Simulation of network representations that answer to different categories of network characteristics

Modeling and Simulation Methods

Nikolas Begetis, ITMB, Feb2017

Categories

Scale-freeNetworks

- Degree distribution follows a power law:
  - $P(k) \sim k^{-\gamma}$, where γ is a constant
- Biological (and social) networks tend to be scale-free

Small World Networks

- Any two arbitrary nodes are connected by a small number of intermediate edges
- The network has an average shortest path length much smaller than the # of nodes in network

Random Networks

- Created by algorithms
- Can be
  - Flat random network Erdos – Renyi
  - High Clusterning coefficient: Watts-Strogatz
  - Scale-Free: Barabasi-Albert
- Nodes have similar degrees
- Useful to compare with regular

Characteristics

Network Measures

- Node degree
  - Node indegree: # of edges for which the node is a target
  - Node outdegree: # of edges for which the node is a source
- Shortest Path Length
  - Shortest traversal distance between 2 nodes
  - Can be weighted if edges have weights
- Clustering Coefficient
  - Measures how close the neighbors of a node are to being a clique (fully connected group)
  - (# of edges connecting a node's neighbors) / (the node's degree)

Centrality

- Measures nodes' importance
- Degree centrality
  - Degree of this node / (# of nodes - 1)
- Betweeness Centrality
  - The average number of shortest paths that go through this node
- Closeness Centrality
  - The sum of all shortest paths between this node and all other nodes / (# of nodes - 1)

Case Study

Subject

- Airplanes – Air Routes
- Modeling the spread of a pathogen
  - Certain amount of time before they recover
- A directed network
  - Airports as vertices
  - Routes as edges
- Multiple strategies - most appropriate # of flight cancelation
  - Edge betweenness
  - centrality
  - clustering coefficient

Subject's network

- Small-world network
  - power-law decaying degree distribution
  - high average clustering coefficient
- Simulations of the spread of a disease and flight cancellation strategies based on
  - betweenness centrality
  - clustering coefficient-based

Network & Data

NetworkX Generator

- Python language software package
  - creation
  - manipulation
  - study of the structure, dynamics, and function of complex networks.
- Load and store networks
  - standard and nonstandard data formats
- Generate many types of random and classic networks
- analyze network structure
- build network models
- design new network algorithms
- draw networks

NetworkX Audience

Who uses NetworkX?

- Mathematicians, physicists, biologists, computer scientists, and social scientists.
- Good reviews of the state-of-the-art in the science of complex networks in:
  - Albert and Barabási,
  - Newman
  - Dorogovtsev and Mendes
- Links:
  - https://networkx.github.io/documentation/networkx-1.10/overview.html
  - https://github.com/networkx/networkx/

Simulations

- Code in
  - Python
  - NetworkX
- Network
  - airports as nodes
  - airline routes as edges
- Data
  - Openflights: openflights.org
  - Nodes without inbound or outbound edges were removed.
  - Redundant edges were removed.

