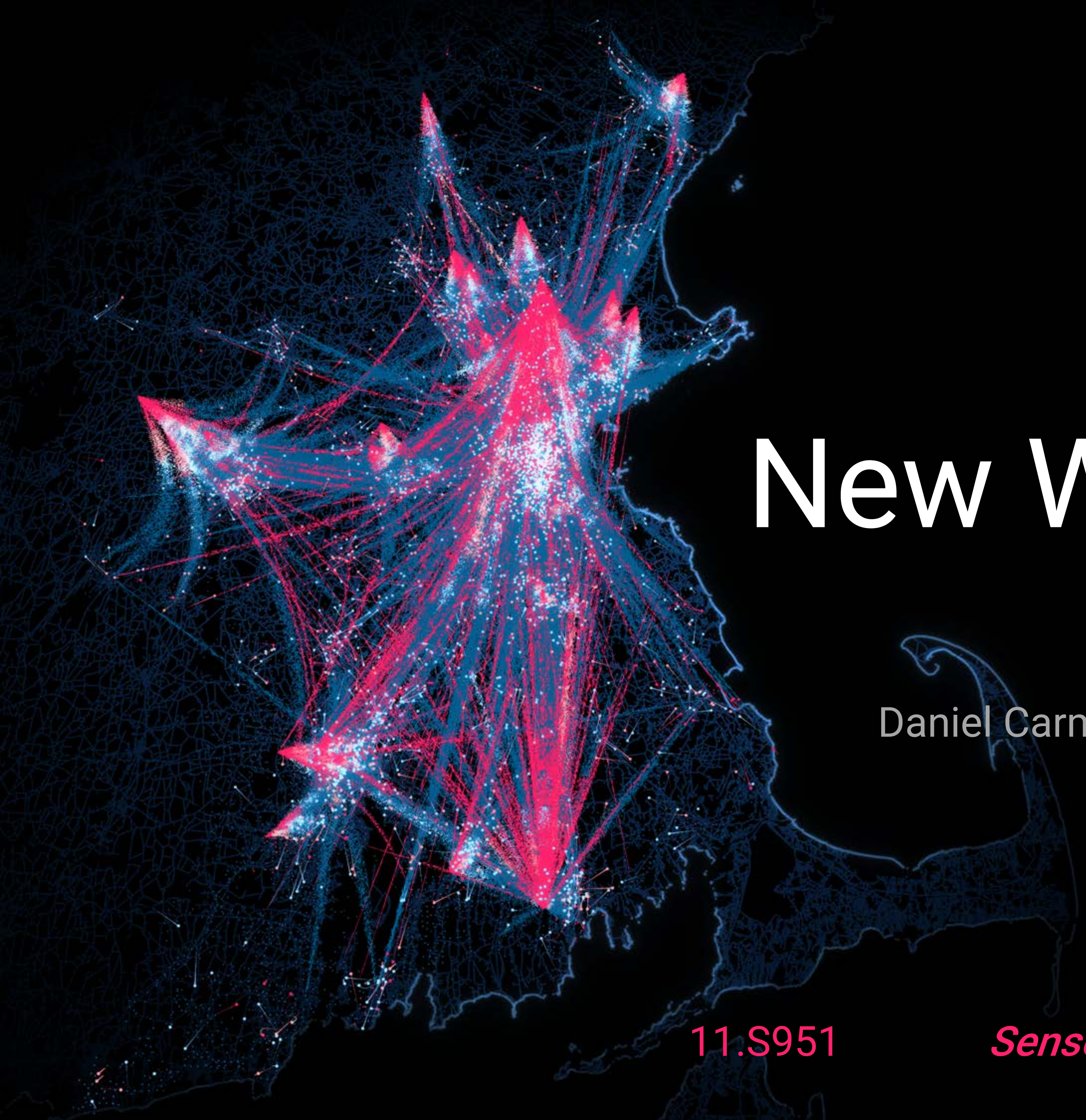


Week 4

New Ways of Living

Daniel Carmody, Martina Mazzarello, Simone Mora

Learning objectives

SENSING THE ENVIRONMENT

Past and contemporary possibilities of scanning the environment

PHYSICAL- DIGITAL LAYERS

Scales, tools, applications to change our ways of living

SENSING HUMAN CONNECTIONS

Digital traces of people's communication in a campus environment

SOCIAL NETWORKS' THEORY

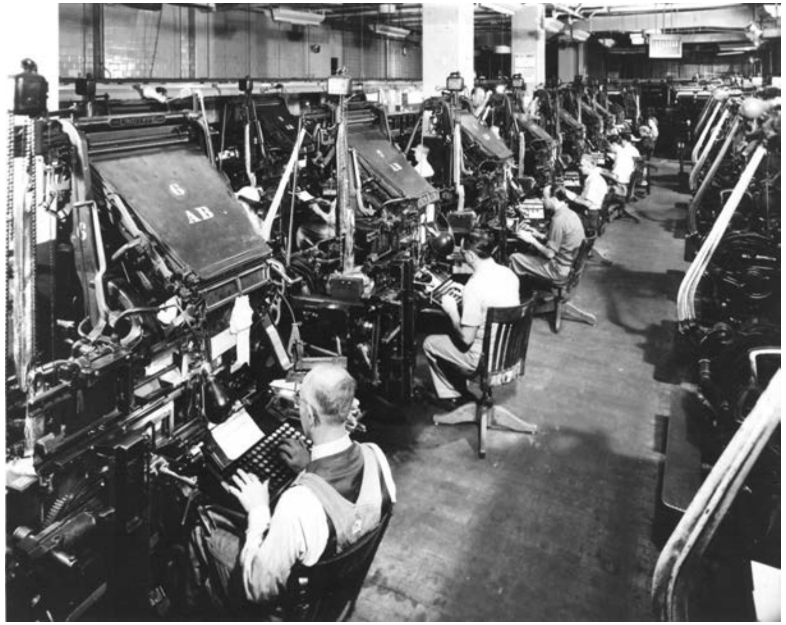
Definitions and applications of network science

NEW WAYS OF LIVING

The view to the south from the Empire State Building on Nov. 24, 1966, one of New York's worst smog days. Photo NYT.

Kansas City during the late 60s affected by both industrial pollution and car smog. Photo EPA Archive.

NEW WAYS OF LIVING



Before the rise of digital technologies, there were specific types of buildings, factories or offices for every occupation: a newspaper, for instance, needed a pressroom, a printing room, and all sorts of equipment to get the paper out on the street every day.

NEW WAYS OF LIVING

A House in a Box You Control by waving Your Hand, a way to turn any small apartment into a more livable one. A project of the MIT Media Lab (2011).

NEW WAYS OF LIVING

Manuel Castells (1950 – 2000) the rise of a digital age society defined by “[...] new forms of spatial arrangements”. With the Digital revolution (2000), Work and leisure in post-industrial cities don’t need a particular spatial configuration anymore

How data can support new ways of living?

How are we tracking these information?

Evolution of sensing

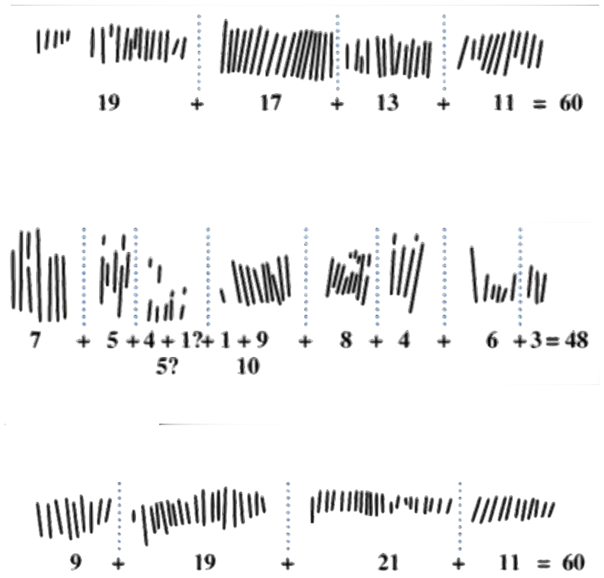
Historical and contemporary efforts in documenting our lives

NEW WAYS OF LIVING

18.000 BCE – 800s
Humans as sensors

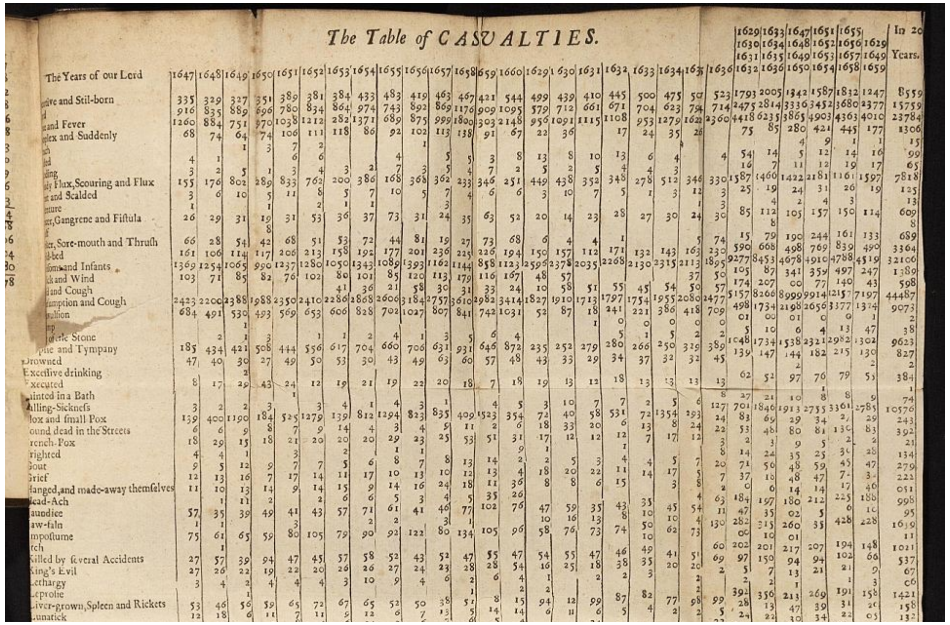
900s - 2022
Analog and Digital sensors

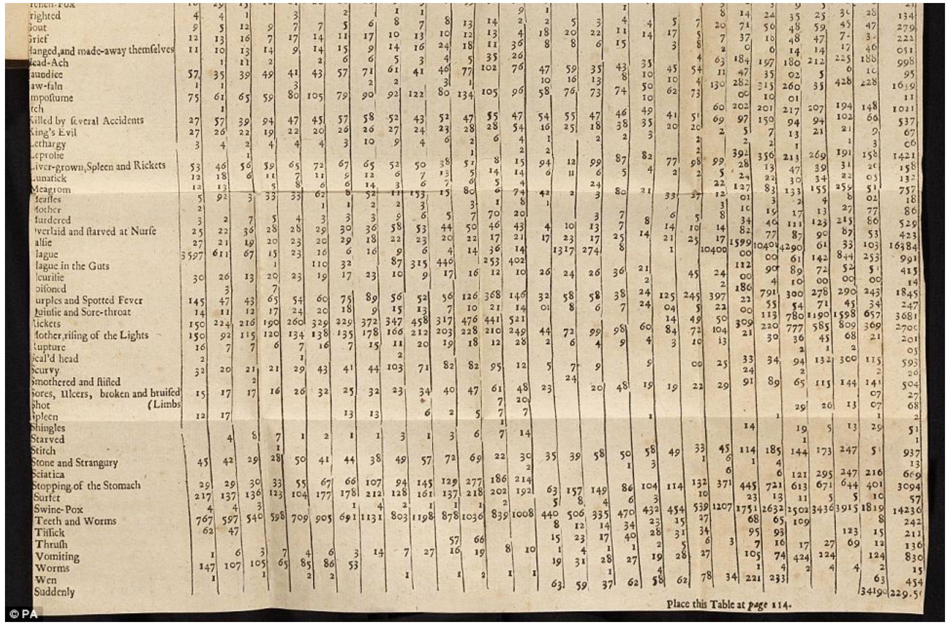
Human as sensor



18.000 BCE, Tally sticks

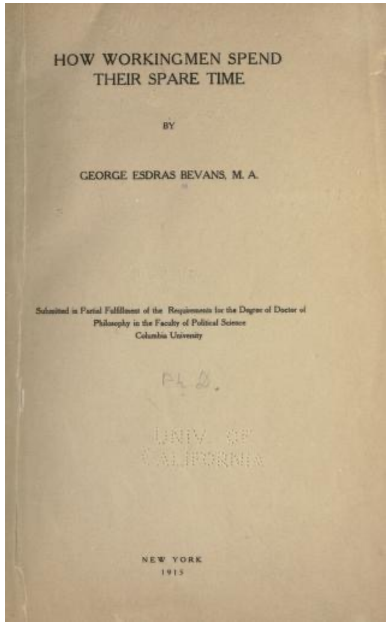
Historical evolution





1663, John Graunt carries out the first recorded experiment in statistical data analysis, data to provide a warning system for bubonic

Historical evolution


HOW WORKINGMEN SPEND THEIR SPARE TIME

BY

GEORGE ESDRAS BEVANS, M. A.

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Faculty of Political Science
Columbia University

NEW YORK
1913

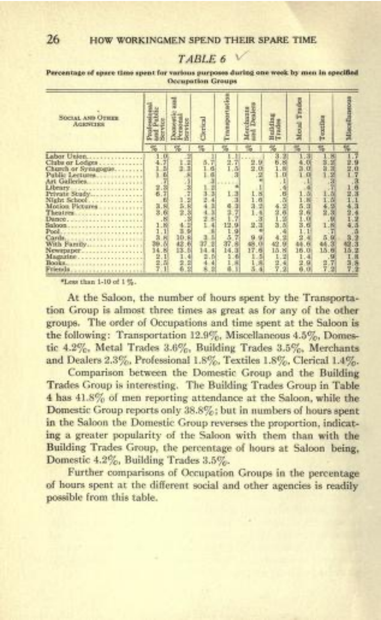
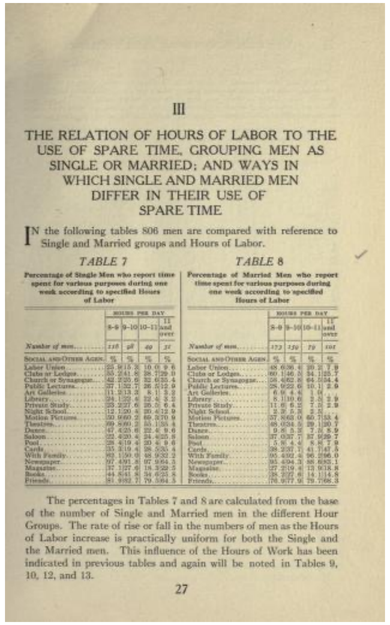

26 HOW WORKINGMEN SPEND THEIR SPARE TIME

TABLE 6


III

THE RELATION OF HOURS OF LABOR TO THE USE OF SPARE TIME, GROUPING MEN AS SINGLE OR MARRIED; AND WAYS IN WHICH SINGLE AND MARRIED MEN DIFFER IN THEIR USE OF SPARE TIME

TABLE 7

TABLE 8

27

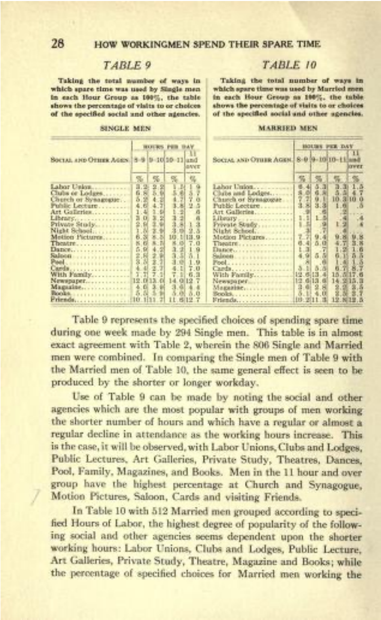

28 HOW WORKINGMEN SPEND THEIR SPARE TIME

TABLE 9

TABLE 10

Bevans’ 1913 Columbia University doctoral thesis on London factory workers, annotating information about working hours and spare time.

https://archive.org/details/howworkingmenspe00bevarich/page/28/mode/2up

Environmental Sensors

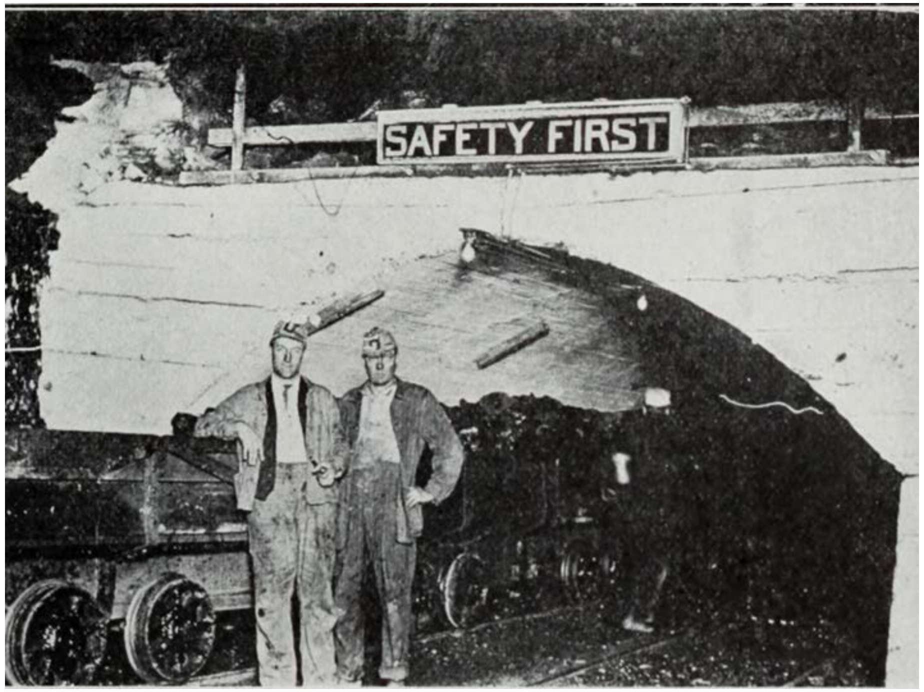


Photo Penn State University Archive

100TH ALWAYS AMERICA FIRST ANNIVERSARY

Chicago Daily Tribune

THE WORLD'S GREATEST NEWSPAPER

42 PAGES CITY FINAL

VOLUME CVI.—NO. 73 C | WEDNESDAY, MARCH 26, 1947 | FOUR CENTS—PAY NO MORE

73 IN MINE! 23 DEAD

Explosion Rips Illinois Pit; 1st Survivor's Story

SENATORS ACT TO ABOLISH LILIENTHAL JOB

Would Give Post to Marshall

BY WILLIAM MOORE

Call for Global Slant in U. S. School Books

BY ROBERT YOUNG

Philadelphia, March 25—

ADMITS GREEK 'CRISIS' KNOWN 5 MONTHS AGO

Told of British Plan Last Fall: Acheson

BY WILLARD EDWARDS

Washington, March 25 —

THE FRONT MAN

House Votes NLRB Slash, Fires Warren

BY JOHN FISHER

Washington, March 25—

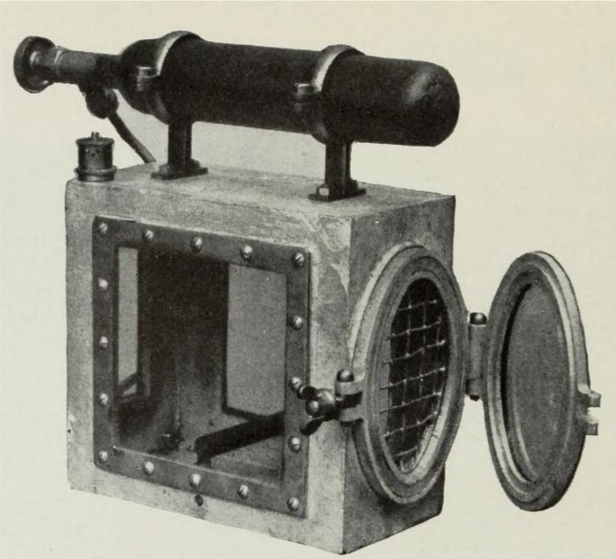
RESCUE CREW BATTLES GAS 540 FEET DOWN

Help Rushed to Disaster Scene

Centralia, Ill., March 26—

Photo Chicago Tribune

First air quality sensor

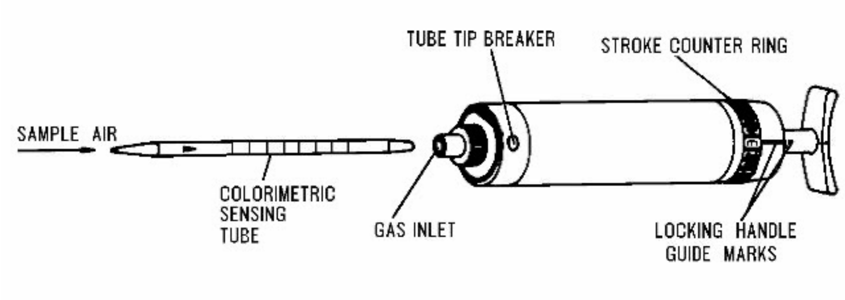


A CAGE FULL OF OXYGEN

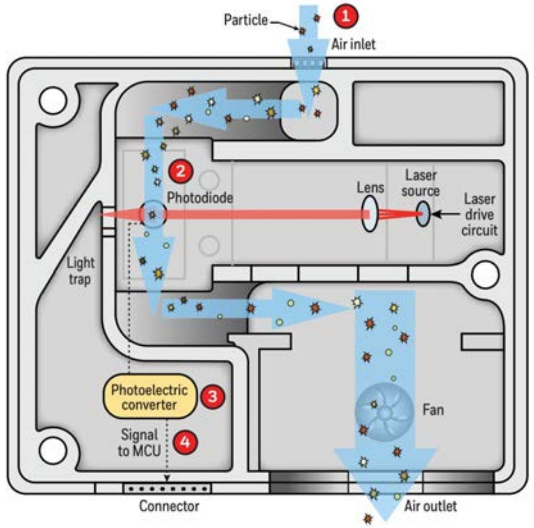
TO REVIVE CANARIES THAT ARE OVERCOME BY GAS IN COAL MINES

Mining worker in the 800s Christal Pollock "The Canary in the Coal Mine," Journal of Avian Medicine and Surgery 30(4), 386-391
https://doi.org/10.1647/1082-6742-30.4.386

First air quality sensor



One of the first portable carbon monoxide sensor. 1937 Drager company.



Light-scattering-based PM sensor, 2022. Treckview.org.

One commercially avalable electrochemical CO sensor. 2022. Alphasense.co.uk

Sensing platforms

smartcitizen.me

purpleair.com

Historical evolution

New world of growing data

How to untap the potential of data
to support new ways of living?

Sensing the environment

The view to the south from the Empire State Building on Nov. 24, 1966, one of New York’s worst smog days. Photo NYT.

Kansas City during the late 60s affected by both industrial pollution and car smog. Photo EPA Archive.

City Scanner

ENVIRONMENT, URBAN SENSING

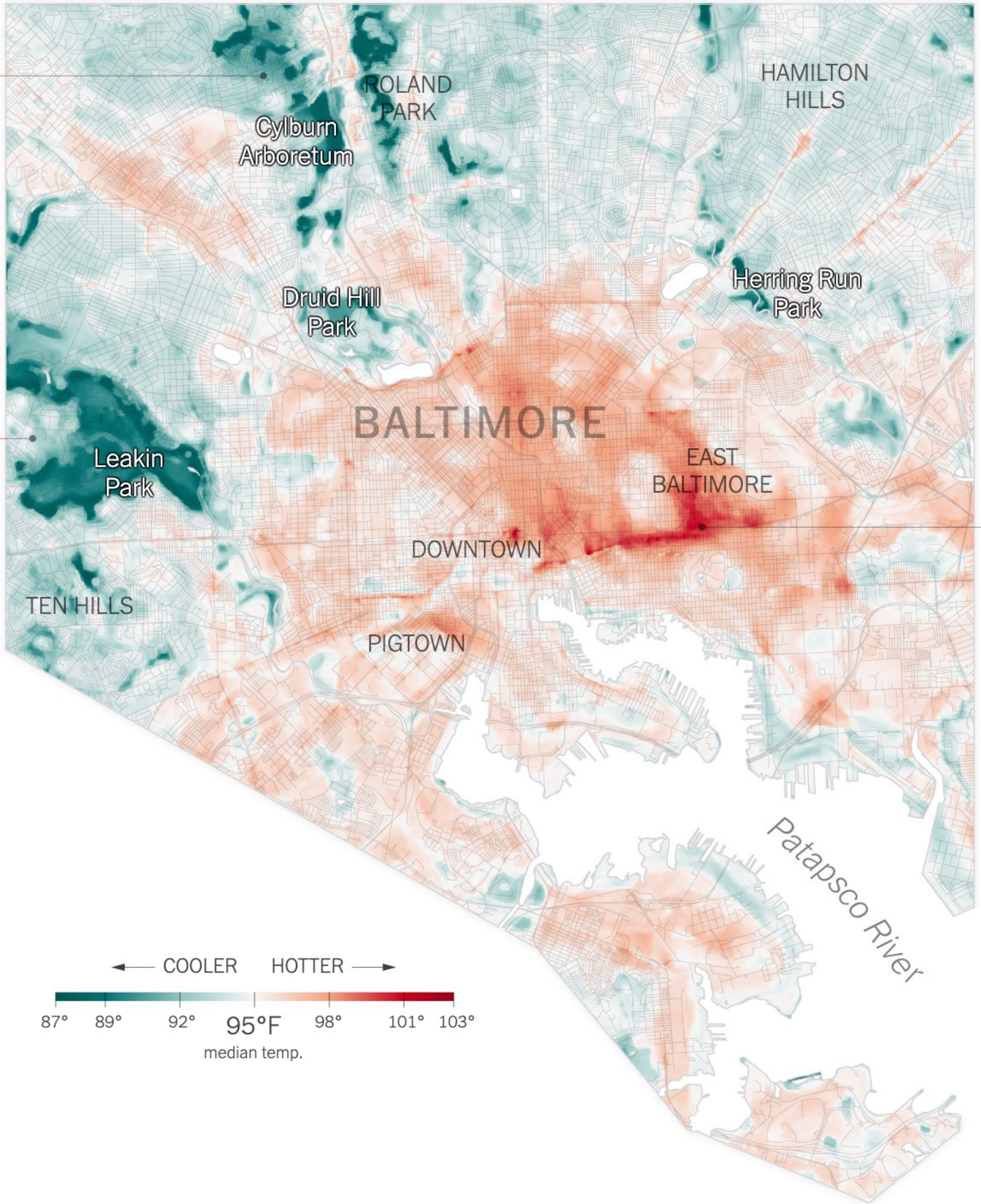
AIR POLLUTION

- 92% of the world population breath unhealthy air (WHO)
- Short term: asthma, cardiovascular diseases
- Long term: cognitive decline and Alzheimer's disease (Killian & Kitazawa, 2018)

- Costs more than US$5 trillion (Word Bank)
- In London, poor AQ leads to 650,000 sick days a year (Kilbane-Dawe et al., 2014)
- Spanish consumers spend up to $50M less on days with poor AQ (Rogers et al., 2016)

- Can vary up to 8x within the same city block

Cooler: Neighborhoods next to parks and those with plenty of tree cover saw significantly cooler temperatures on a hot summer afternoon: **as low as 87°F.**



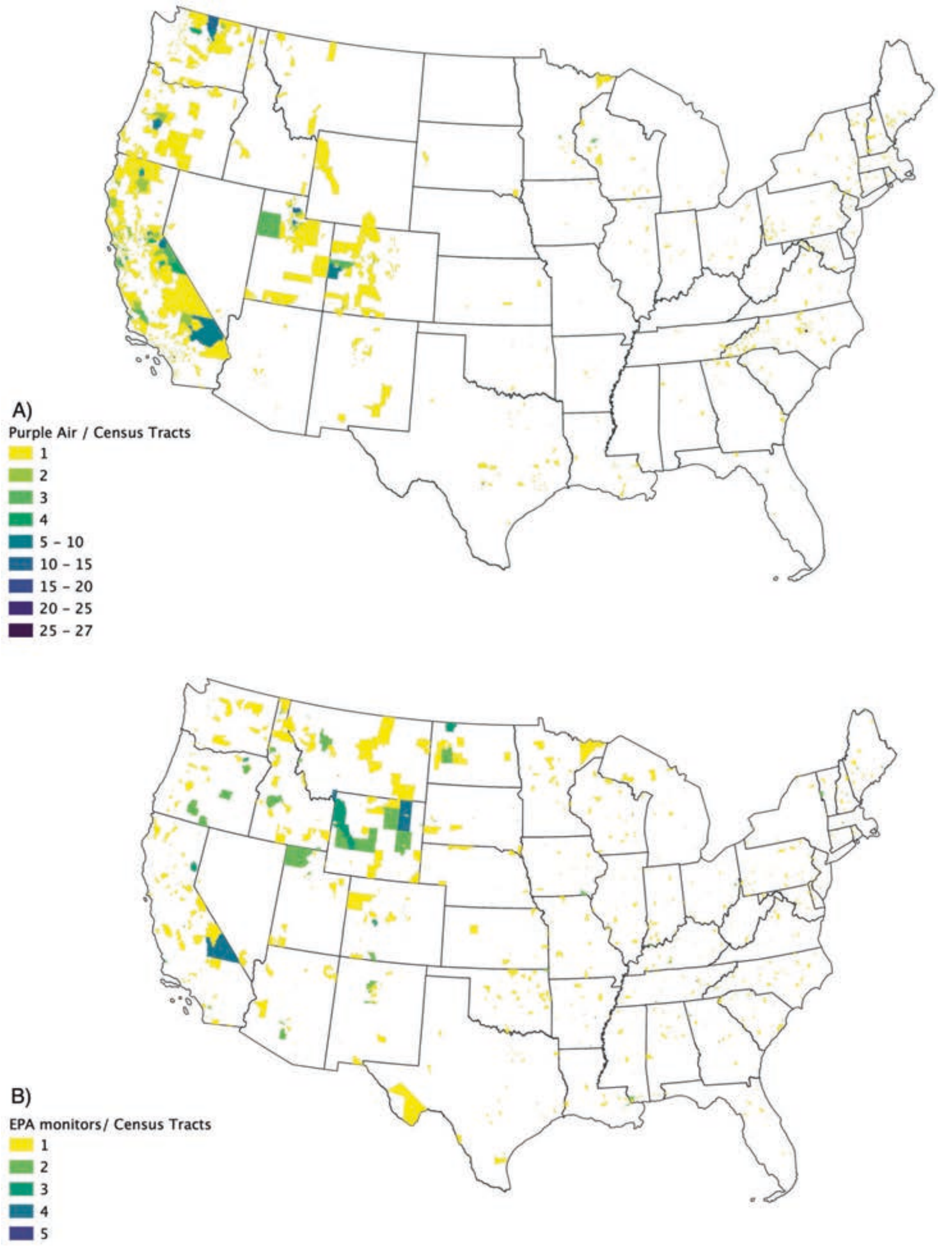
Hotter: On the same day, residential neighborhoods east of downtown saw hotspots reach **over 101°F.**

Nadja Popovich and Christopher Flavelle
The New York Times

EXTREME HEAT

- Associated with higher rates of cardiovascular diseases, cancer
- Compounds the negative effects of air pollution
- Widely varies as a function of socioeconomic status and race/ethnicity
- Every year extreme heat events kill more Americans than other extreme weather combined

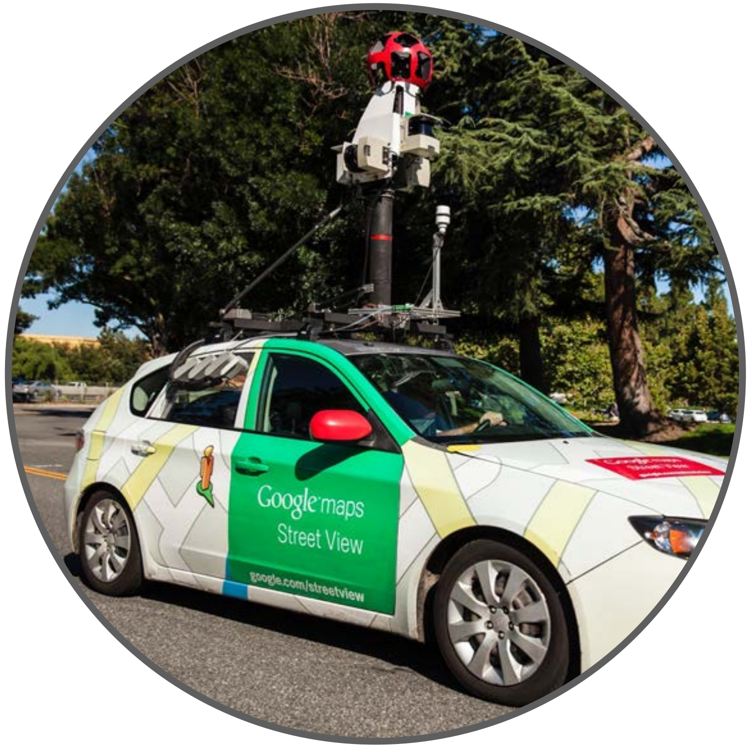
Fig. 1 Maps showing the distributions of PurpleAir and EPA mointors. a Number of PurpleAir sensors/census tract in the United States as of Feb 22, 2020. **b** Number of EPA monitors that report $PM_{2.5}$ from 2015 to Feb 22, 2020 per census tract in the United States Only census tracts with monitors are shown in this analysis.

