

# RECON

A graph-based evidence synthesis approach for outbreak detection

---

Thibaut Jombart

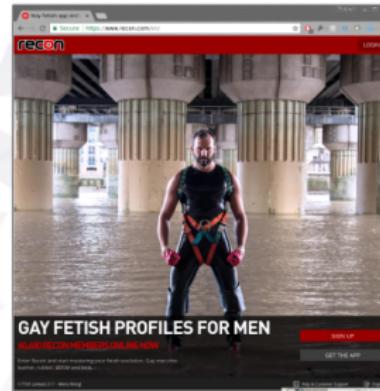
20th September 2017

Imperial College London  
MRC Centre for Outbreak Analysis and Modelling

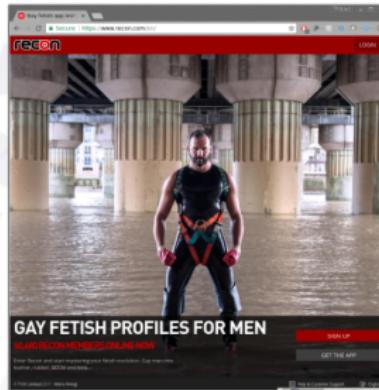
# Topics of the day

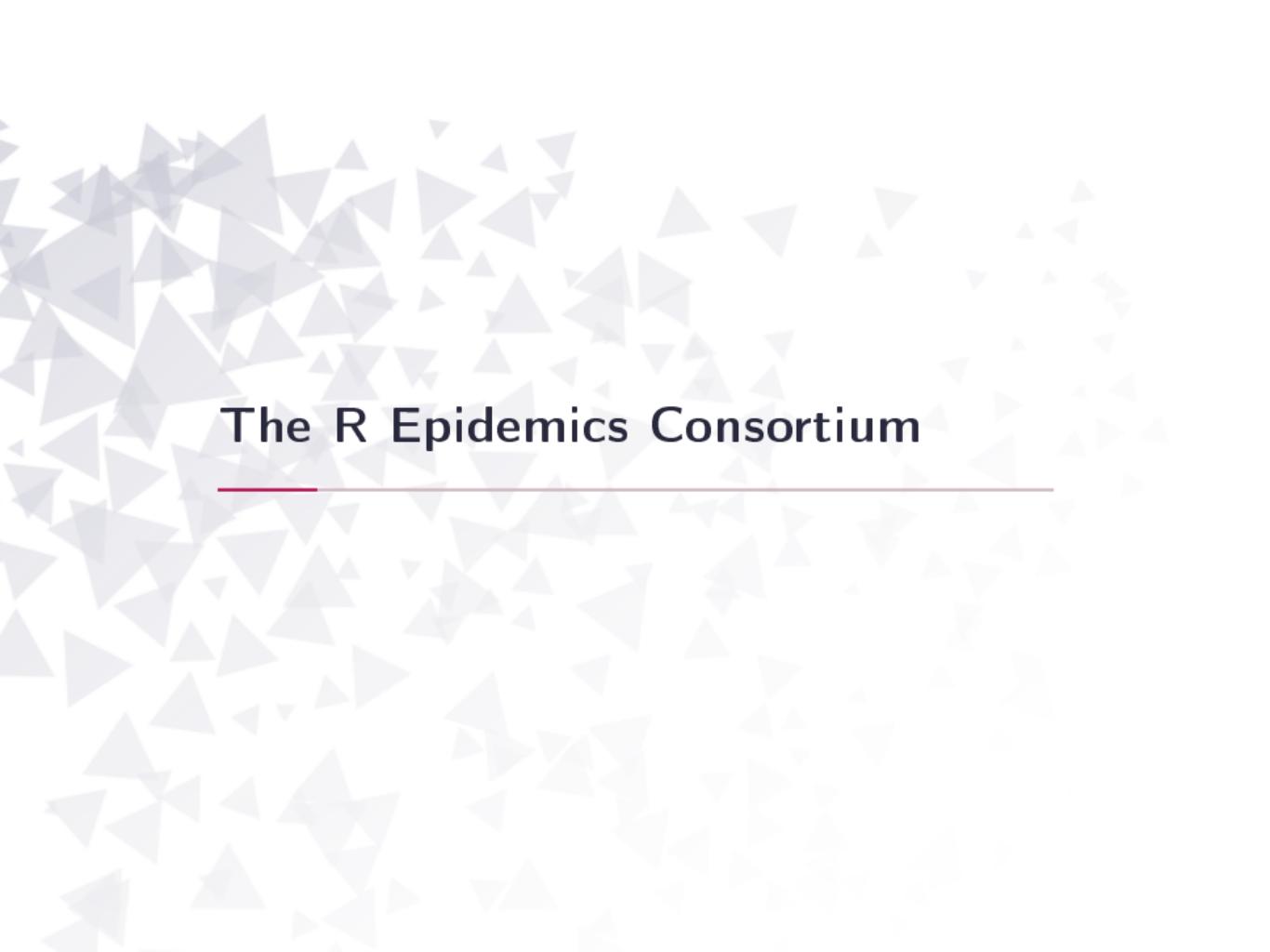


# Topics of the day



# Topics of the day





# The R Epidemics Consortium

---

# Lessons learnt from the Ebola response



# Lessons learnt from the Ebola response



# Lessons learnt from the Ebola response

**WHO Ebola response team**

**Help improving situation awareness**

Imperial College Ebola team

Timeline:

