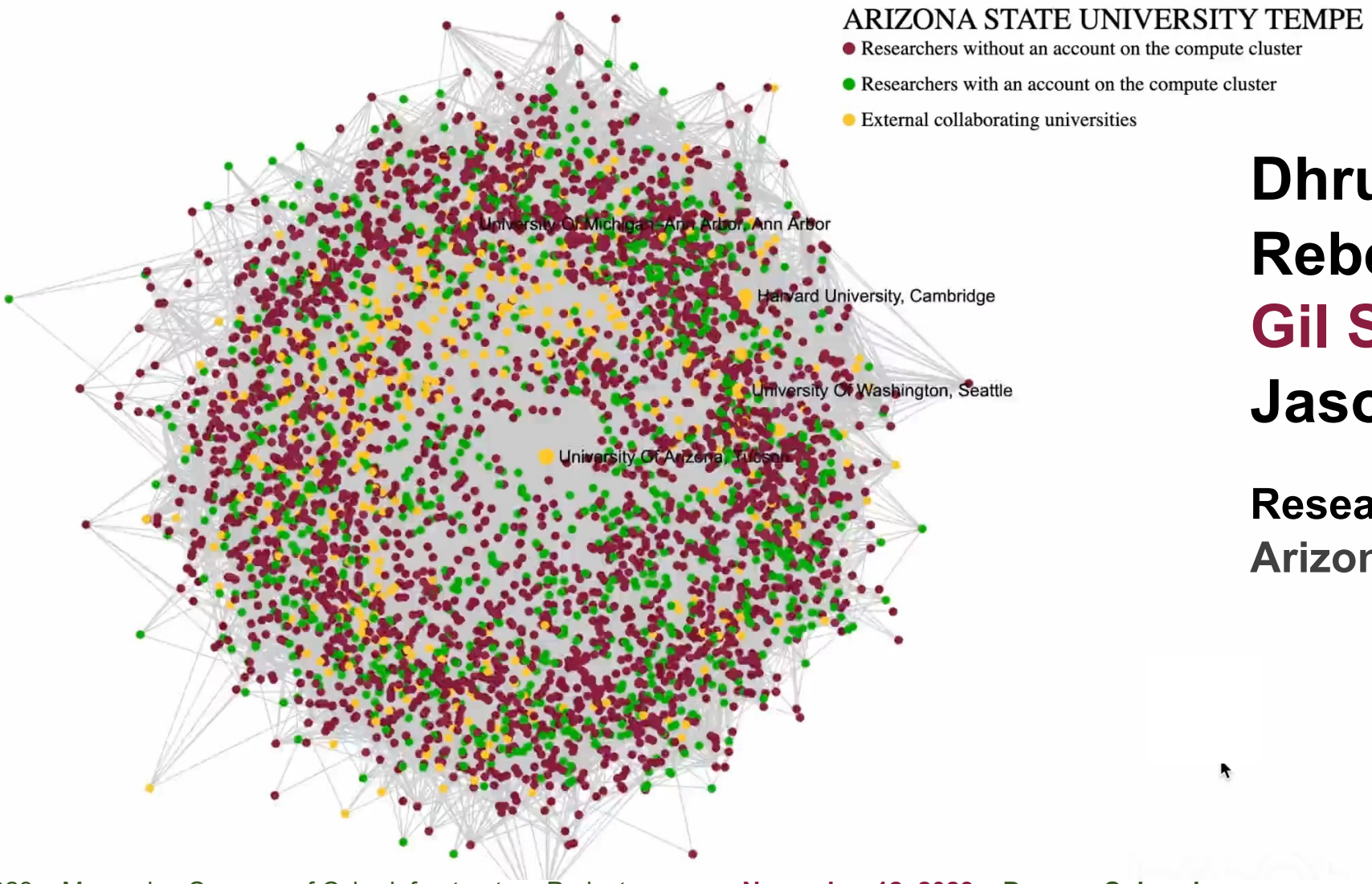
Measuring the impact of centralized high-performance computing with research collaboration networks



Dhruvil Shah
Rebecca Belshe
Gil Speyer
Jason Yalim

Research Technology Office
Arizona State University

ASU Knowledge Enterprise
Arizona State University

Motivation

- Recurring investment in large-scale computation maintains the need for various metrics to identify **ROI**
- **Collaboration** is increasingly made a priority by universities seeking to pursue larger and more various funding opportunities that require teams of investigators, specifically **interdisciplinary** collaboration
- Investigate of **temporal trends** reveal the impact of technology investment

Prior Work

- Economic production model
- Statistical analysis of research outputs to availability of local HPC resources
- Collaboration networks
  - KMAP
  - Relationship to productivity
  - Communities of practice

Dataset

- Dimensions data 2015-2022
  - 137M Publications: Coauthors
  - 7M Awards: Coinvestigators
  - 155M Patents: Coinventors

- Research Computing accounts

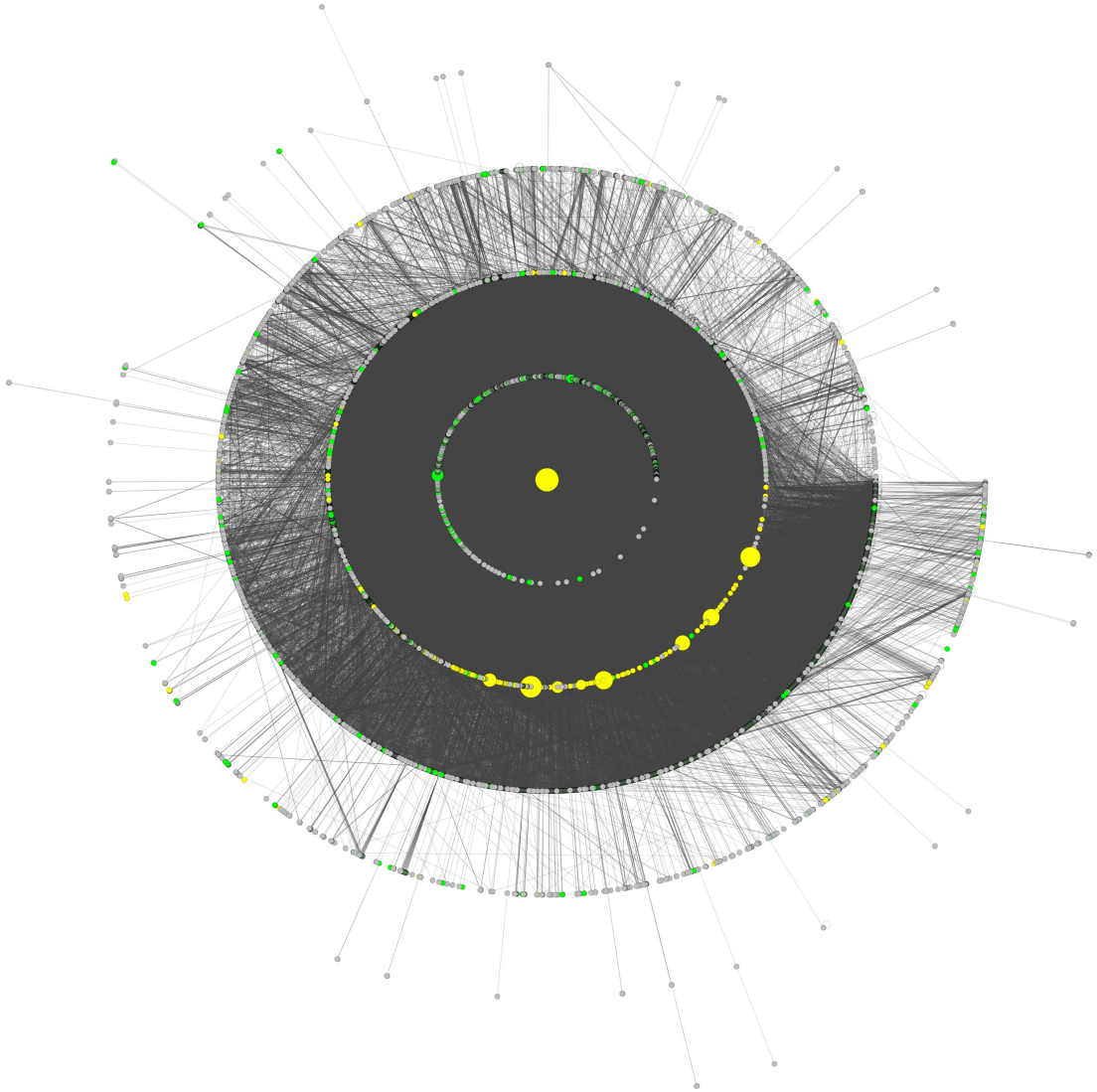
- Record of when ASU users registered for accounts