Data Distribution

Data: Distribution Degree

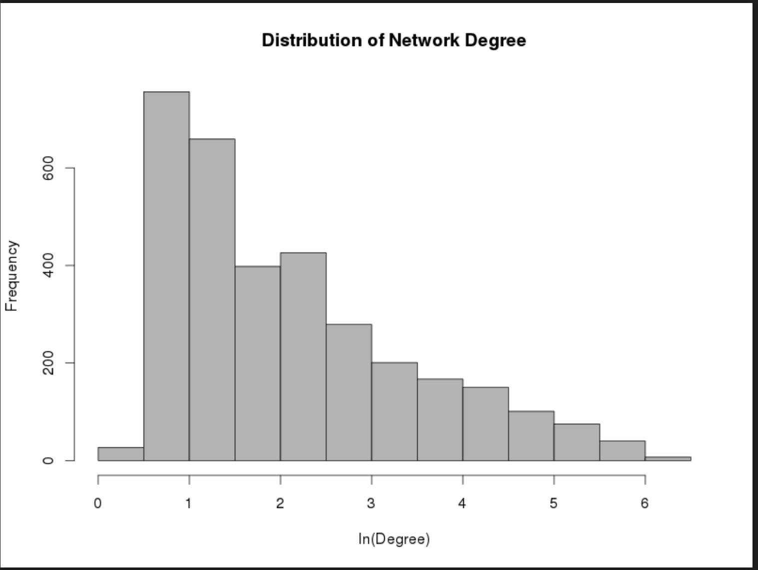


Fig.1

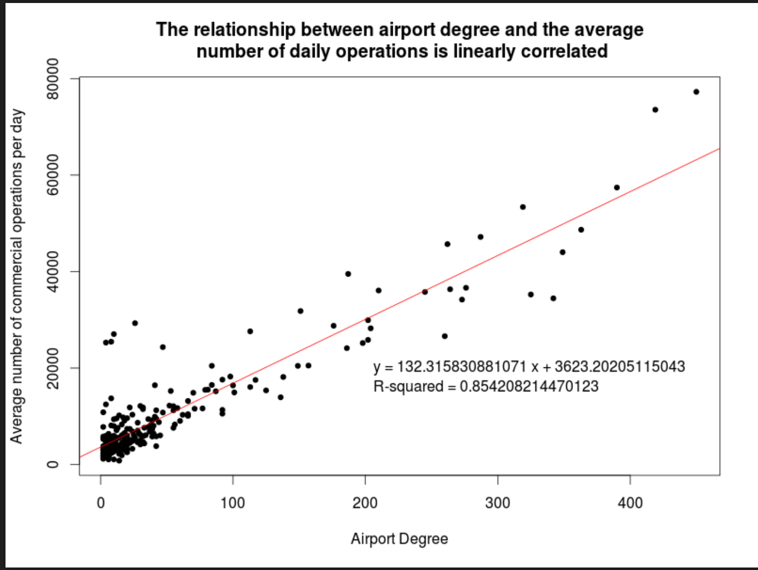


Fig.2

- 39,467 edges, 3,308 vertices
- Power Low (Fig. 1) – Degree 0-3: ~2500 nodes
- majority of airports: 2 - 20 inbound/outbound flights (Fig. 2)
- Few airports: 300 - 400 inbound/outbound flights (Fig. 2)

