

RISK-BASED MONITORING USING AI/ML DETECTION OF SYSTEMATIC BIAS IN CLINICAL DATA



PHUSE US Connect 2024

Kostiantyn Drach, PhD

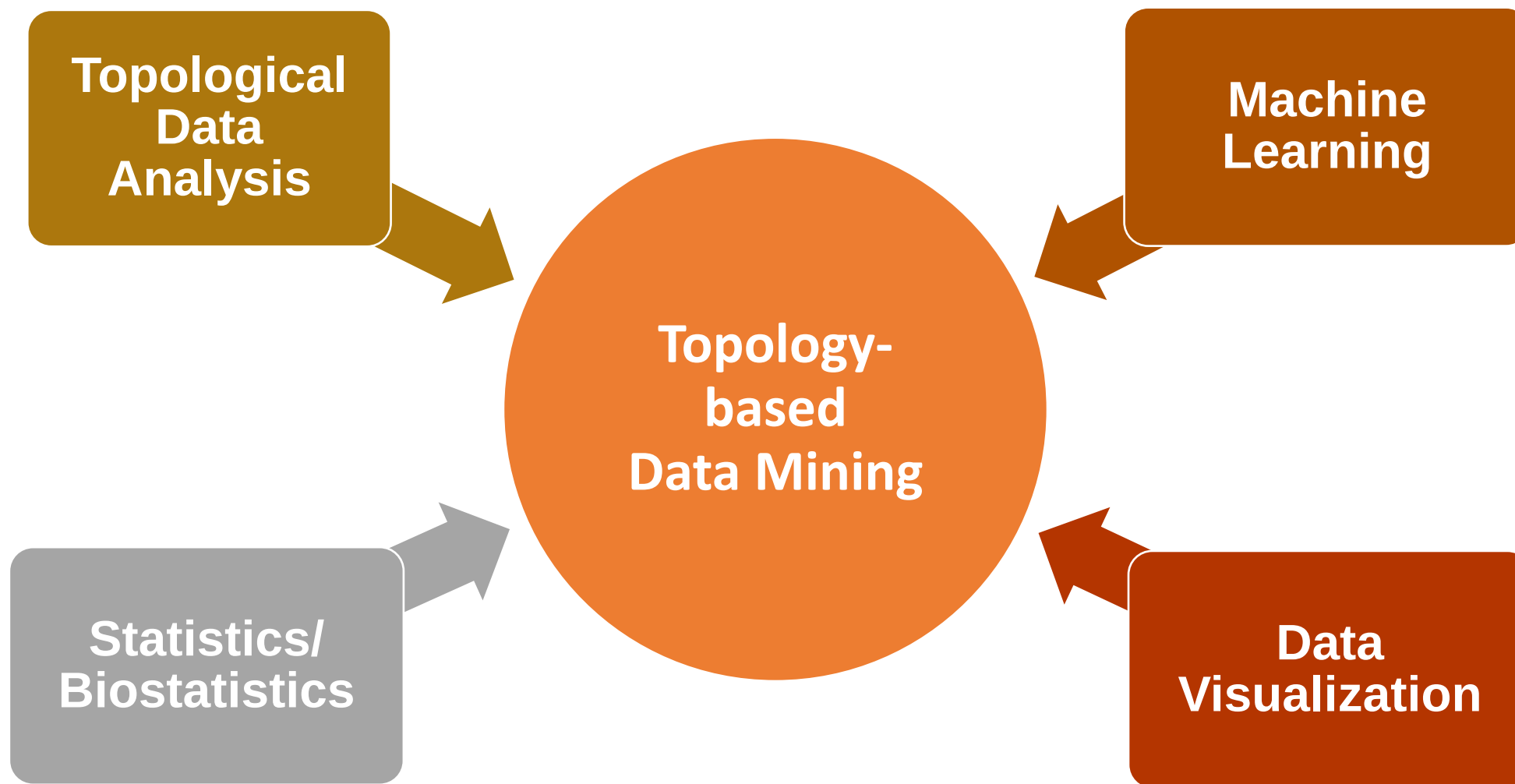
Assistant professor, University of Barcelona, Spain / Intego Group, Kharkiv, Ukraine

Kostiantyn graduated from Kharkiv National University with Master's degree in Mathematics. He obtained his PhD in Mathematics with focus on geometry and topology from the Institute for Low Temperature Physics in Kharkiv, Ukraine. Over the last several years he has participated in various research projects in pure and applied mathematics. Kostiantyn is the core external member of an R&D team of data scientists studying clinical data using methods of applied topology.

Sergey Glushakov

Managing Director, Intego Group, Orlando, FL, USA

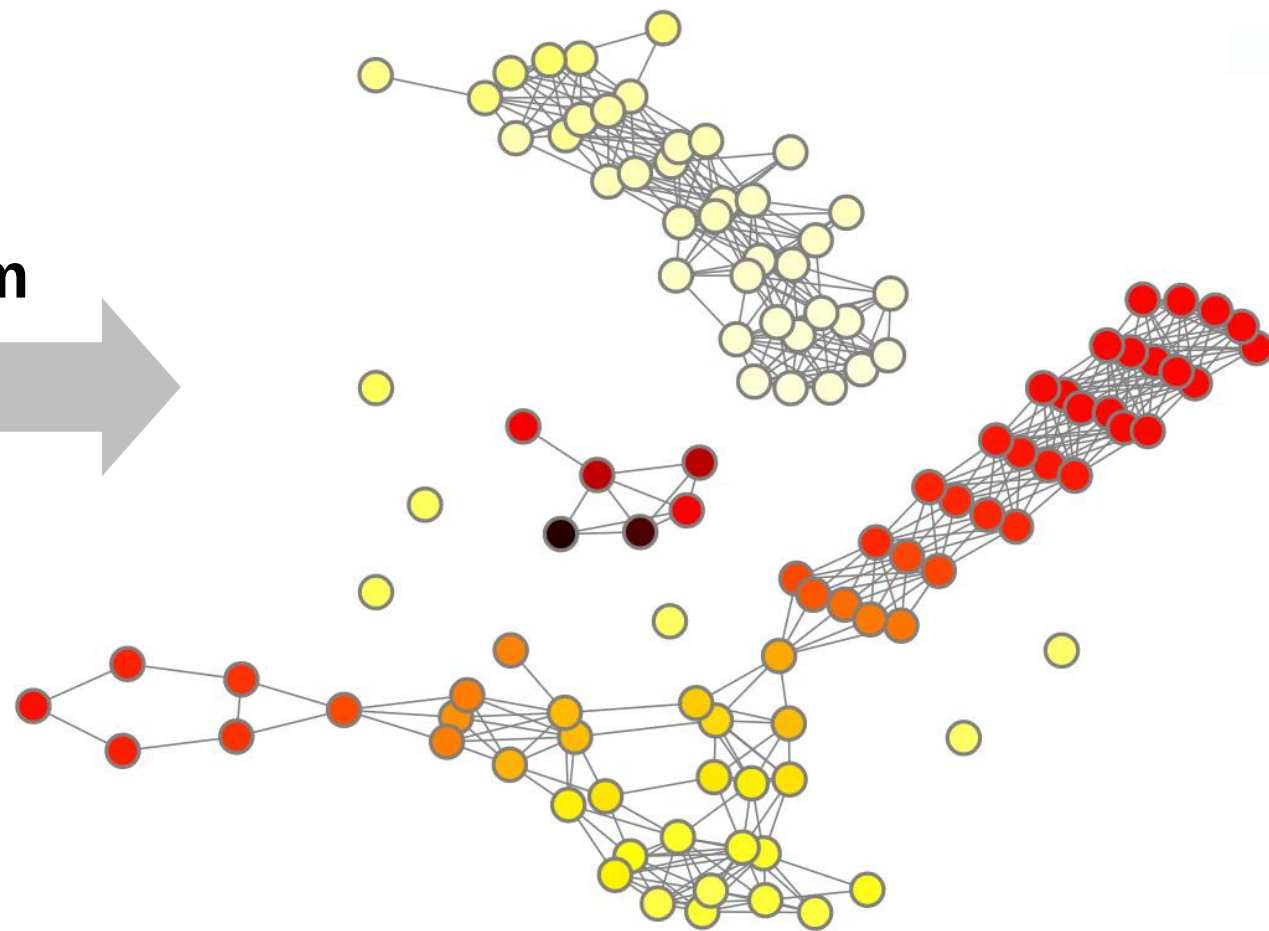
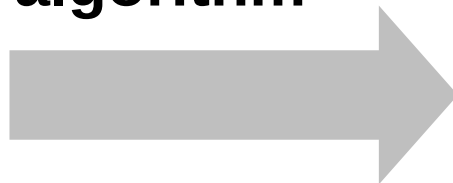
Sergey has an extensive educational background, combined with advanced experience in key areas of business, technologies, and management. He holds an M.Sc. (Applied Math) and a Ph.D. (Engineering) from top Ukrainian universities, and an MBA from Duke University. Sergey has held various positions in the information technology, software, and outsourcing industries. He began his career as a software engineer and researcher before advancing to a C-level executive position and co-founding several companies.





**Unstructured
raw data**

**TDA
algorithm**



**Topological model
(TDA graph)**

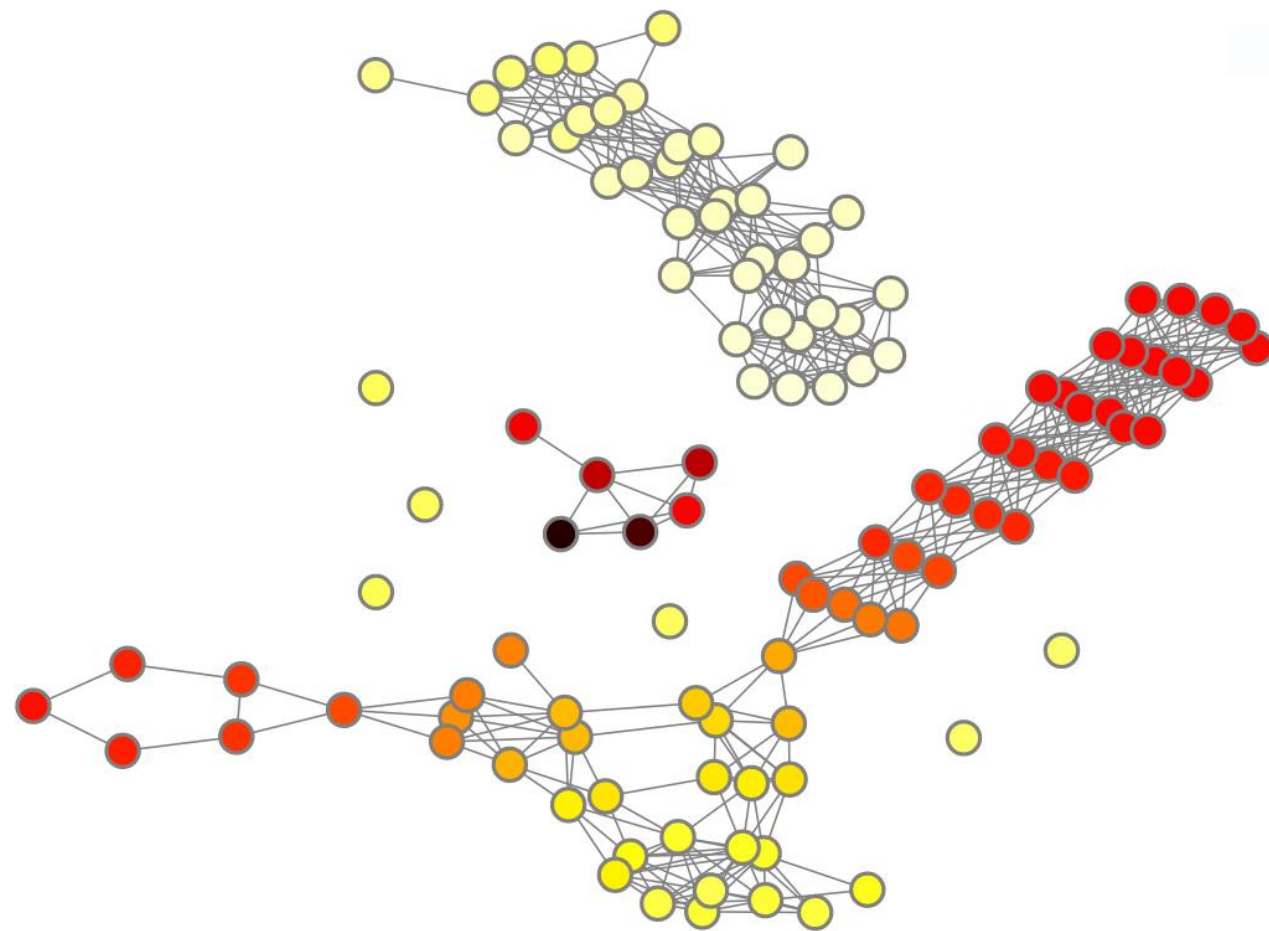
What is a topological model?

Each node represents patient(s)

TDA produces topological models (graphs) where nodes correspond to either individual patients or group of patients within a clinical study

Similar nodes are connected

Two nodes representing similar patients (in terms of predefined outcomes within certain stratifying bins) are connected with an edge



What is a topological model?

Each node represents patient(s)

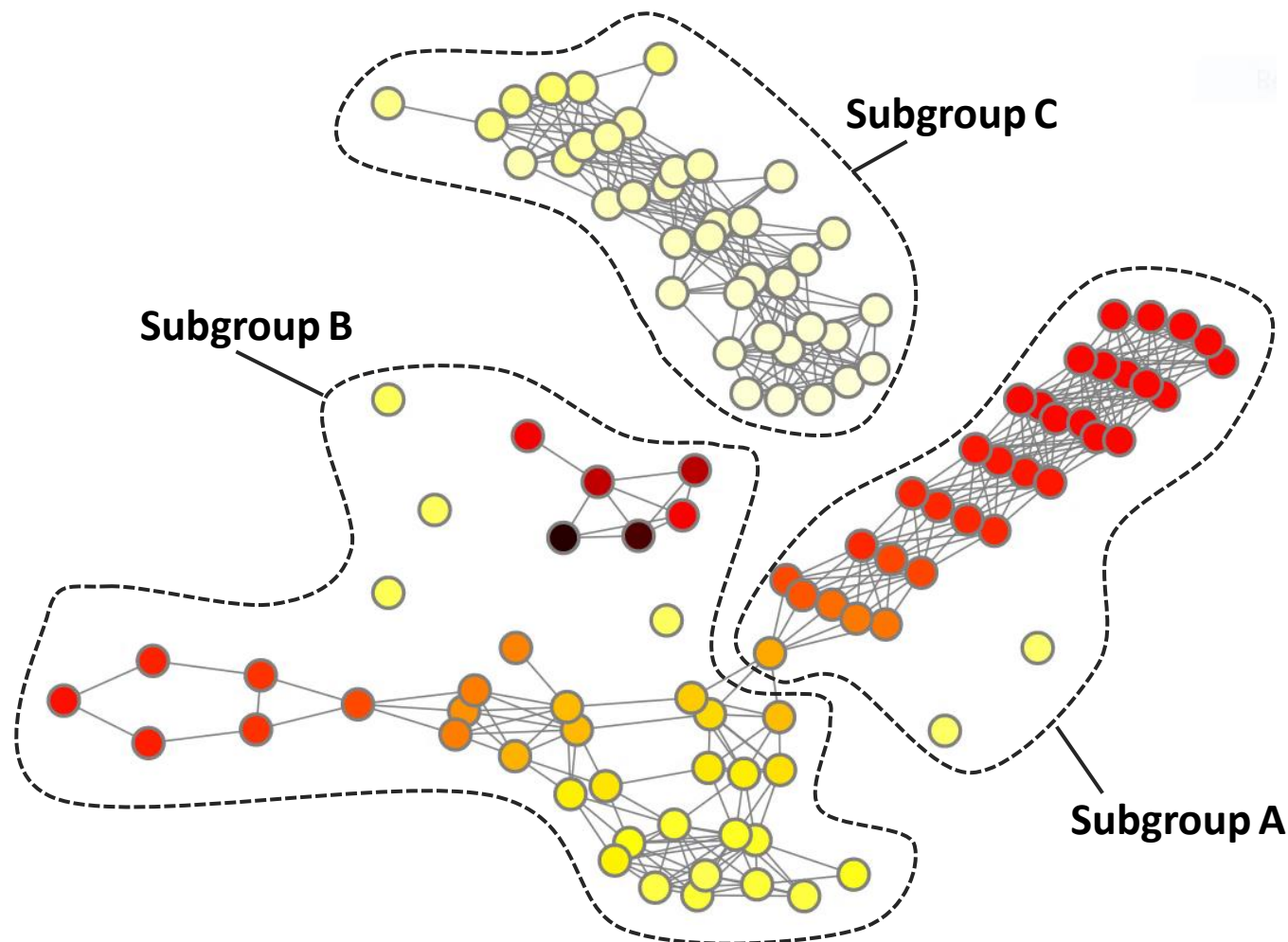
TDA produces topological models (graphs) where nodes correspond to either individual patients or group of patients within a clinical study

