Graphing Healthcare Networks: Data, Analytics & Use Cases

Martin S. Zand MD PhD

Professor of Medicine and Public Health Sciences

Rochester Center for Health Informatics

University of Rochester Medical Center

Eighth Annual TUC Meeting

June 22, 2015

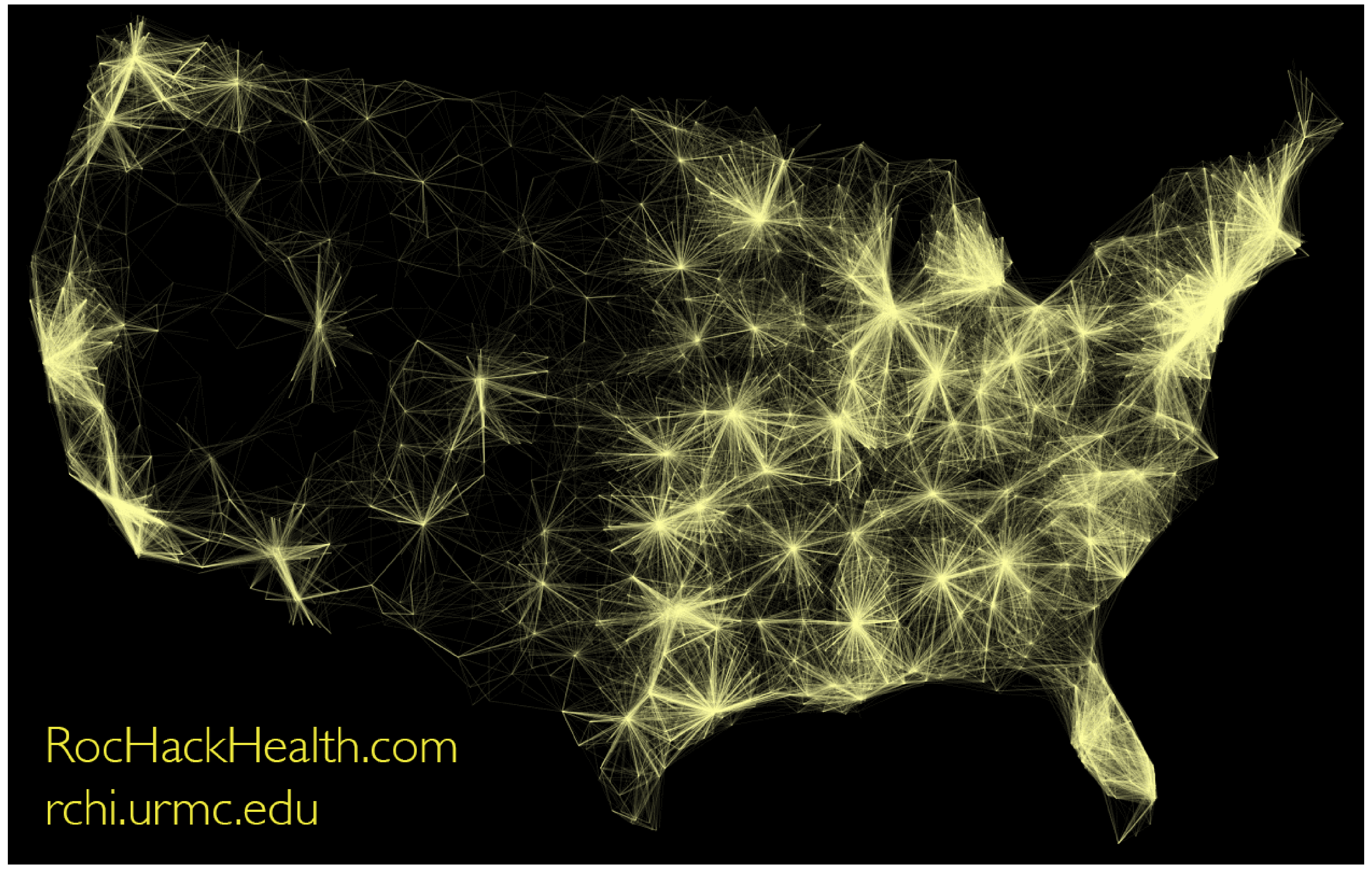
UNIVERSITY of ROCHESTER

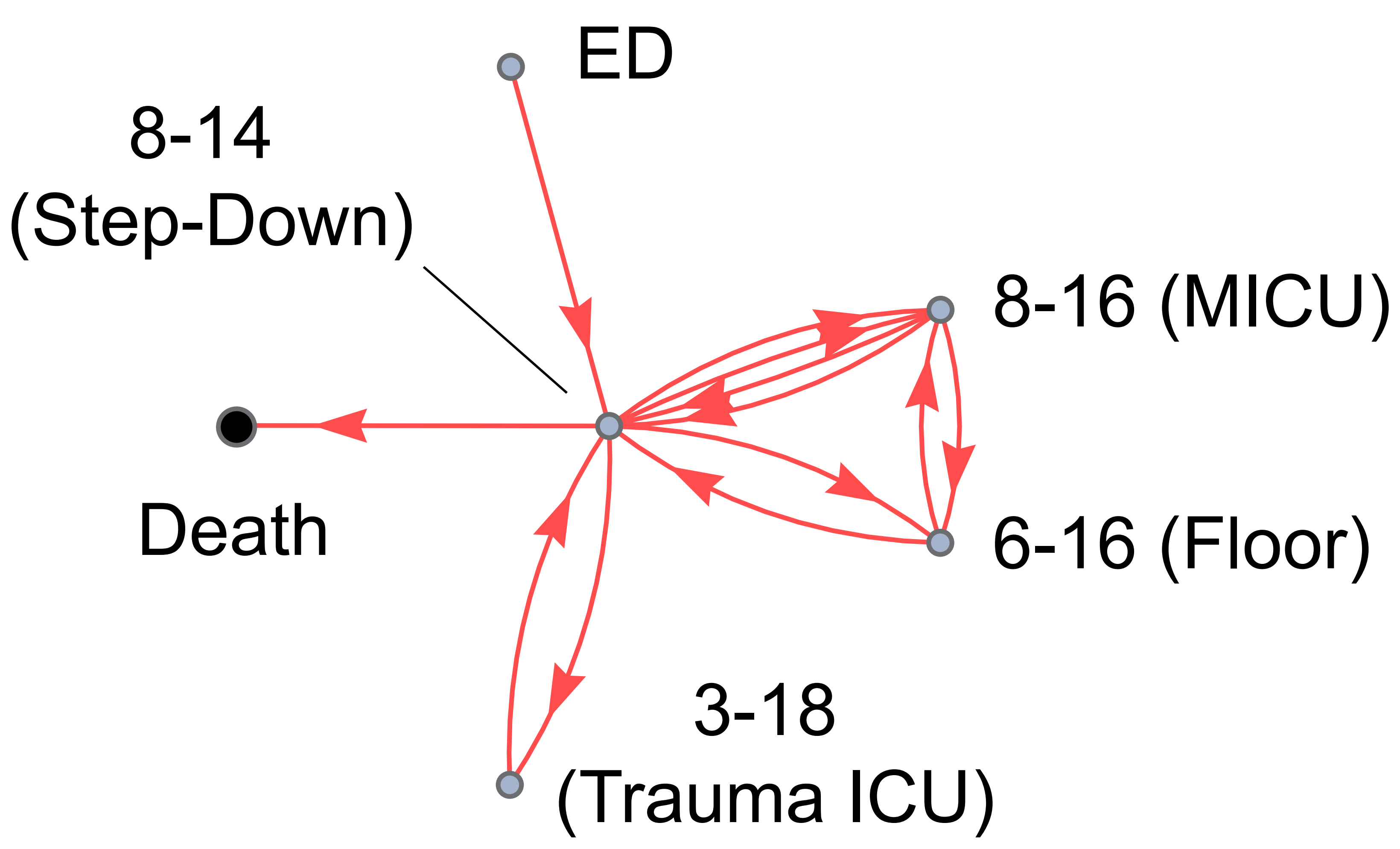
LDBC The graph & RDF benchmark reference

rchi.urmc.edu

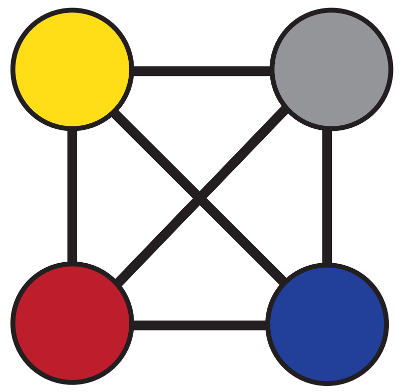
Why Graph Databases in Healthcare?

- Healthcare is delivered by networks
  - We don't understand their toplogy or dynamics
- Patients traverse those networks
  - A patient journey is a multipartite, directed, temporal subgraph of the larger network
- Network topology influences outcomes
  - Provider-provider networks determine referrals
  - Homophily among providers



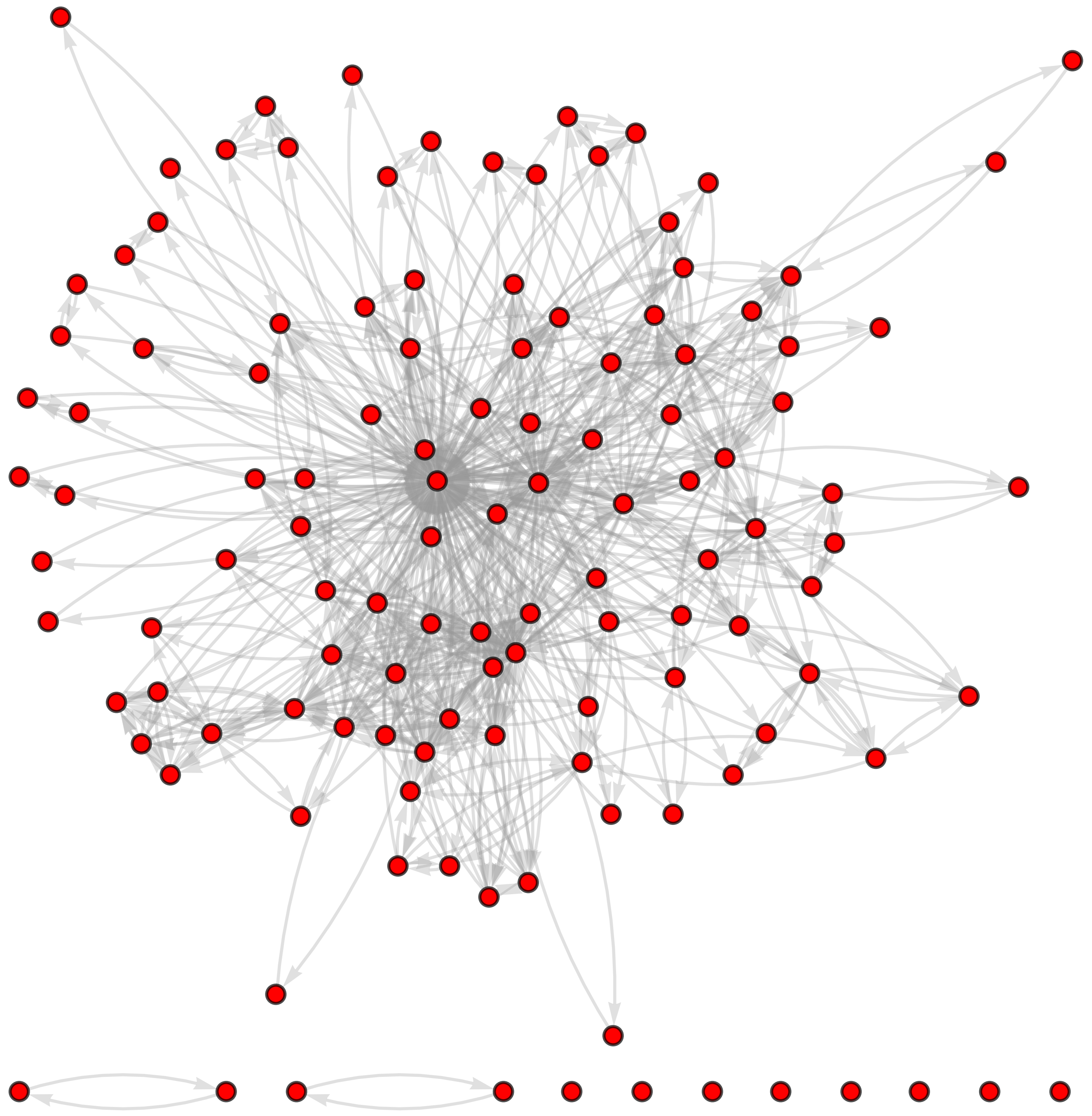


Fundamental Structure of Healthcare Data

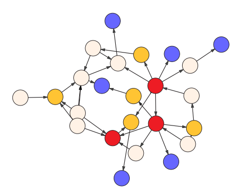


+ t

{Patient, Provider, Location, Event, Time}



Age	Location	In/Out-patient	Encounter	Absolute
Sex	Specialty	Geo Location	Admission	Interval
Race	Organization	County	Procedure	.
Address	Identifier	State	Diagnosis	.
Insurance	Education	Census Tract	Prescription	.
Attributes	Certification	Service Area	Birth/Death	
.	.	.	.	
.	.	.	.	
.	.	.	.	



Data Set

- Basic data structure
  - 209,173 ER visits and hospital admissions to a single hospital
  - All locations and duration of "dwell" for each admissions

- Associated data
  - Discharge destinations (including death)
  - Diagnoses, procedures
  - All medications, lab test results, bacterial and viral culture results

- Outcomes
  - Hospital "disposition"
  - Discharge destinations (including death)

Use Case 1: Intensive Care Unit Readmissions

Questions:

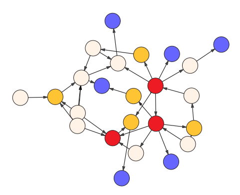
- Are patient outcomes associated with graph toplogy?
- Can we identify groups of patients based on graph topology?