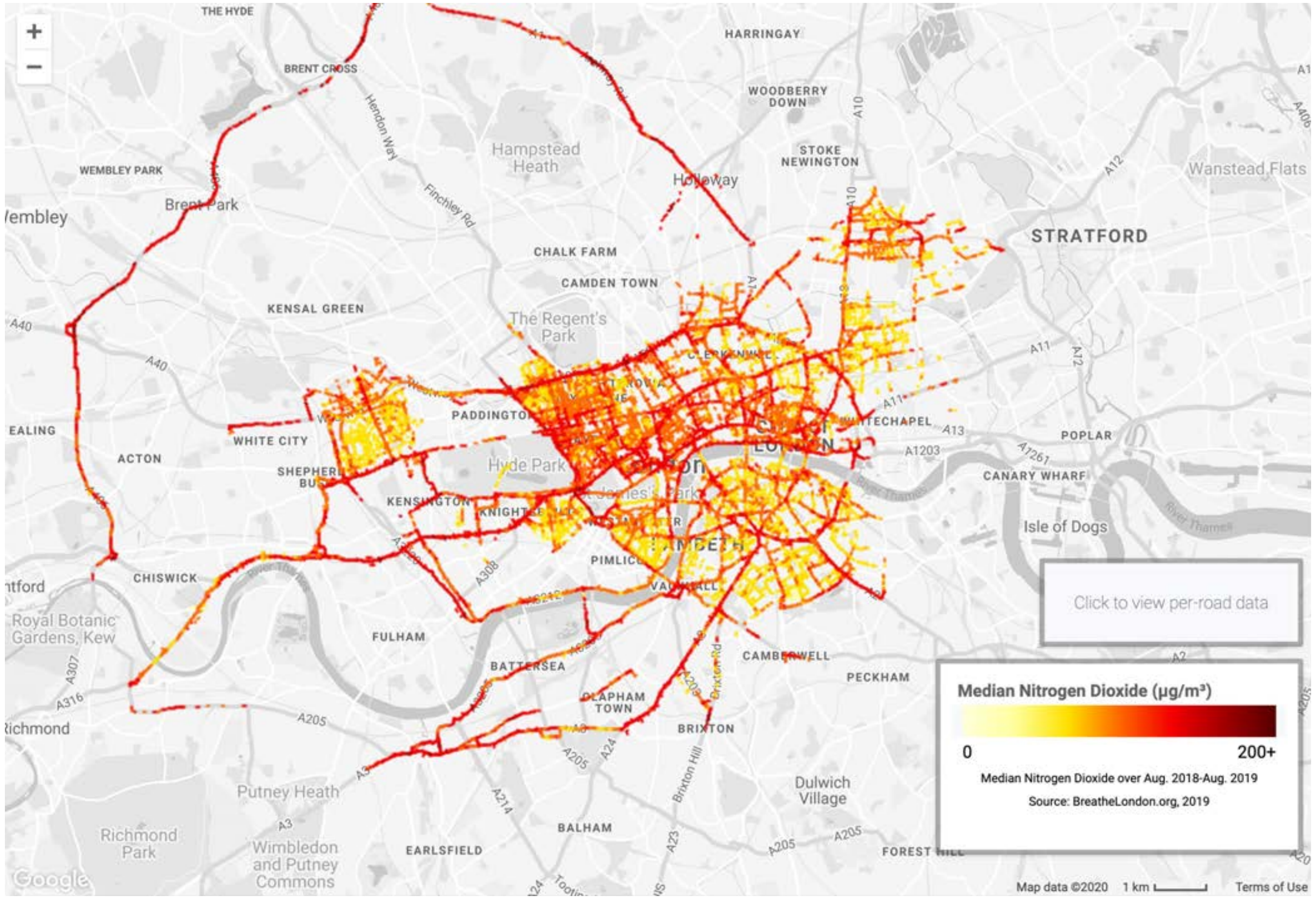


deSouza, Priyanka, and Patrick L. Kinney. "On the distribution of low-cost PM2. 5 sensors in the US: demographic and air quality associations." *Journal of exposure science & environmental epidemiology* (2021)

Can we use mobile sensors to
map environmental data cities?

ISTOCK.COM/BANKSPHOTOS

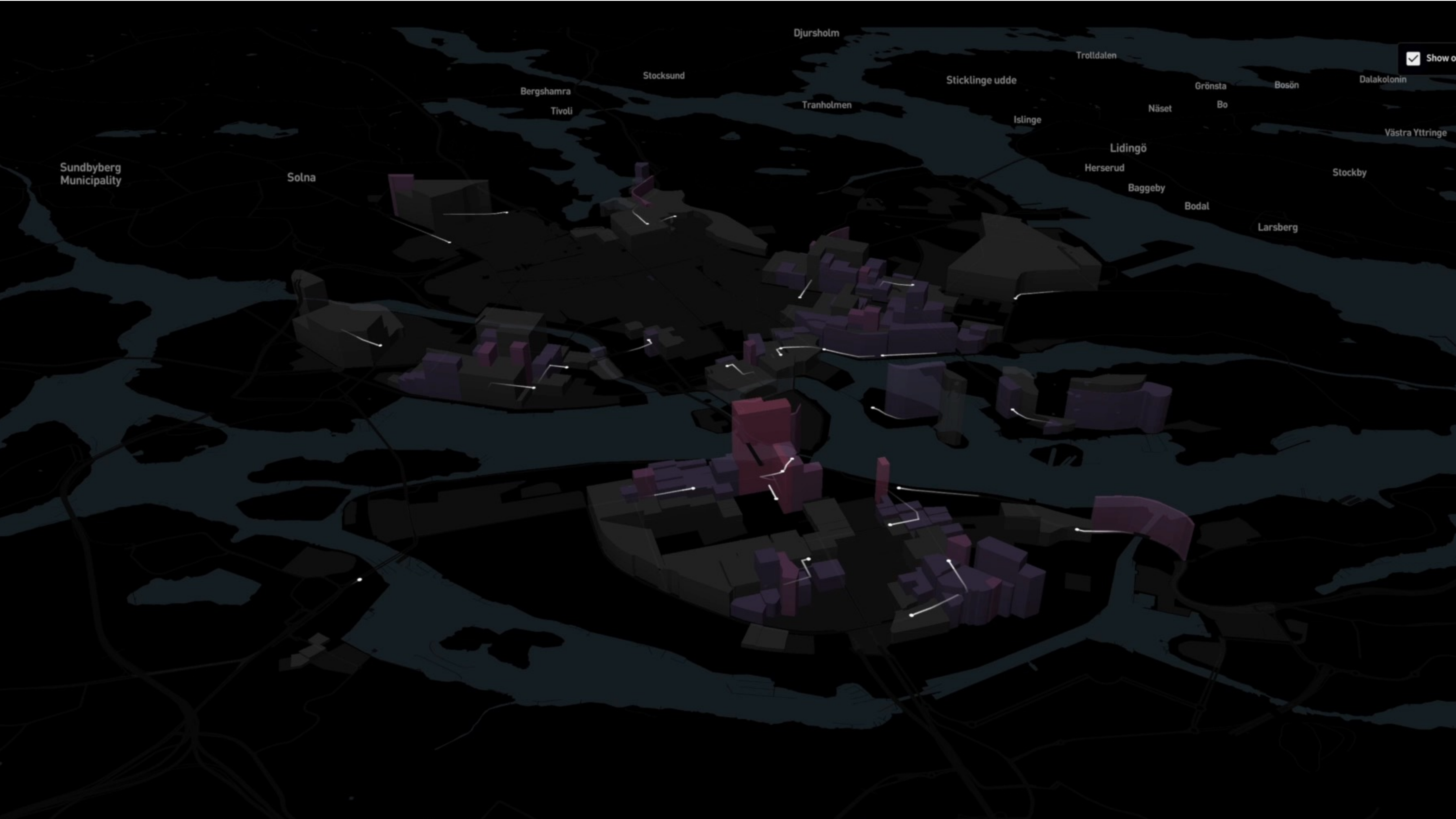




Lab-on-wheels approach

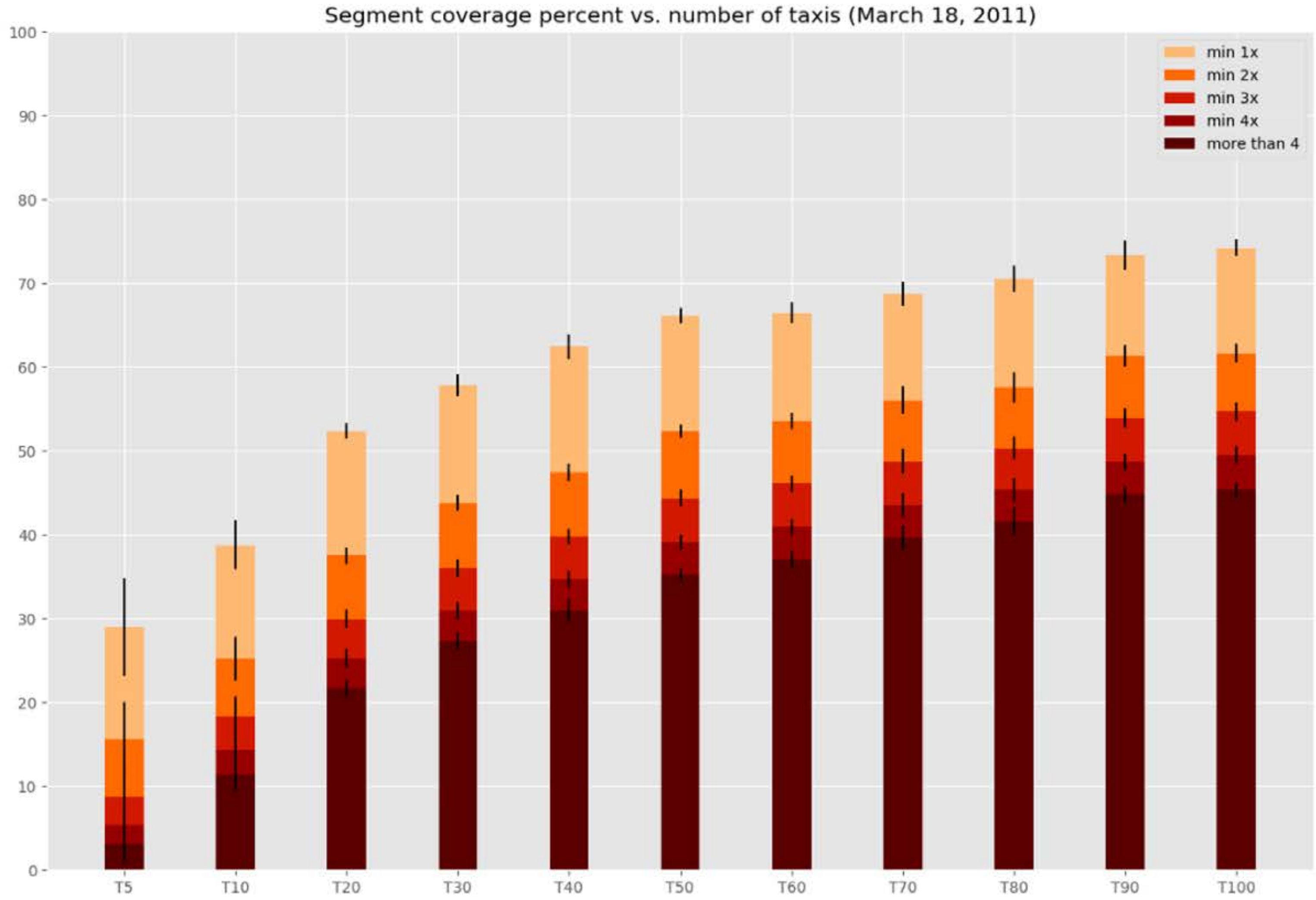
Can we turn urban vehicles into sensing platforms?

Challenge #1
Feasibility – How many sensors?



Djursholm
Trolldalen
Show o
Stocksund
Bergshamra
Tivoli
Sticklinge udde
Tranholmen
Grönsta
Bosön
Dalakolonin
Näset
Bo
Islinge
Västra Yttringe
Lidingö
Sundbyberg Municipality
Solna
Herserud
Stockby
Baggeby
Bodal
Larsberg

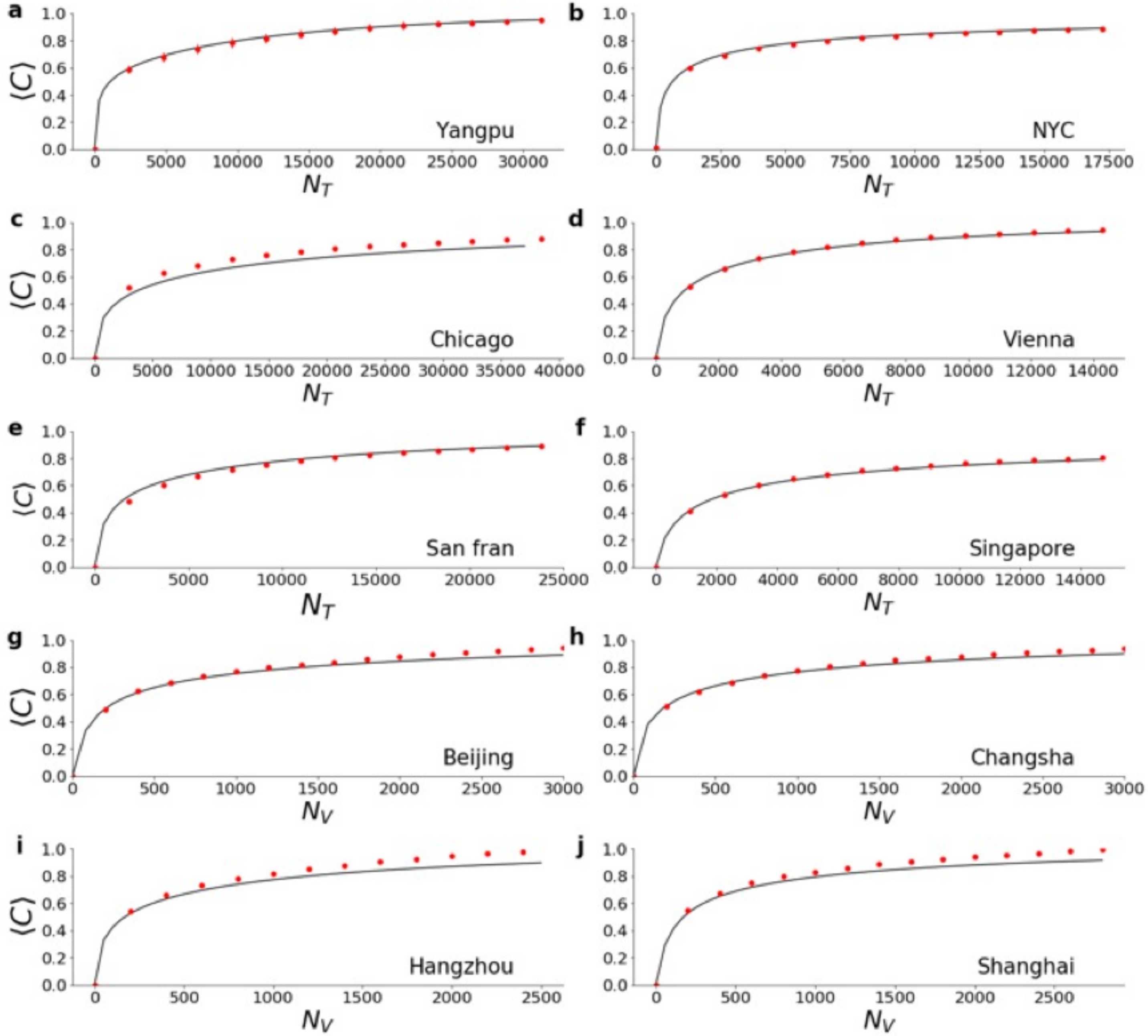
How many sensors do we need to cover a city?



How many daily trips cover the city?

just 1% of trips to get the desired coverage of 50% of street segments

Average coverage

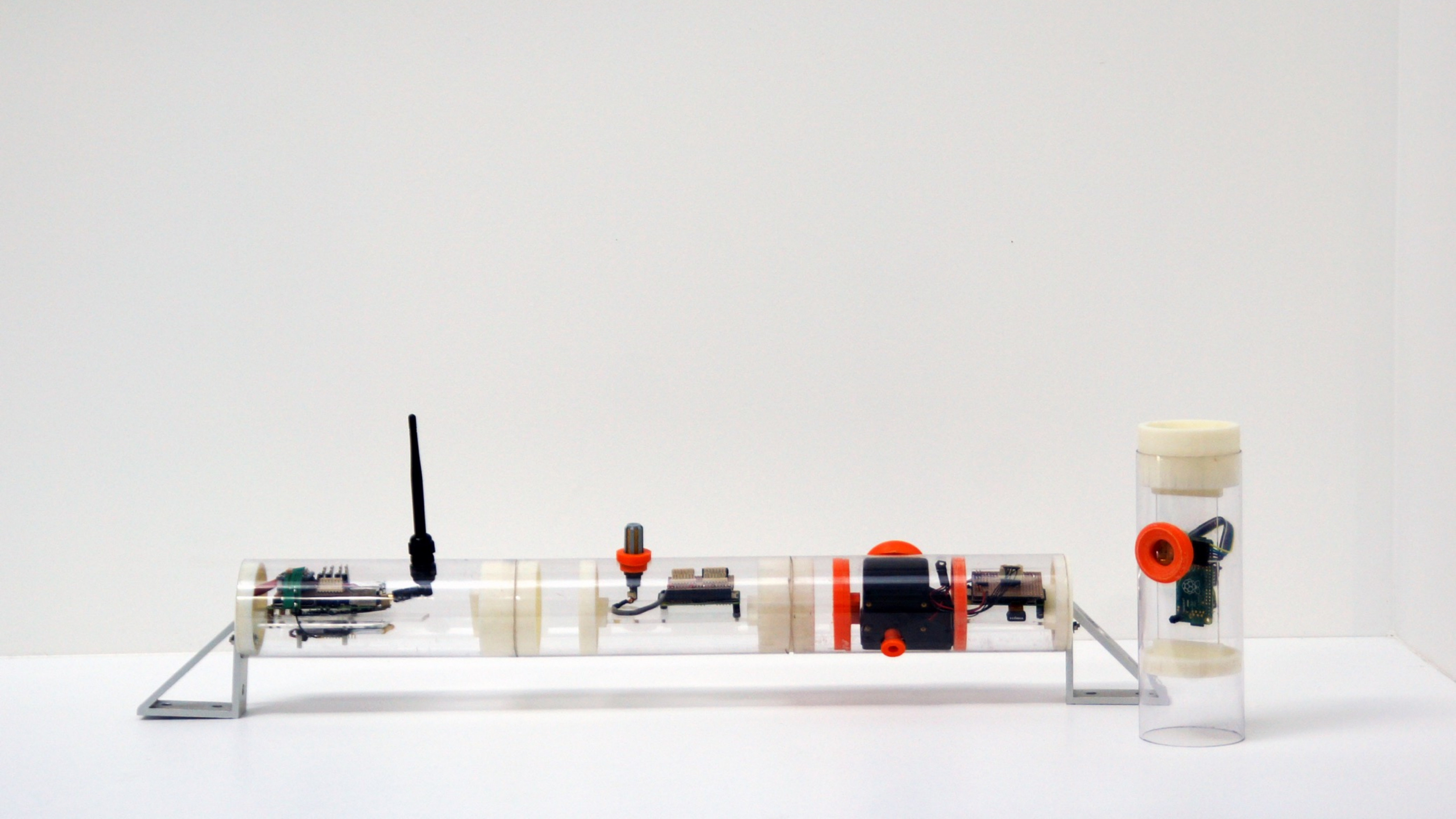


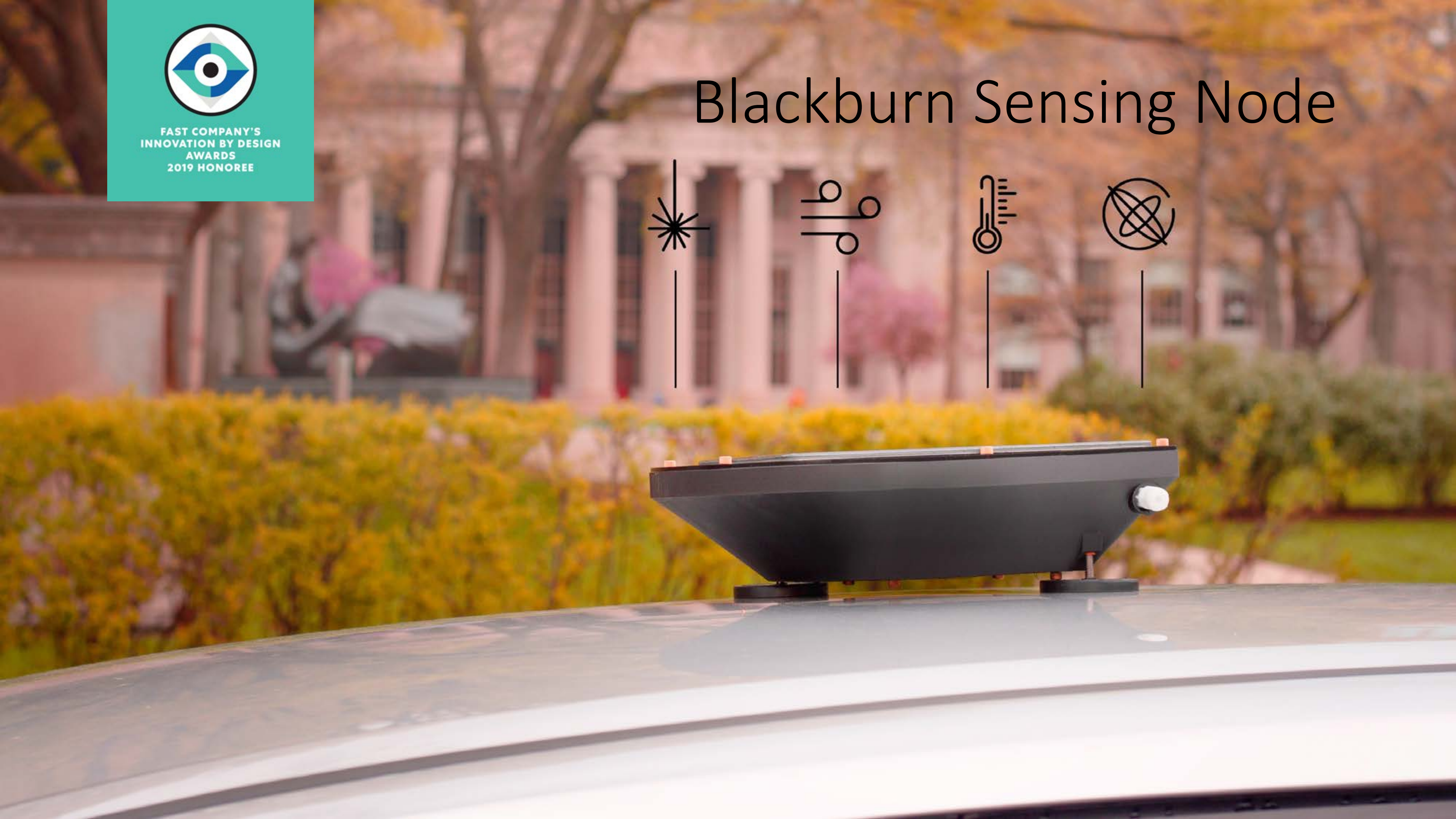
red curve : model prediction
black curve : real data

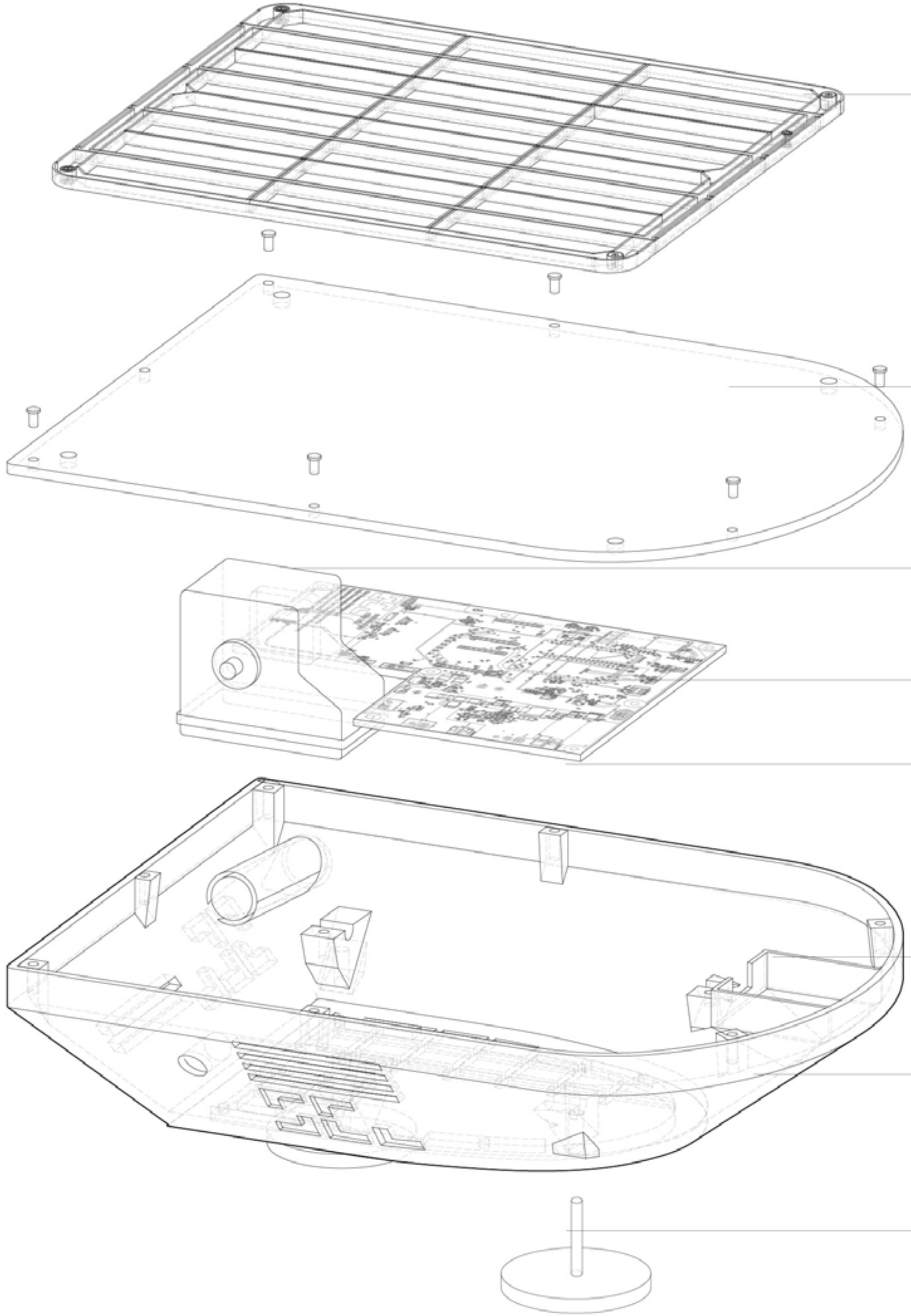
(# vehicles / # trips)

O'Keeffe, Kevin P., et al. "Quantifying the sensing power of vehicle fleets." *Proceedings of the National Academy of Sciences* 116.26 (2019): 12752-12757.

Challenge #2
Prototyping




Blackburn Sensing Node
FAST COMPANY'S
INNOVATION BY DESIGN
AWARDS
2019 HONOREE



Solar-powered
High-efficiency photovoltaic panels (PVs) can be tiled to fit irradiance characteristics of different cities, enabling continuous operation.

4mm Perspex
A layer that will act as the roof and add robustness to the device.

Core services onboard
Include 3G modem, GPS, temperature & humidity, accelerometer. During test deployments a particulate counter (OPC-N2) was included.

Adaptive real-time streaming
The device can adapt data sampling and broadcasting

Multi-purpose, customizable architecture
Support a wide range of sensors: e.g. particle counters, gas meters and thermal cameras.

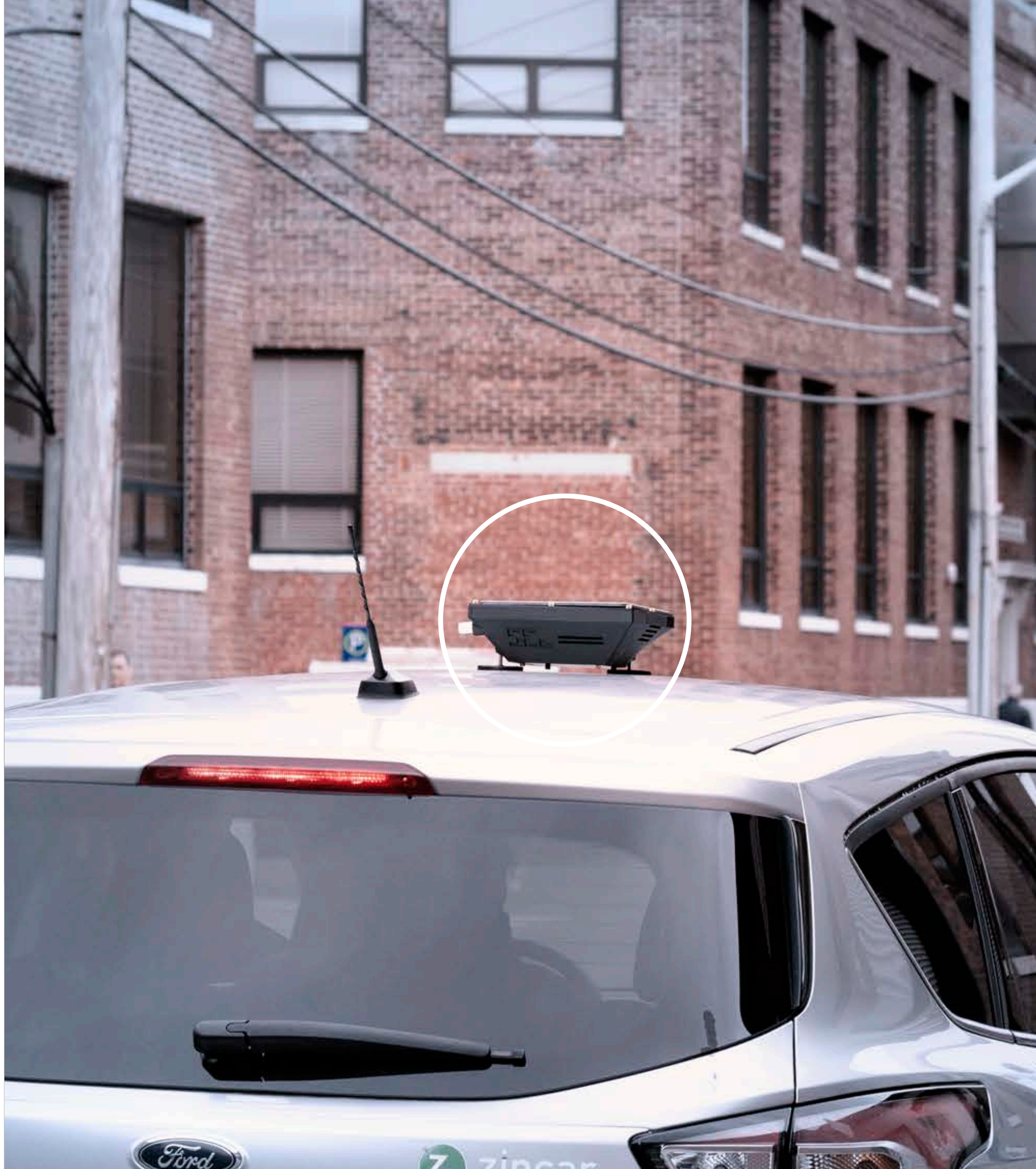
GPS and Cellular Antennas
Space for antennas that isn't directly beneath the solar panel which can block connectivity.

Shock Resistance
3D-printed shell in carbon fiber reinforced nylon, provides resistance and lightweight.

Magnetic bindings
For easy anchoring to the vehicle, allowing to reconfiguration the sensing fleet on-demand. Each magnet develops a force of circa 200N.

