Specific Communication Network Measure Distribution Estimation

Daniel P. Baller, Joshua Lospinoso

17 June 2008

13th ICCRTS

USMA Network Science Center
http://www.netscience.usma.edu/

Agenda

- Background & Motivation
- ORA Visualization
- Assumptions
- NPM
- Simulated Networks
- Results
- Future Research

Background & Motivation

- Simulating random instances of networks is a hot topic in today's Network Science research
  - Erdos-Renyi, Watts-Strogatz, Barabasi-Alberts, NPM
- How do networks arrive at **structure**? How do we explore these structures?

- Many methods of simulating **random networks** exist
- No sound methodology exists for measuring "goodness"

- We propose a methodology for **testing** how well simulations perform under a rigorous **statistical framework**, and execute testing for two data sets under the *Network Probability Matrix* and *Erdos-Renyi*.

Data Set 1 (Net07)

- Warfighting Simulation run at FT Leavenworth, KS in April 2007.
- 68 Mid Career Army Officers
- 4 day simulated exercise
- Self Reported Communications survey
- Surveys conducted 2 x per day
- 8 Total Data Sets

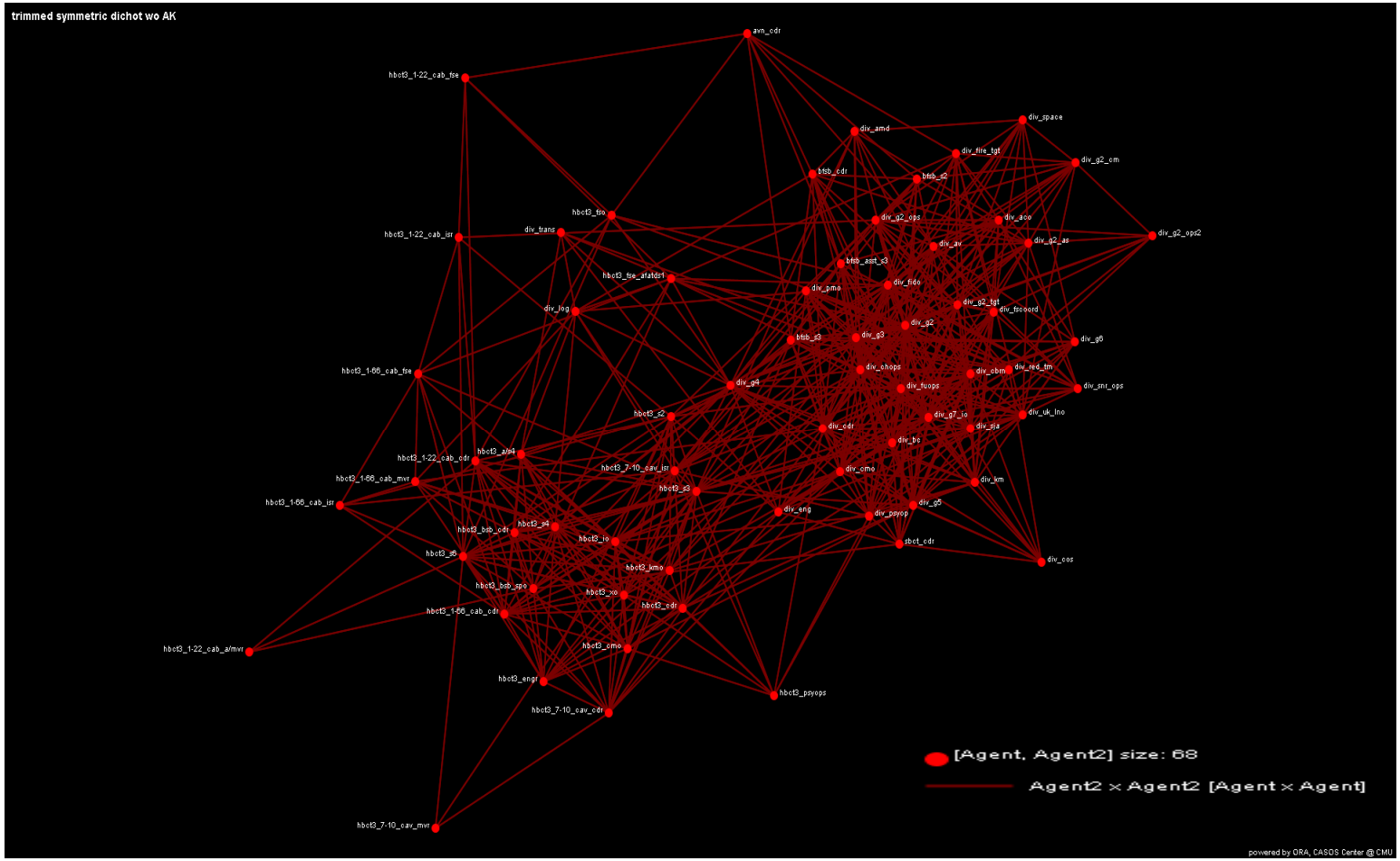
Data Set 2 (Net05)

- Warfighting Simulation run at FT Leavenworth, KS in 2005.
- 156 Mid Career Army Officers
- 5 day simulated exercise
- Self Reported Communications survey
- Surveys conducted 2 x per day
- 9 Total data sets

