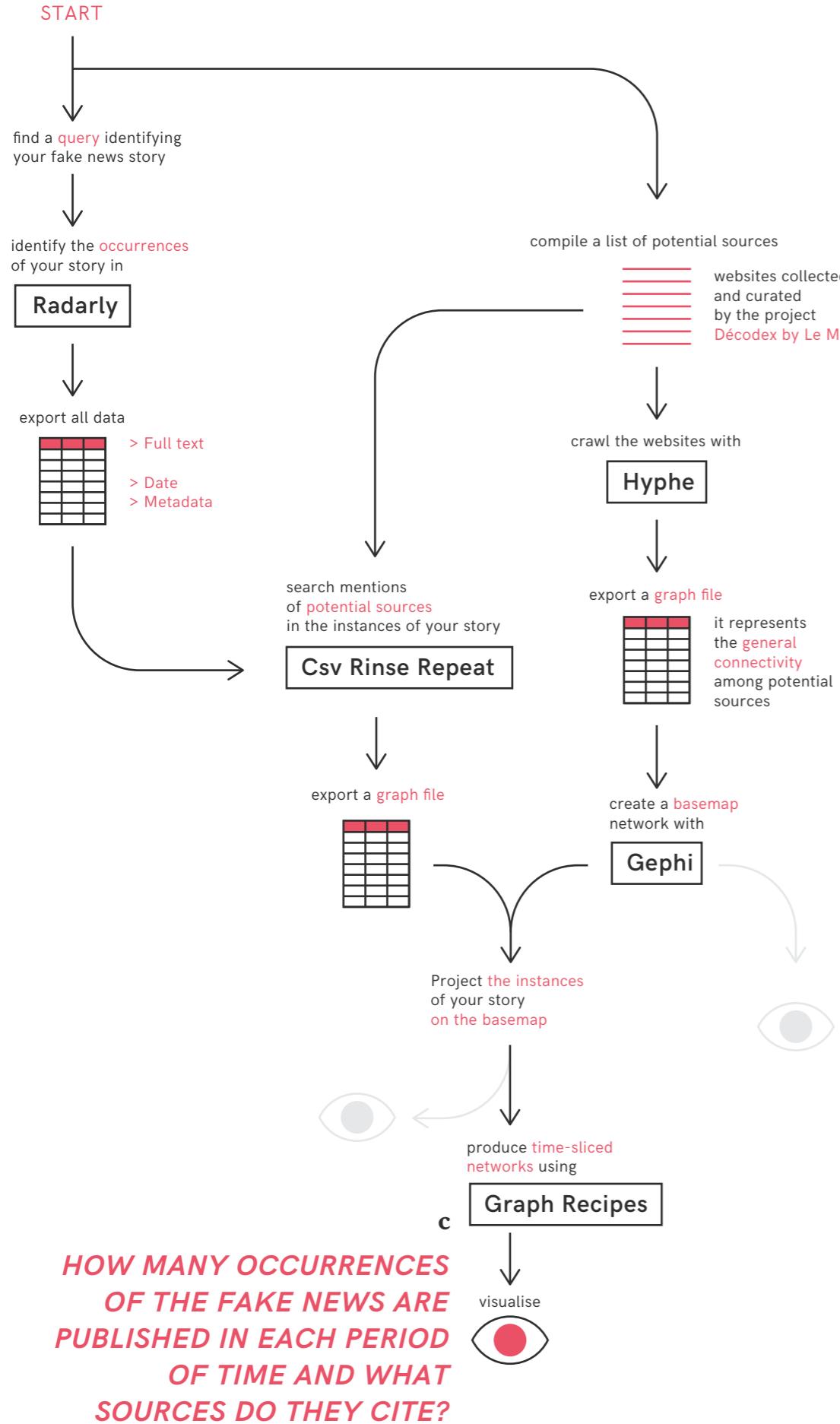
In this step we will explore how fake news stories are associated with different sources. This is interesting, as while a fake news story might – for example – start out its life as a piece of satire, as it travels it can become more prominently associated with non-satirical sources. Here we will identify 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 mentions of the sources of your base map, searching for the 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 merging them with the source network and 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.