Approach:

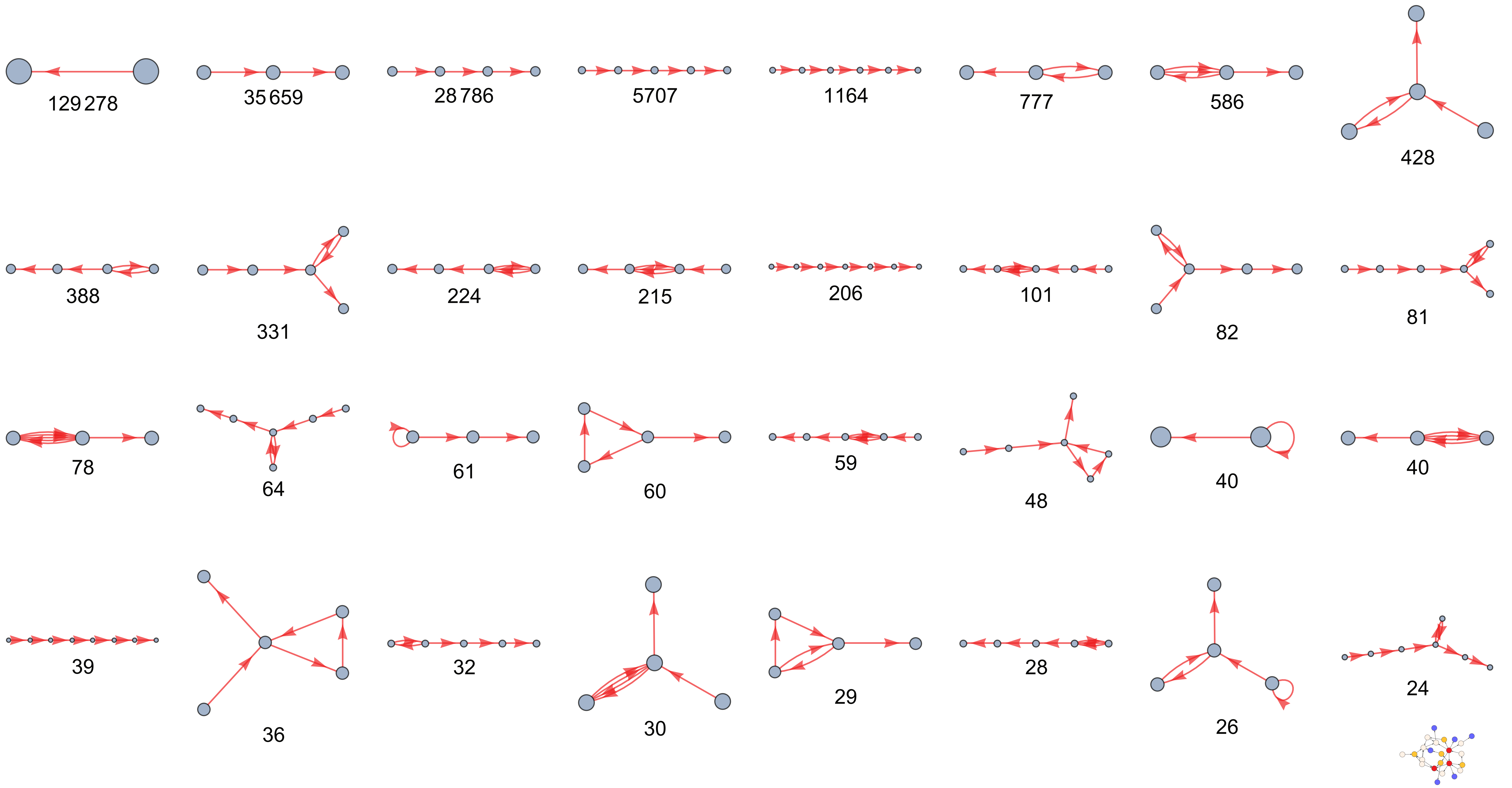
- Pathgraph mapping of patient journeys

Application:

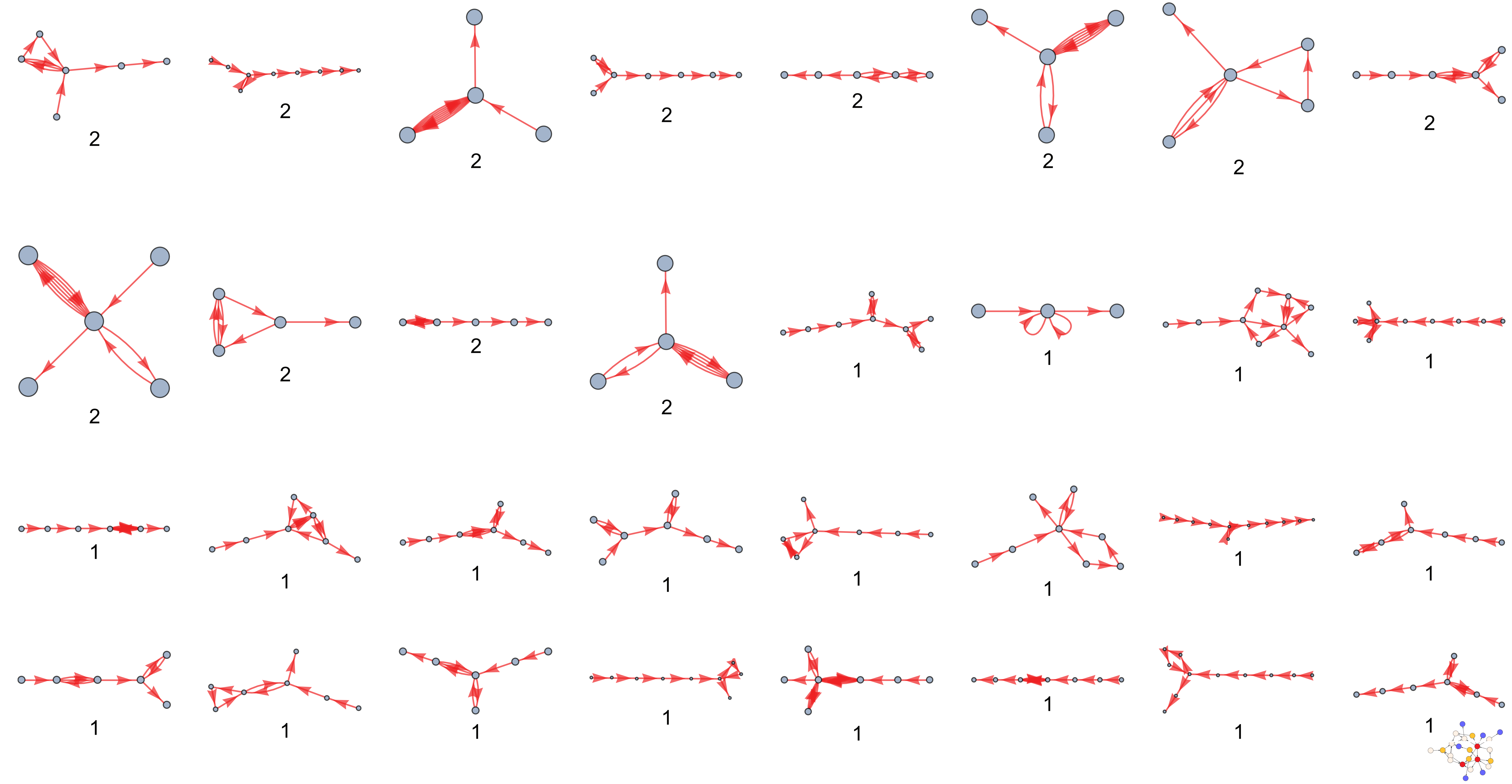
- Use journey topologies for predictive modeling and to design early interventions



Use Case I: Graph Isomorphism

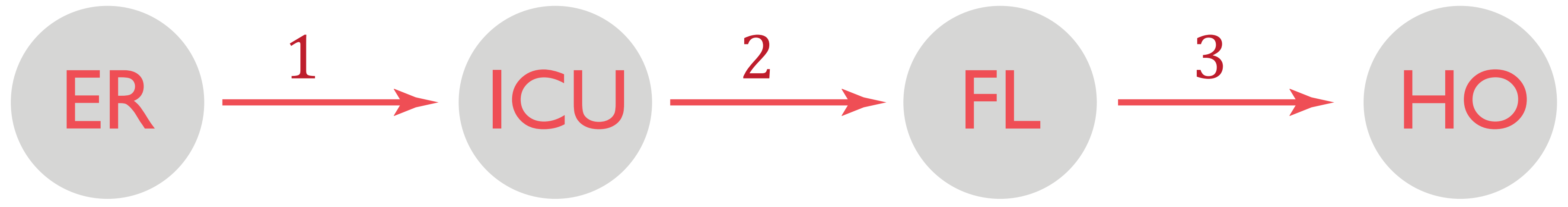


Use Case I: Graph Isomorphism

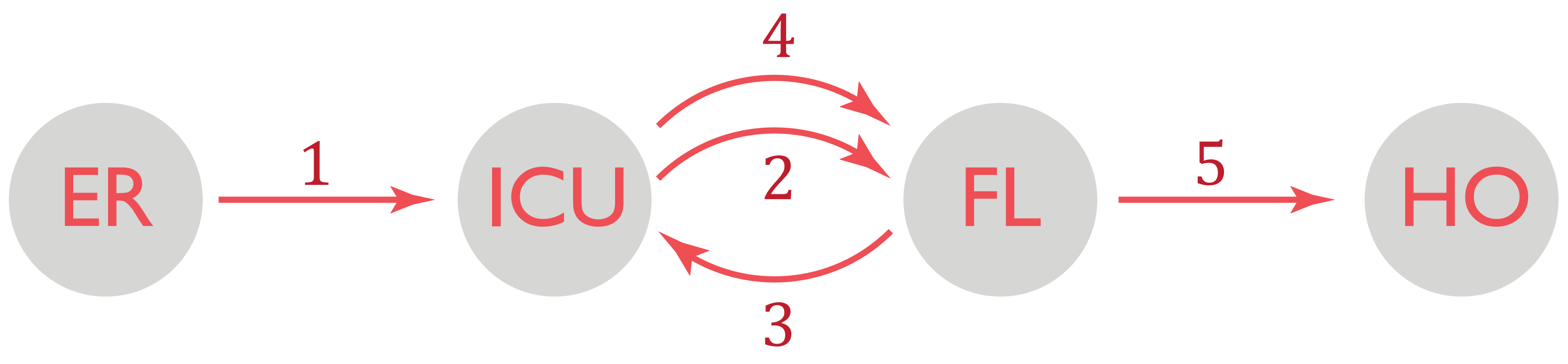


Basic Patient Journey Topology

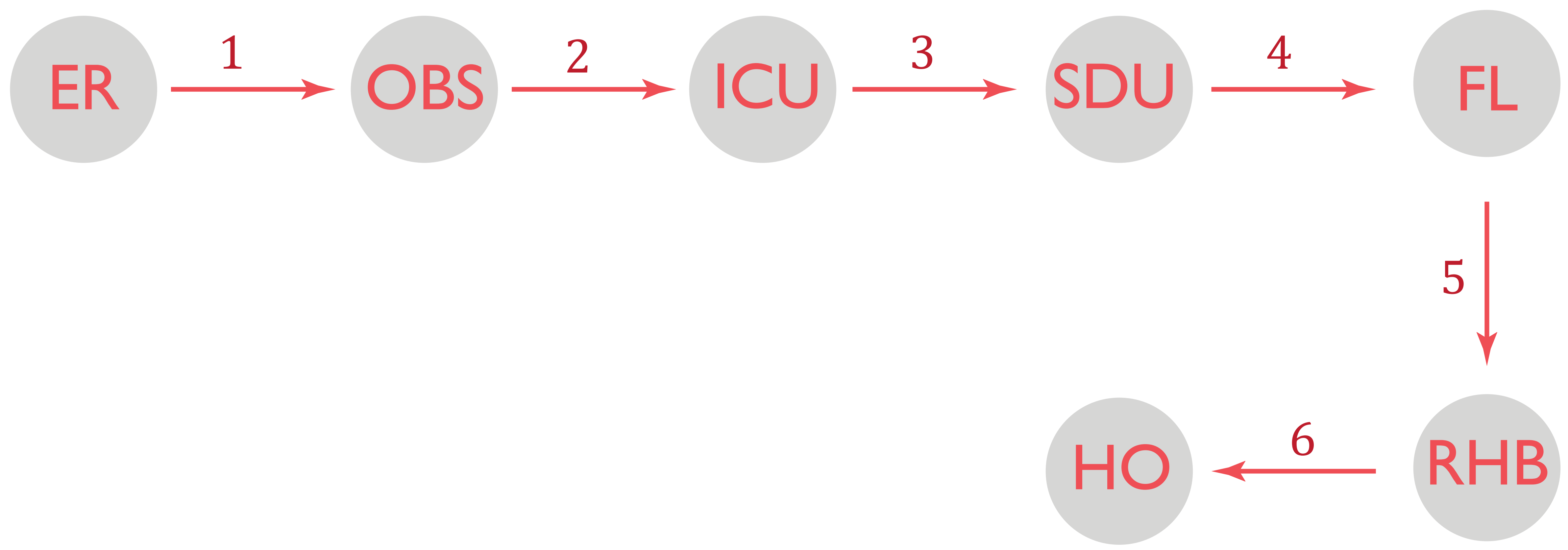
Linear Journey



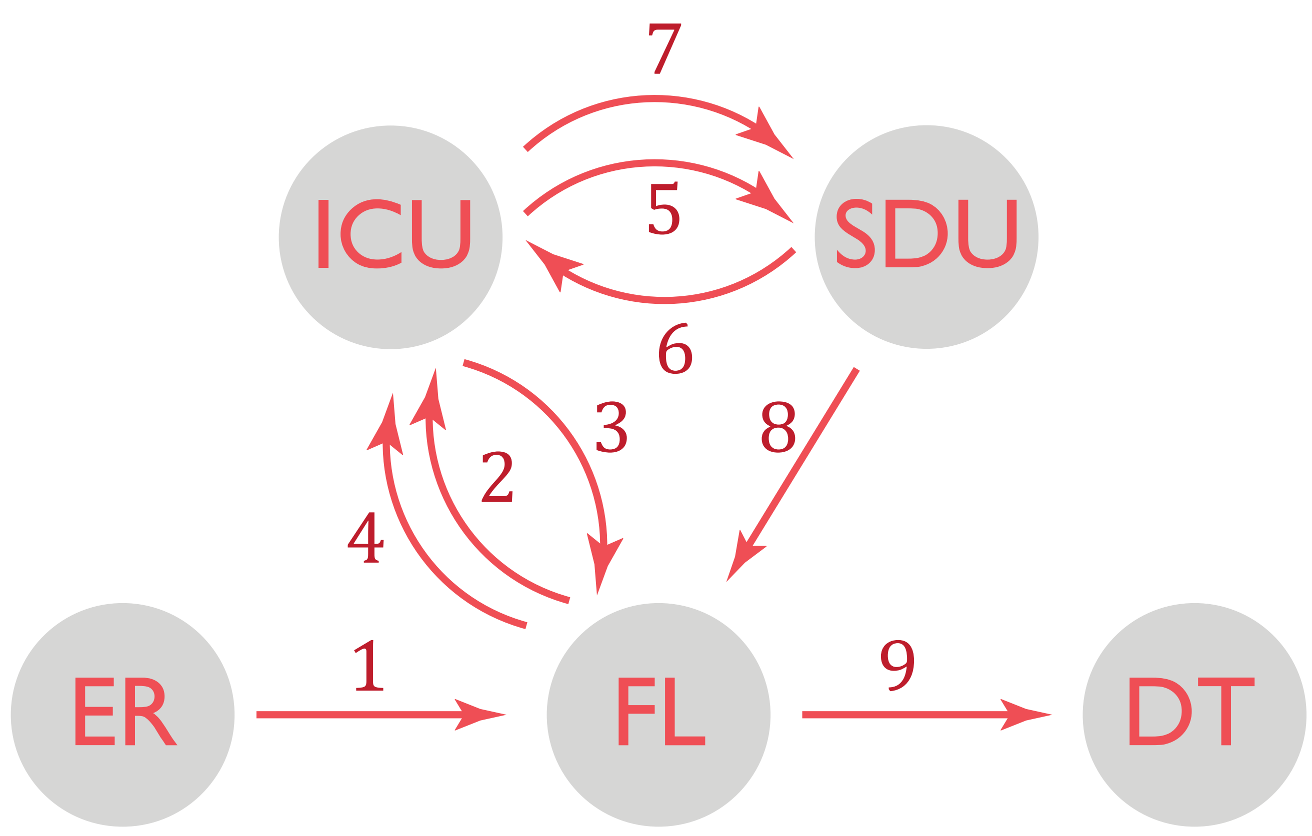
Cyclic Journey



Complex Linear Journey



Complex Cyclic Journey



Is a cyclic patient journey associated with a poor outcome?