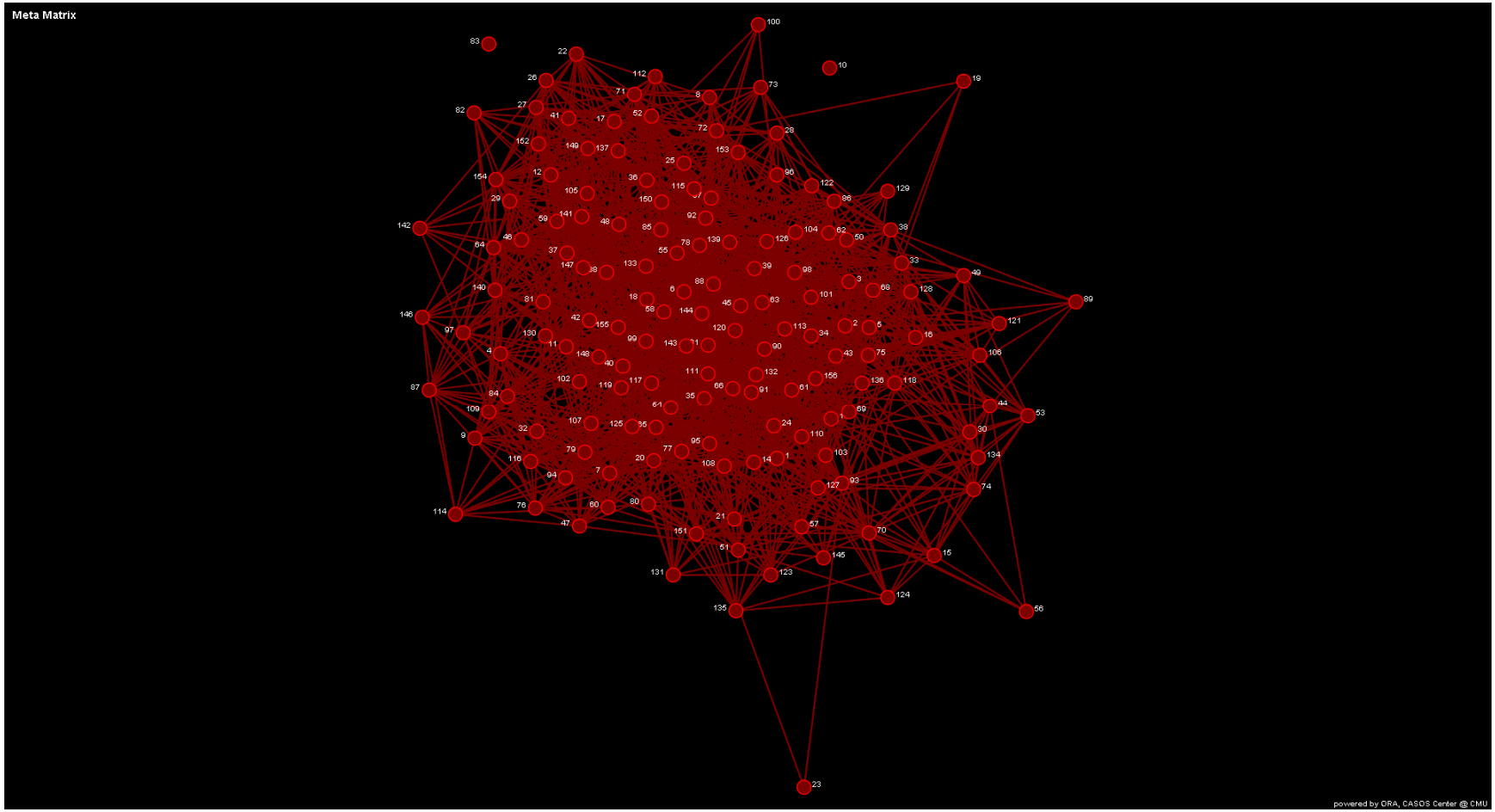
Data

- Data symmetrized and dichotomized
- Square symmetric matrix
- Over time data compiled by agent

ORA Visualization (Net07)