Internal network construction

- Researchers map to 14,606 nodes
- Collaborations map to 55,362 edges
- Caveats about interpretation of degree
- Researchers with RC accounts in green (2,032)

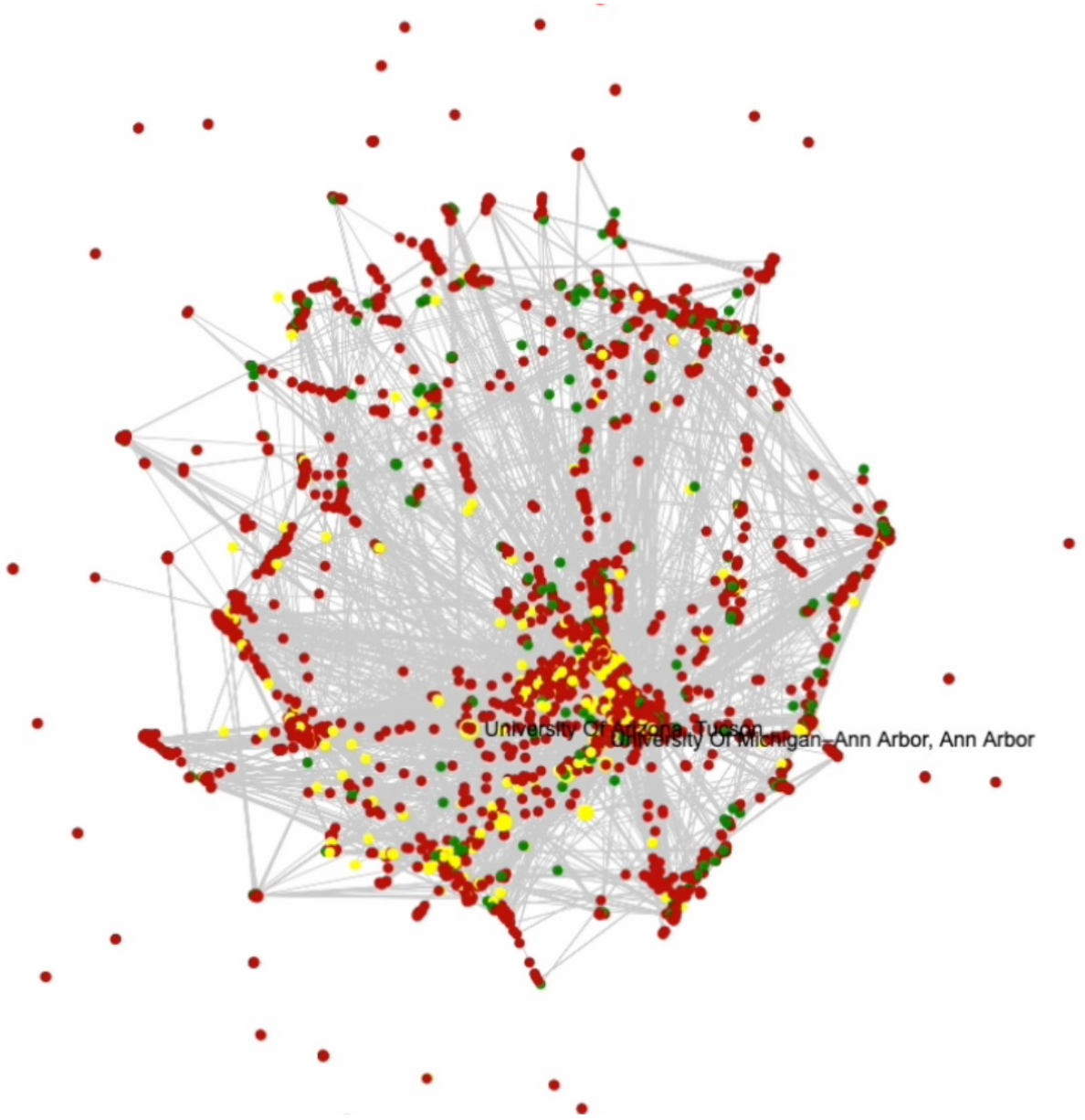
Including external collaborations

- Consolidate external collaborators into 587 "university" nodes
- ~98% now in Giant Component
- Edges increase from 55,362 to 104,394

