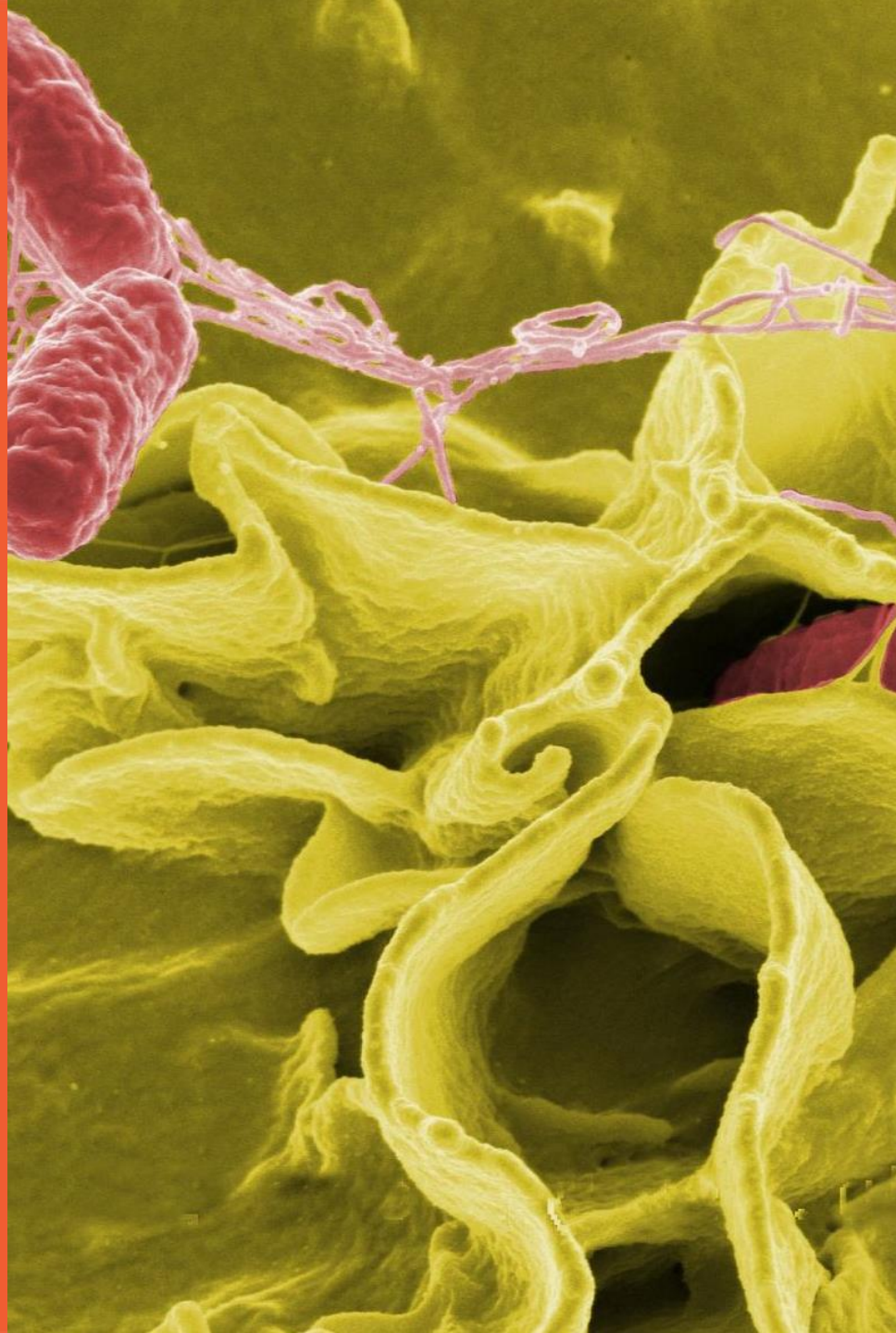
Tracing the evolution of non-typhoidal Salmonella: Is a new superbug emerging?

Presented by

Oliver Cliff

Complex Systems Research Group

THE UNIVERSITY OF SYDNEY



Why salmonella in NSW?

- **93.8 million cases** and **155,000 deaths** each year globally
  - More cases in Australia than anywhere else
  - Most fatal airborne disease in Australia

- ***S*. Typhimuriam (STM)** is the dominant subspecies of non-typhoidal salmonellosis

- Drivers of their evolution remain poorly understood

THE UNIVERSITY OF SYDNEY

Dataset

- Nine consecutive seasons of instances in NSW
  - 1st January 2008 to 31st December 2016
- **17,107 isolates** of STM
  - 99.3% of all STM found from human cases in NSW over 3,287 days
- Genotyped through **MLVA**
  - MLVA = **M**ultiple-**L**ocus **V**ariable-number tandem repeat (VNTR) **A**nalysis
  - Allows for more diversity of strains to be identified
  - 1675 unique MLVAs identified

What is MLVA?