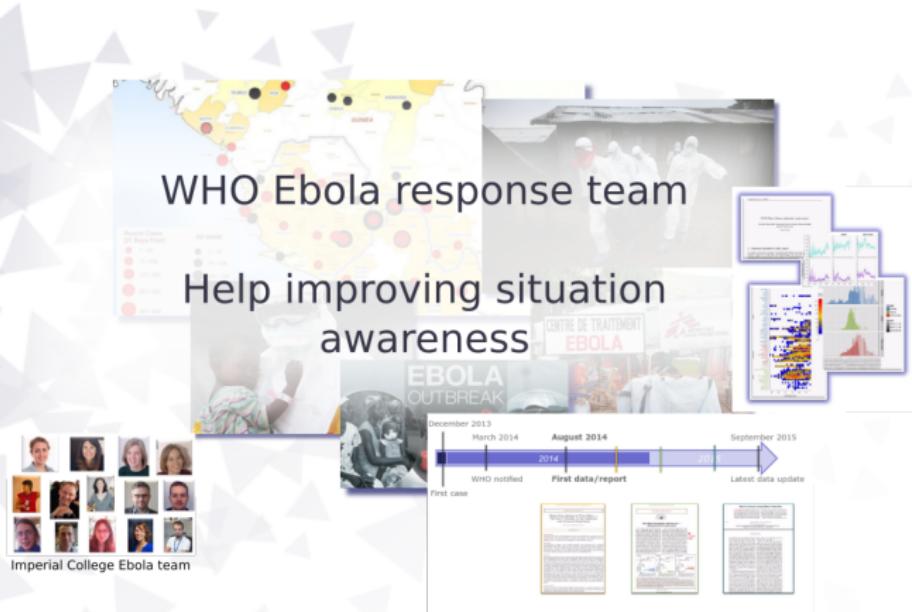
- December 2013: First case
- March 2014: WHO notified
- August 2014: First data/report
- September 2015: Latest data update

Centre de Traitement EBOLA

EBOLA OUTBREAK

WHO Ebola response team dashboard

# Lessons learnt from the Ebola response



*Most statistical/modelling tools for situation awareness were missing.*

# Who do we need to develop these tools?



# Who do we need to develop these tools?

Public Health



# Who do we need to develop these tools?



# Who do we need to develop these tools?



# From a hack to a pack



Hackout 3, summer 2016, Berkeley

# From a hack to a pack



Hackout 3, summer 2016, Berkeley



A word cloud illustrating the evolution of the project from a hackathon to a comprehensive platform. The words are arranged in a grid-like structure, with larger words indicating higher frequency or importance.

Key terms include: functional, incubation, userfriendly, secure, dictionary, systems, testing, automated, continuous, collection, series, repository, fast, tools, outbreaks, parsing, number, secured, bias, reporting, integration, gui, vhl, integration, reporting, delay, interface tree, cdc, site, estimation, censoring, transmission, package, reproducible, tools, reproduction, period, exposure, mutations, linelist, situations, synchronised, anonymised, epistemic, contact, epiinfo, compiled, clean, time, interface tree, fellow, symptoms, outbreaker, tracing, shiny, automation, incidence, cdc, cleaning, bayesian, sitereport, contacts, security, peak, report, estimation, curation, model, ggplot, clusters, rates, reliable, contacttracing, parallel, parameters, epidemics, genomics, distribution, edc, lonica, annotations, encrypted, distributions, environments.

# From a hack to a pack



Hackout 3, summer 2016, Berkeley

functional  
incubation  
userfriendly secure dictionary  
systems testing automated continuous  
collection series repository  
rpp efficiency number fast  
secured bias outbreaks  
parsing code integration  
reporting gui  
unit data delay  
epidemiology security peak  
situation anonymised  
opensource contact  
epiinfo clean time  
compiled interface tree  
outbreaker interface  
symptoms fellow  
lineelist tracing shiny  
automation cdc  
epicontacts edic  
ggplot clusters rates  
dashboard reliable  
parallel contacttracing  
parameters epidemics  
genomics distribution  
incidence bayesian site report  
cleaning estimation  
security transmission  
peak curation model  
contacts package  
report reproduction  
functional incubation  
userfriendly secure dictionary  
systems testing automated continuous  
collection series repository  
rpp efficiency number fast  
secured bias outbreaks  
parsing code integration  
reporting gui  
unit data delay  
epidemiology security peak  
situation anonymised  
opensource contact  
epiinfo clean time  
compiled interface tree  
outbreaker interface  
symptoms fellow  
lineelist tracing shiny  
automation cdc  
epicontacts edic  
ggplot clusters rates  
dashboard reliable  
parallel contacttracing  
parameters epidemics  
genomics distribution  
incidence bayesian site report  
cleaning estimation  
security transmission  
peak curation model  
contacts package  
report reproduction

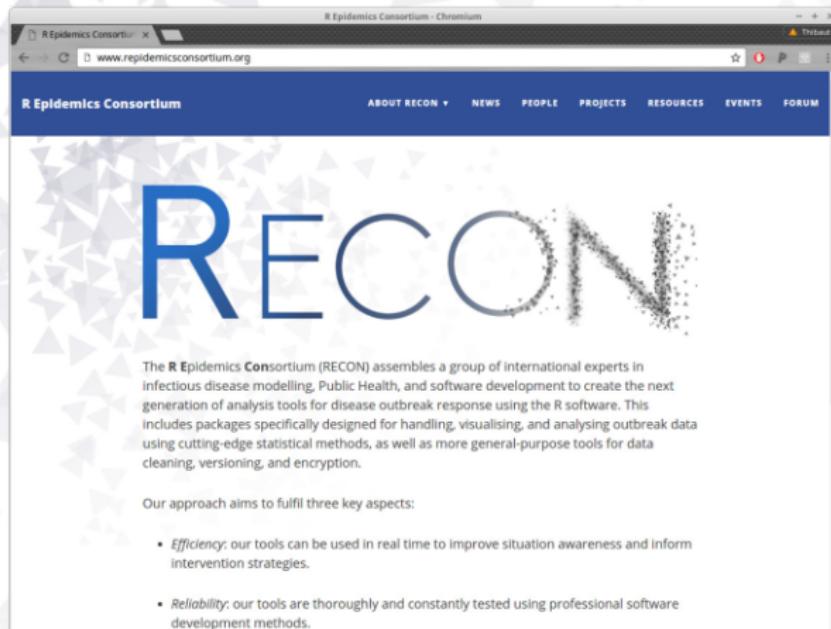
**RECON**  
[The R Epidemics Consortium]

# From a hack to a pack



# RECON: the R Epidemics Consortium

A taskforce to build a new generation of outbreak response tools in .