Similar nodes are connected

Two nodes representing similar patients (in terms of predefined outcomes within certain stratifying bins) are connected with an edge

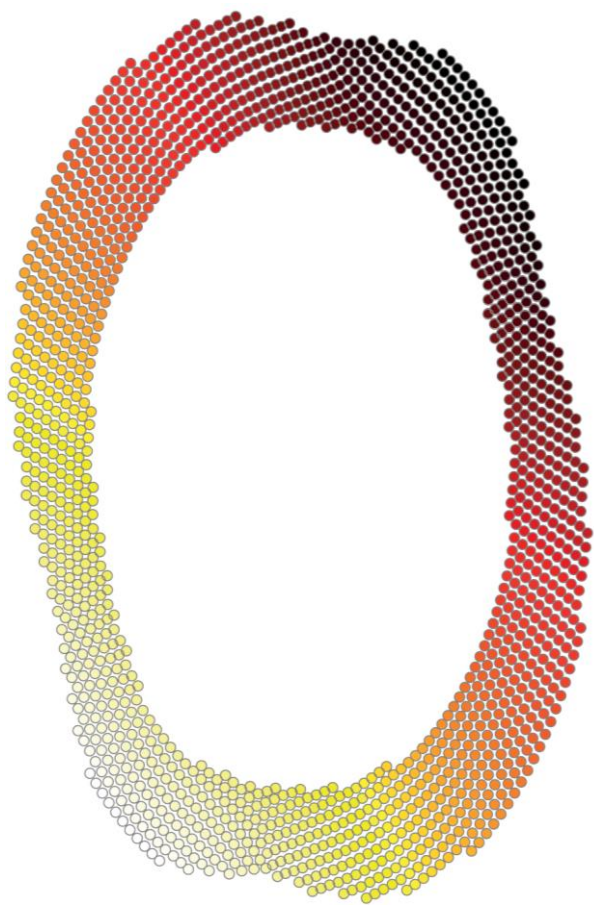
Discovery of subgroups

Communities of nodes in topological model reflect segmentation of patients which may indicate robust patterns within the data



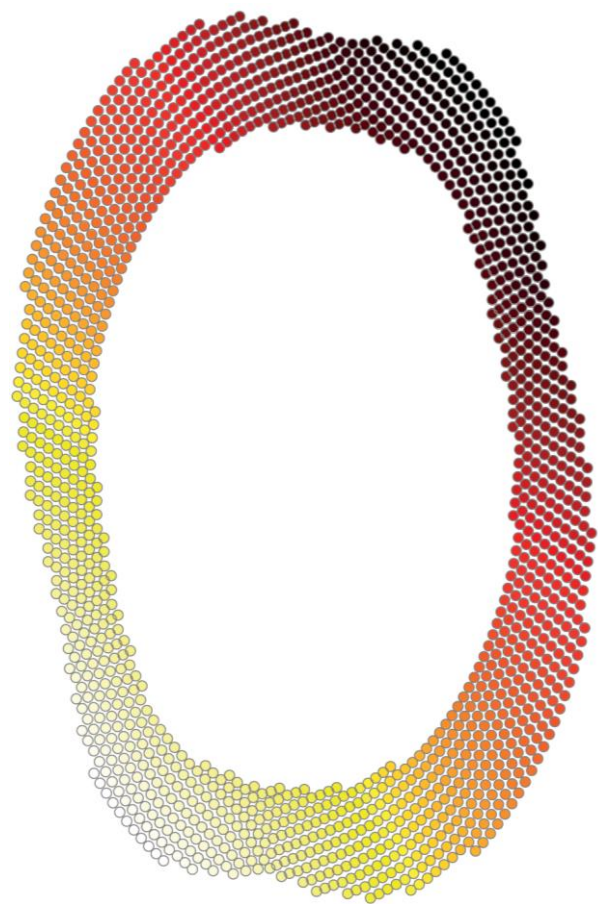
TDA zooms in on robust patterns in data that are not affected by random noise!

