Formation of Scale-free networks: A Discrete-Event Simulation Approach

Rex K. Kincaid
Department of Mathematics
College of William and Mary

NASA Faculty Fellowship Program
July 30, 2004

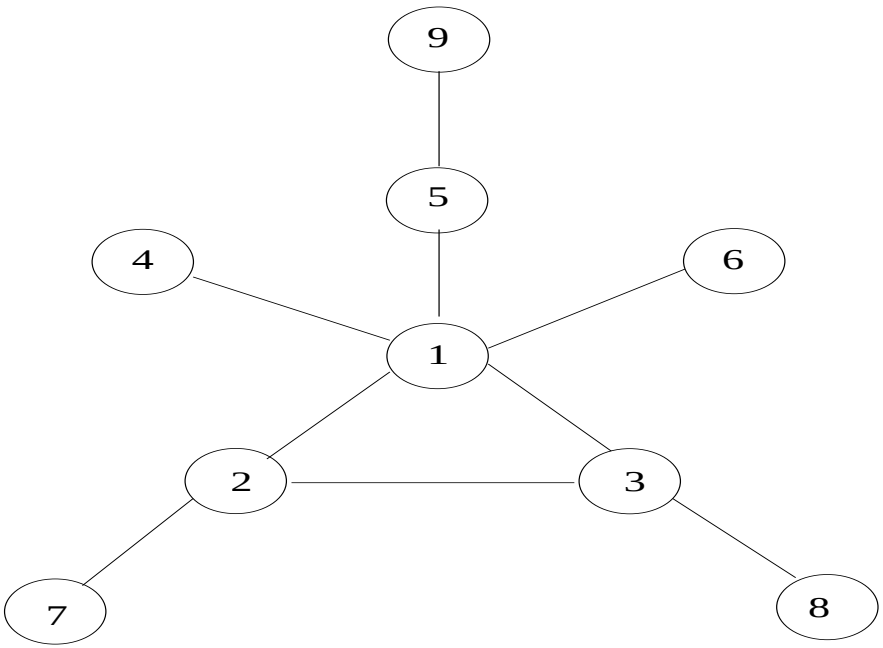
OUTLINE

- **Review of Definitions**
- **Some Results from Last Year**
- **Research Question**
- **Description of Simulation Model**
- **Computational Results**
- **Conclusions and Future Work**

Random Graphs—$G_{n,p}$

- **Erdos and Renya (1959,1960)**
- n **nodes**
- p **probability edge between two nodes**
- $p \cdot n(n-1)/2$ **expected number of edges**
- **degree distribution is binomial**
- **average shortest path** $log\ n$
- **clustering coefficient** $log\ n/n$

Example: degree distribution



- degree of node 1 is 5; degree sequence $\{1, 2, 3, 5\}$
- fraction of nodes of each degree k, $p(k) = \{5/9, 1/9, 2/9, 1/9\}$
- fraction of nodes degree k or larger, $P(k) = \{1, 4/9, 3/9, 1/9\}$—cdf

Clustering Coefficient

Transitivity: friend of your friend is also your friend

Goal: [0,1] number associated with number of triangles

Method 1:

$$C^{(1)} = \frac{\text{[three times total number of triangles]}}{\text{[total number of connected three tuples of nodes].}}$$

Method 2:

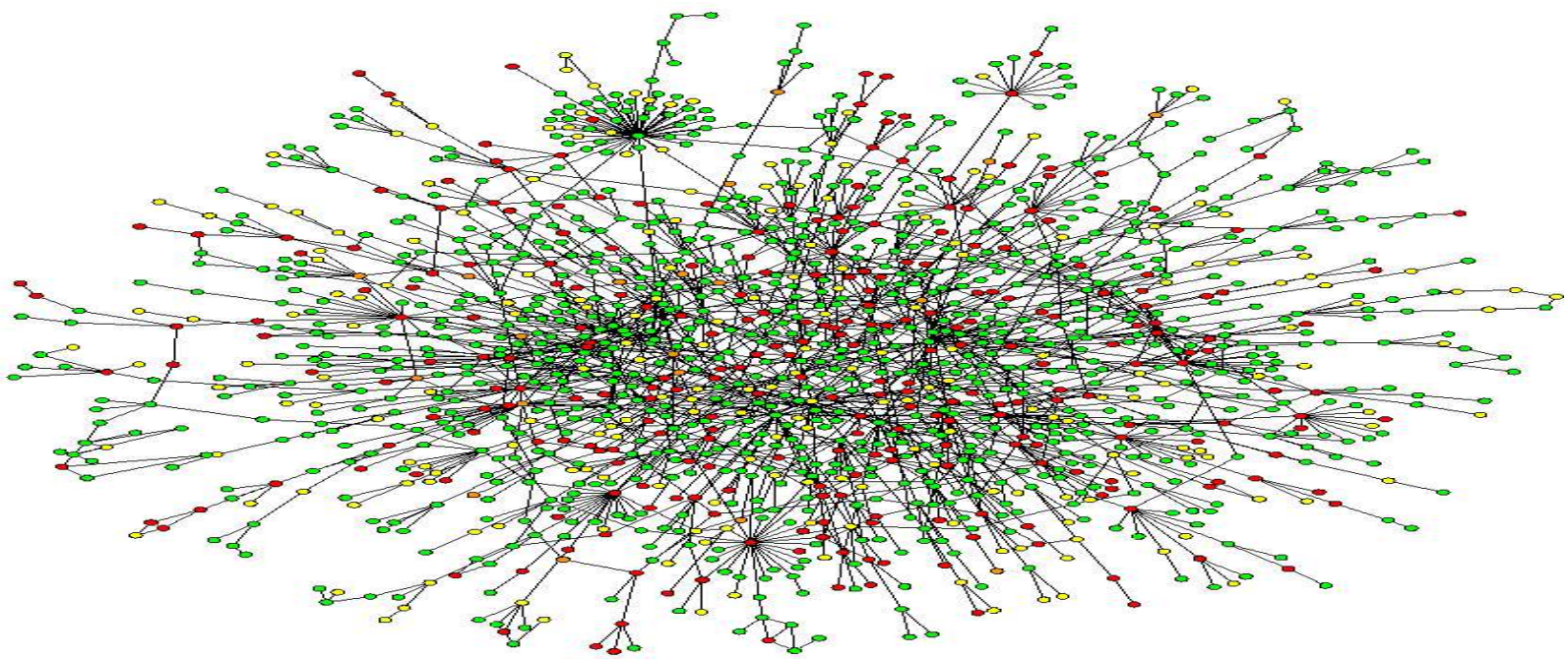
$$C_i = \frac{\text{[number of triangles in which node } i \text{ is incident]}}{\text{[number of three tuples of connected nodes centered on node } i\text{].}}$$

$$C^{(2)} = 1/n \cdot \Sigma_i C_i$$

- $C^{(1)} = 1/5$ and $C^{(2)} = 11/90$ in example

Properties of Scale-Free Graphs

- **Small-world: average shortest path $log\ n$**
- **Power law: tail of $P(k)$ is linear log-log plot**
- **Exponential: $P(k)$ tail is linear for log plot**
- **Clustering Coefficient: constant for large n**
- **Power laws $\in [2, 3]$ have $log\ log\ n$ diameter**



Popularity of Scale-free Graphs

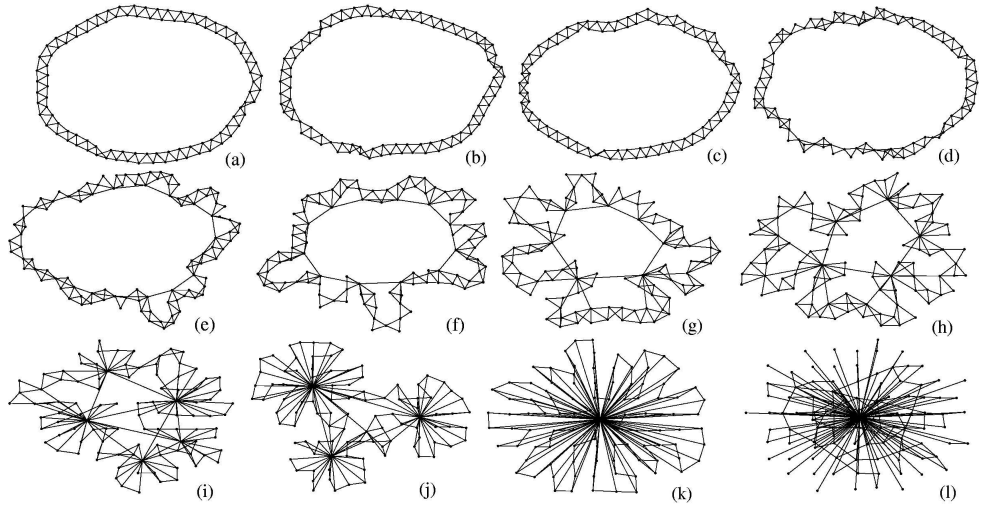
Graph	Nodes	Edges	AvgDeg	AvgSPD	$C^{(1)}$	$C^{(2)}$
film actor	449,913	25,516,482	113.443	3.48	.20	.78
Altavista	203,549,046	2,130,000,000	10.46	16.18	—	—
Internet	10,697	31,992	5.98	3.31	.035	.39
protein	2,115	2,240	2.12	6.80	.072	.071

- 6 books—Linked, Nexus, Six Degrees.
- More than 500 research articles since 1997.
- Expected C value for random film actor graph is .000252.

How and Why do Scale-Free Networks Form?