Mora et al., 2019
IEEE IoT World Forum Best Paper Award


n3



Challenge #3
Deployment

Tech innovators should pay attention to NYC's new air pollution monitoring pilot

Environmental Defense Fund Follow
Jan 22 · 4 min read

By Harold Rickenbacker, Manager, Clean Air & Innovation, Environmental Defense Fund

POLITICS NEWS

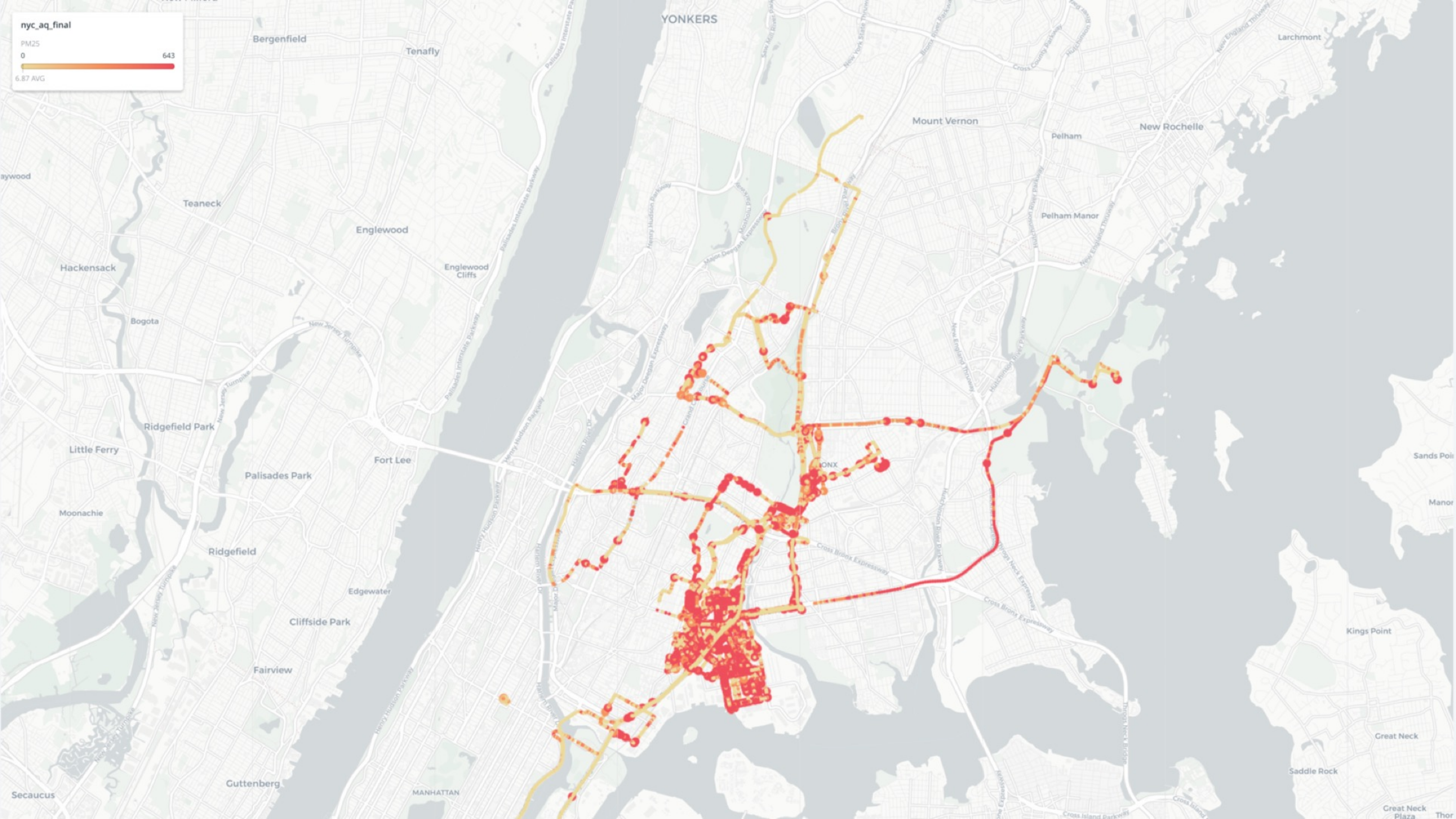
NYC municipal vehicles to test local air quality for pollution in South Bronx with mobile sensors

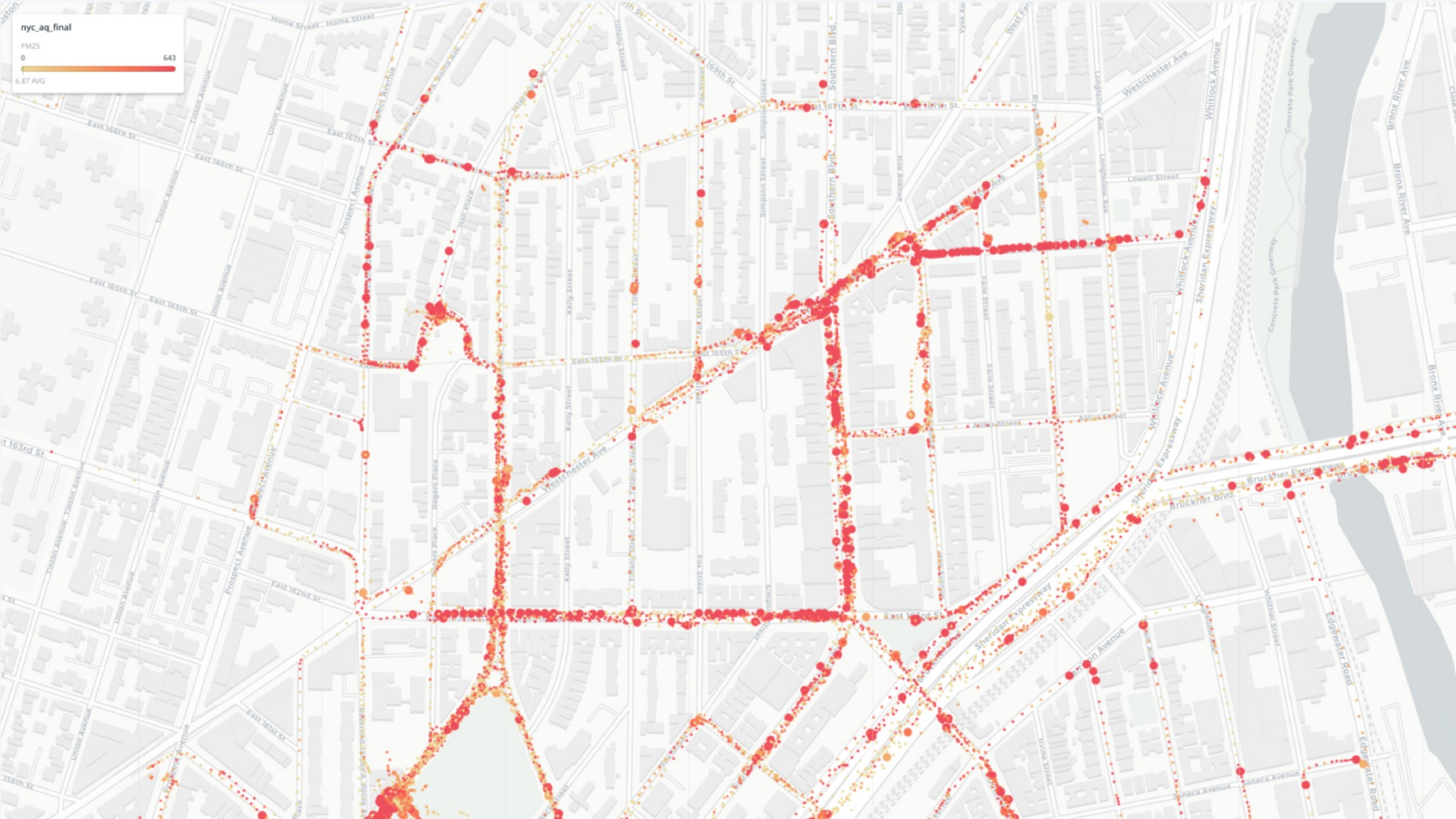
By ANNA SANDERS
NEW YORK DAILY NEWS | JAN 21, 2020 | 1:00 PM

New York City vehicles will be equipped with mobile air quality sensors developed by the Senseable City Lab at MIT to find pollution hotspots.

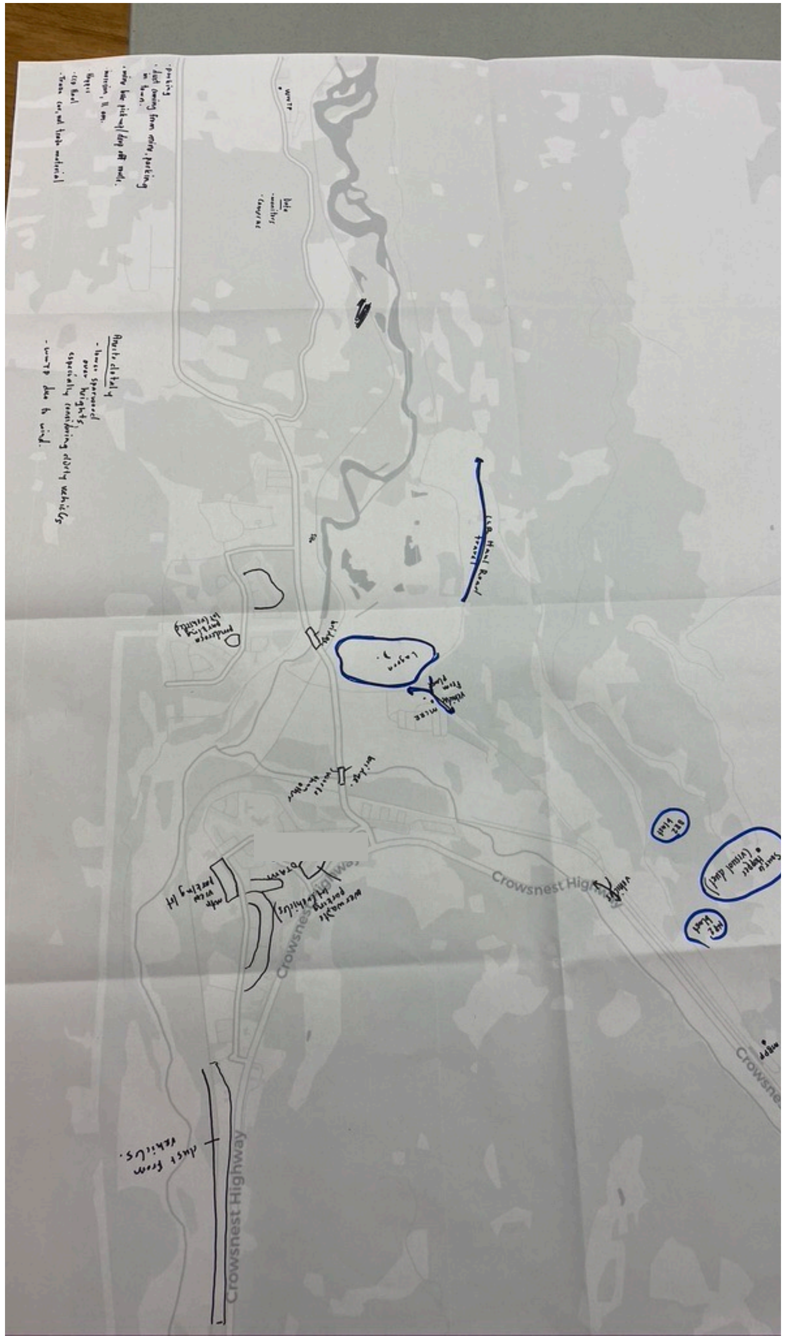
PILOT STUDIES

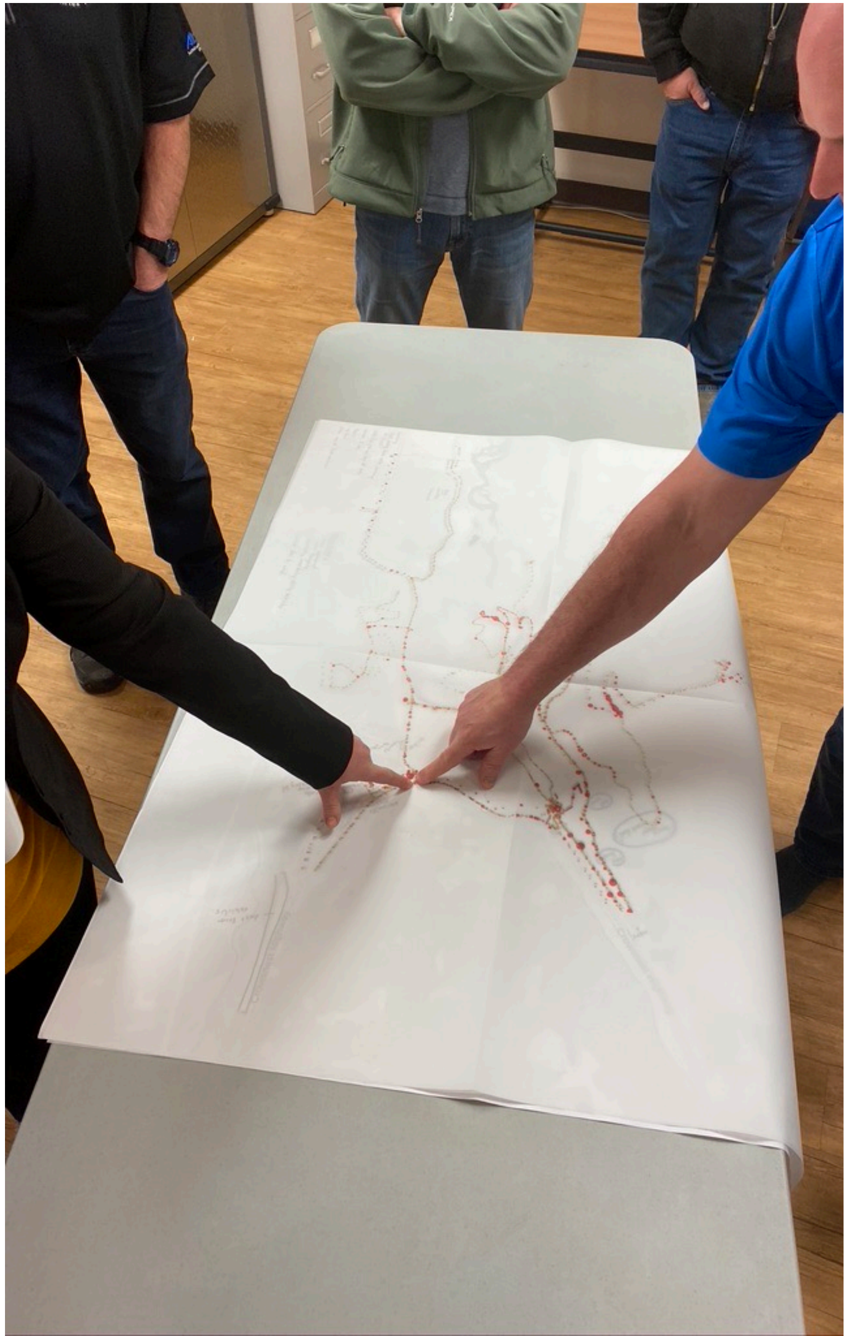
Challenge #4
Use cases – what research questions can we answer?


nyc_aq_final
PM25
0
643
6.87 AVG
YONKERS
Bergenfield
Tenafly
Mount Vernon
New Rochelle
Larchmont
Pelham
Pelham Manor
Teaneck
Englewood
Englewood Cliffs
Hackensack
Bogota
Ridgefield Park
Little Ferry
Fort Lee
Palisades Park
Moonachie
Ridgefield
Edgewater
Cliffside Park
Fairview
Guttenberg
Secaucus
MANHATTAN
BRONX
Cross Bronx Expressway
Sands Point
Manor
Kings Point
Great Neck
Saddle Rock
Great Neck Plaza


nyc_aq_final
PM25
0
643
6.87 AVG
Home Street
East 169th St
East 167th St
East 166th St
East 165th St
East 163rd St
East 162nd St
East 161st St
Westchester Ave
Southern Blvd
Simpson Street
Fox Street
Kelly Street
Tiffany Street
Prospect Avenue
Union Avenue
Tinton Avenue
Hall Place
Rogers Place
Longfellow Ave
Lowell Street
Bryant Ave
Faile Street
Hoe Avenue
Aldus Street
Whitlock Avenue
Sheridan Expressway
Bruckner Expressway
Bruckner Blvd
Concrete Park Greenway
Bronx River Ave
Whittier Street
Edgewater Road
Irvine Street
Seneca Avenue

Community Engagement





- Integrating local, qualitative knowledge with quantitative data
- Towards pollution source estimation
- Use case definition

Sensing people's lives

A House in a Box You Control by Waving Your Hand, a way to turn any small apartment into a more livable one. A project of the MIT Media Lab (2011).

BUSAN EDC

Data for rent, sensing the neighborhood

Busan EDC

In Busan, South Korea, 300 people started to experiment a new way of living

Busan EDC

Korea JoongAng Daily

In association with The New York Times

All Articles | Why | National | Business | Culture | Entertainment | Sports

Monday
February 7, 2022

dictionary +A -A

From home appliances to telemedicine, smart city project offers next-generation living

The Eco Delta Smart City in Busan [K-WATER]

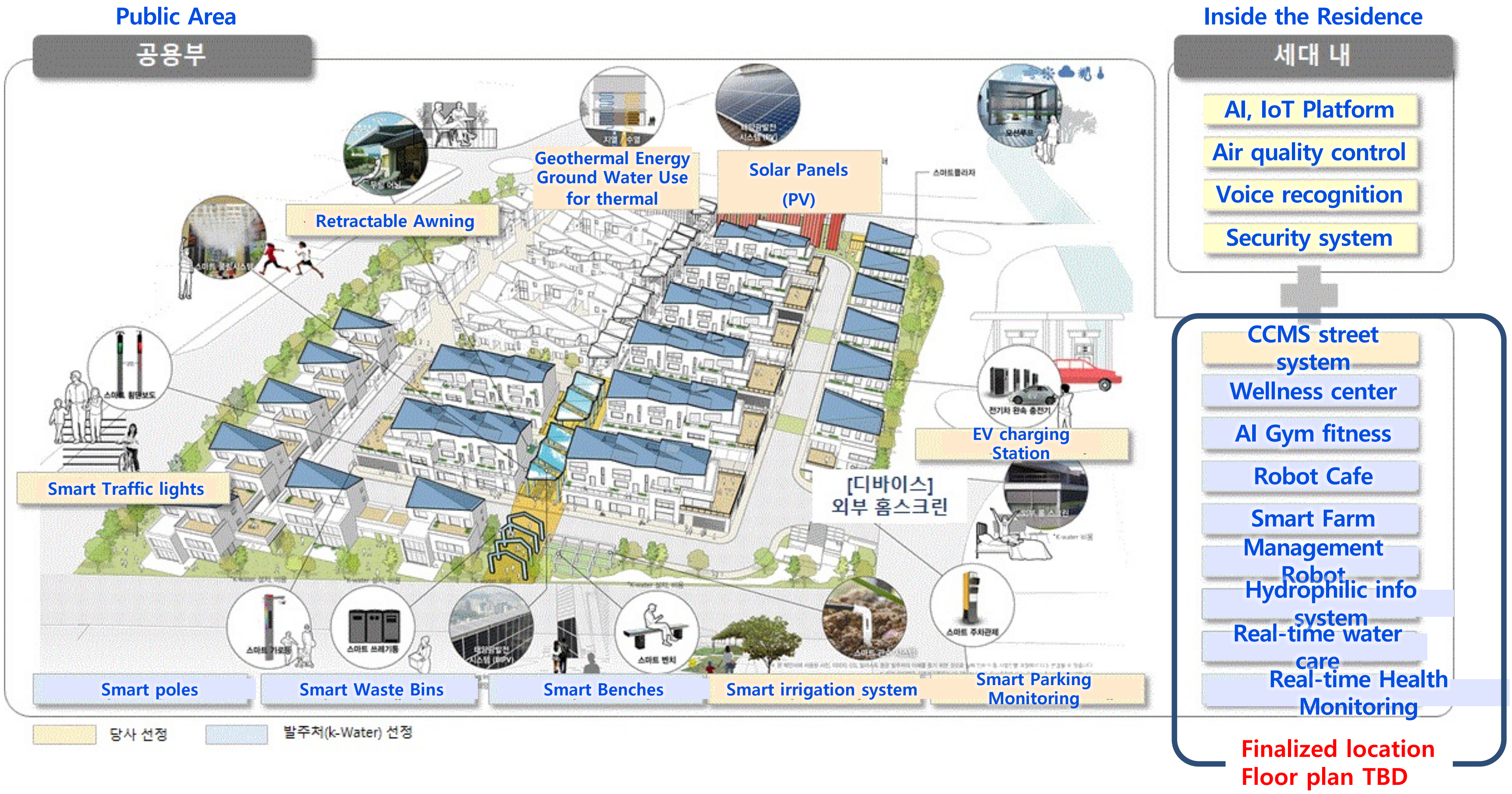
a unique
concept of «Data for Rent» ,
$1.8 billion government-funded smart city project

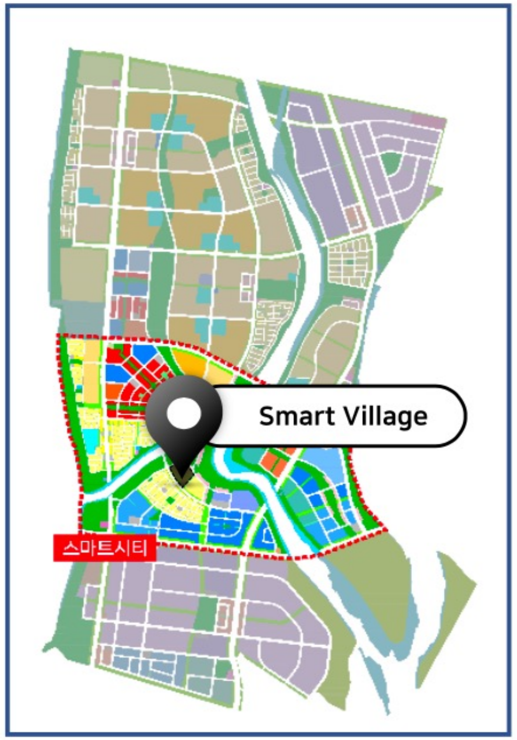
3000 applications
54 households,
300 residents
for the next 3 years
willing to live for free
in exchange for their data.

Busan EDC households

Rental expenses for data - ranging from
energy consumption patterns to health data,
home appliance usage and other behavioral information

Data possibilities





1 Smart Pole [●]

Recognizes motions of people and vehicles, and controls CCTVs

Smart Pole (8 Location)

Fall type(1)	W type(1)	O type(6)

2 Smart Waste [■]

Real-time info. of status such as current load, compress & overflow

Smart Waste (8 Location, Trash collected to separate location)

3 Smart Bench [▲]

Solar-powered bench, provide amenity such as cellphone charge

Smart Bench (4 location)

A Type (3 spots)	B Type (1 spot)

4 Smart Farm [♦]

Eco-friendly farm utilizing rainwater, Growing crops and vegetables

Smart Farm (2 location)
[Cultivation Bldg. (1), Service Bldg. (1)]

Data possibilities

(Official K-Water Selection)

Category	Key Technolgy
Water	① Hydrophilic information platform
	② Real-time water care
Healthcare	③ Real-time health management
	④ Wellness center
	⑤ AI sports center
Neighborhood	⑥ Smart pole
	⑦ Robot café
	⑧ Smart management robot
	⑨ Smart bench
Lifestyle	⑩ Smart trash can
	⑪ Smart farm

\+

(Samsung C&T Proposed)

Category	Key Technolgy
Private Space	① smart home - energy
	② smart home - air quality
	③ smart home - safety / security
	④ smart home - convenience
Private Outdoor Space	⑤ Moving awning / motion roof
Public Space	⑥ EV Charging Station
	⑦ smart parking control & monitor
	⑧ smart pedestrian crosswalk
	⑨ smart irrigation system
	⑩ smart solar energy
	⑪ Intelligent video management system (security/surveillance)

Residents completed moving in on Jan. 15, 2022

Data collection has started

People’s point of view

From the project to the reality

People’s point of view

Lee is a student at the department of civil engineering at Pusan National University

People’s point of view

“The biggest difference that I feel now is that I don’t have to get up from the bed to turn the light off at night,” Lee said.
“I can command it with my voice, which is actually more convenient than you think, once you get used to it.”

People's point of view

most convenient thing for Lee is the TV, which "tells us when our laundry is done or when the oven's finished with cooking."

Data possibilities

How do we design a network of urban sensors centered on creating knowledge?

Data possibilities

Data can provide knowledge on lives and behaviors

monitoring and alerting | services & experiences | research

Data possibilities

How to create engaging spatial experiences in the Busan Eco City while collecting data at the same time?

Sensing human connections

Work and leisure in post-industrial cities don't need a particular spatial configuration anymore, how people are communicating?

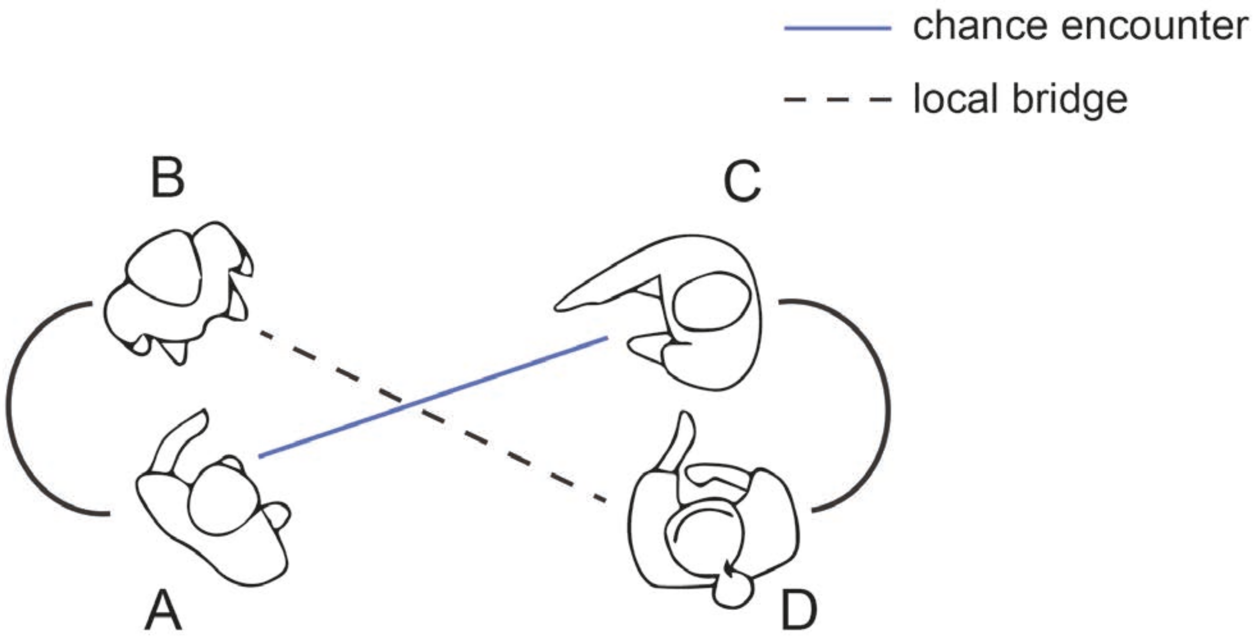
Understanding new ways of living

Digital communication, human connections

Random encounters

Imagine the following scenario:

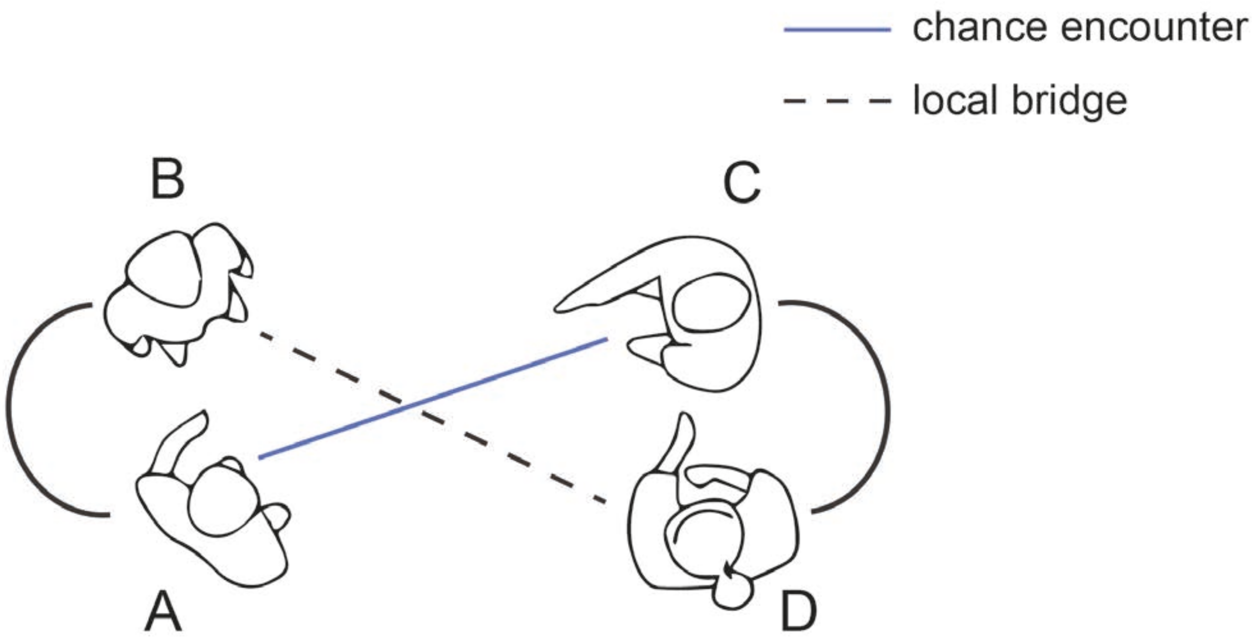
You (A) go to lunch with one of your friends (B).



Group participation

Definition: A local bridge in a network is an edge which is not part of any triangle in the network.

Question: What are other scenarios under which local bridges might form in a social network?

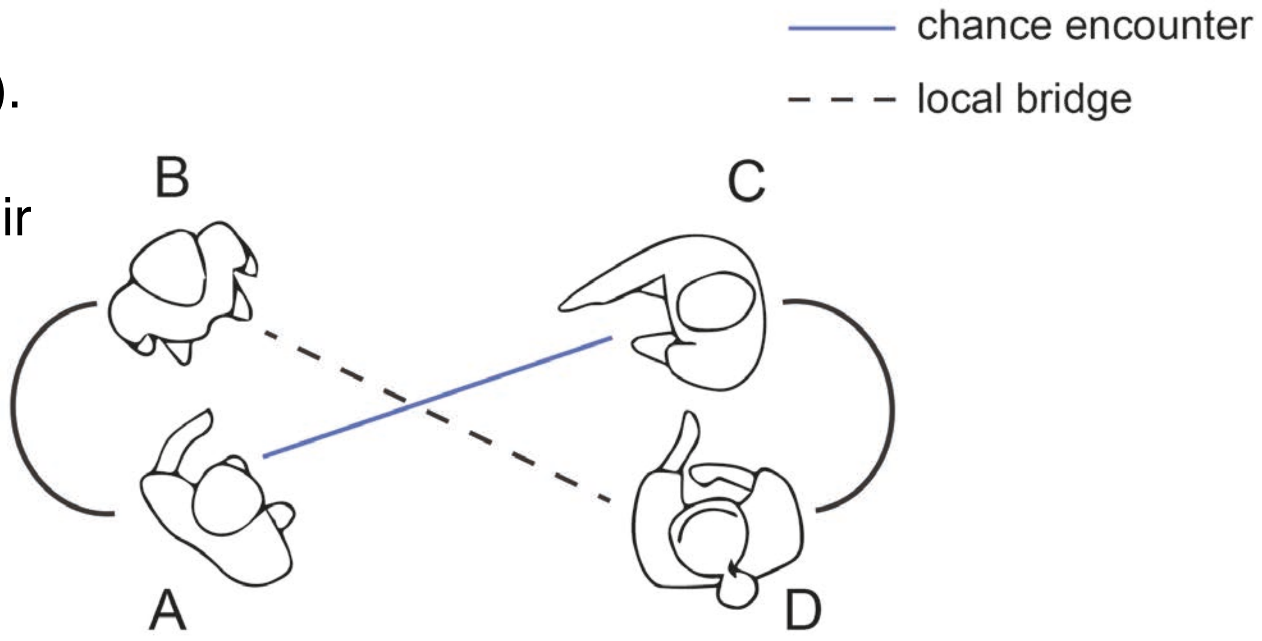


Random encounters

Imagine the following scenario:

You (A) go to lunch with one of your friends (B).

Your coauthor (C) goes to lunch with one of their friends (D), who you don't know.



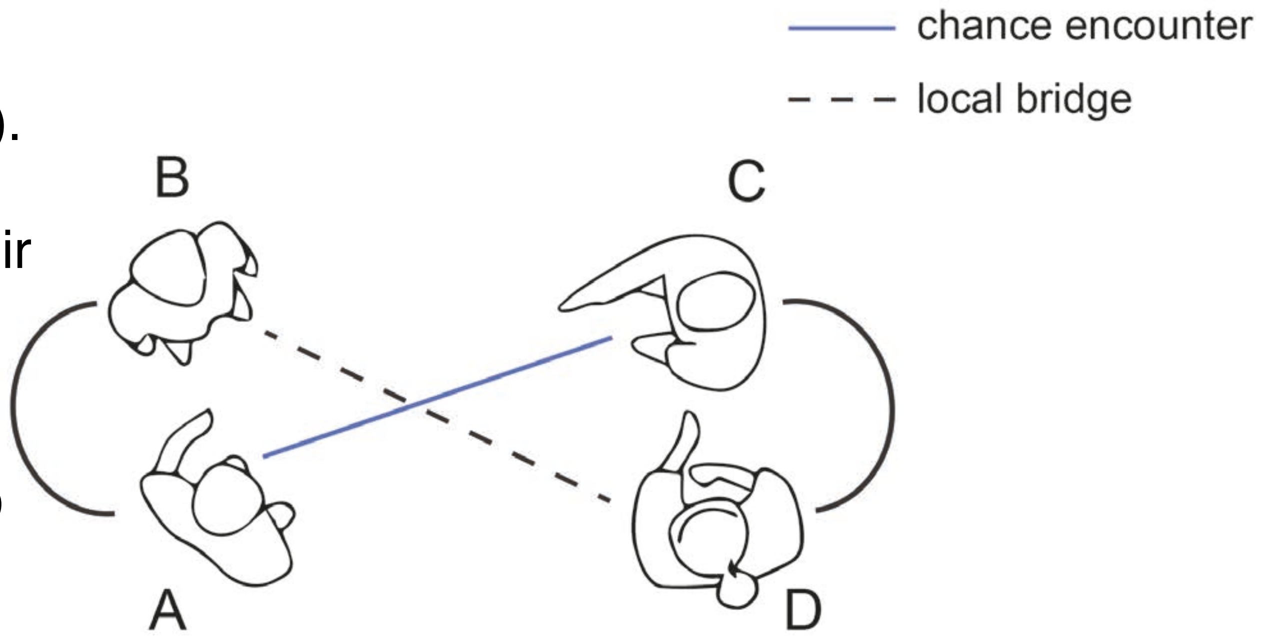
Random encounters

Imagine the following scenario:

You (A) go to lunch with one of your friends (B).