Robust geometric properties of dataset



*No missing data
(starting dataset)*

Robust geometric properties of dataset

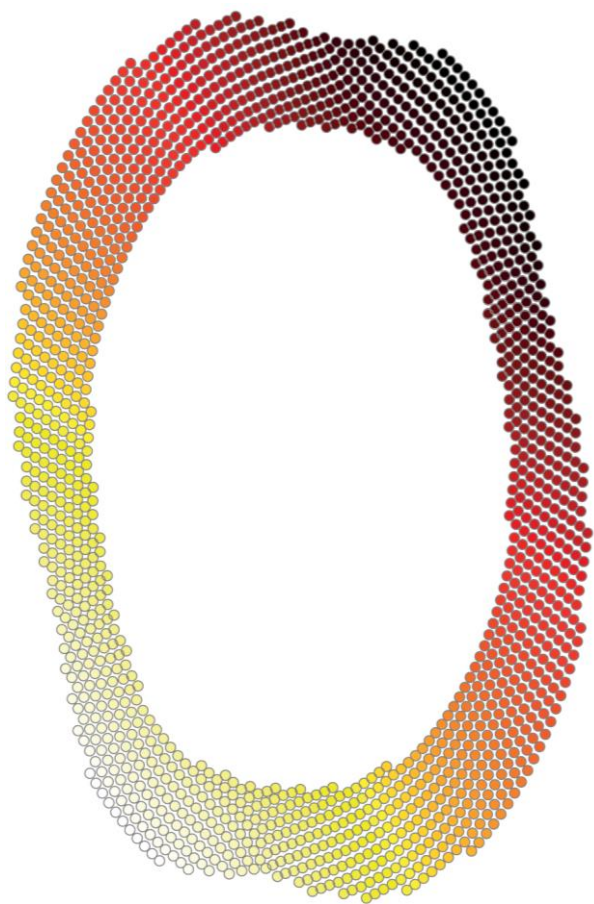


*No missing data
(starting dataset)*



50% missingness

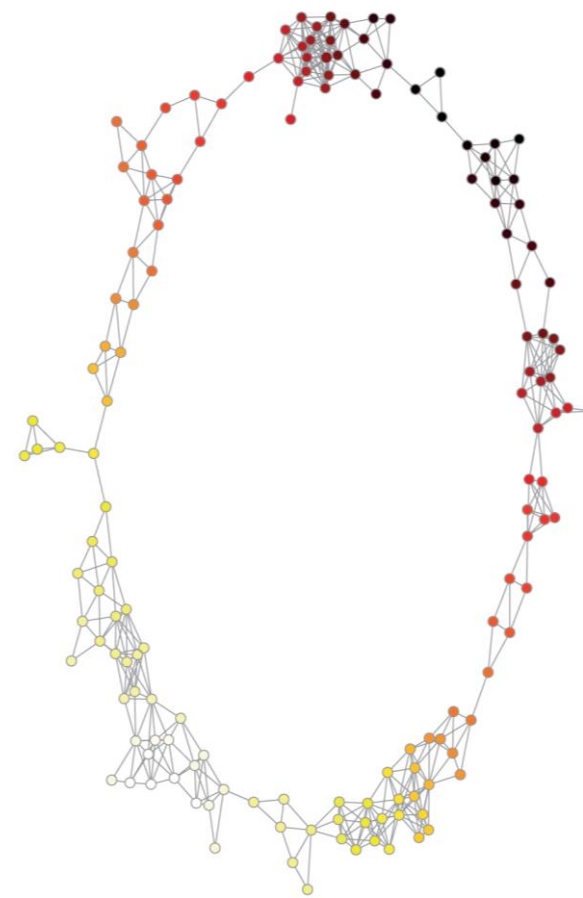
Robust geometric properties of dataset



*No missing data
(starting dataset)*



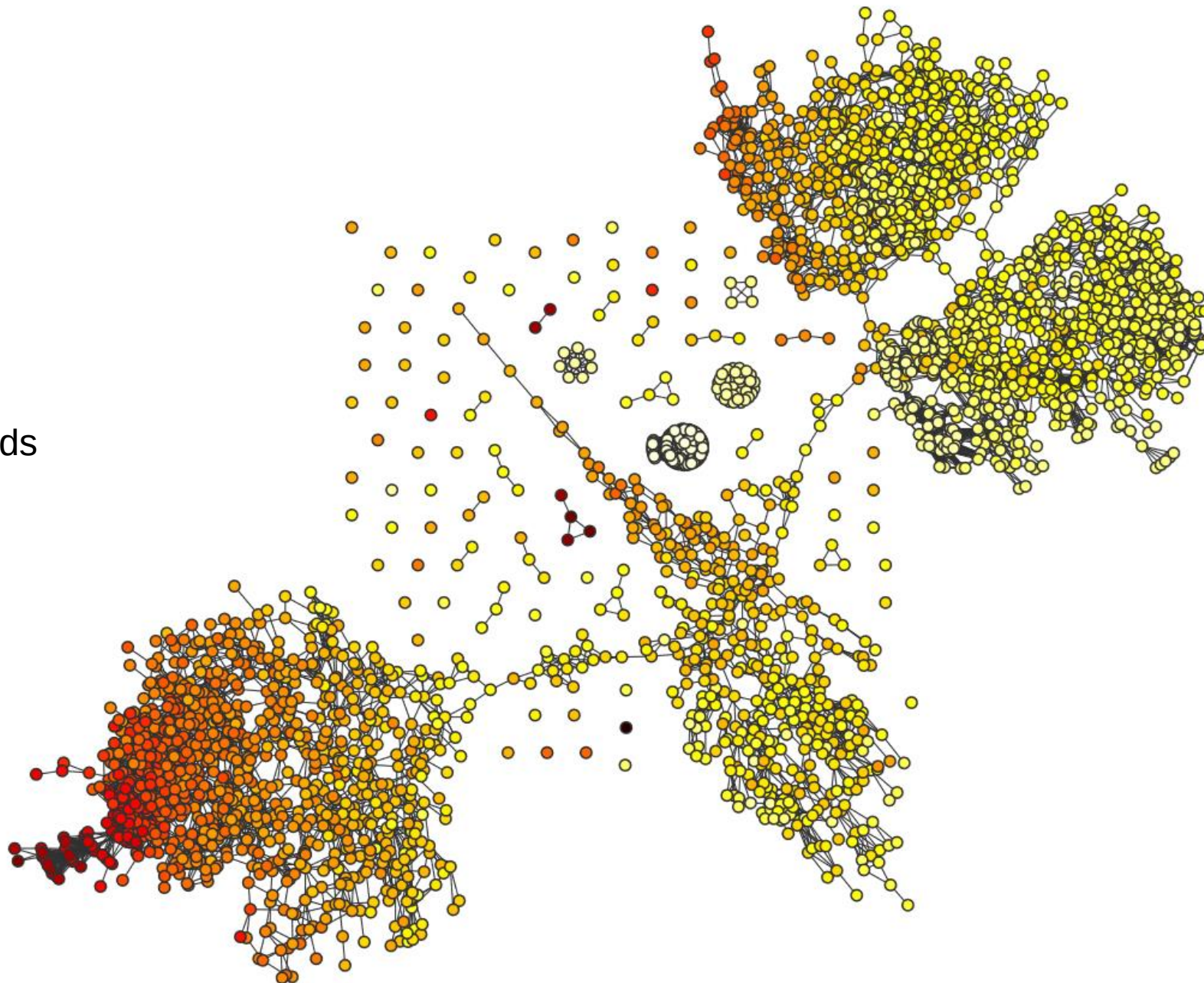
50% missingness



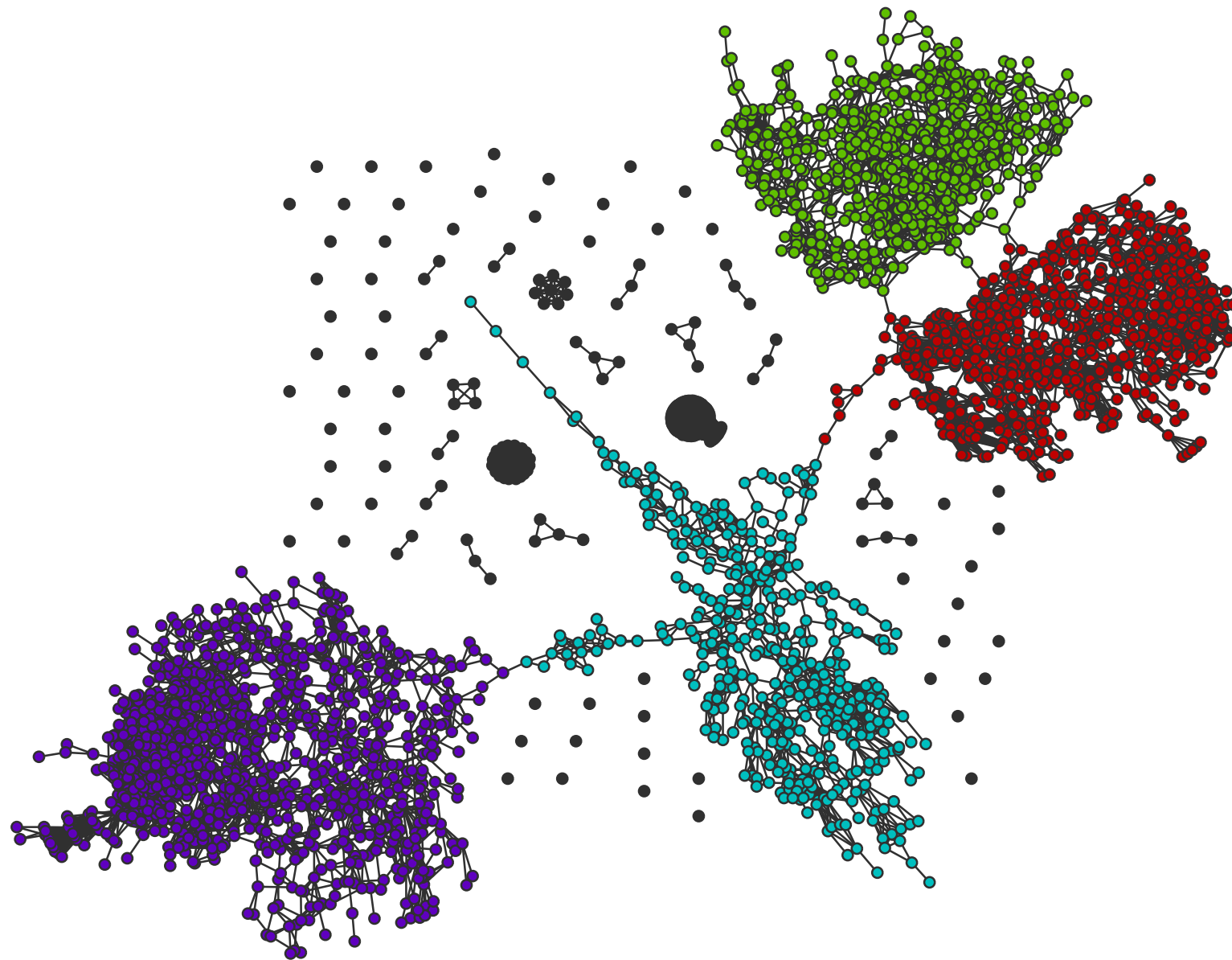
90% missingness

Topological model of COVID-19 spread over the US

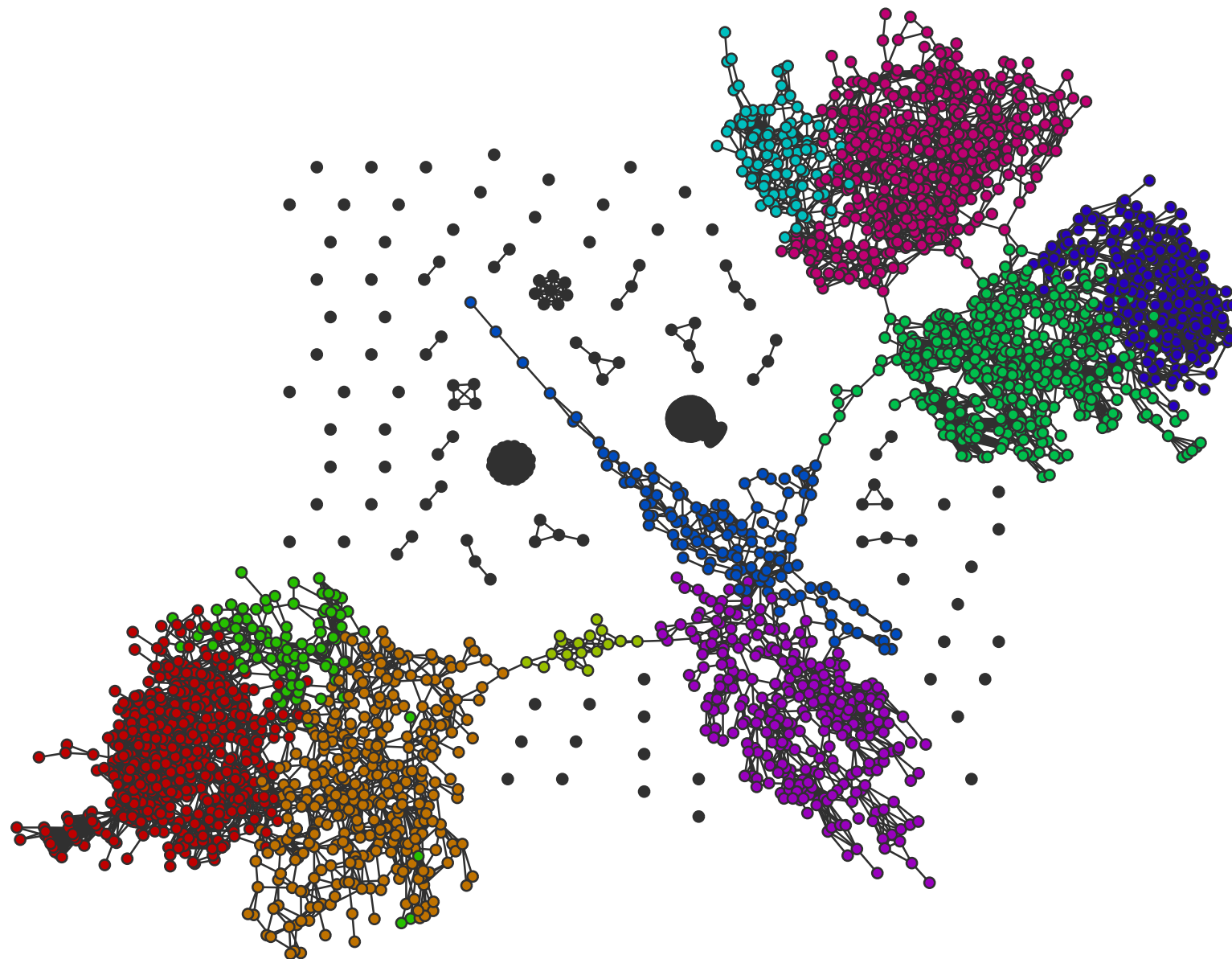
Each node on the graph corresponds
to one county of the United States
(total 3,142 nodes)



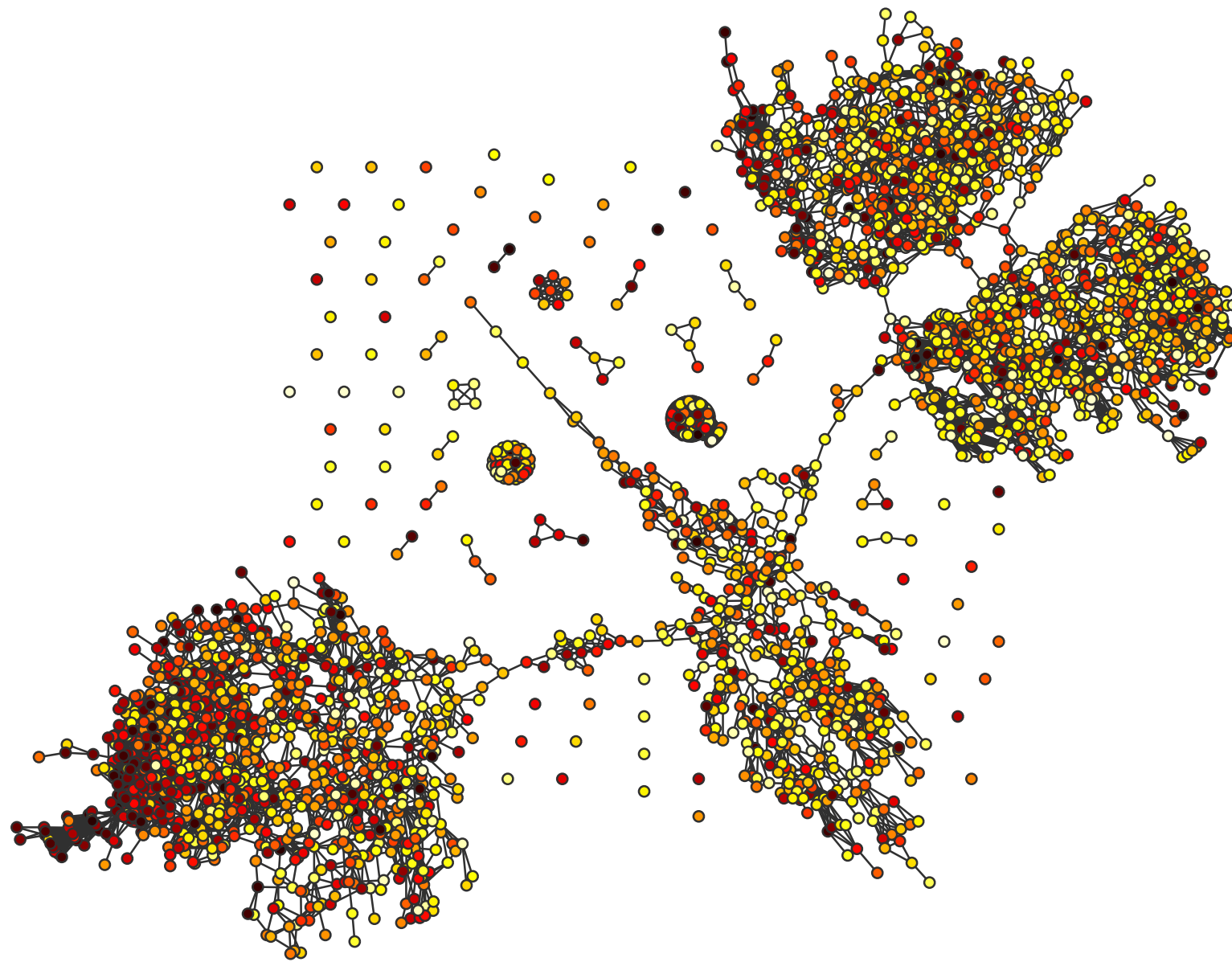
Topological model of COVID-19 spread over the US



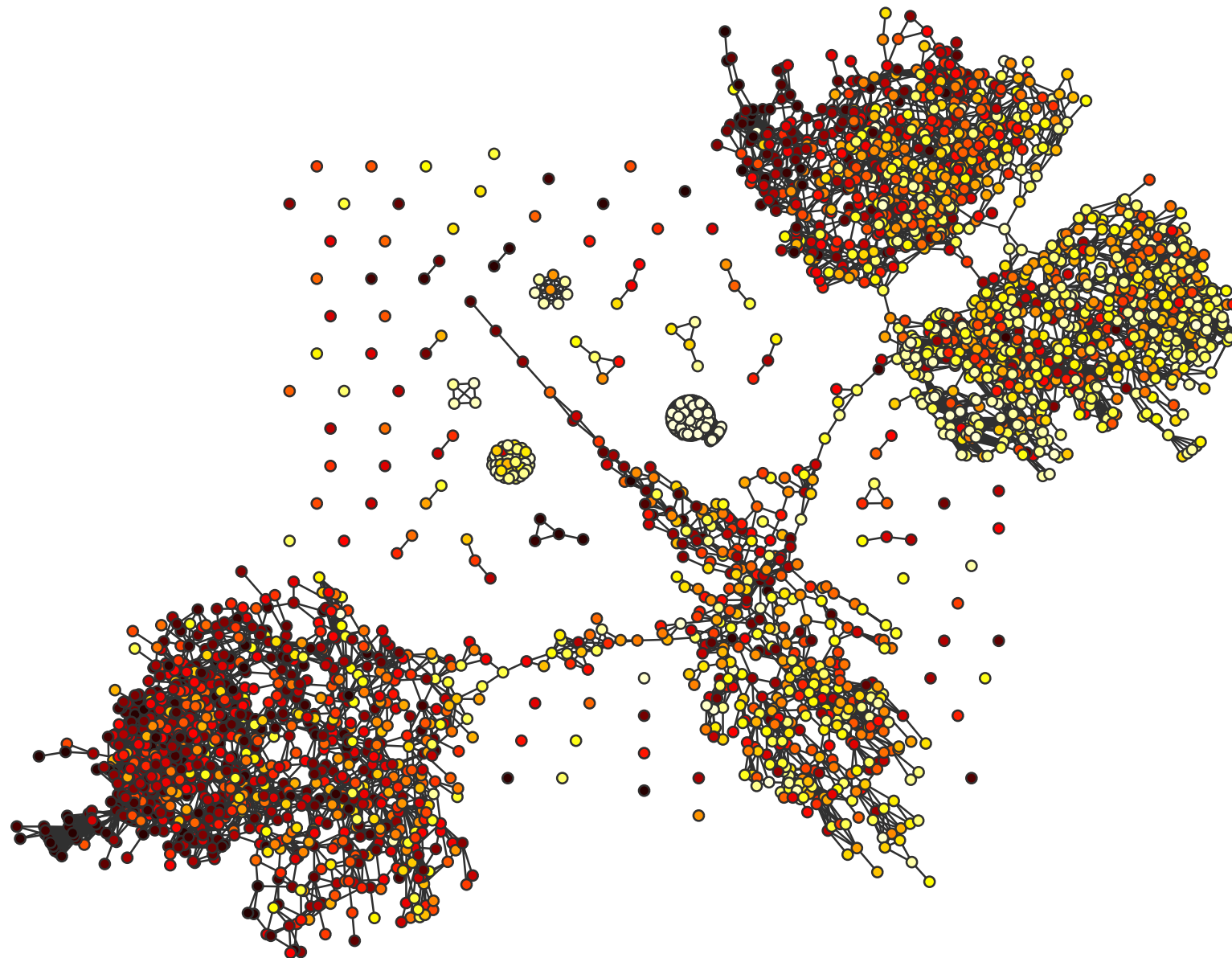
Topological model of COVID-19 spread over the US



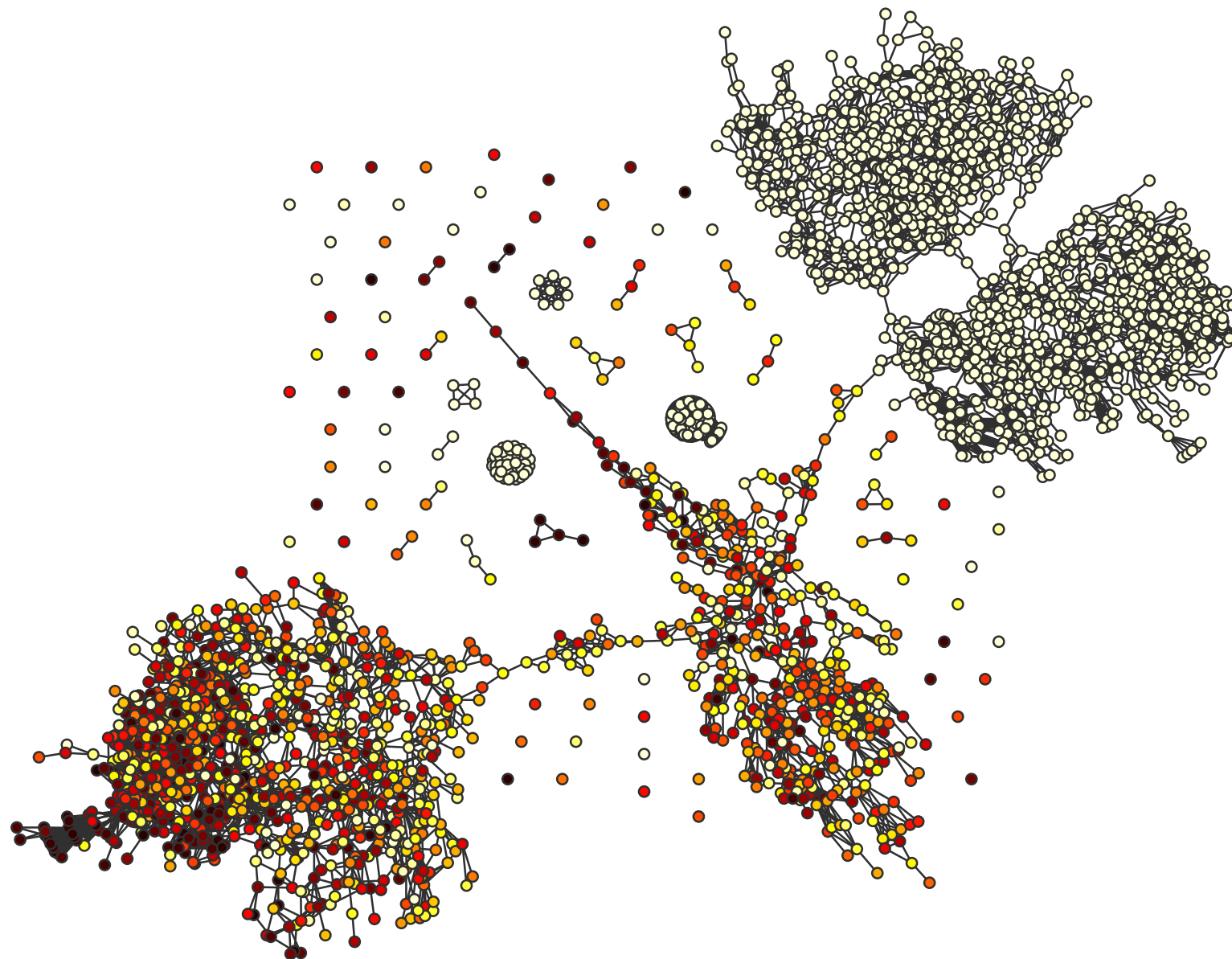
Topological model of COVID-19 spread over the US



