ERGM Specifications
- social circuit models

Peng Wang

SWINBURNE UNIVERSITY OF TECHNOLOGY

The good news

GREAT!

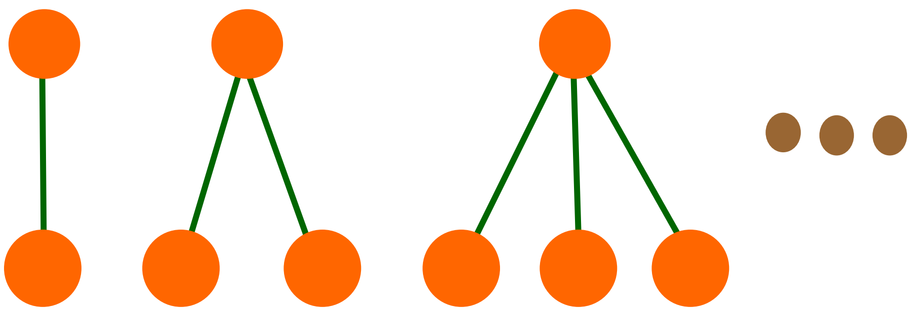
- Markov random graph distributi… statistical models for social networks based on plausible assumptions and importantly can represent clustering through the triangle parameter!

The BAD news

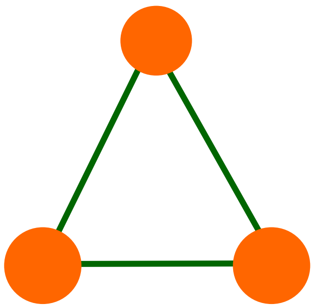
- THEY DON’T ALWAYS WORK!
  - For some parameter values they are just not coherent models (they can imply two entirely different types of graphs)
  - They have great problems in attaining sensible parameter estimates when clustering is high.

The BAD news

Often for **Markov model**



degree distribution; preferential attachment, etc



friends meet through friends; clustering; etc

Matching model $E_{\hat{q}_{MLE}}\{z(X)\} = z(x_{obs})$ **With data**

hard

or

Impossible!

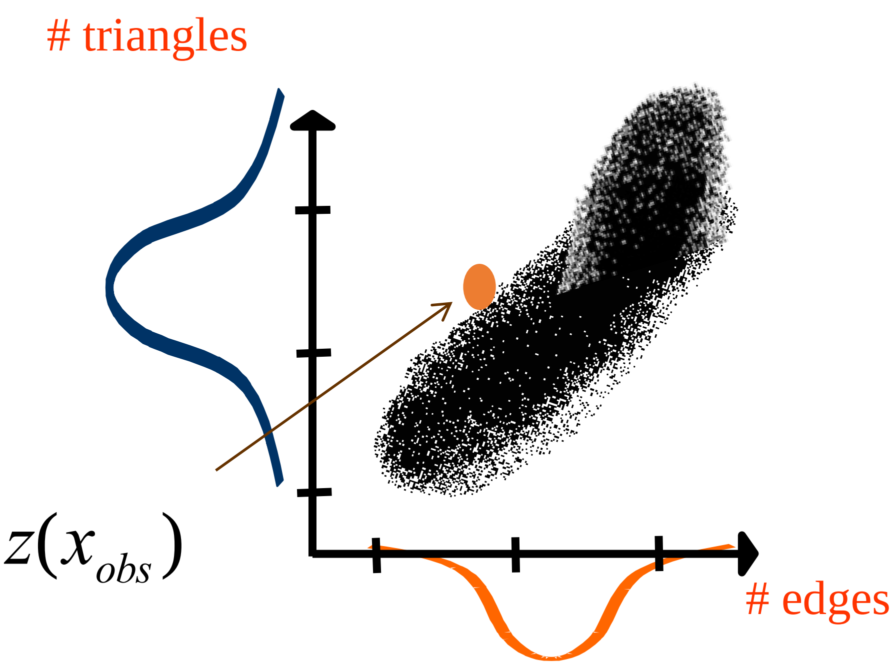
The BAD news

Matching model $E_{\hat{q}_{MLE}}\{z(X)\} = z(x_{obs})$ **With data**

hard

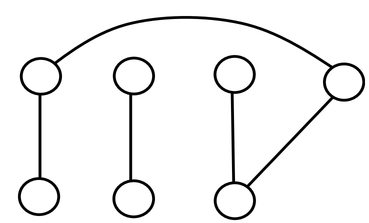
or

Impossible!