- **Preferential Attachment**
- **Bi-objective Optimization**
- **Dynamical Systems**
- **Statistical Mechanics**
- **Local search and shortest paths**
- **Nodes assigned fitness values (*Nature*, 2003)**
  - **Gradient network is scale-free**
  - **If $G(n,p)$ then maximal congestion**
  - **If scale-free then almost no congestion**

Small World Behaviour Emerging from Optimization *(continued)*



(a) $\lambda = 0$. (b) $\lambda = 5 \times 10^{-4}$. (c) $\lambda = 5 \times 10^{-3}$. (d) $\lambda = 0.0125$. (e) $\lambda = 0.025$. (f) $\lambda = 0.05$. (g) $\lambda = 0.125$. (h) $\lambda = 0.25$. (i) $\lambda = 0.5$. (j) $\lambda = 0.75$. (k) $\lambda = 1$.

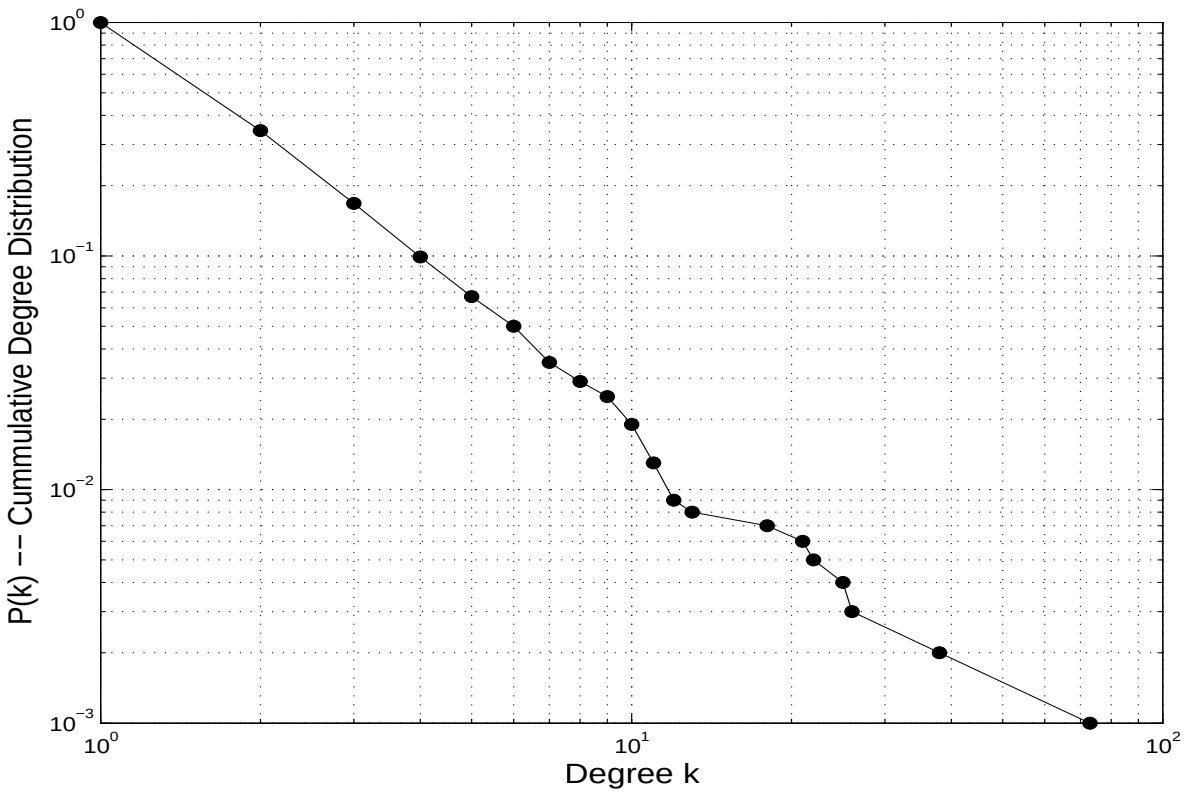
http://www.tp.umu.se/~kim/Network/holme1.pdf 10 *Umeå University, Sweden*

Bi-objective model: local vs. global

—Roger Guimera (NWU)

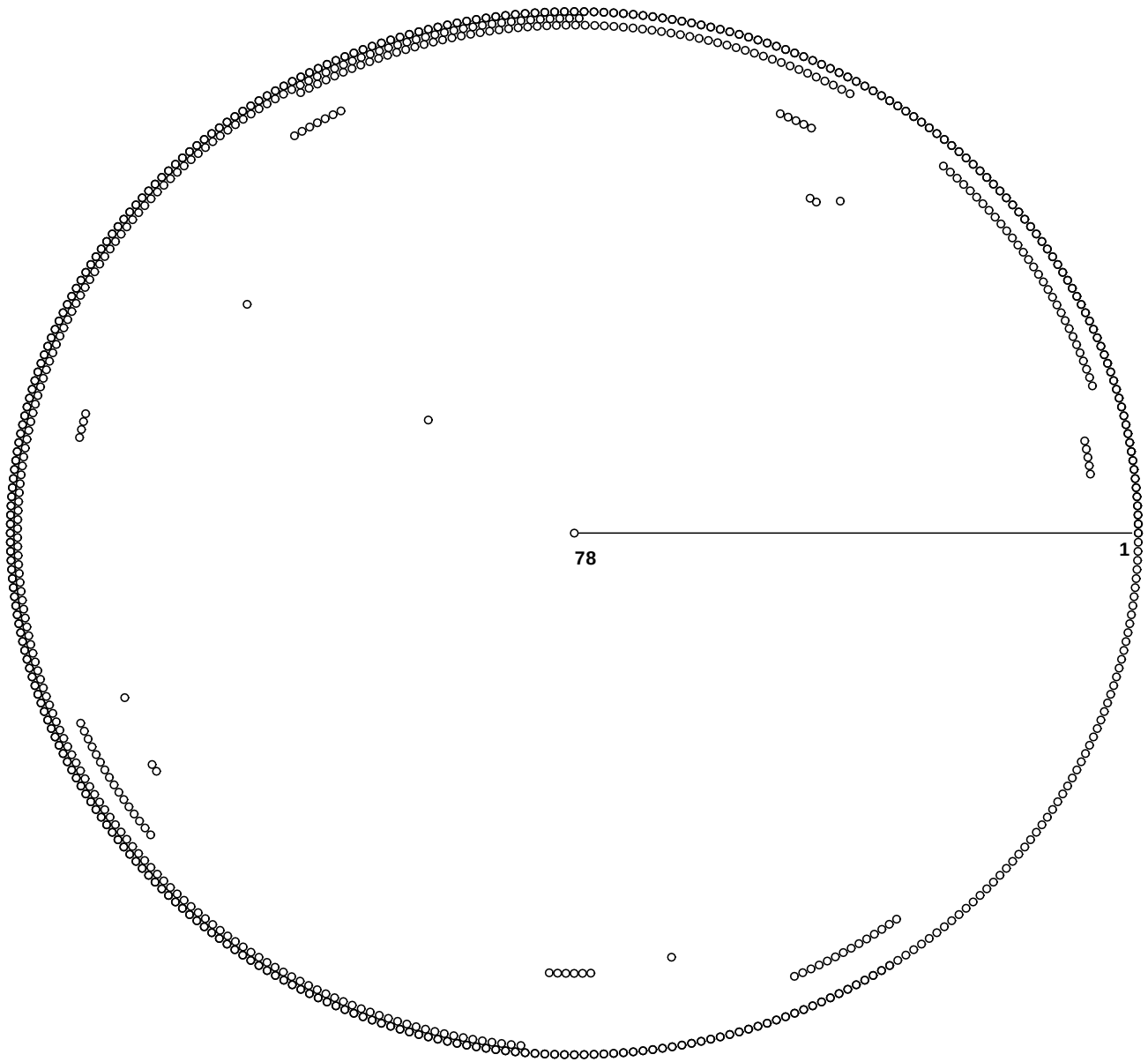
How do Scale-Free Networks Form?