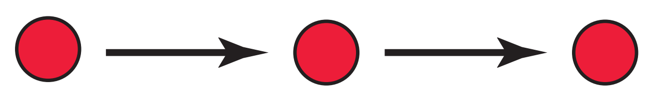
123,184 patients presenting to URMC
209,173 separate encounters
January 2012 - December 2014

Sequential patient exclusion criteria

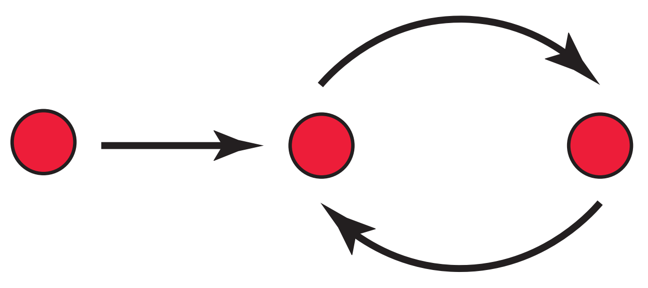
- Patients never admitted to ICU or STDU (n=91,035)
- Index admission only (n=8,468)
- Age < 18 yo (n=3,215)
- Age ≥ 18 but admitted to a non-adult ICU (n=22)
- Patient died during index ICU admission (n=620)
- Patients admitted to STDU only (n=6,885)

4,756 survived first ICU admission
eligible for re-admission to ICU

Linear trajectory
(n=3,277)



Cyclic trajectory
(n=1,479)

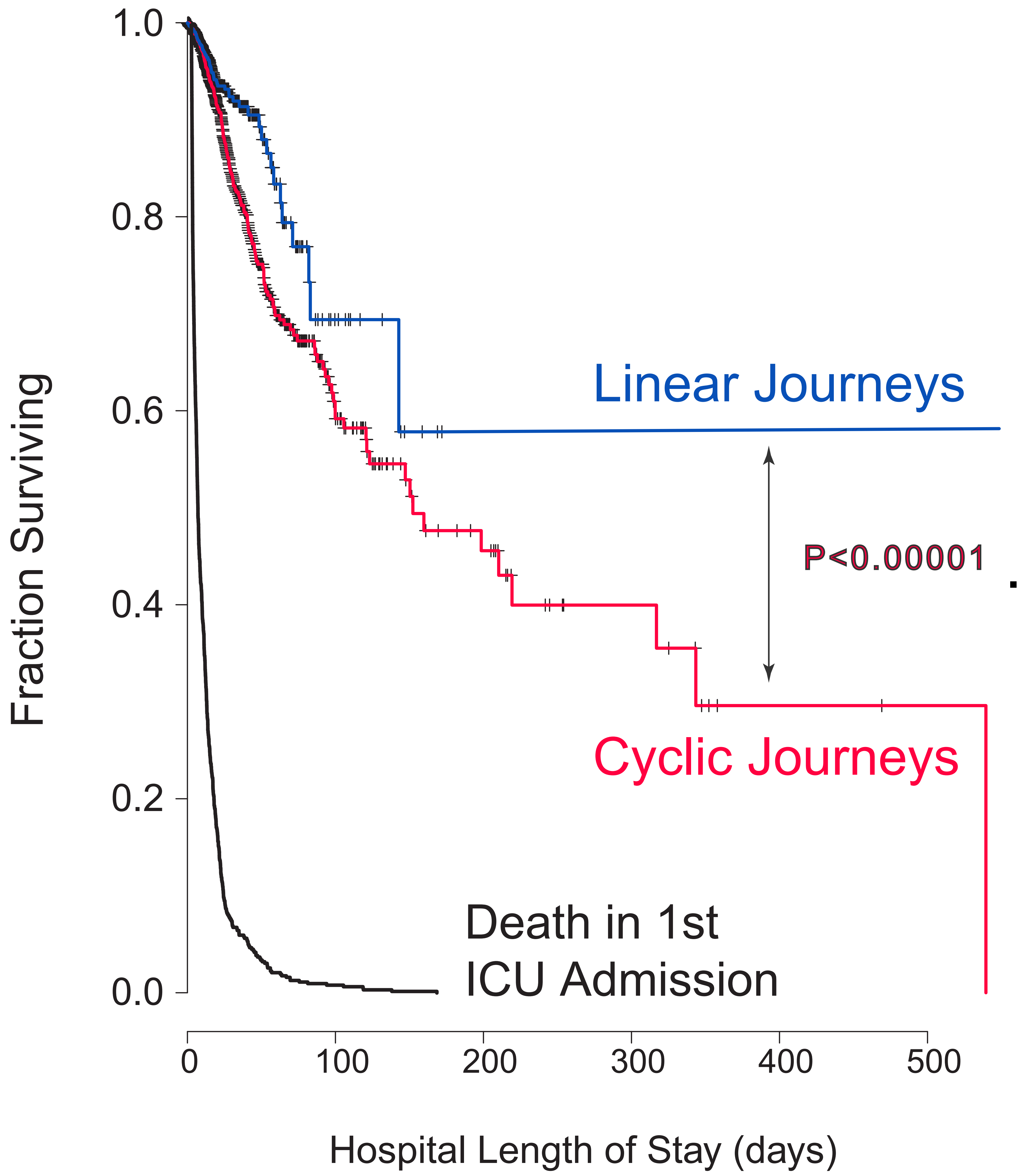


Quill C ... Zand MS (2016) *in preparation*

Graph topology matters for survival

	Totals (n=)	Linear (%) (n=3277)	Cyclic (%) (n=1479)	Significance (p-value)
Diabetes	1503	29.0	37.4	p < 0.000001
Malnutrition	105	1.5	3.8	
Respiratory Failure	1434	24.8	42.1	p < 0.000001
CHF	1055	17.3	33.1	
Vent	973	16.4	29.5	
COPD	670	12.5	17.7	
Asthma	364	7.8	7.4	N.S.
Lipid Metabolism	1420	28.2	33.5	p<0.0003
PVD	1095	19.5	30.9	p < 0.000001
MI	384	6.3	12.0	
CVD	671	14.1	14.1	N.S.
Acute Kidney Injury	1255	21.6	36.9	p < 0.000001
Chronic Kidney Disease	737	12.5	22.2	
ESRD	188	3.0	6.1	
Anemia	1447	26.0	40.2	p < 0.000001
Cancer	800	14.4	22.2	
Coagulation Disorder	517	8.3	16.5	
Sepsis	911	14.3	29.8	p < 0.000001
HIV	46	1.0	1.0	N.S.
Liver Disease	412	6.4	13.7	p < 0.000001
Hepatitis	110	2.1	2.8	N.S.

Quill C ... Zand MS (2016) *in preparation*



Use Case 2: Hospital Flow Mapping

Questions:

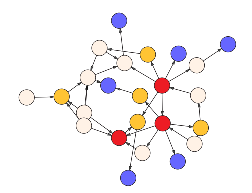
- How are hospital floors/units *really* linked by patient transfers?
- Can we model patient flow between units?

Approach:

- Traceroute graphing of the hospital using patient admission traces

Applications:

- Managing hospital patient distribution
- Graph "module" based feature vectors for classifying hospitals



Graph inspiration....

$ traceroute **74.125.226.18**

24.93.10.62

↓

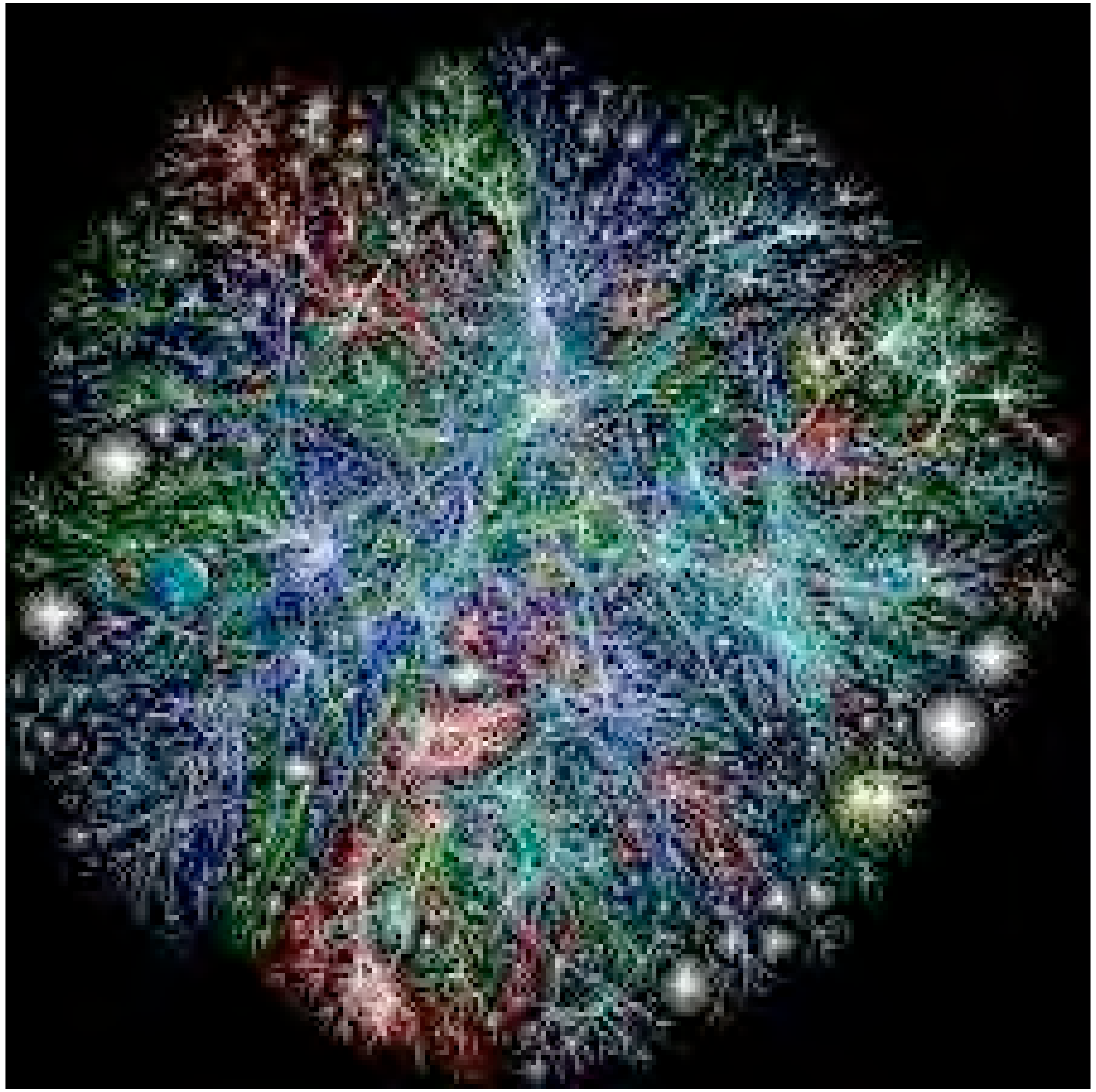
107.14.19.100

↓

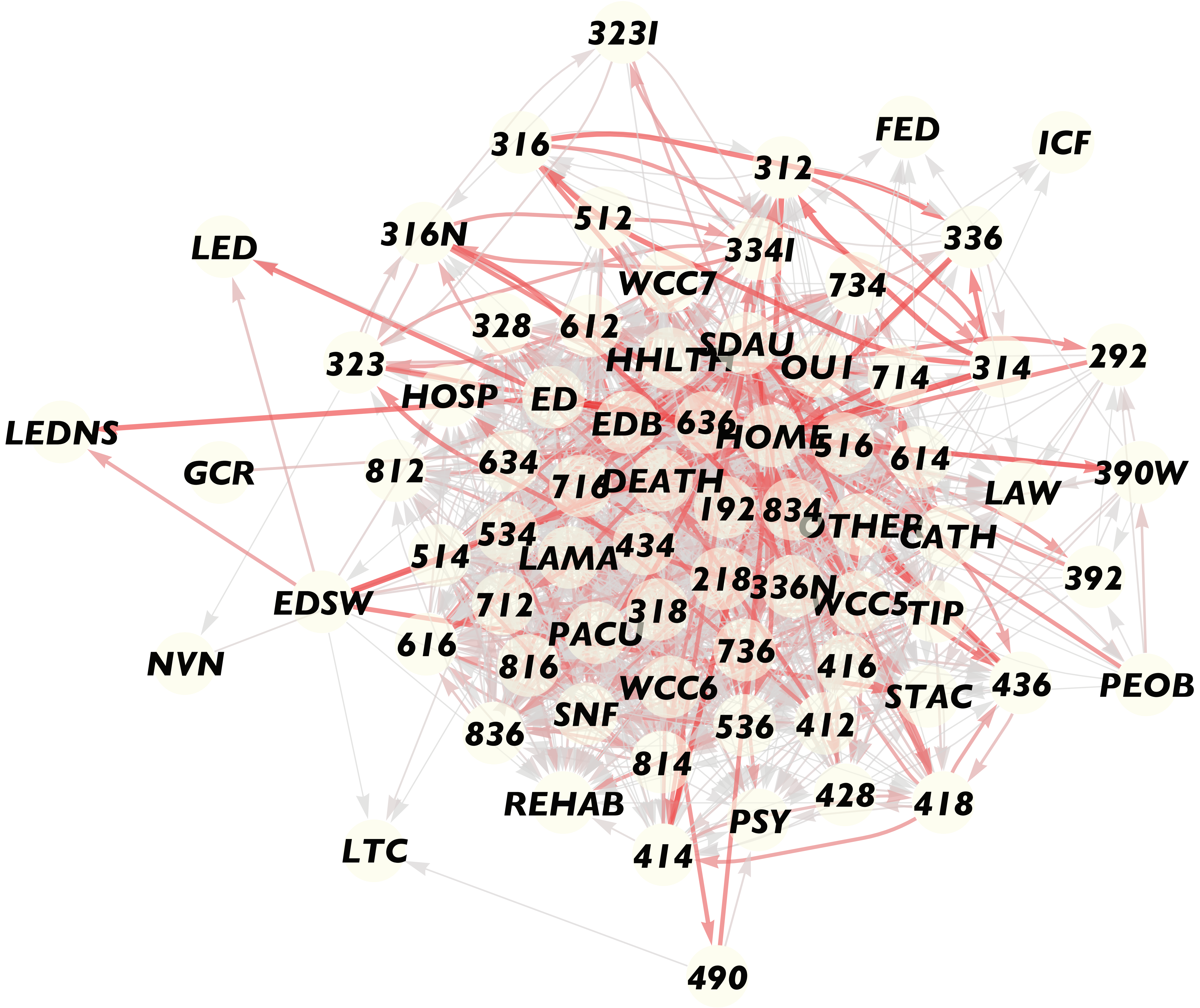
216.156.16.133

⇣

74.125.226.18



Trace-route Hospital Mapping



- Hospital unit linkage
  - by patient transfers
  - patient flow model
- Patients traverse this network
  - Multiple inter-ICU transfers
  - Implications for load balancing
  - Some units more "isolated" from hospital flow
- Hairball visualizations don't help

- **Visualization**
  - Adjacency matrix map
  - Topical clustering of units
- **Hospital Simulations/Modeling**
  - Models of patient flow
  - Disaster modeling
  - Infection control prediction
- **Compare hospitals by topology**
  - Like gene regulatory networks
  - Find functional modules
  - Academic vs. community, etc.

rchi.urmc.edu

Use Case 3: Mapping the US Healthcare System

Questions:

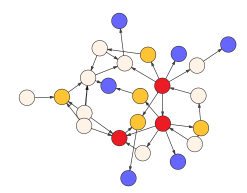
- How are healthcare providers linked by common patients?
- Can we model patient flow throught the US healthcare system?