The screenshot shows a web browser window for the "R Epidemics Consortium" website. The title bar reads "R Epidemics Consortium - Chromium". The address bar shows "www.repidemicsconsortium.org". The main content area features a large, stylized blue "RECON" logo where the letters are composed of small dots. Below the logo, a paragraph of text describes the consortium's mission. A sidebar on the left lists navigation links: ABOUT RECON, NEWS, PEOPLE, PROJECTS, RESOURCES, EVENTS, and FORUM. A footer at the bottom contains social media icons for GitHub, LinkedIn, and YouTube.

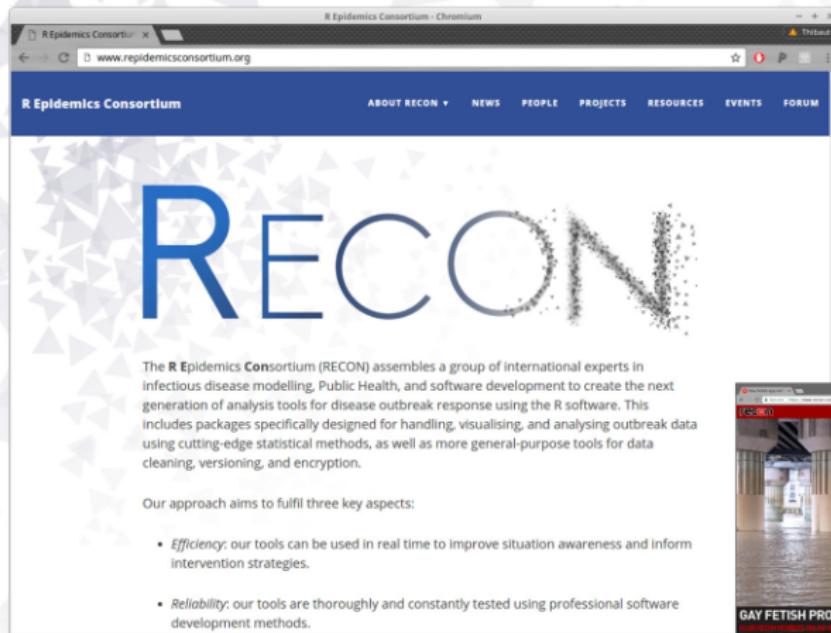
The R Epidemics Consortium (RECON) assembles a group of international experts in infectious disease modelling, Public Health, and software development to create the next generation of analysis tools for disease outbreak response using the R software. This includes packages specifically designed for handling, visualising, and analysing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption.

Our approach aims to fulfil three key aspects:

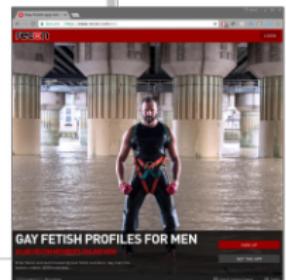
- *Efficiency*: our tools can be used in real time to improve situation awareness and inform intervention strategies.
- *Reliability*: our tools are thoroughly and constantly tested using professional software development methods.

# RECON: the R Epidemics Consortium

A taskforce to build a new generation of outbreak response tools in .

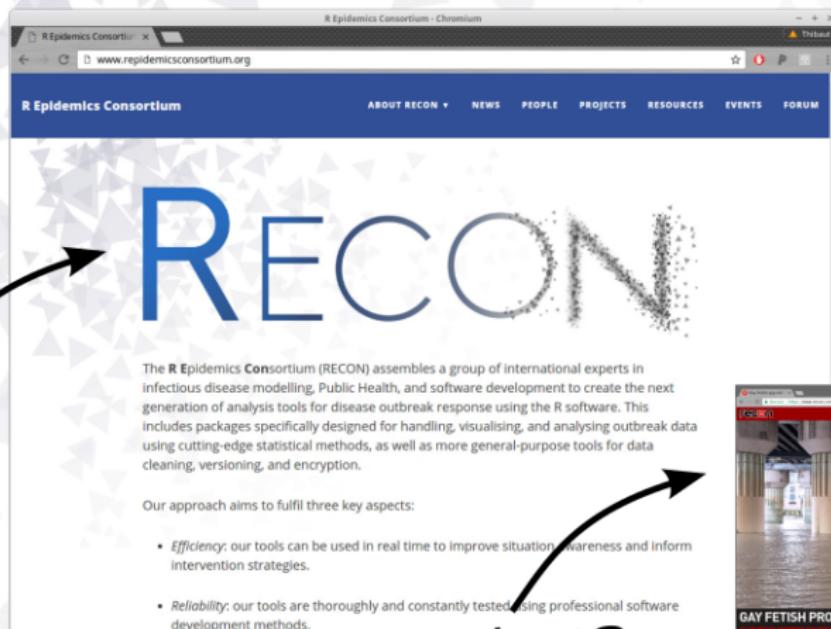


The screenshot shows the homepage of the RECON website. The header features the text "R Epidemics Consortium" and "RECON". Below the header is a large graphic where the letters of "RECON" are composed of small dots or particles. A descriptive paragraph explains the consortium's mission: "The R Epidemics Consortium (RECON) assembles a group of international experts in infectious disease modelling, Public Health, and software development to create the next generation of analysis tools for disease outbreak response using the R software. This includes packages specifically designed for handling, visualising, and analysing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption." Below this text, a section titled "Our approach aims to fulfil three key aspects:" lists three bullet points: "Efficiency: our tools can be used in real time to improve situation awareness and inform intervention strategies.", "Reliability: our tools are thoroughly and constantly tested using professional software development methods.", and "Scalability: our tools are designed to handle large amounts of data and can be scaled up as needed."



# RECON: the R Epidemics Consortium

A taskforce to build a new generation of outbreak response tools in  .



The screenshot shows the homepage of the R Epidemics Consortium (RECON) website. The header includes the logo and navigation links: ABOUT RECON, NEWS, PEOPLE, PROJECTS, RESOURCES, EVENTS, and FORUM. The main feature is a large, stylized title "RECON" where the letters are composed of small dots, suggesting data points or individuals. Below the title is a descriptive paragraph about the consortium's mission to assemble international experts in infectious disease modelling, public health, and software development to create tools for outbreak response using R. It highlights the use of cutting-edge statistical methods and general-purpose tools for data cleaning, versioning, and encryption. A bulleted list at the bottom outlines three key aspects: efficiency, reliability, and real-time improvement of intervention strategies.