Tandem repeats

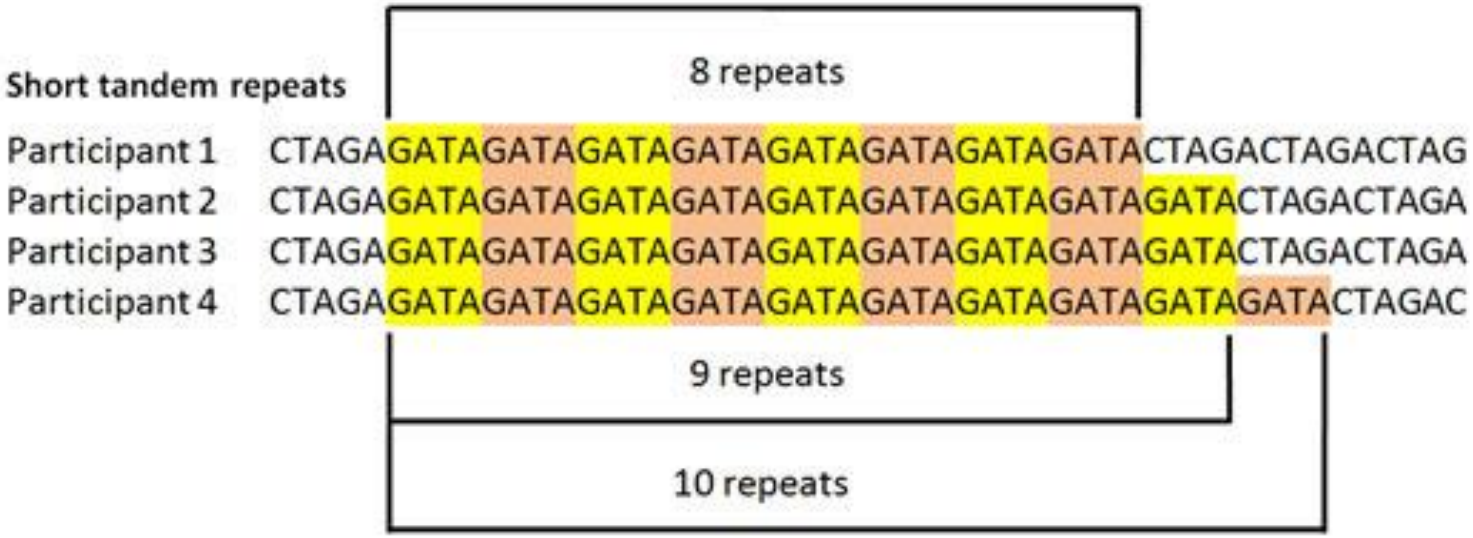
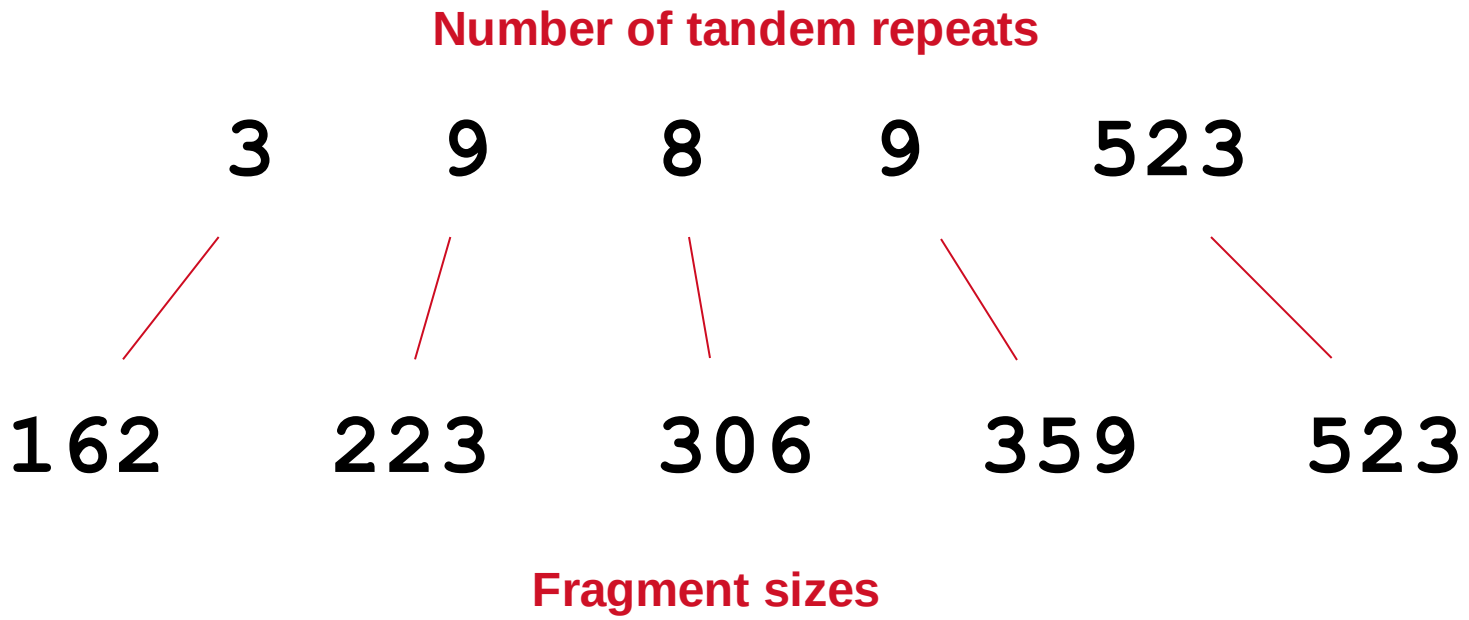


Image source: https://www.stewartsociety.org/images/bannockburn-tandem-repeats.jpg

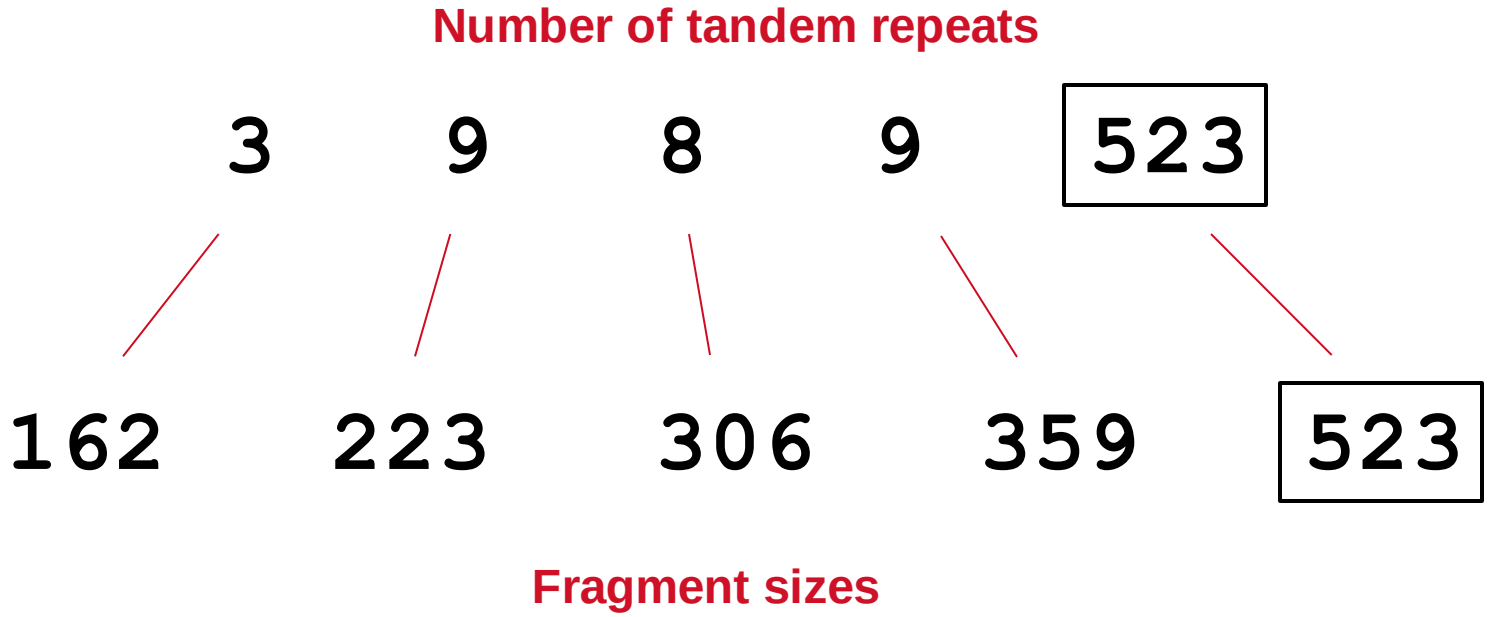
What is MLVA?