Your coauthor (C) goes to lunch with one of their friends (D), who you don’t know.

At lunch you and your coauthor run into one another, and to be polite you introduce B and D to each other.



Random encounters

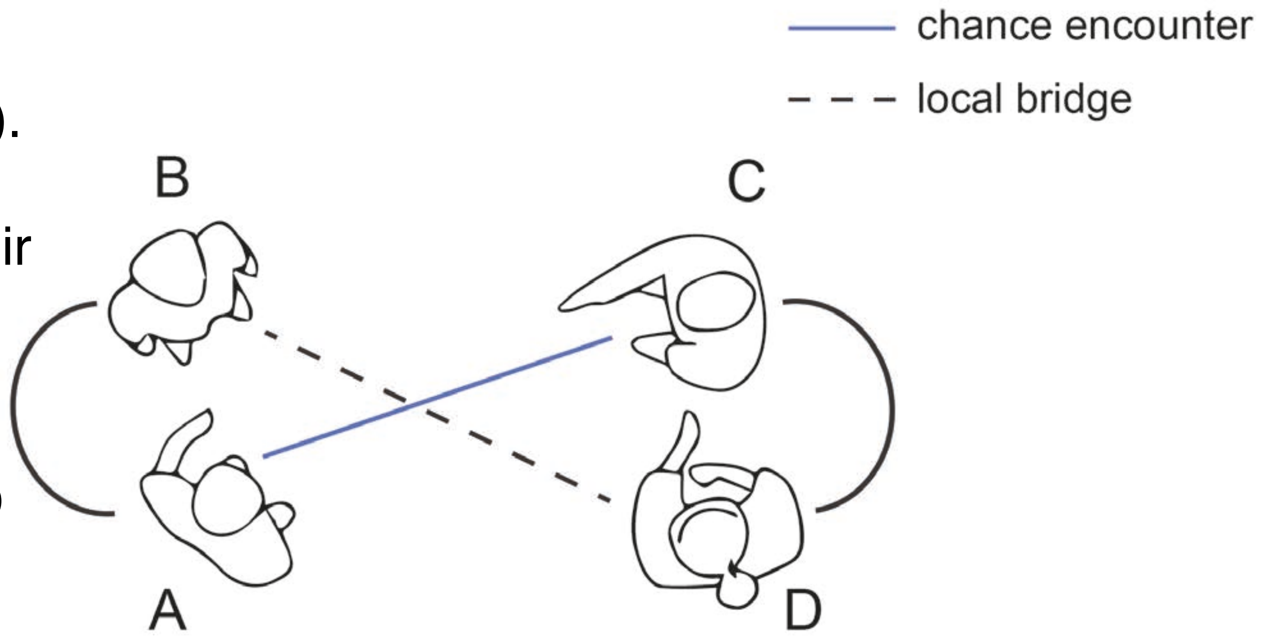
Imagine the following scenario:

You (A) go to lunch with one of your friends (B).

Your coauthor (C) goes to lunch with one of their friends (D), who you don’t know.

At lunch you and your coauthor run into one another, and to be polite you introduce B and D to each other.

Now B and D have formed a connection **even though they have no common friend**.



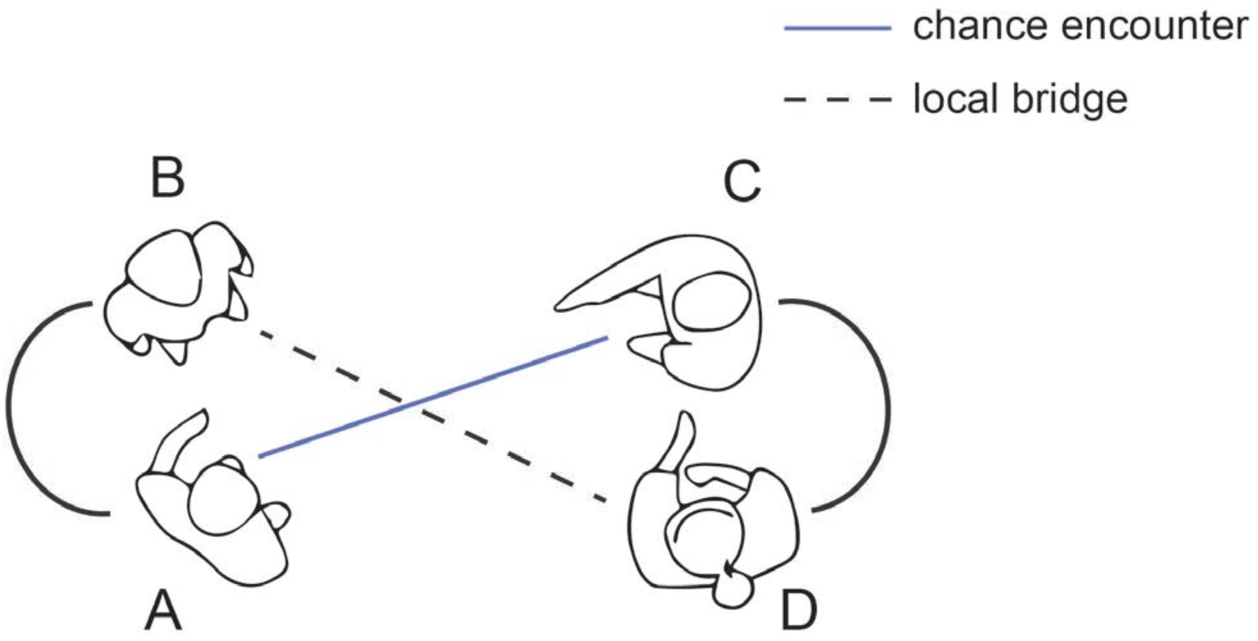
Local bridges

Definition: A local bridge in a network is an edge which is not part of any triangle in the network.

In other words, a local bridge is a connection between people who have no mutual friends.

Local bridges are topologically "weak ties" in the sense of Granovetter.

We will use the phrase "local bridge" and "weak tie" interchangeably.

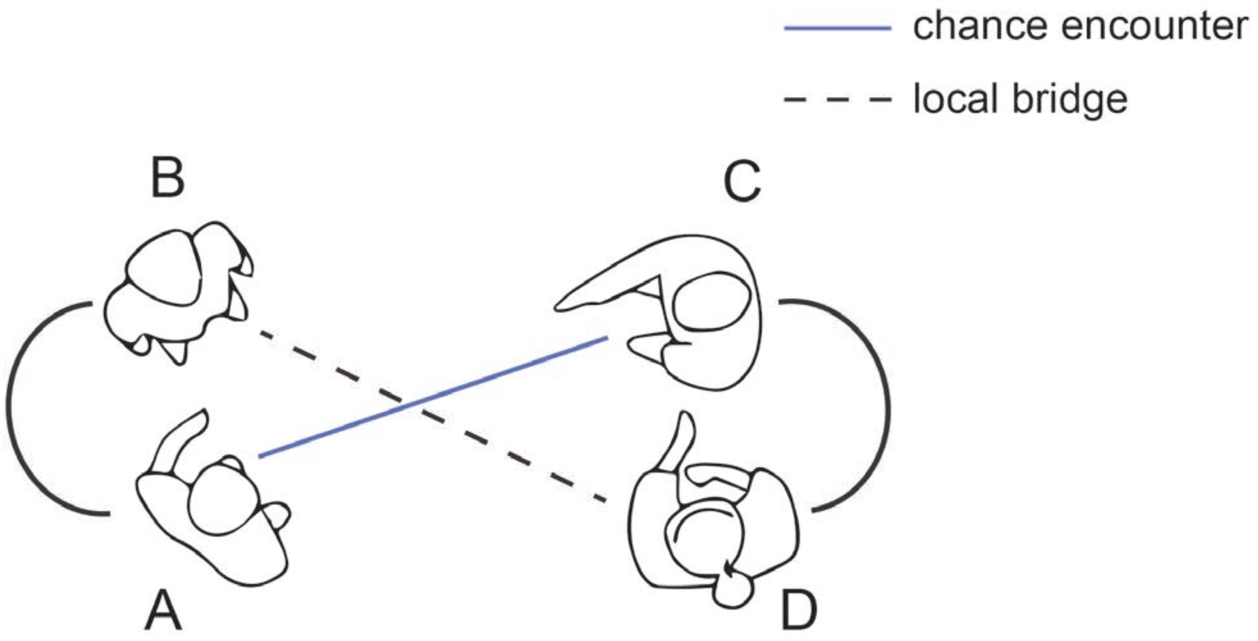


Local bridges

Definition: A local bridge in a network is an edge which is not part of any triangle in the network.

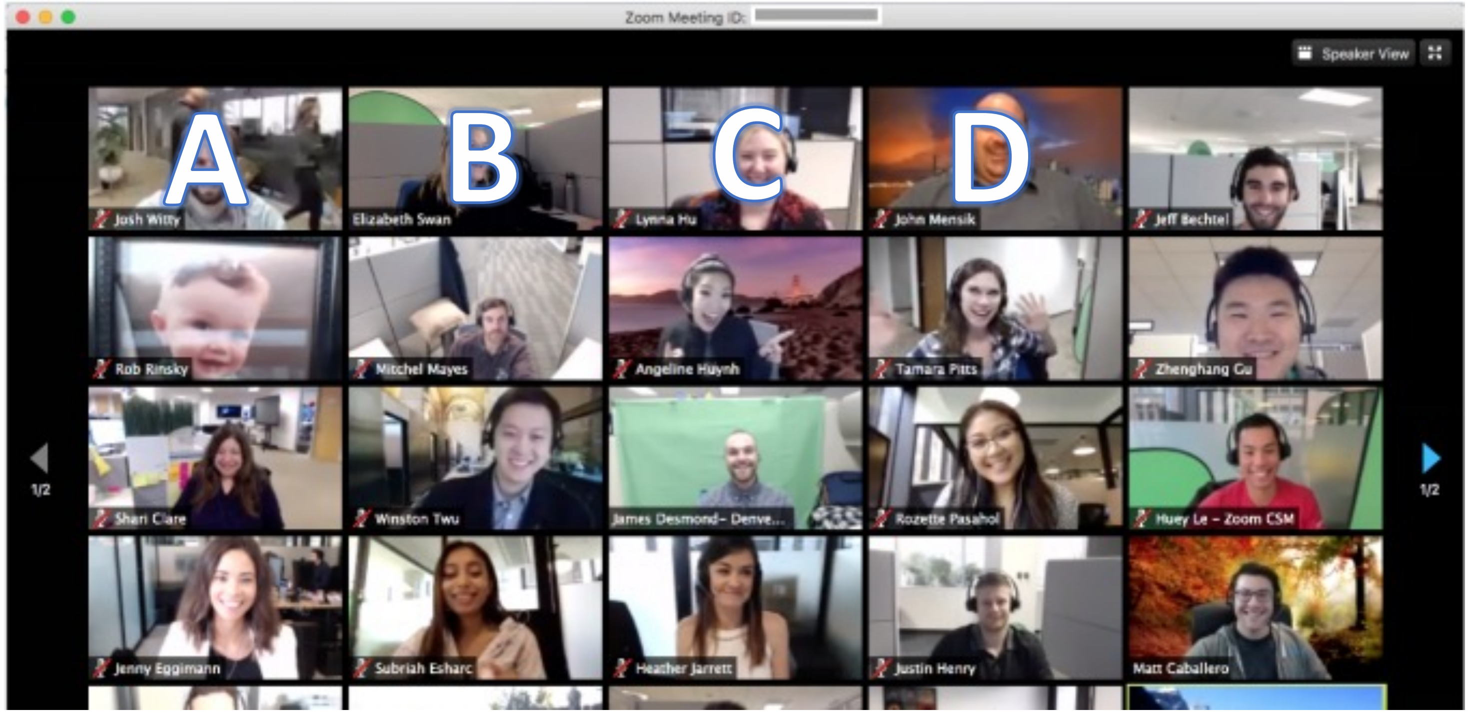
Local bridges are important for the spread of information in networks.

By definition, removing local bridges increases the average shortest path length in a network more than removing edges embedded in triangles (with the same betweenness centrality).



Random encounters without co-location

Consider the following modification of the original scenario:



You (A) go to a Zoom seminar with your friend (B).

Your coauthor (C) goes to the same seminar with their friend (D).

You (A) see that (C) is connected, but you have no way of knowing that they are friends with (D), and (B) and (D) are never introduced.

Broad question: Does co-location promote the formation of local bridges in communication networks?

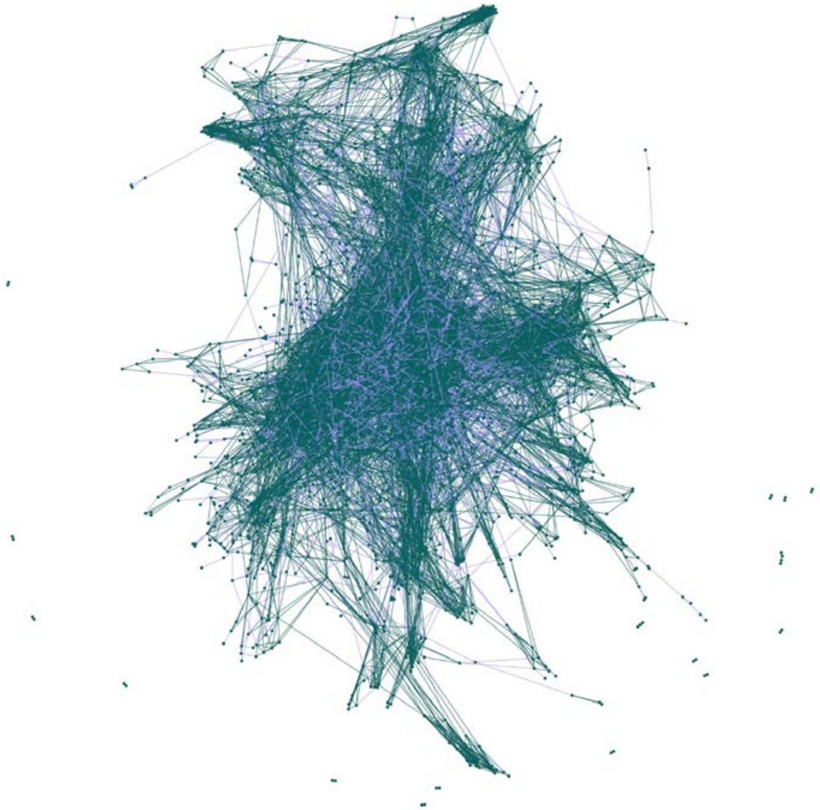
MIT COVID-19 policy

MIT implemented a mandatory remote-work policy which went into full effect midway through the Spring 2020 semester on **March 23, 2020**.

The Fall 2020 and Spring 2021 semesters were completely remote.

At the start of the Fall 2021 semester on **September 8, 2021** MIT partially re-opened its campus, with many researchers going to their offices 2-3 times per week.

Experimental setup

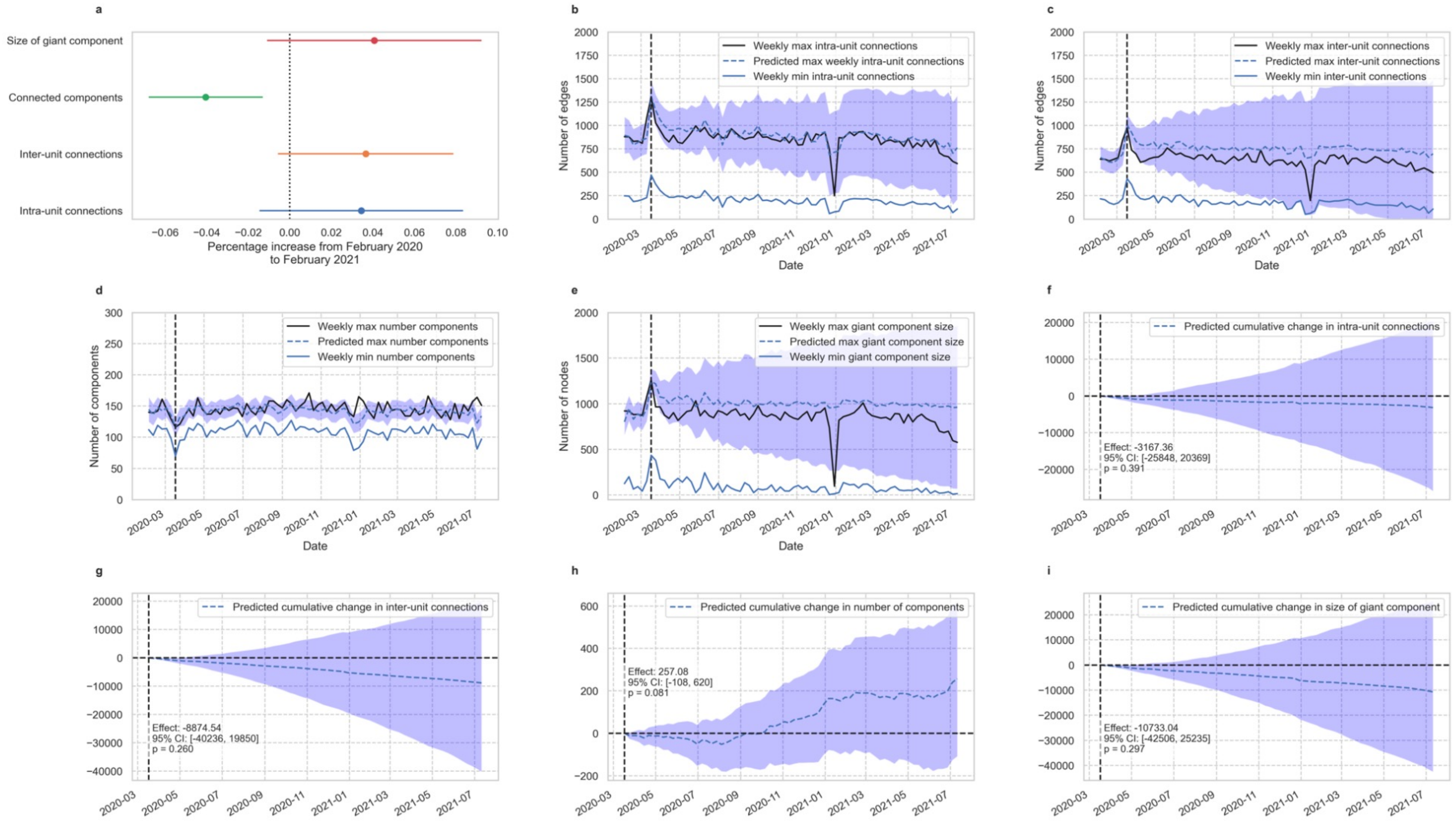


We study the daily email networks of MIT researchers from December 2019-October 2021.

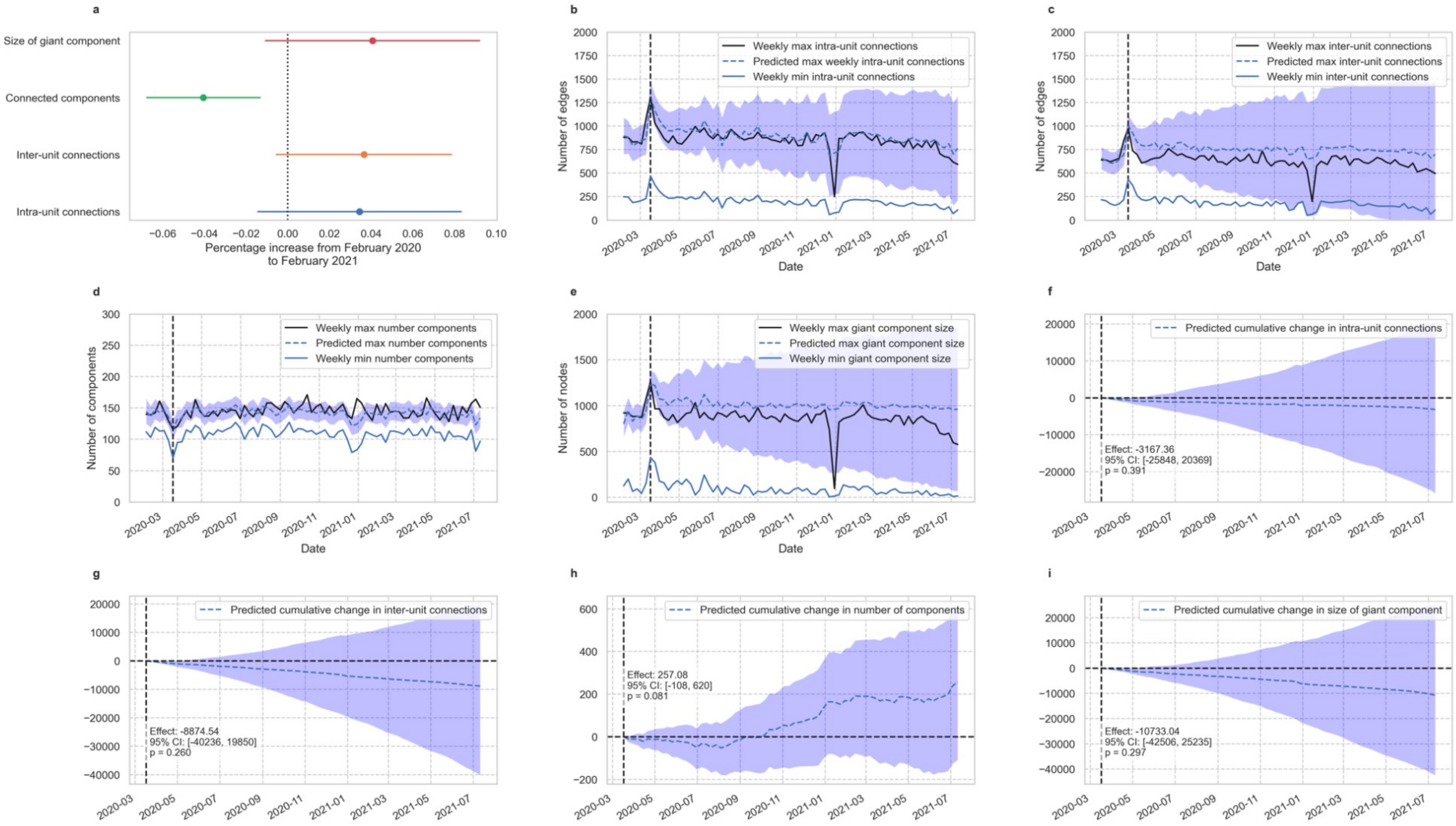
There is an edge between researchers on a given day if both researchers sent an email to one another that day.

The shift to remote work on March 23, 2020 acted as an intervention, so we can study its causal effect on local bridges in the email network.

Are the networks obviously damaged by remote work?



Are the networks obviously damaged by remote work?

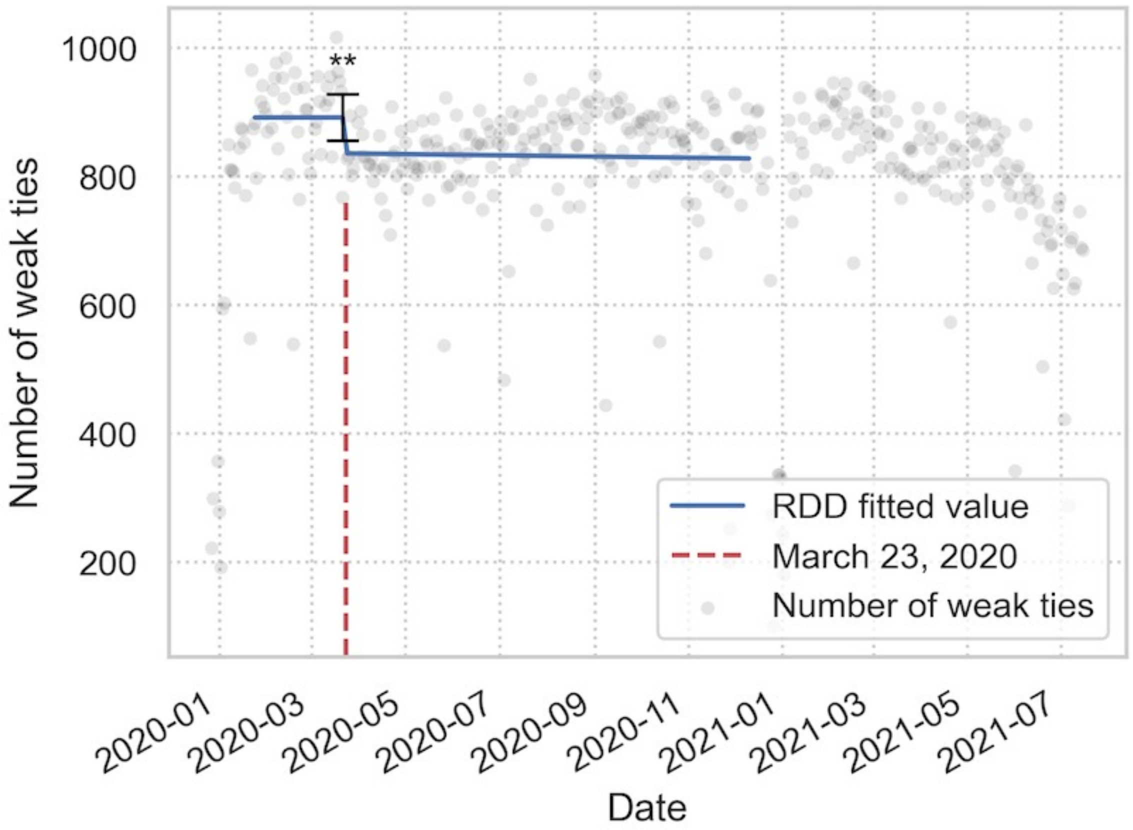


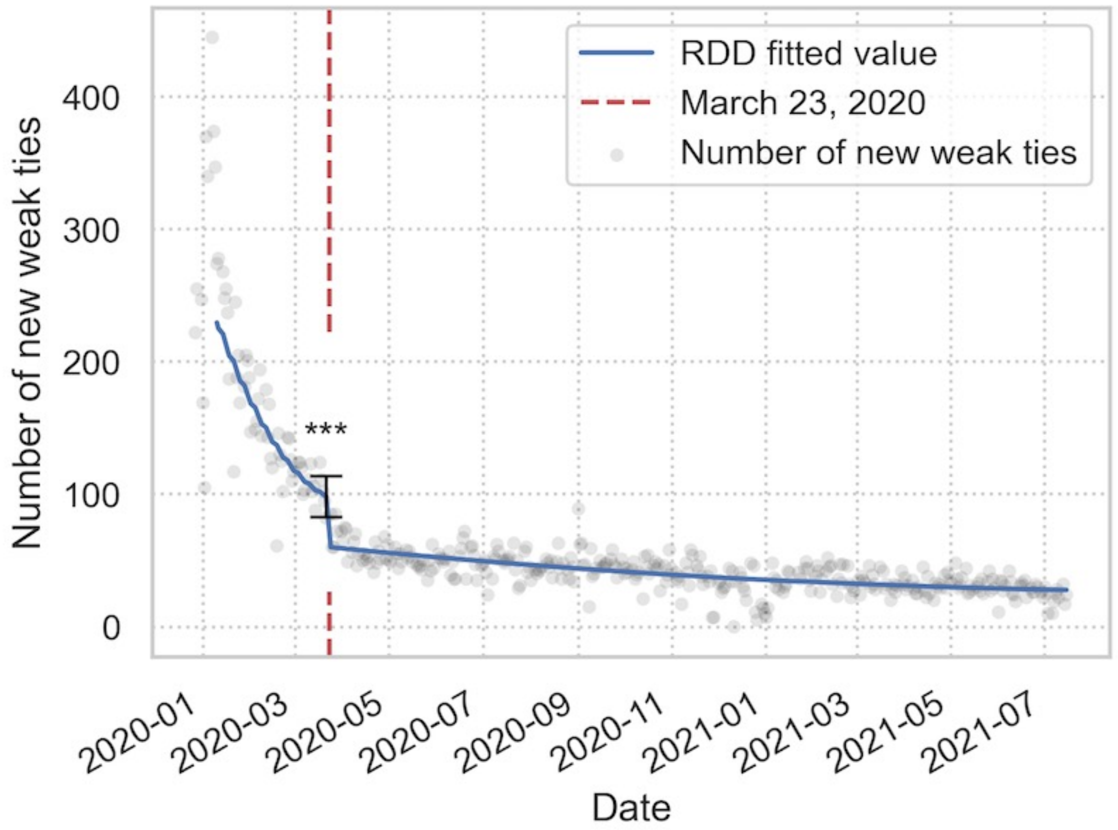
Not really.

Refined question: Does working nearby on campus on a given day **cause** an increase in the probability to form a local bridge in the email network that day?