—preferential attachment

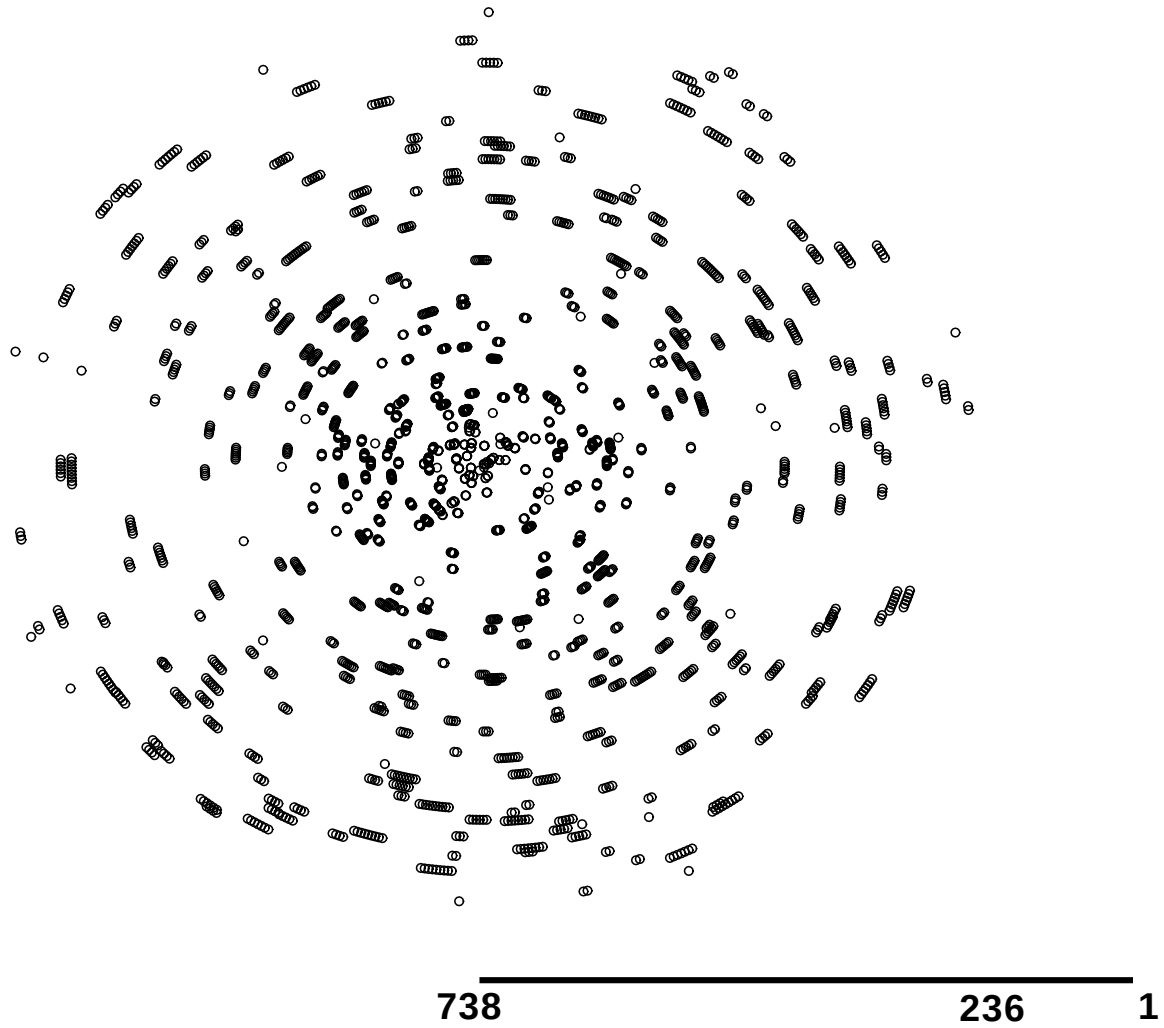


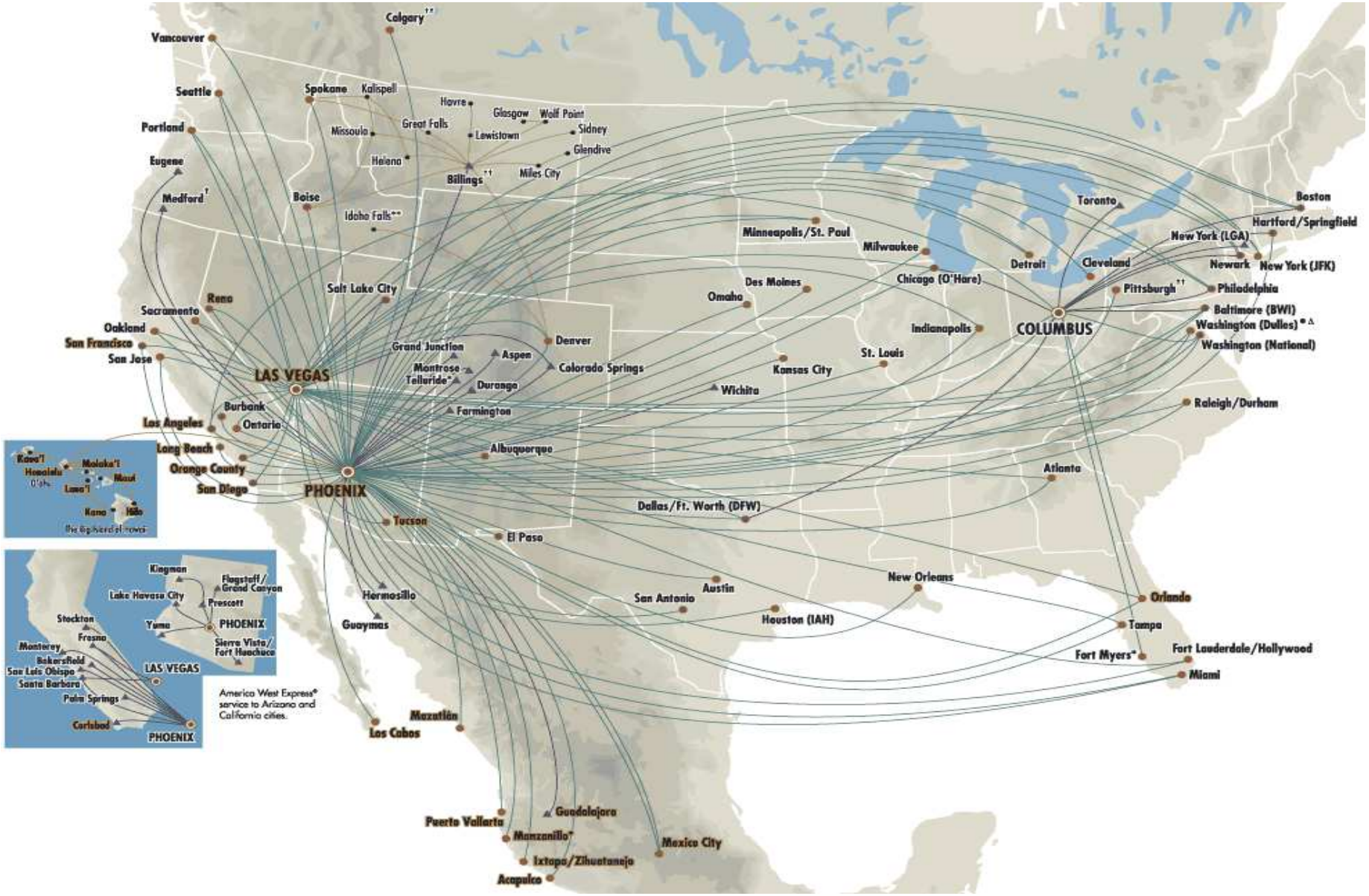
- **Start with 3 node 2 edge path. Add nodes one at a time: average degree**
- **Roulette wheel selection based on $p(k)$—1000 nodes**