CHAPTER 2 → RECIPE 2

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

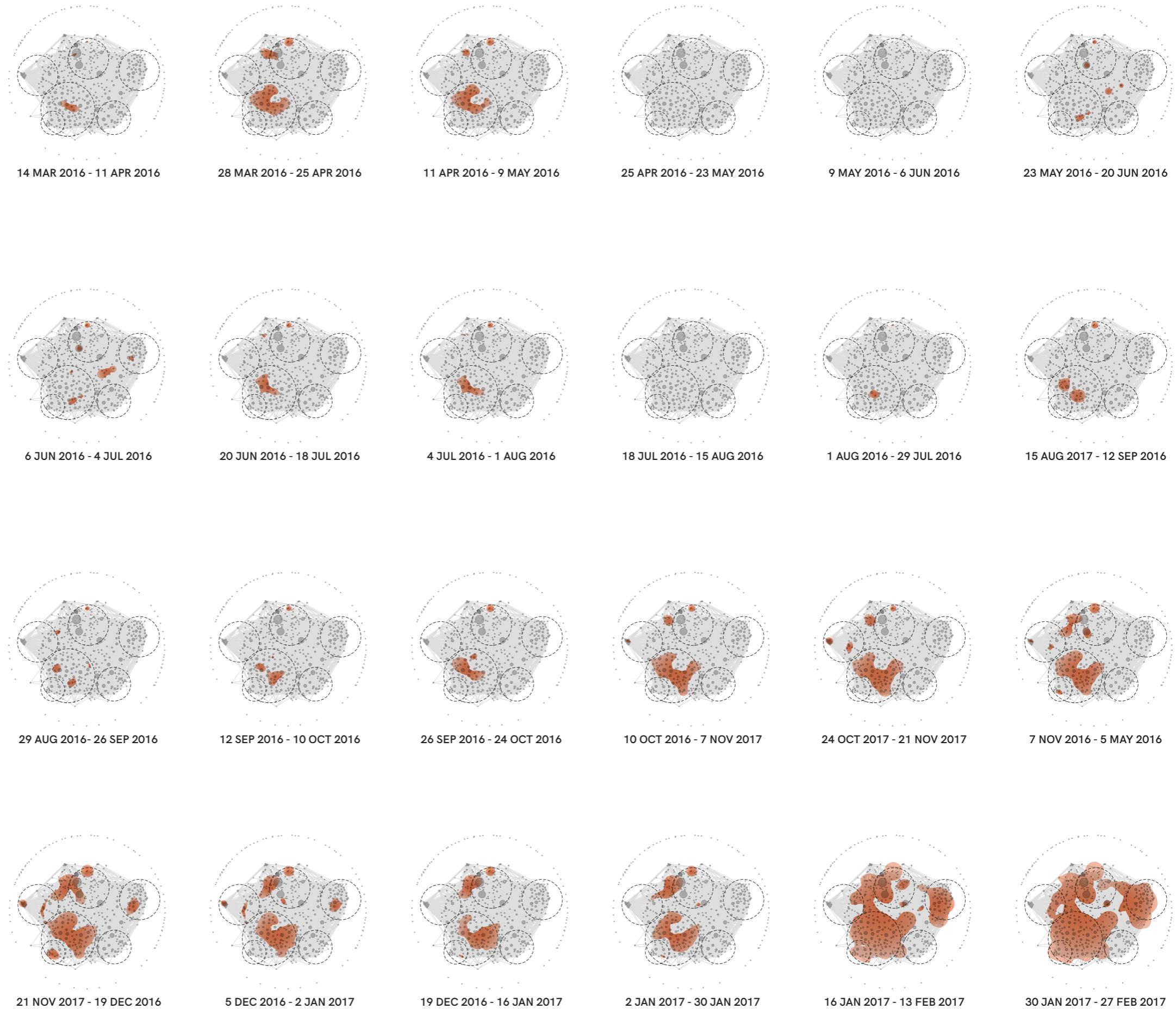
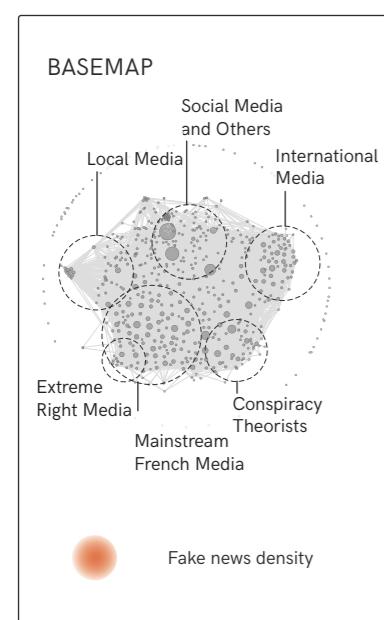
In this step, you will reveal how the reference patterns identified in the previous step evolve over time.

- ◊ 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.

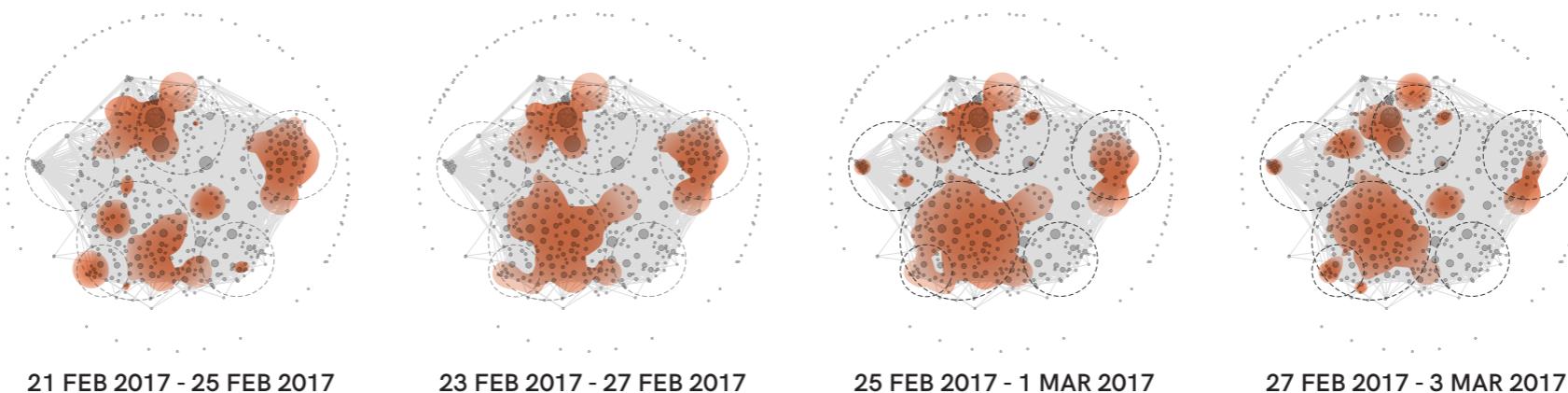
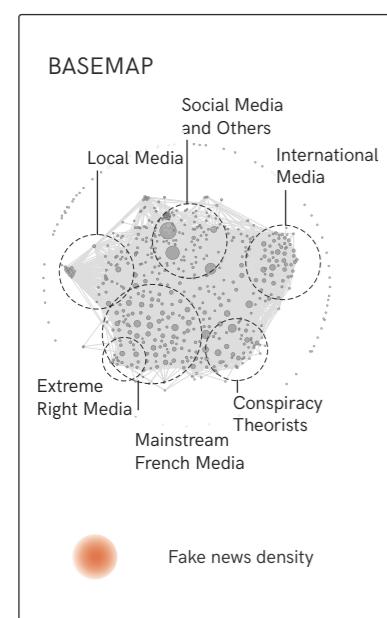
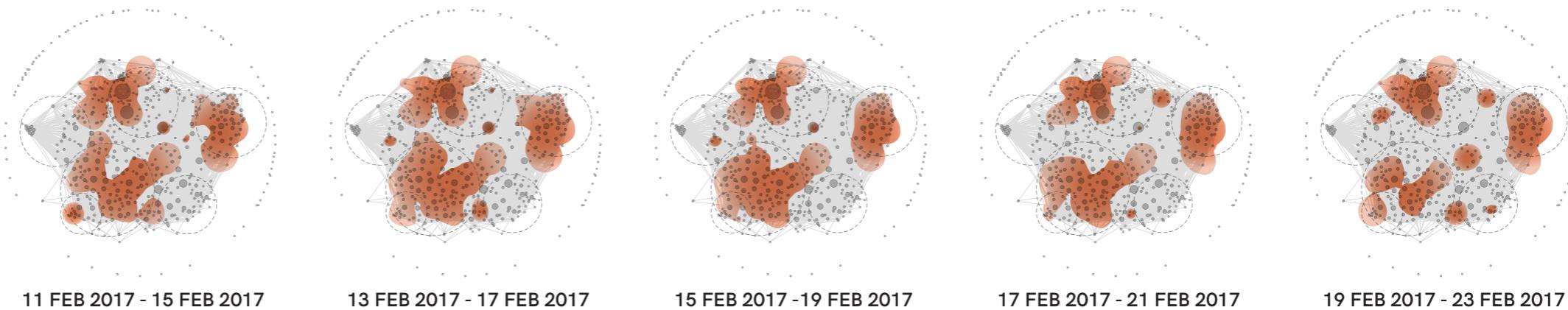
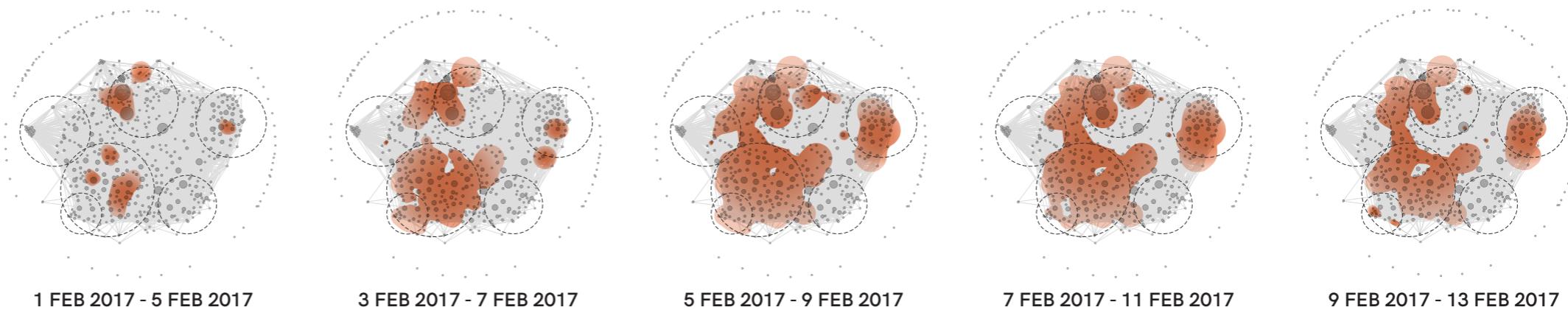
In this image, the occurrences of the fake news story are divided in slices of 4 weeks (with an overlap of two weeks) and represented as a density heat map rather than as individual points. Though mentions of the story have been present for more than one year, its circulation appears to spike up in February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