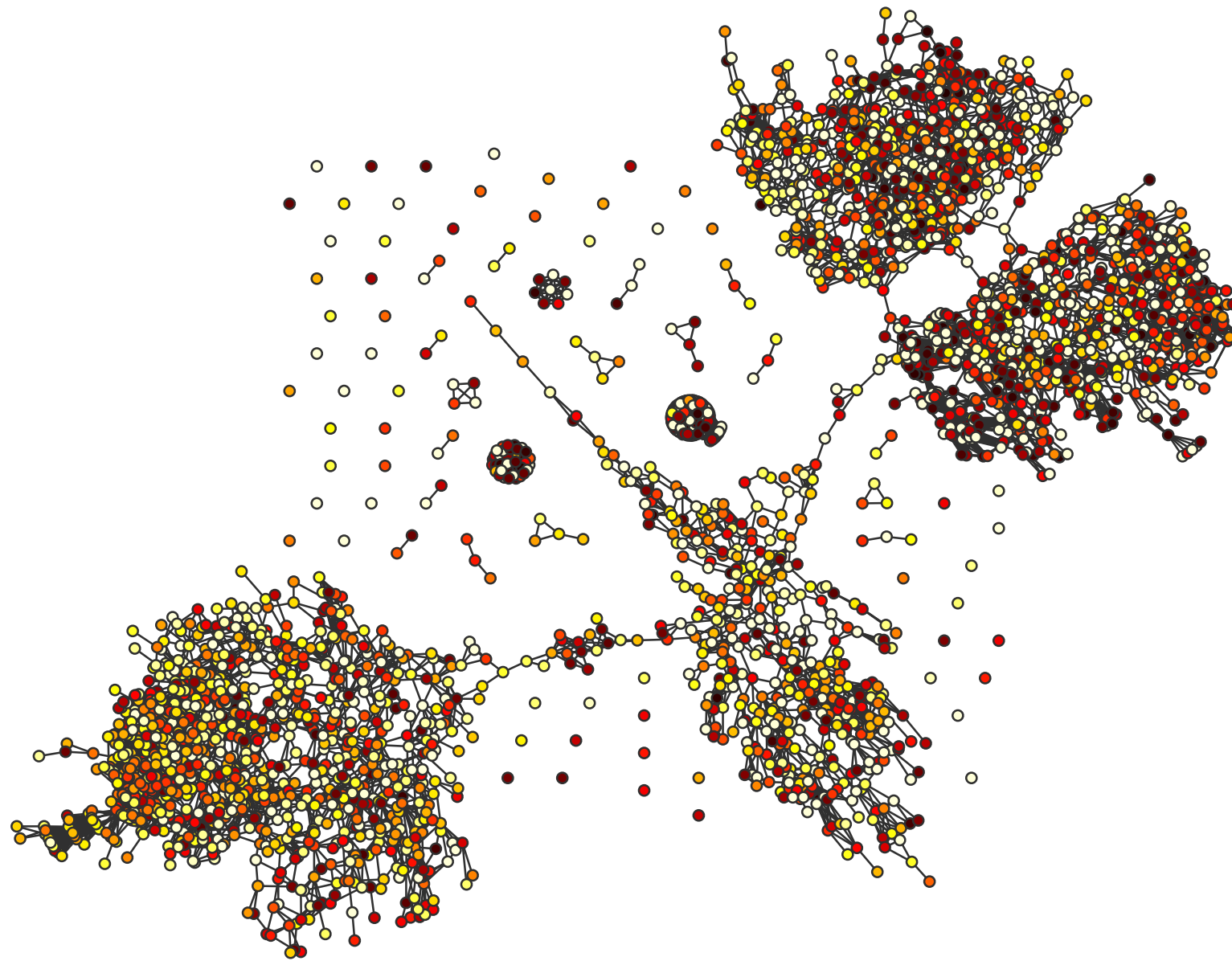
Topological model of COVID-19 spread over the US



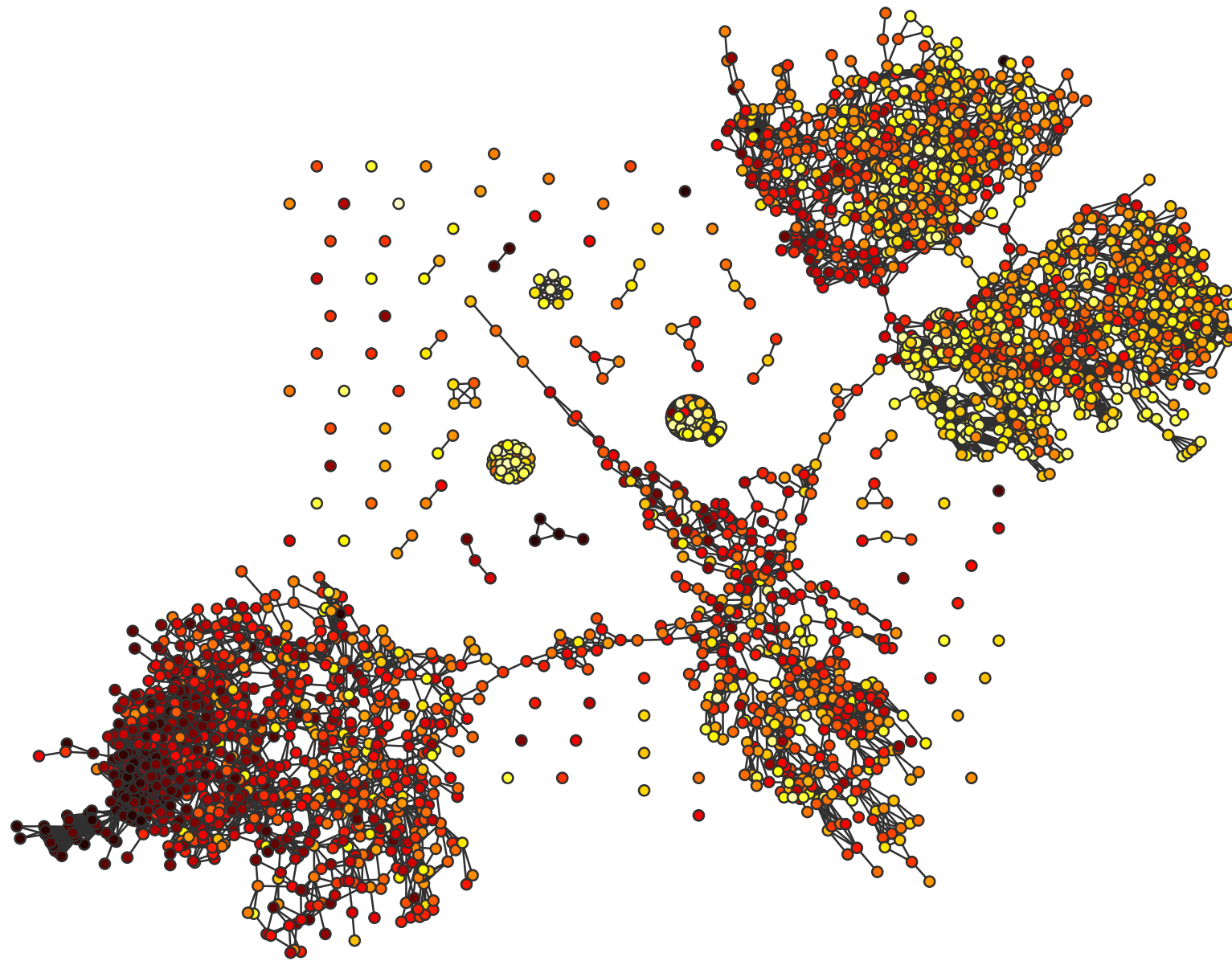
Topological model of COVID-19 spread over the US



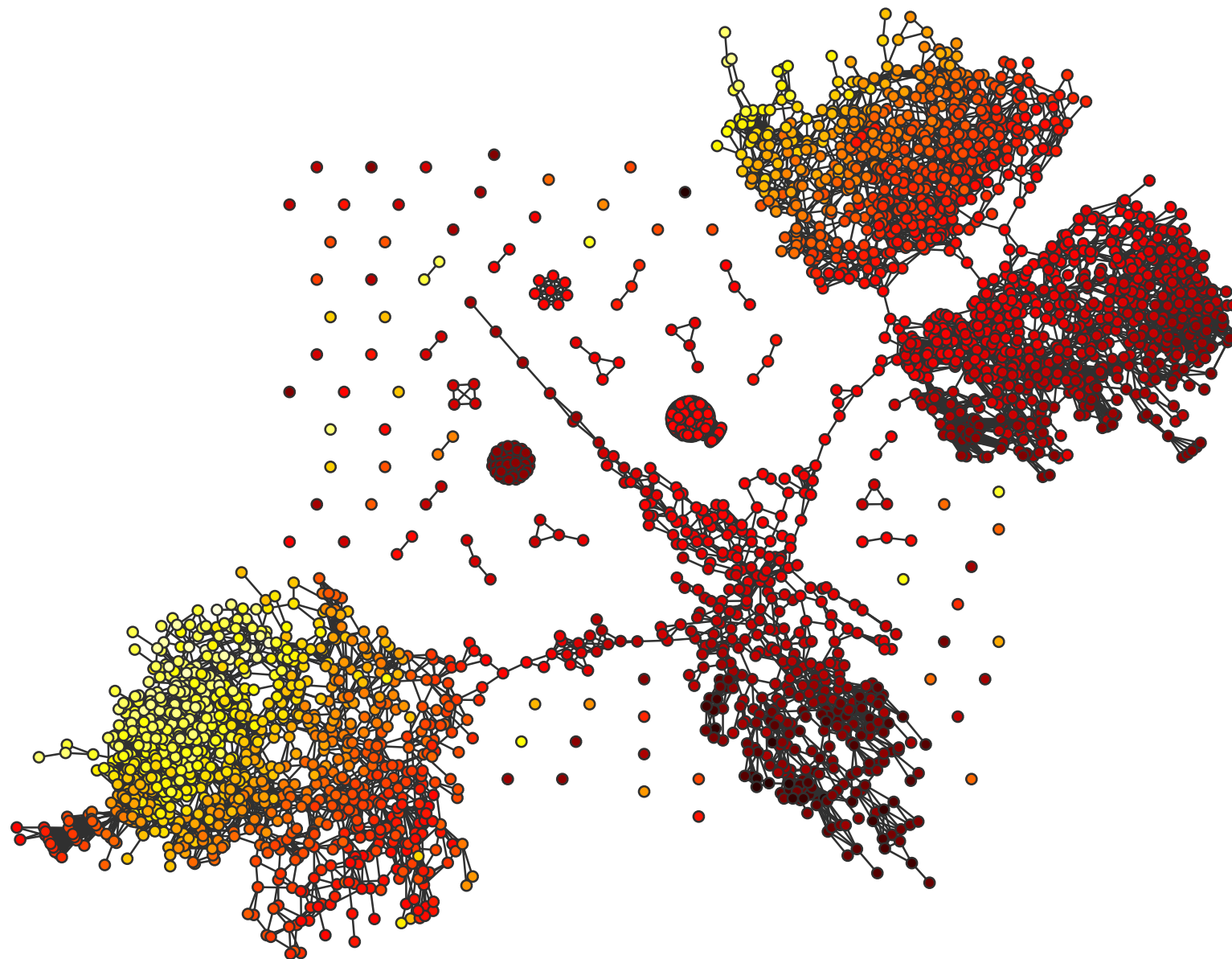
Topological model of COVID-19 spread over the US



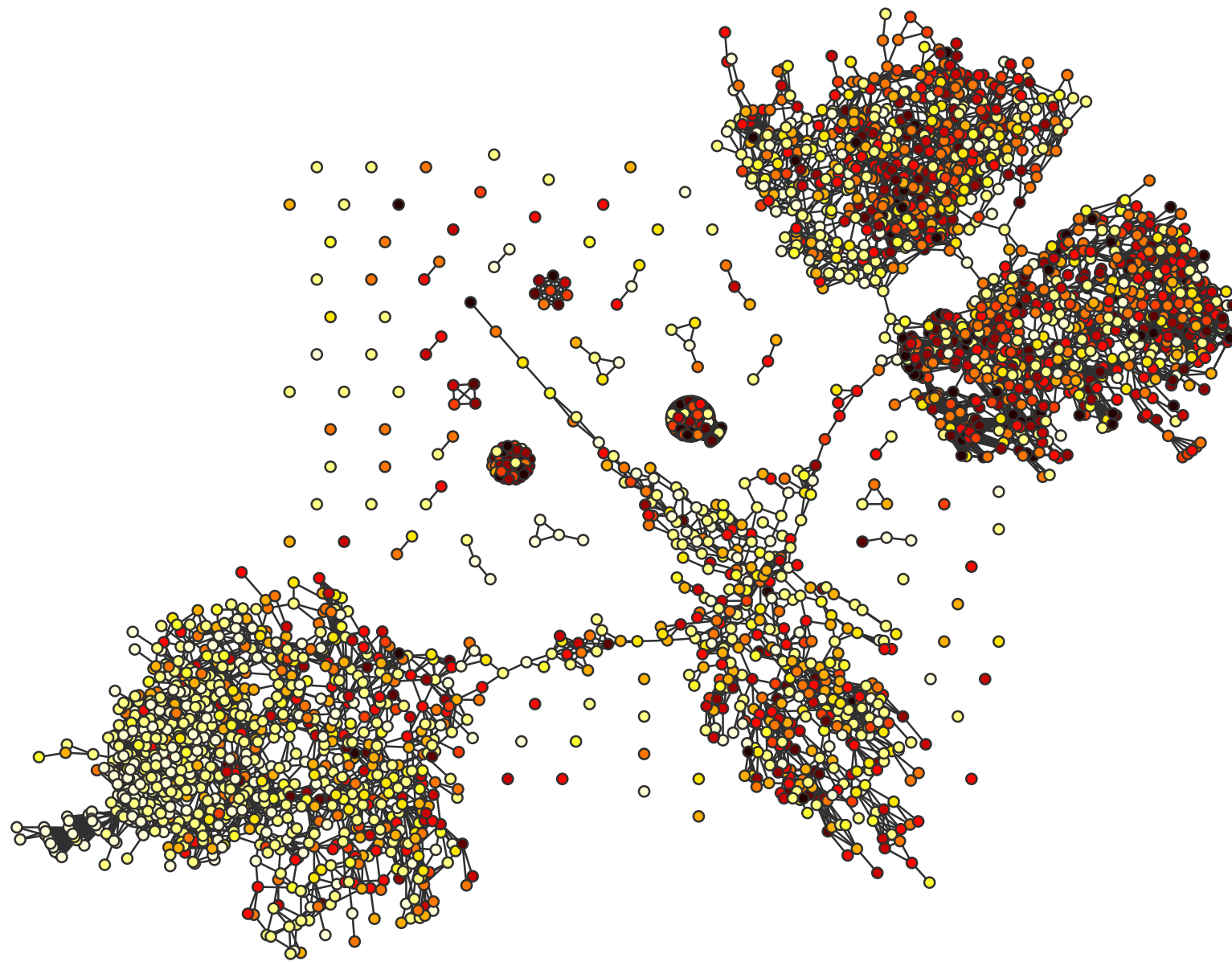
Topological model of COVID-19 spread over the US



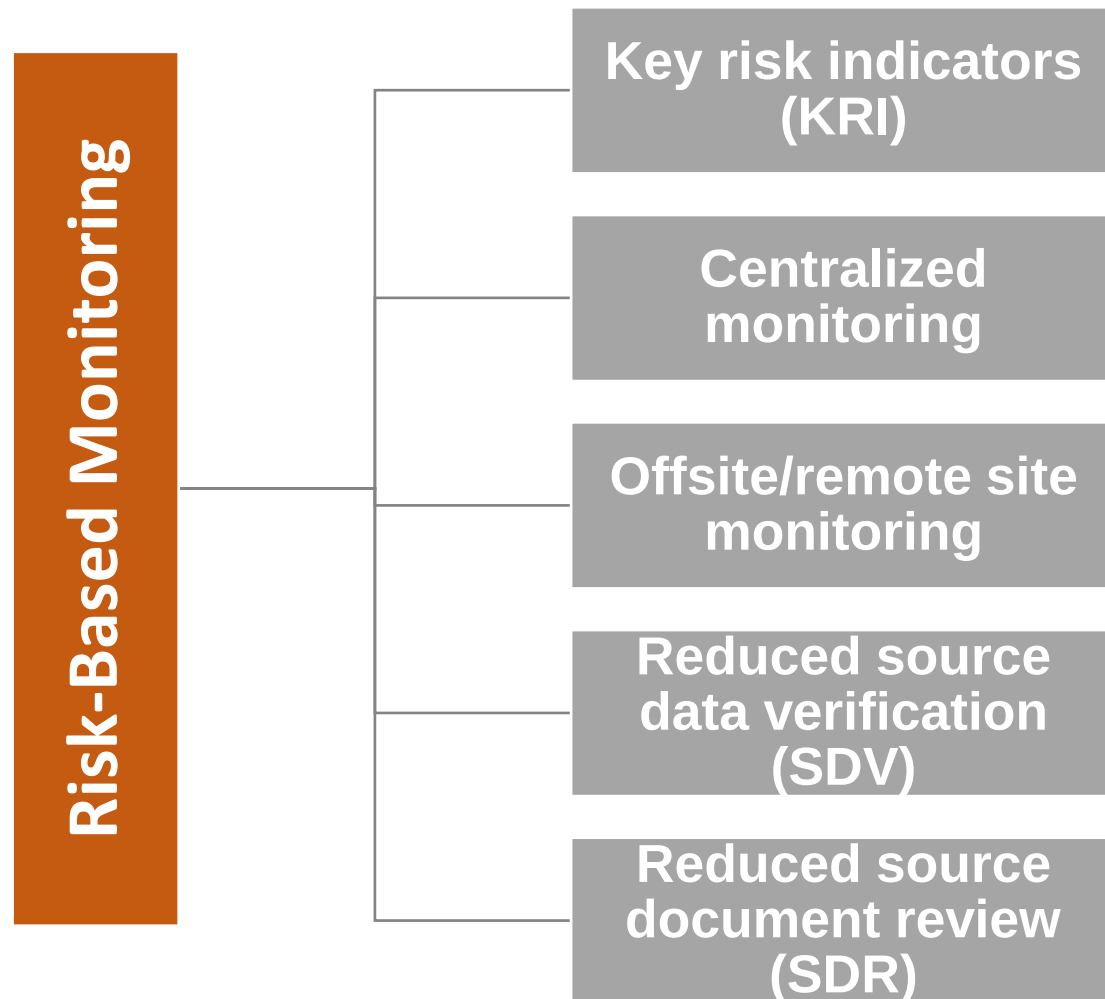
Topological model of COVID-19 spread over the US