The existence of a causal link

Interrupted time series (regression discontinuity design)

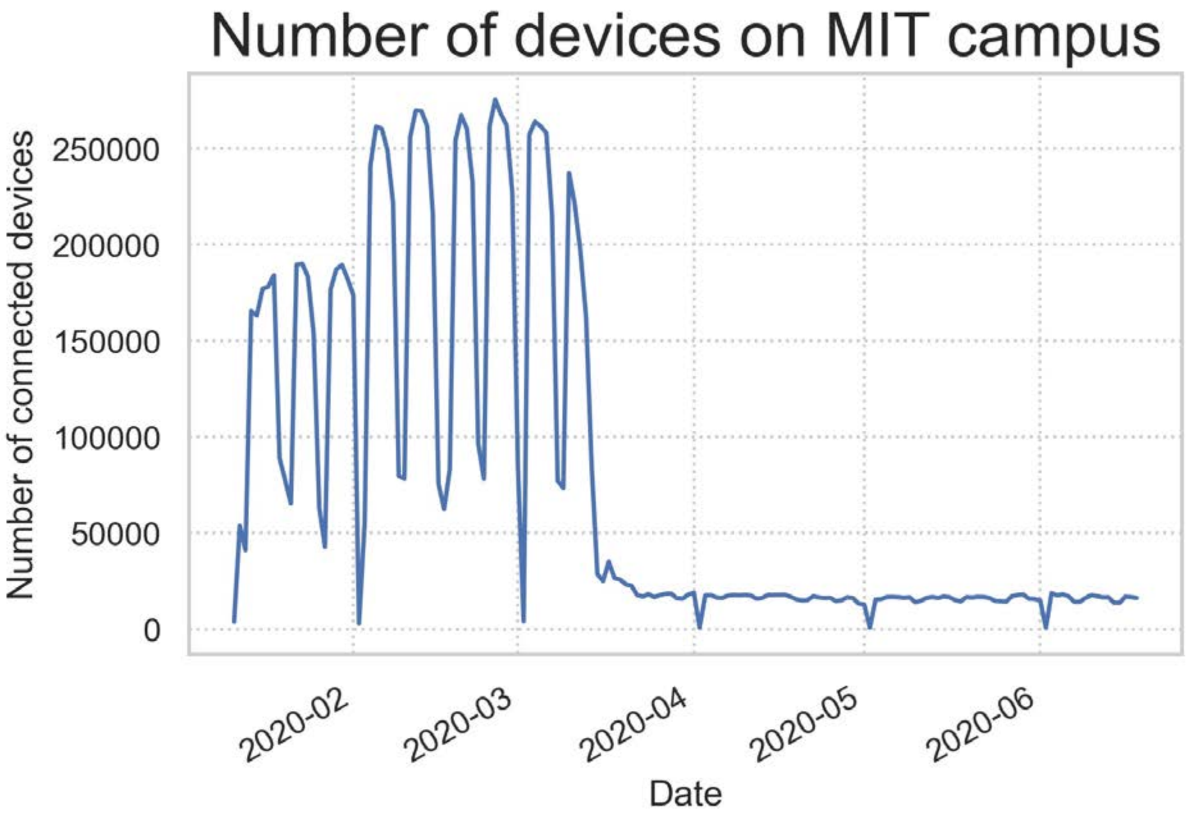




Assumptions:

1. The response variable is continuous with respect to time near the cutoff on March 23, 2020
2. Subjects cannot precisely manipulate the assignment variable to determine their treatment status

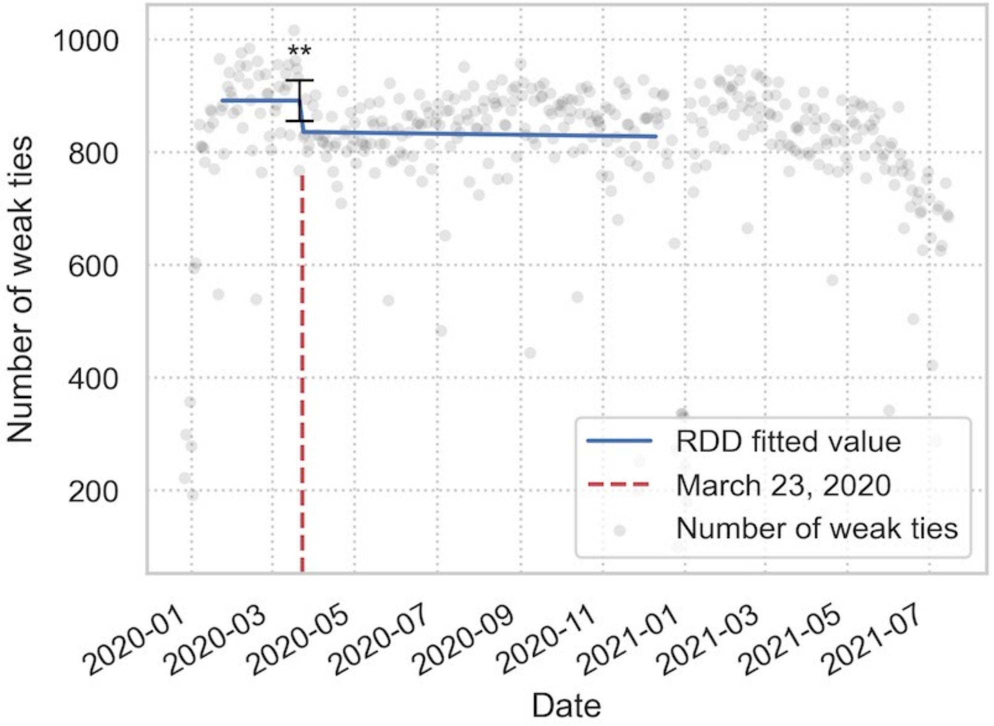
Interrupted time series (regression discontinuity design)

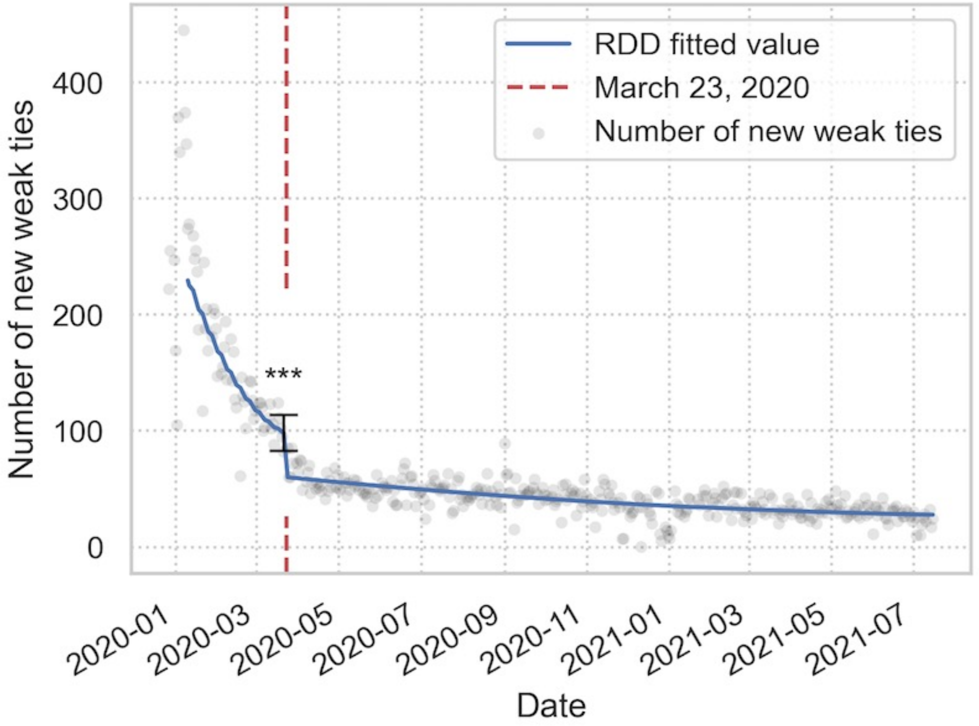


Assumptions:

1. The response variable is continuous with respect to time near the cutoff on March 23, 2020
2. **Subjects cannot precisely manipulate the assignment variable to determine their treatment status**

Interrupted time series (regression discontinuity design)



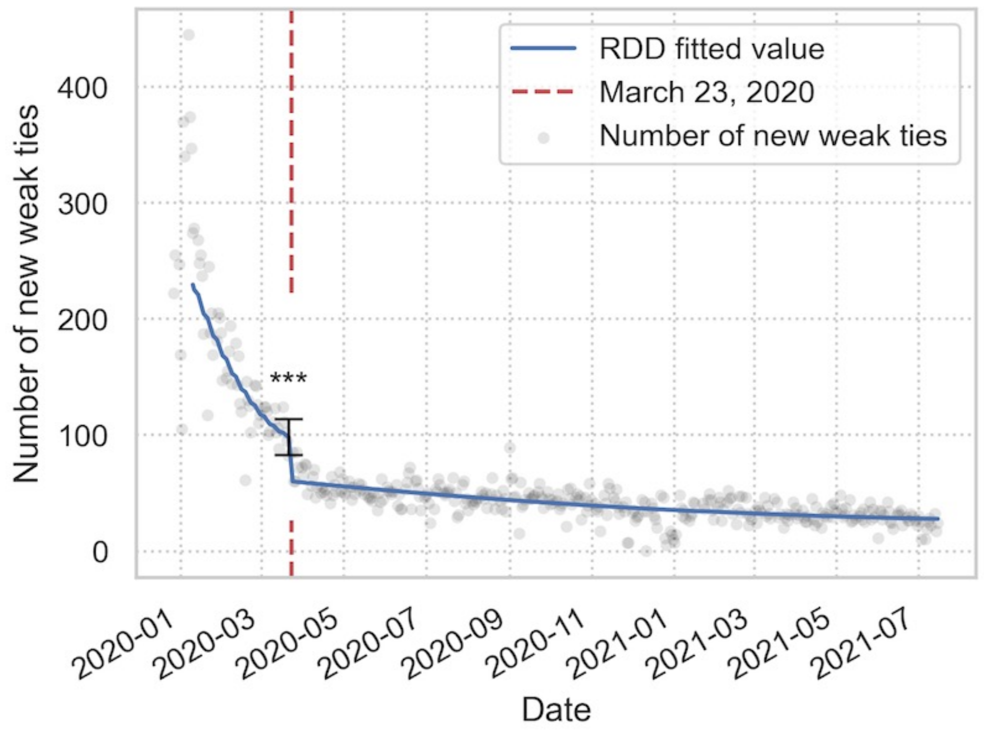


τ is the impact of the policy

$$Y = \alpha + \tau D + \beta_2 D(X - c) + \epsilon$$

$$Y = \alpha + \tau D + \beta_1 (X - c) + \beta_2 (X - c)^2 + \beta_3 D(X - c) + \beta_4 D(X - c)^2 + \epsilon$$

Interrupted time series (regression discontinuity design)

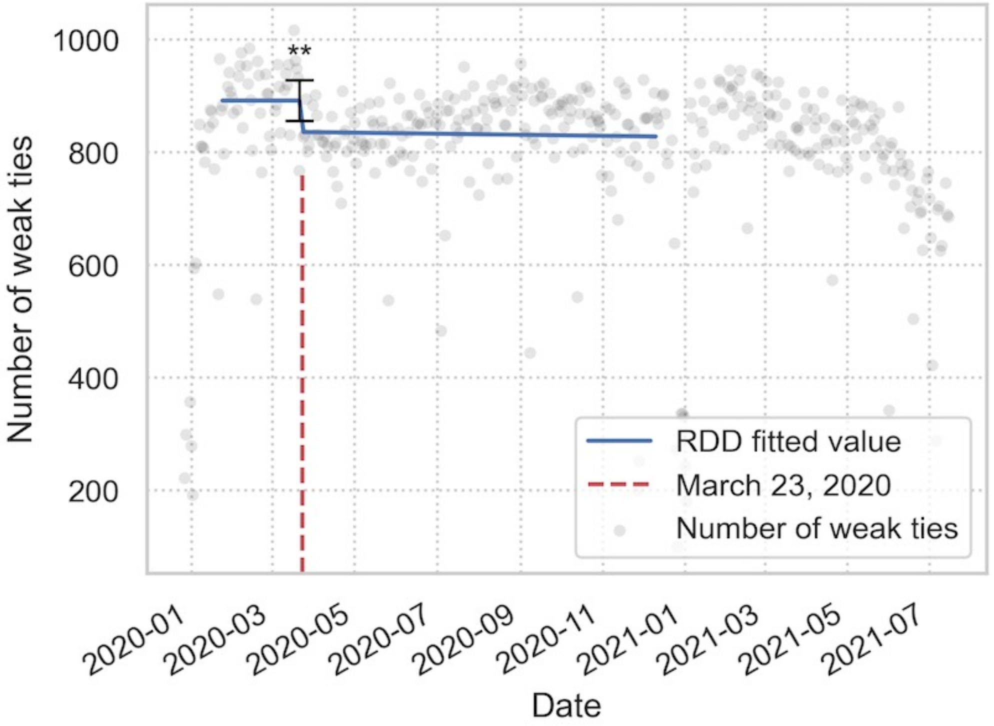


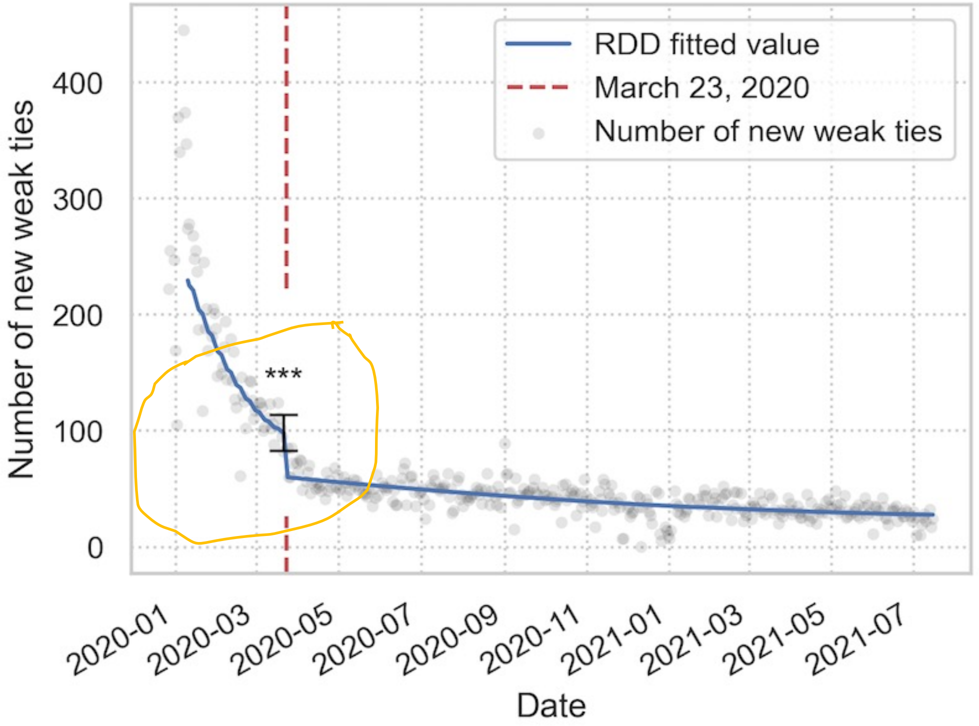
τ is the impact of the policy

$$Y = \alpha + \tau D + \beta_2 D(X - c) + \epsilon$$

$$Y = \alpha + \tau D + \beta_1 (X - c) + \beta_2 (X - c)^2 + \beta_3 D(X - c) + \beta_4 D(X - c)^2 + \epsilon$$

Interrupted time series (regression discontinuity design)





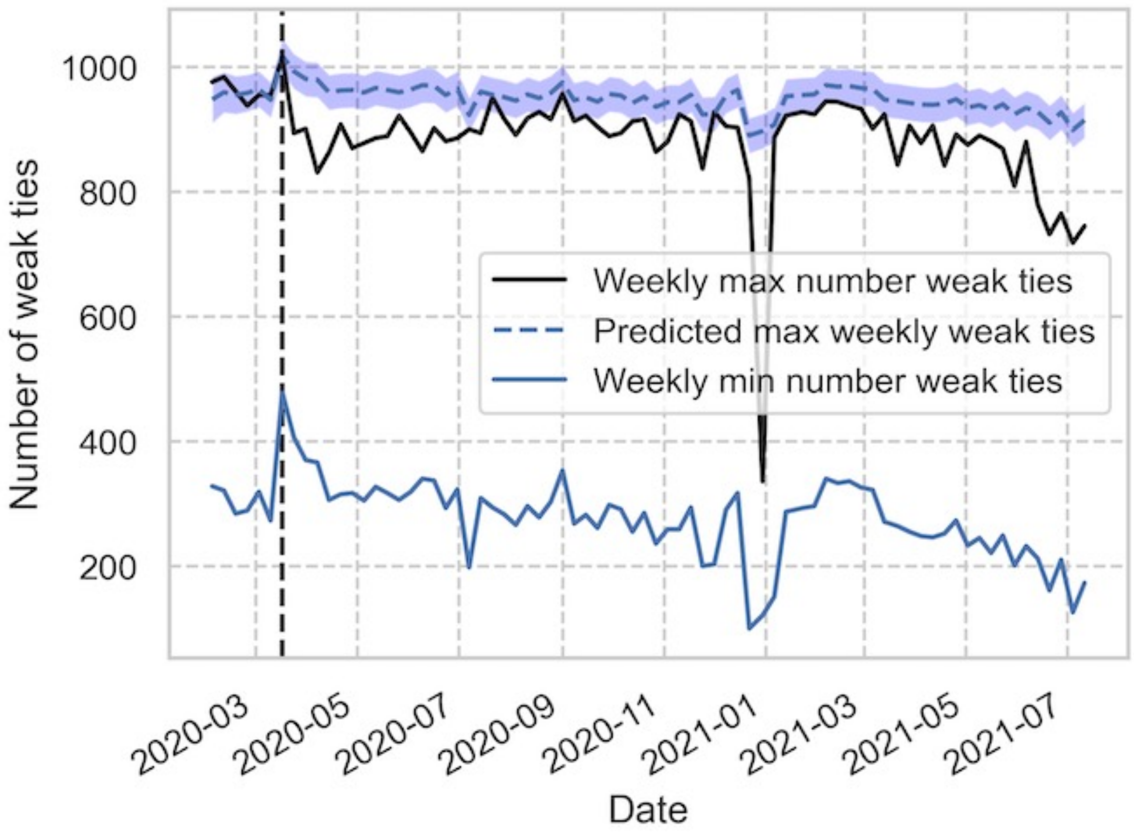
τ is the impact of the policy

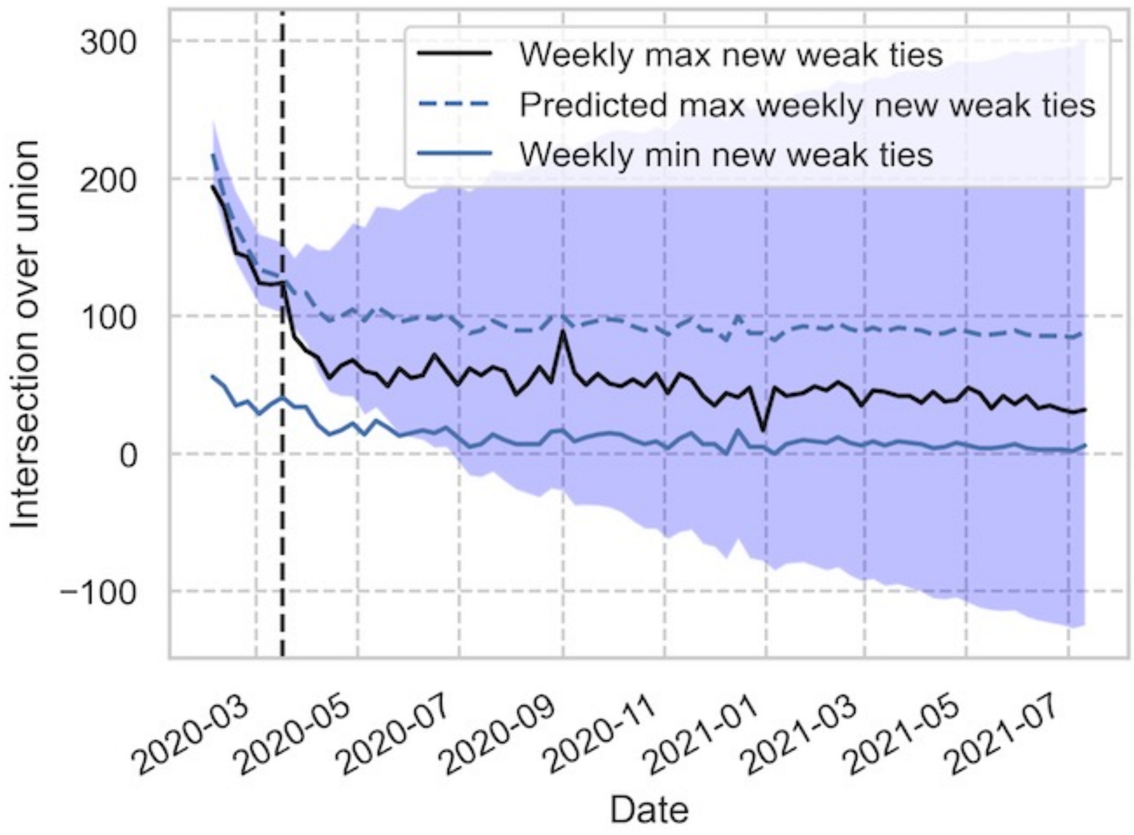
$$Y = \alpha + \tau D + \beta_2 D(X - c) + \epsilon$$

$$Y = \alpha + \tau D + \beta_1(X - c) + \beta_2(X - c)^2 + \beta_3 D(X - c) + \beta_4 D(X - c)^2 + \epsilon$$

Bayesian structural time series

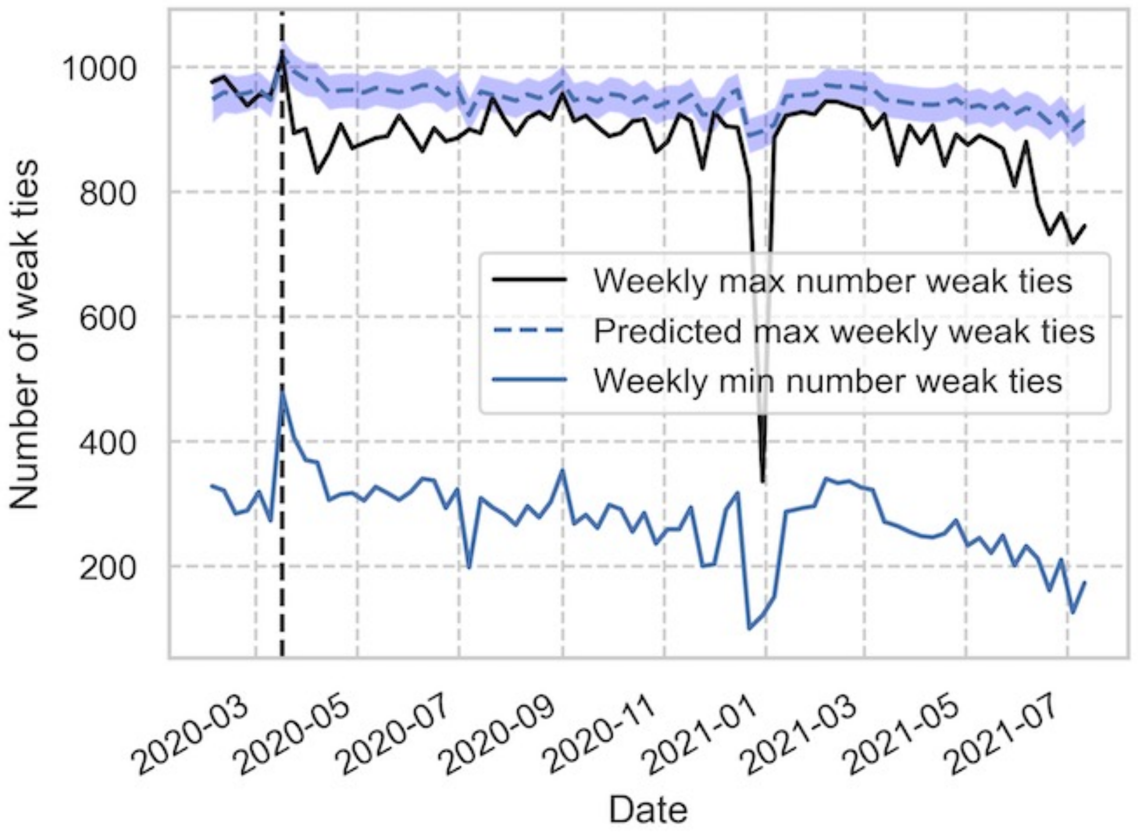
We construct a synthetic counterfactual from values of the time series prior to the intervention as well as **weekend data** (when most researchers are not in the office) to predict the effect of banning office-work during the weekdays.

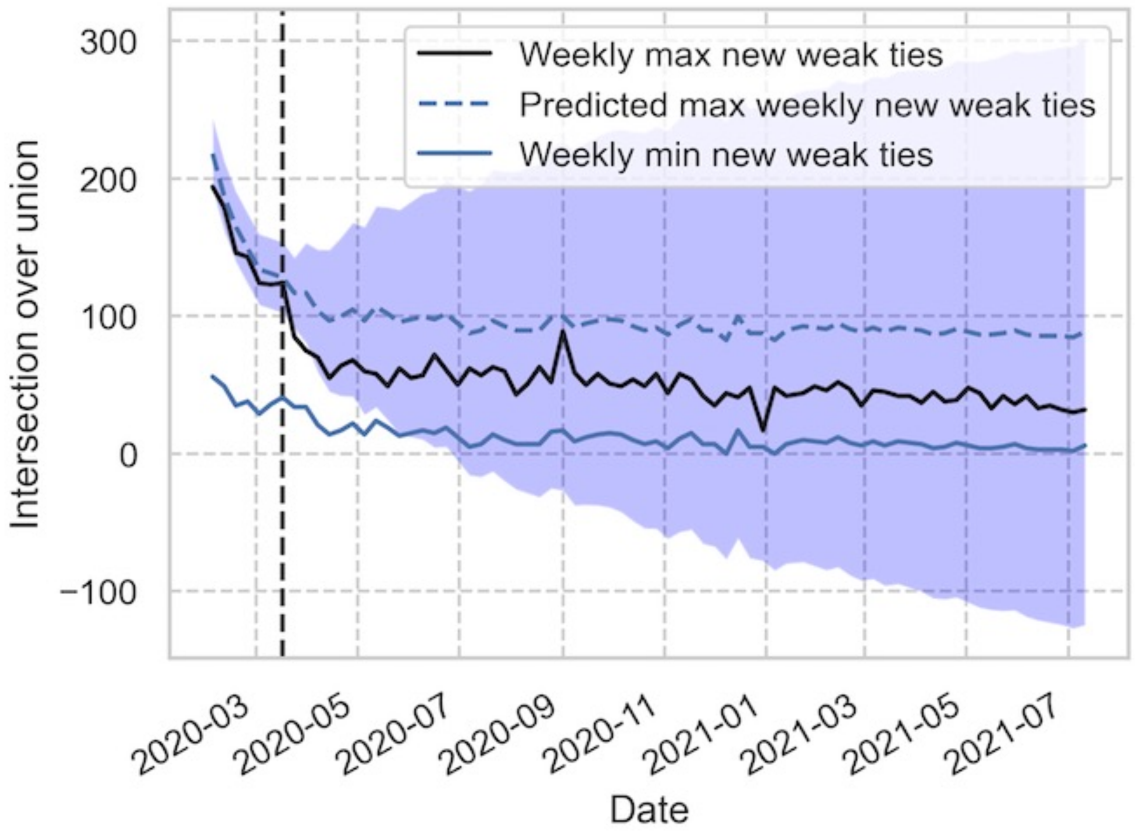




Bayesian structural time series

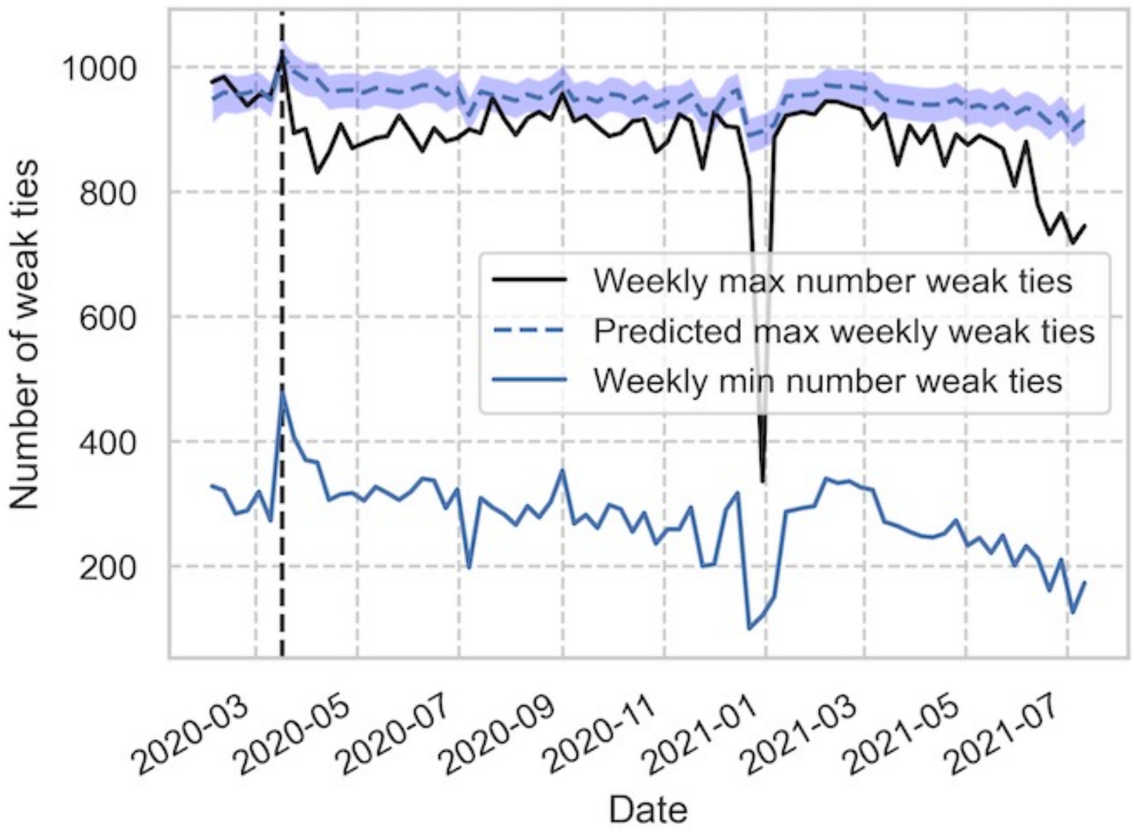
The shaded regions show a 95% posterior predictive interval, we want the shaded regions away from the black line in order to conclude statistically significant results.

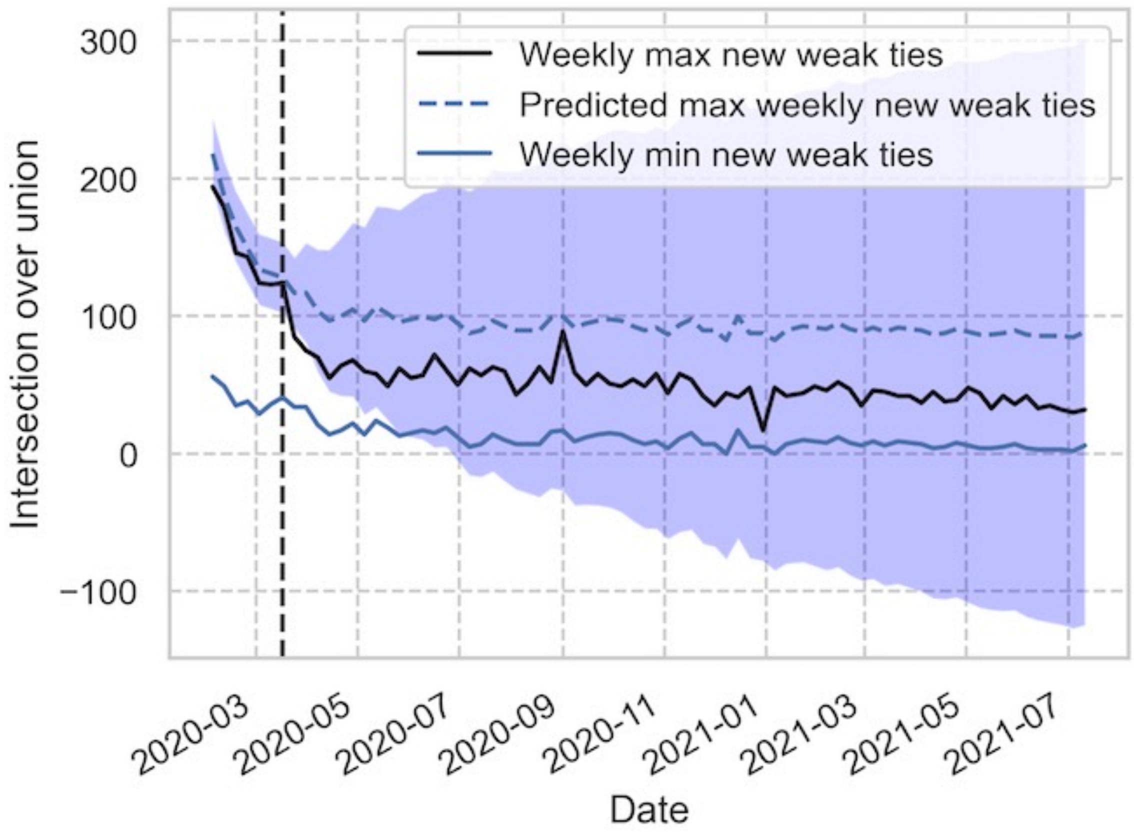




Bayesian structural time series

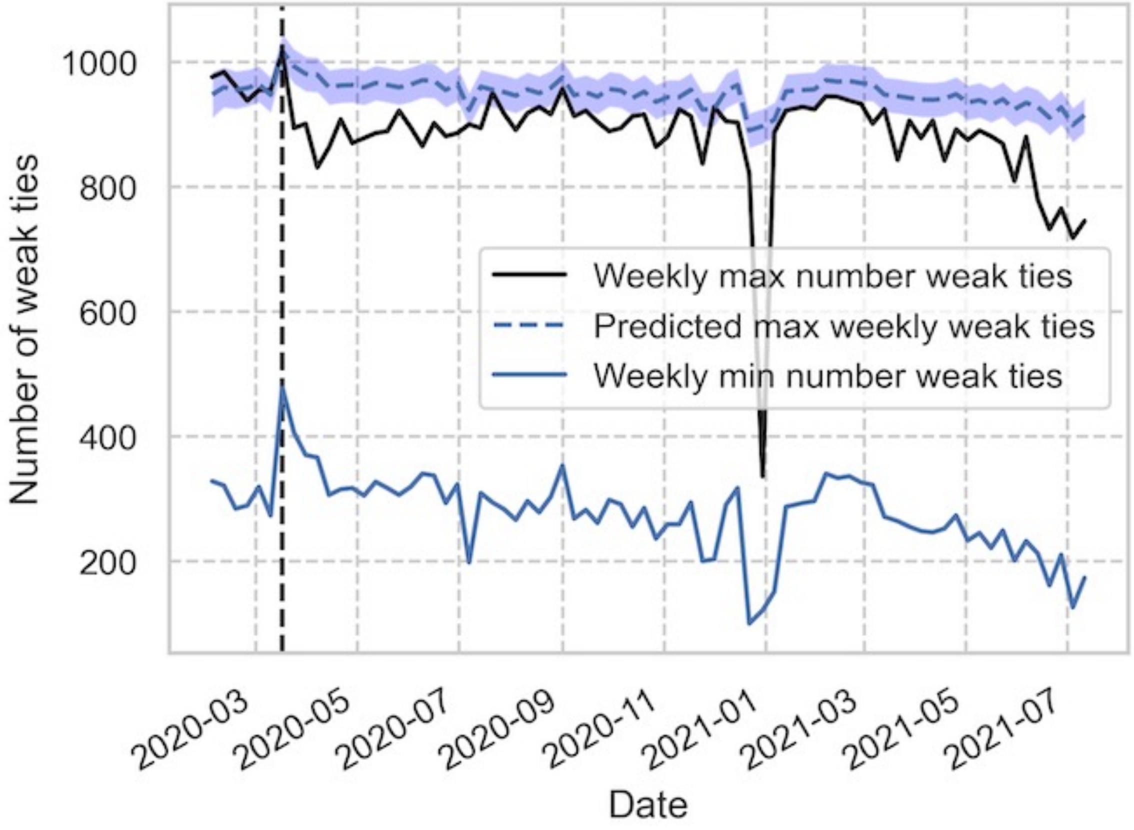
The number of local bridges after the implementation of mandatory remote work is significantly below the predicted values, indicating a significant and lost-lasting drop in the number of weak ties due to mandatory remote work.