Data: Distribution Degree

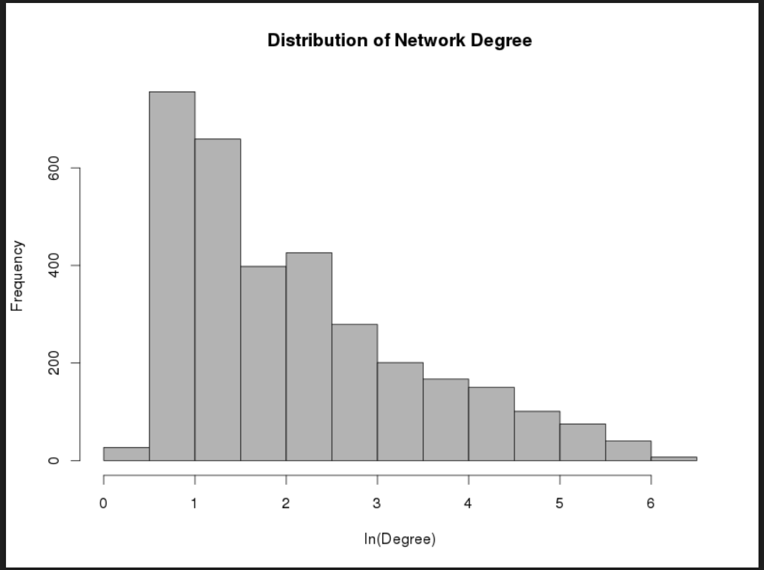


Fig.1

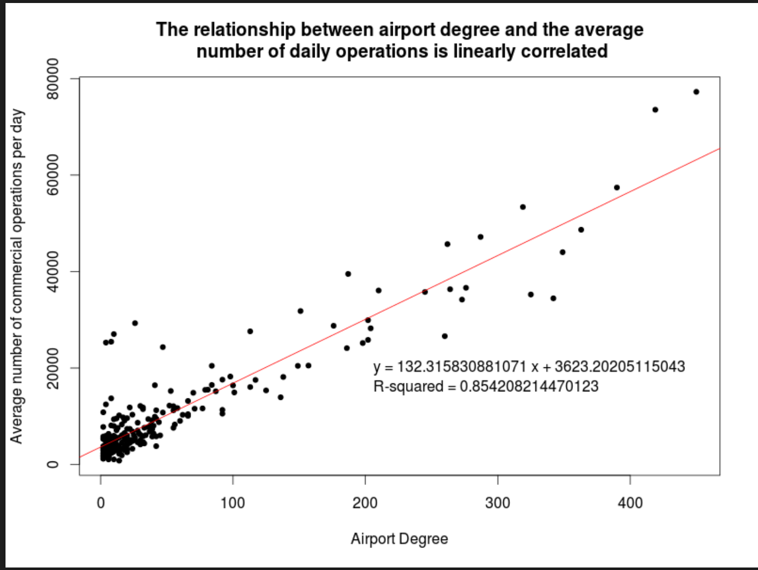
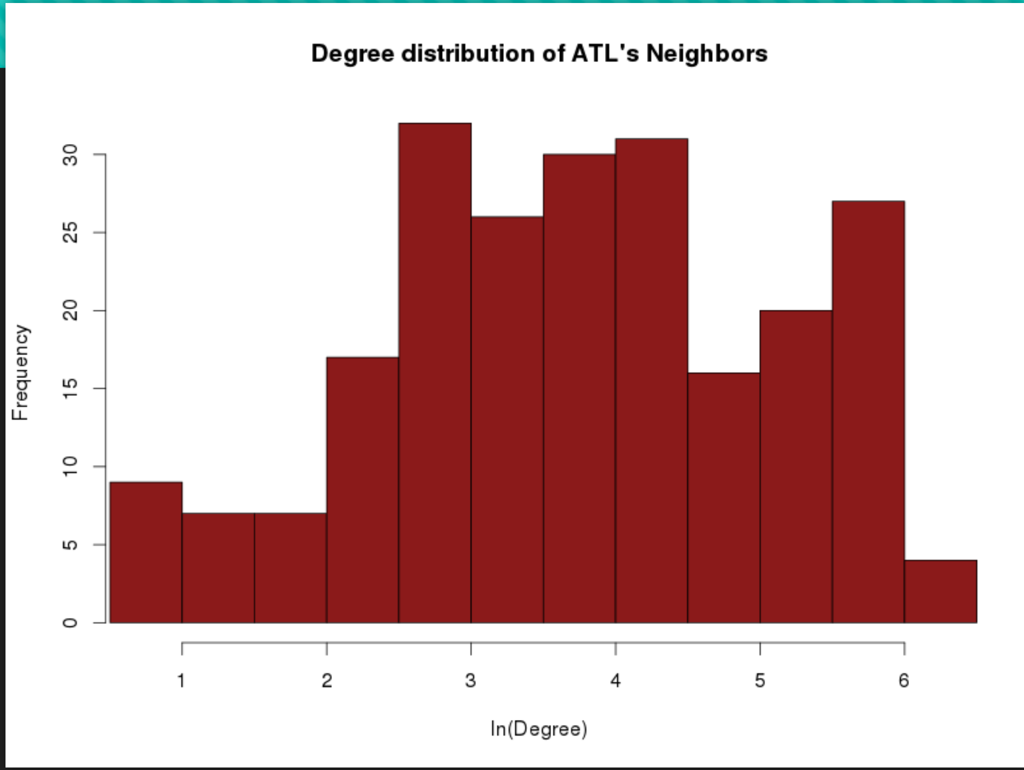