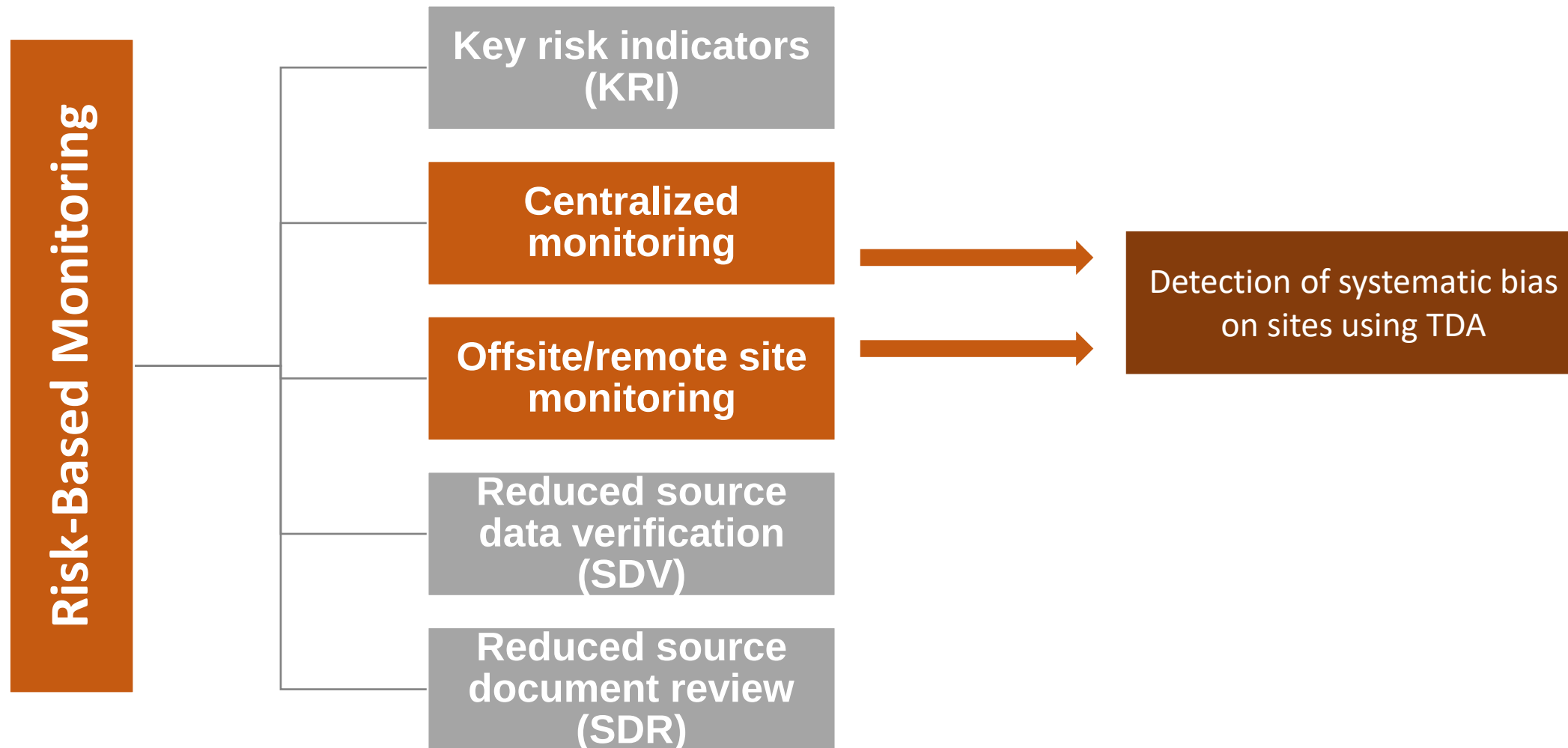
Topological model of COVID-19 spread over the US



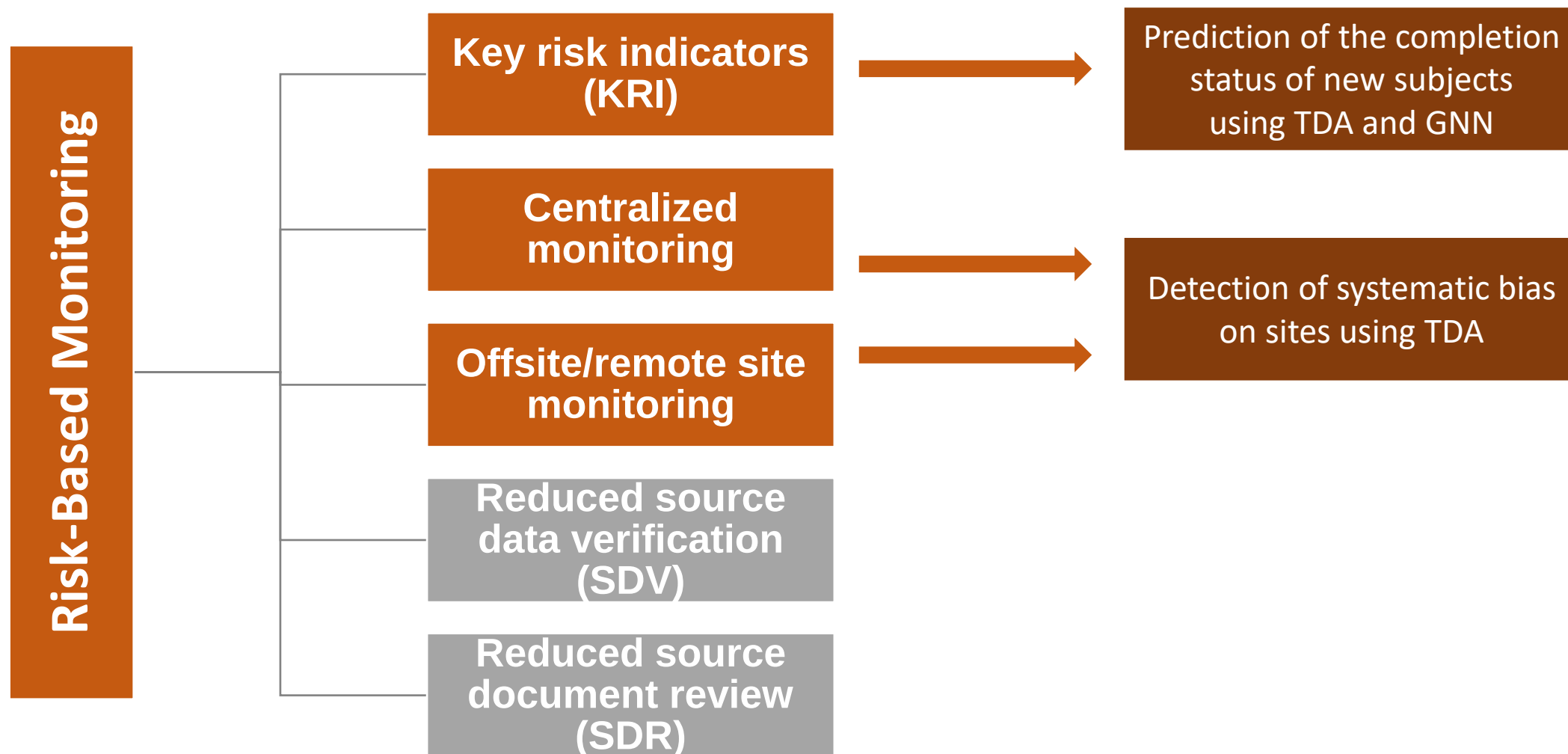
Key components of Risk-Based Monitoring and Solutions



Key components of Risk-Based Monitoring and Solutions

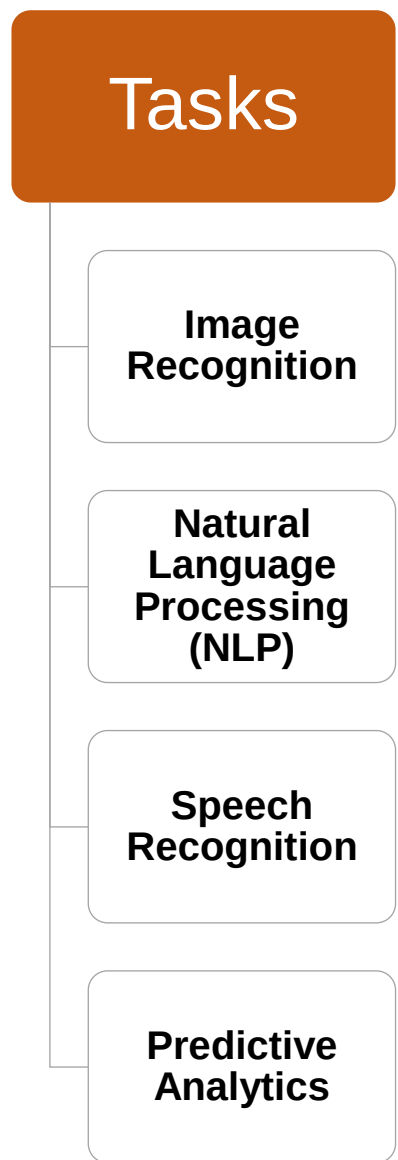


Key components of Risk-Based Monitoring and Solutions



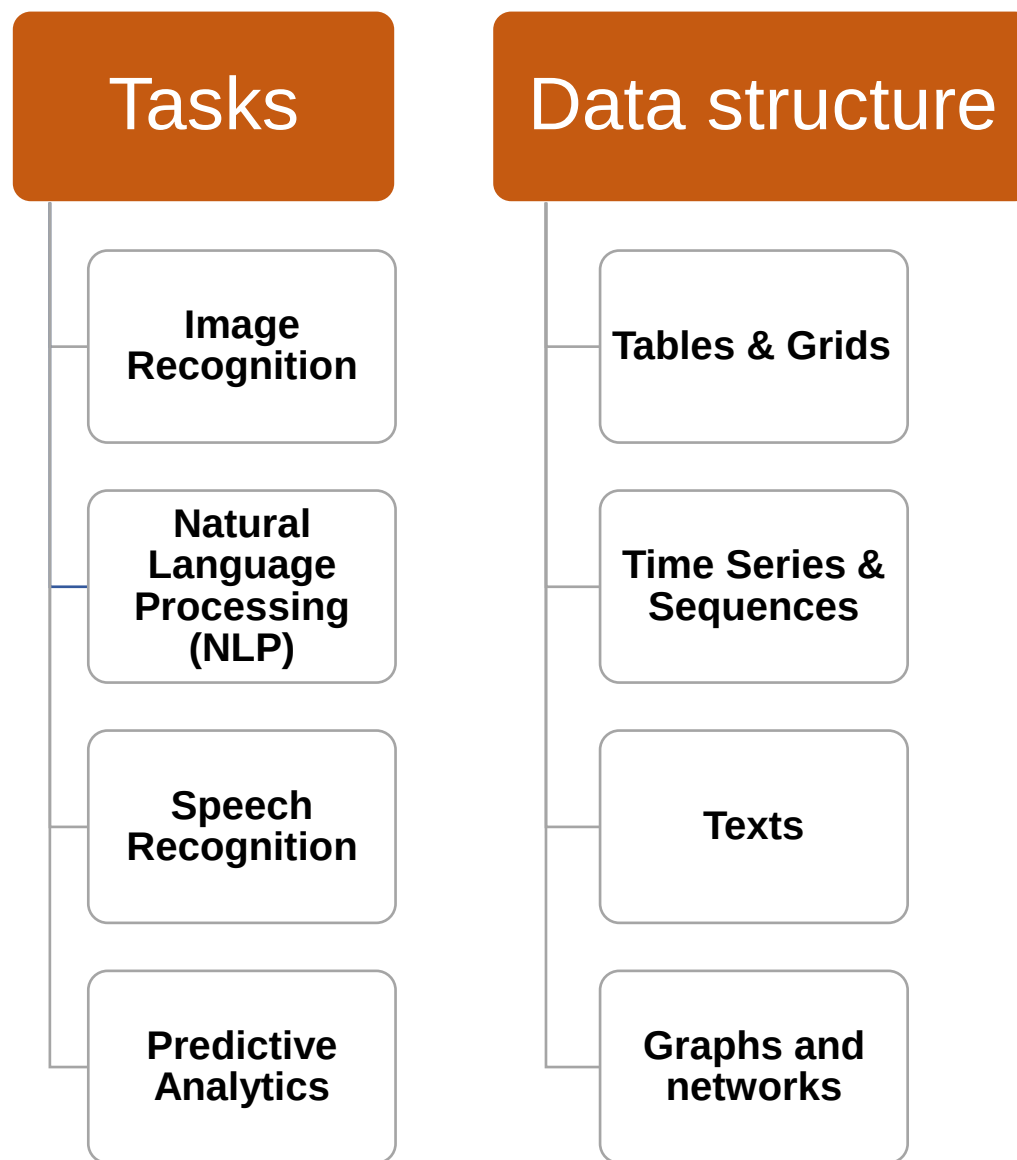
Introduction to Neural Networks

- NN can deal with a variety of **tasks**



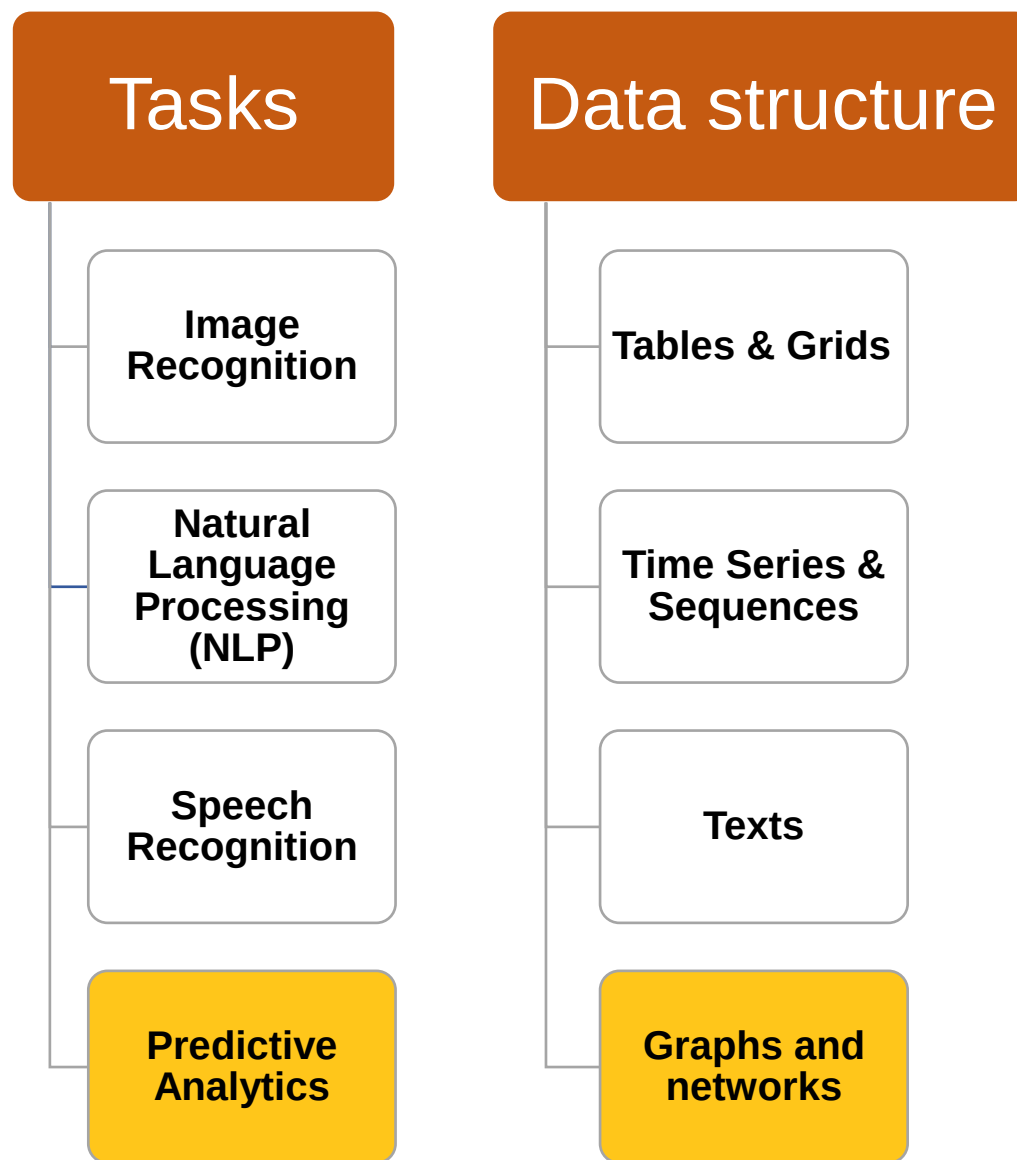
Introduction to Neural Networks

- NN can deal with a variety of **tasks**
- NN can be trained on data with different **structure**

