

***PROTECT PURDUE: DATA DRIVEN  
UNIVERSITY OPERATIONS IN A  
PANDEMIC***

OCTOBER 7, 2021

# *PROTECT PURDUE COVID DATA EFFORTS*

## **IDA+A – Institutional Data Analytics and Assessment**

- Department became the “Hub” of all COVID Reporting and Analytics
- Protect Purdue Proximity System (Digital Contact Tracing)
  - In March 2020 we started building a Purdue-specific contact tracing system
- Surveillance Testing
  - In late August, IDA+A began managing sampling and follow-up for campus surveillance testing efforts

# IDA+A

**CONTACT TRACING**

# *Protect Purdue Proximity System*

## Overview

- Built in-house from existing data in Greenplum
  - All contact tracing compute done in-database (postgresql)
- Required pulling many disparate data sources together
  - Housing contracts
  - Wireless network logs
  - Dining card swipes
  - Class enrollments
  - Research Lab SOPs
  - Fraternity, Sorority, Cooperative Life (FSCL) membership and housing

# *Protect Purdue Proximity System*

## Digital Contact Tracing

- Significant Contact
  - Two or more **moderate** contact activities (e.g., sharing a typical classroom, sharing an access point)
  - One **close** contact activity (e.g., living together, eating together)
- System stores over 4.9B potential contacts since start of 2020
- Significant contacts are identified in six categories: Living, Dining, Proximity, Classroom, Research Labs, and FSCL

# Protect Purdue Proximity System

## Contact Definitions

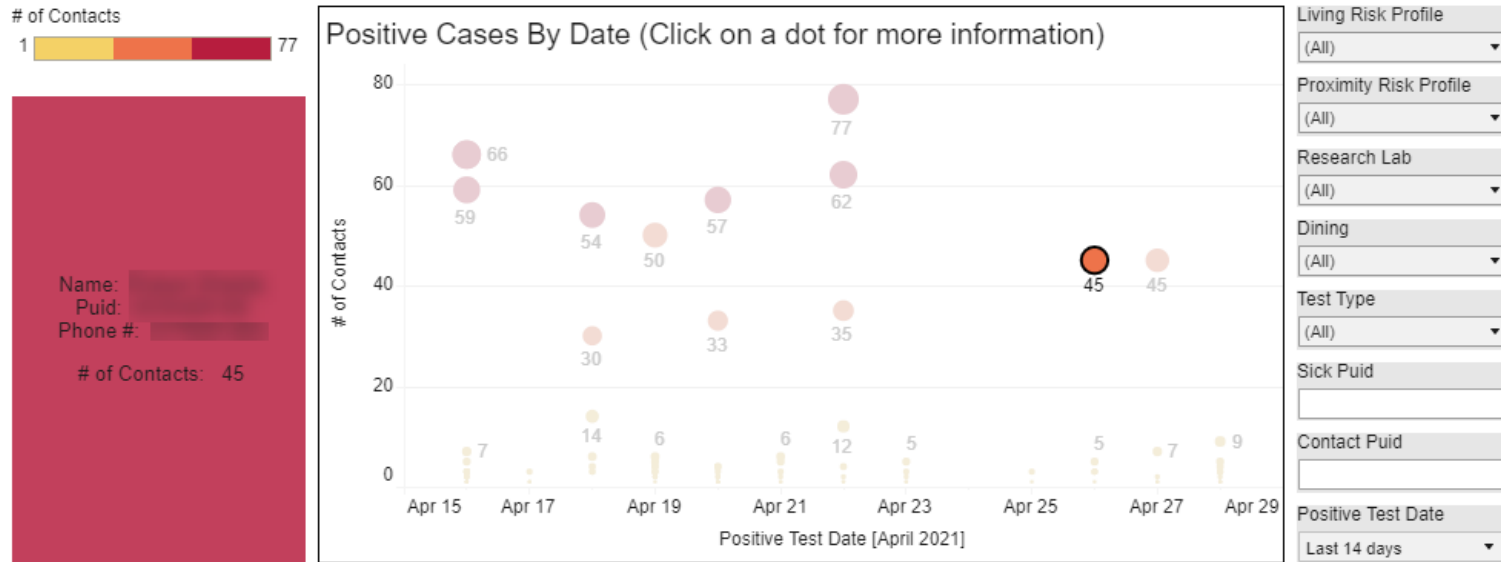
- Living Contact
  - On-campus housing and off-campus housemates
  - Roommate/housemate is a **close** contact
  - Section mate is **moderate** contact
- Dining Contact
  - 2 or more swipes within 10 seconds of each other in the past week is **close** contact
- Proximity Contact
  - 60+ minutes shared on a Wi-Fi access point is **close**, less is **moderate**
  - Only sessions overlapping by at least 15 minutes are considered at all