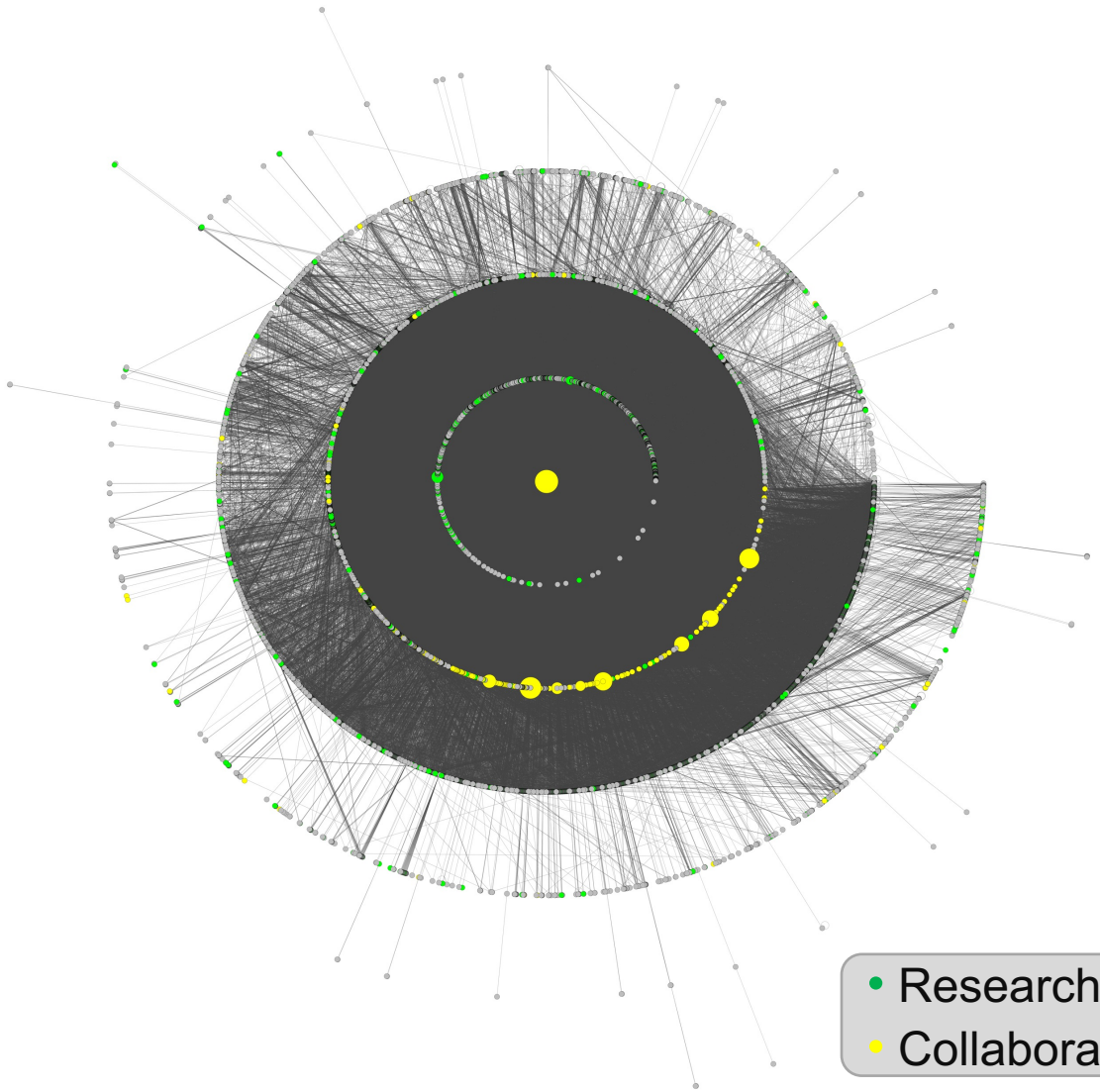


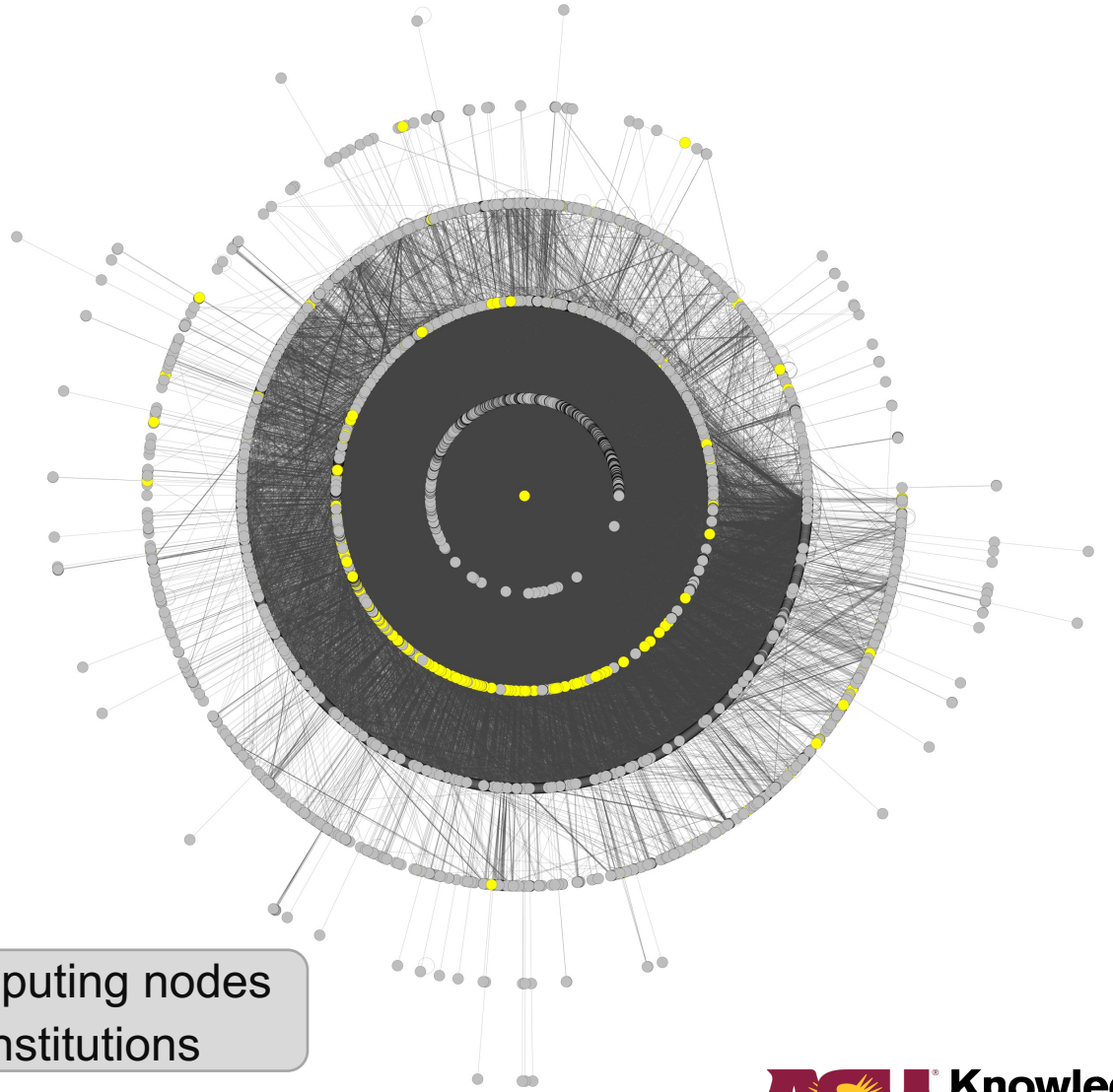
Small Worlds

- Network structure defined by 2 ratios to random Erdos-Renyi network
  - clustering coefficients
  - mean inter-node distance
- Robustness to node removal
- Small densely connected subnetworks with "**bridge nodes**"