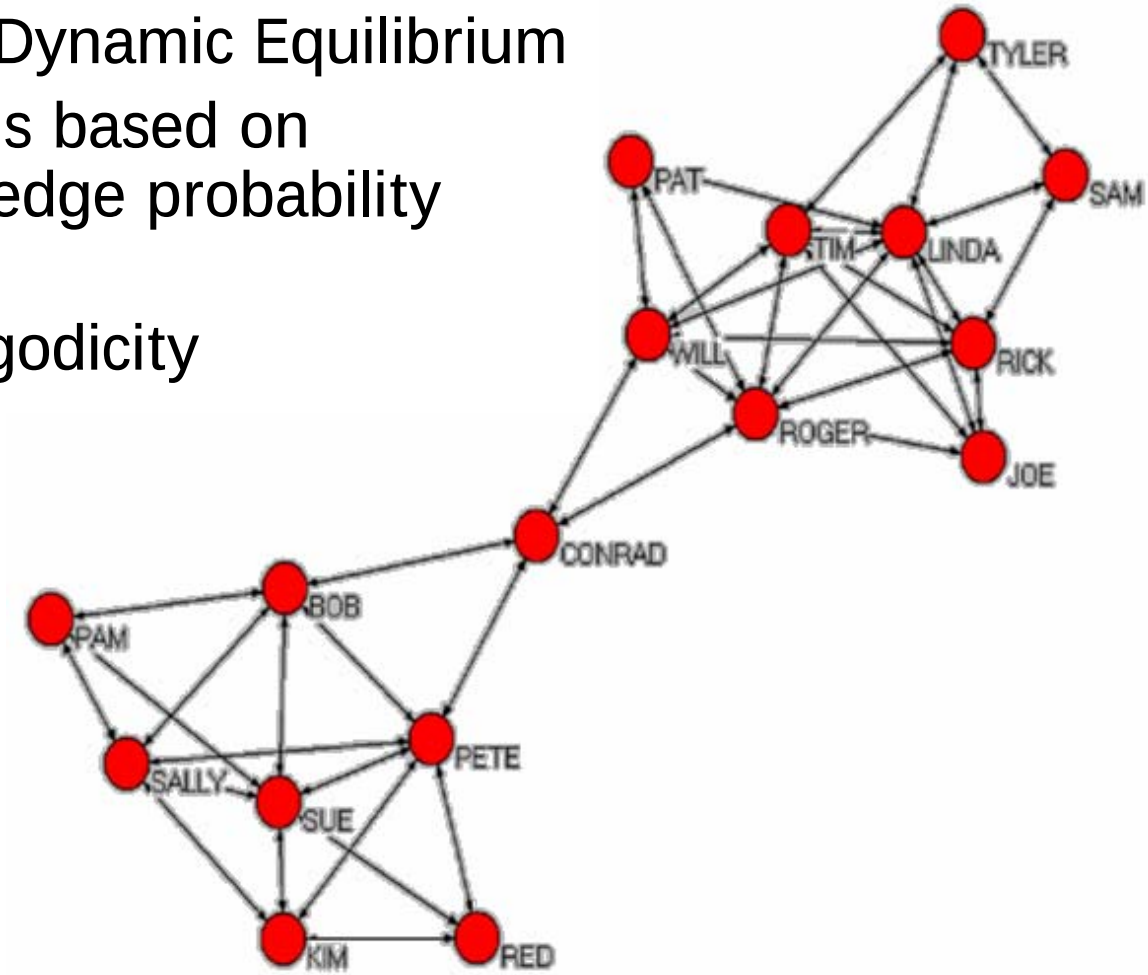


ORA Visualization (Net05)



Assumptions

- Network in Dynamic Equilibrium
- Observations based on Underlying edge probability structure
- Maintain ergodicity



Graph courtesy of David Krackhardt

Network Probability Matrix

Microsoft Excel - prob3.csv

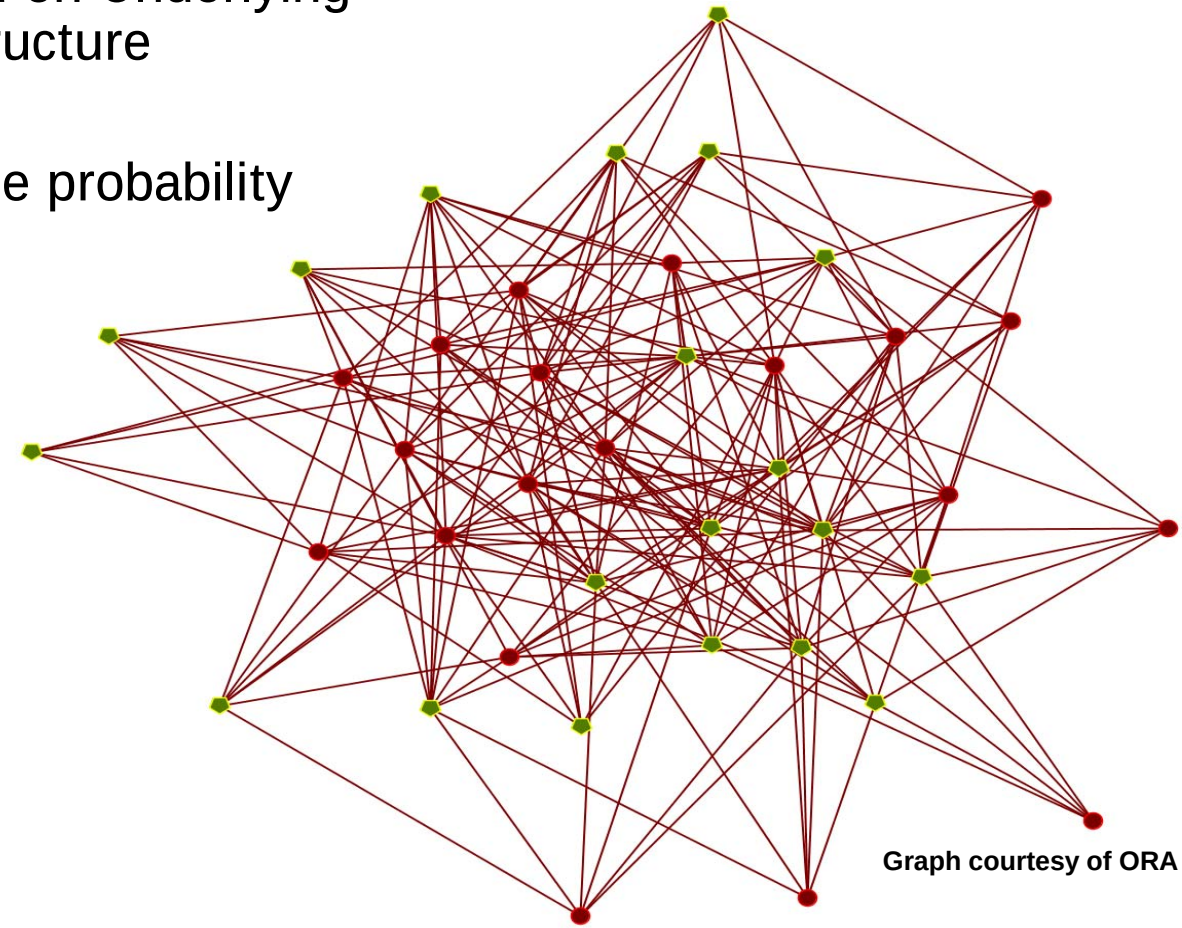
File Edit View Insert Format Tools Data Window Help