› MLVA profile and fragment sizes



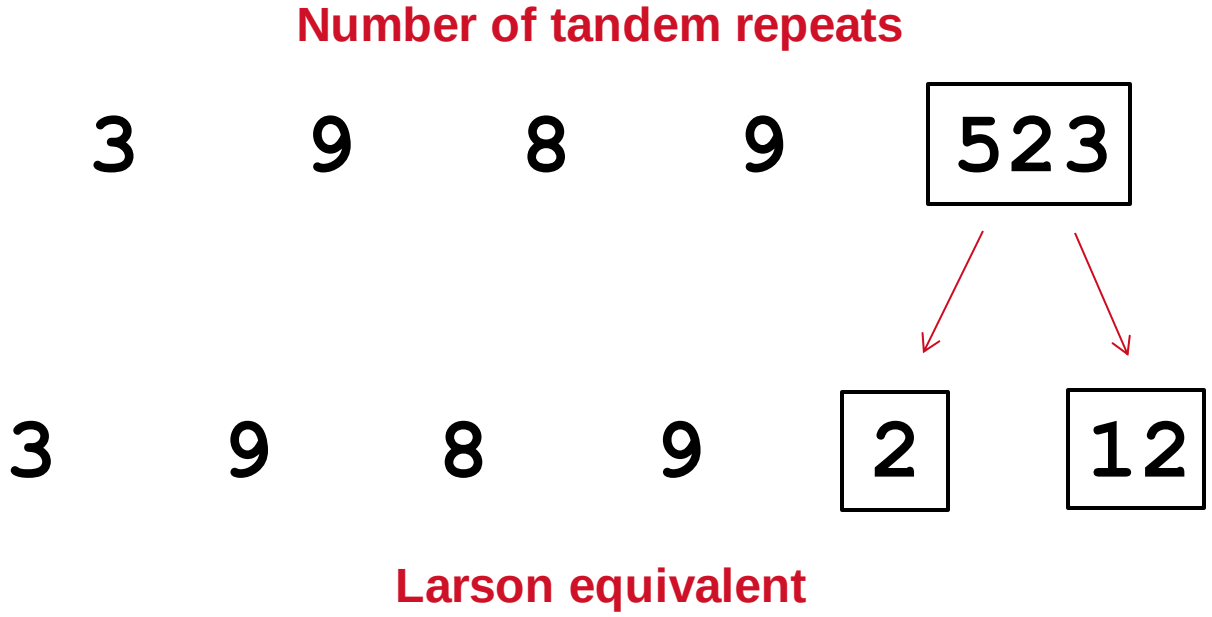
What is MLVA?

› MLVA profile and fragment sizes

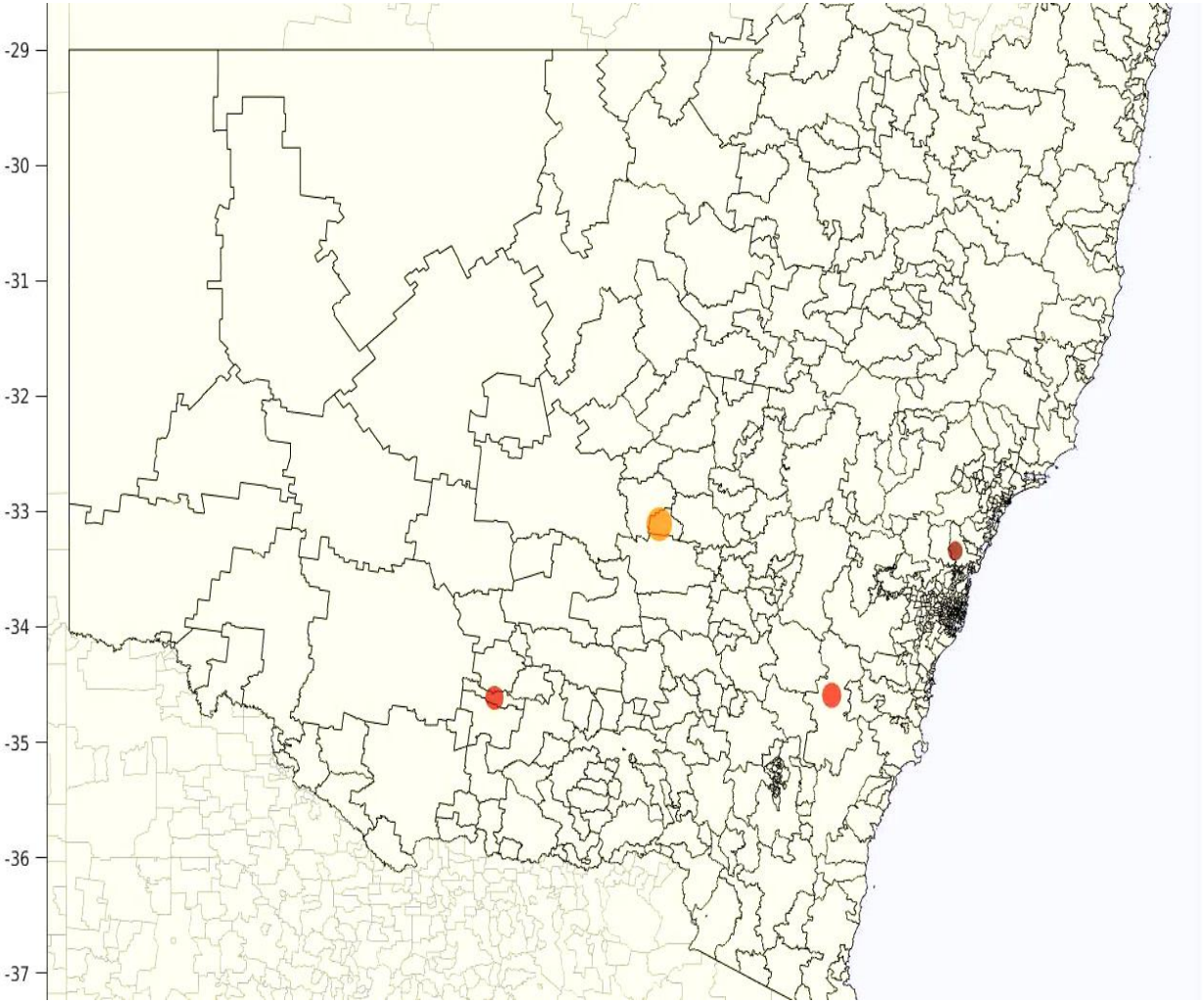


What is MLVA?