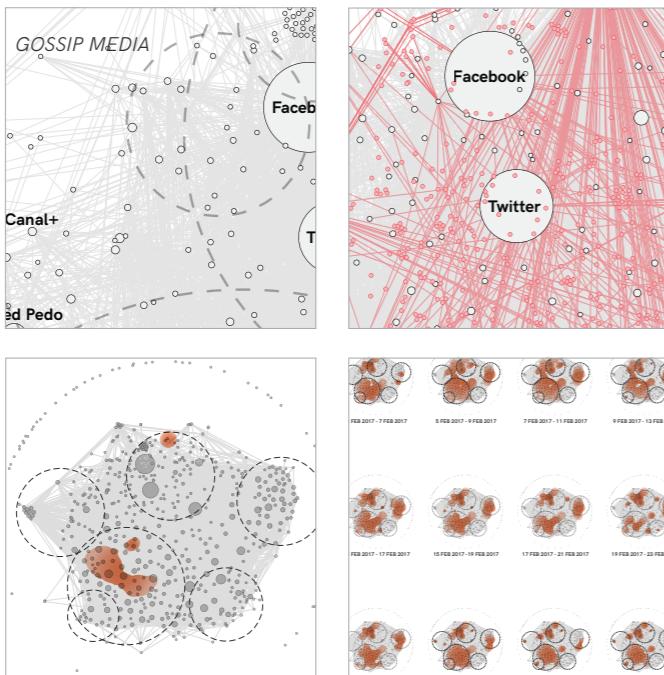
This image represents a 'temporal zoom' of the previous one. Here the occurrences of the fake news story are break up in slices of 4 days (with an overlap of two days).



## Chapter #3

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

CHAPTER 2 → RECIPE 2



### SERVING SUGGESTIONS

This recipe may be used to identify which websites appear most often when a fake news story is mentioned in different settings. 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).

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 misinformation sites vary across language spheres?