# *Protect Purdue Proximity System*

## Contact Definitions

- Class Contact
  - Based on the enrollment number, size of room, time in the room, type of course (e.g., Lab course has higher level of contact)
  - Accounts for staggered attendance plans, multi-instructor courses
  - Originally took many shared classes for **close** contact, now not used
- Lab Contact
  - Everyone assigned to the same room in an SOP is a **moderate** contact
- Fraternity, Sorority & Cooperative Life (FSCL) Contact
  - Living in the same FSCL house is **close** contact
  - Belonging to the same FSCL organization is **moderate** contact

# Protect Purdue Proximity System



## Contact Tracing Dashboard

- Accessible to medical team, contact tracing team

Contacts

Risk Profile	Contact Name	Contact Puid	Contact Phone	Contact Email
Lives in greek house with;				
In greek organization with; Lives in greek house with;				

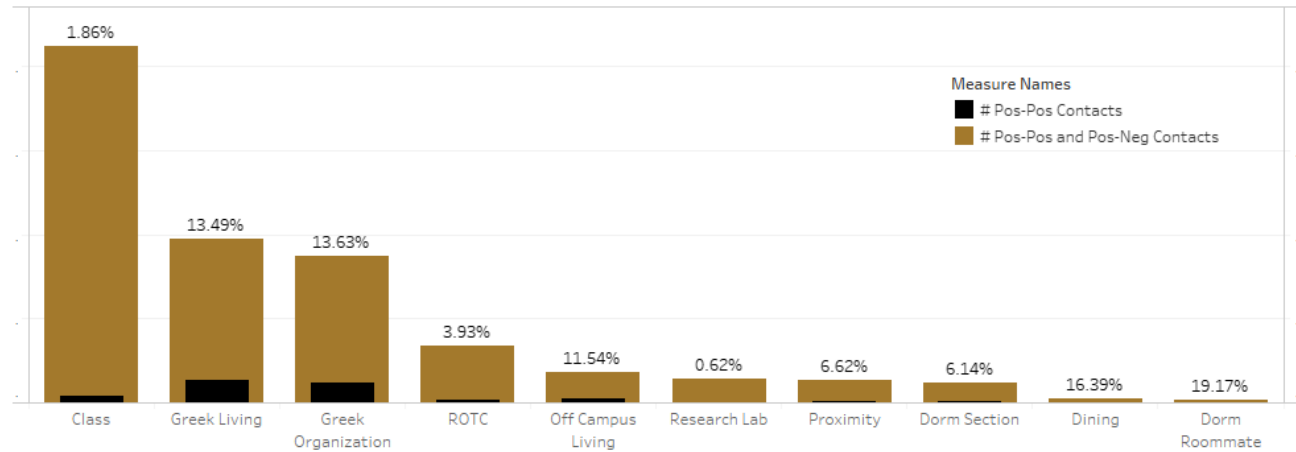


# PROTECT PURDUE PROXIMITY SYSTEM

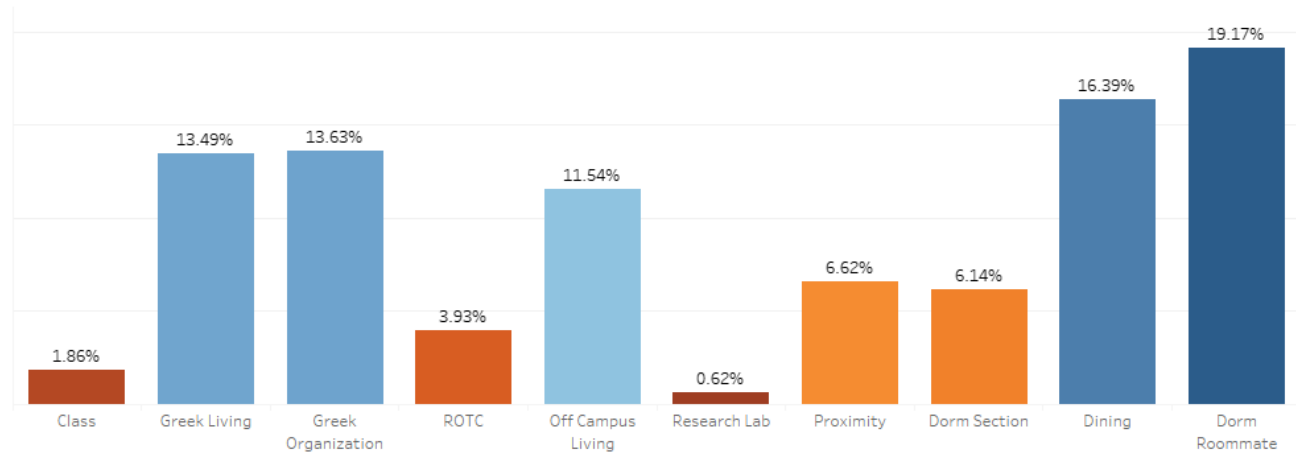
## Post-hoc Analysis Fall 2020

- 1<sup>st</sup> August, 2020 to 18<sup>th</sup> December, 2020
- Measured all connections within 2 weeks of a positive test
- Labeled 'Positive-Positive' if both sides of the connection tested positive
- Can combine: Dining + Proximity + Dorm Roommate is 27%