Preferential Attachment: 1000 nodes, m=1



Uniform Random Graph: 2103 Nodes

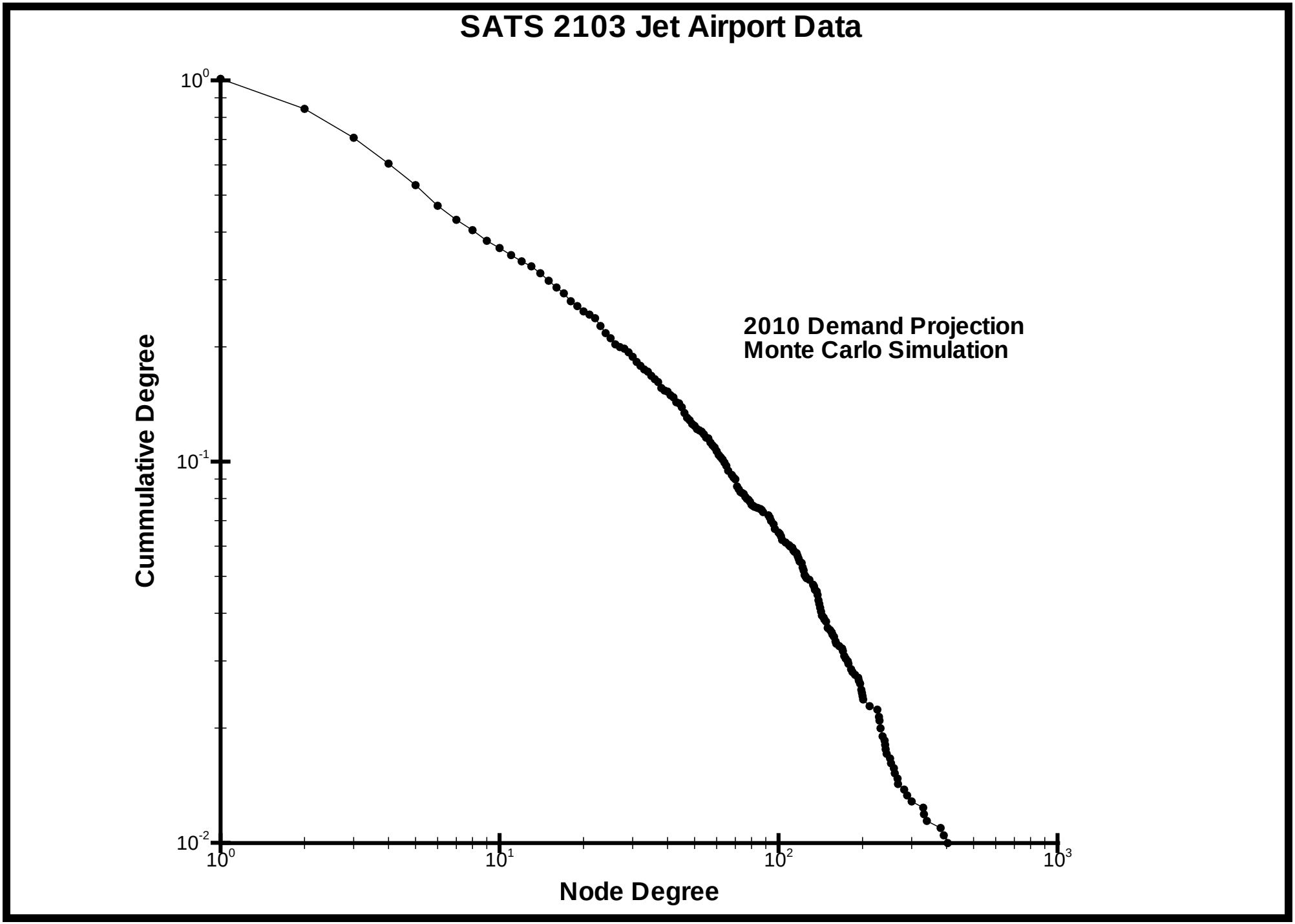



Calgary
Vancouver
Seattle
Spokane
Kalispell
Havre
Glasgow
Wolf Point
Portland
Missoula
Great Falls
Lewistown
Sidney
Glendive
Eugene
Helena
Billings
Miles City
Medford
Boise
Idaho Falls
Toronto
Boston
Hartford/Springfield
Minneapolis/St. Paul
Milwaukee
New York (LGA)
Newark
New York (JFK)
Detroit
Cleveland
Chicago (O'Hare)
Des Moines
Omaha
Pittsburgh
Philadelphia
Salt Lake City
Reno
Sacramento
Oakland
San Francisco
San Jose
Baltimore (BWI)
Washington (Dulles)
Washington (National)
Indianapolis
COLUMBUS
Denver
Grand Junction
Aspen
Montrose
Telluride
Durango
Colorado Springs
St. Louis
Kansas City
LAS VEGAS
Wichita
Burbank
Farmington
Raleigh/Durham
Los Angeles
Ontario
Long Beach
Albuquerque
Orange County
San Diego
PHOENIX
Atlanta
Dallas/Ft. Worth (DFW)
Tucson
El Paso
Kingman
Flagstaff/Grand Canyon
Lake Havasu City
Prescott
Stockton
Yuma
PHOENIX
Fresno
Sierra Vista/Fort Huachuca
Monterey
Bakersfield
San Luis Obispo
Santa Barbara
LAS VEGAS
Palm Springs
Carlsbad
PHOENIX
America West Express service to Arizona and California cities.
Hermosillo
Guaymas
Austin
San Antonio
New Orleans
Houston (IAH)
Orlando
Tampa
Fort Myers
Fort Lauderdale/Hollywood
Miami
Mazatlán
Los Cabos
Guadalajara
Puerto Vallarta
Manzanillo
Mexico City
Ixtapa/Zihuatanejo
Acapulco

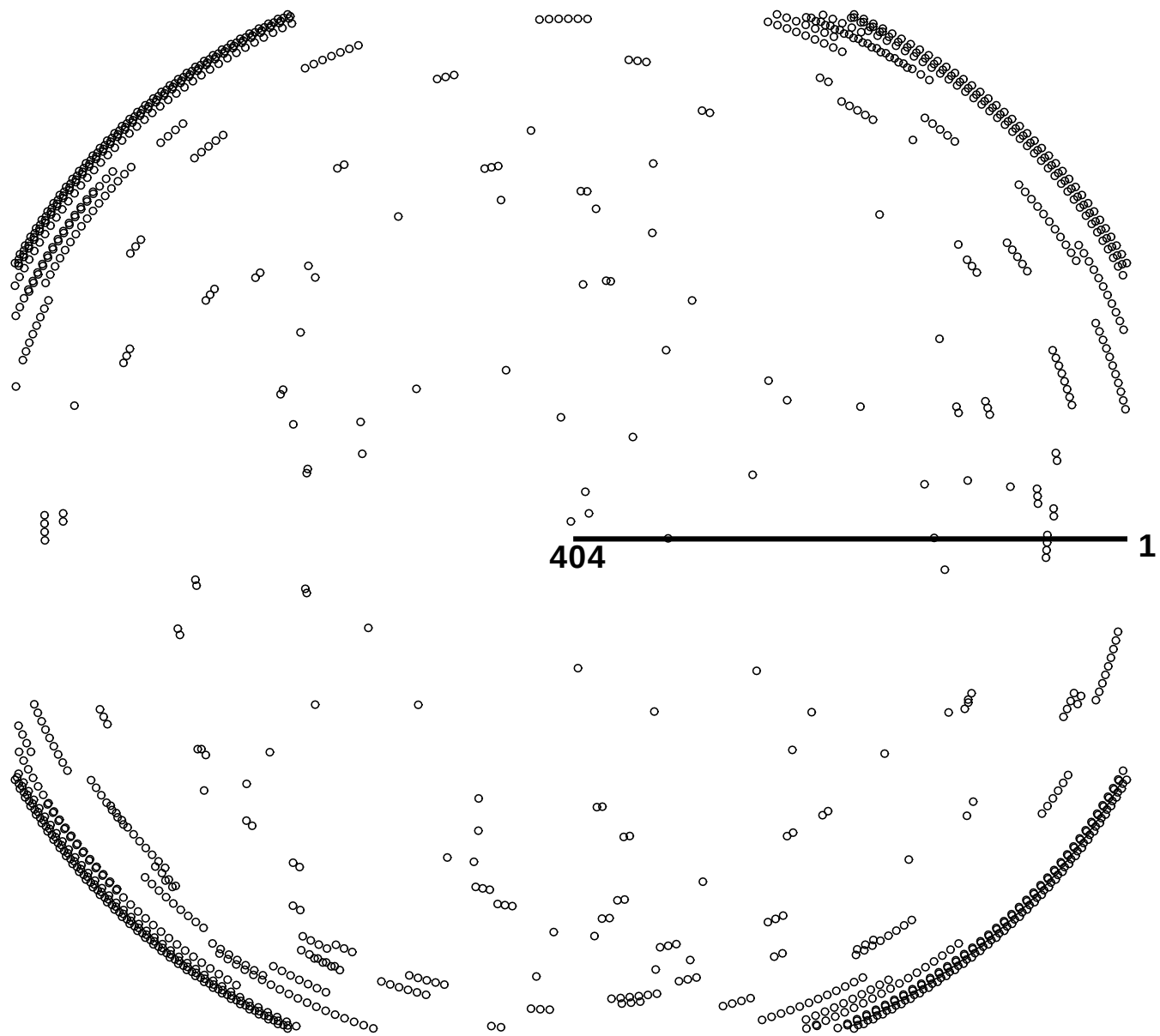
Scale-Free Metrics for SATS General Aviation Model

- **Gravity model for flow between airports ($log_{10}\ 2102 = 3.33$)**
- **Monte Carlo simulation of high traffic day**

Data Set	# of Airports	Avg Shortest Hops	$C^{(1)}$
Jets	2102	2.8	0.195
SingleIFR	2323	4.94	0.07
SingleVFR	2985	2.93	0.158
MultiVFR	2415	3.86	0.057
Piston	2274	4.14	0.047
Turbo	2266	4.17	0.047


SATS 2103 Jet Airport Data
2010 Demand Projection
Monte Carlo Simulation
Cummulative Degree
Node Degree
10^0
10^{-1}
10^{-2}
10^0
10^1
10^2
10^3