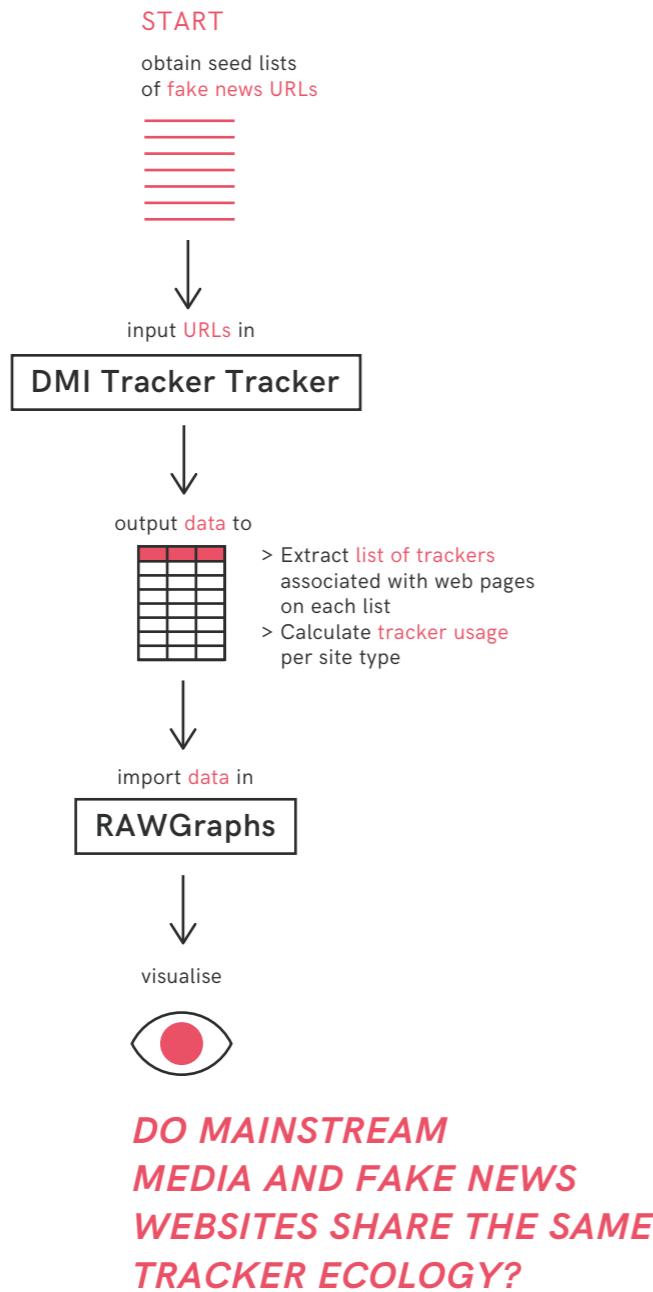
**Introduction** - Over the past few decades many responses to misinformation have focused on mapping and debunking claims made or repeated by politicians, journalists or other public figures. What are the prospects of mapping fake news online not just by looking at the circulation of claims, but by 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 misinformation 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.



CHAPTER 3 → RECIPE 1

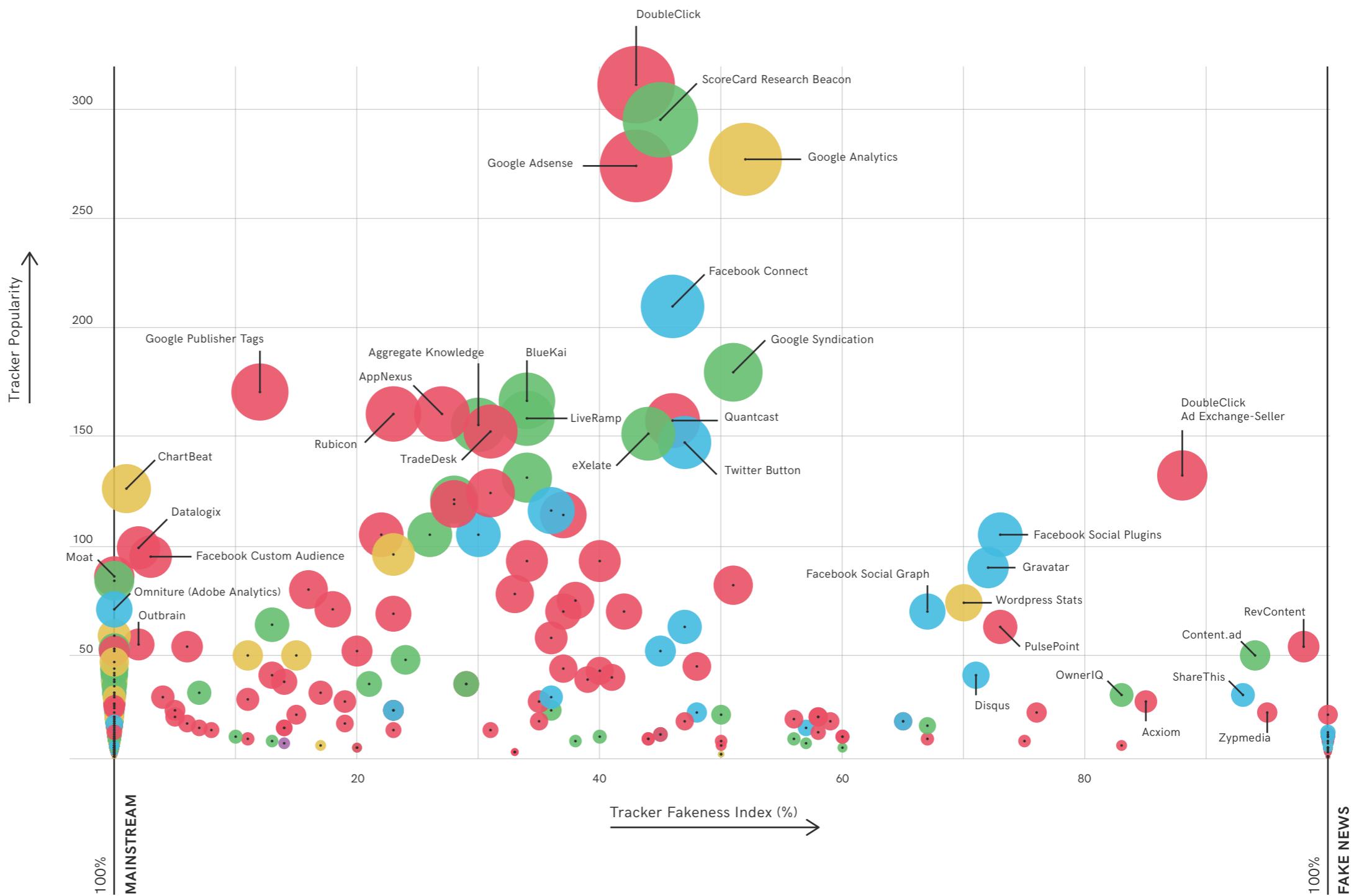
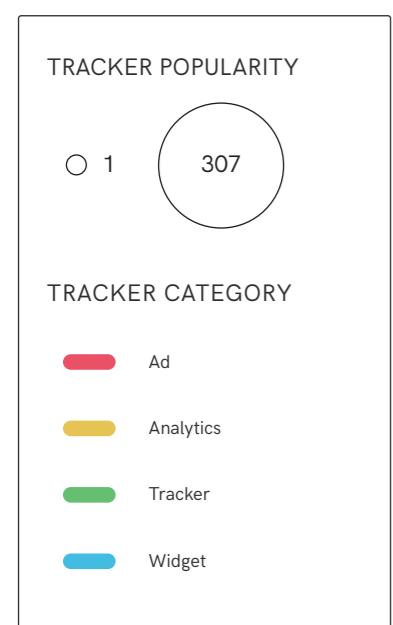
## CALCULATE TRACKER USAGE PER SITE TYPE

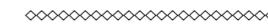
From the source code of web pages it is often possible to see which third-party tracking services are used.