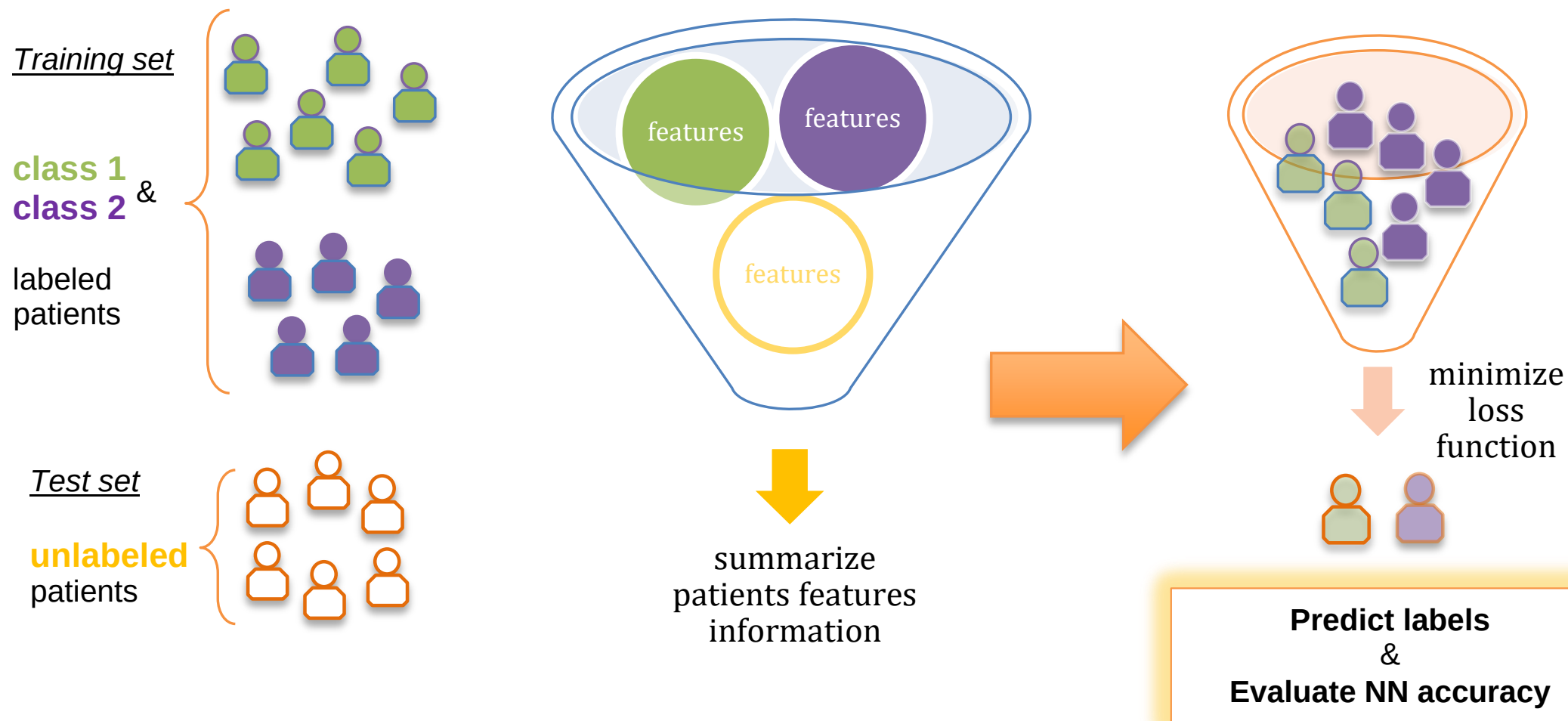


Introduction to Neural Networks

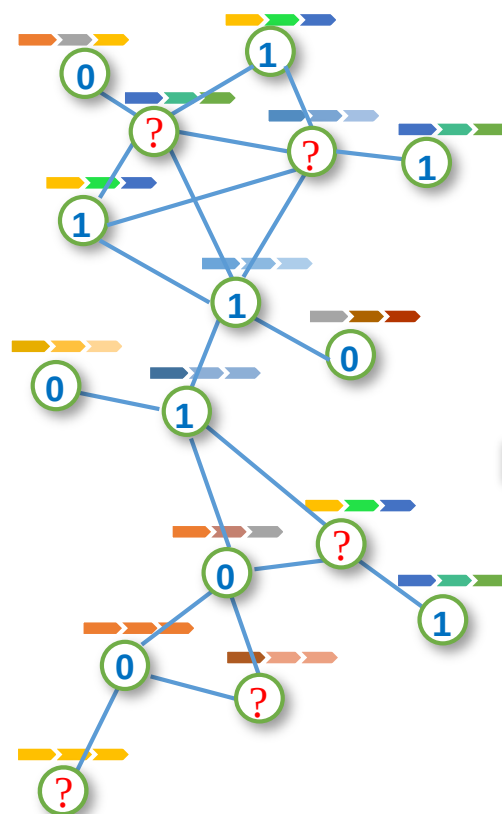
- NN can deal with a variety of **tasks**
- NN can be trained on data with different **structure**



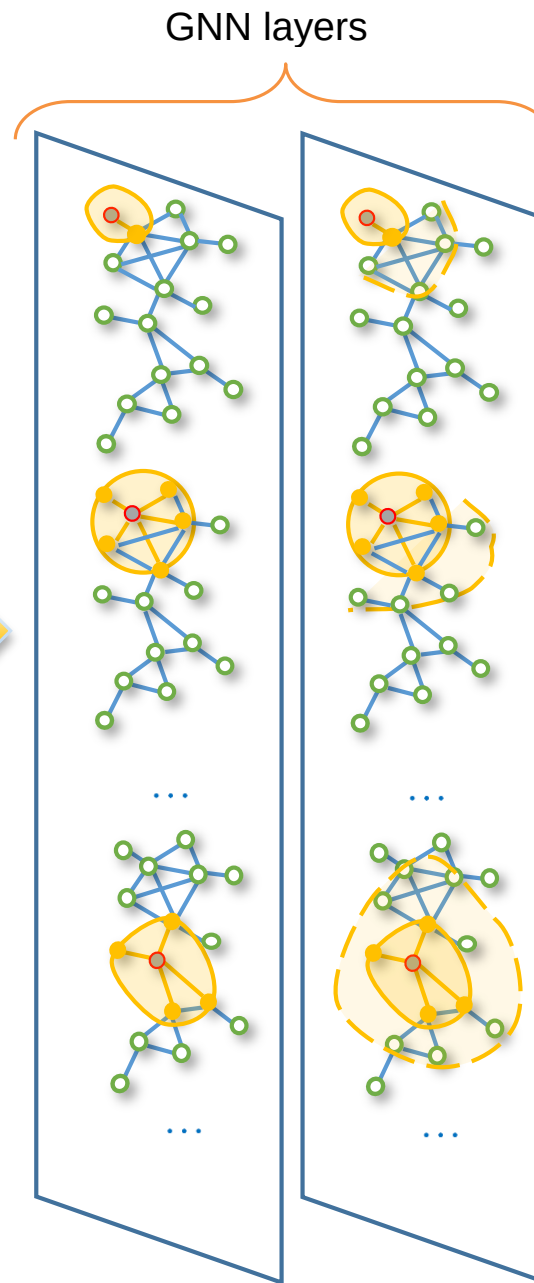
Neural Networks for predictive classification



Graph Neural Networks for predictive classification



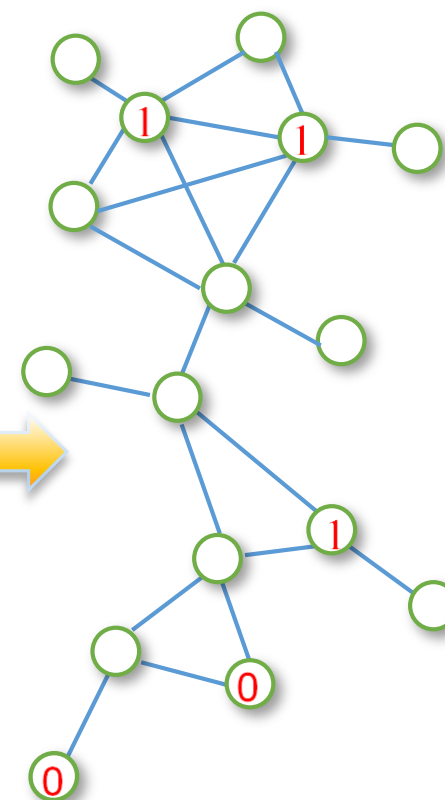
Initial graph with node features and labels



...



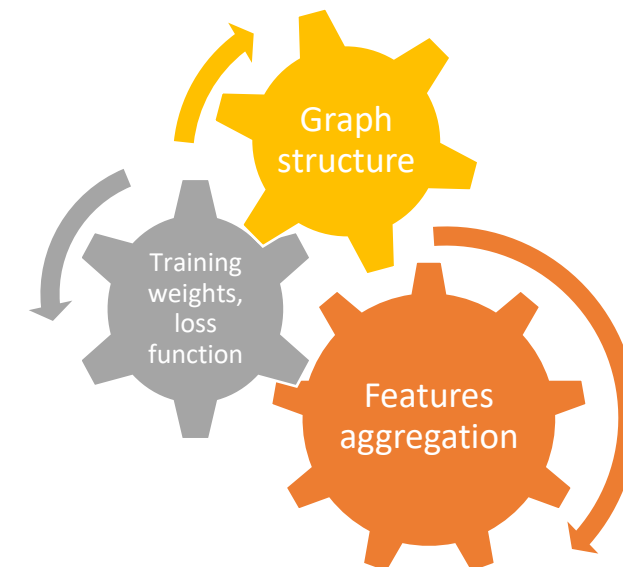
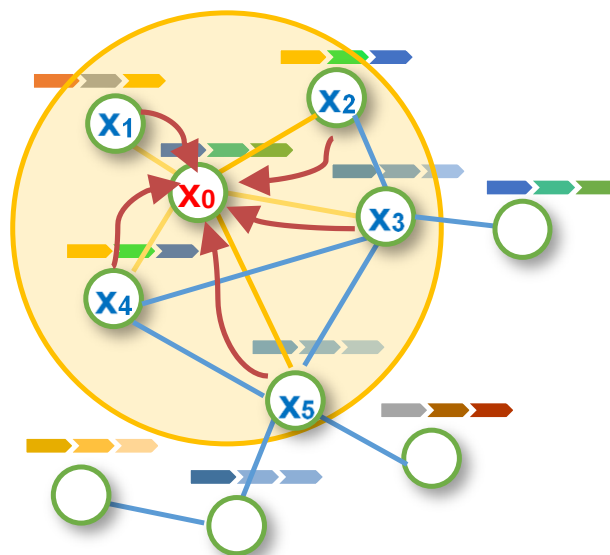
feature1 → feature2 → ... Features
1 0 Labeled nodes
? Unlabeled node



The graph with predicted node labels

GNN iteratively aggregates information from each **node's local neighborhood**, incorporating both the **topological structure** and **node feature information**

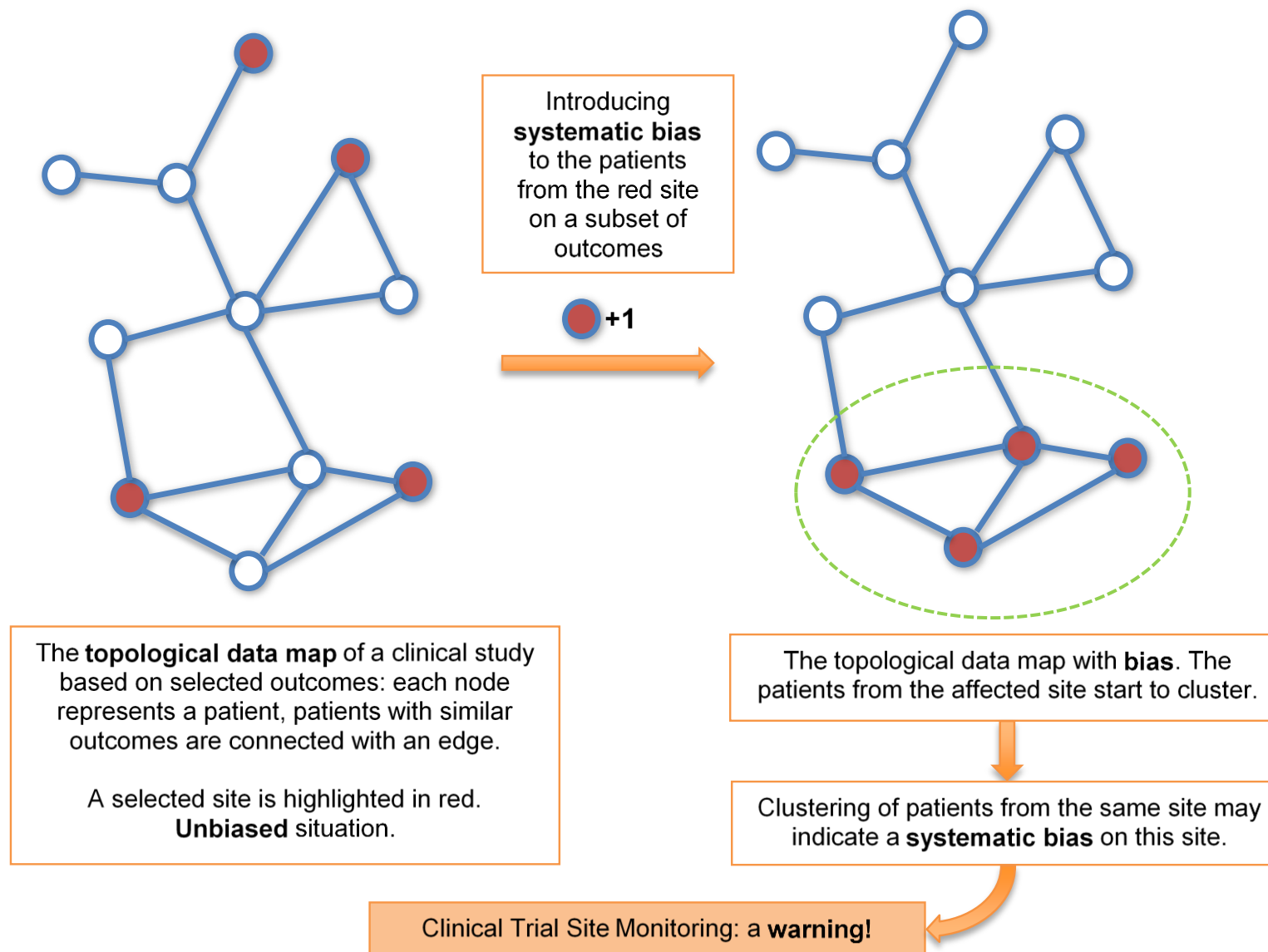
Graph Neural Networks layer



Node x_0 receives messages sent by all of its immediate neighbors x_1 to x_5 .