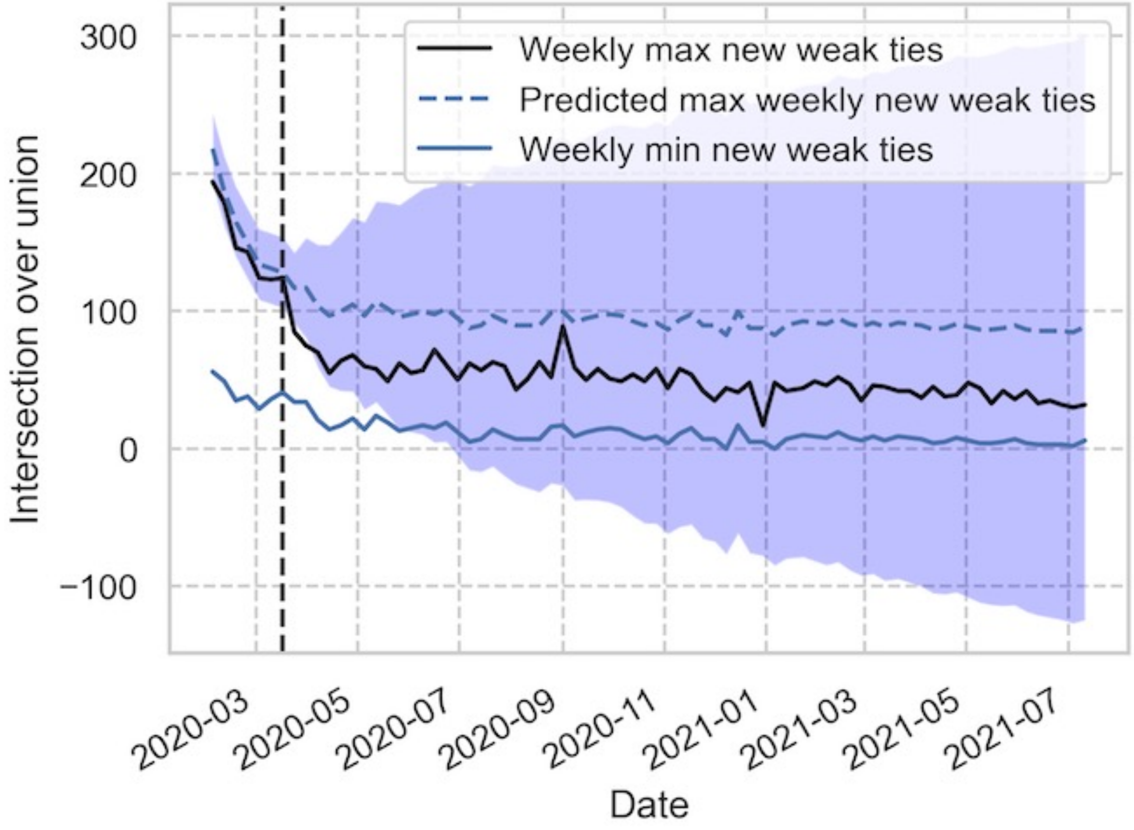




Bayesian structural time series

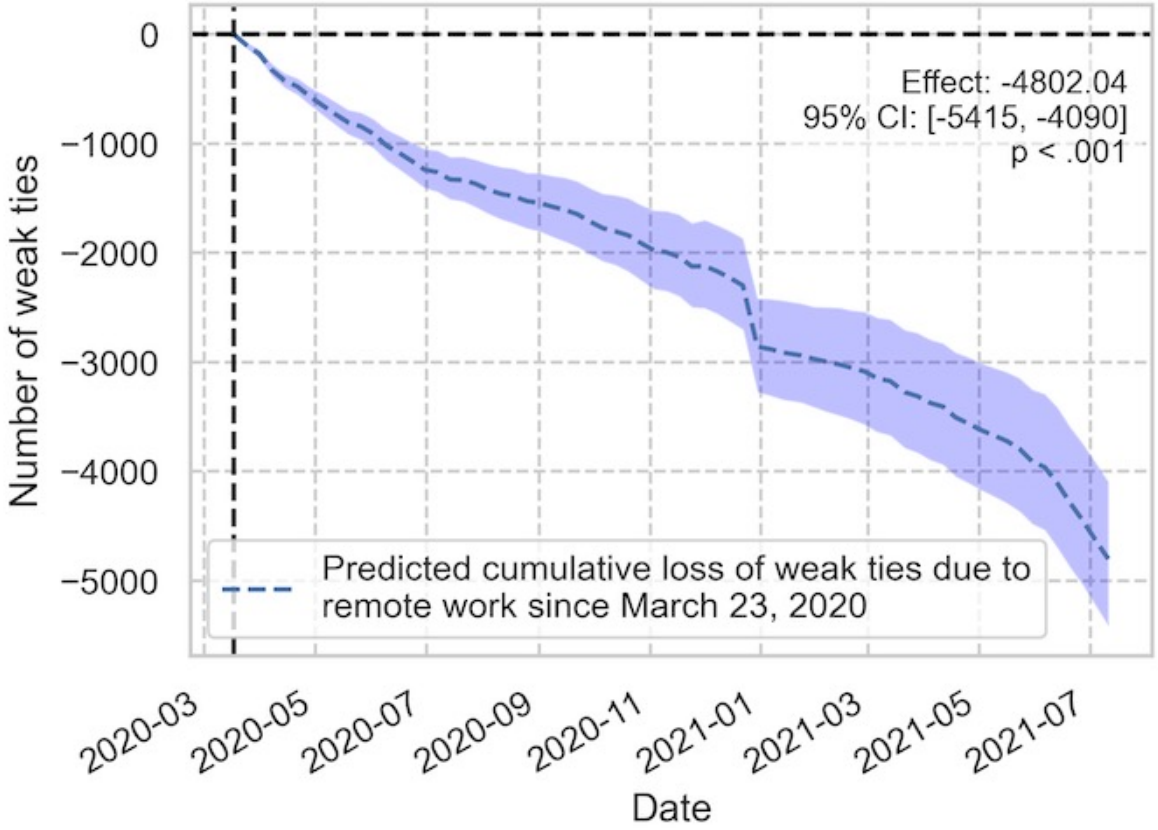
On the other hand, we don’t yet see any statistically significant effect of mandatory remote-work on the number of new weak ties. We’ll return to this soon.

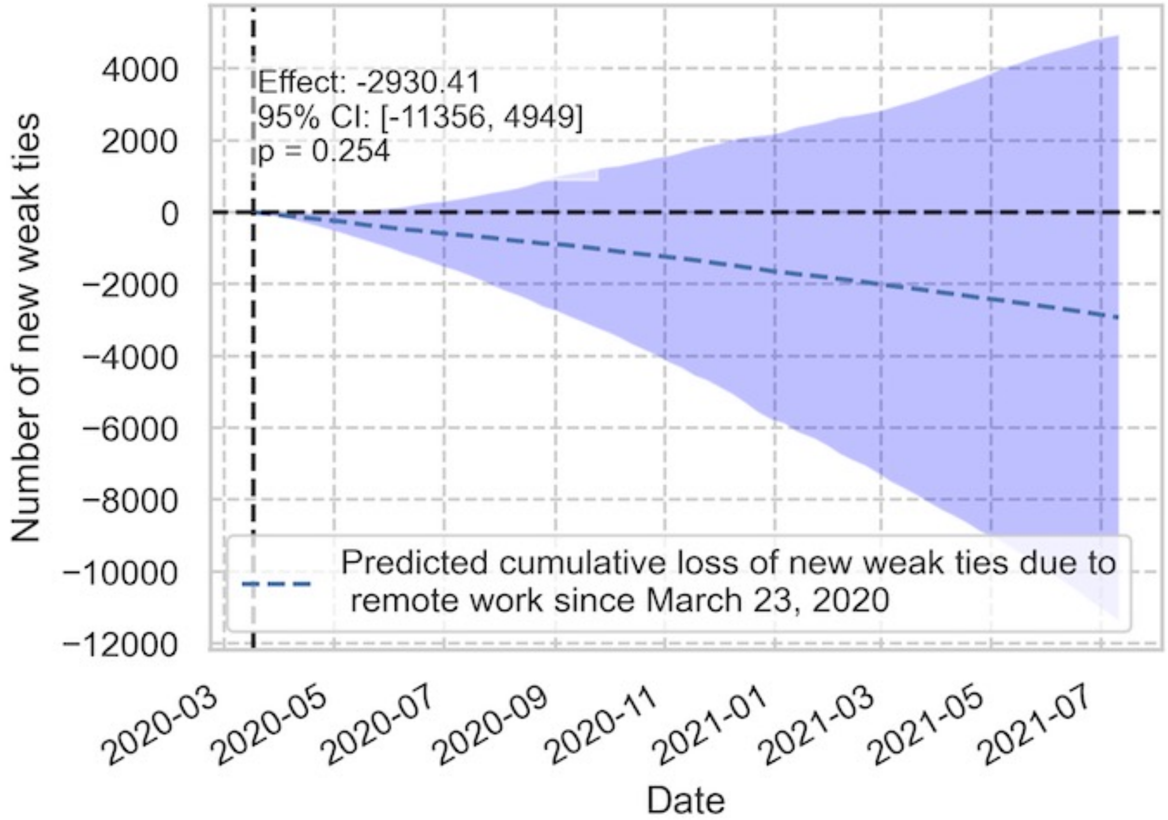




Bayesian structural time series: cumulative effect

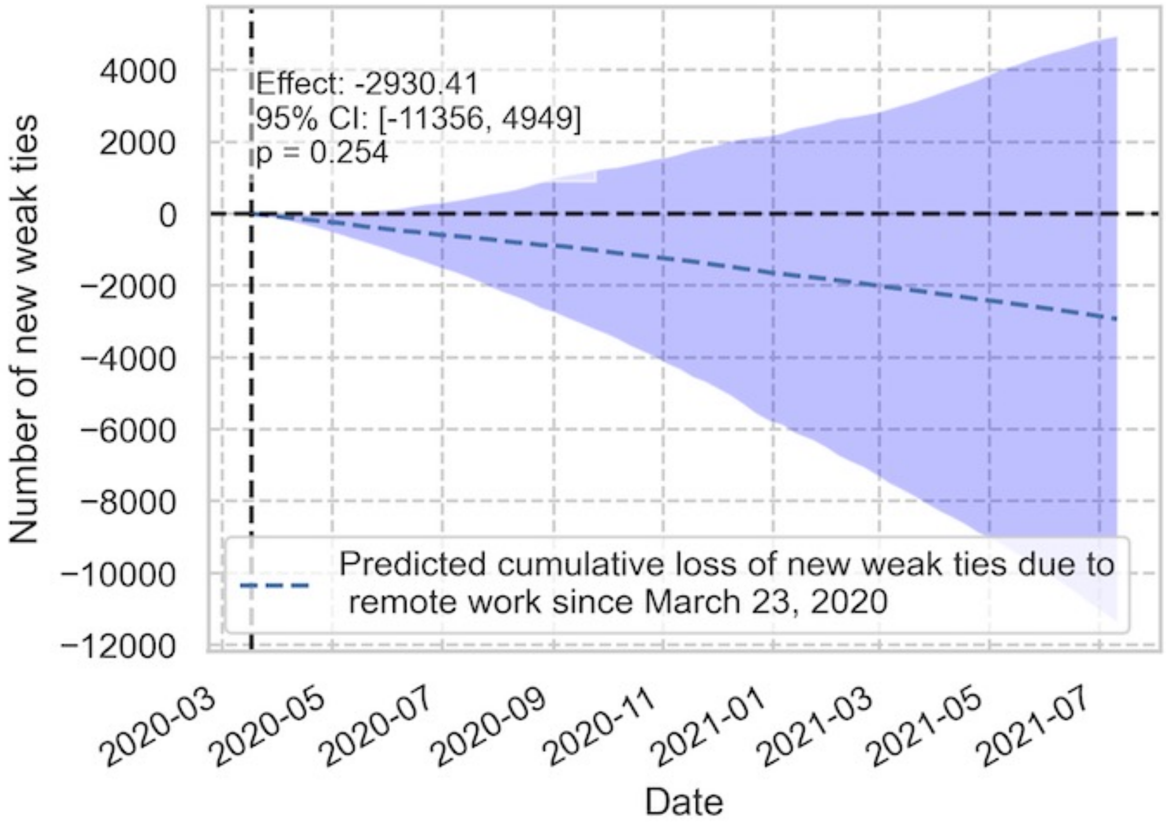
We can also plot the cumulative effect over time. In particular there is a statistically significant **drop of more than 4800** local bridges due to mandatory remote work over the course of the data.





But wait…

It appears as if there is no significant causal effect of mandatory remote work on the formation of new local bridges, is our hypothesis incorrect?



Stratifying by office distance

Let W_d denote the collection of local bridges in the daily email network on day d which have not appeared on any previous day.

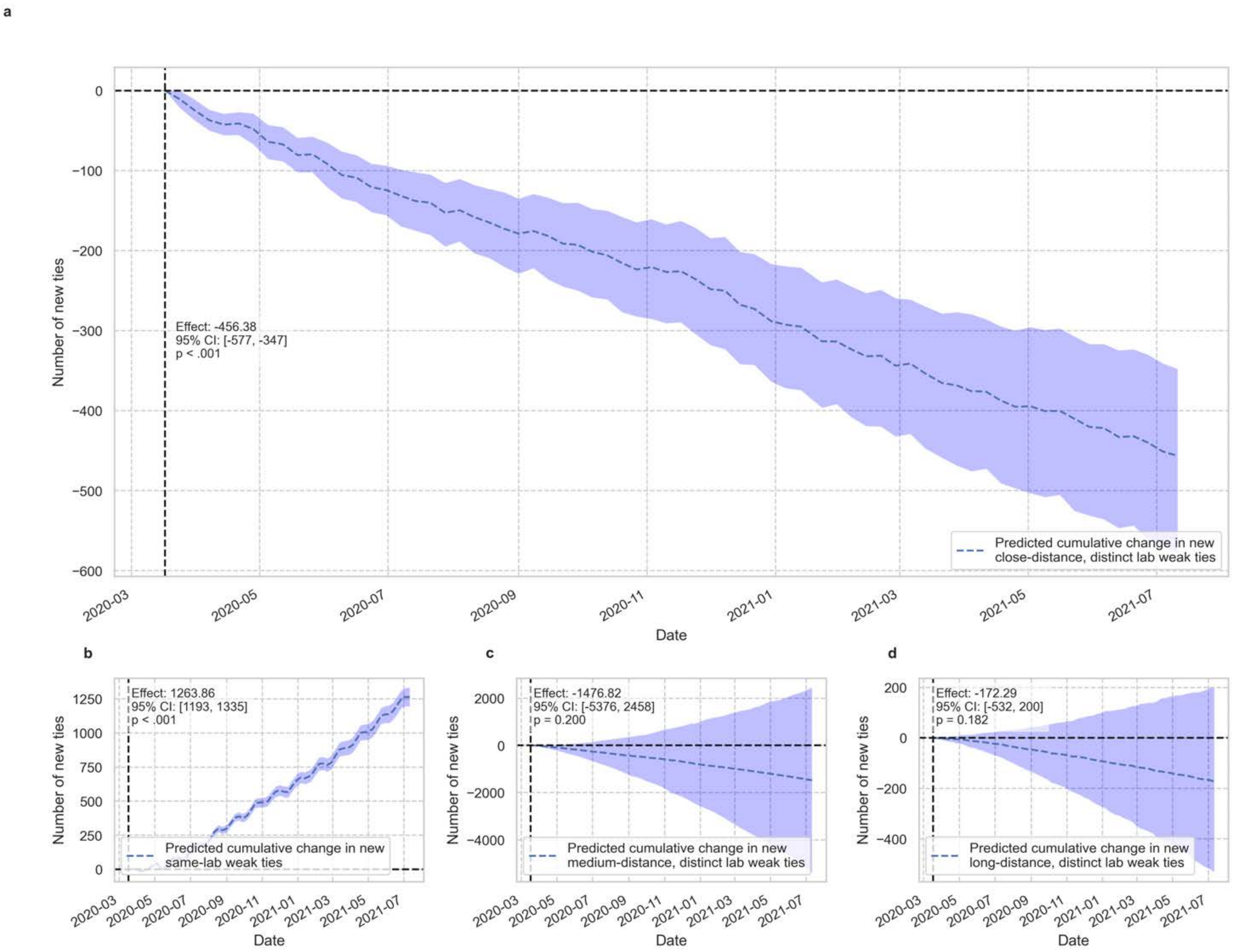
Divide W_d into four strata:

Same-office: ties between people with offices in the **same room**

Close-distance: ties between people whose offices are **between 0 and 150** meters apart

Medium-distance: ties between people whose offices are **between 150 and 650** meters apart

Long-distance: ties between people whose offices are between **more than 650** meters apart



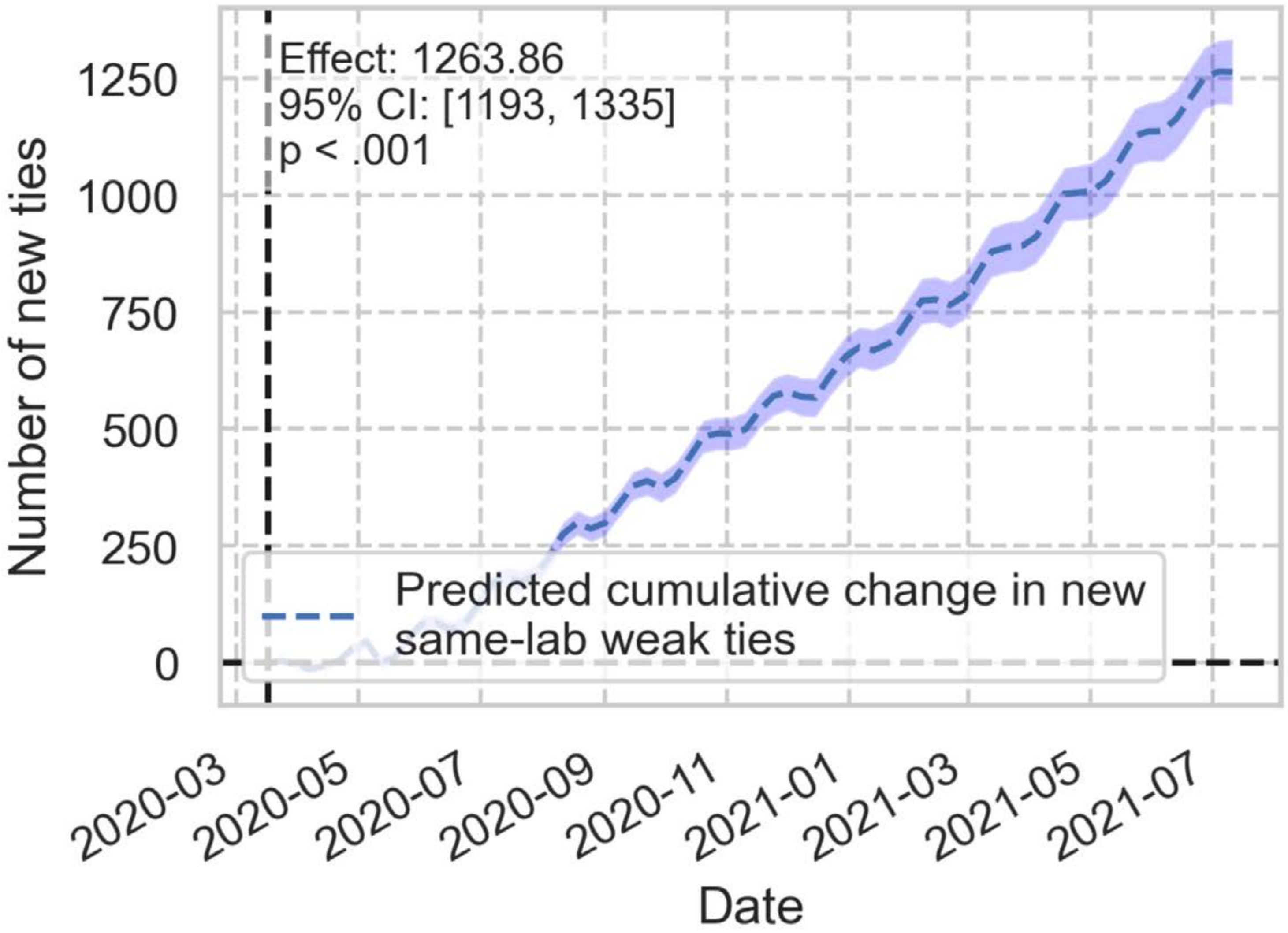
Same-office: ↑

Close-distance: ↓

Medium-distance: N.S.

Long-distance: N.S.

The number of new local bridges between **same-office** researchers increases compared to what's expected!



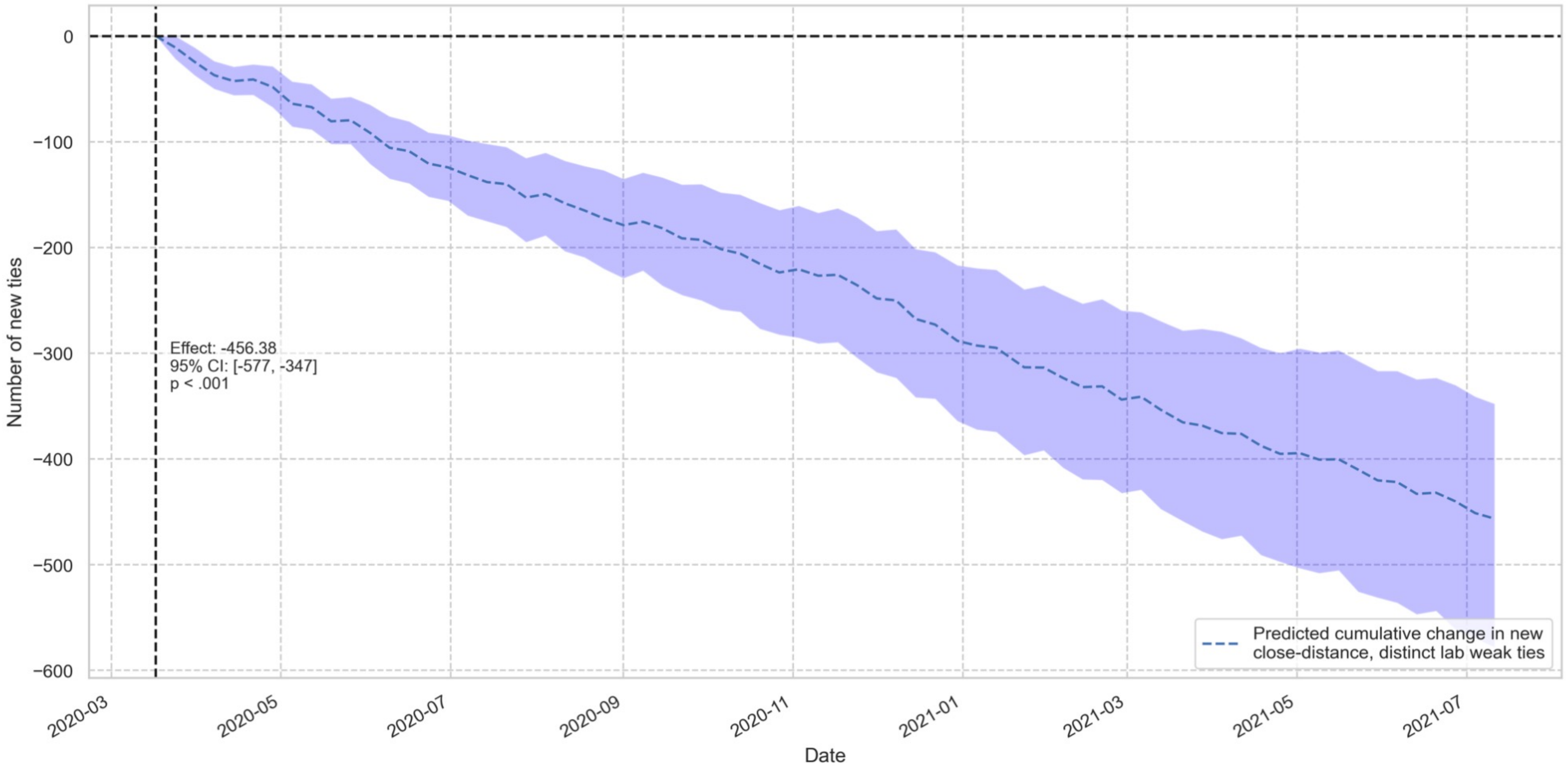
This is less interesting than it seems.

Every person in our dataset was required to be active before the pandemic, so people in the same lab would almost certainly already have met.

This could correspond to researchers who were previously working together having to use email to schedule Zoom meetings.

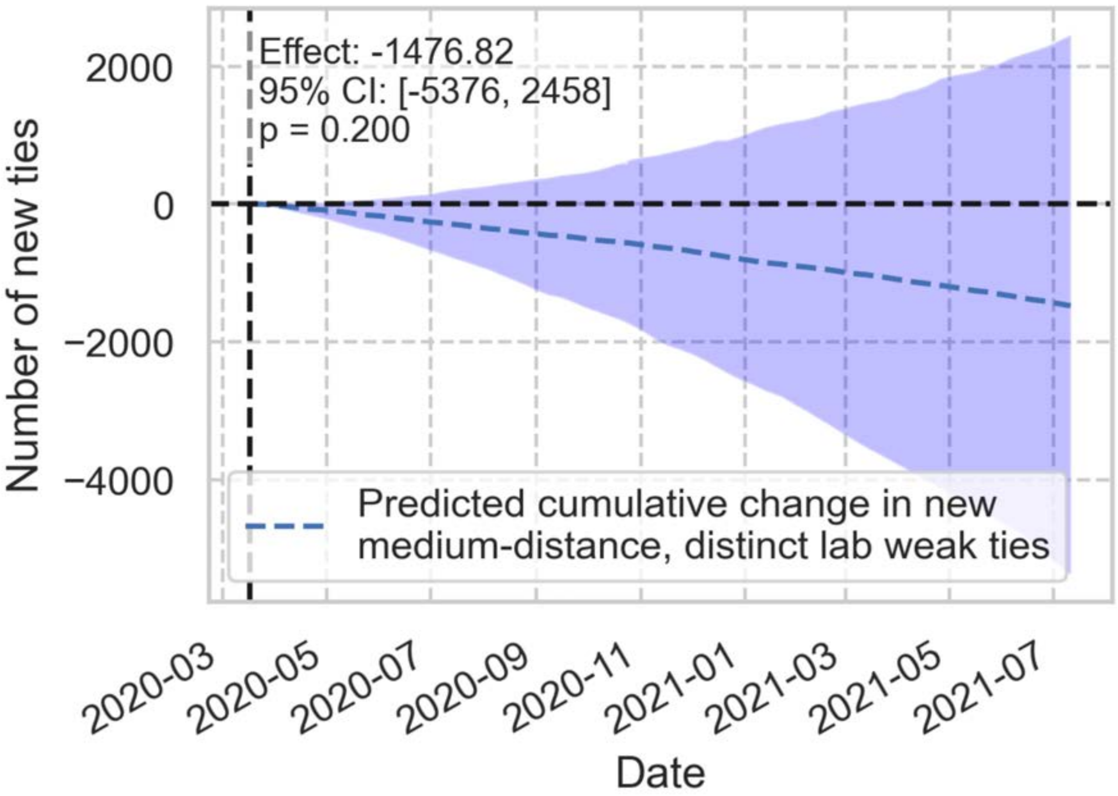
The cumulative number of new local bridges between **close-office** researchers decreases significantly.

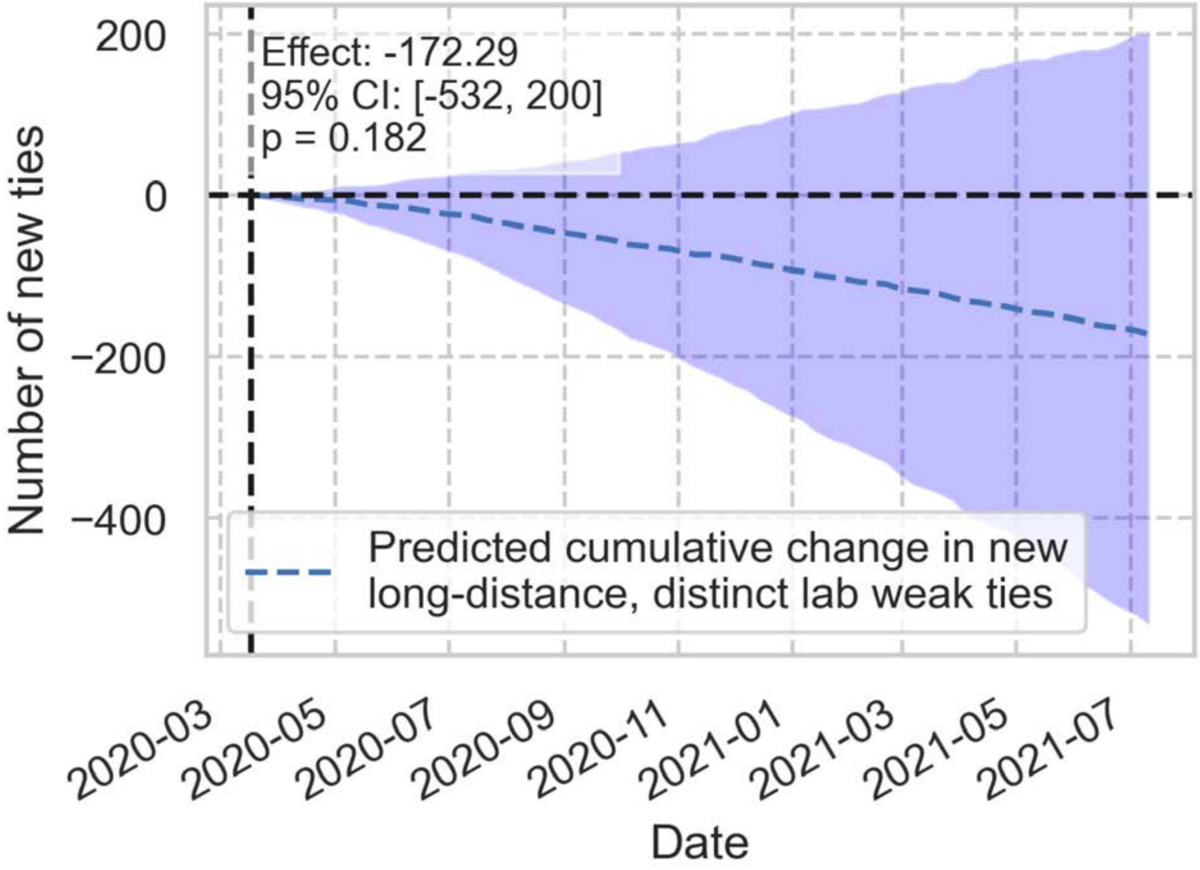
This is consistent with the idea that co-location causes new weak tie formation.



The number of new local bridges between **medium/long-distance** researchers does not change.

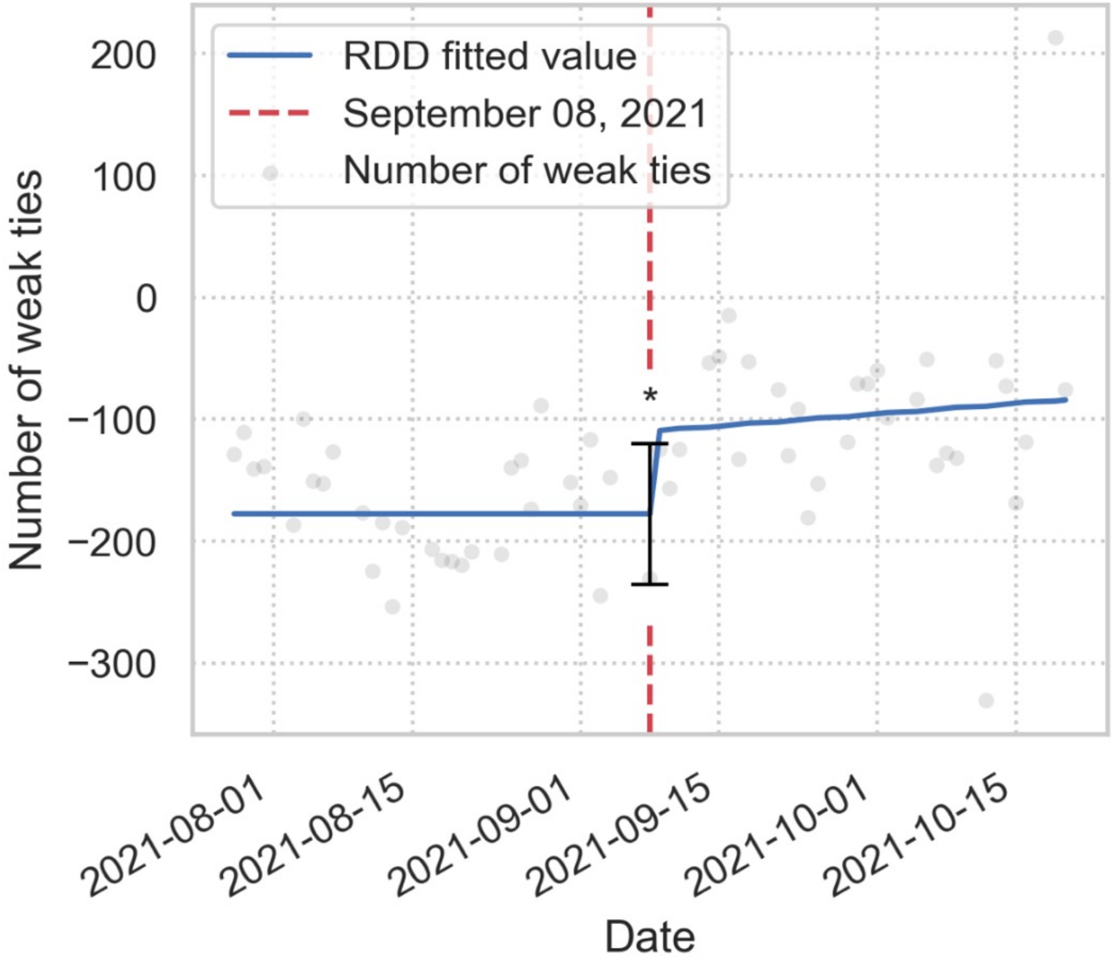
This is also consistent with the idea that co-location causes new weak tie formation, as we wouldn't expect researchers who work far away on campus to be affected by co-location even before the pandemic.