Comparing node cohorts - node deletion

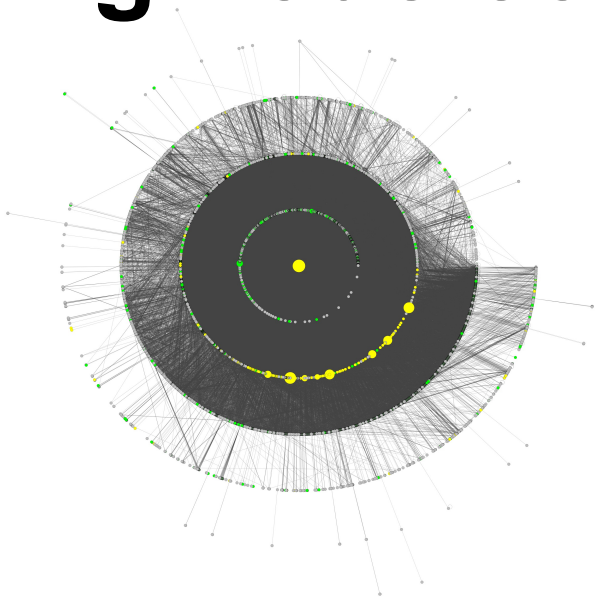


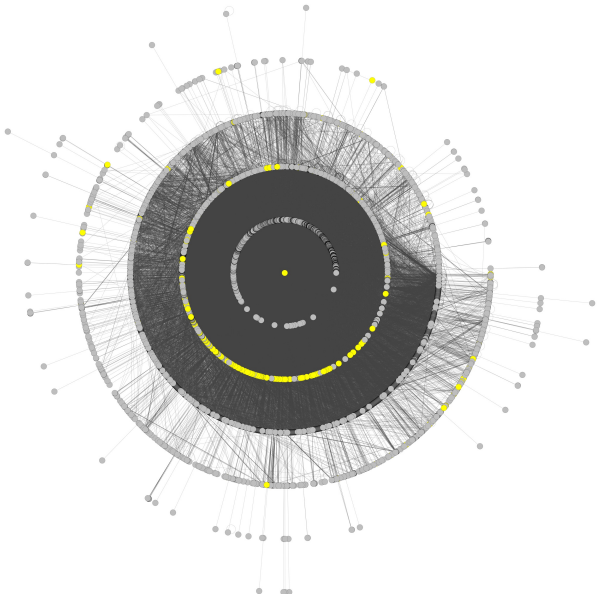


- Research Computing nodes
- Collaborating institutions

ASU Knowledge Enterprise
Arizona State University

Comparing node cohorts - node deletion



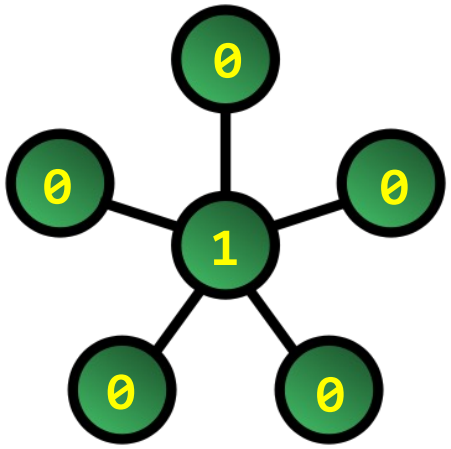


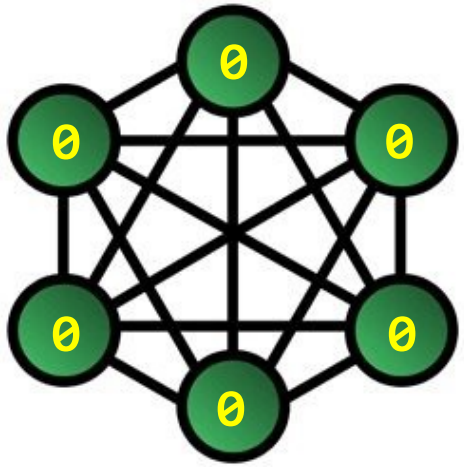
Network measures	Complete network	Research Computing (RC) nodes removed
Number of Nodes (including 587 Universities)	15 193	13 161
Number of Edges	105 840	70 941
Clustering Coefficient	0.54	0.48
Network Diameter	10	11
Average Shortest Path Length	3.59	3.80
Number of nodes in giant component (including 587 Universities)	14 947	12 923
Number of edges in giant component	105 632	70 740