- ◊ Collect data about trackers associated with the web pages on each list. You may use the → **DMI Tracker 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?

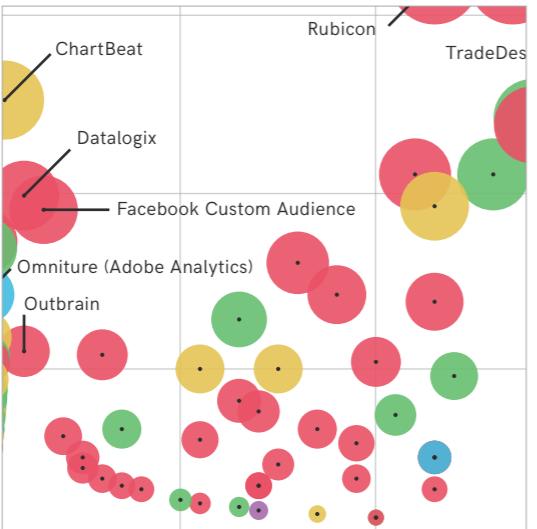
**Scatterplot representing tracker usage on a series of fake news and mainstream media sites.** While fake news sites and mainstream media sites share popular tracker services such as Google Adsense, DoubleClick and Google Analytics, mainstream media sites appears more mature and sophisticated in its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





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

CHAPTER 3 → RECIPE 1

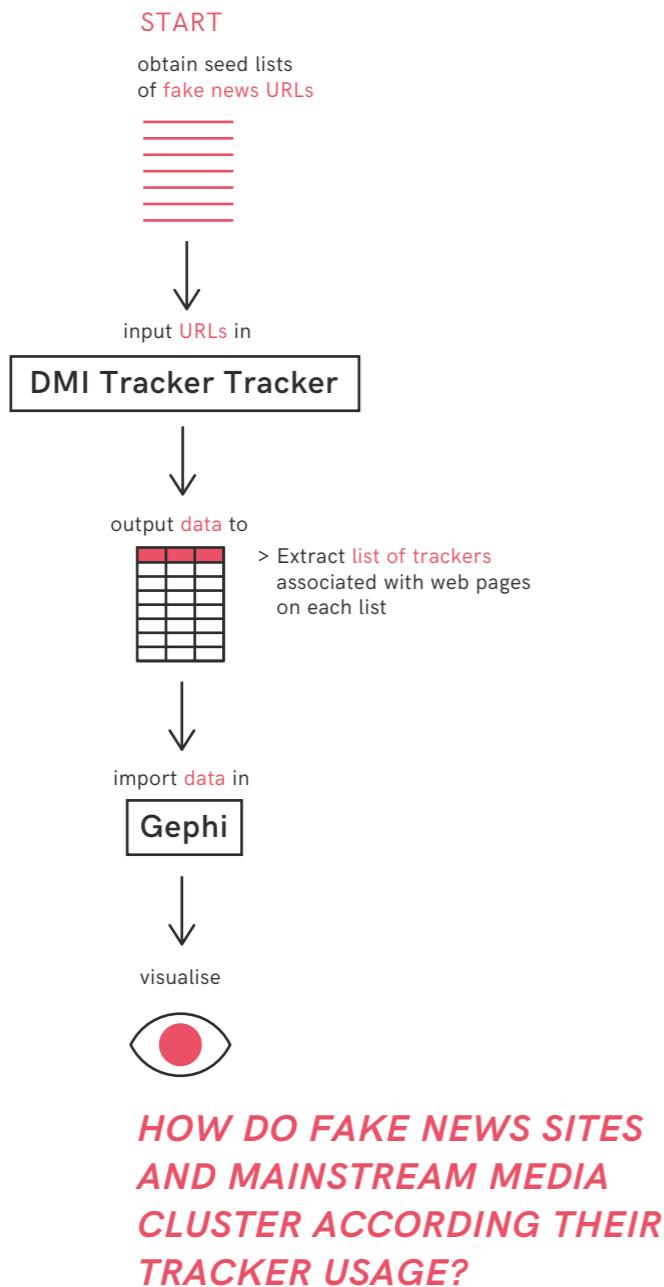


## SERVING SUGGESTIONS

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.