Basic Risk Results



Basic Risk Positivity Rate



# IDA+A

**COVID SURVEILLANCE TESTING**

# COVID Surveillance Testing Methods

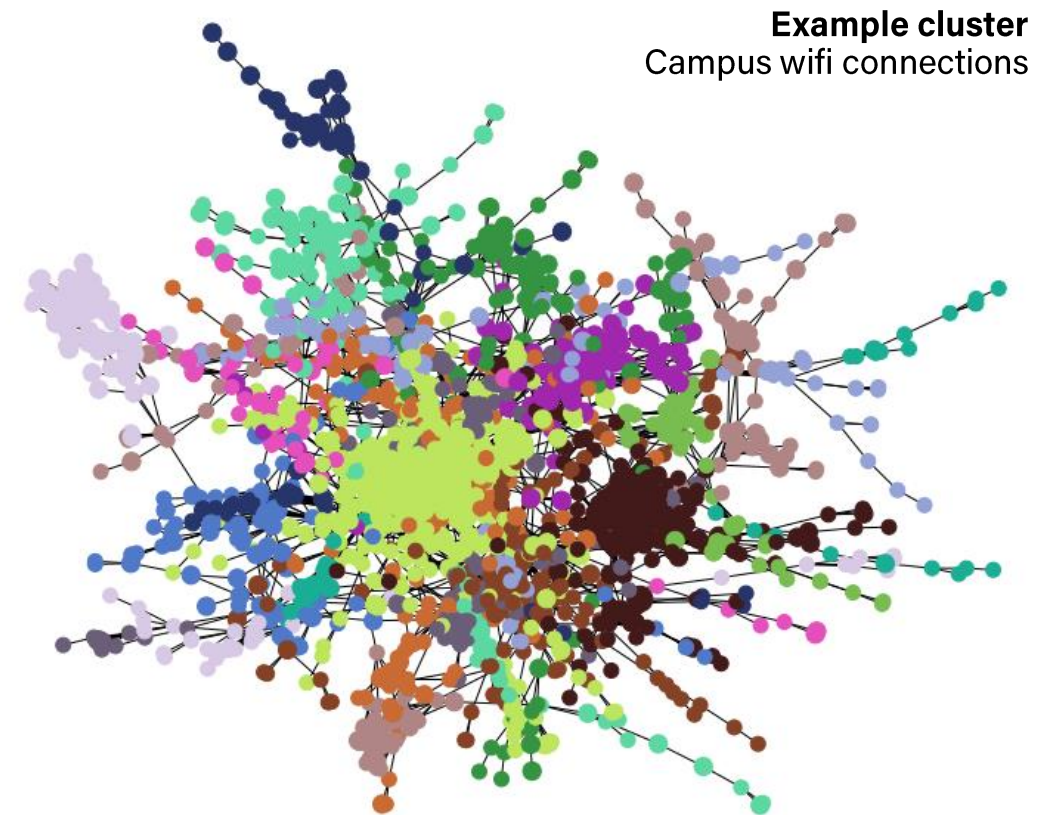
## Necessity Leads to Invention

- Limited testing capacity of ~5,600 per week could be done on-site
  - On Campus Population ~49,700
- How could we monitor this population effectively?
  - Smart sampling strategies to maximize utility of available tests
  - Rebalance sample sizes based on actual testing data
- Employees – based on type of job and working location
  - Critical – weekly testing
  - Other On Campus – random sampling
- Students – based upon where the student lived
  - Residents Halls and FSCL – stratified random sampling
  - Off Campus – cluster weighted sampling

# Off Campus Random Sampling

## Cluster Weighted Sampling

- For off campus students, sampling is hard
  - Some students will have many touchpoints to campus, some hardly any
- The contact tracing system 'clustered' students based on the various tracing criteria
- More connected students more likely to be sampled
- Concept: Catch the COVID cases most likely to threaten Purdue students and staff



# ***APPENDIX***

**Additional Information as Needed**

# PROTECT PURDUE COVID DATA EFFORTS

## IDA+A & Collaborators

- IDA+A COVID Team
  - 3 Data Analysts
  - 2 Data Engineers
  - 2 Data Scientists
- Partnerships
  - Protect Purdue Health Center (PPHC)
  - Medical Advisory Team
  - Office of the Registrar
  - Residential Life
  - Marketing and Media
  - ITaP
  - Office of the Dean of Students
  - Testing Vendors
  - Human Resources
  - Many More!