Comparing node cohorts - centralities

Calculate measures of individual nodes

- Degree – raw count of edges
- Betweenness centrality – "hub"-ness of node





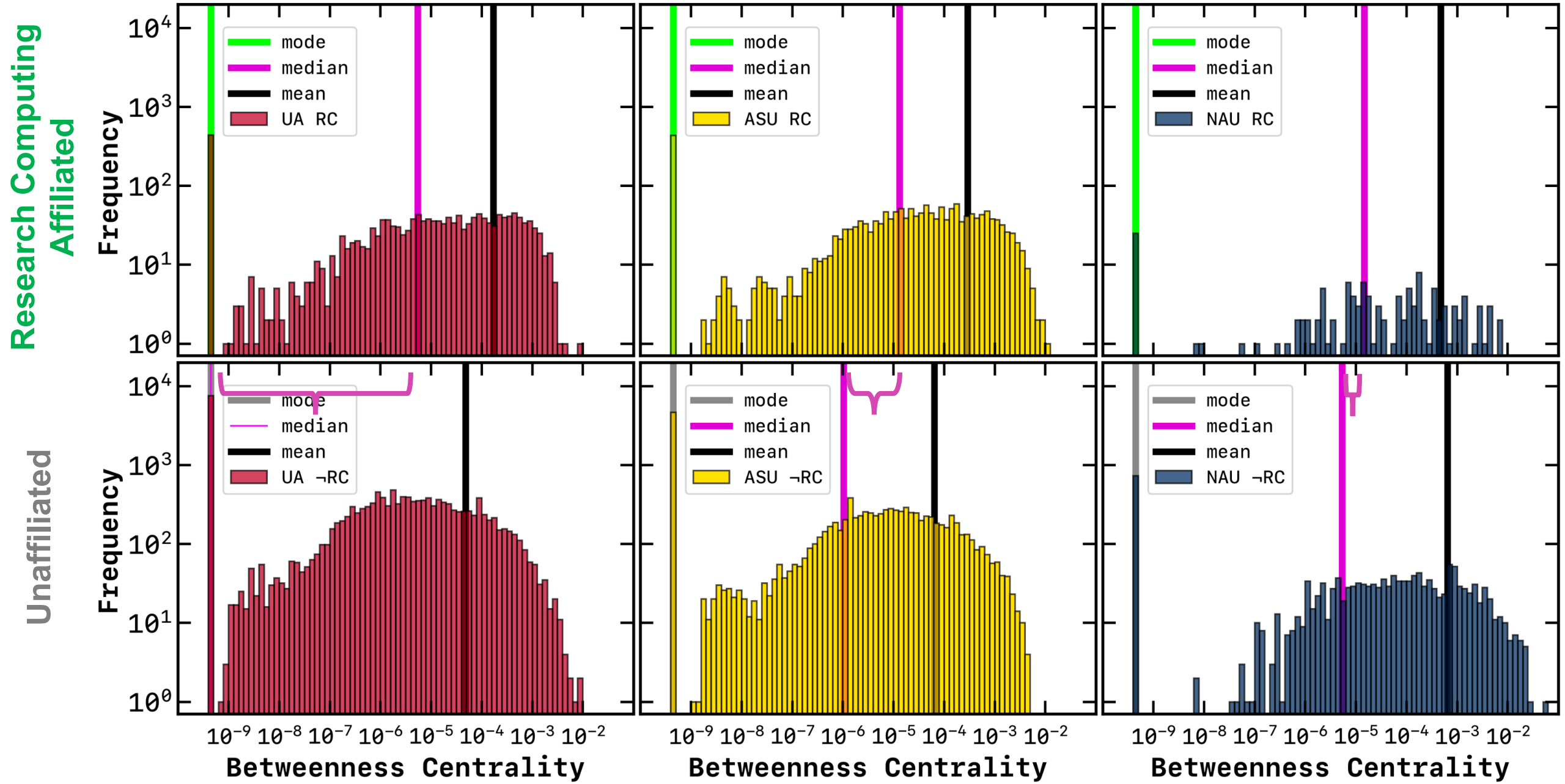
Comparing node cohorts - centralities

Calculate measures of individual nodes

- Degree – raw count of edges
- Closeness centrality – distance to all other nodes

	Mean		Median		
Measure	RC nodes	Non-RC nodes	RC nodes	Non-RC nodes	p-value
Degree	19.8	9.79	11.0	6.00	<0.001
Betweenness centrality	2.9×10^{-4}	6.4×10^{-5}	1.3×10^{-5}	1.0×10^{-6}	<0.001

Arizona Networks – U. of AZ, ASU, & NAU



- Analysis of "Research Software Engineers" – a group that is not formally centralized

COMPARISON OF DEGREE AND BETWEENNESS CENTRALITY MEASURES BETWEEN RSE AND NON-RSE COHORTS.

Mean statistic	**RSE nodes**	**Non-RSE nodes**	***p*-value**
Degree	15.63	11.09	<0.001
Betweenness centrality	8.1×10^{-5}	9.6×10^{-5}	<0.001

Network statistics over time

Network Measure	2015	2016	2017	2018	2019	2020	2021	2022
# Nodes	4173	4735	4475	4612	4519	4954	5080	4807
# RC Nodes	200	353	436	616	806	998	1126	1072
# Edges	13658	17258	15238	16390	15378	18985	20225	19511
Network Diameter	14	12	11	15	13	12	11	11
Clustering Coeff	0.51	0.53	0.50	0.52	0.51	0.53	0.51	0.52
Avg. Shortest Path Len.	4.65	4.46	4.51	4.50	4.50	4.34	4.23	4.29
% nodes in GC	86.77	87.65	87.15	87.06	87.05	89.56	89.15	88.64

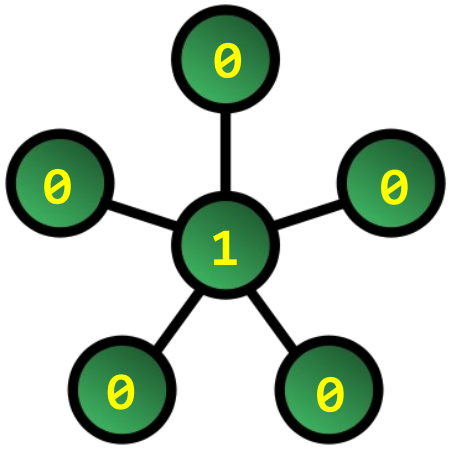
Network statistics over time

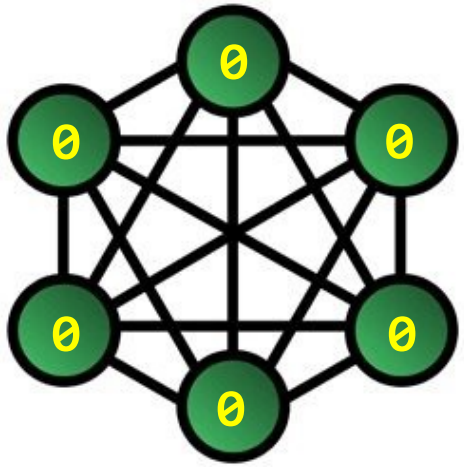
Network Measure	2015	2016	2017	2018	2019	2020	2021	2022
# Nodes	4173	4735	4475	4612	4519	4954	5080	4807
# RC Nodes	200	353	436	616	806	998	1126	1072
# Edges	13658	17258	15238	16390	15378	18985	20225	19511
Network Diameter	14	12	11	15	13	12	11	11
Clustering Coeff	0.51	0.53	0.50	0.52	0.51	0.53	0.51	0.52
Avg. Shortest Path Len.	4.65	4.46	4.51	4.50	4.50	4.34	4.23	4.29
% nodes in GC	86.77	87.65	87.15	87.06	87.05	89.56	89.15	88.64