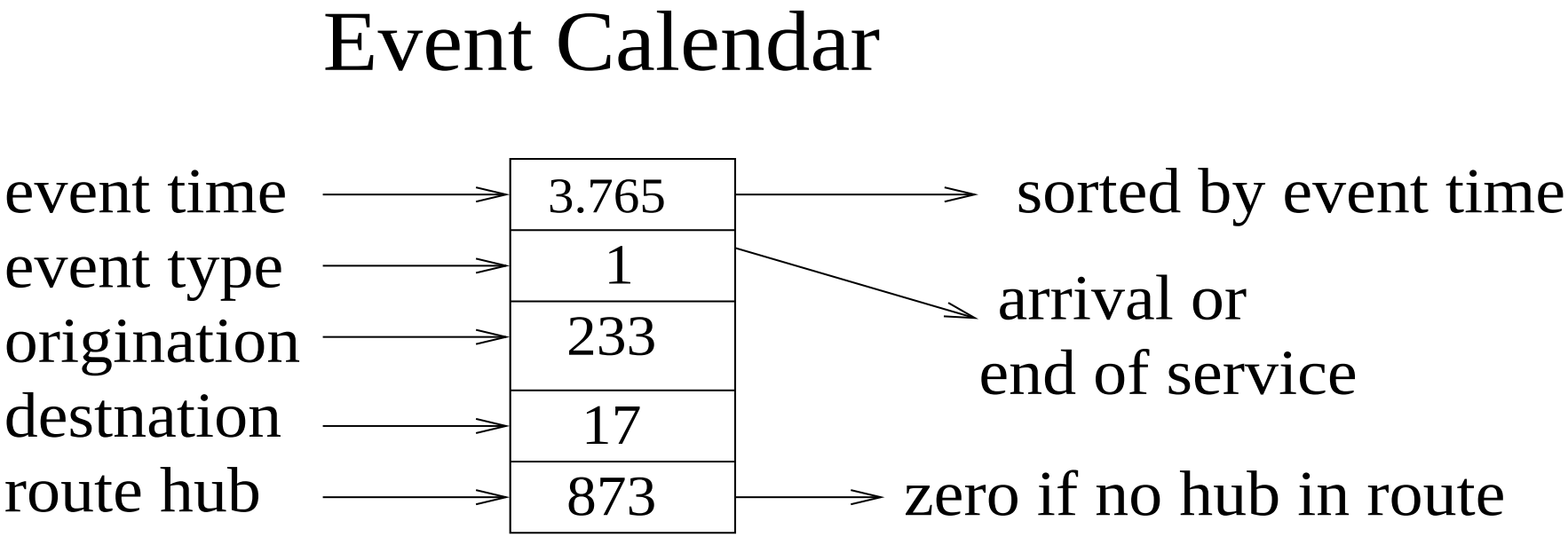
2103 Airports: SATS 2010 Jet Data Set



Discrete-Event Simulation

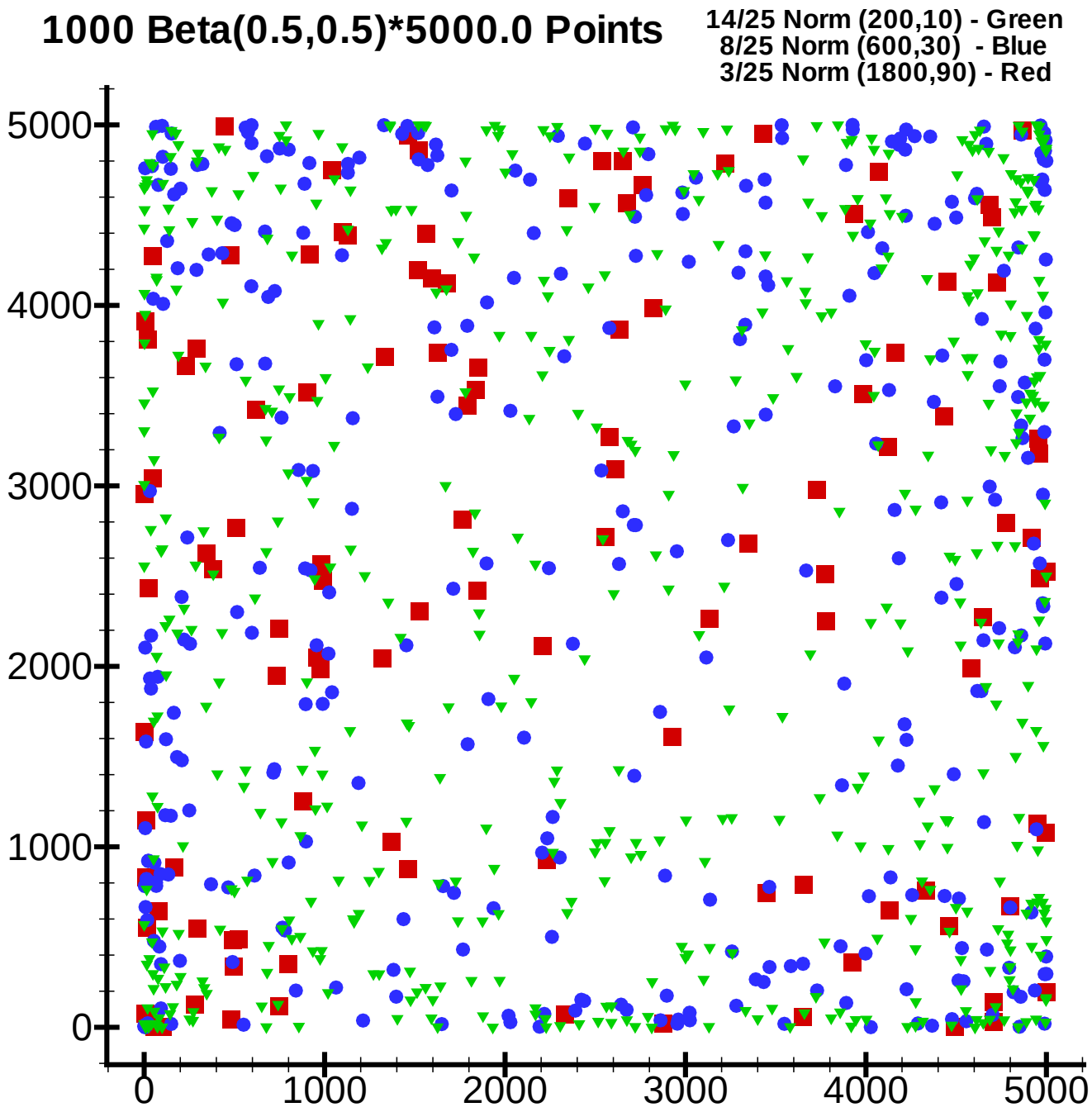
- **Entities are Planes (sorted, fixed length list)**
- **Beta Distributed Points are Cities**
- **Each City has a Normally distributed Pop_Size**
- **Euclidean Distance between Cities**
- **O/D selected preferentially based on Pop_Size**
- **Power law fitness generated via simulation**

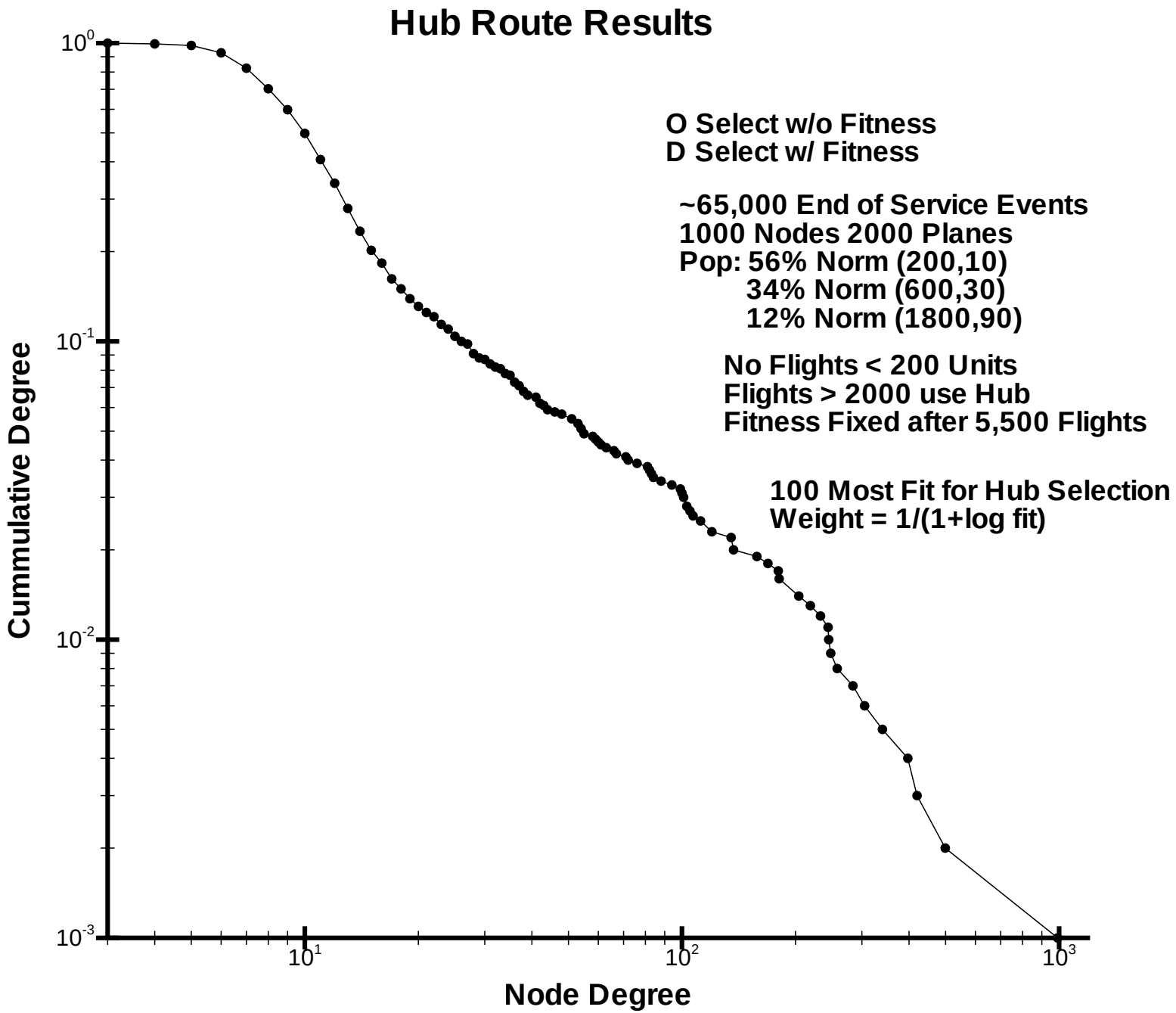
How is time advanced?