› MLVA profile and fragment sizes



Dataset: STM in NSW 2008-2016



MLVA isolates as a complex network

- Considered as a complete graph
  - 1675 nodes (unique isolates)
  - Edge weights are inverse of pairwise MLVA distance
- Get closeness centrality of MLVA isolate in network
- Cluster isolates
  - **Partition** (linkage) clusters
  - **Overlapping** clusters
- Study evolution (incidence) of global network and individual clusters

Complex network analysis

- Edge weights
  - Inverse of Manhattan distance (L1-norm)

x	3	9	8	9	2	12

Complex network analysis

- Edge weights
  - Inverse of Manhattan distance (L1-norm)

x	3	9	8	9	2	12
y	3	10	8	8	2	12

- Edge weights
  - Inverse of Manhattan distance (L1-norm)

x	3		9	8	9	2	12
y	3		10	8	8	2	12
$d(x,y)$	0	$\|+1\|$	0	$\|-1\|$	0	0

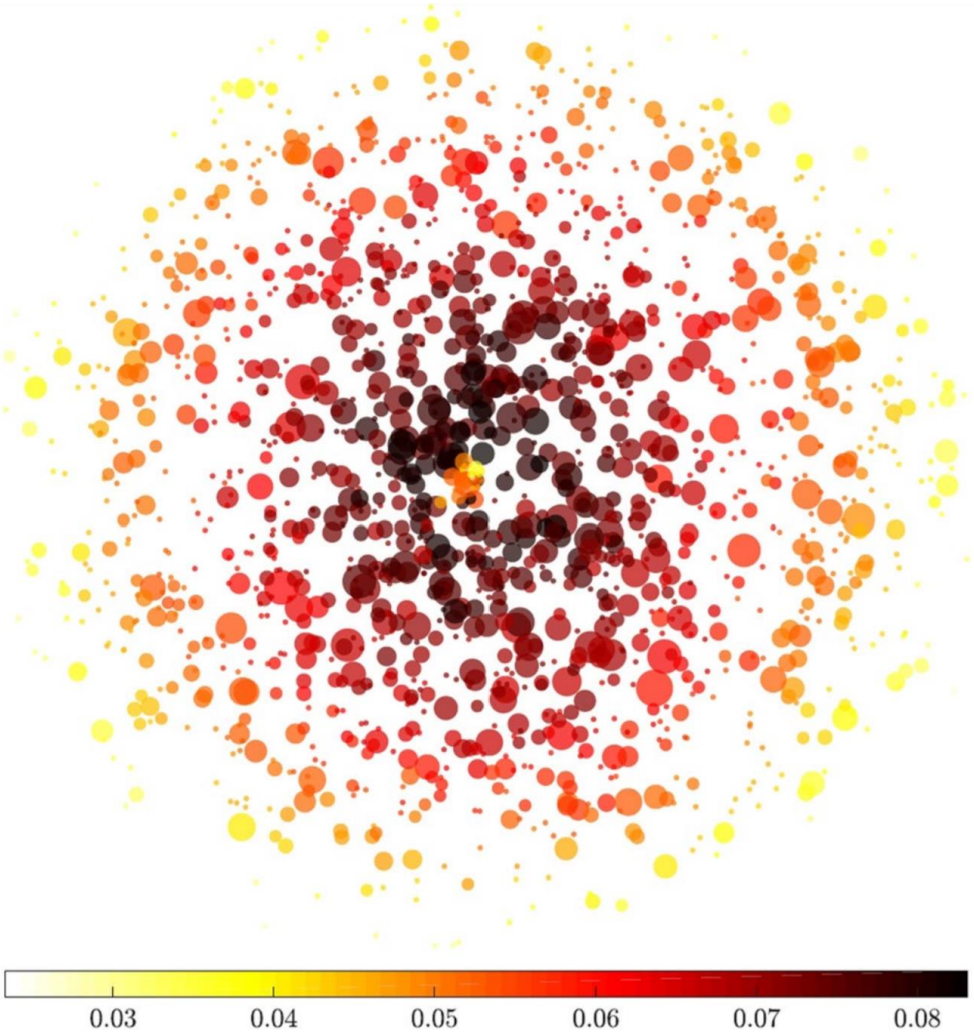
- Edge weights
  - Inverse of Manhattan distance (L1-norm)

x	3		9	8	9	2	12
y	3		10	8	8	2	12
$d(x,y)$	0	$\lvert+1\rvert$		0	$\lvert-1\rvert$	0	0