Comparing node cohorts - centralities

Calculate measures of individual nodes

- Degree – raw count of edges
- Betweenness centrality – "hub"-ness of node

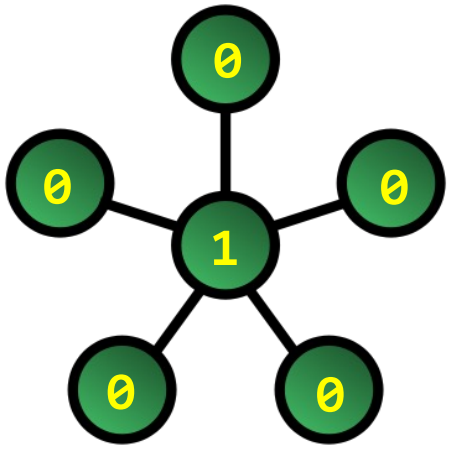


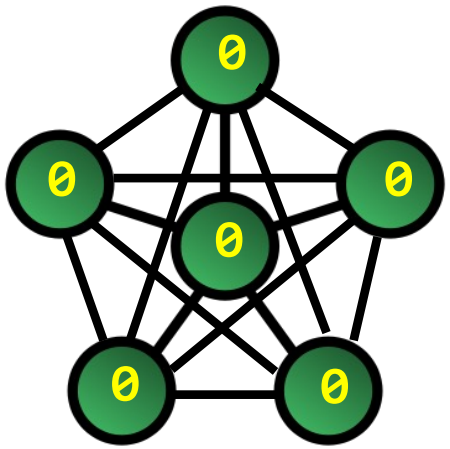


Comparing node cohorts - centralities

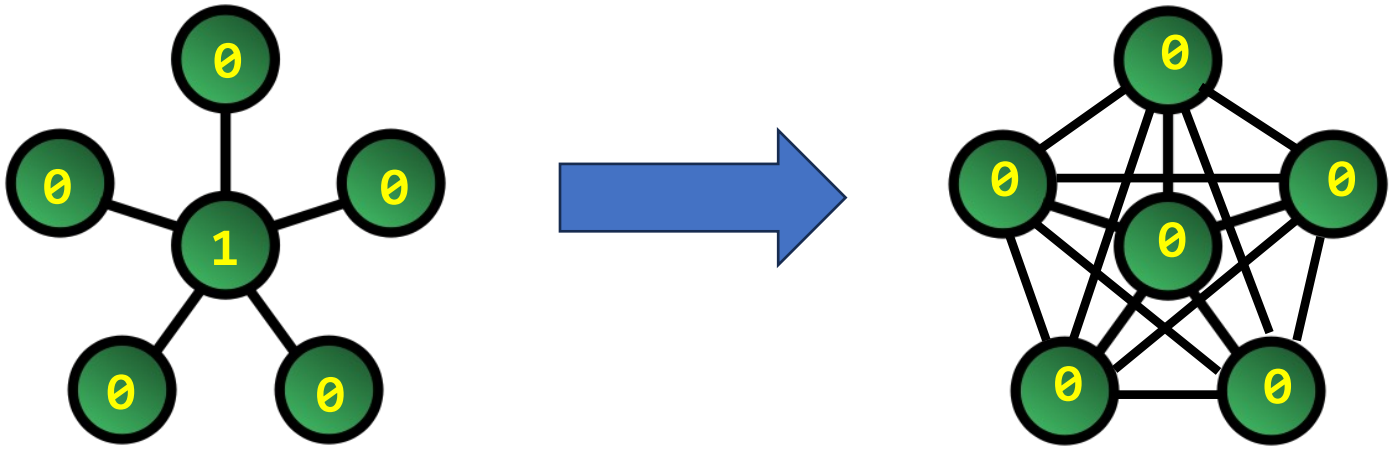