State1 → State2 → ... feature/state vector

Clinical trial site monitoring and bias detection



Experiment data summary

NIDA-CSP-1018

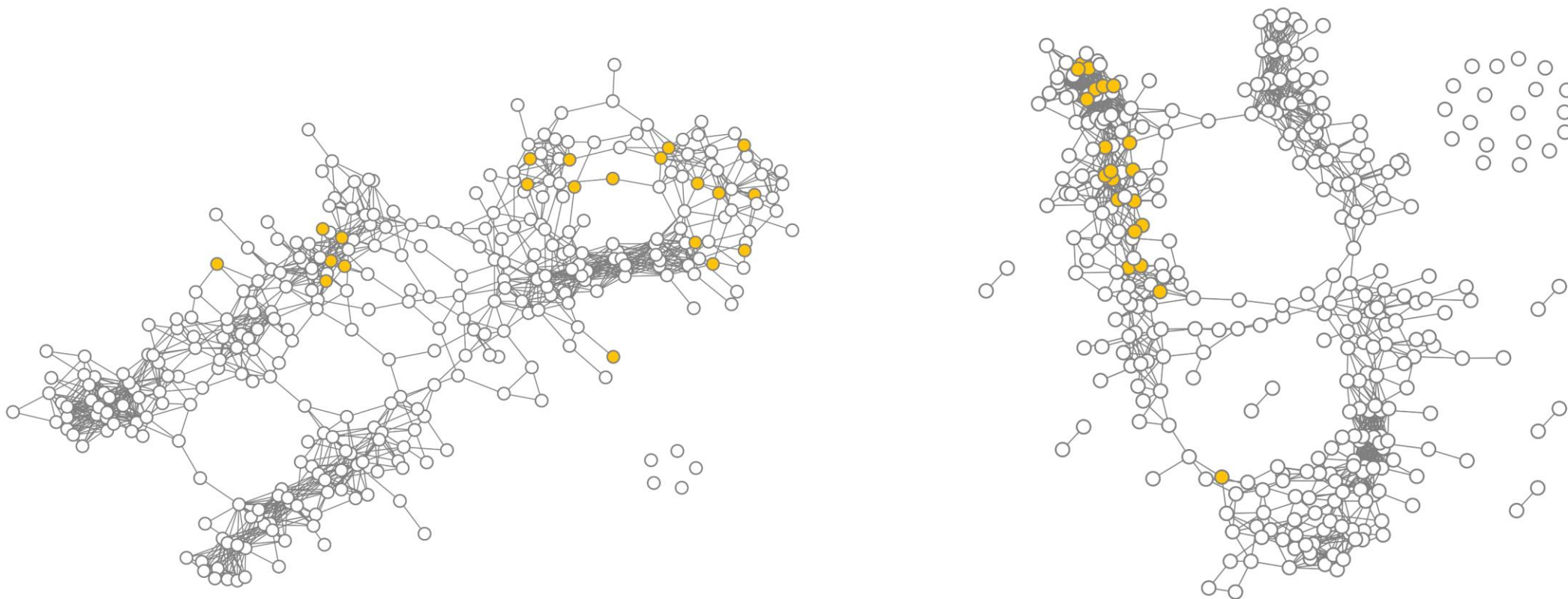
a multicenter safety trial of buprenorphine/naloxone for the treatment of opiate dependence

- **582** patients, **38** sites
- Study participation up to 12 months
- **>280** features: general health and vital signs, drug addiction, family history, employment status, etc.

Bias detection experiment

- **339** patients with **baseline + 9 visits** focused on the first 9 weeks of the study, no more than one missed visit
- Missing data imputed using LOCF method
- Data included **binary** features (e.g., presence/absence of certain symptoms of opiate addiction) and **numerical** features (e.g., temperature, respiratory rate)

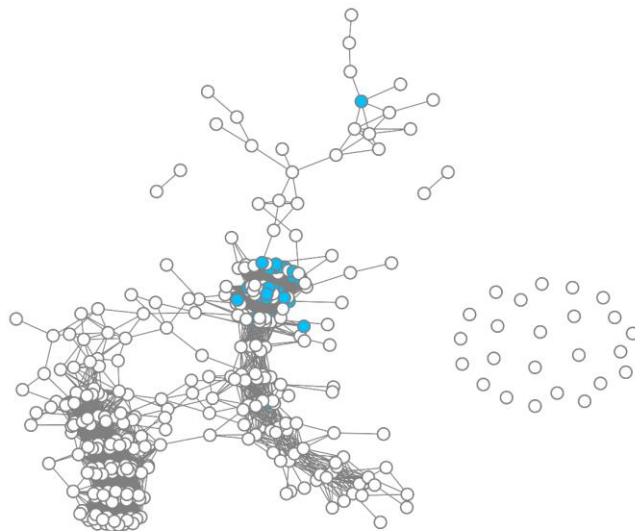
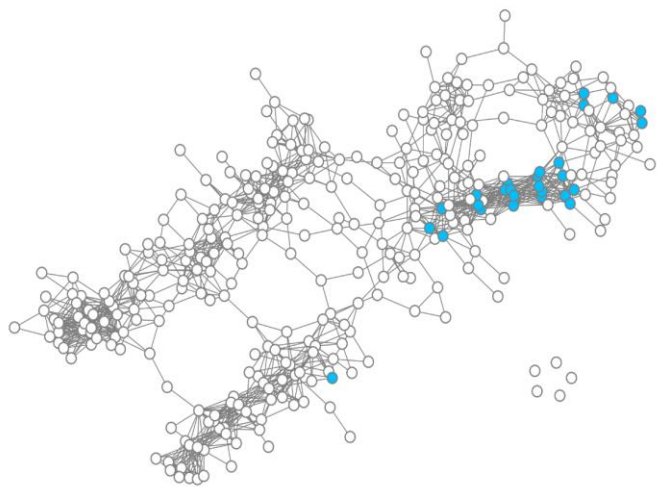
Bias detection experiment



TDA graph for pulse dynamics.

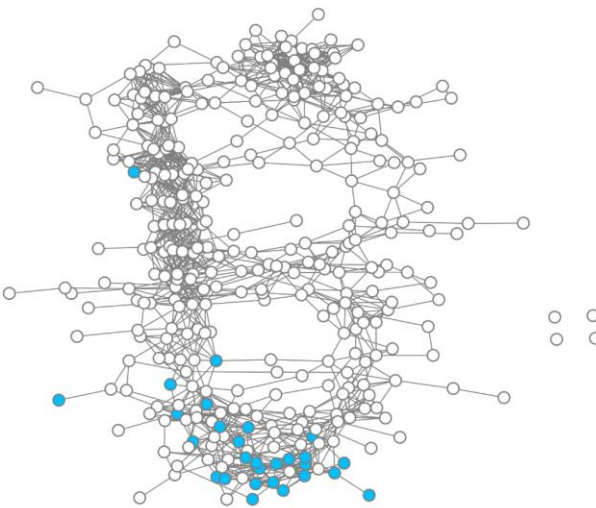
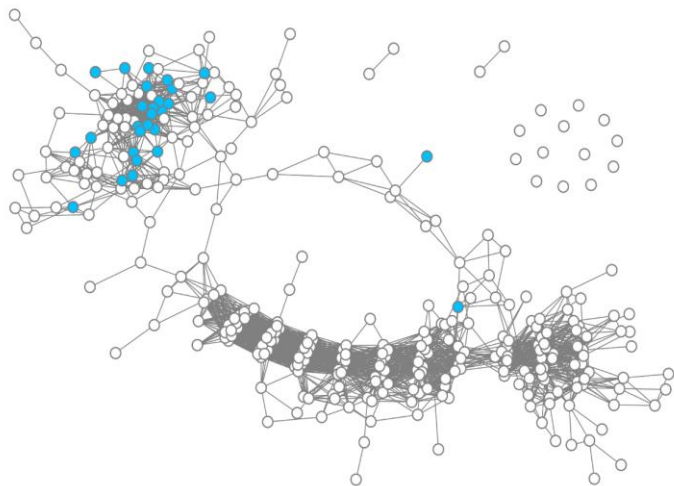
The patients from site 622 are colored. The bias introduced is a constant shift by +10, the connectivity index of the unbiased data (left) is 0.03, the connectivity index of the biased data (right) is 0.28

Bias detection experiment

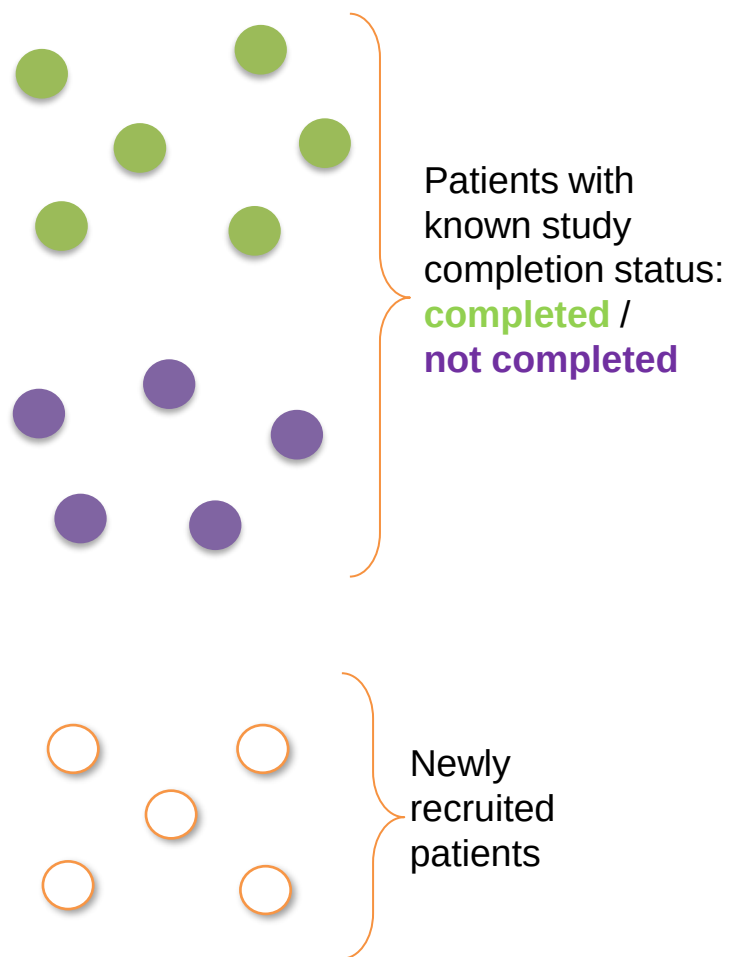


TDA graph with unbiased data for **site 178** by pulse dynamics, respiration frequency, number of opiate symptoms and body temperature.

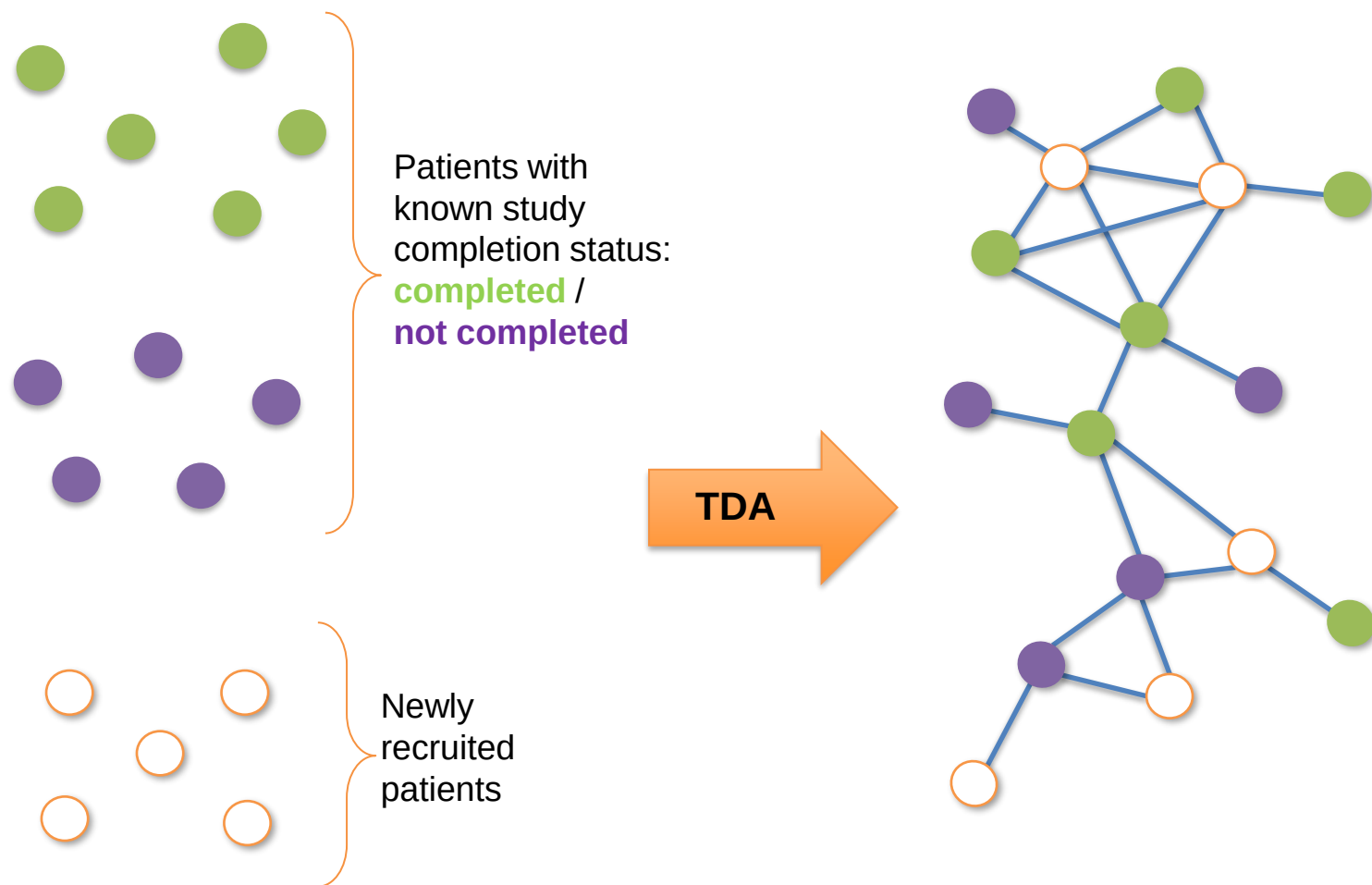
The similarity of the patients from site 178 by a number of different features is a signal to check the data gathering and processing on this site