- Closeness centrality $C(x) = \frac{1}{\sum_y d(x,y)}$

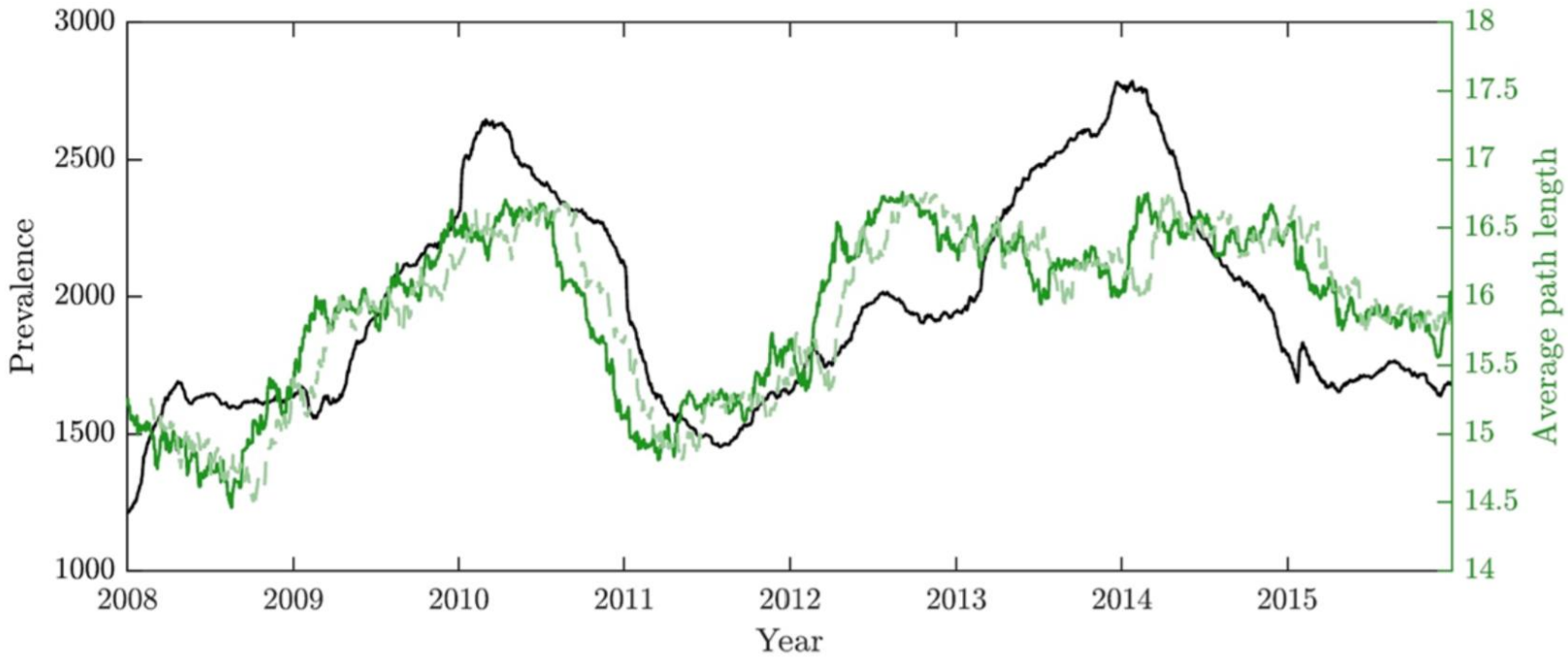
THE UNIVERSITY OF SYDNEY

Closeness centrality of MLVA profiles



Global network properties

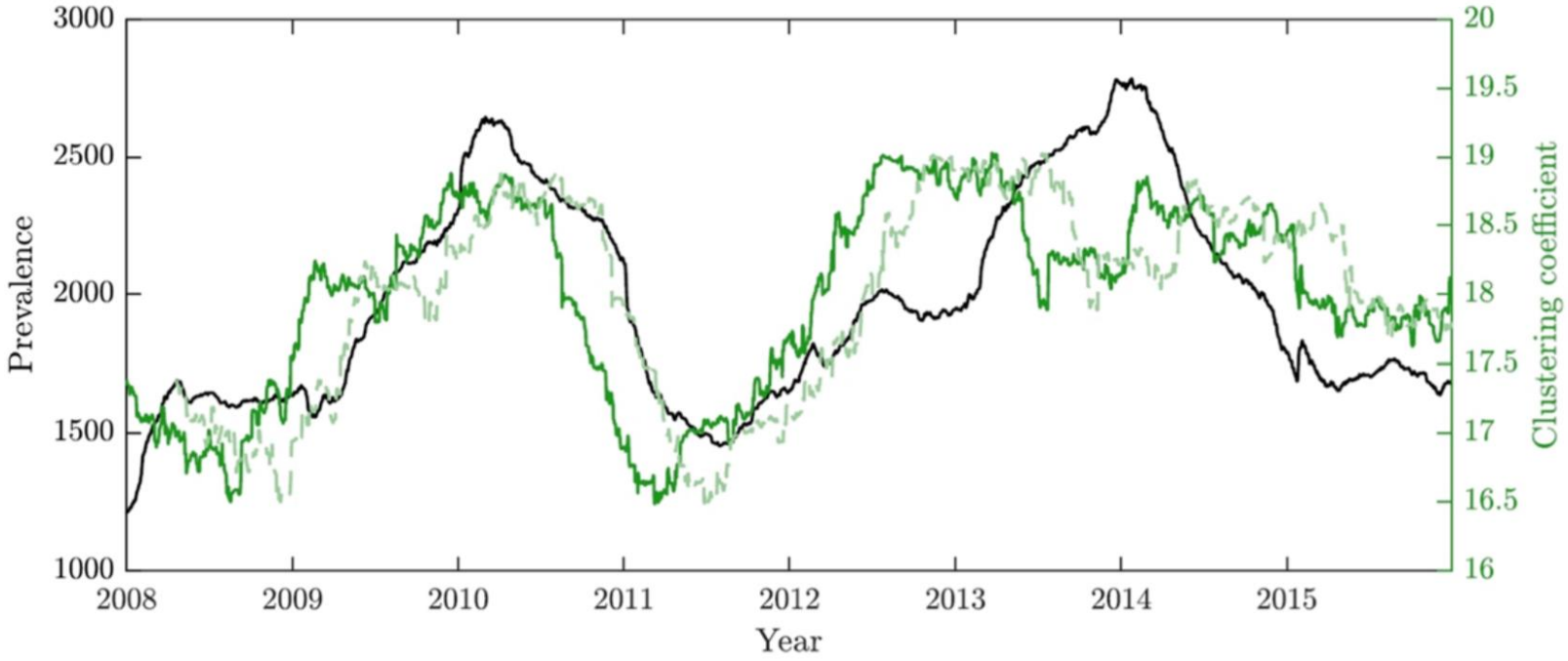
Path length and prevalence ($\rho \approx 0.7$ at $\sim$100 days)



› Path length: average distance to all other nodes

Global network properties

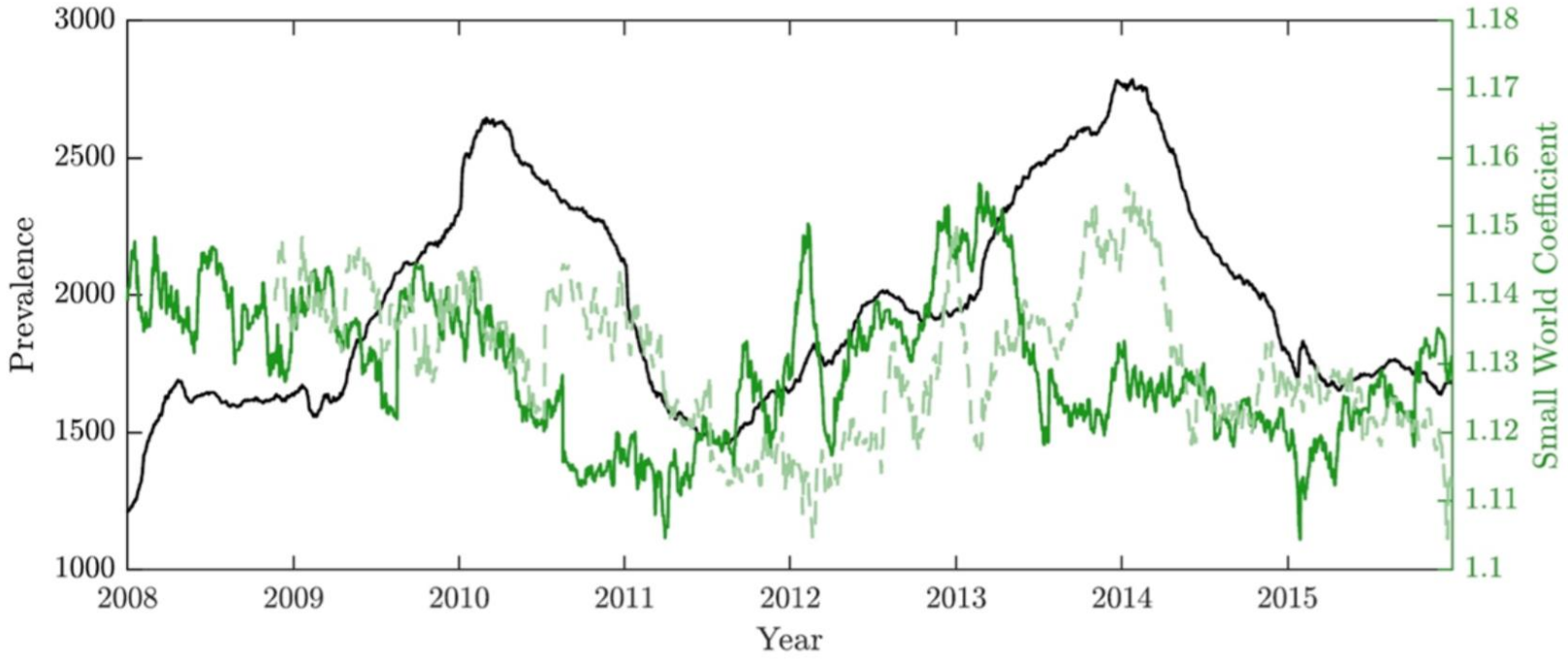
Clustering coefficient and prevalence ($\rho \approx 0.7$ at ~ 50 days)