Calculate measures of individual nodes

- Degree – raw count of edges
- Betweenness centrality – "hub"-ness of node

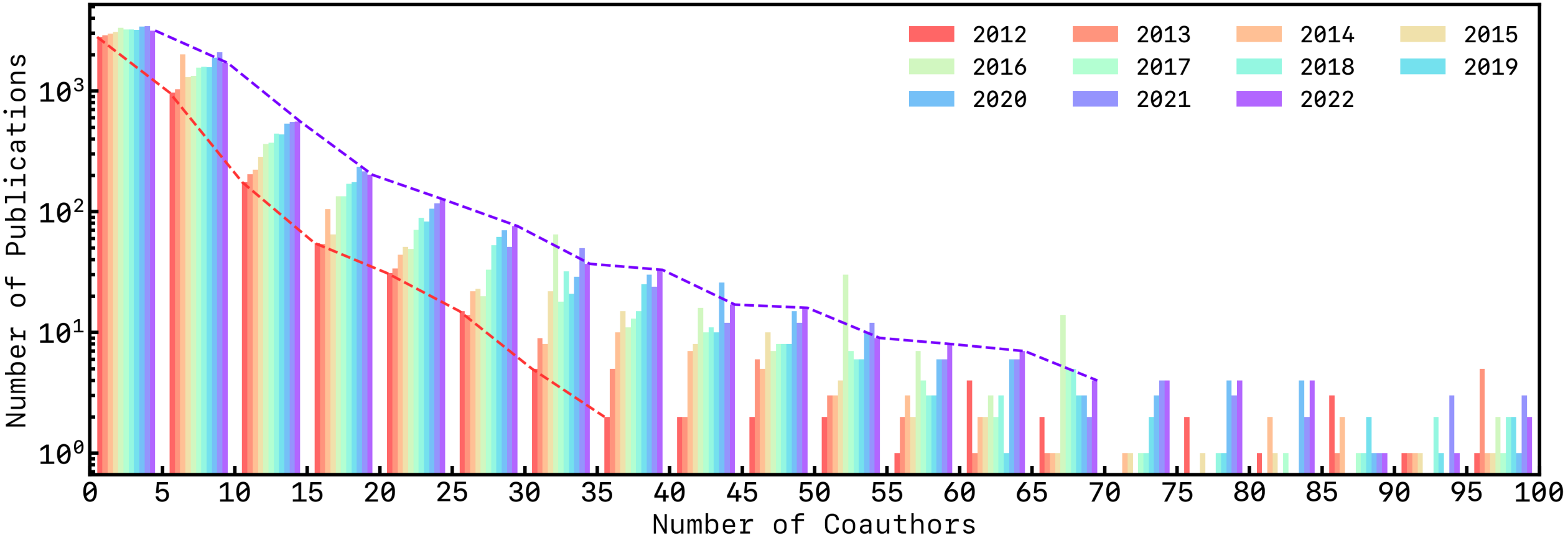




Centrality trends over time

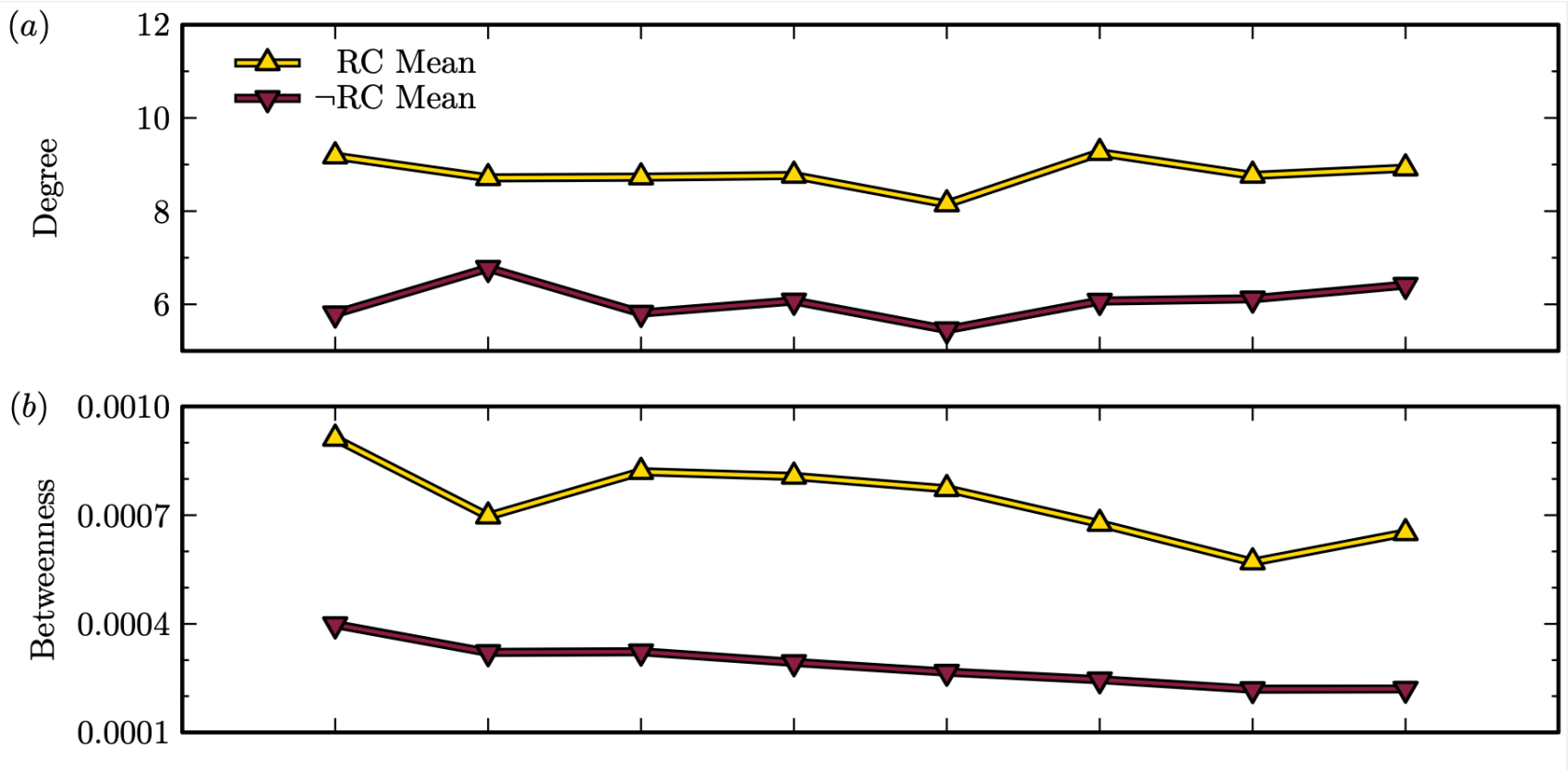


Dataset over time



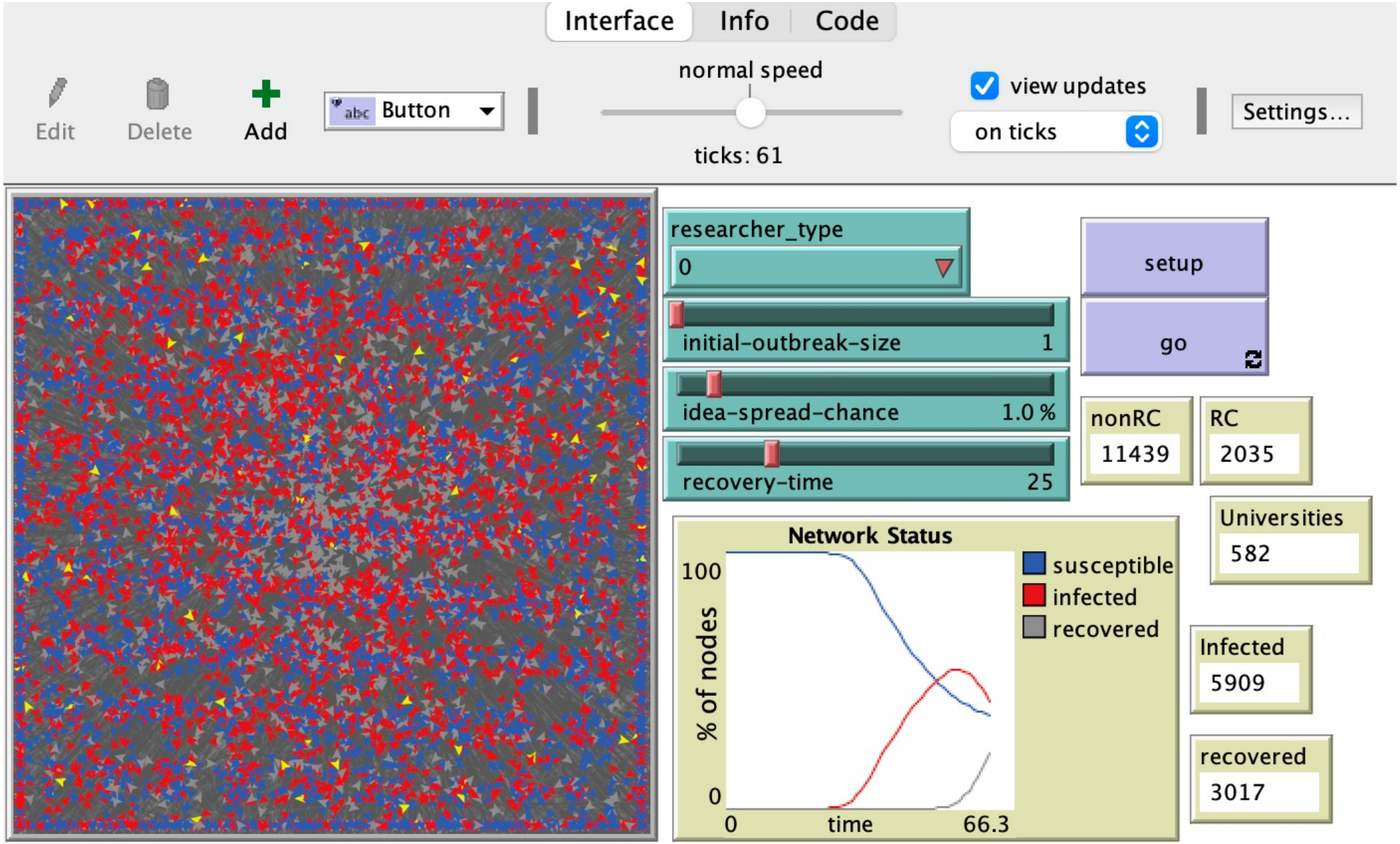
ASU coauthorship 2012-2022

Centrality trends over time