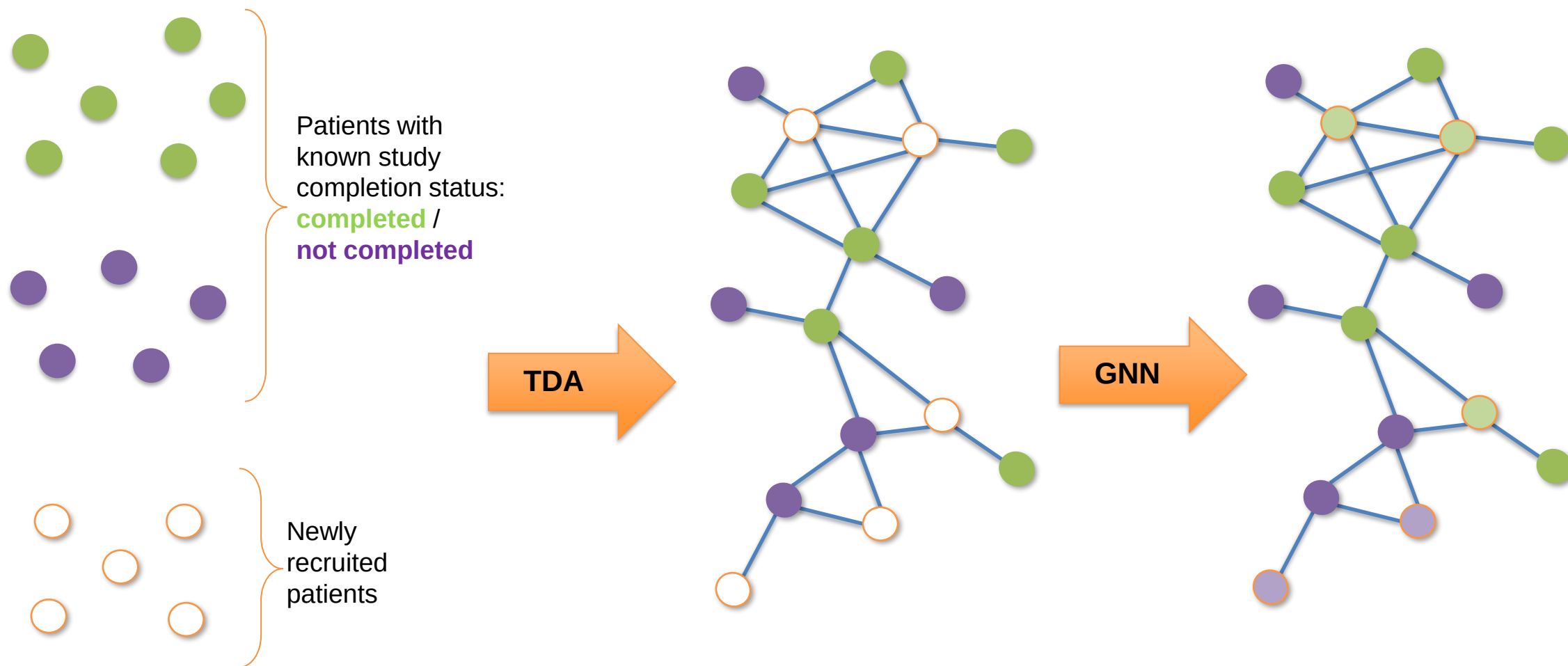
Prediction of patients' study completion status



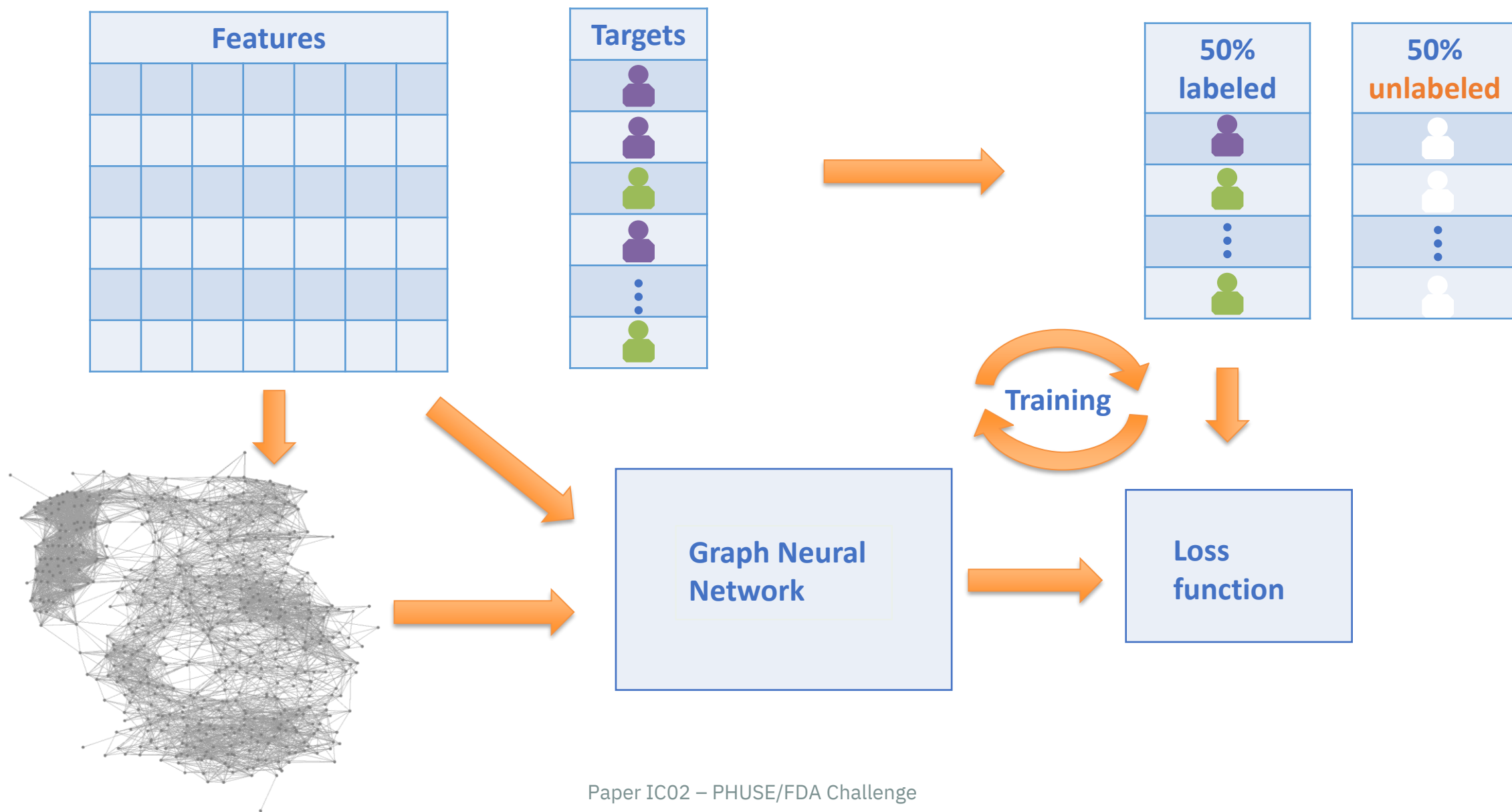
Prediction of patients' study completion status



Prediction of patients' study completion status

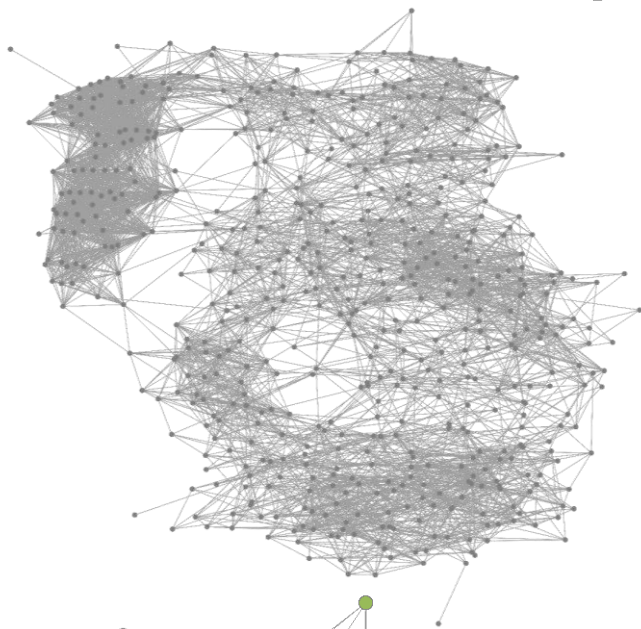


Prediction of patients' study completion status

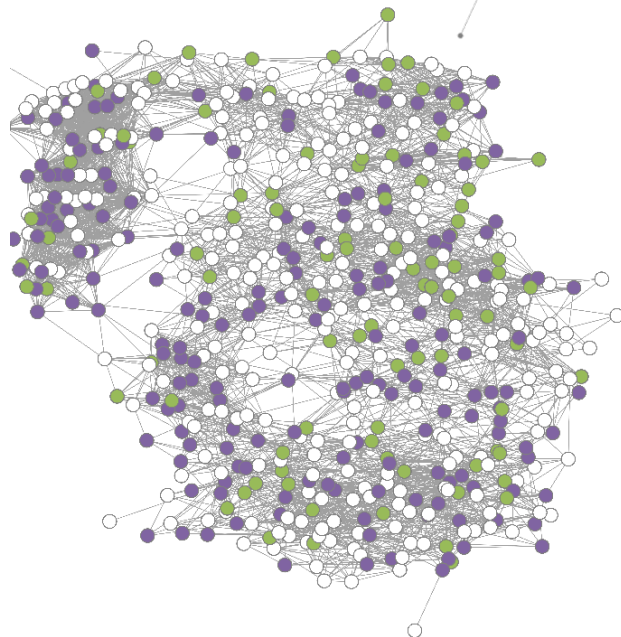


Prediction of patients' study completion status

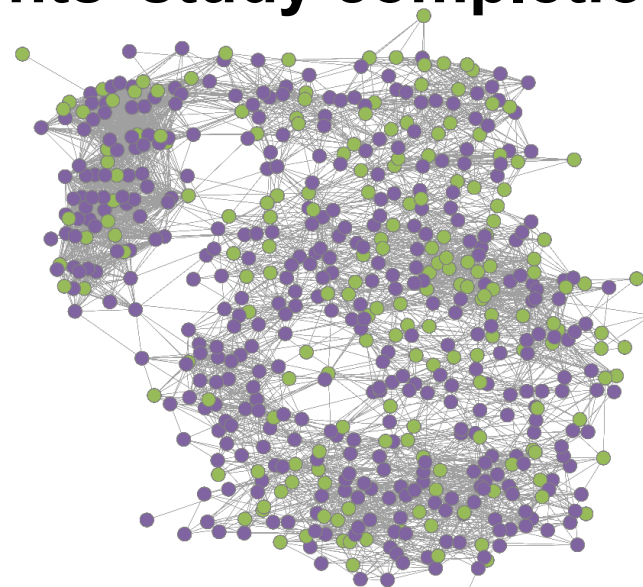
Starting graph
(Topological Model
of data)



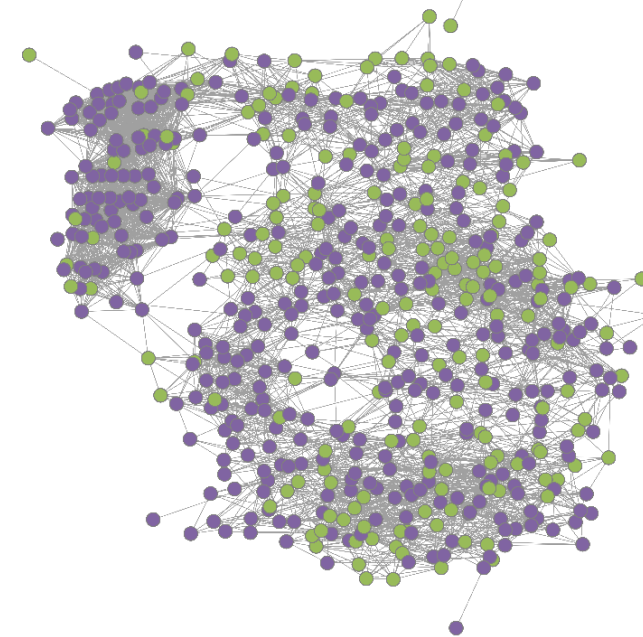
White nodes model
new patients (50%)



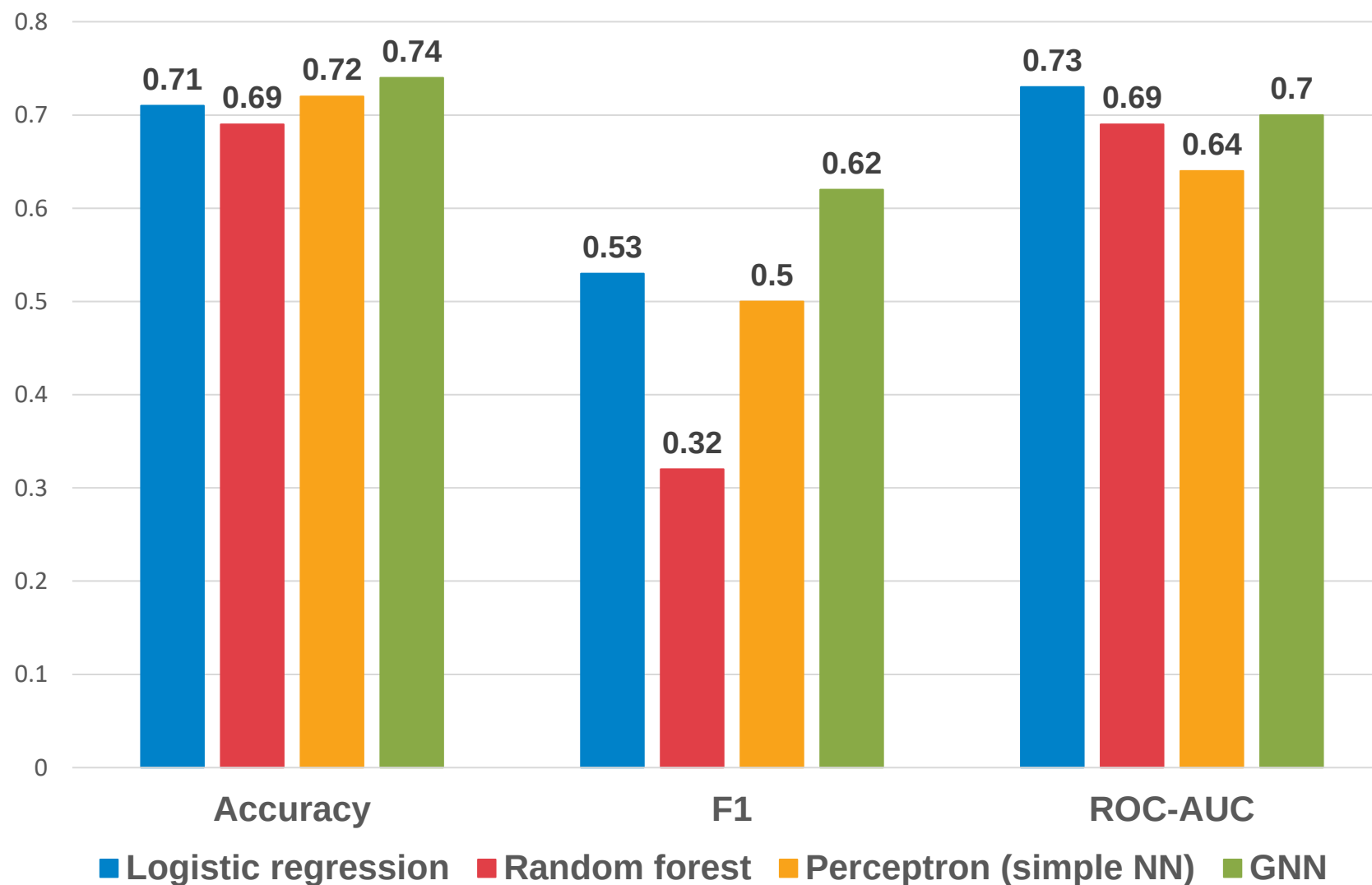
Coloring by completion:
completed /
not completed



Labels predicted by GNN



Comparison to other label prediction AI/ML methods



Presentation takeaways

- We focus on two important challenges of RBM:
 - clinical trial site monitoring
 - ongoing risk assessment related to inclusion of new patients
- Several effective AI-enhanced tools are proposed to address these challenges based on a combination of **Topological Data Analysis** and **Graph Neural Networks**
- Two successful experiments are carried out to demonstrate and support proposed solutions

*See our paper IC02 “**Risk-based Monitoring Using AI/ML Detection of Systematic Bias in Clinical Data**” for more details*

THANK YOU

Contact the authors:

Kostiantyn Drach kostya.drach@intego-group.com

Sergey Glushakov sergey.glushakov@intego-group.com

Intego Group, LLC

2300 Maitland Center Pkwy, Ste. 240,

Maitland, FL 32751

Phone: +1 (407) 512-1006

<https://integoclinical.com/data-science>