H10 0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	0	0.833333	0	0.5	0	0	0	0	0	0.5	1	0.833333	0	1	0	0.166667	
2	0.833333	0	0	0	0	0	0	0	0	0.166667	0	0.166667	0	0	0	0	
3	0	0	0	0	0	0	0	0.166667	0.166667	0.166667	0	0.5	0.5	0.5	0	0.333333	0.33333
4	0.5	0	0	0	0.833333	1	1	0.5	1	0.166667	1	1	0.666667	0	0.333333	0.666667	0.16666
5	0	0	0	0.833333	0	0.666667	0.333333	1	0.666667	0	0	0.166667	0	0	0.166667	0.166667	
6	0	0	0	1	0.666667	0	1	0.5	1	0	0	0	0	0	0	0	
7	0	0	0	1	0.333333	1	0	0.166667	0.666667	0	0.166667	0.5	0.333333	0	0	0	
8	0	0	0.166667	0.5	1	0.5	0.166667	0	0.166667	0	0	0.166667	0	0	0.166667	0.333333	
9	0	0	0.166667	1	0.666667	1	0.666667	0.166667	0	0	0	0.5	0.5	0	0	0.333333	0.33333
10	0.5	0.166667	0.166667	0.166667	0	0	0	0	0	0	0.666667	0.666667	0.333333	0.666667	0	0.166667	
11	1	0	0	1	0	0	0.166667	0	0	0.666667	0	1	1	1	0.666667	0.833333	
12	0.833333	0.166667	0.5	1	0.166667	0	0.5	0.166667	0.5	0.666667	1	0	1	0.5	0	0.666667	
13	0	0	0.5	0.666667	0	0	0.333333	0	0.5	0.333333	1	1	0	1	0.166667	0.166667	
14	1	0	0.5	0	0	0	0	0	0	0.666667	1	0.5	1	0	0	0	0.33333
15	0	0	0	0.333333	0.166667	0	0	0.166667	0	0	0.666667	0	0.166667	0	0	1	0.83333
16	0.166667	0	0.333333	0.666667	0.166667	0	0	0.333333	0.333333	0.166667	0.833333	0.666667	0.166667	0	1	0	
17	0	0	0.333333	0.166667	0	0	0	0	0.333333	0	0	0	0	0.333333	0.833333	1	
18	0	0	0	0.333333	1	0	0.166667	0.666667	0.666667	0	0.166667	0.166667	0	0	0	0	
19	0.166667	0.166667	0.166667	0.5	0	0	0.166667	0	0	0.666667	0.833333	0.666667	0	0.833333	0	0	
20	0	0	0	0	0.666667	0.166667	0.666667	0	0	0.166667	0.5	0	0	0.5	0.666667	0.166667	0.
21	0	0	0.166667	0	0.166667	0	0.833333	0	0	0	0	0	0	0	0.166667	0	
22	0.833333	0.5	0.166667	0.333333	0.333333	0.333333	0.166667	0	0.333333	0.333333	0.666667	0.5	0	0.666667	0	0	0.16666
23	0	0	0	0.166667	0	0	0.333333	0	0	0.5	0.5	0	0	0.333333	0	0	
24	0.166667	0	0	0.166667	0	0	0.166667	0.666667	0	0	1	0.166667	0.166667	0	0.5	0	0.16666
25	0.333333	0	0	0.166667	0	0	0	0	0	0.166667	0.833333	0.666667	0.333333	0.166667	0	0.5	
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	0	0.166667	0.166667	0.333333	0.166667	0	0.333333	0	0	0	0	0.666667	0.333333	0.5	0	0	
29	0	0	0.166667	0	0	0	0.166667	0	0	0	0	0.833333	0.833333	0.166667	0	0	