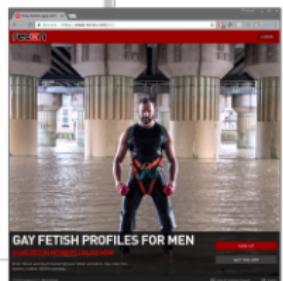
Us

The R Epidemics Consortium (RECON) assembles a group of international experts in infectious disease modelling, Public Health, and software development to create the next generation of analysis tools for disease outbreak response using the R software. This includes packages specifically designed for handling, visualising, and analysing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption.

Our approach aims to fulfil three key aspects:

- *Efficiency*: our tools can be used in real time to improve situation awareness and inform intervention strategies.
- *Reliability*: our tools are thoroughly and constantly tested using professional software development methods.

Not us



# RECON

*www.repidemicsconsortium.org*

- started 6th September 2016
- ~70 members
- 17 countries, > 40 institutions
- ~ 3 packages released, 20 under development
- public forum, blog, online resources



## Statistical software development

- **efficiency**: useful for improving situation awareness in real time
- **reliability**: outputs can be trusted
- **accessibility**: widely available, easy learning curve



## Statistical software development

- **efficiency**: useful for improving situation awareness in real time
- **reliability**: outputs can be trusted
- **accessibility**: widely available, easy learning curve

## Translation

- **disseminating knowledge**: free online training material, involvement with FETPs, workshops
- **outbreak response**: deployment to the field
- **RECON deployer**: portable data analysis environment

# RECON: projects

The screenshot shows the 'Projects' section of the R Epidemics Consortium website. The page features a grid of 14 project entries, each represented by a circular icon containing gears and a brief description. The projects are:

- epiflows**: Visualisation and analysis of passenger flows.
- epimaps**: Helpers and wrappers for mapping diseases.
- epimatch**: Finding matching patient records across tabular data sets.
- epitrix**: Small utility functions for epidemiology.
- gisfirstaid**: Tutorials and code gists for mapping infectious diseases.
- incidence.ul**: Graphical user interface for incidence.
- nomad**: Pack up R to take away.
- outbreaker2**: Inferring transmission chains by integrating epidemiological and genetic data.
- projections**: Projections of future incidence.
- recon.ul**: Template shiny GUI for RECON packages.
- recontools**: Tools to develop RECON packages.
- vimes**: Visualisation and Monitoring of Epidemics, including some outbreak detection algorithms.

A red circle highlights the **vimes** project entry.



# An integrative approach for outbreak detection

---



Aims: develop a new method which..



---

<sup>1</sup> well, really, I made that up because I was reading 'Snuff' at the time; at least this one is not a dodgy website (yet); incidentally, Terry Pratchett was a huge fan of using long footnotes, which were often quite entertaining to read; note that it does not apply here: if you are still reading this, you probably missed what I just said



Aims: develop a new method which..

- **detects clusters** i.e. groups of related cases (on the same transmission chain)

---

<sup>1</sup> well, really, I made that up because I was reading 'Snuff' at the time; at least this one is not a dodgy website (yet); incidentally, Terry Pratchett was a huge fan of using long footnotes, which were often quite entertaining to read; note that it does not apply here: if you are still reading this, you probably missed what I just said



Aims: develop a new method which..

- **detects clusters** i.e. groups of related cases (on the same transmission chain)
- **integrates different data:** temporal, spatial, genetic, etc.

<sup>1</sup> well, really, I made that up because I was reading 'Snuff' at the time; at least this one is not a dodgy website (yet); incidentally, Terry Pratchett was a huge fan of using long footnotes, which were often quite entertaining to read; note that it does not apply here: if you are still reading this, you probably missed what I just said

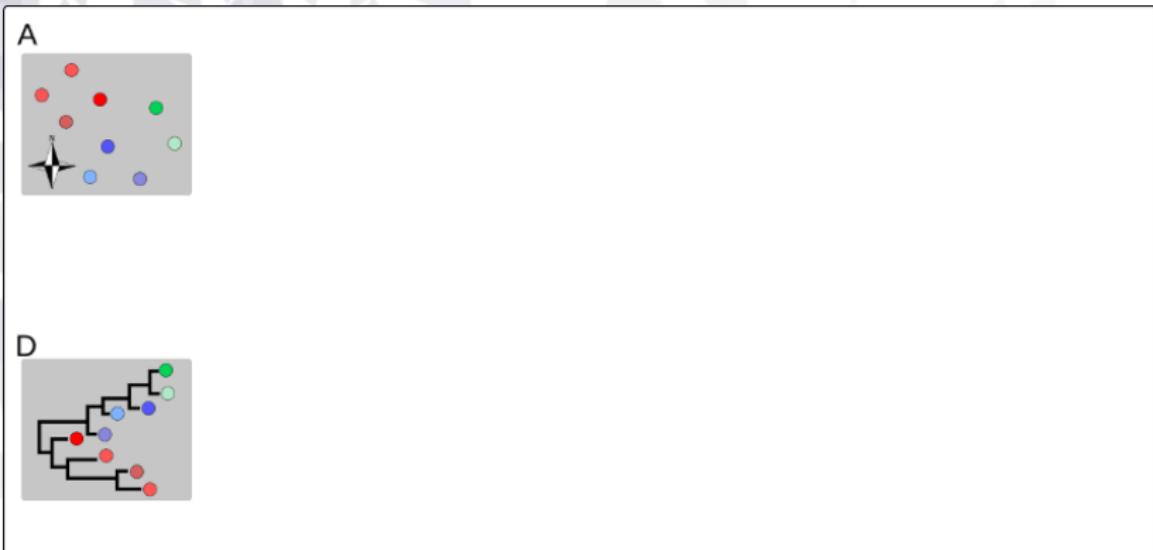


Aims: develop a new method which..