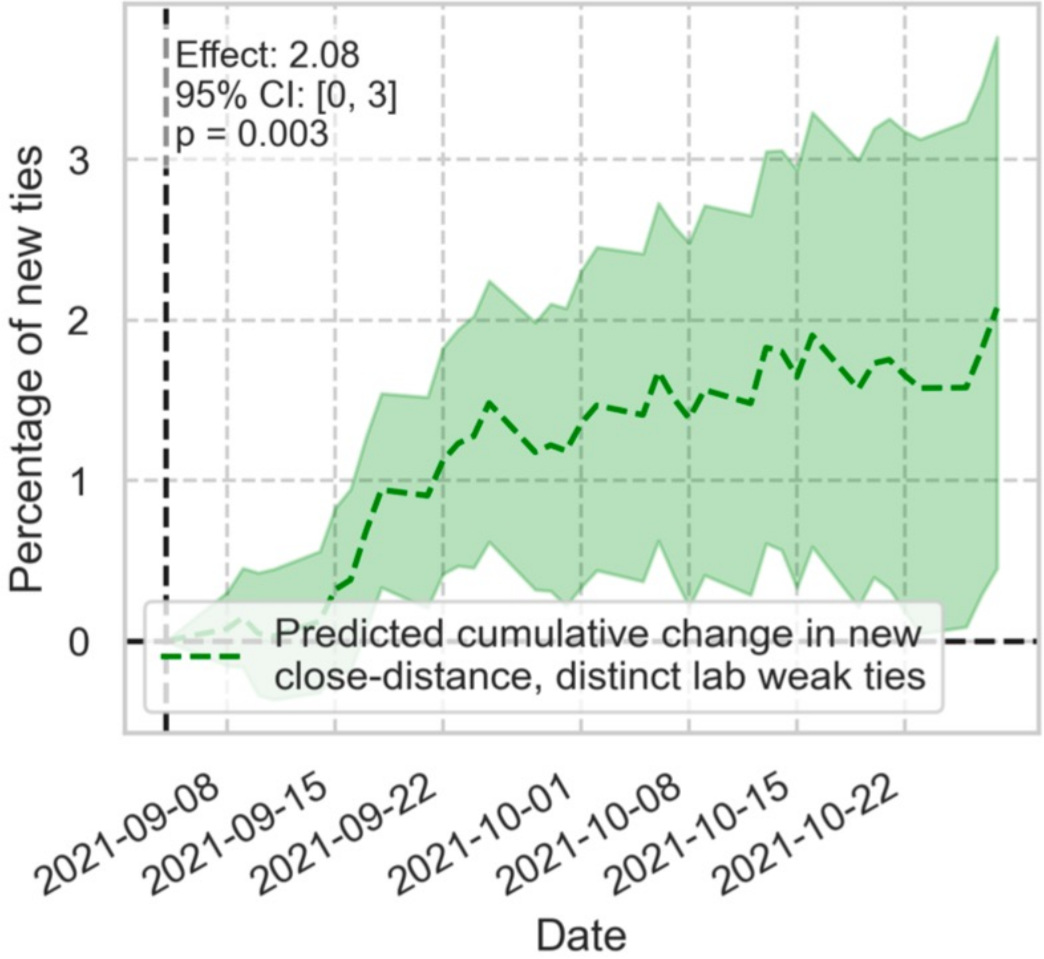




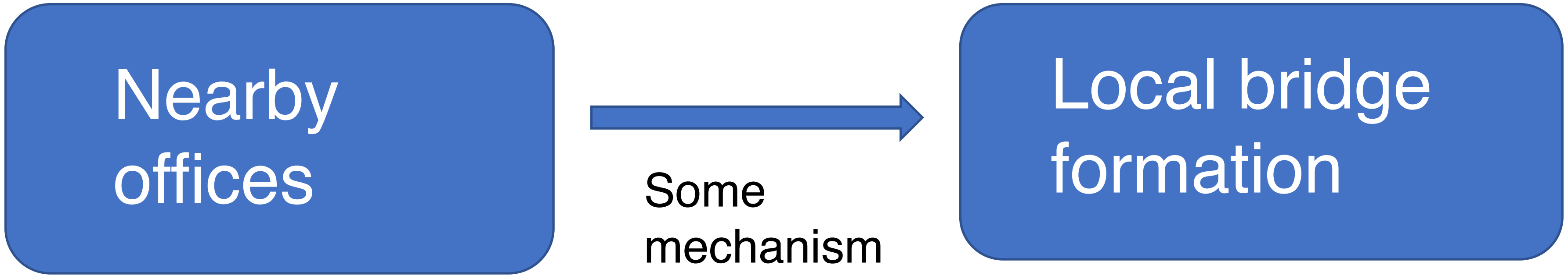
What happens when we re-introduce co-location?

There is a weak, but statistically significant increase in the number of weak ties at the start of the Fall 2021 semester compared to the Fall 2020 semester.





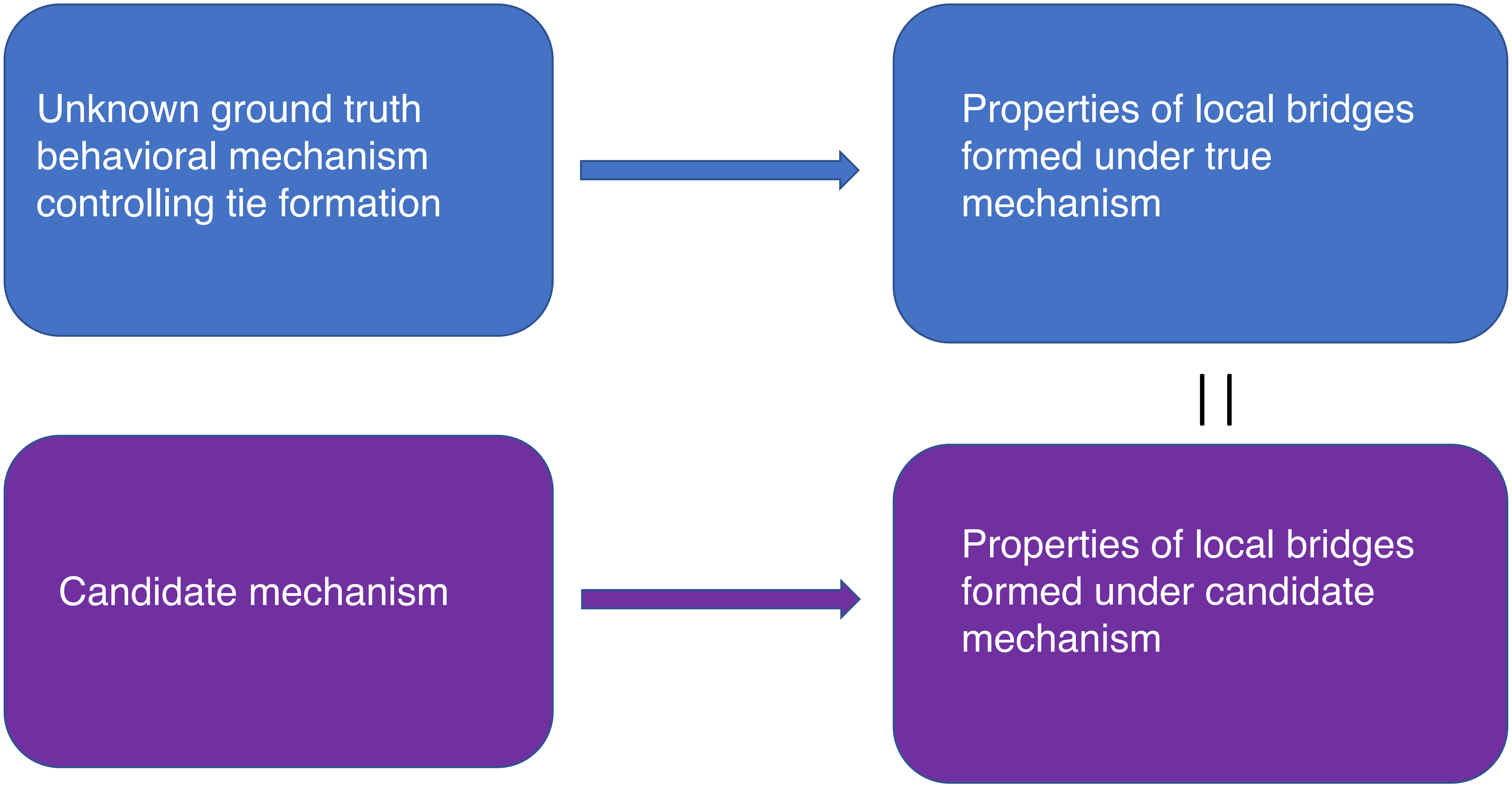
The story so far



Results so far are consistent with the **existence** of a mechanism via which lack of co-location causes local bridge deterioration.

Identifying a mechanism

Goals of a candidate mechanism



The proposed mechanism

Fix once and for all a collection of nodes N and a bucket of possible edges E between those nodes.

Each day, form a network by performing two steps of weighted draws without replacement from the bucket of edges.

In the first step, the probability of an edge is determined by:

The proposed mechanism

Fix once and for all a collection of nodes N and a bucket of possible edges E between those nodes.

Each day, form a network by performing two steps of weighted draws without replacement from the bucket of edges.

In the first step, the probability of an edge is determined by:

- **Focal closure** (are the researchers in the same unit?)

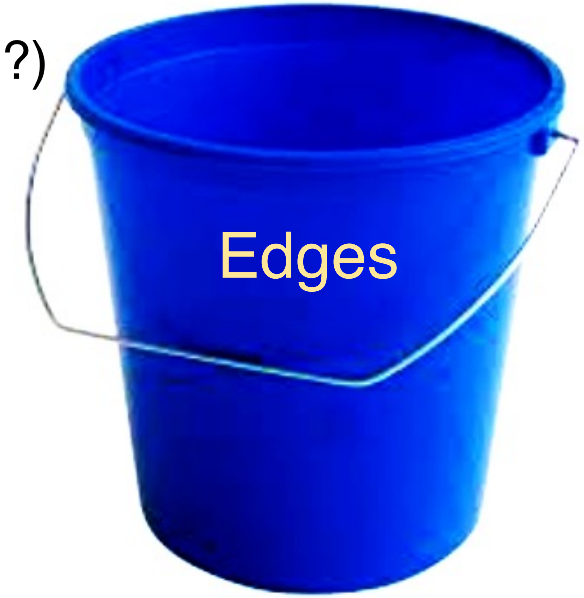
The proposed mechanism

Fix once and for all a collection of nodes N and a bucket of possible edges E between those nodes.

Each day, form a network by performing two steps of weighted draws without replacement from the bucket of edges.

In the first step, the probability of an edge is determined by:

- **Focal closure** (are the researchers in the same unit?)
- **Link centric preferential attachment** (has the edge been seen before?)



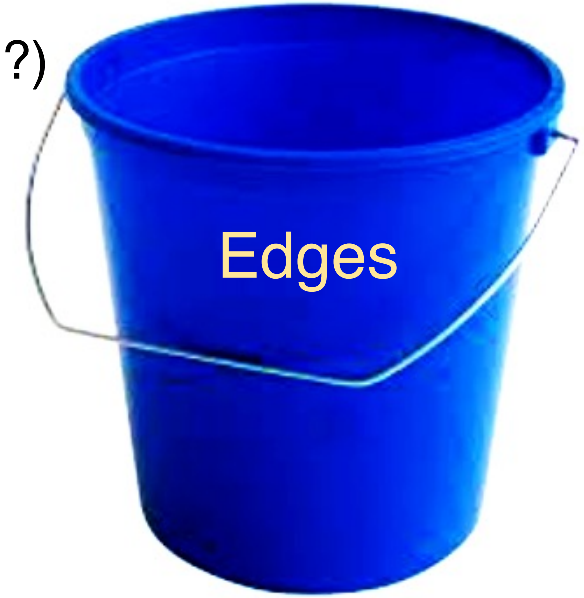
The proposed mechanism

Fix once and for all a collection of nodes N and a bucket of possible edges E between those nodes.

Each day, form a network by performing two steps of weighted draws without replacement from the bucket of edges.

In the first step, the probability of an edge is determined by:

- **Focal closure** (are the researchers in the same unit?)
- **Link centric preferential attachment** (has the edge been seen before?)
- **Co-location** (are the offices of the researchers close?)



The proposed mechanism

Fix once and for all a collection of nodes N and a bucket of possible edges E between those nodes.

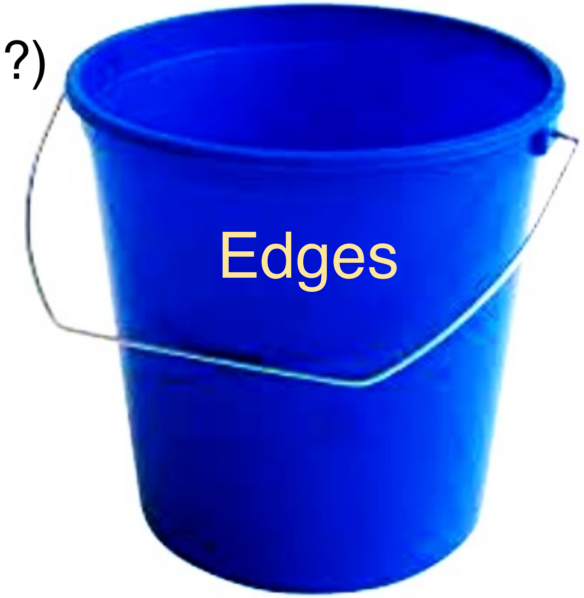
Each day, form a network by performing two steps of weighted draws without replacement from the bucket of edges.

In the first step, the probability of an edge is determined by:

- **Focal closure** (are the researchers in the same unit?)
- **Link centric preferential attachment** (has the edge been seen before?)
- **Co-location** (are the offices of the researchers close?)

In the second step, the probability of an edge is determined by the same factors plus

- **Triadic closure** (does the edge close a triangle in the network from step 1?)



The proposed mechanism

Step 1:

$$p_e \propto \tau(e) \sum_{Q\in\{P,O,N,D\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q\in\{P,O,N,D\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

Step 2:

$$p_e \propto \tau(e) \sum_{Q\in\{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q\in\{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

The co-location factors C_Q are multiplicative factors which either amplify or dampen the effects of the other factors Q based on whether the edge is between co-located researchers, represented by the binary variable $\tau(e)$.

The proposed mechanism

Step 1:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q \in \{P,O,N,D\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

Step 2:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q \in \{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

P controls the weekly periodicity of the model – edges are more likely to be selected if they appeared exactly one week ago.

The proposed mechanism

Step 1:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q \in \{P,O,N,D\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

Step 2:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q \in \{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

O corresponds to link-centric preferential attachment – edges are more likely to be selected depending on their frequency of past appearance.

The proposed mechanism

Step 1:

$$p_e \propto \tau(e) \sum_{Q\in\{P,O,N,D\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q\in\{P,O,N,D\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

Step 2:

$$p_e \propto \tau(e) \sum_{Q\in\{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q\in\{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

N is a small constant corresponding to the probability of choosing a previously unseen edge.

The proposed mechanism

Step 1:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D\}} C_Q \mathbb{1}_Q(e) Q + (1 - \tau(e)) \sum_{Q \in \{P,O,N,D\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

Step 2:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1 - \tau(e)) \sum_{Q \in \{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

D controls the probability of choosing an edge between people in the same research unit.

The proposed mechanism

Step 1:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D\}} C_Q \mathbb{1}_Q(e) Q + (1 - \tau(e)) \sum_{Q \in \{P,O,N,D\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

Step 2:

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1 - \tau(e)) \sum_{Q \in \{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

F is a large constant which makes edges that close triangles in the network formed in step 1 more likely to be chosen during step 2.

Simulated experiment

We simulate the empirical conditions as follows:

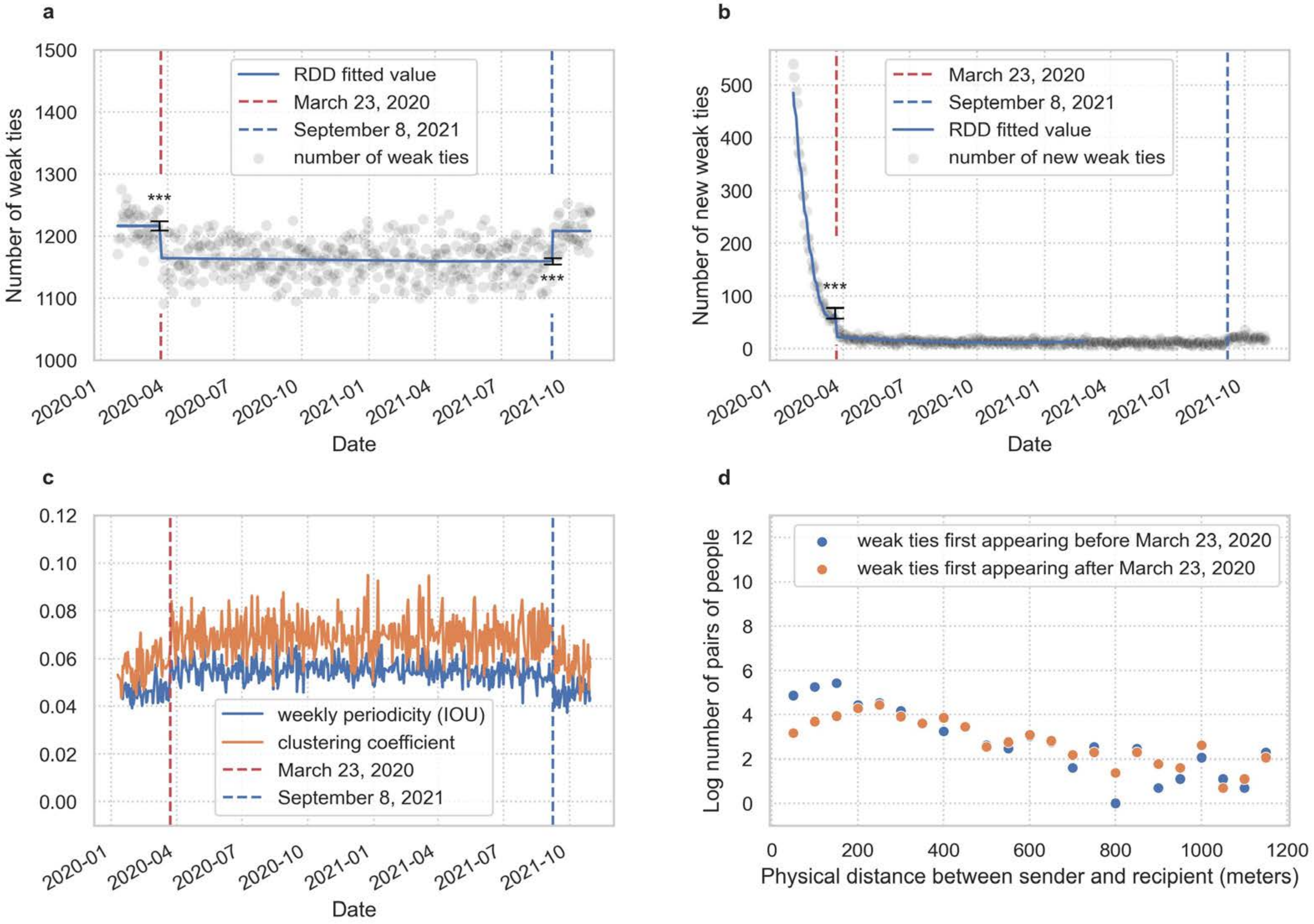
1. Initialize an edge memory dictionary with two weeks of real, weekday data from February 2020

2. Each day form a new network by the drawing edges according to the distribution outlined above, updating the edge memory dictionary as we go

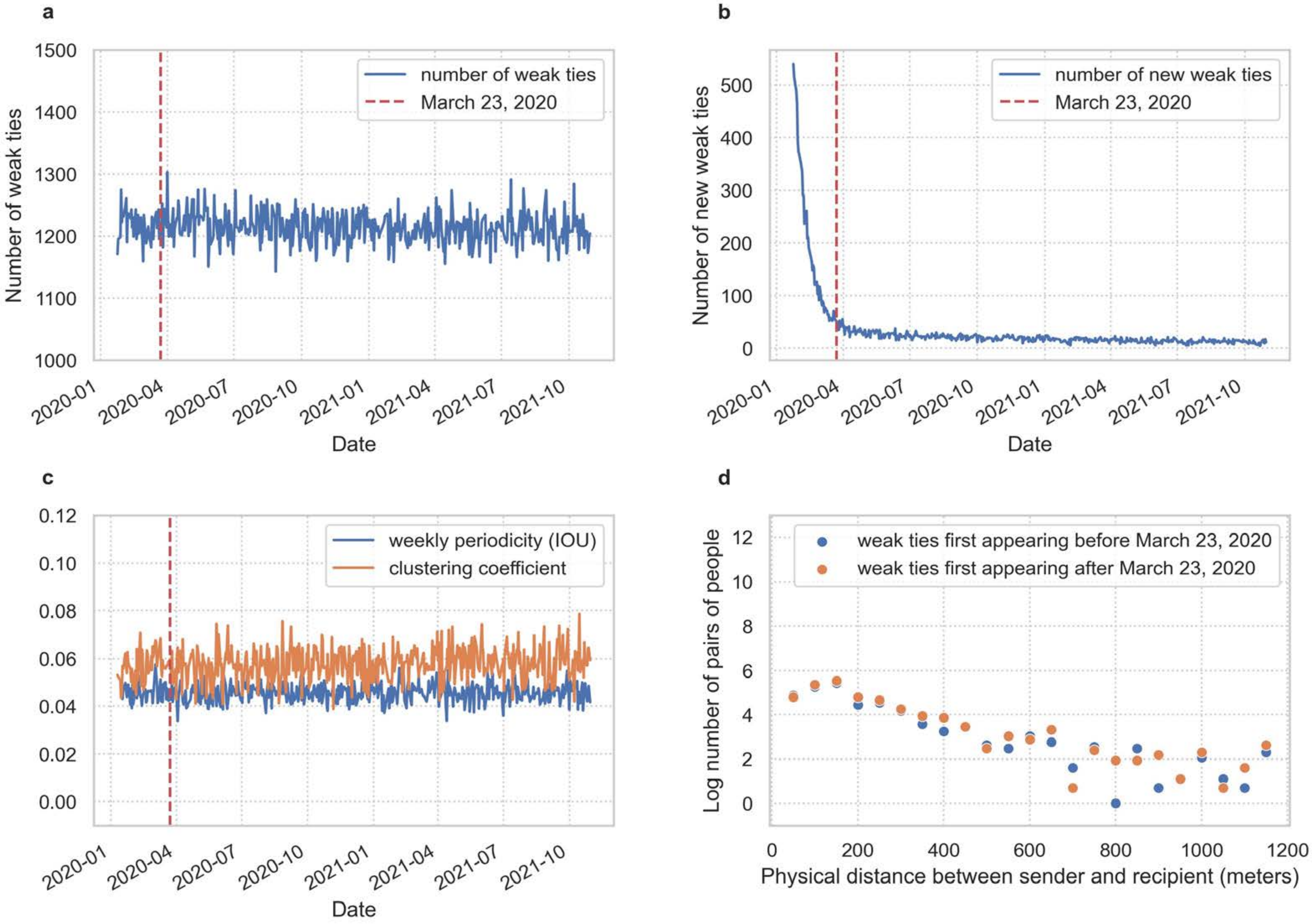
3. On March 23, 2020 remove the possibility for co-location by setting $\tau(e) = 0$ for all candidate edges e.

4. On September 8, 2021 add back the possibility for co-location by restoring $\tau(e)$ to its original value.

Simulated experiment: Does removing the possibility for co-location reproduce the **dynamics** observed in the empirical data?



By using our model and removing the possibility for co-location (setting τ to zero), we reproduce the empirical features of the data.



As a robustness check, if we leave τ unchanged, we observe no drops in the number of local bridges.

How does co-location affect each factor?

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q \in \{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

$C_P < 1$: co-location inhibits periodicity

$C_O = 1$: co-location has no effect on already established connections

$C_N > 1$: co-location promotes the formation of new ties

$C_D < 1$: co-location inhibits within-lab emails (because people talk in-person instead)

$C_F < 1$: co-location leads to less cliquey behavior

Implications

$$p_e \propto \tau(e) \sum_{Q \in \{P,O,N,D,F\}} C_Q \mathbb{1}_Q(e) Q + (1-\tau(e)) \sum_{Q \in \{P,O,N,D,F\}} \frac{1}{C_Q} \mathbb{1}_Q(e) Q$$

$C_P < 1$: co-location reduces redundancy of information

$C_N > 1$: co-location promotes the formation of new ties

$C_F < 1$: co-location leads to less cliquey behavior

Co-location is important for updating the sources from which researchers receive novel information.

Co-location is important for re-organization of research networks over time.

Implications

Co-location is important for updating the sources from which researchers receive novel information.

- Given that information tends to spread more slowly through email networks than predicted by typical epidemic models (Iribarren-Moro), missing local bridges which are capable of spreading information to distant corners of a network is disastrous.

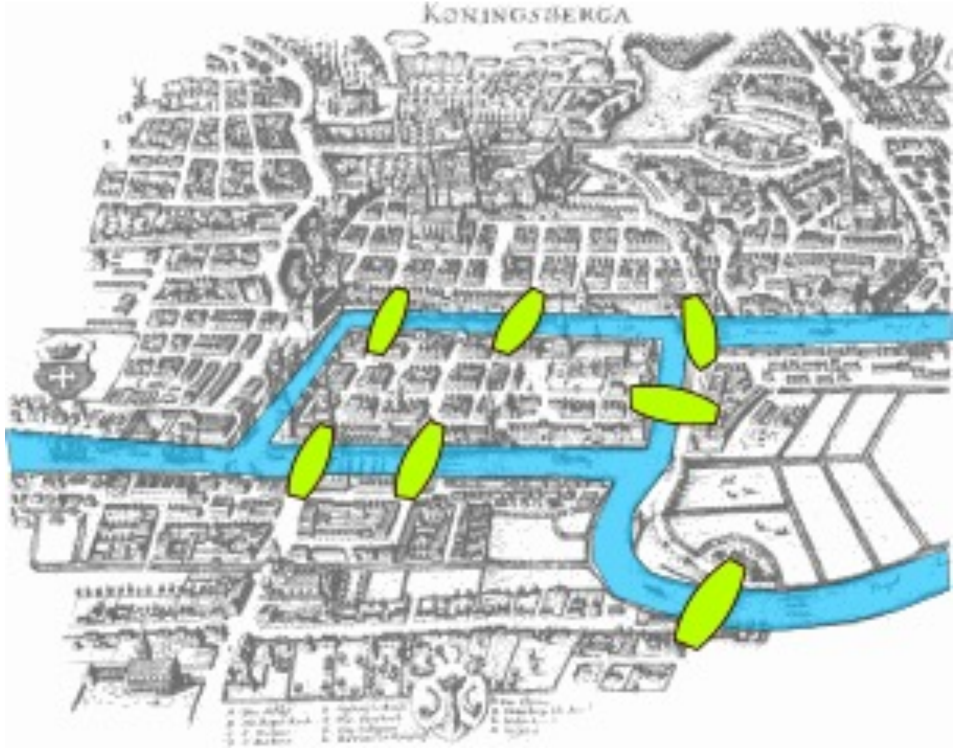
Co-location is important for re-organization of research networks over time.

- The ability to re-organize is vital for large-scale human cooperation when approaching complex tasks (Rand-Arbesman-Christakis).

A brief introduction to networks

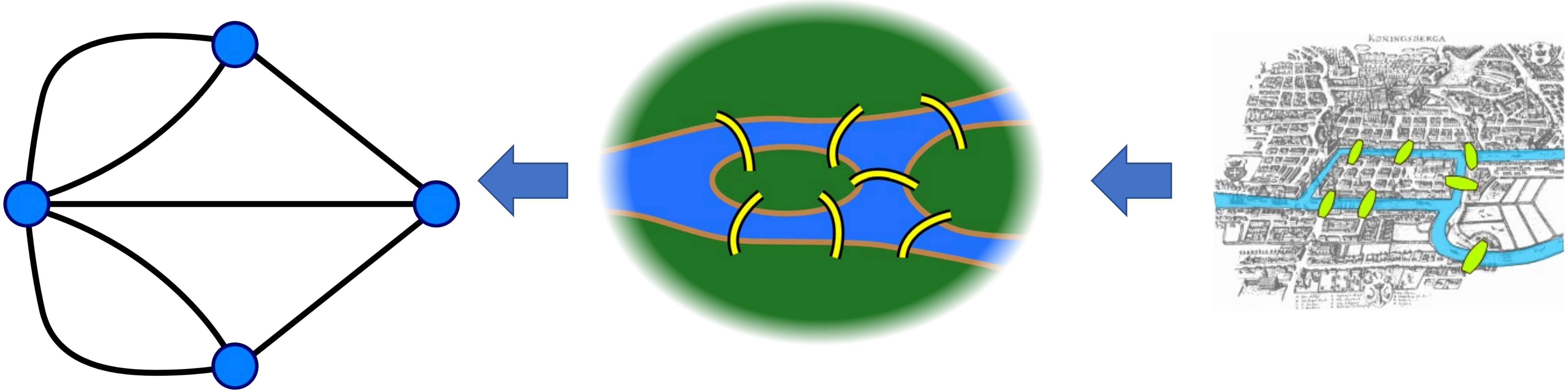
The history of network science

7 bridges of Köningsberg

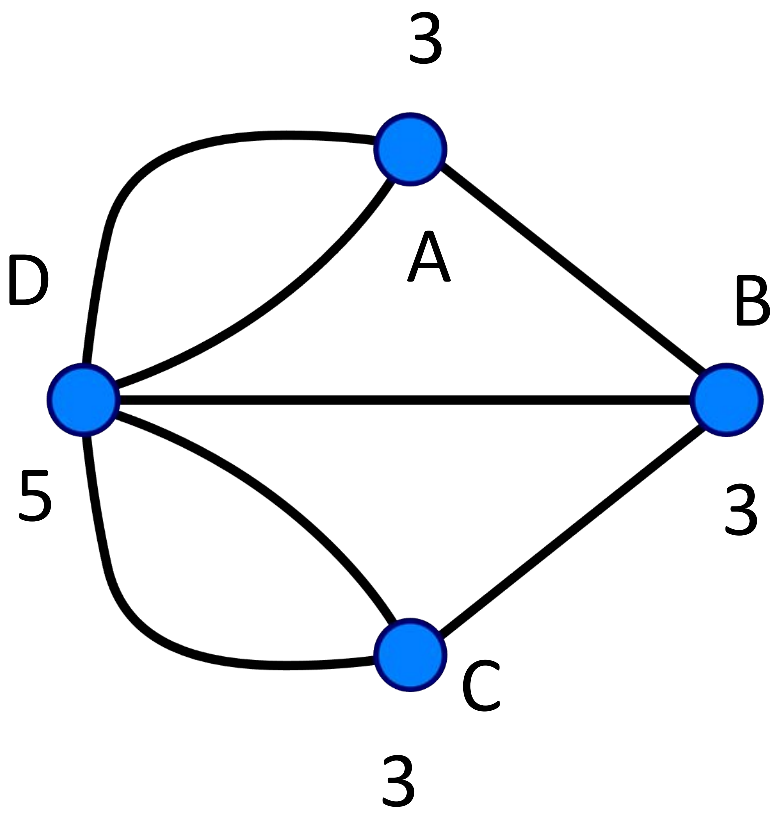


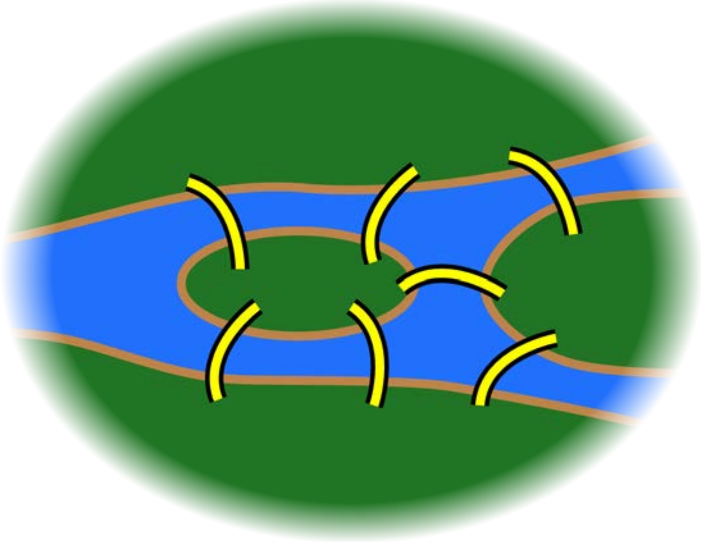
Leonhard Euler and the 7 bridges of Köningsberg

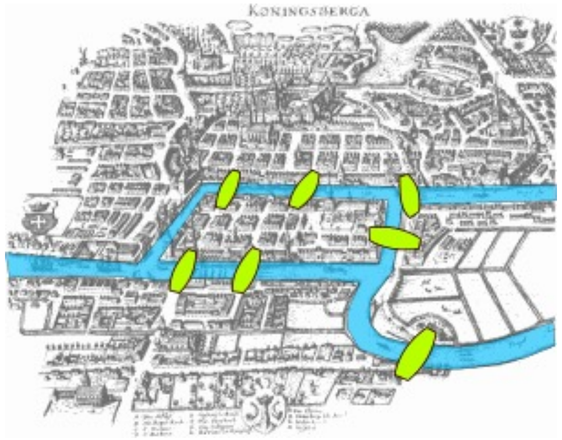
7 bridges of Köningsberg



7 bridges of Köningsberg





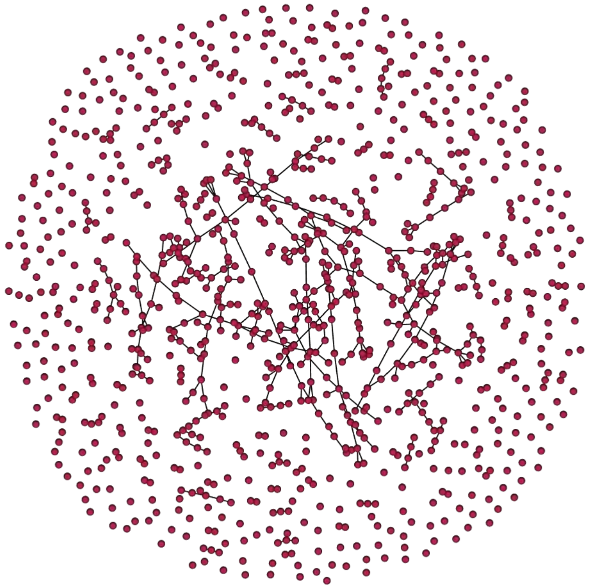


Random graphs

In 1959, Erdős and Renyi (in parallel with Gilbert) began the systematic study of random graphs

Today, the phrase « random graph » typically refers to G(n,p) – a graph with n nodes such that each pair of nodes is connected independently with probability p.