## 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.



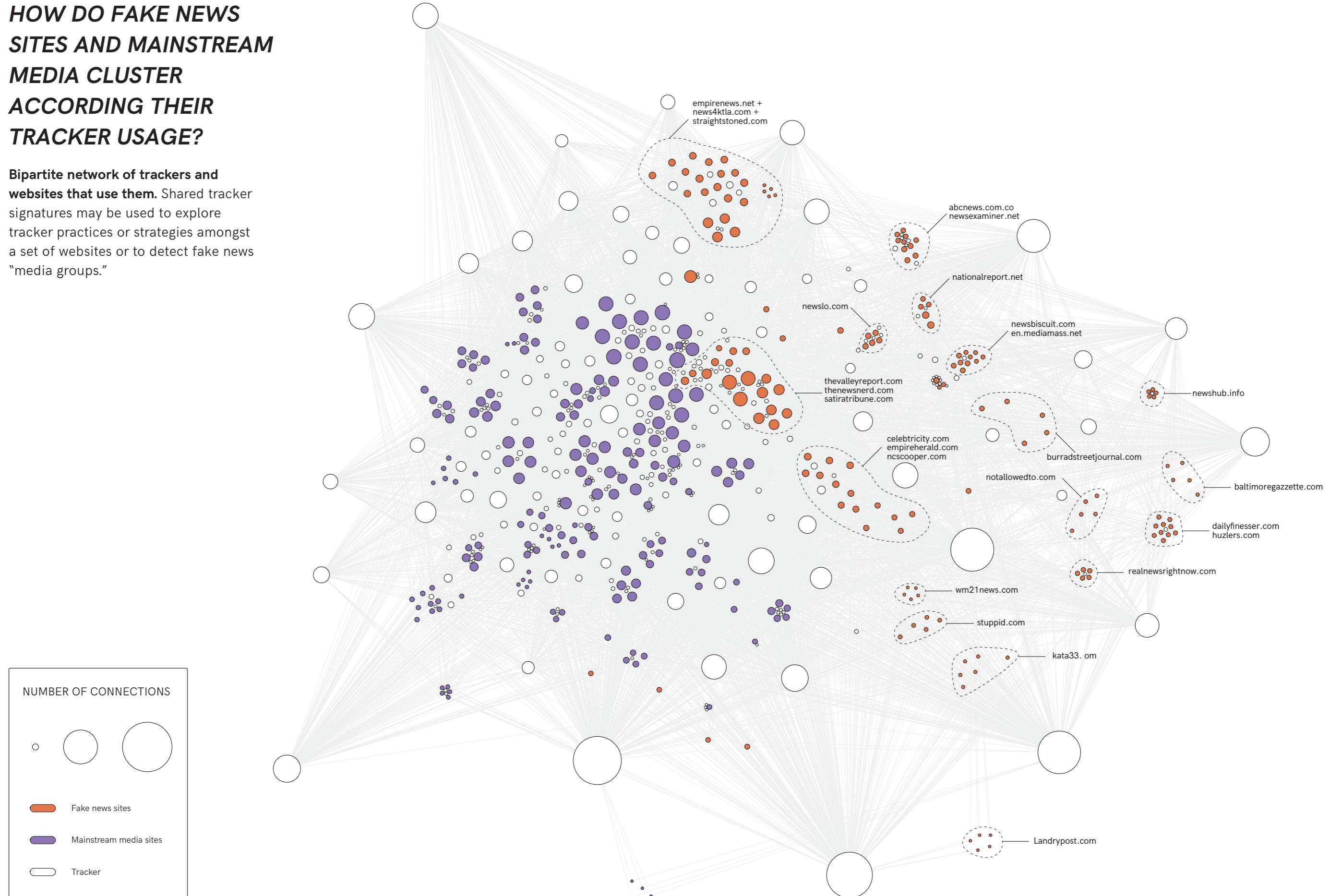
## GRAPH RELATIONS BETWEEN PAGES AND TRACKERS

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 cluster 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.”





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

CHAPTER 3 → RECIPE 2

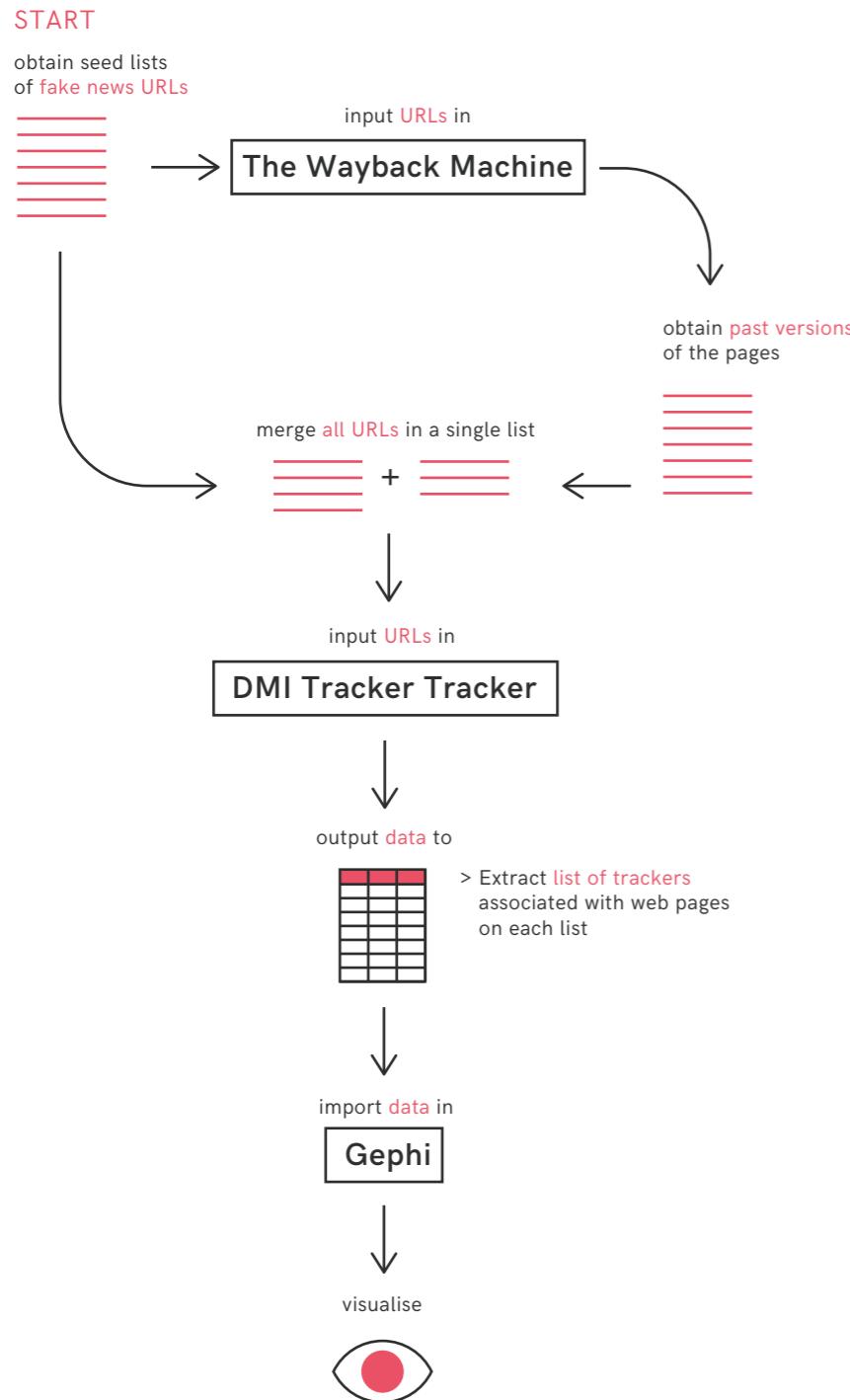


## SERVING SUGGESTIONS

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.