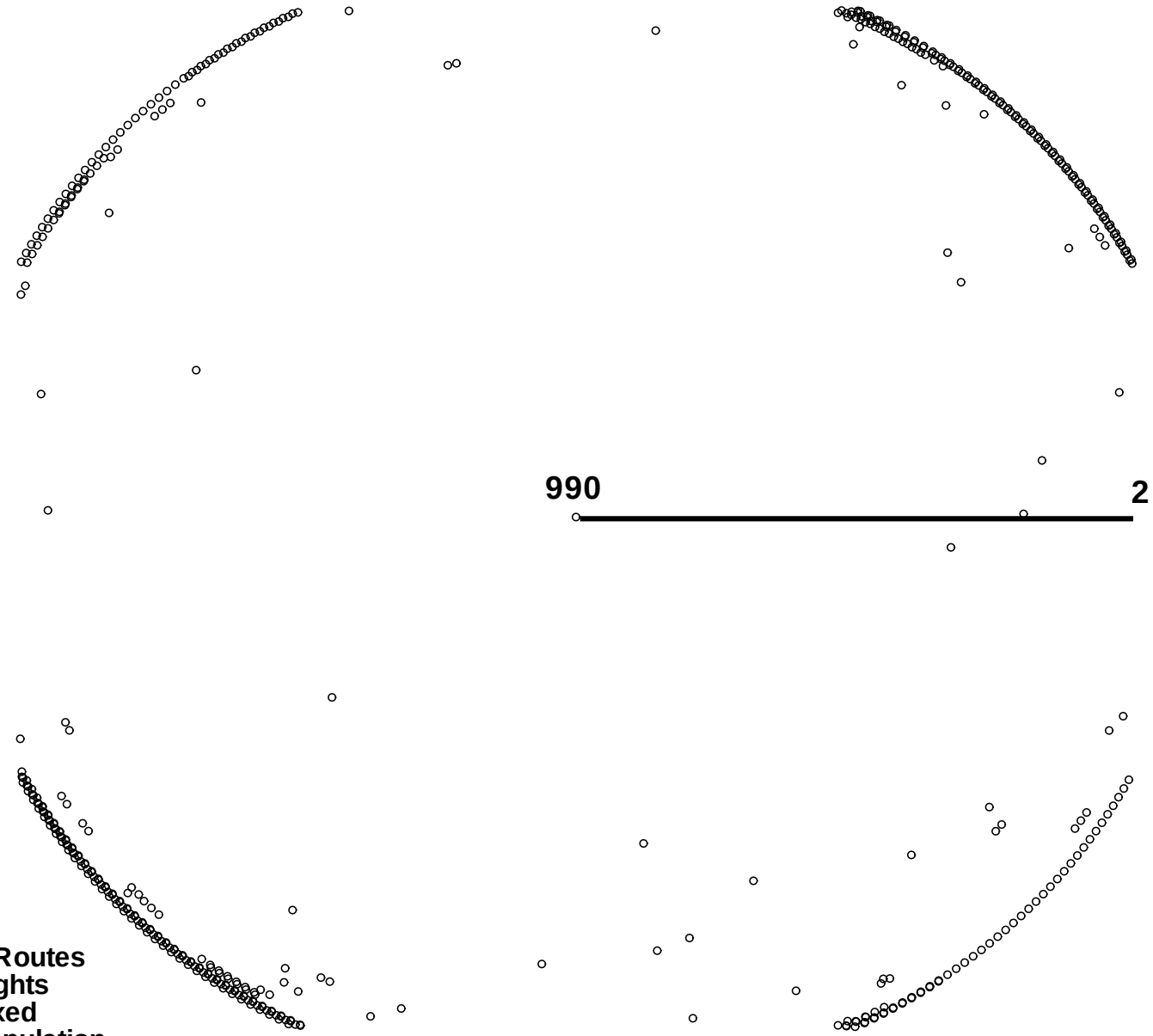
Only arrival events or end of service events trigger time advance.

The event list must be sorted after each new event is added to the calendar.

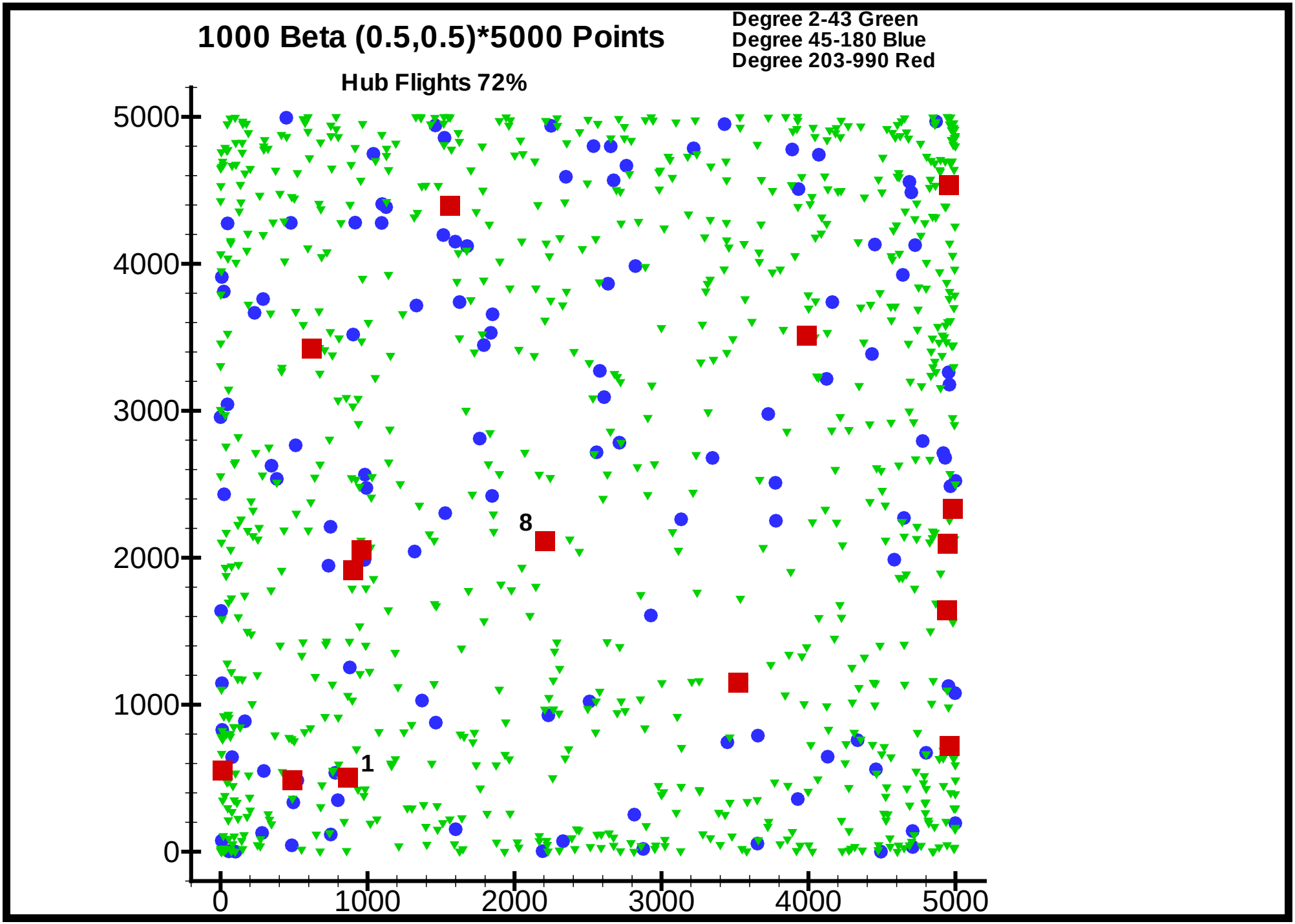

1000 Beta(0.5,0.5)*5000.0 Points
14/25 Norm (200,10) - Green
8/25 Norm (600,30) - Blue
3/25 Norm (1800,90) - Red
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000

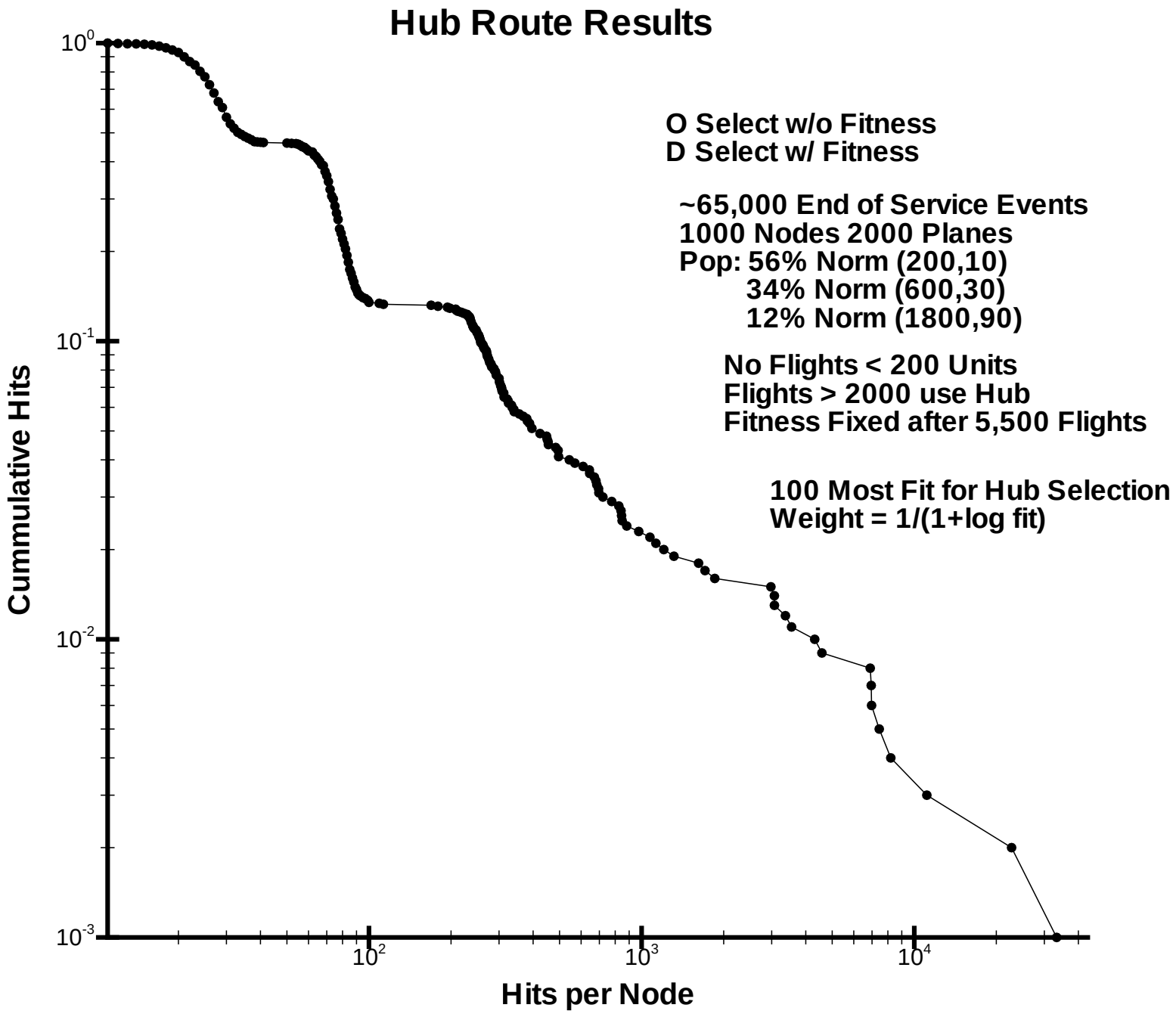

Hub Route Results
O Select w/o Fitness
D Select w/ Fitness
~65,000 End of Service Events
1000 Nodes 2000 Planes
Pop: 56% Norm (200,10)
34% Norm (600,30)
12% Norm (1800,90)
No Flights < 200 Units
Flights > 2000 use Hub
Fitness Fixed after 5,500 Flights
100 Most Fit for Hub Selection
Weight = 1/(1+log fit)
Cummulative Degree
Node Degree
10^0
10^{-1}
10^{-2}
10^{-3}
10^1
10^2
10^3