Erdsös-Réni Random Graph Assumptions

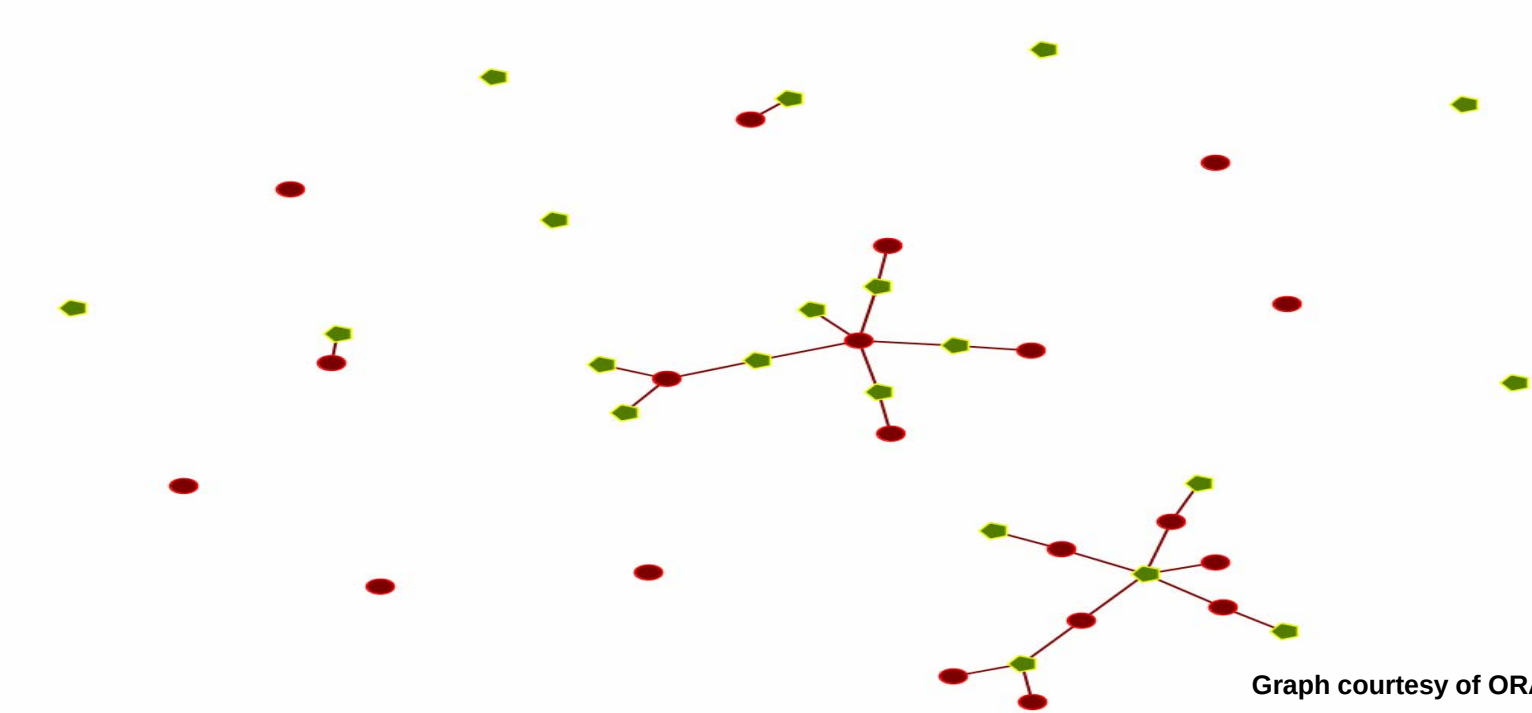
- Network in Dynamic Equilibrium
- Observations based on Underlying edge probability structure
- Maintain ergodicity
- All nodes have same probability distribution.