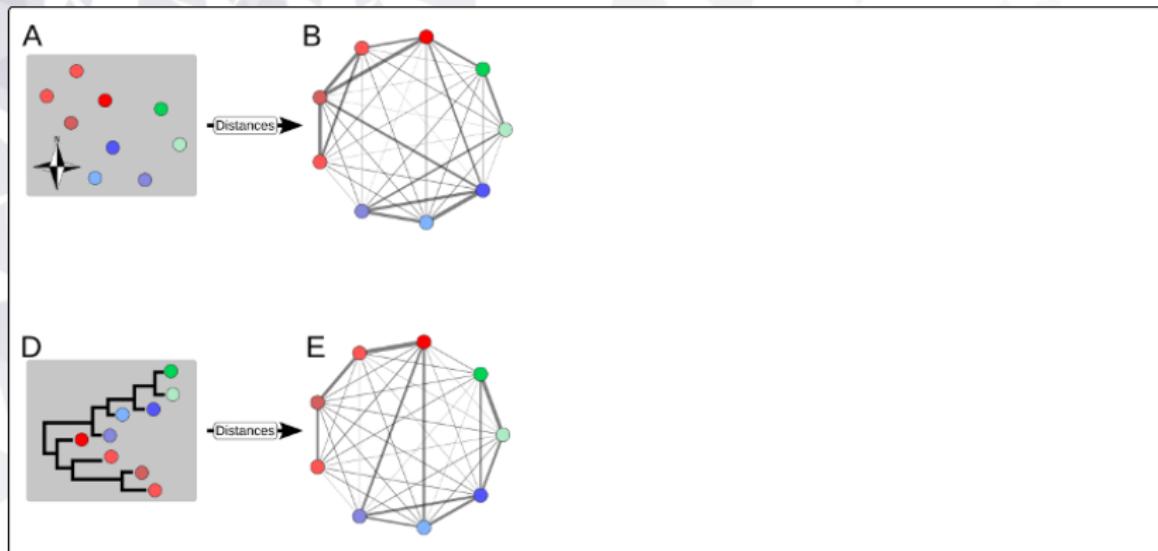
- **detects clusters** i.e. groups of related cases (on the same transmission chain)
- **integrates different data:** temporal, spatial, genetic, etc.
- **works fast, scales well:** so that it can be used for real-time outbreak detection

<sup>1</sup> well, really, I made that up because I was reading 'Snuff' at the time; at least this one is not a dodgy website (yet); incidentally, Terry Pratchett was a huge fan of using long footnotes, which were often quite entertaining to read; note that it does not apply here: if you are still reading this, you probably missed what I just said

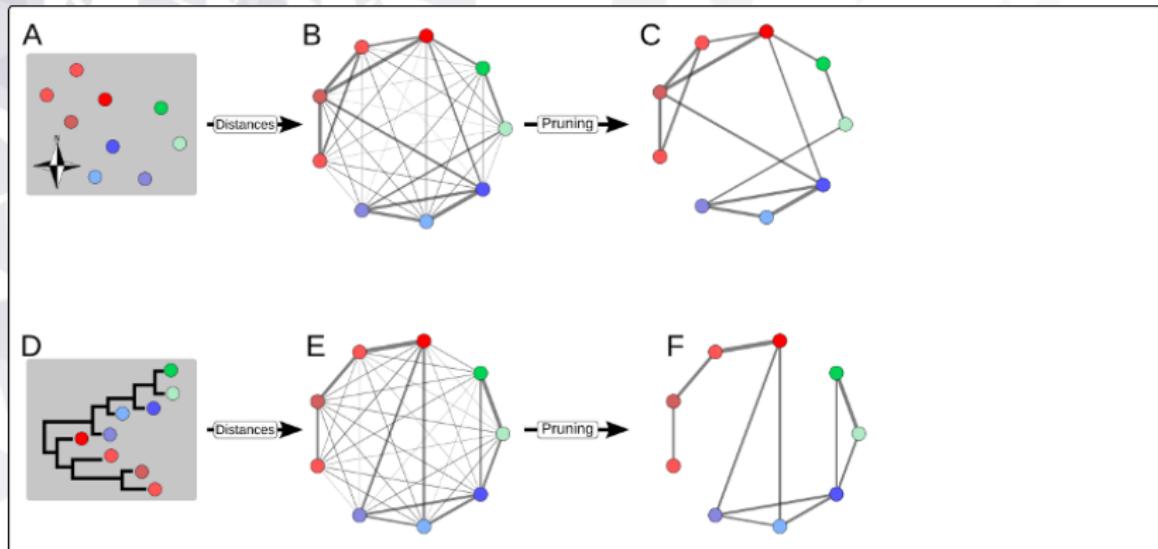
# A graph-based evidence synthesis approach



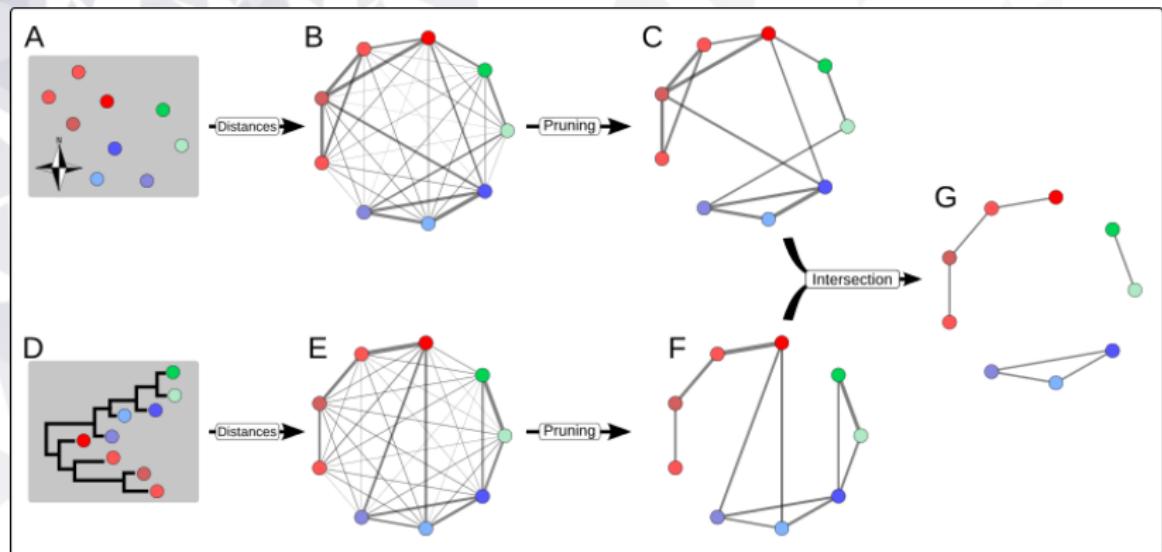
# A graph-based evidence synthesis approach