Degree Distribution: 1000 Nodes, 2000 Planes

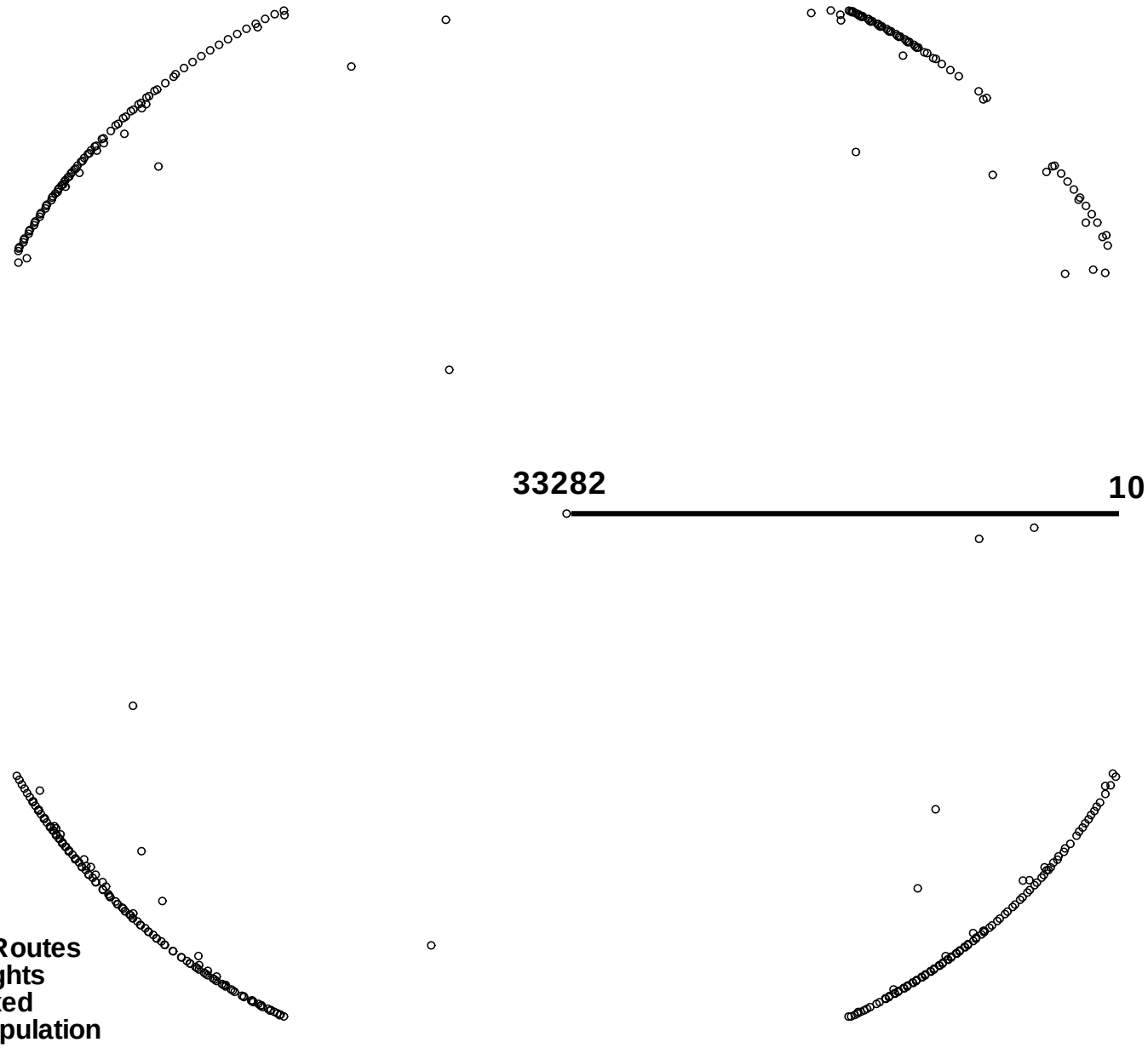


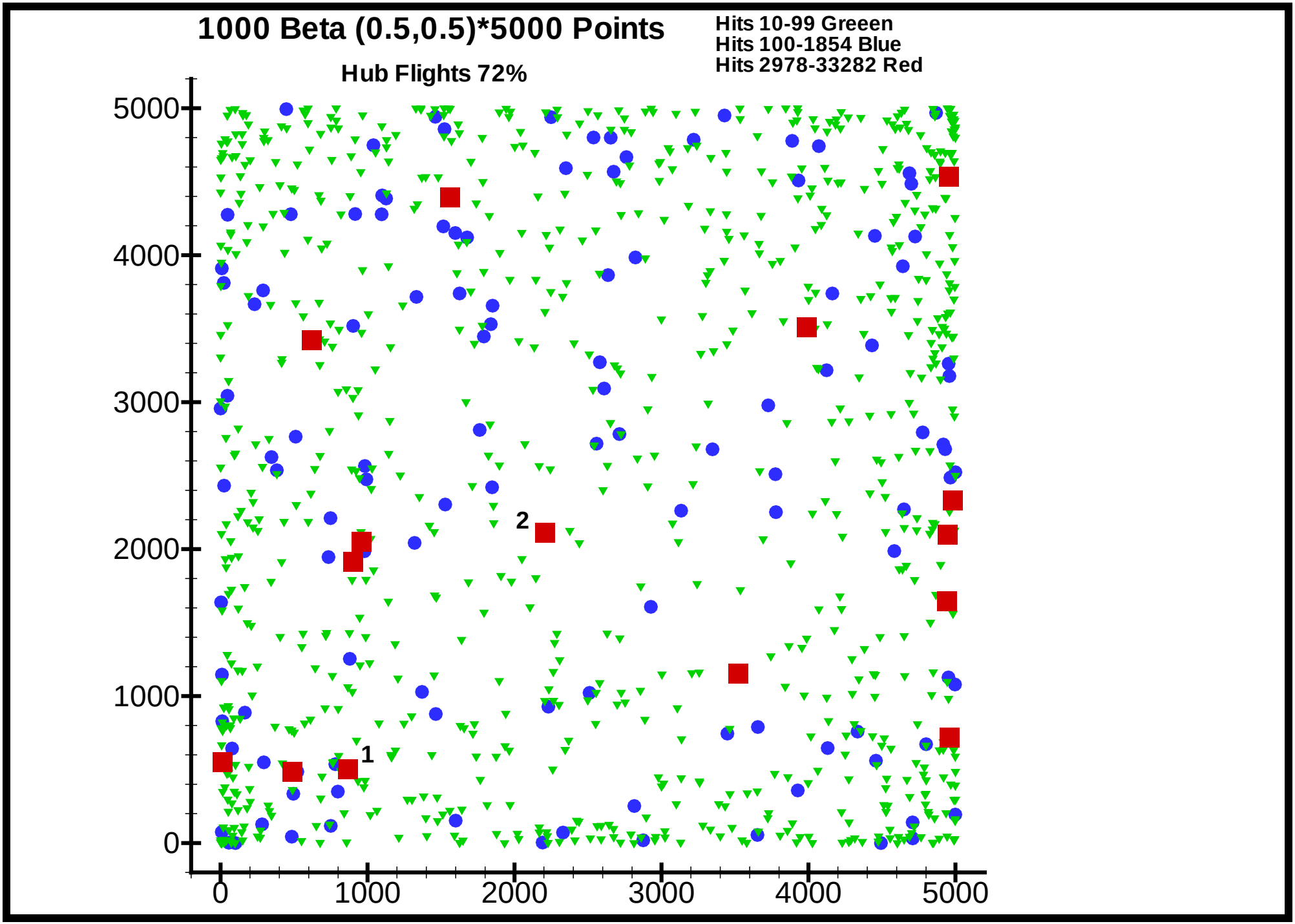
72% Hub Routes
65,000 Flights
Fitness Fixed
3 Norm Population