Approach:

- Traceroute graphing of Medicare using outpatient claims traces

Application:

- Identifying self-organized provider teaming networks
- Improving outcomes for care of complex medical condition

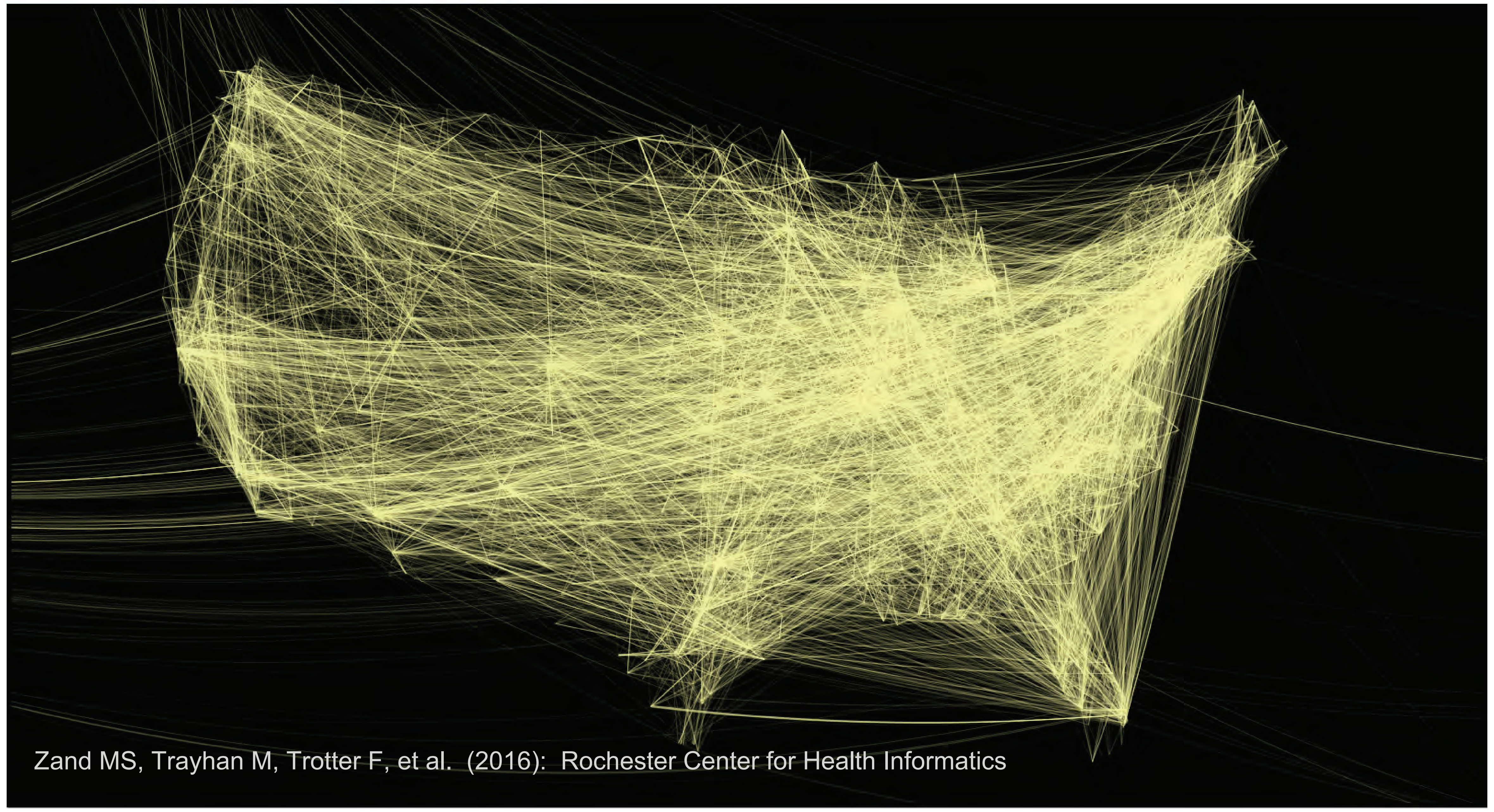


Mapping Medicare Properly

DocGraph: Open social doctor data

An inside look at DocGraph, a data project that shows how the U.S. health care system delivers care.

by Fred Trotter | @fredtrotter | +Fred Trotter | Comment | November 19, 2012

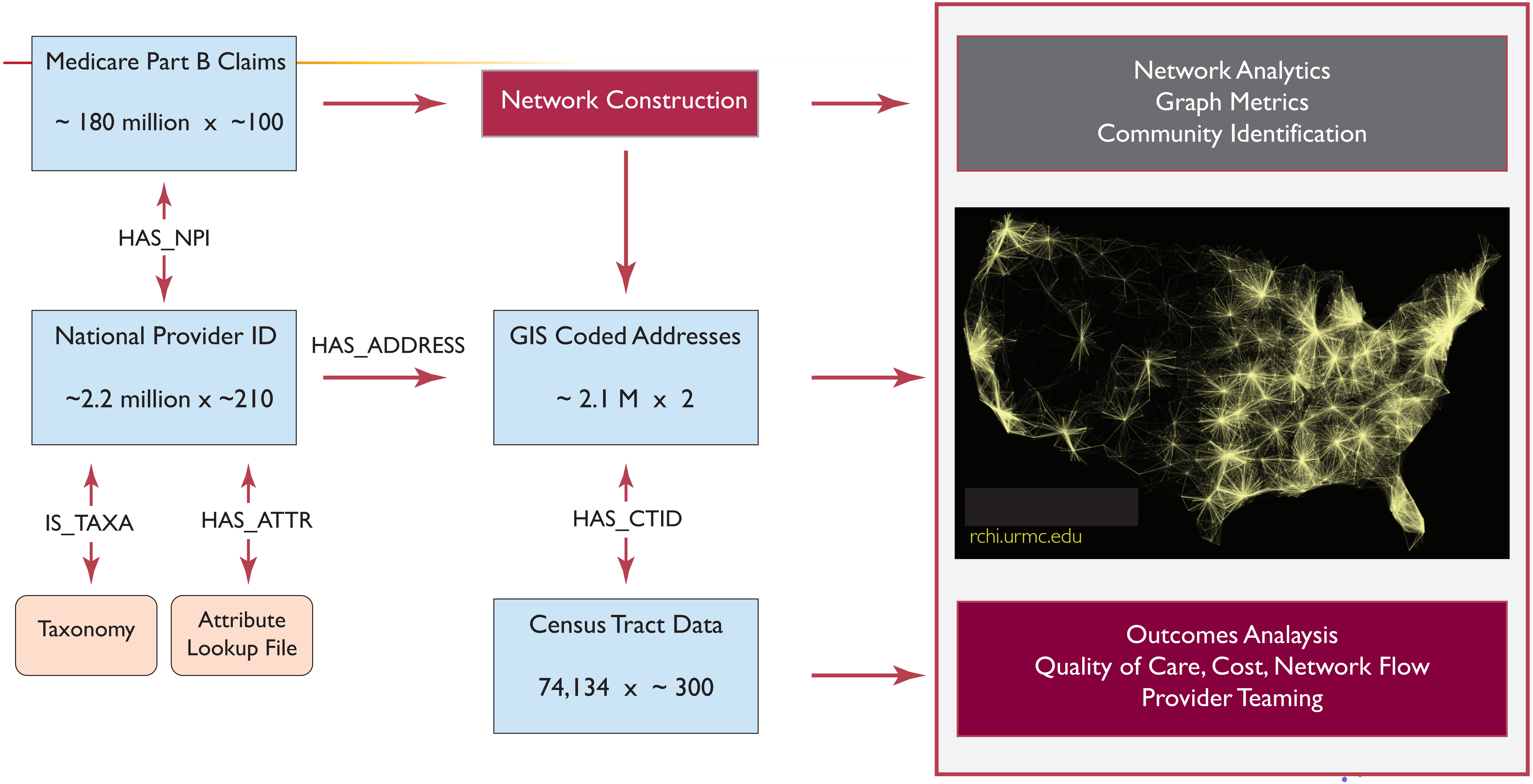


Zand MS, Trayhan M, Trotter F, et al. (2016): Rochester Center for Health Informatics

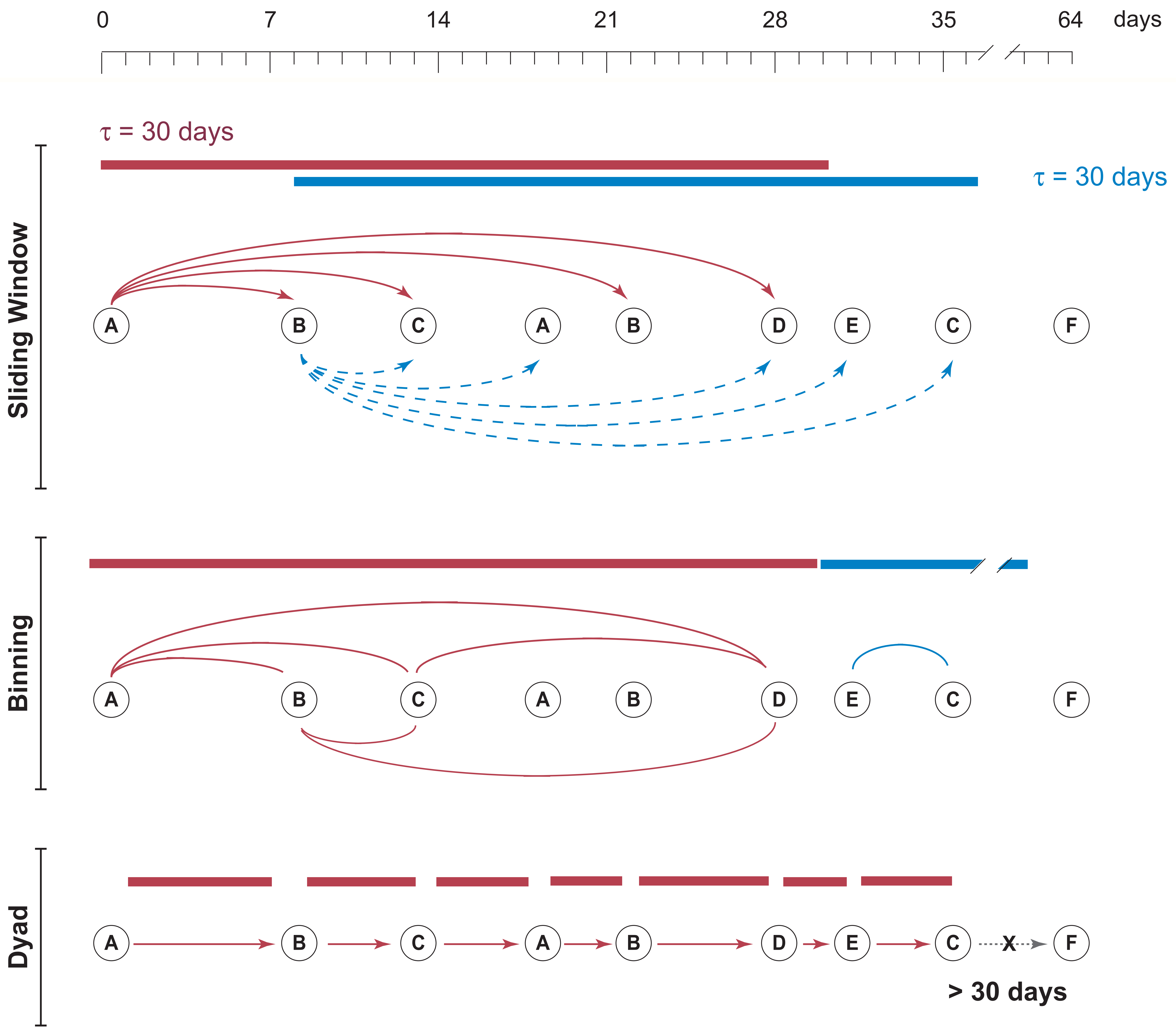
- Freedom of Information request to use Medicare claims data from 2009 to build networks
- Designed to study provider "referrals"
- Algorithm described, but actual code not released
- DocGraph 2009 =
  - 1.1 M vertices (provider + organization)
  - 93.7 million edges

Data Sets

- **Medicare Part B Claims Datasets** (dnav.cms.gov)
  - ~180 million annual claims x ~40M patients
  - 100+ data elements per claim
  - Data is temporal
  - Requires HIPAA compliant compute platform
- **National Provider Identifier Dataset** (download.cms.gov/nppes/NPI_Files.html)
  - ~22 million providers by ~210 data elements
  - Data is mostly static
- **Benchmark?**
  - Bipartite (provider → patient) to Unipartite (provider → provider) projection



Medicare Part B Claims
~ 180 million x ~100
Network Construction
Network Analytics
Graph Metrics
Community Identification
HAS_NPI
National Provider ID
~2.2 million x ~210
HAS_ADDRESS
GIS Coded Addresses
~ 2.1 M x 2
IS_TAXA
HAS_ATTR
HAS_CTID
Taxonomy
Attribute
Lookup File
Census Tract Data
74,134 x ~ 300
[redacted]
rchi.urmc.edu
Outcomes Analaysis
Quality of Care, Cost, Network Flow
Provider Teaming

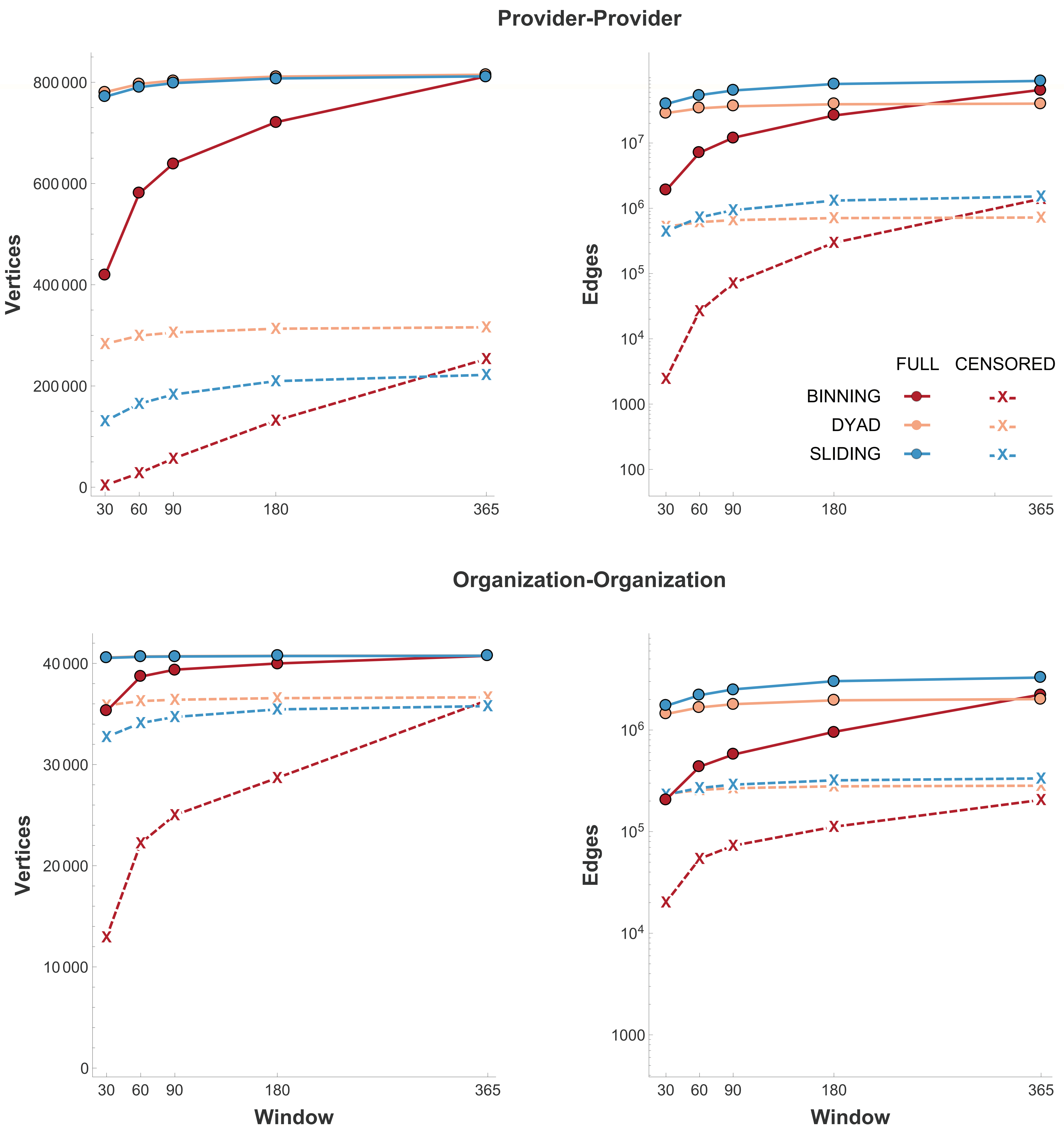


- **Algorithms**
  - Sort claims by patient
  - Next order claims temporally
  - Iterate through claims using a frame/window τ
- **Sliding Algorithm**
  - Similar to Medicare algorithm
  - Slides window between claims
- **Binning Algorithm**
  - Complete graphs within frame τ
- **Dyad Algorithm**
  - Enforces sequential edge order
  - Traceroute mapping