# A graph-based evidence synthesis approach

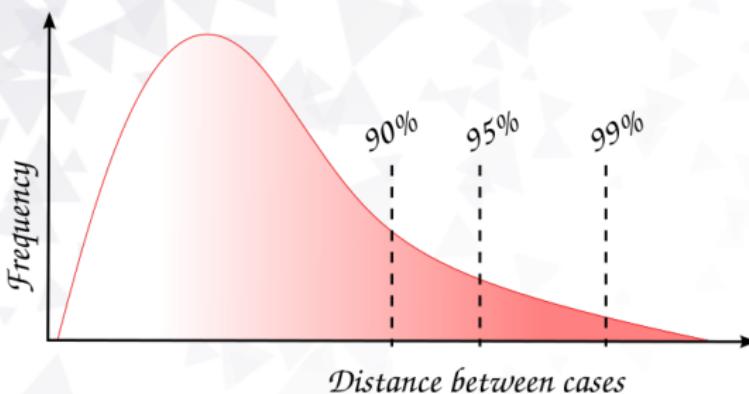


# A graph-based evidence synthesis approach

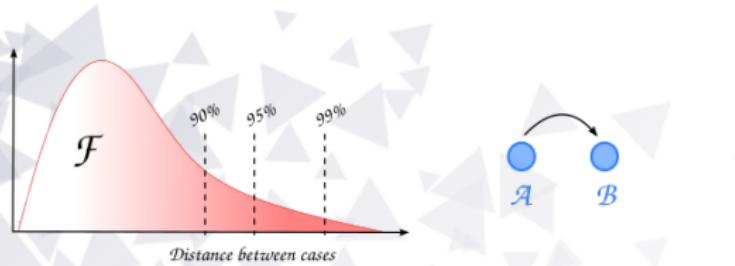


# Pruning graphs: where to cut?

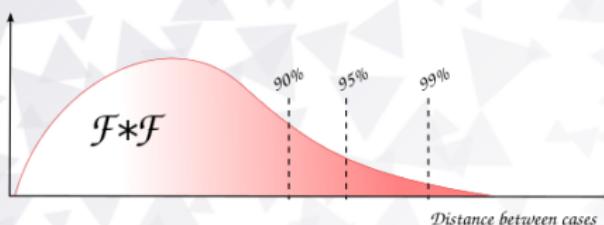
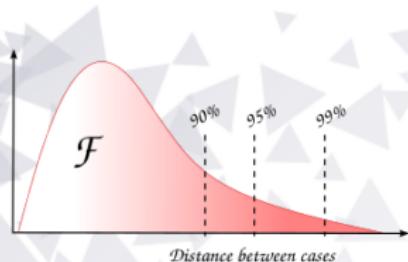
Assuming a known expected distribution between pairs of cases (e.g. serial interval, spatial kernel, molecular clock), different quantiles can be used:



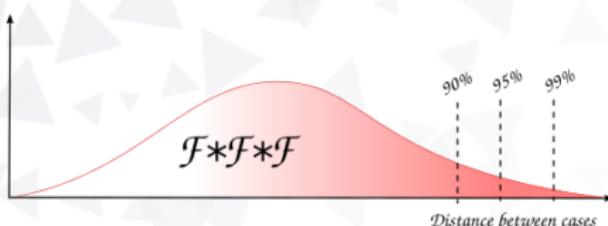
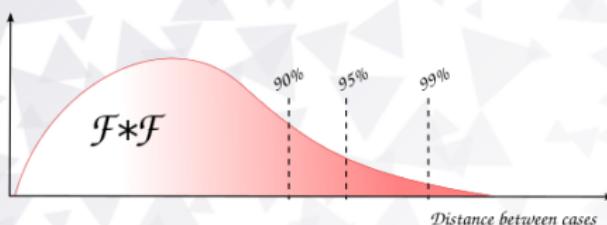
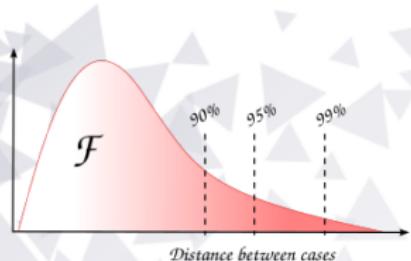
# Pruning graphs: where to cut?



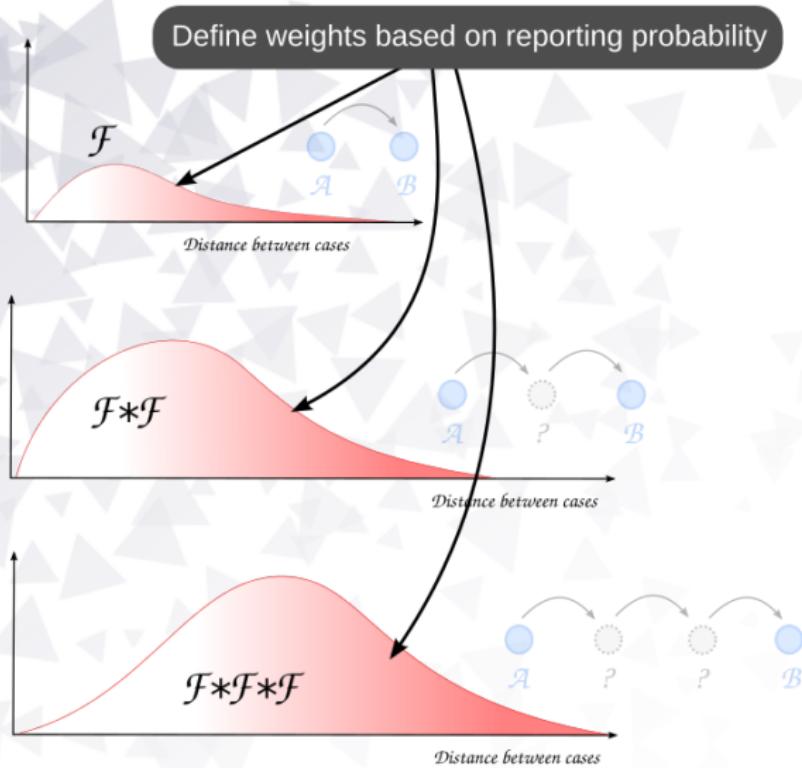
# Pruning graphs: where to cut?