Fig.2

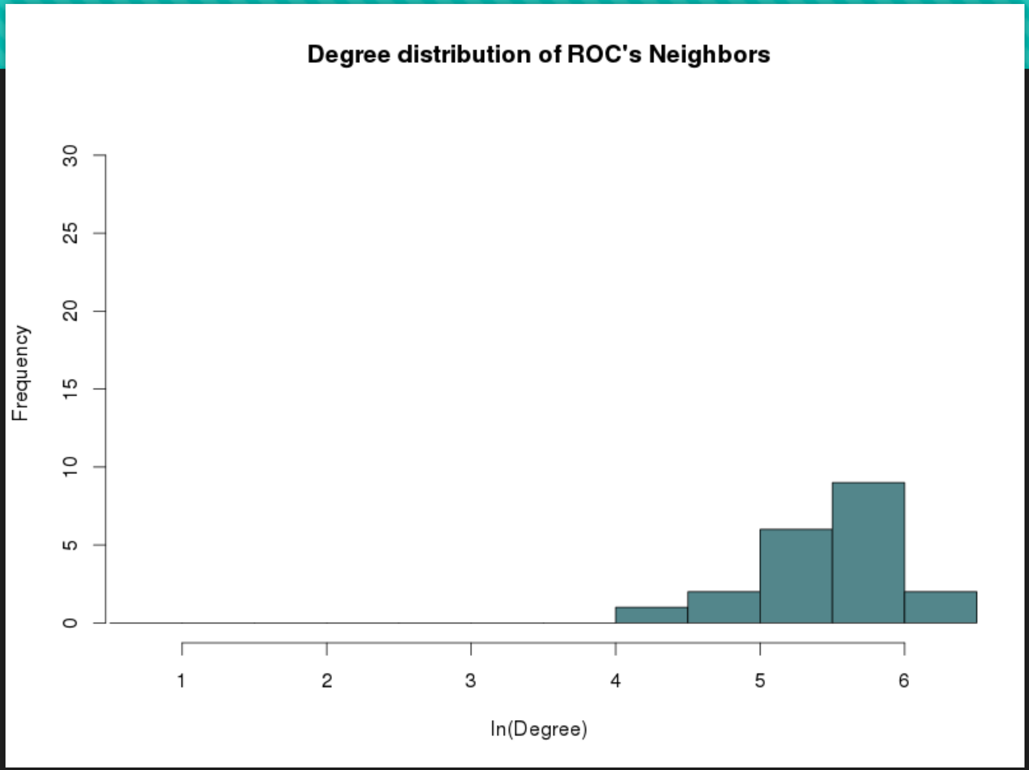
- Fig.1 and Fig.2:
  - International hubs carrying traffic to smaller regional airports

Data: High Degree Airport



- Hartfield-Jackson Atlanta International Airport (ATL) – 230 Flights / Airport Degree
- Wide variety of airports
  - low-degree airports (regional airports)
  - high-degree airports (international hubs)

Data: Low Degree Airport



- Greater Rochester International Airport (ROC) – 22 Flights / Airport Degree
- Almost entirely to airports with a degrees between 200 and 500
  - high-degree airports (international hubs)

Methodology

Methodology: Rules

- Example disease
  - influenza A
- Four states:
  - Susceptible
  - Exposed
  - Infectious
  - Recovered
- 3 days to incubate
- 7 days to recover from
- Basic reproductive rate of 2.32

Methodology: Algorithm

- Starts randomly 10 airports
- In every loop
  - All nodes: one plane leaves on outbound route and arrives at its destination
- Airports are receivers and transmitters
  - Idea: Transient passengers are able to infect permanent employees of the airport, who are then able to pass the infection on to other individuals in the airport.
- Routes are directional
  - Each edge is differently: Probability of carrying infectious individuals based upon the degree of the source and destination airports.
- Edge weights are recalculated after each cancellation strategy
  - To properly model the flow of individuals around cancellations

Network Metrics

Network Metrics

- Betweenness centrality

$$b(e) = \sum_{i \neq e \neq j} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$$

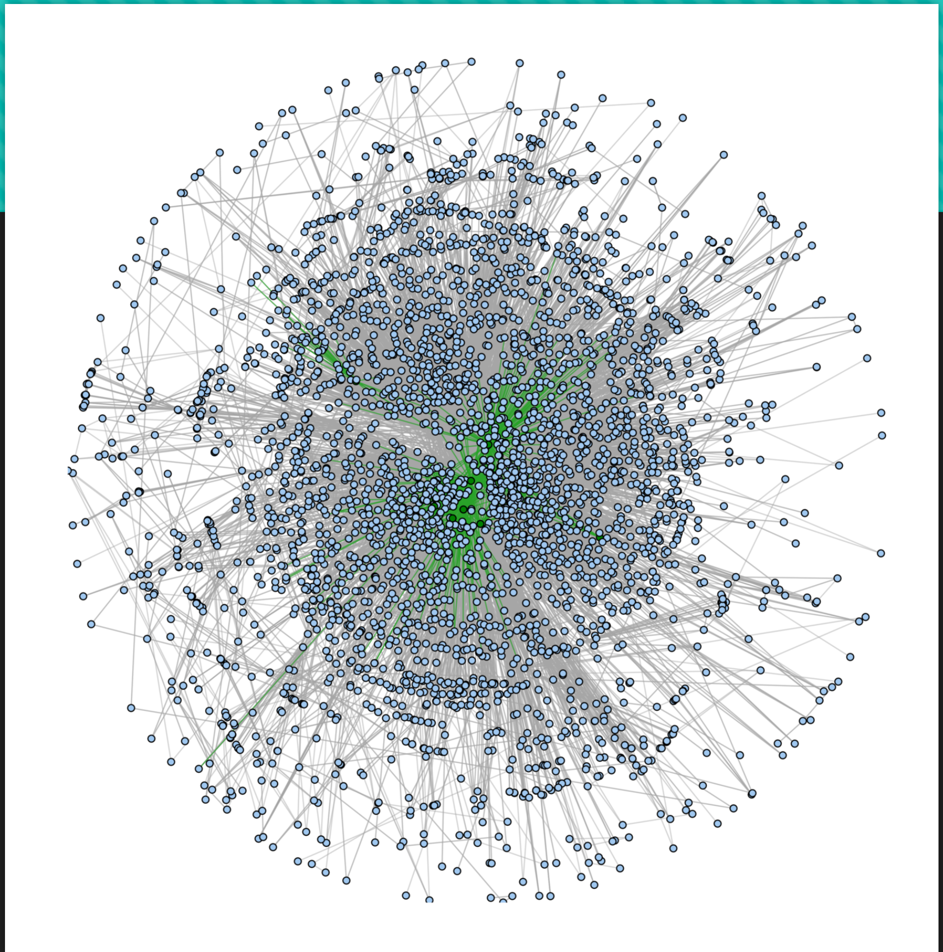
  - from each edge *i* to each other edge *j*, where σ_{ij} is the number of shortest paths from *i* to *j*