› Clustering coefficient: how close a node is to being a clique (complete graph)

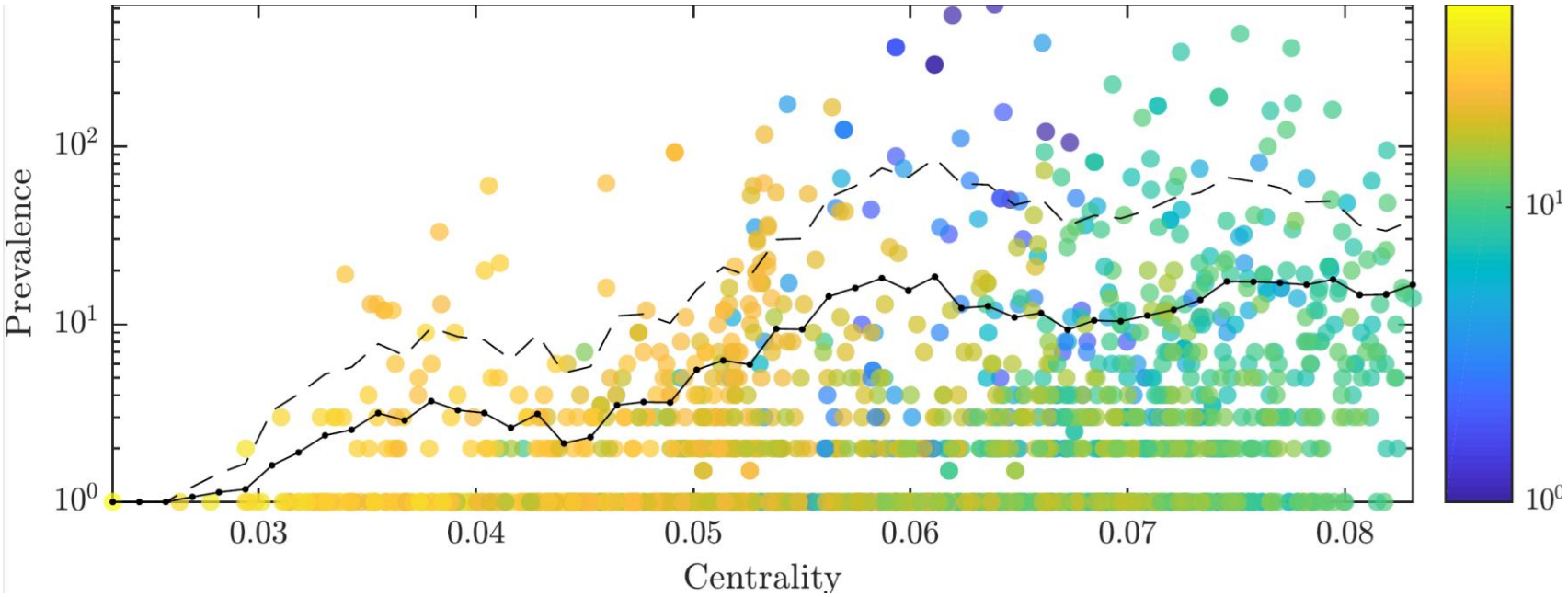
Global network properties

Small world coefficient and prevalence ($\rho \approx 0.6$ at $\sim$300 days)