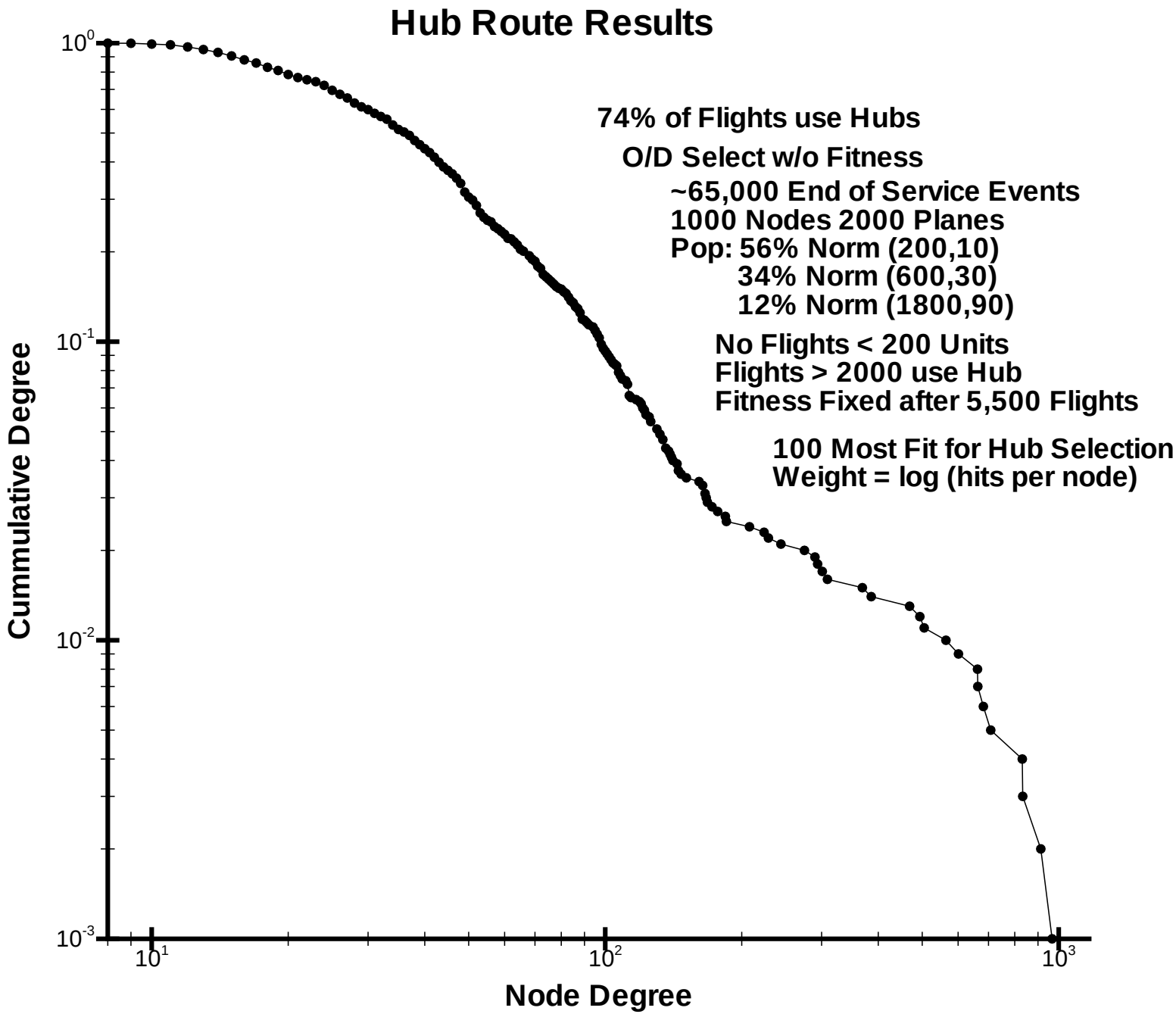

1000 Beta (0.5,0.5)*5000 Points
Degree 2-43 Green
Degree 45-180 Blue
Degree 203-990 Red
Hub Flights 72%
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
8
1

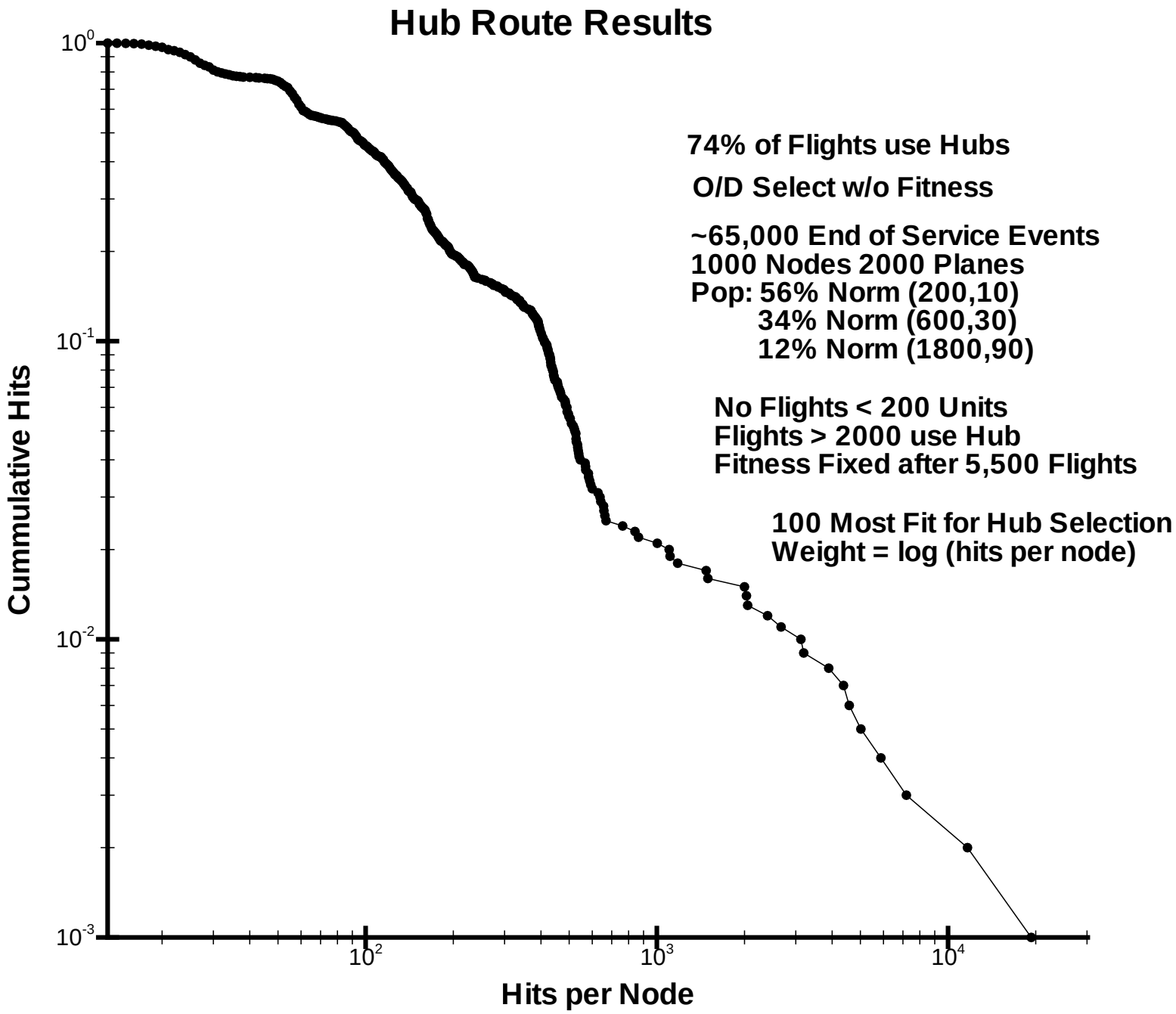

Hub Route Results
O Select w/o Fitness
D Select w/ Fitness
~65,000 End of Service Events
1000 Nodes 2000 Planes
Pop: 56% Norm (200,10)
34% Norm (600,30)
12% Norm (1800,90)
No Flights < 200 Units
Flights > 2000 use Hub
Fitness Fixed after 5,500 Flights
100 Most Fit for Hub Selection
Weight = 1/(1+log fit)
Cummulative Hits
Hits per Node
10^0
10^{-1}
10^{-2}
10^{-3}
10^2
10^3
10^4

Hit Distribution: 1000 Nodes, 2000 Planes




1000 Beta (0.5,0.5)*5000 Points
Hits 10-99 Greeen
Hits 100-1854 Blue
Hits 2978-33282 Red
Hub Flights 72%
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
1
2


Hub Route Results
74% of Flights use Hubs
O/D Select w/o Fitness
~65,000 End of Service Events
1000 Nodes 2000 Planes
Pop: 56% Norm (200,10)
34% Norm (600,30)
12% Norm (1800,90)
No Flights < 200 Units
Flights > 2000 use Hub
Fitness Fixed after 5,500 Flights
100 Most Fit for Hub Selection
Weight = log (hits per node)
Cummulative Degree
Node Degree
10^0
10^{-1}
10^{-2}
10^{-3}
10^1
10^2
10^3


Hub Route Results
74% of Flights use Hubs
O/D Select w/o Fitness
~65,000 End of Service Events
1000 Nodes 2000 Planes
Pop: 56% Norm (200,10)
34% Norm (600,30)
12% Norm (1800,90)
No Flights < 200 Units
Flights > 2000 use Hub
Fitness Fixed after 5,500 Flights
100 Most Fit for Hub Selection
Weight = log (hits per node)
Cummulative Hits
Hits per Node
10^0
10^{-1}
10^{-2}
10^{-3}
10^2
10^3
10^4

Conclusions

- **Population Density, Flight Delay not enough for Scale-Free**
- **Power Law cdf for Node Fitness needed to observe Scale-Free Network**
- **Visualization scheme for degree distribution**
- **Insight is goal of simulation**
- **Simulation code is in C (about 500 lines)**
  - **rngs.c and rvgs.c libraries (S. Park)**
  - **approx. 300 lines of codes for utilities**
- **NFFP program benefits**

Future Directions

- Capture transition from ??? to scale-free
- Use 1000 U.S. cities (coordinates, population)
- Add capacity restrictions to cities
- Experiment with current model
  - different distributions
  - different classes of aircraft
  - add double linked list for event calendar
- Add queueing effects (ARENA needed)
- Examine flow effects on network structure