- Clustering coefficient summation
  - assigned to edge_{ij} as the sum of the clustering coefficients of vertices *i* and *j*

- Random strategy
  - randomly closes an edge

Demo

Simulations Initial State



- 10 most connected airports with infectious individuals
- Connected edges colored in green

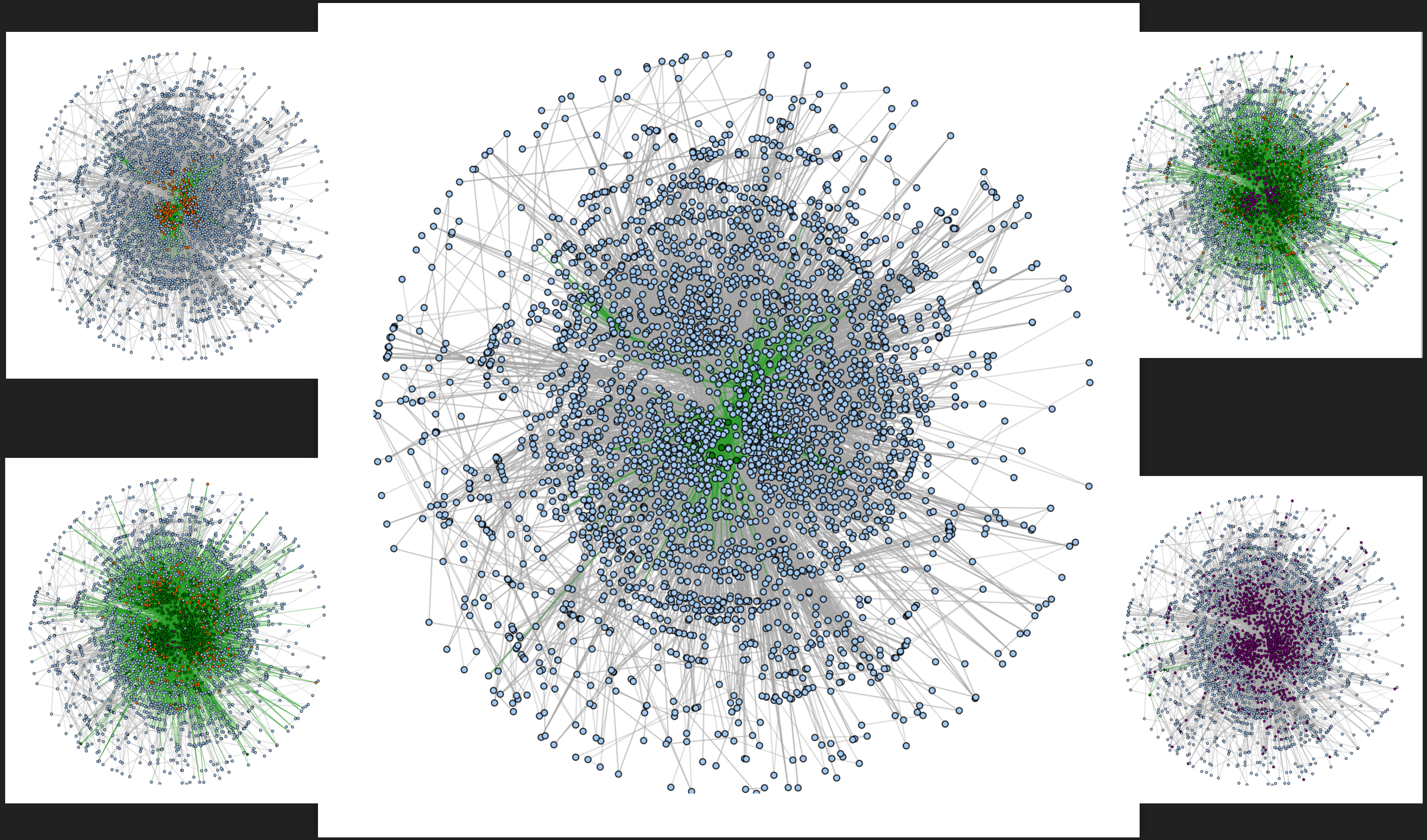
Disease Outspread Day1-Day10

Susceptible

Exposed

Infectious

Recovered

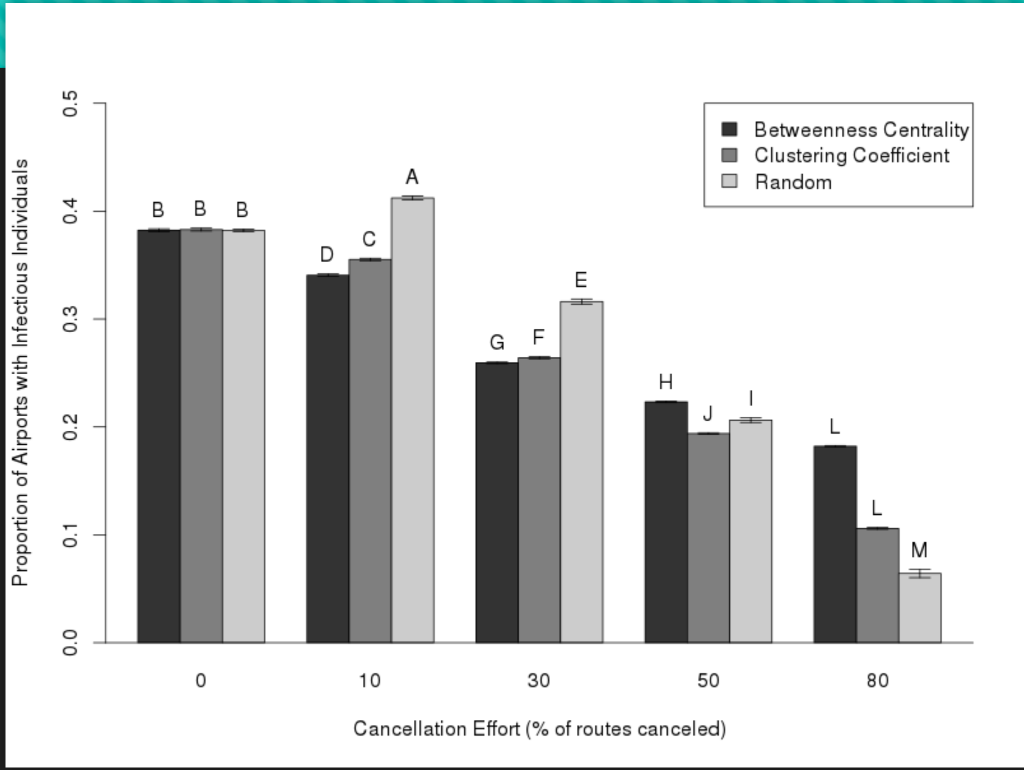


Simulation Examination

Simulation Examination

- A total of 20,580 simulations were performed
  - Optimal strategy based on the number of airports harboring infected individuals
- 343 simulations
  - for each of 3 strategies
  - at 20 cancellation efforts (5% interval)
- Examined 5 cancellation efforts:
  - 0%
  - 10%
  - 30%
  - 50%
  - 80%

Simulation Examination: Cancellation Comparison



- Tukey HSD Test

Thank You ☺