# Data Science Infrastructure

## Mature Computing Environment

- Data lake for university data, in-database compute for modelling & analytics



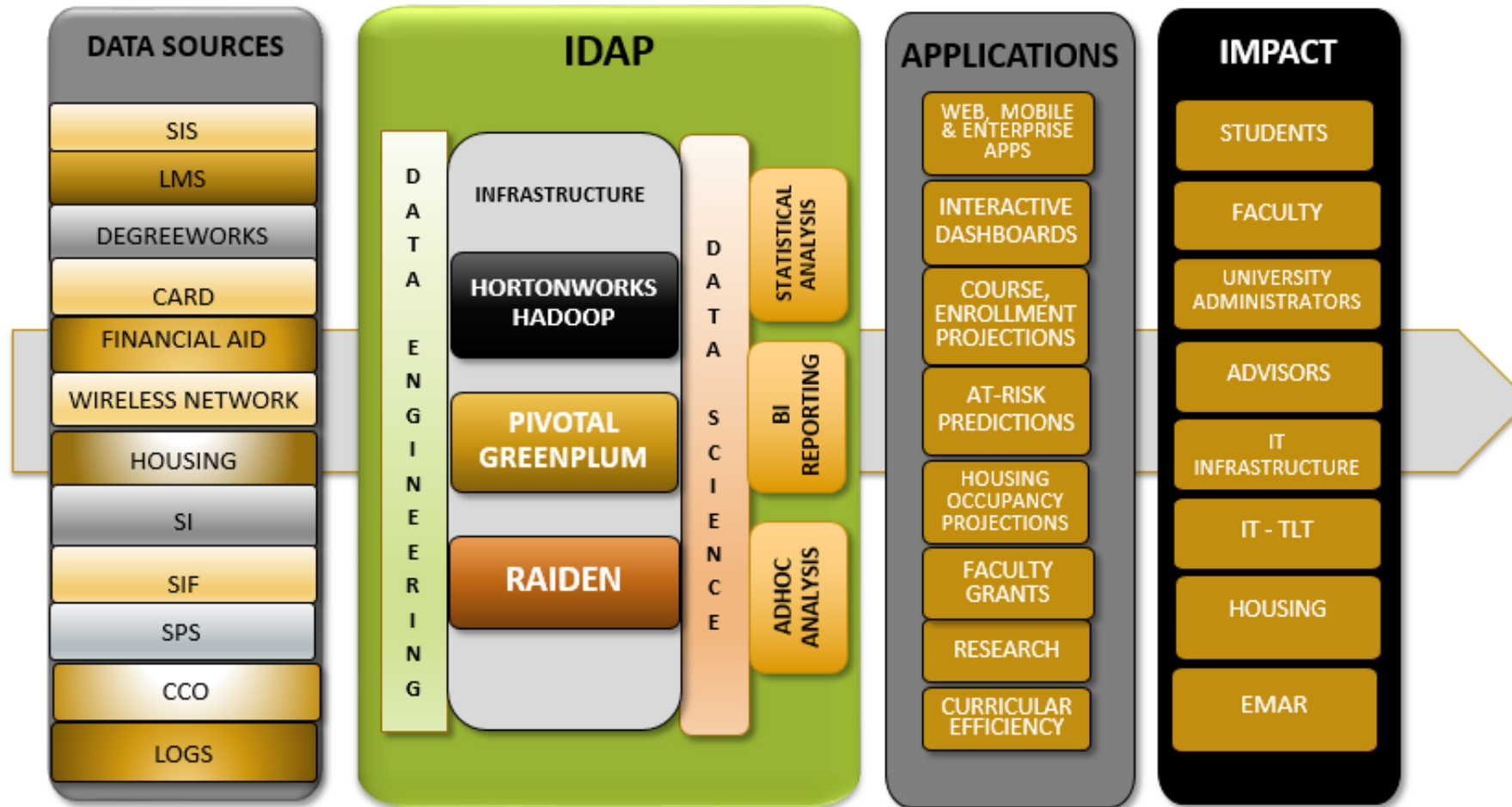
- 
- Mass storage space for unstructured data, has allowed us to utilize data other universities haven't



- 
- Extremely powerful machine for cutting-edge experimentation as well as faculty research



# Institutional Data Analytics Platform





# COVID Test Result Pipeline