## 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?**

CHAPTER 3 → RECIPE 3

## 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.

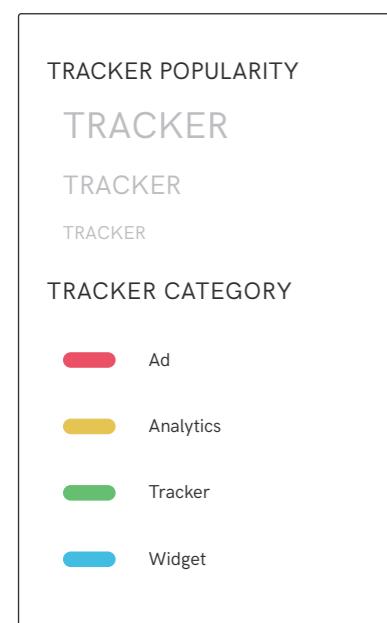
- ◊ Obtain archived copies of a webpage. You may use → **The Wayback Machine** to see how a given page changed over time.
- ◊ Identify associated trackers with the current and previous version of the page. You may use the → **DMI Tracker Tracker** tool to collect such information.
- ◊ Identify the trackers 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).



Admarvel  
Clicksor  
Drawbridge  
Facebook Custom Audience  
Google Publisher Tags  
Gumgum  
Media.net  
Sekindo  
Doubleverify  
Visible Measures  
Omniture (Adobe Analytics)

TRACKERS PRESENT ONLY BEFORE DECEMBER

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  
Acloudimages  
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

TRACKERS ALWAYS PRESENT

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 PRESENT ONLY AFTER JANUARY



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

CHAPTER 3 → RECIPE 3

Google Analy  
Google Adser  
Doubleclick Ad  
Google Syndicati  
Scorecard Researc  
Gravatar

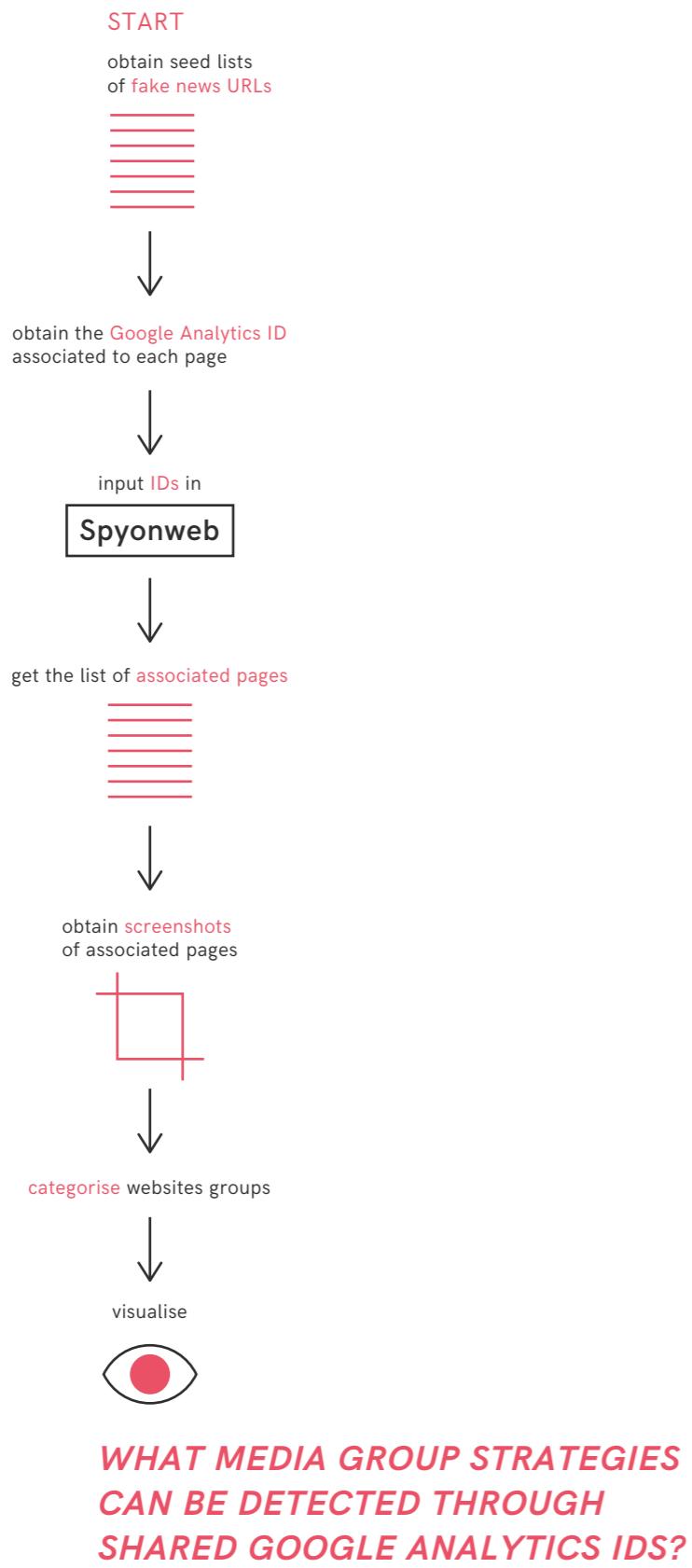
### SERVING SUGGESTIONS

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.

### BEFORE STARTING

Before you start you will need to compile or identify seed lists of fake news or other misinformation 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.