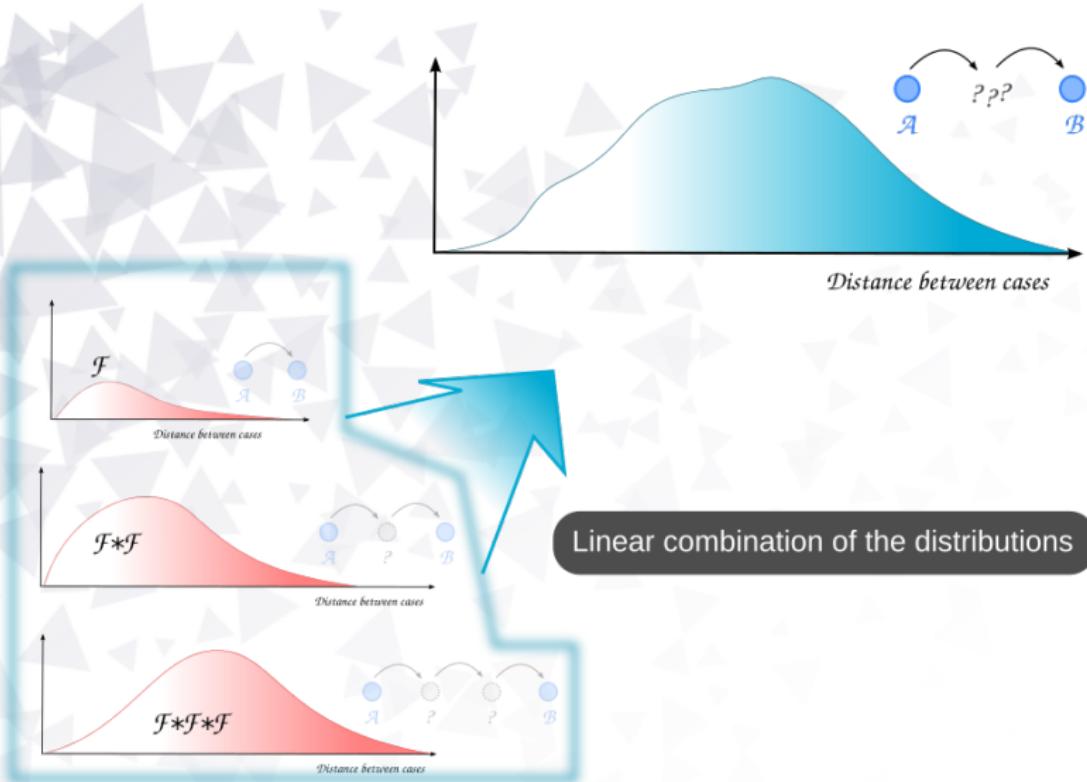
# Pruning graphs: where to cut?



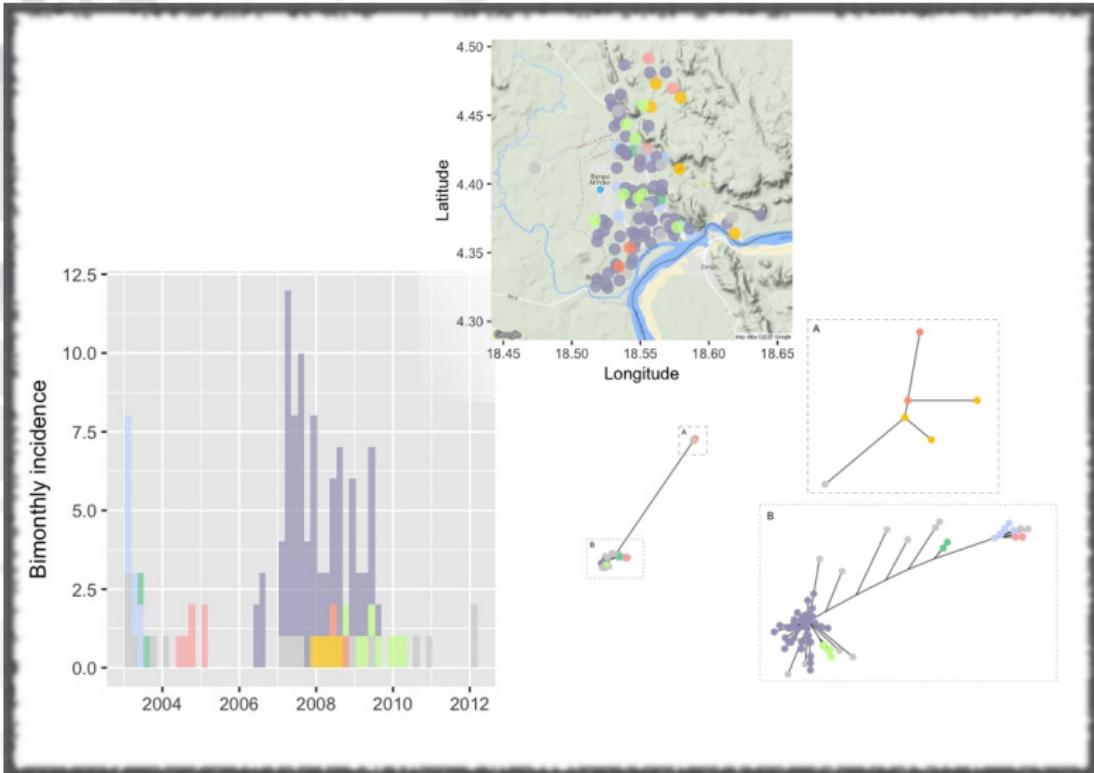
# Pruning graphs: where to cut?