Table 1. Comparison of characteristics for patient co-care networks generated by different algorithms with τ =365 days

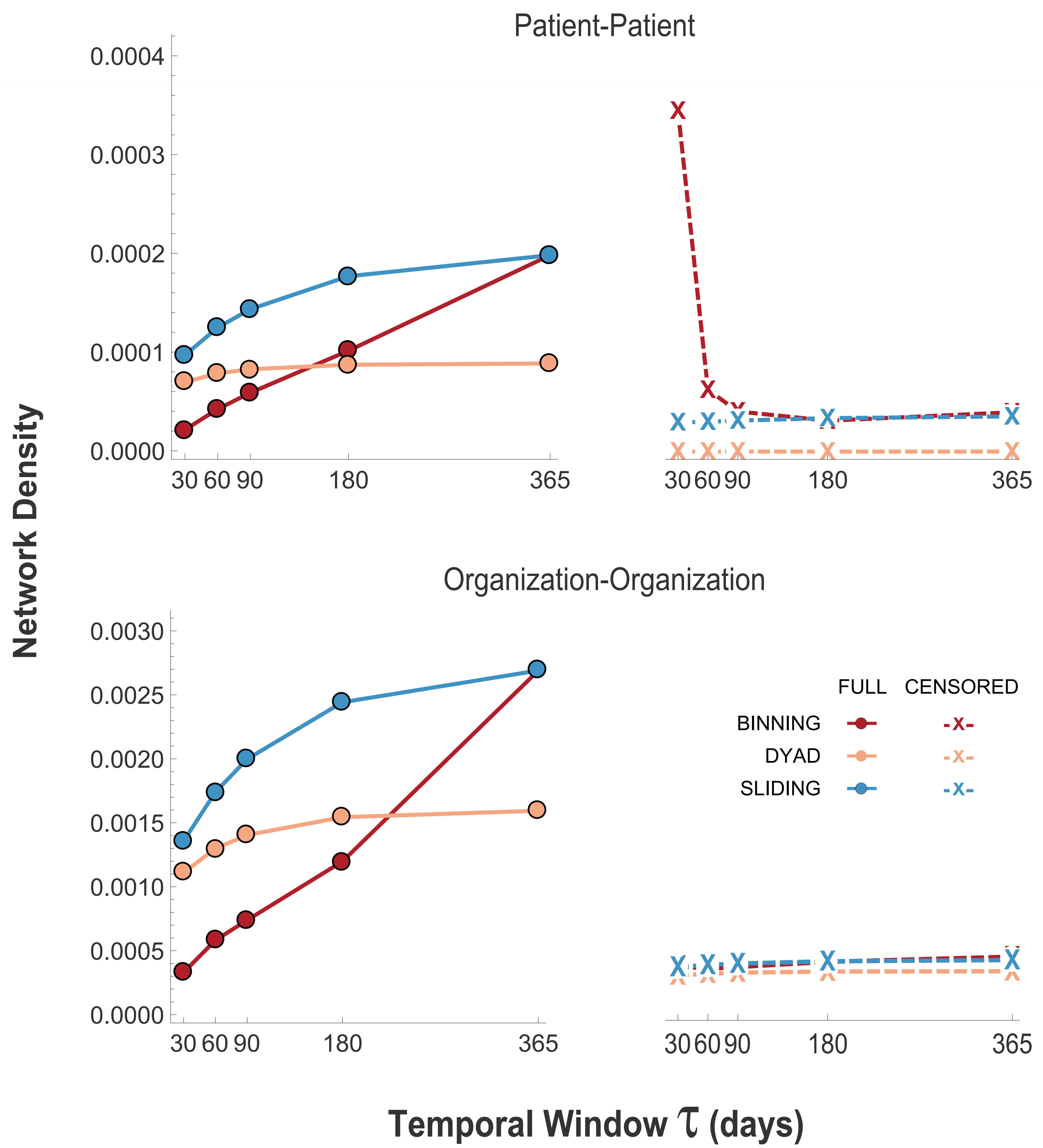
	Provider-Provider Networks			Organization-Organization Networks		
Metric	Sliding	Binning‡	Dyad	Sliding	Binning‡	Dyad
Edges (E)	89,377,290	65,287,590	40,077,297	3,282,133	2,233,601	2,014,859
Edge Type	Directed	Undirected	Directed	Directed	Undirected	Directed
Vertices (V)	811,784	811,784	814,917	40,749	40,749	40,768
E_{loop}/E †	-	-	0.411	-	-	0.122
V_{loop}/V †	-	-	0.938	-	-	0.943
d	51	29	89	6	13	10
r	0.05534	0.06985	0.02521	0.14768	0.16542	0.15915
ρ	0.56975	1.0	0.77929	0.86637	1.0	0.97295
C	0.28097	0.0	0.21598	0.53721	0.0	0.57809
D	0.00014	0.00010	0.00005	0.00198	0.00135	0.00119
lco	811,099	811,099	810,952	40,749	40,749	40,749
Max. V degree.	19,320	12,836	10,857	8,905	5,248	6,485
Mean V deg.	126.4	12.13	67.34	7.982	2.057	3.364
Max. E weight.	75,985	3,128	22,166	376,808	32,039	472,774
Mean. E weight.	7.982	2.057	3.364	126.4	12.13	67.34

d: network diameter, r: assortivity, ρ: reciprocity, C: mean clustering coefficient, D: network density, lco: number of nodes in the largest component. ‡Metrics for undirected graph, †Algorithm explicitly excludes self-loops.



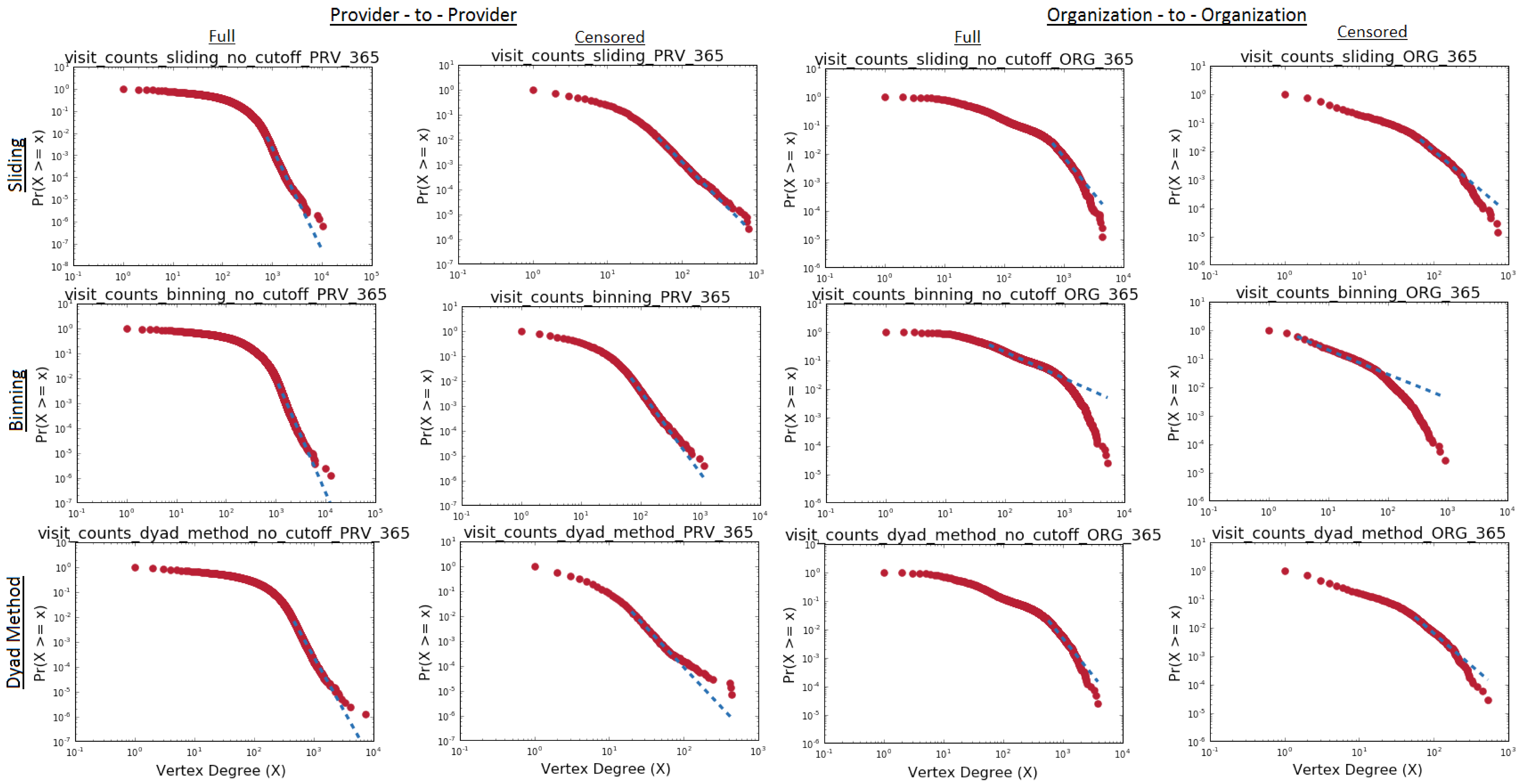
- Temporal windows
  - Capture network by τ=180 days
  - Binning method highly variable
- Censoring
  - Limit edge weights to > 10
  - Greatly decreases V, E
- Dyad and Sliding Algorithms
  - Very stable with respect to scope
- Organization analysis
  - Capture essentially same V

rchi.urmc.edu



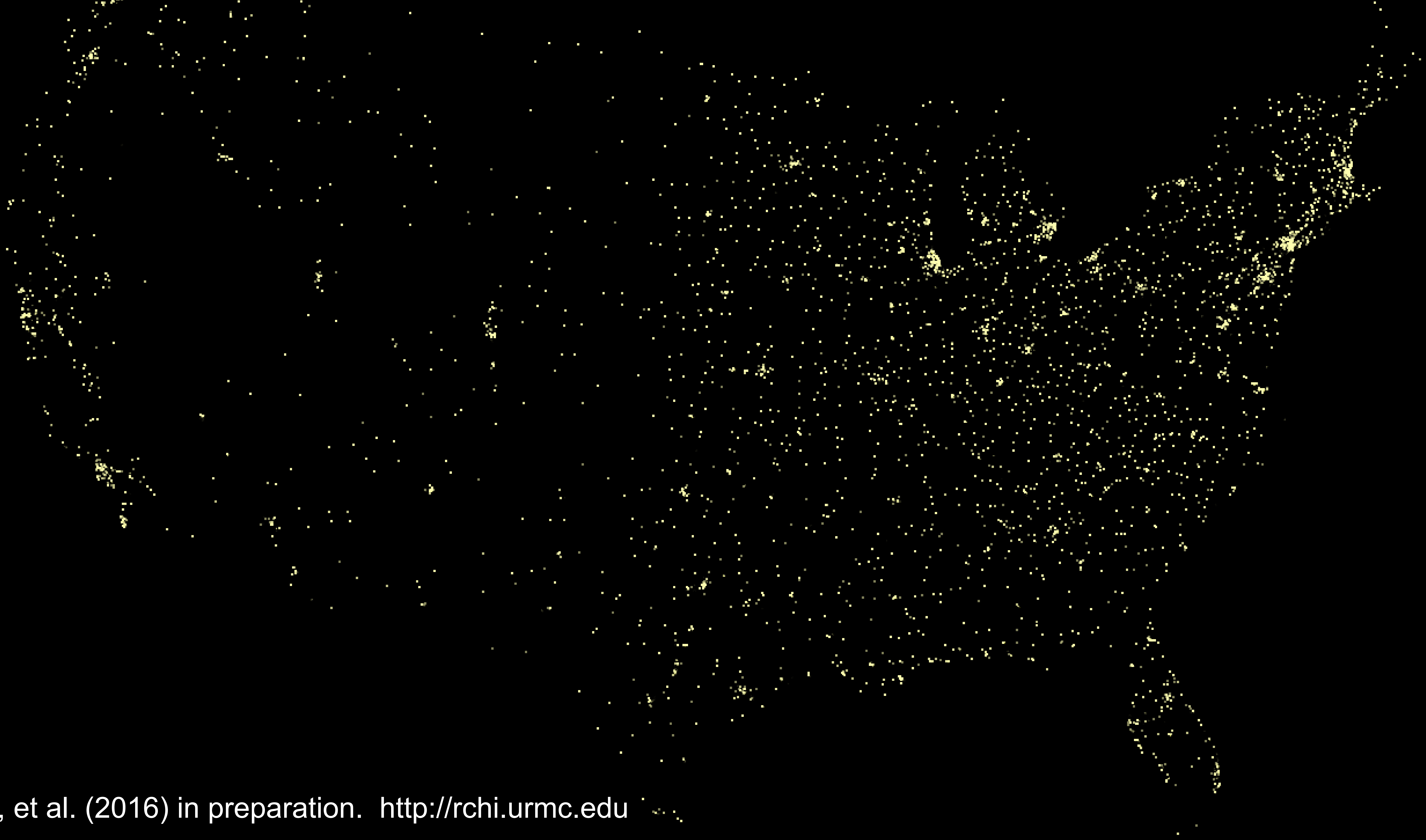
- **Network Density**
  - Capture network by τ=180 days with dyad and sliding window
  - Binning method highly variable
- **Censoring**
  - Greatly decreases density for all methods
- **Provider Analysis**
  - Much lower network densities
  - Regional effect?
- **Organization analysis**
  - Very dense networks (expected)