Erdos and Renyi showed that many properties of random graphs satisfy thresholding phenomena – there is a critical threshold of edge probability where the graph property suddenly changes.

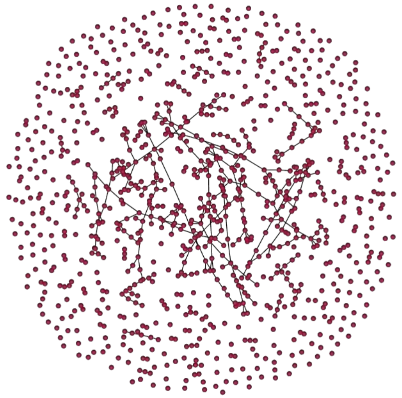


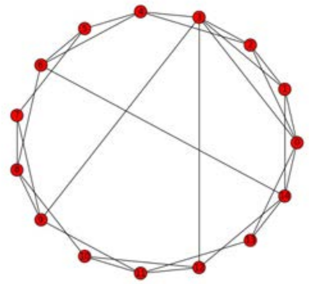
Network models

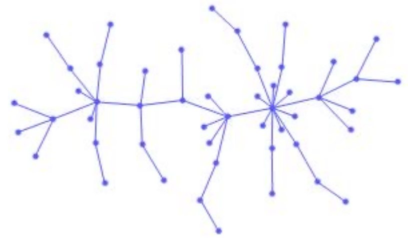
The Erdős-Renyi-Gilbert model of random graphs produces networks with different properties than most real-world social networks

In 1998 Watts and Strogatz introduced a new network model better representing community structures observed in real life networks.

In 1965 Price introduced a new network model explaining the observed power law degree distribution of citation networks. In 1998 this model was popularized by Albert and Barabsi, who introduced the phrase « preferential attachment ».







Learning objectives

Understand the basic definitions of network science

Load and manipulate social networks in python

Compute standard network metrics for given communication/social networks

Understand the real-world implications of network structure

Why networks?

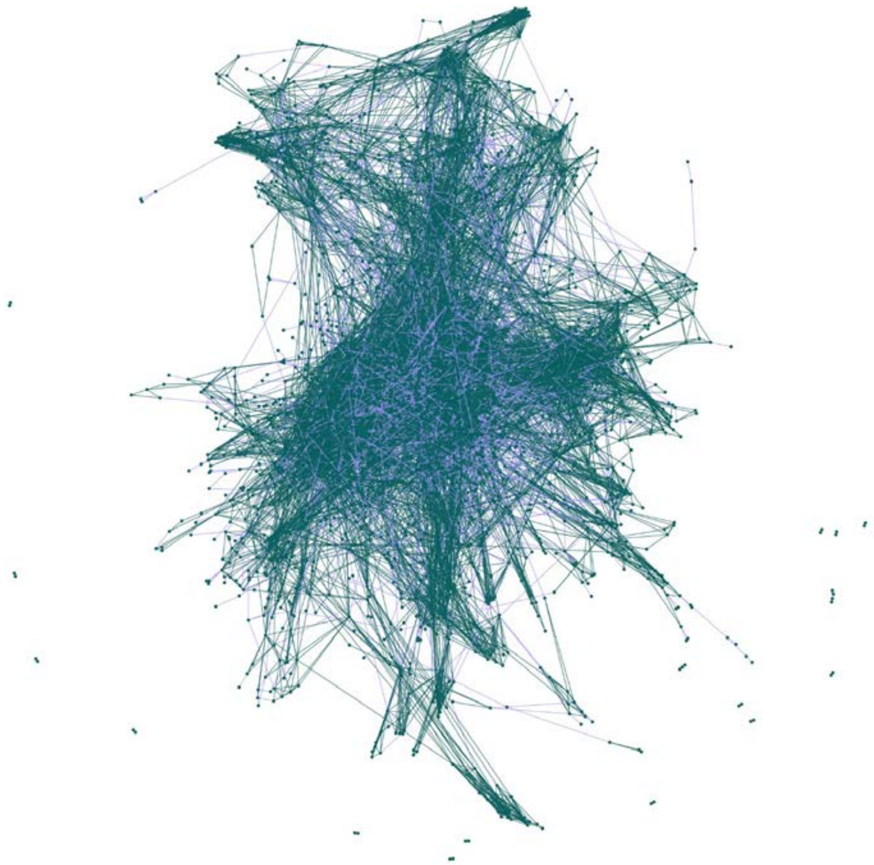
In general, asking the question, « Does this data have a network representation? » can be extremely fruitful.

The study of networks is both mathematically and algorithmically mature, so phrasing problems in the language of networks gives one access to a host of tools and methodologies.

A good example of using the language of networks to get results on a seemingly unrelated problem was the lab’s work on the « minimum fleet problem » (which I was not part of).

Networks

Definition: A network G is a set N of nodes together with a set $E \subseteq 2^N$ of pairs of nodes called edges.



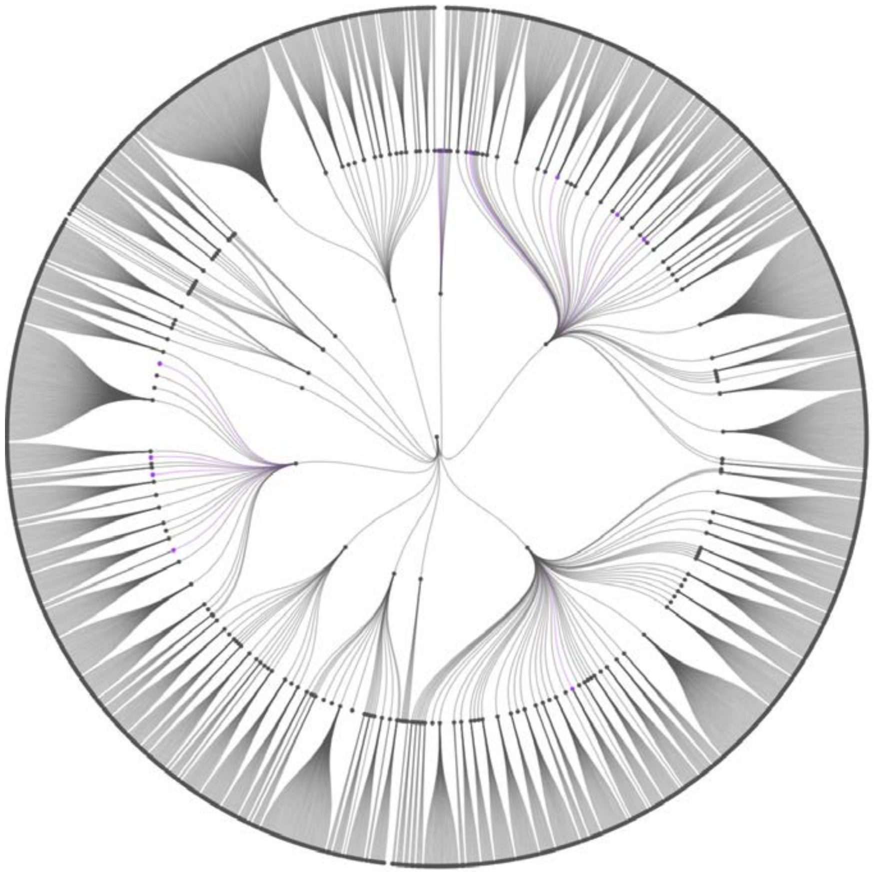
Networks are useful for representing **symmetric relationships**.

For example, a network might represent:

- landmasses and bridges connecting them
- friendship relations in a social network
- coauthor relationships between researchers

Directed networks

Definition: A directed network G is a set N of nodes together with a set $E \subseteq N \times N$ of directed edges.



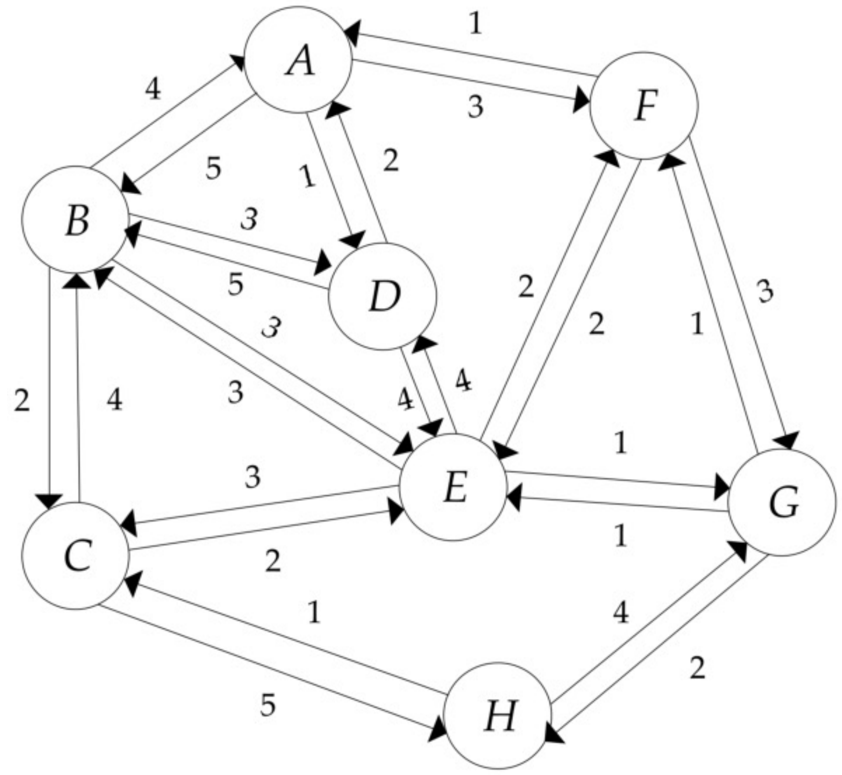
Directed networks are useful for representing **actions, transitions,** and **causal relationships**.

For example, a directed network might represent:

- paper citations (node A cites node B)
- human migrations (people from location A travel to location B)
- Neural networks (the activation of neuron A causes the activation of neuron B)

Weighted networks

Definition: A (directed) network G is weighted if there is a function $w : E \rightarrow \mathbb{R}$ which assigns to each edge a weight.



Both directed and undirected networks can be weighted. Weights may represent things like counts, speeds, capacities, relationship strength, etc.

Group participation

Question: What are some other examples of phenomena or data that can be represented with a **weighted network**?

Question: What are some other examples of phenomena or data that can be represented with a **weighted directed network**?

Basic network metrics

Number of nodes:

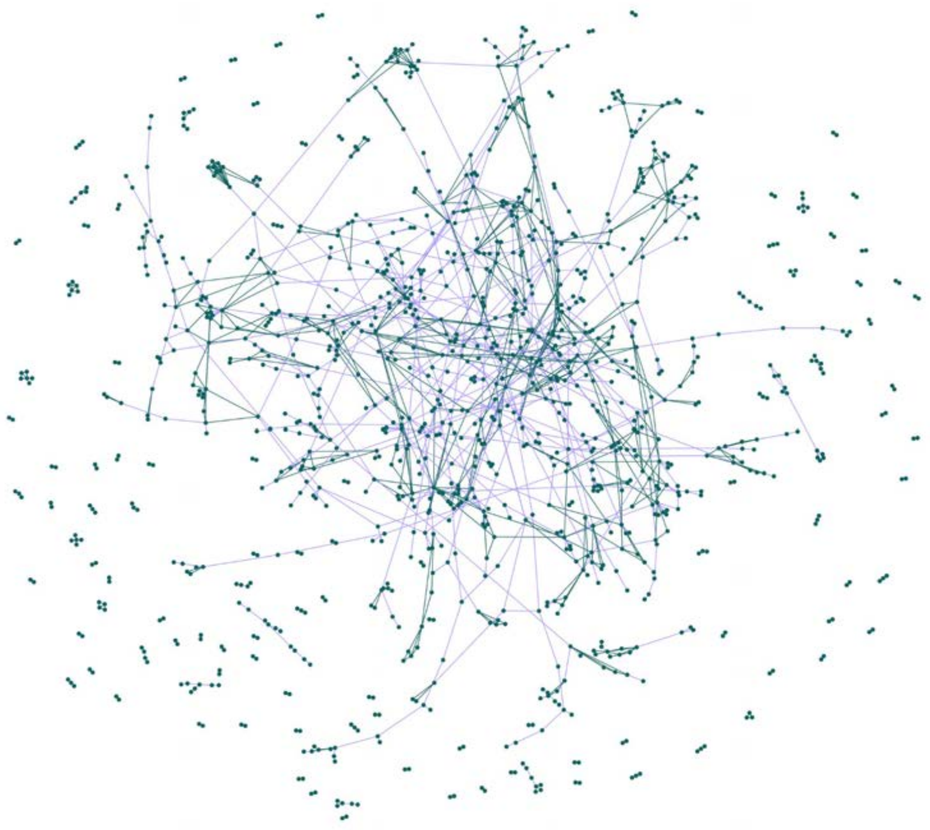
Number of edges:

Number of components:

Size of largest component:

Average degree:

Clustering coefficient:



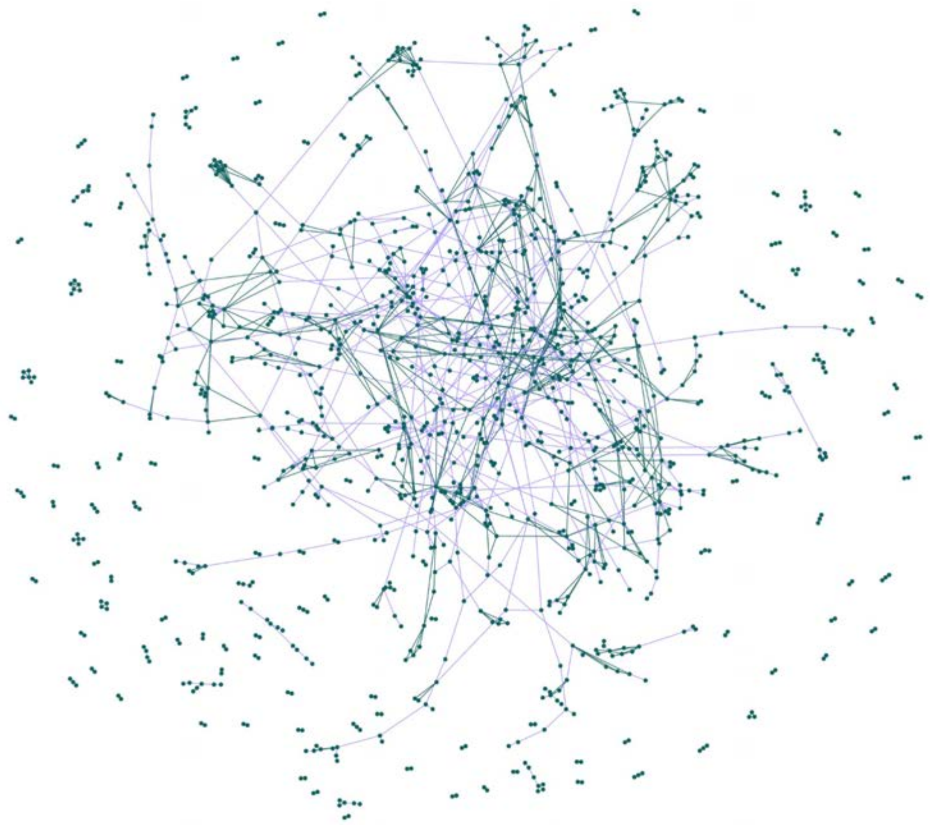
Basic network metrics

The number of nodes and number of edges are self explanatory.

- The number of components is the number of “islands” in the network – the maximal subsets such that any node in the subset can be reached from any other node in the subset through a path.

- The degree of a node is the number of edges connected to that node. For communication networks this answers the question “on average how many others does each person talk to?”

- The average clustering coefficient is the average proportion of triangles that each node belongs to. “What percentage of the people I talk to talk to each other?”



Basic network metrics

Number of nodes: 543

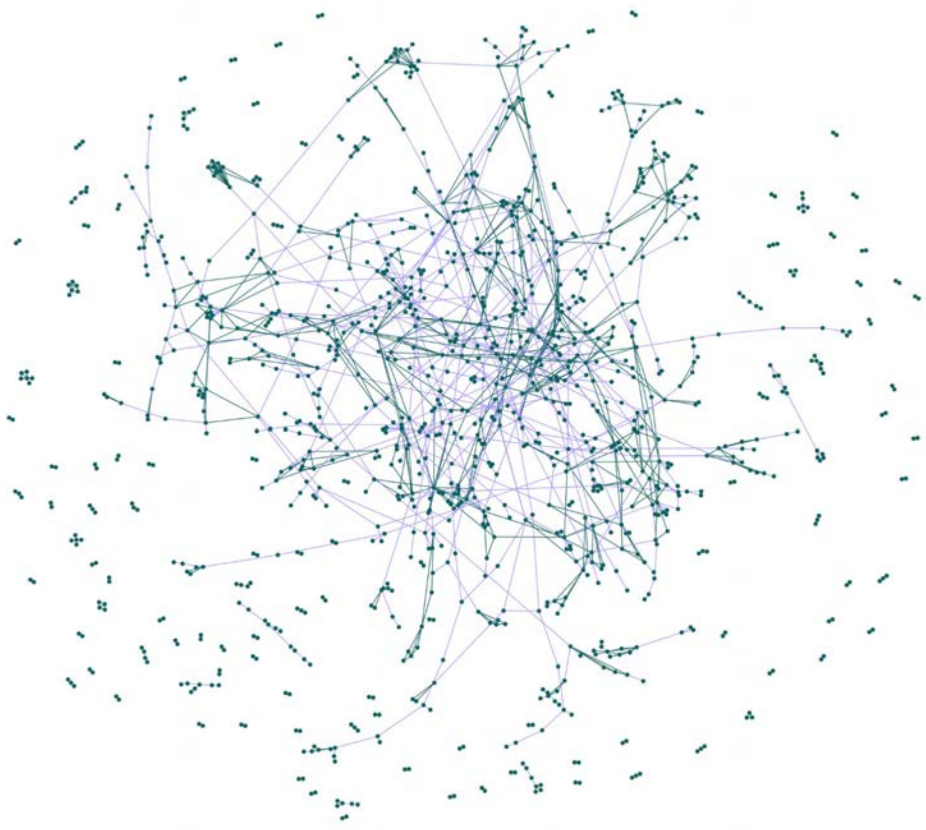
Number of edges: 480

Number of components: 121

Size of largest component: 176

Average degree: 1.77

Clustering coefficient: .09



Centrality measures

Which nodes are central in the network?

It depends on the definition of central….

Two typical measures are **closeness** and **betweenness**

Closeness centrality: How long does it take to reach other nodes from a given node?

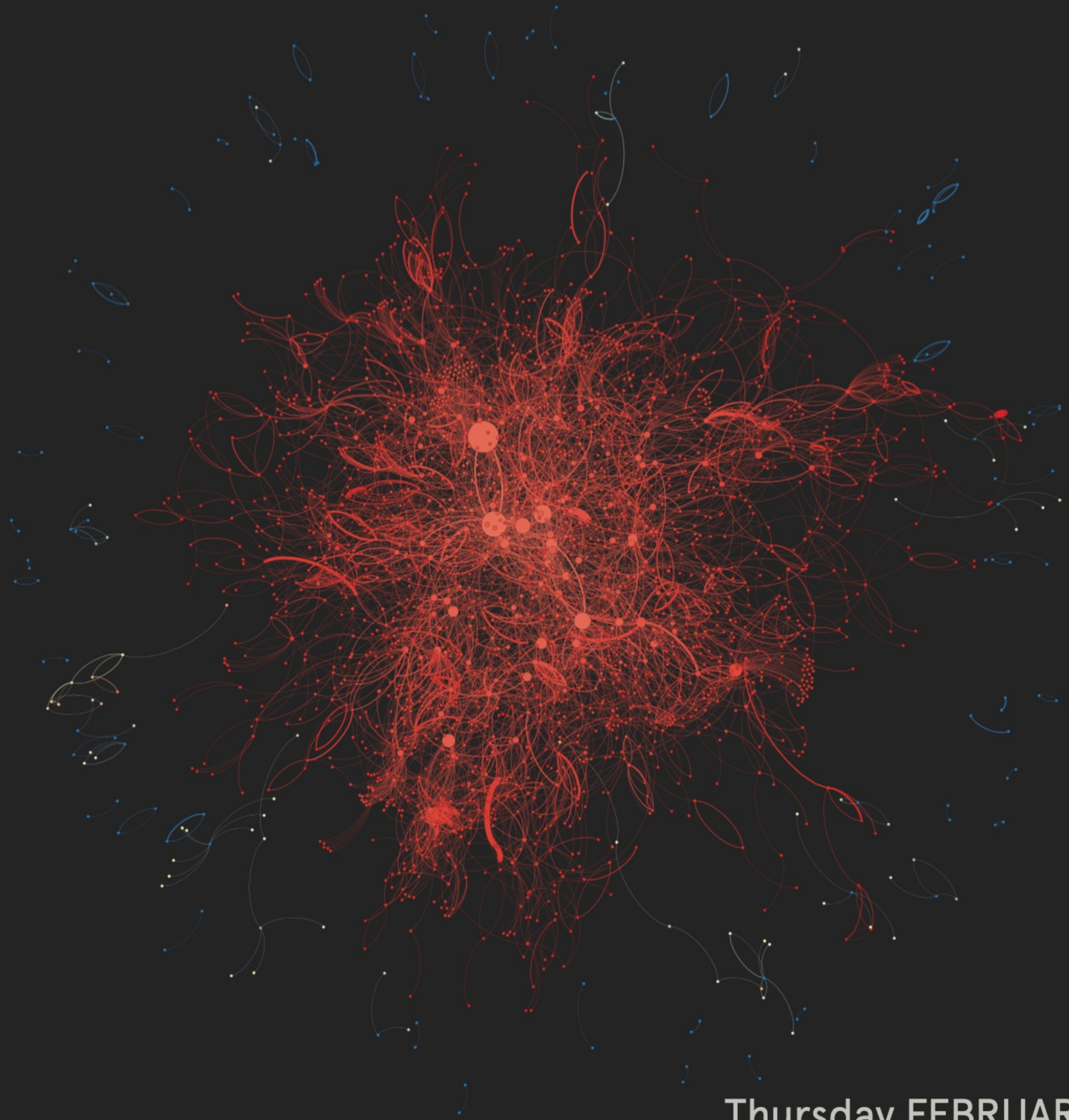
Betweenness centrality: How many shortest paths go through the given node?

MIT email collaboration

—> Nodes color by Closeness centrality

—> Nodes dimension by Betweenness centrality (scale 10-100)

—> Edges thickness by n. Emails exchanged



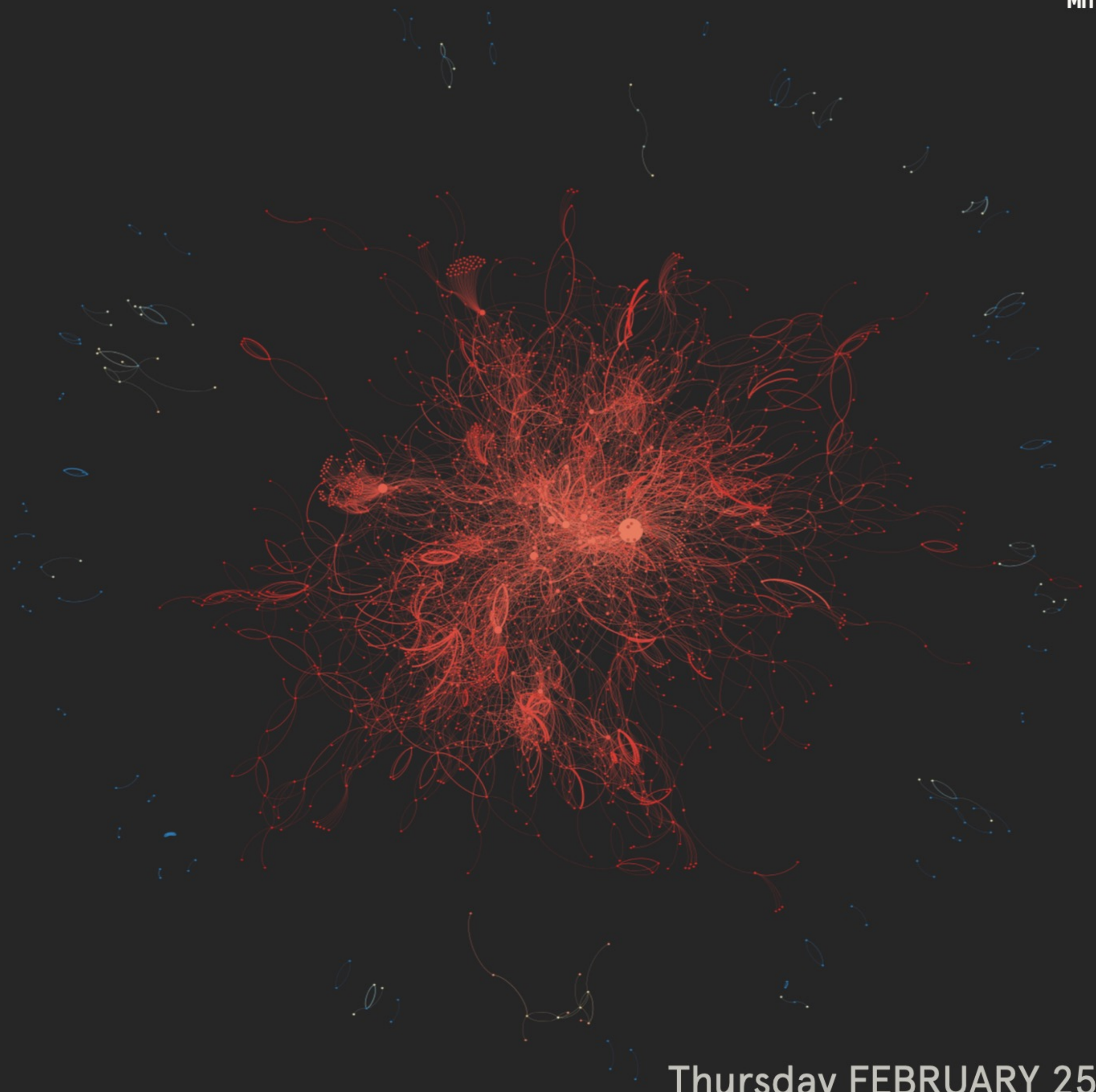
Thursday FEBRUARY 20th, 2020

MIT email collaboration

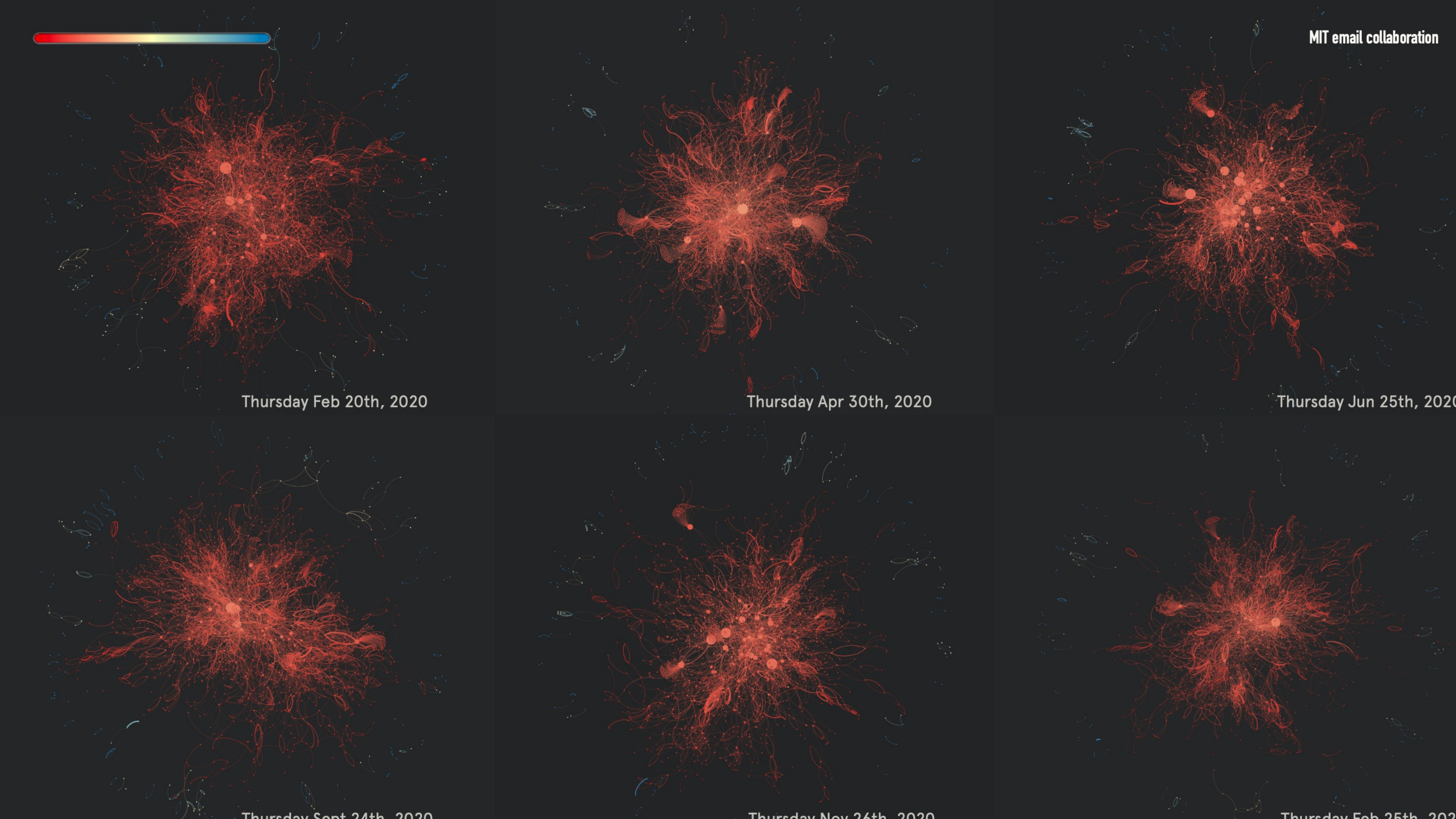
—> Nodes color by Closeness centrality

—> Nodes dimension by Betweenness centrality (scale 10-100)

—> Edges thickness by n. Emails exchanged



Thursday FEBRUARY 25th, 2021


MIT email collaboration
Thursday Feb 20th, 2020
Thursday Apr 30th, 2020
Thursday Jun 25th, 202