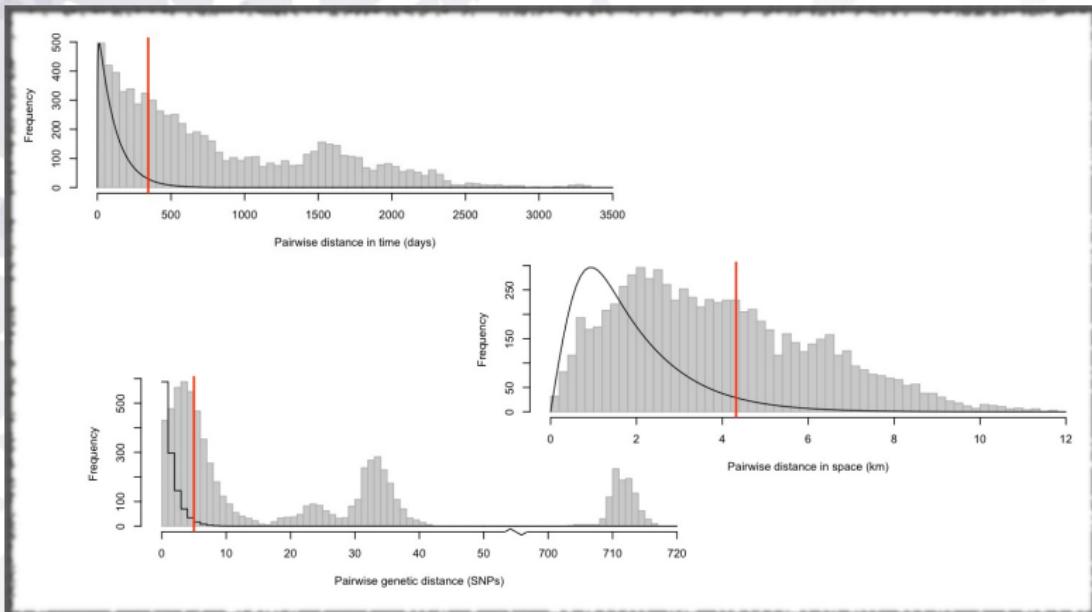
# Pruning graphs: where to cut?



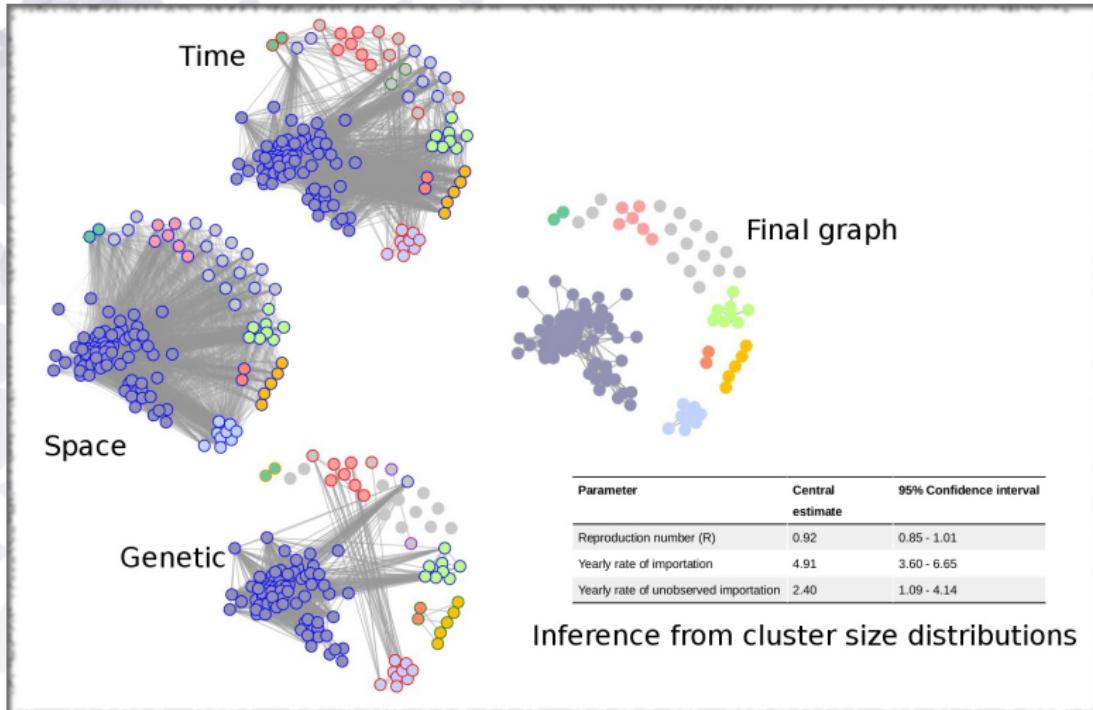
# Application: dog rabies epidemics, Central African Republic



# Distributions of distances between cases



# Results



# Perspectives

- **flexible** approach for detecting outbreaks using different data types

# Perspectives

- **flexible** approach for detecting outbreaks using different data types
- **threshold**: unsatisfying, but sensitivity study easy

# Perspectives

- **flexible** approach for detecting outbreaks using different data types
- **threshold**: unsatisfying, but sensitivity study easy
- **fast and scalable**: possible integration in routine surveillance

# Perspectives

- **flexible** approach for detecting outbreaks using different data types
- **threshold**: unsatisfying, but sensitivity study easy
- **fast and scalable**: possible integration in routine surveillance
- can serve as basis to other methods for integrating different data sources

# Thanks

- Conference organisers
- Colleagues: Anne Cori, Pierre Nouvellet, Tini Garske, Hervé Bourhy, Emmanuel Nakouné
- Groups: WHO Ebola Response Team, Hackout 1/2/3, RECON members, GOARN
- funding: HPRU-NIHR, MRC

The logo consists of the word "RECON" in a large, bold, blue sans-serif font. The letter "O" is unique, composed of numerous small, dark grey dots arranged in a circular pattern.

*[www.repidemicsconsortium.org](http://www.repidemicsconsortium.org)*

The logo consists of the word "vimes" in a pink, lowercase, sans-serif font.

*[www.repidemicsconsortium.org/vimes](http://www.repidemicsconsortium.org/vimes)*