› Small world coefficient: ratio of clustering to path length

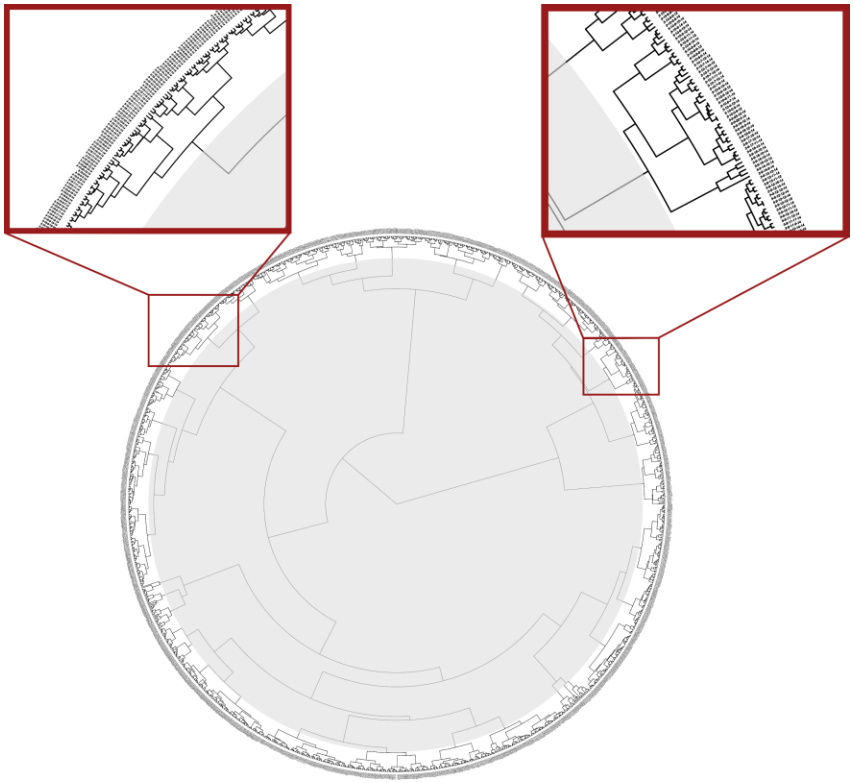
Closeness centrality of MLVA profiles



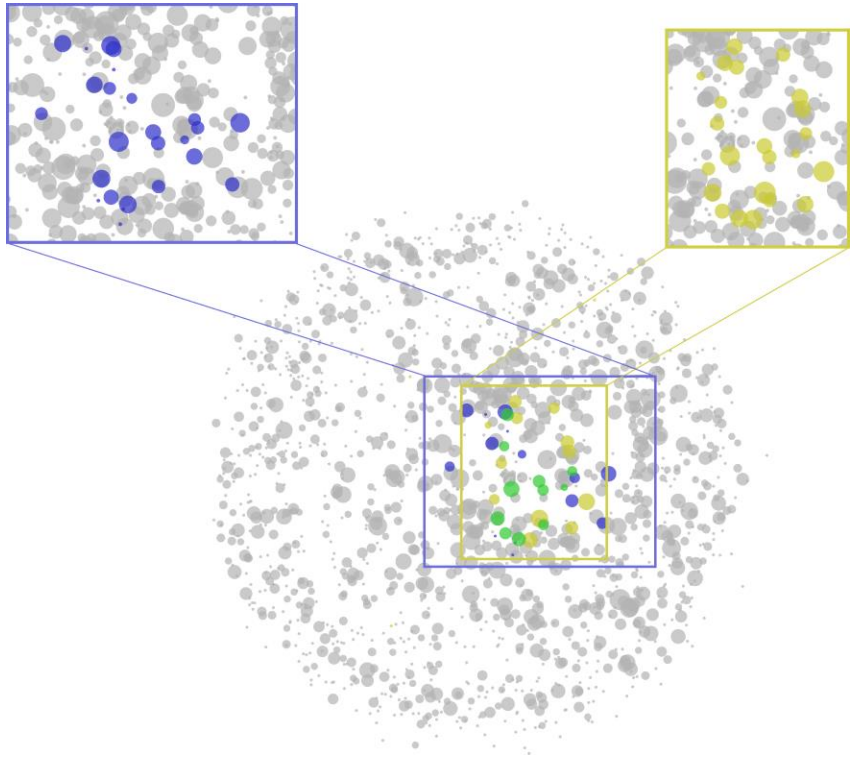
- Black line traces binned averages
- Colours denote distance to most prevalent node
- Higher centrality correlates to higher prevalence