rchi.urmc.edu



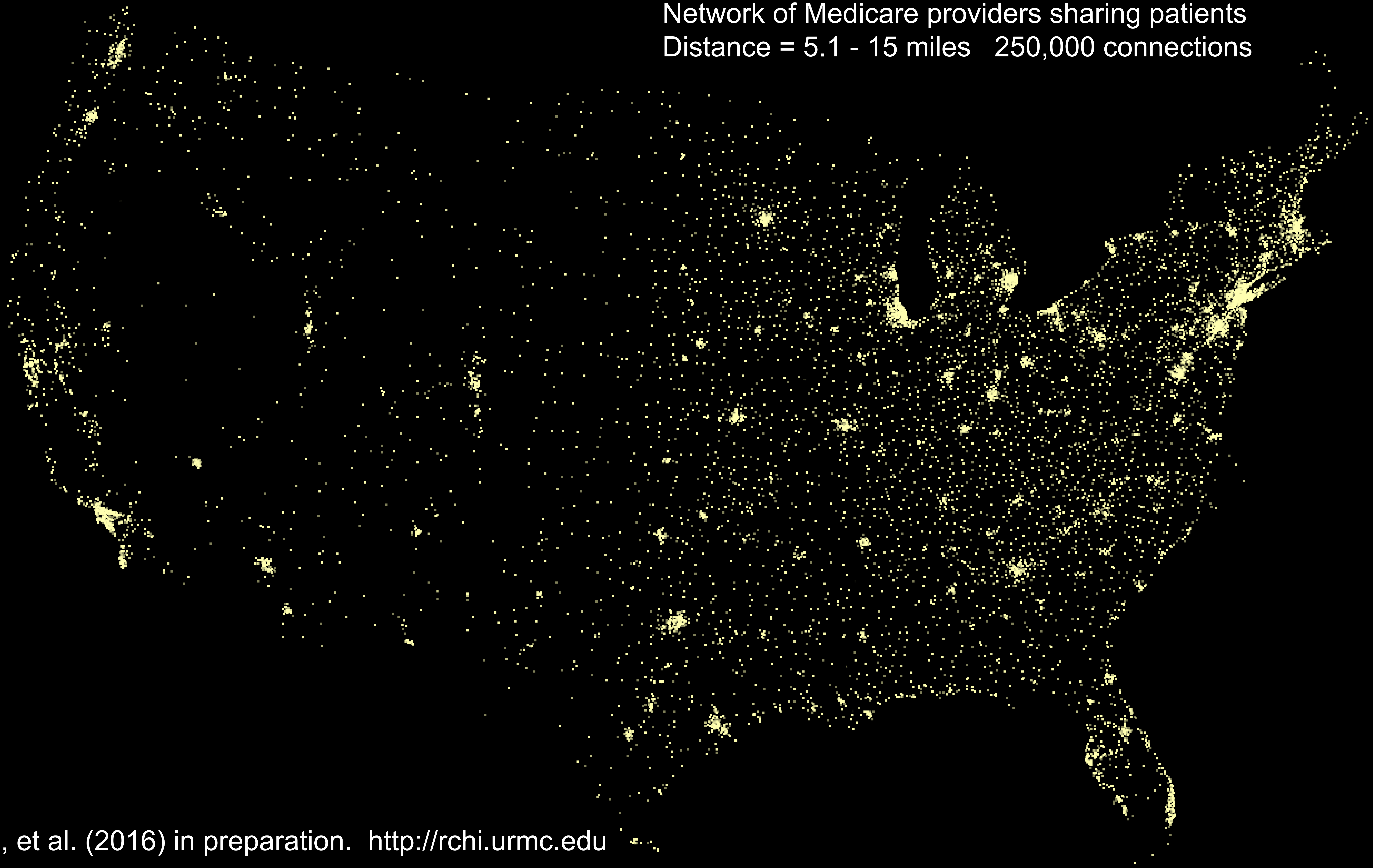
Provider - to - Provider
Organization - to - Organization
Full
Censored
Sliding
Binning
Dyad Method
visit_counts_sliding_no_cutoff_PRV_365
visit_counts_sliding_PRV_365
visit_counts_sliding_no_cutoff_ORG_365
visit_counts_sliding_ORG_365
visit_counts_binning_no_cutoff_PRV_365
visit_counts_binning_PRV_365
visit_counts_binning_no_cutoff_ORG_365
visit_counts_binning_ORG_365
visit_counts_dyad_method_no_cutoff_PRV_365
visit_counts_dyad_method_PRV_365
visit_counts_dyad_method_no_cutoff_ORG_365
visit_counts_dyad_method_ORG_365
Pr(X >= x)
Vertex Degree (X)

Network of Medicare providers sharing patients
Distance = 0.1 - 5 miles   250,000 connections



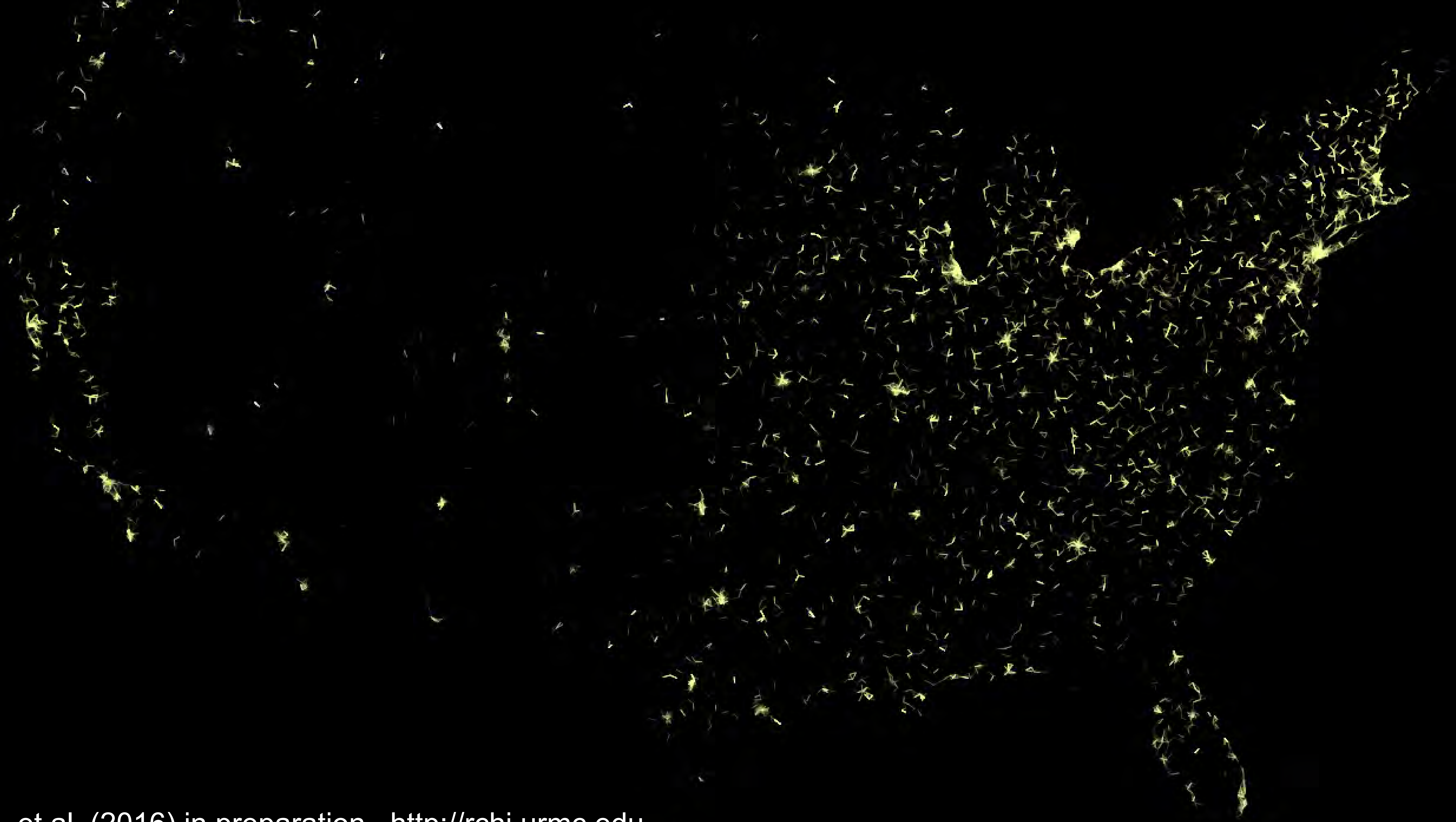
MS Zand, et al. (2016) in preparation. http://rcchi.urmc.edu

Network of Medicare providers sharing patients
Distance = 5.1 - 15 miles   250,000 connections



MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu

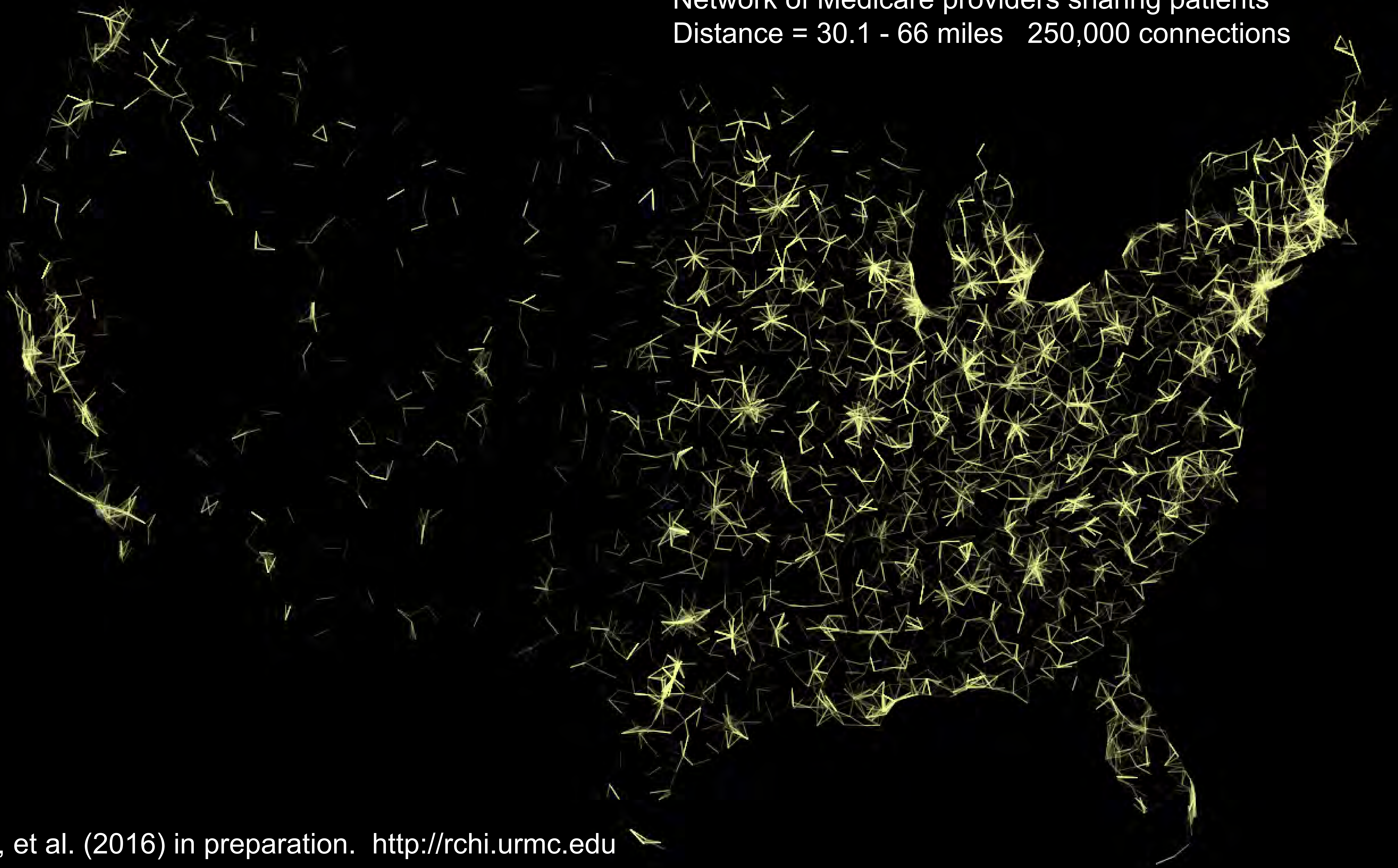
Network of Medicare providers sharing patients
Distance = 15.1 - 22.6 miles   250,000 connections



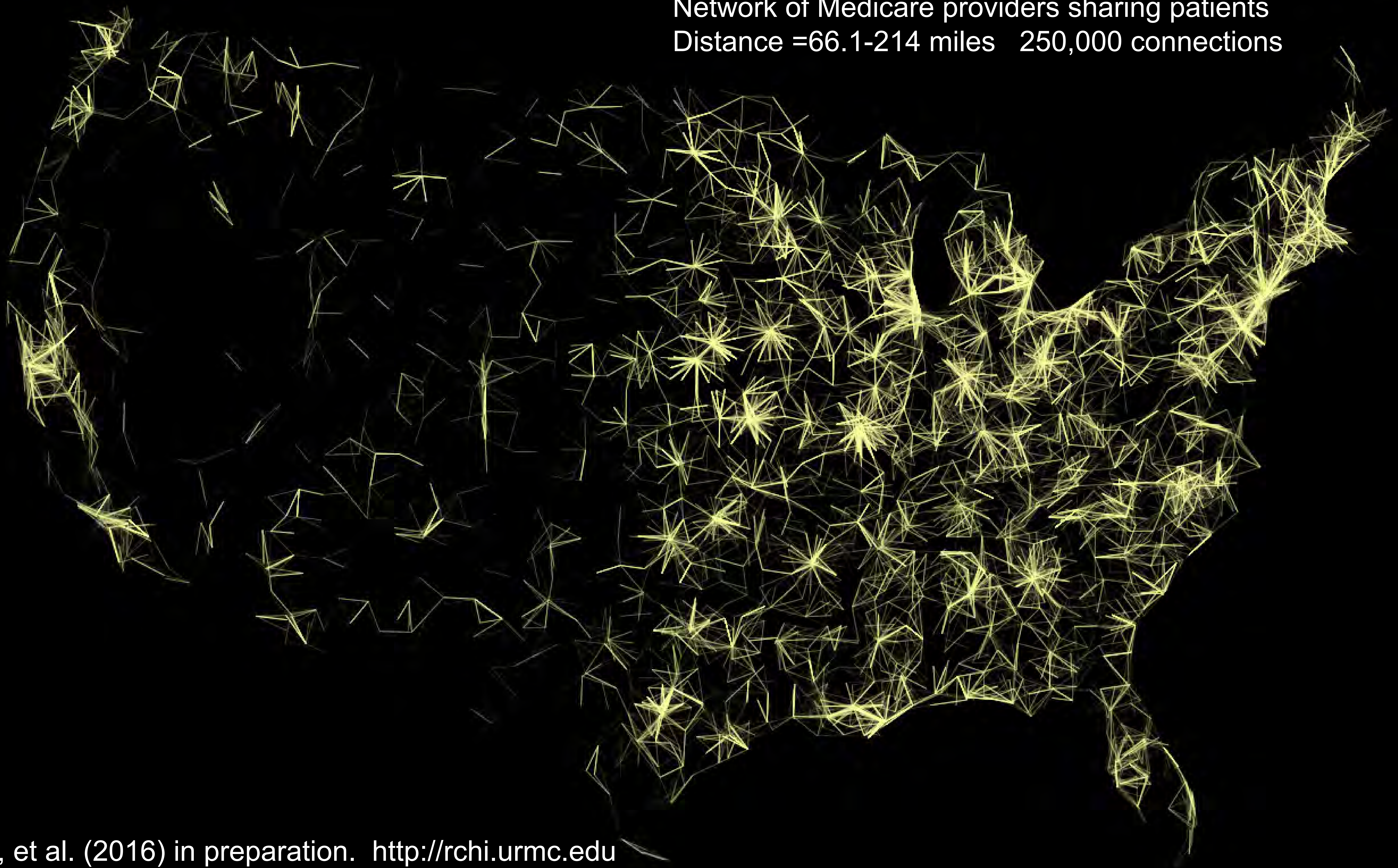
MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu


Network of Medicare providers sharing patients
Distance = 22.7 - 30 miles   250,000 connections
MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu

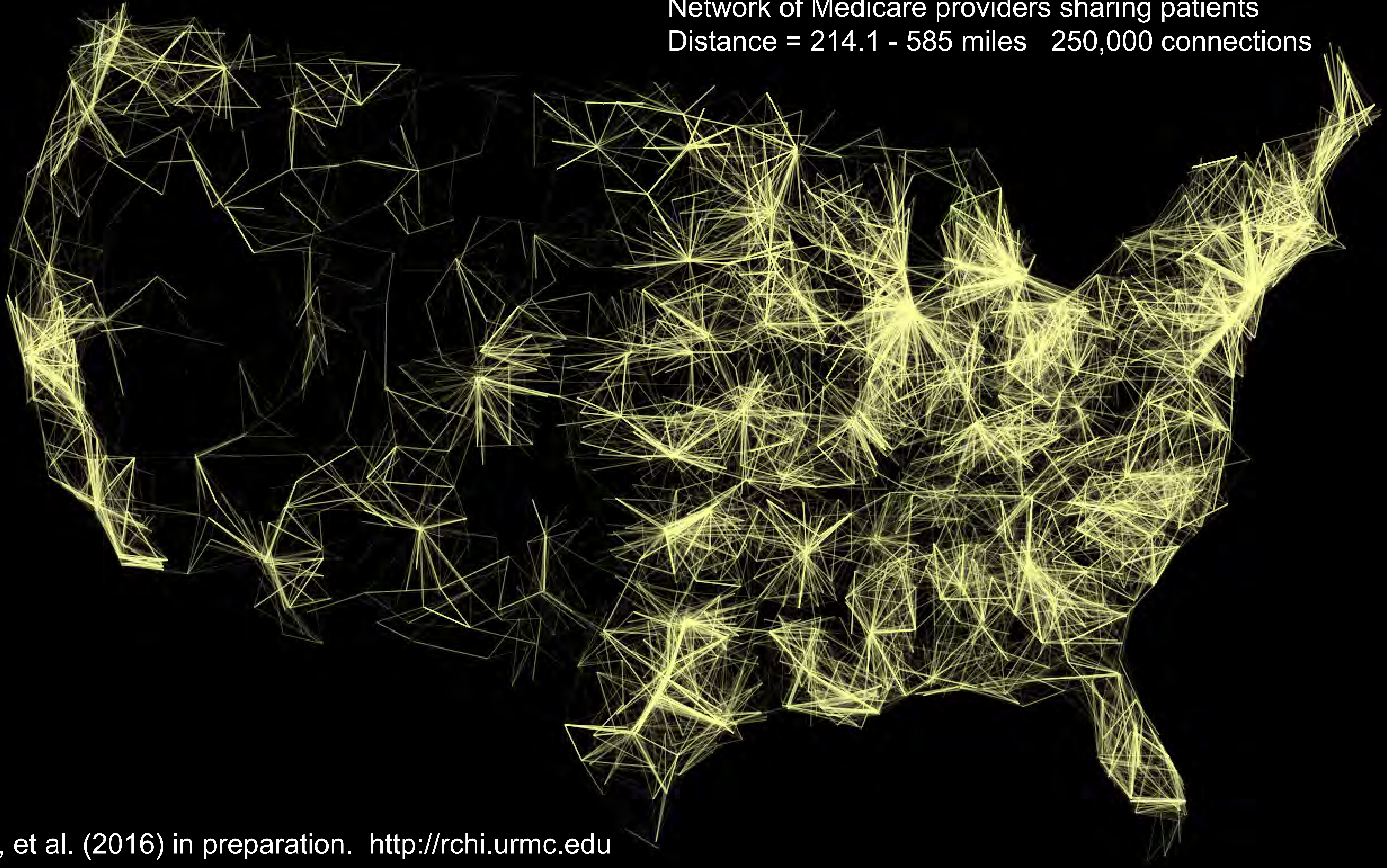
Network of Medicare providers sharing patients
Distance = 30.1 - 66 miles   250,000 connections



MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu

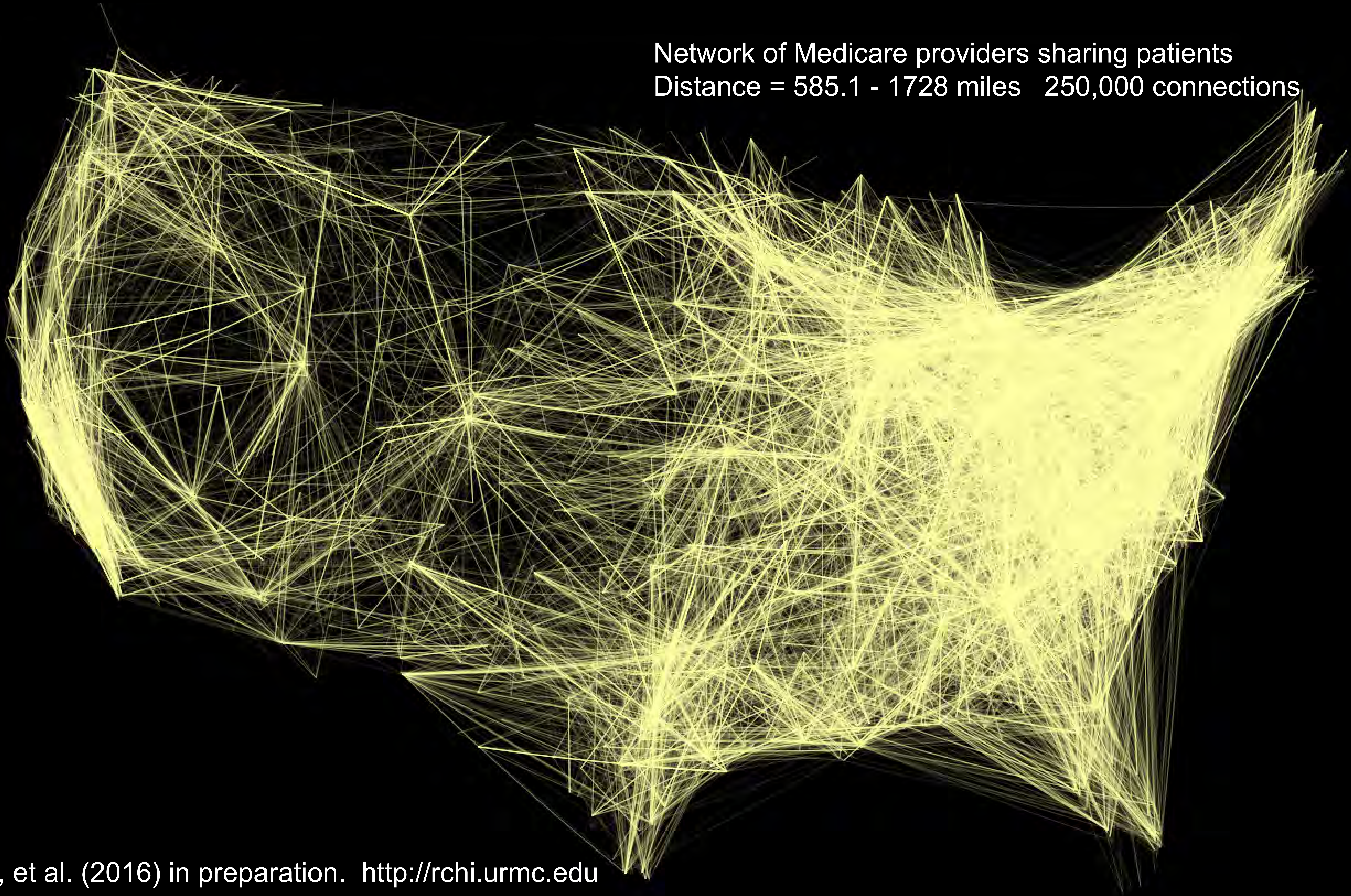

Network of Medicare providers sharing patients
Distance =66.1-214 miles   250,000 connections
MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu

Network of Medicare providers sharing patients
Distance = 214.1 - 585 miles   250,000 connections



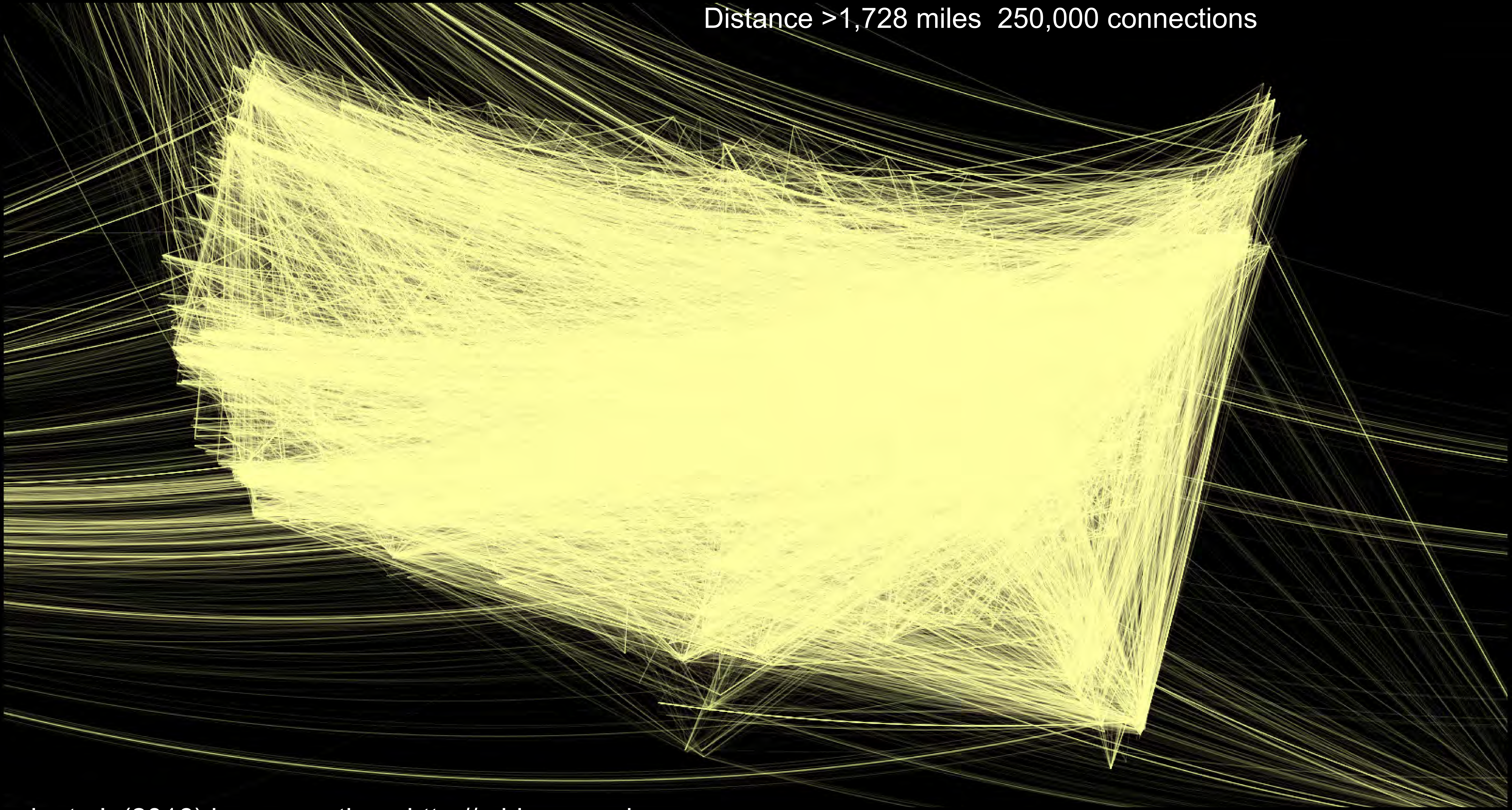
MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu

Network of Medicare providers sharing patients
Distance = 585.1 - 1728 miles   250,000 connections