Graph courtesy of ORA

Erdsös-Réni Random Graph

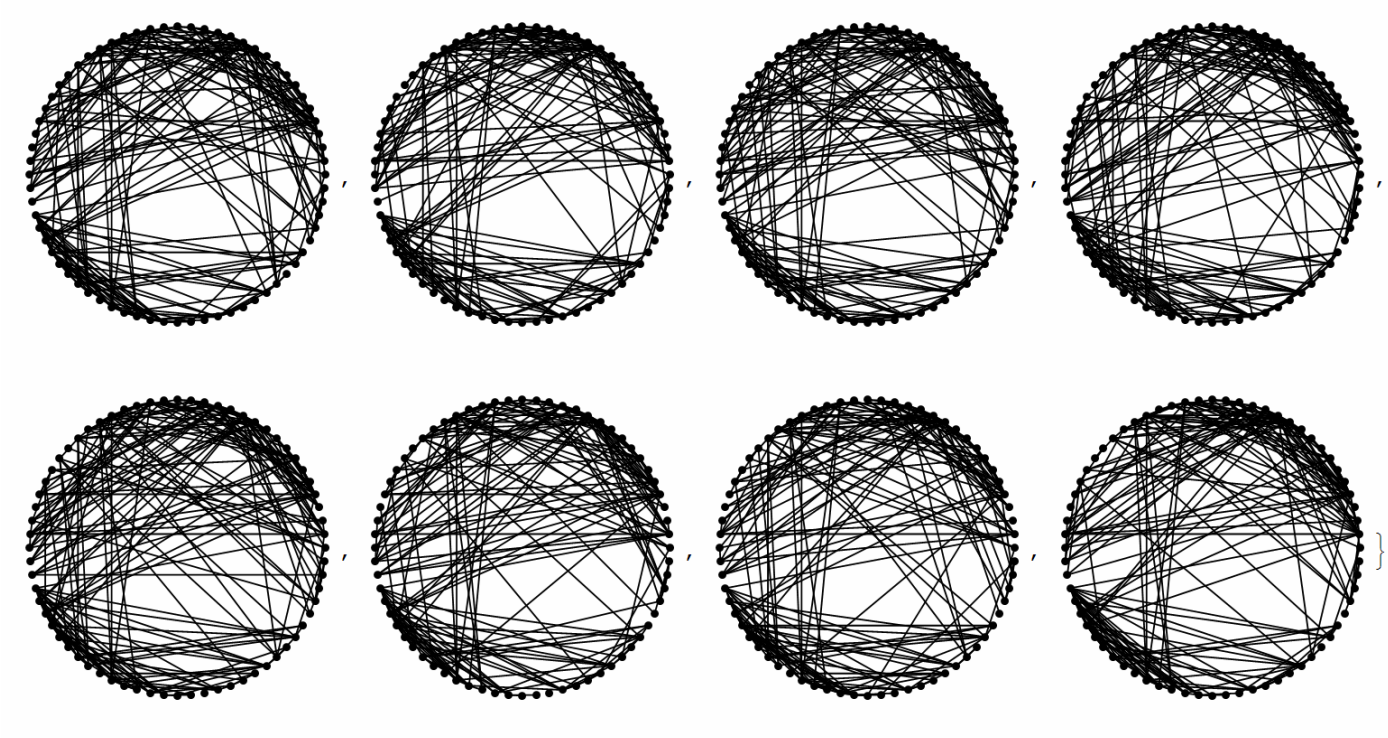
- All edges have the same probability
- Same random Bernoulli trial
- Connectivity threshold $\frac{(1-\varepsilon)\ln(n)}{n}$



Graph courtesy of ORA

Simulated Network

- Monte Carlo Simulation
- N = 100,000



The Statistical Test

- We use Hamming Distance (HD) as a measure of similarity between networks.
- We find HD for all combinations of time periods in the empirical data.
- We take all of the simulated graphs, and find HD between them and each time period from empiricals.
- Perform **T-TEST** between these two sets of HDs.
- This test tells us if the simulation (NPM/E-R) does a BETTER JOB at explaining a given time period than the rest of the empirical data does.

The Statistical Test

1. Create a vector of hamming distances between all possible combinations of **empirical** vectors. Group them by time period.
2. Create a vector of hamming distances between all simulated graphs and **each time period**, grouping the vectors by time period.
3. Perform a T-test between each corresponding vector to answer the following question:

Does the NPM/E-R, on average, **more closely match** any particular time period from the empirical data than the rest of the empirical data?

Results (NET07): NPM

- Comparison of simulated vs. empirical data

	M	*8*	*N*	*60000*		
Data Set	**Mean Hamming Distance to Empirical Networks**	**Standard Deviation of Hamming Distance to Empirical Networks**	**Mean Hamming Distance to Simulated Networks**	**Standard Deviation of Hamming Distance to Simulated Networks**	**t-test**	**p-value**
1	409.286	38.560	358.094	12.775	3.755	0.00
2	365.857	18.298	320.097	12.739	7.073	0.00
3	365.857	29.043	320.164	12.793	4.450	0.00
4	377.857	38.247	330.674	12.773	3.489	0.00
5	375.286	36.100	328.377	12.796	3.675	0.00
6	349.857	38.159	306.078	12.785	3.245	0.00
7	373.8571	48.45076	327.0728	12.82622	2.731135	0.01
8	362.4286	55.63529	317.1509	12.77754	2.301849	0.02

Results (NET07): E-R

- Comparison of simulated vs. empirical data

	M	*8*	*N*	*60000*		
Data Set	**Mean Hamming Distance to Empirical Networks**	**Standard Deviation of Hamming Distance to Empirical Networks**	**Mean Hamming Distance to Simulated Networks**	**Standard Deviation of Hamming Distance to Simulated Networks**	**t-test**	**p-value**
1	409.286	38.560	1127.379	17.41762	-3.9167	0.00
2	365.857	18.298	1116.303	21.54558	-4.23399	0.00
3	365.857	29.043	1193.895	18.60198	-3.73844	0.00
4	377.857	38.247	1252.086	16.82216	-4.40049	0.00
5	375.286	36.100	1169.254	18.88182	-3.64695	0.00
6	349.857	38.159	1209.797	17.59757	-3.60082	0.00
7	373.8571	48.45076	1110.78	17.31786	-3.44968	0.00
8	362.4286	55.63529	1192.288	17.44347	-3.461	0.00

Results (NET05): NPM

	M	*9*	*N*	*60000*		
Data Set	**Mean Hamming Distance to Empirical Networks**	**Standard Deviation of Hamming Distance to Empirical Networks**	**Mean Hamming Distance to Simulated Networks**	**Standard Deviation of Hamming Distance to Simulated Networks**	**t-test**	**p-value**
1	1445	84.774	1284.338	23.747	3.467	0.001
2	1394.75	67.487	1239.647	23.703	3.765	0.000
3	1296.125	85.436	1151.946	23.671	3.287	0.001
4	1315.875	153.533	1169.665	23.718	2.421	0.015
5	1191.25	112.324	1058.99	23.667	2.732	0.006
6	1204.875	207.944	1071.116	23.623	1.912	0.056
7	1167.375	190.431	1037.713	23.695	1.98	0.048
8	1159.625	204.465	1030.815	23.732	1.888	0.059
9	1170.125	195.266	1040.142	23.618	1.953	0.051