Clustering

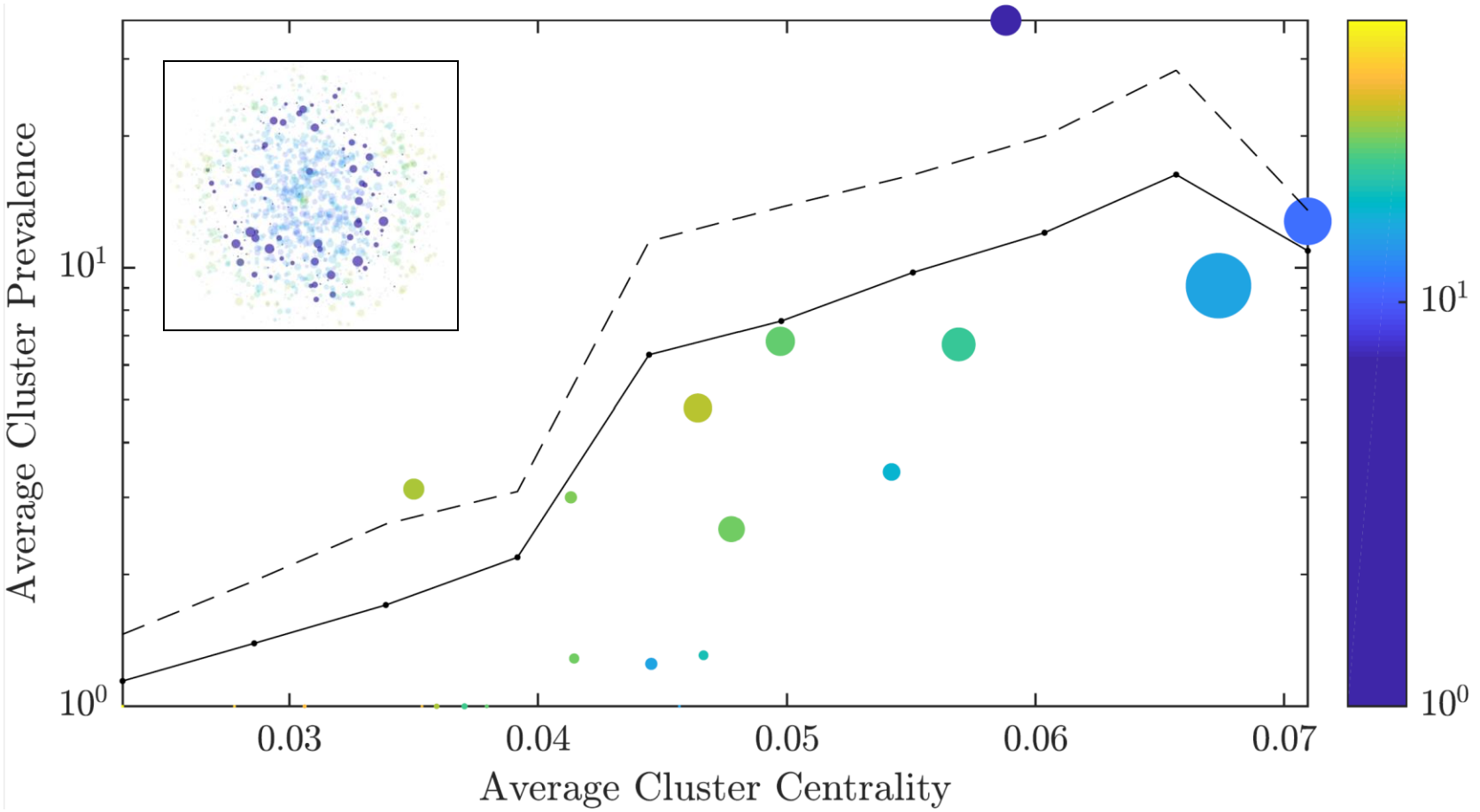
Linkage (partitioning) approach



Overlapping approach

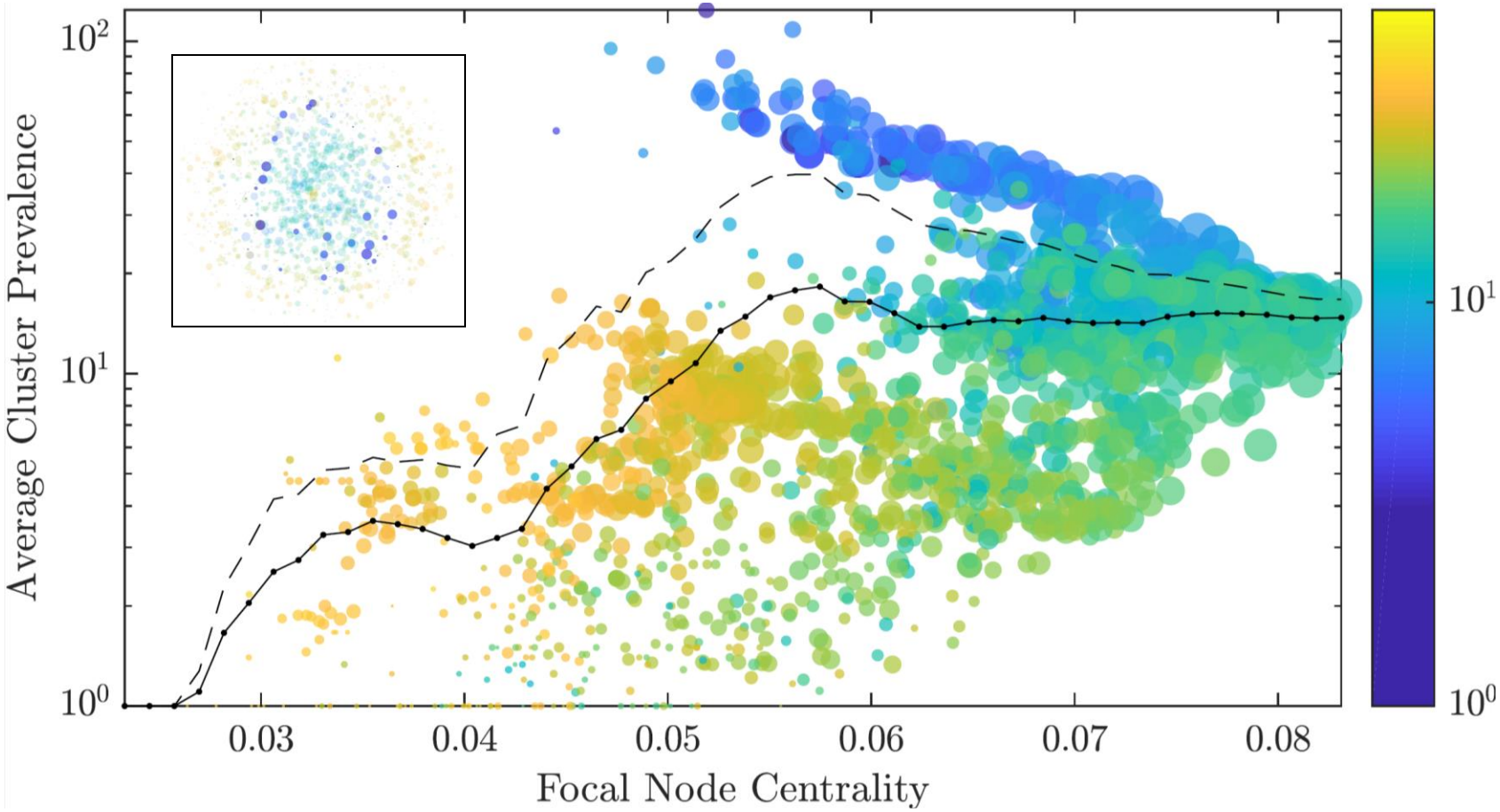


Incidence-Centrality plot (partitioning clusters)

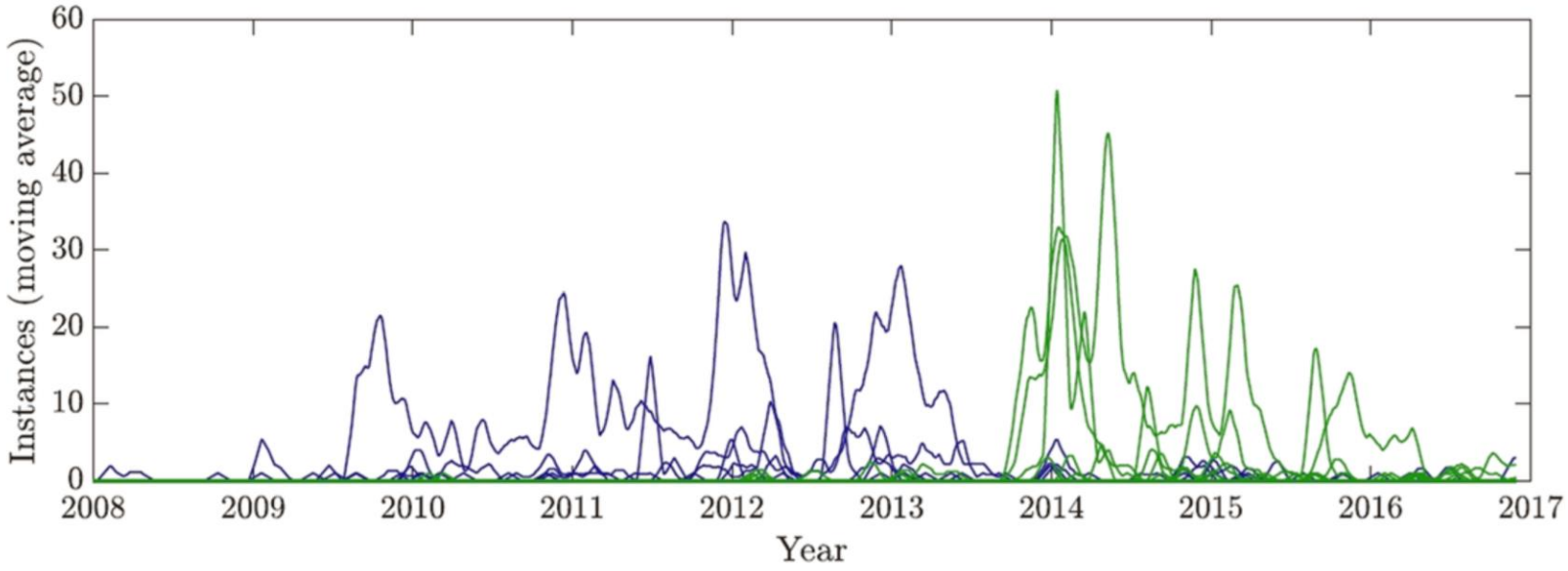


THE UNIVERSITY OF SYDNEY

Incidence-Centrality plot (overlapping clusters)

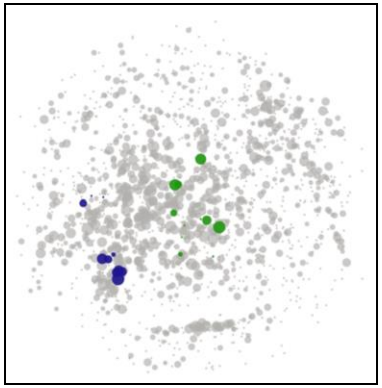


Temporal evolution of clusters



- 3-9-8-13-523
- 3-9-8-14-523
- 3-9-9-13-523
- 3-9-9-13-525
- 3-9-9-14-523

- 3-9-9-14-523
- 3-9-9-15-523
- 3-10-8-13-523
- 3-10-8-14-523
- 3-10-9-13-523
- 3-10-9-14-523



- 3-16-9-11-523
- 3-16-10-10-523
- 3-16-10-11-523
- 3-17-9-11-523
- 3-17-10-11-523
- 4-16-10-11-517

- 2-16-9-11-523
- 3-15-9-10-496
- 3-15-9-10-523
- 3-15-9-11-523
- 3-16-9-10-496
- 3-16-9-10-523