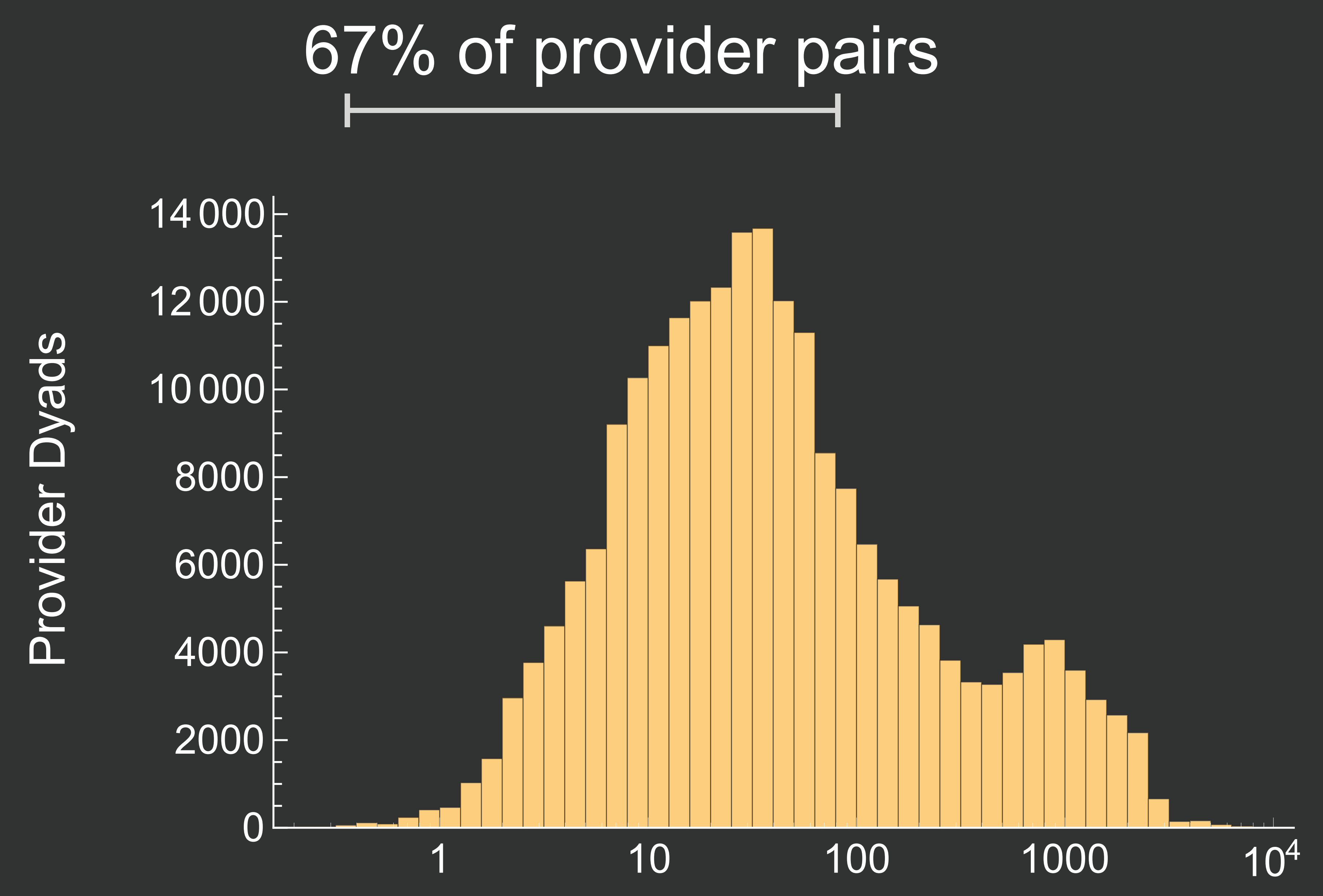


MS Zand, et al. (2016) in preparation. http://rcHi.urmc.edu

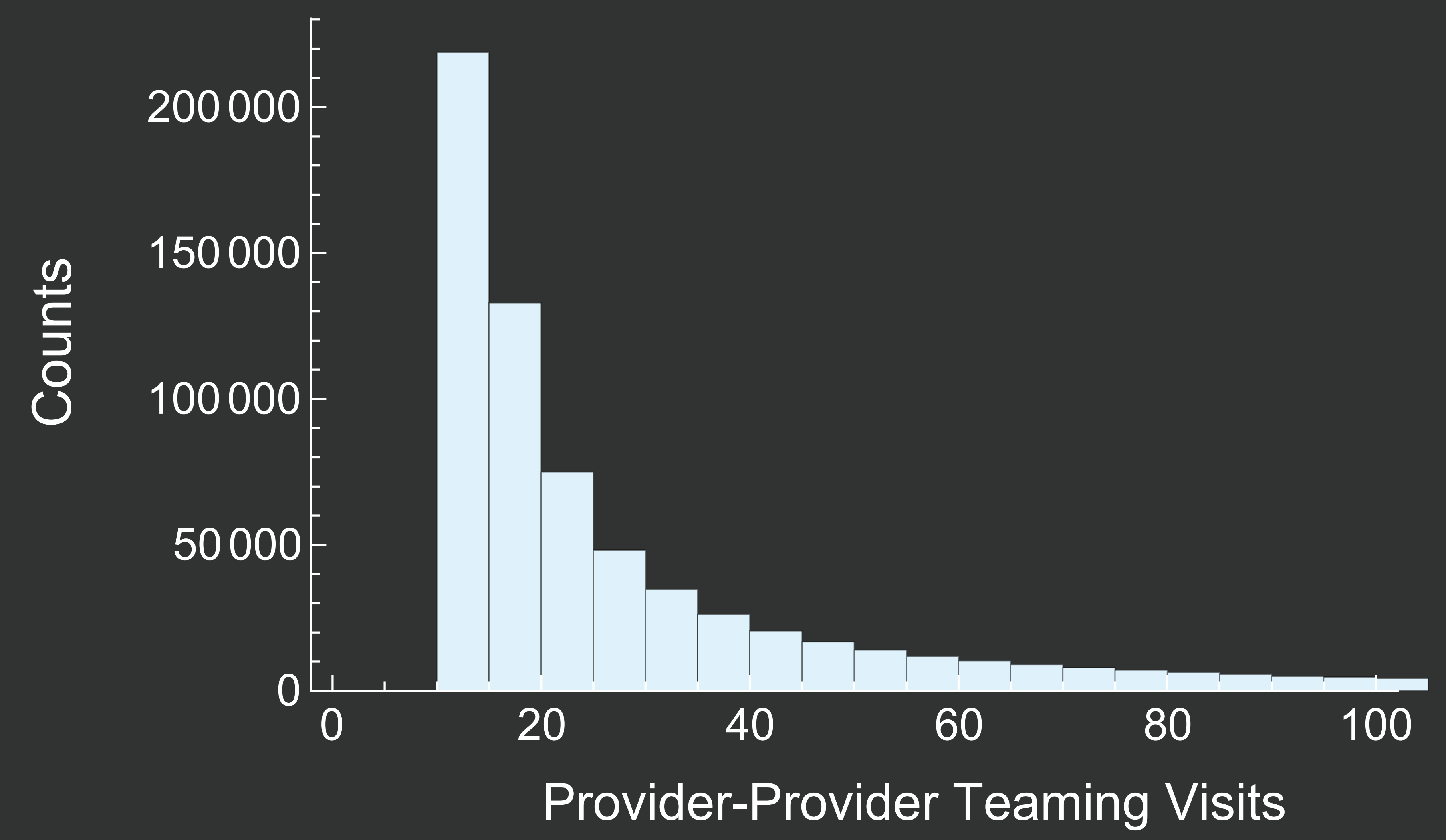
Network of Medicare providers sharing patients
Distance >1,728 miles 250,000 connections

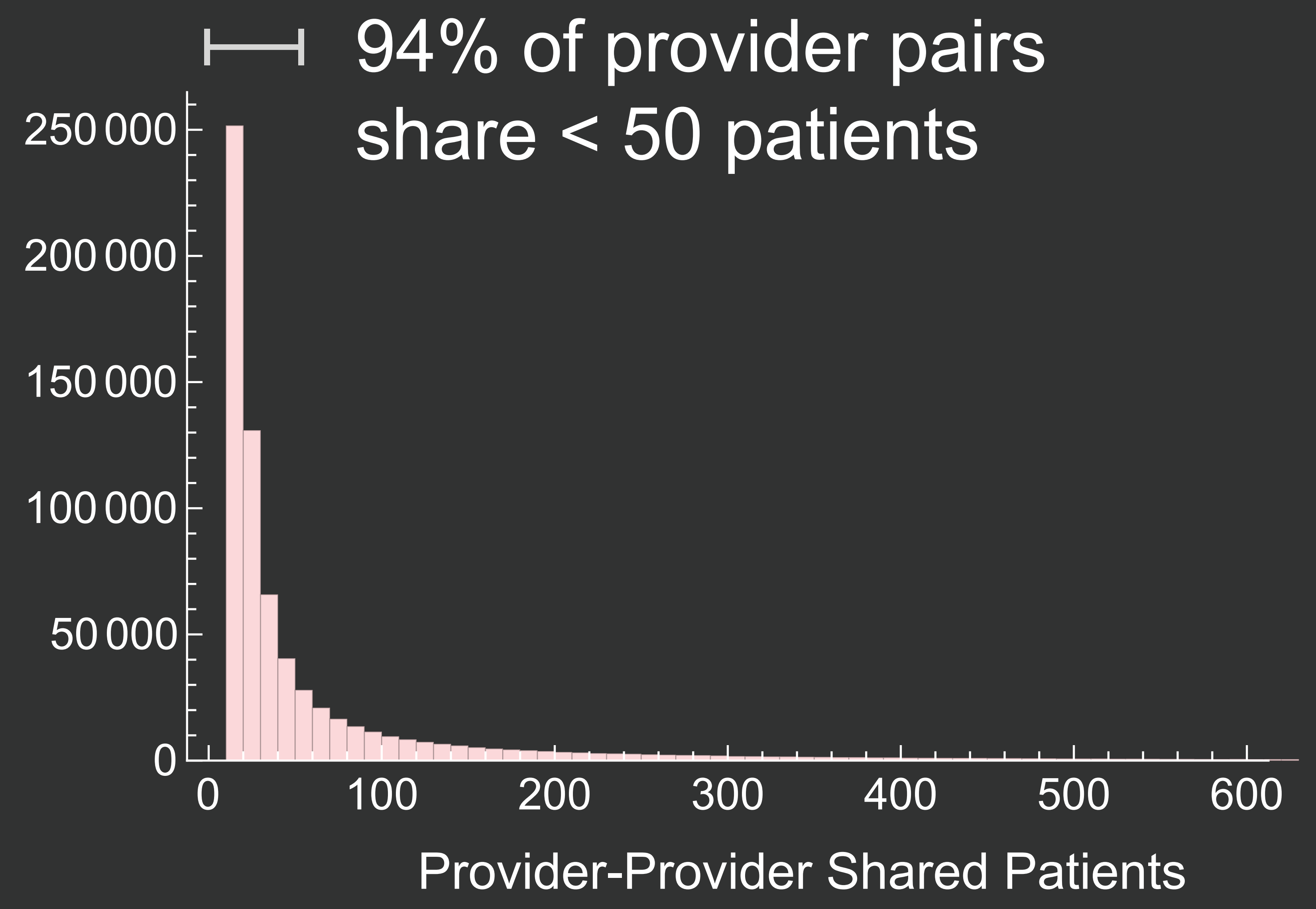


MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu



- Analysis using the “dyad counting” method
- >50% of provider pairs are within 40 miles of each other
- Dyad counting method with 11 visit threshold also does not capture multiple visits to the same provider
- Primary care providers now have highest netowrk centrality with this method



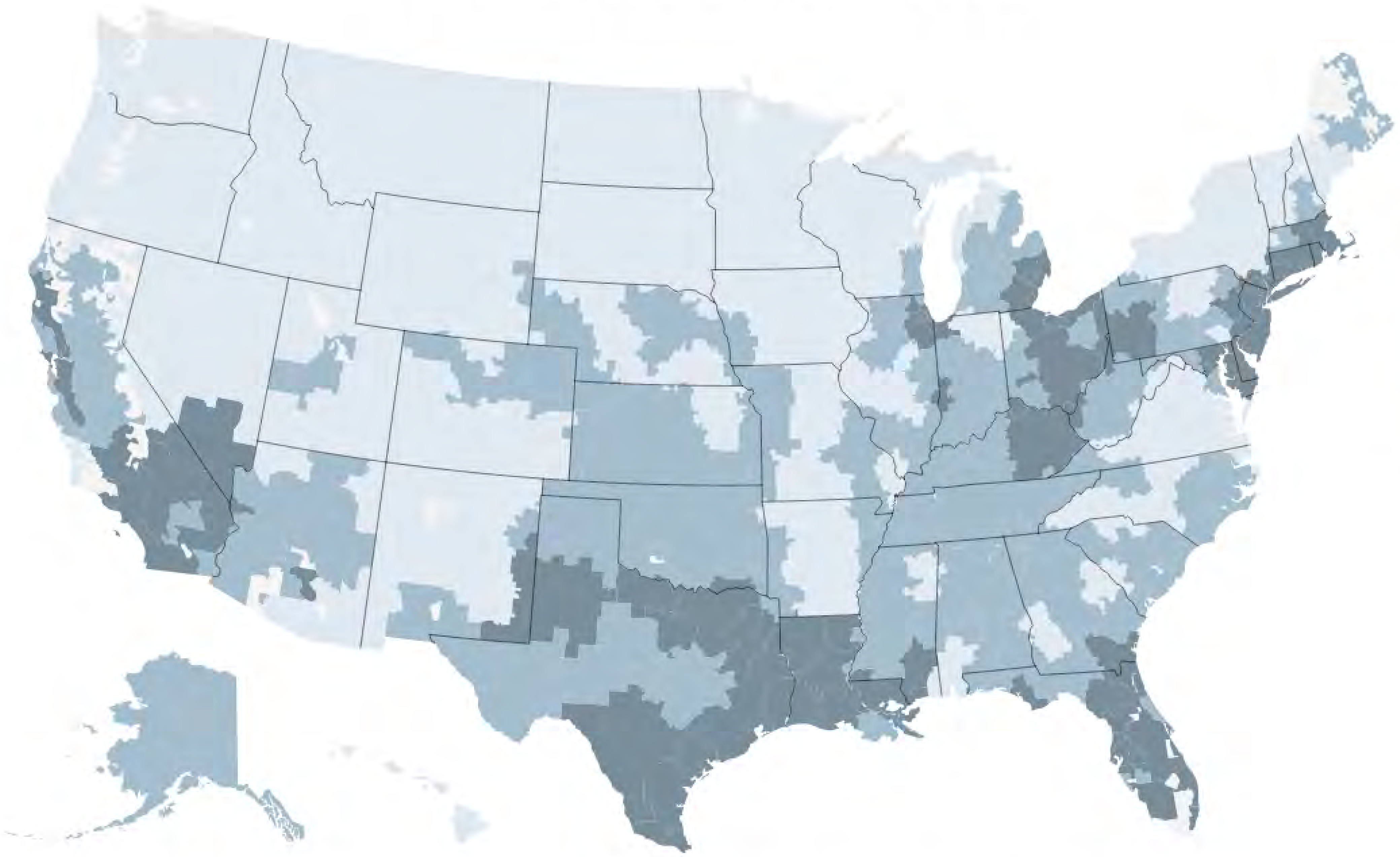


MS Zand, et al. (2016) in preparation. http://rchi.urmc.edu

Transformation Opportunity: Population Health, Cost, Outcomes

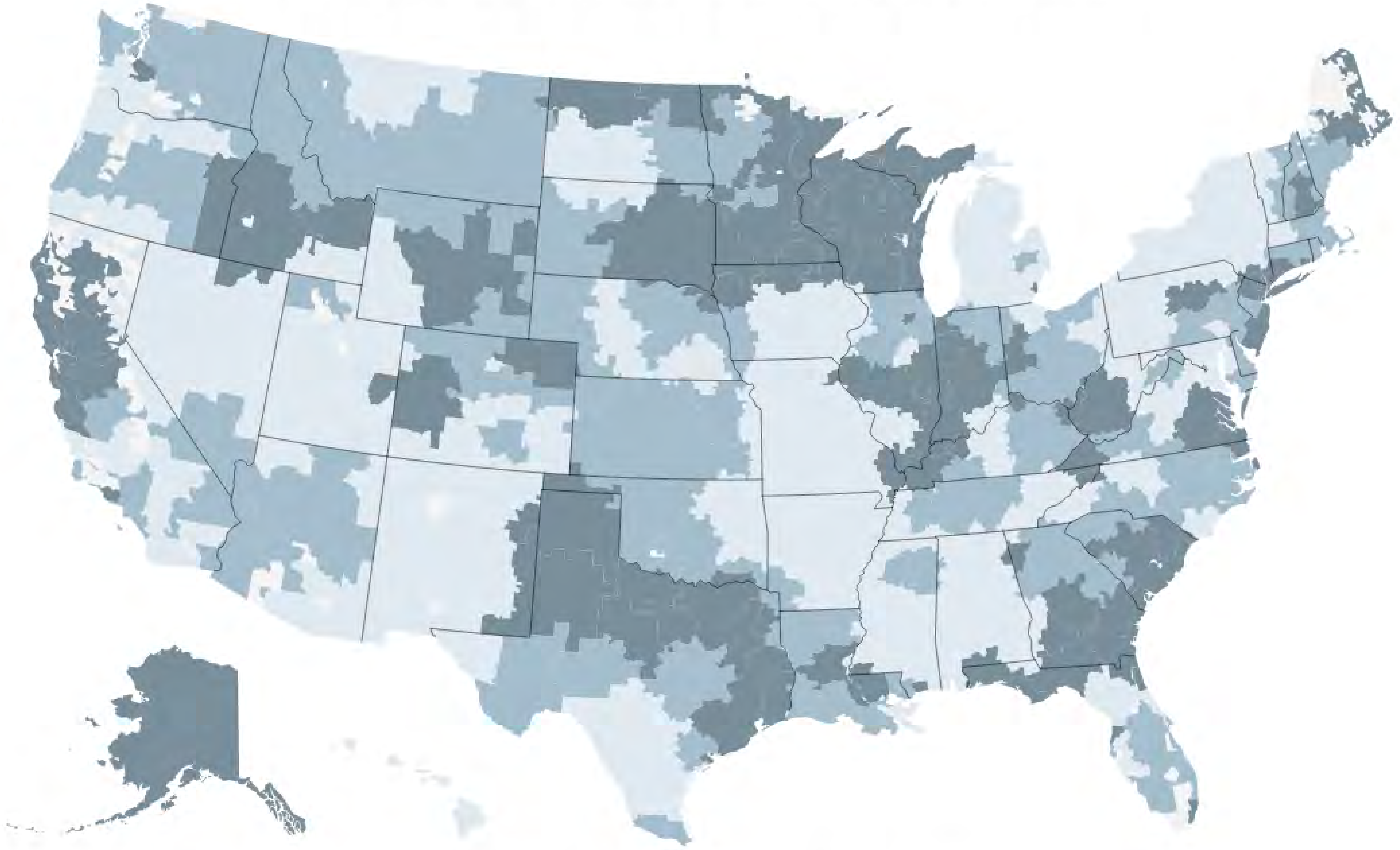
These maps look nothing alike. Their big differences are forcing health experts to rethink what they know about health costs in Rochester, N.Y. and across the country.

Medicare spending per capita



A lot of what we know about health care costs comes from Medicare data.

Private insurance spending per capita



But a new study suggests that places spending less on Medicare do not necessarily spend less on health care over all.

rchi.urmc.edu

Graph analytics useful for healthcare dataset analysis

- Community identification algorithms
  - Self-organized provider cliques, clans, and communities
  - Patient cohort identification
  - Geospatial, social, and "non-medical" edges
- Network construction algorithms
  - Multipartite and lower dimensional projections
  - Routines for constructing time-varying graphs
  - Clustering and cohort identification
  - Hypergraphs

Software and hardware we use

Oracle Labs
PGX

Wolfram Mathematica 10

Gephi

python

ANACONDA

University of Rochester HPC:
178 compute nodes, each with:
2 Intel Xeon E5-2695 v2 processors (12 core)

UNIVERSITY of ROCHESTER MEDICAL CENTER

RCHI ROCHESTER CENTER FOR HEALTH INFORMATICS