Results (NET05): E-R

	M	*9*	*N*	*60000*		
Data Set	**Mean Hamming Distance to Empirical Networks**	**Standard Deviation of Hamming Distance to Empirical Networks**	**Mean Hamming Distance to Simulated Networks**	**Standard Deviation of Hamming Distance to Simulated Networks**	**t-test**	**p-value**
1	1445.000	84.774	2253.82	34.26138	-6.8034	0.00
2	1394.750	67.487	2232.07	41.48661	-7.46798	0.00
3	1296.125	85.436	2385.99	35.58944	-6.47687	0.00
4	1315.875	153.533	2503.9	32.87007	-7.80098	0.00
5	1191.250	112.324	2336.64	36.9779	-6.2939	0.00
6	1204.875	207.944	2419.19	34.87729	-6.20163	0.00
7	1167.375	190.431	2219.81	33.75171	-5.89936	0.00
8	1159.625	204.465	2383.33	33.89981	-5.99199	0.00
9	1170.125	195.266	2453.82	36.2168	-7.1034	0.00

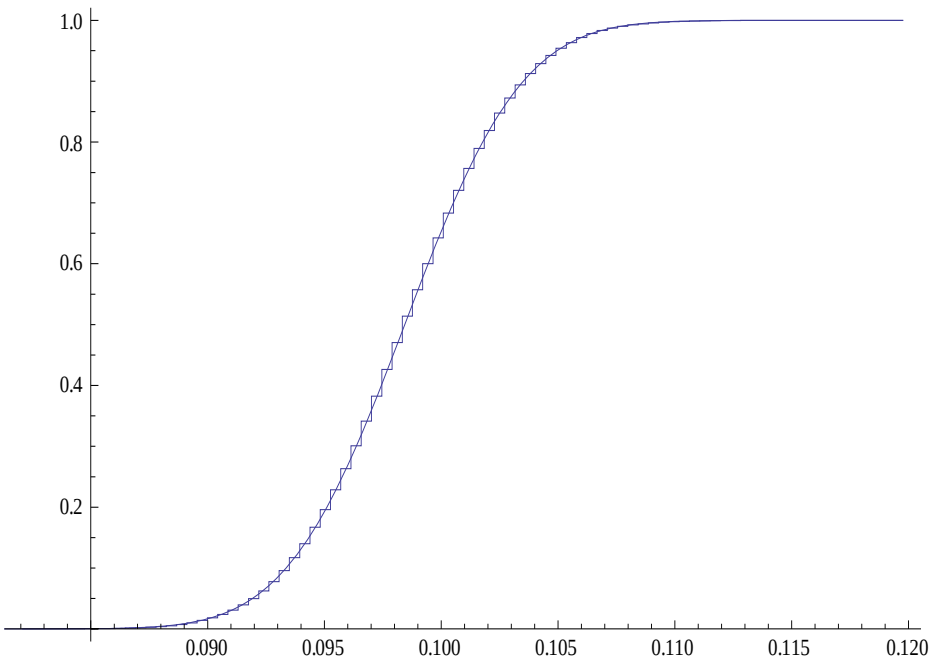
Significance

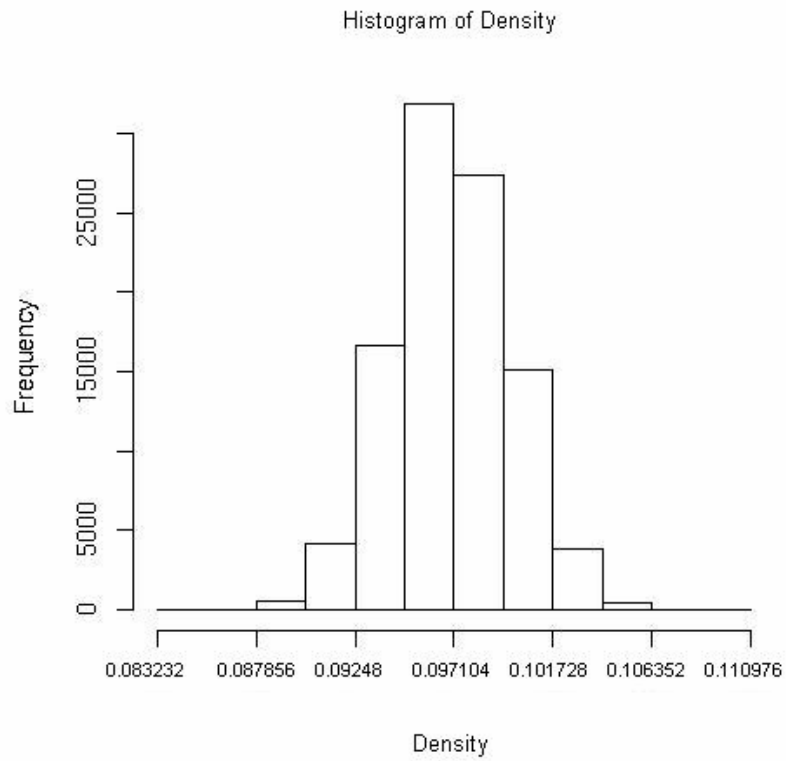
- NPM performs well, E-R does not.
  - Why? (Net-07 Clustering)
- Since the NPM does a good job of representing the laws which govern the network, we can use simulation to:
  - Explore large numbers of “instances” of the graphs
  - Create distributions of network and agent-level measures

- With a validated simulation, we facilitate further statistical analysis of the network and its measures!
  - Statistical Process Control: When has the network undergone a significant change?
  - Percolation: What is the likelihood that a rumor/ideology/belief spreads throughout the network?