CHAPTER 3 → RECIPE 4

## 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-xxxxxx”) 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.** 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 online misinformation.

rt.com

19 disinformation stories  
UA-5773642



Media group (Russia Today)

xryshaygh.com

1 disinformation story  
UA-4839940



Lone Webmaster

defensenews.com

1 disinformation story  
UA-841082



newcoldwar.org

7 disinformation stories  
UA-15942468



almanach.cz

3 disinformation stories  
UA-3004323



Themed network (Mysticism, Liberland)



assange.rt.com



asco.gr



doc.rt.com



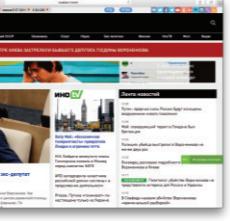
royalparadise.gr



learnrussian.rt.com



discoverthassos.com



russian.rt.com



catalog.rt.com



armedforcesjournal.com



oakvillendp.ca



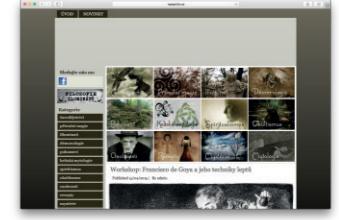
grand-mystical-lodge.com



sightlinemediagroup.com



socialiststudies.com



malachim.cz



militarytimes.com



ndpsocialists.ca



liberlandpress.com



airforcetimes.com



ccu-csc.ca



illuminati-journal.com



marinecorpstimes.com



oldm.cz



# DO TRACKERS ASSOCIATED WITH HYPER- PARTISAN AND MISINFORMATION SITES VARY ACROSS LANGUAGE SPHERES?

CHAPTER 3 → RECIPE 4



asco.gr



royalparadise.gr

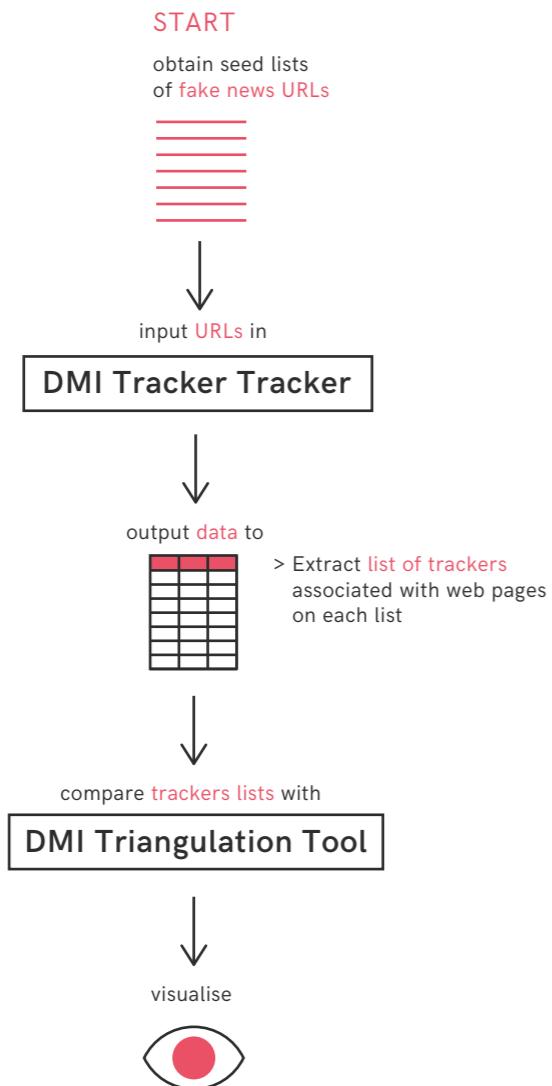
## SERVING SUGGESTIONS

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

## BEFORE STARTING

For this recipe, you will need lists of fake news, hyper-partisan or misinformation 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 misinformation sites in Dutch, English and German language spheres.



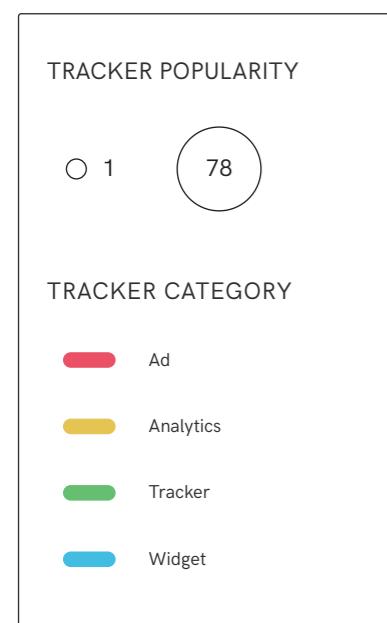
**DO MISINFORMATION AND  
HYPER-PARTISAN WEBSITES  
IN DIFFERENT LANGUAGE  
SPHERES HAVE DISTINCT  
TRACKER ECOLOGIES?**

## IDENTIFY TRACKERS PER LANGUAGE SPHERE

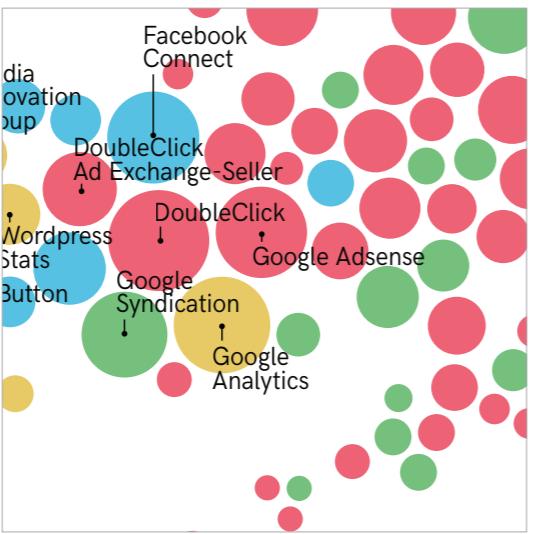
- ◊ Extract trackers associated with lists of the web pages for each language sphere. We did this using the → **DMI Tracker Tracker** tool.
- ◊ Identify the trackers which are shared across and which are unique to different languages spheres within the dataset. We did this using the → **DMI Triangulation** 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 MISINFORMATION AND HYPER-PARTISAN WEBSITES IN DIFFERENT LANGUAGE SPHERES HAVE DISTINCT TRACKER ECOLOGIES?

Visualisation of tracker ecologies associated with hyper-partisan or misinformation 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.