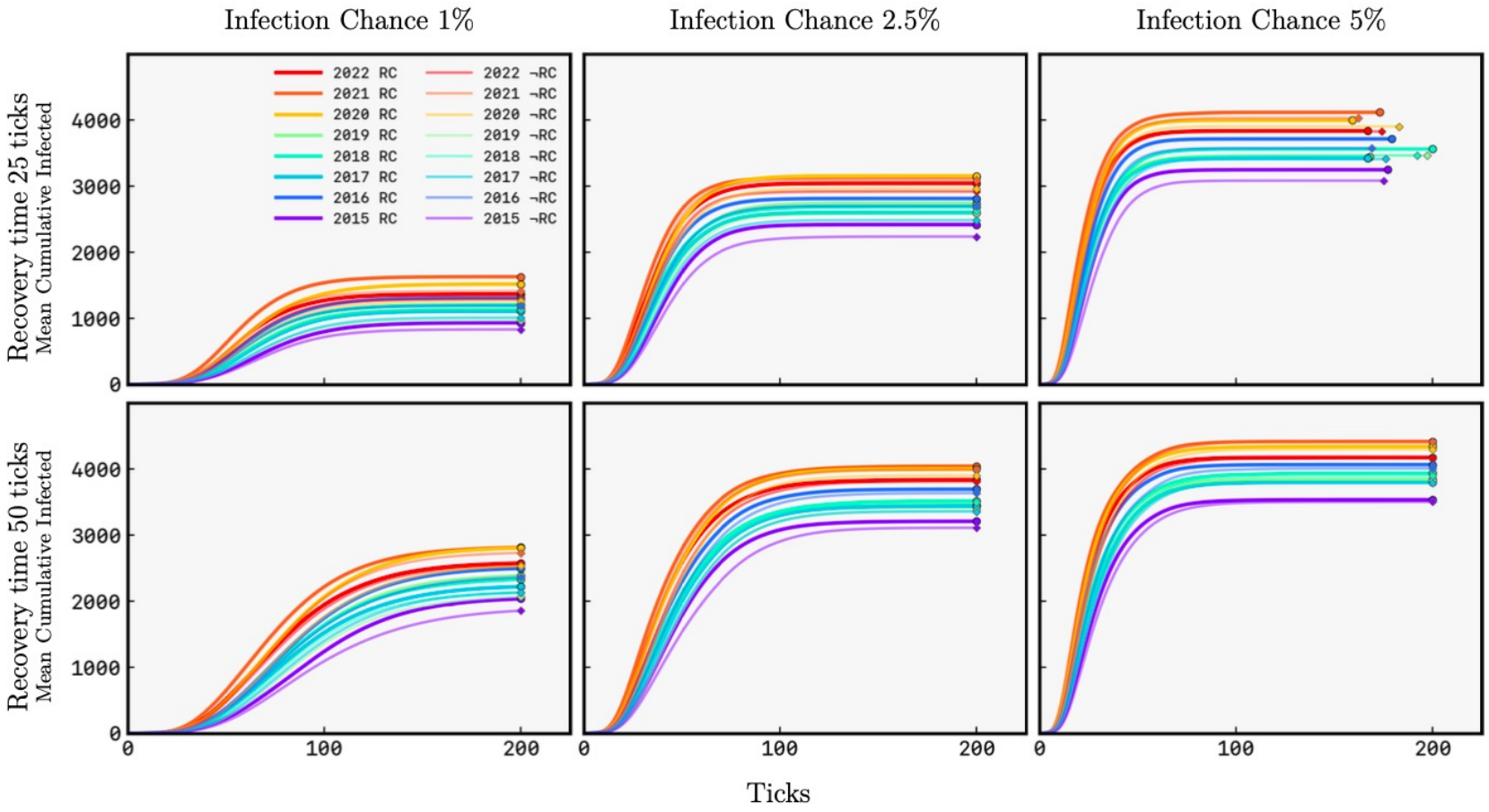


Agent-based modeling

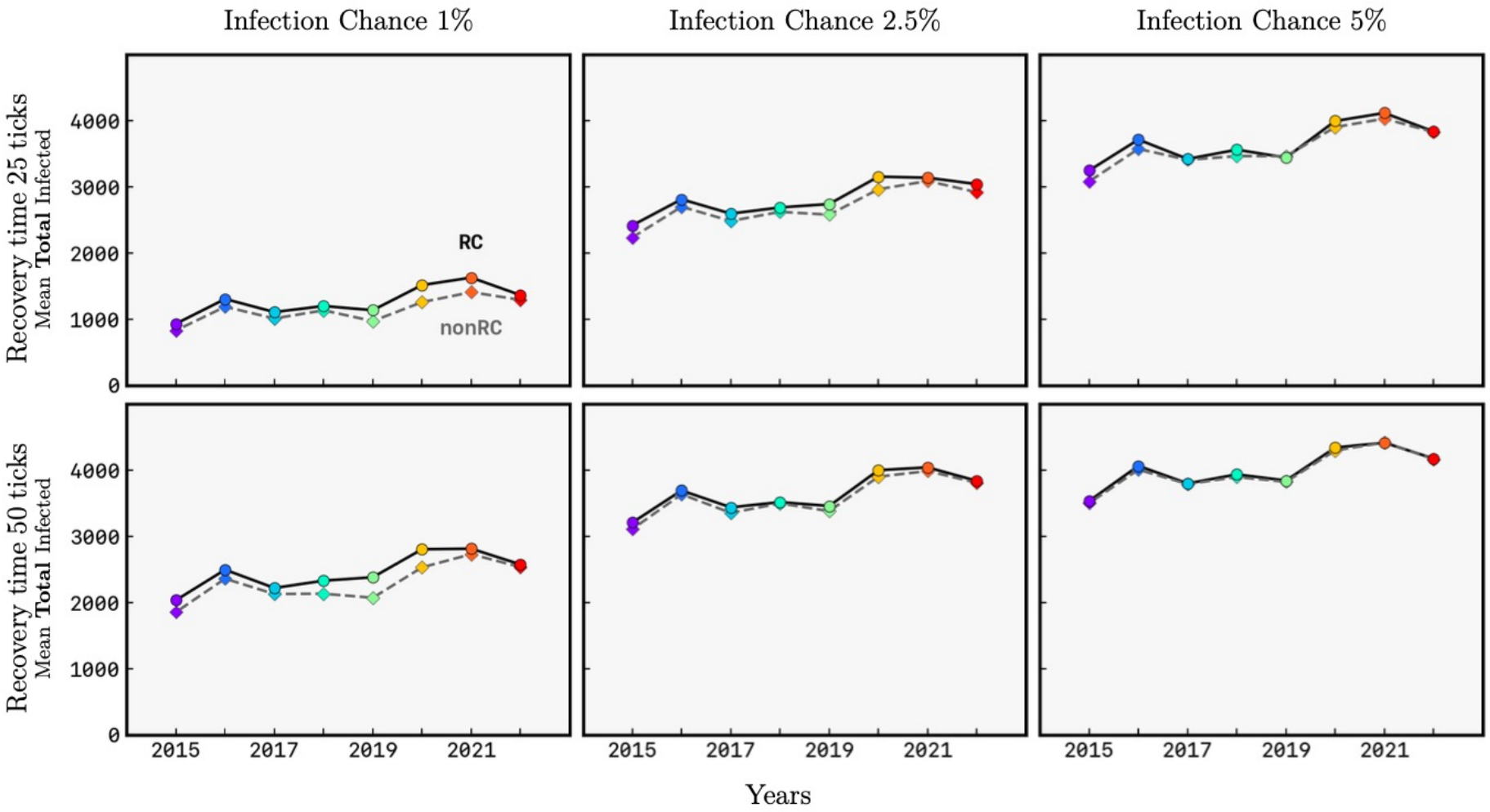
Measure the *spread of an idea* using an SEIR model



Spread of an idea over time



Spread of an idea over time



Acknowledgments

- Collaboration and interdisciplinarity metrics
- Network analysis over time