Results (NET07)

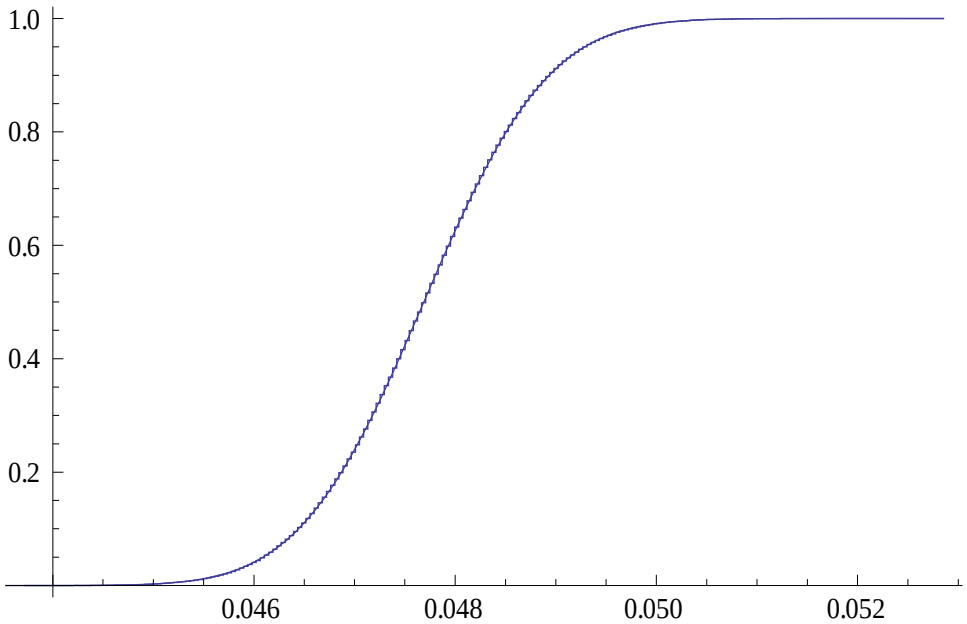
- Fit a normal distribution to densities

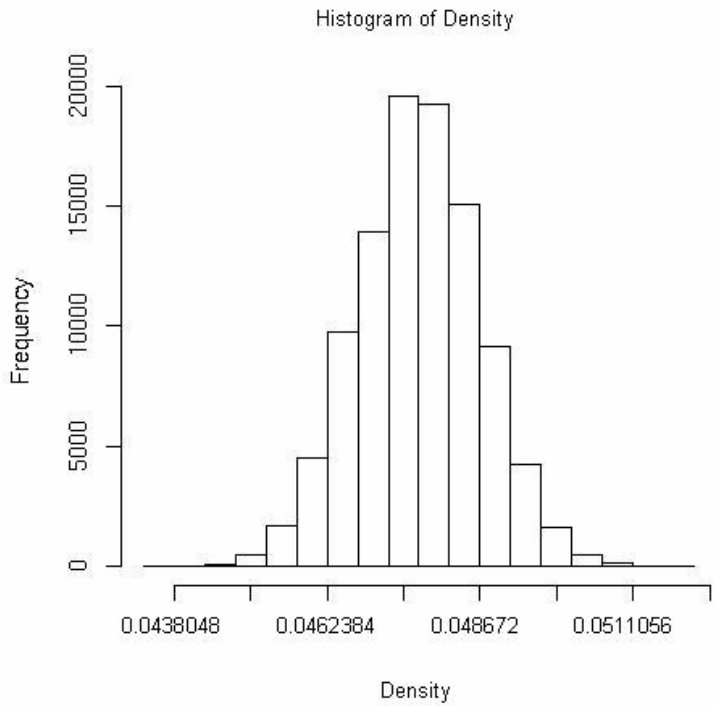




Results (NET05)

- Fit a normal distribution to densities





Conclusion

- When highly complex systems are being simulated, and empirical data is available, we can use this methodology to test whether our simulation is *at least as close* to each time series in a data set as the rest of the time periods are.
  - Which model (Erdos, Watts, Barabasi, NPM) most accurately describes the empirical data?

- The “simple case” of the NPM is shown to be a viable explanation of social networks.

Acknowledgements

- U.S. Army Research Institute
- U.S. Military Academy Network Science Center
- U.S. Army Research Labs
- Dr. Kathleen Carley
- COL Steve Horton
- LTC John Graham
- MAJ Tony Johnson
- MAJ Ian McCulloh
- Dr. Craig Schrieber