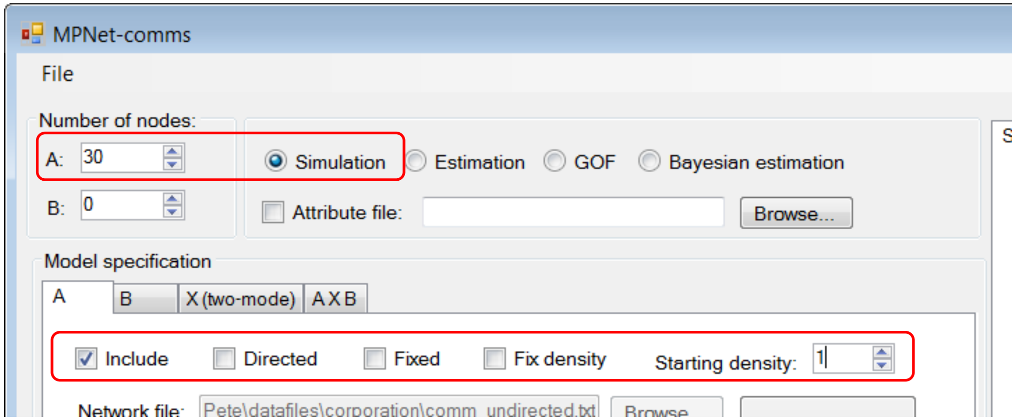
Some statistic *k*

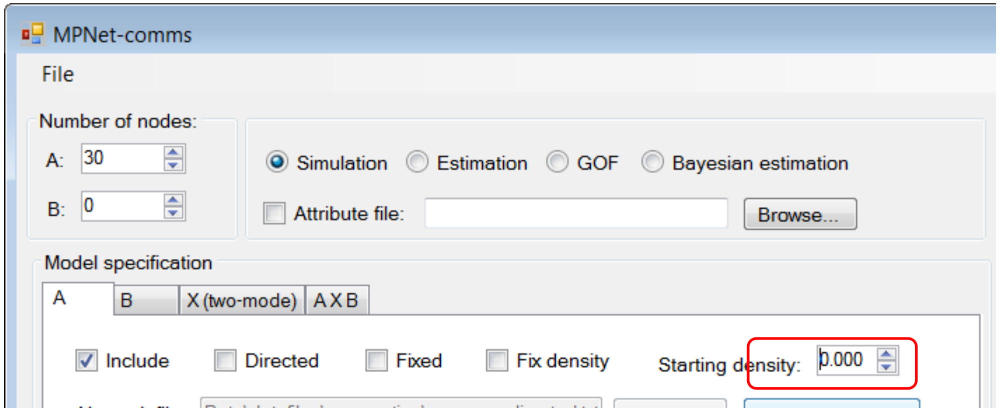
$z_k(x_{obs}) = 0$ or $z_k(x_{obs}) = z_k^{\max}$



Exercise:
Simulating an edge/triangle model

- Simulate a model on 30 nodes with only the following parameters:
  - Edge = - 3, triangle = 0.75
- In two ways:
  - Starting graph density = 1
  - Starting graph density = 0
- Draw the final graph in Pajek.





Example of a degenerate model:
Edge/Triangle model for 30 nodes:

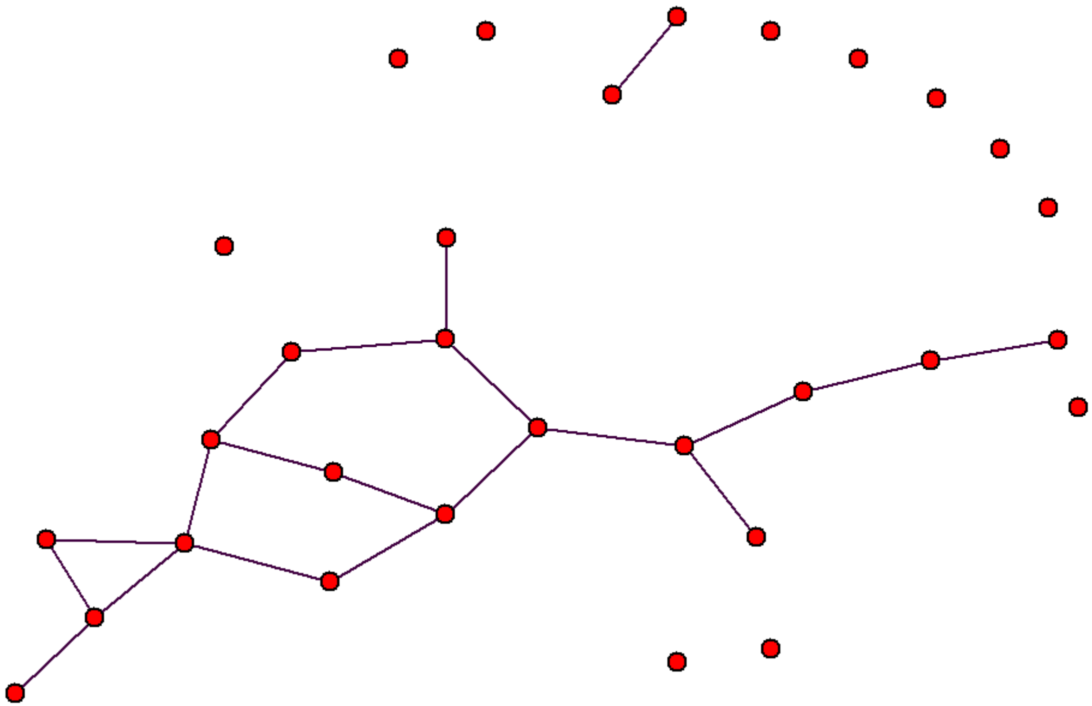
θ = - 3; τ = **0.75**

Start from empty graph

Burn-in 50,000

200,000 simulations

Final graph



Example of a degenerate model:
Edge/Triangle model for 30 nodes:

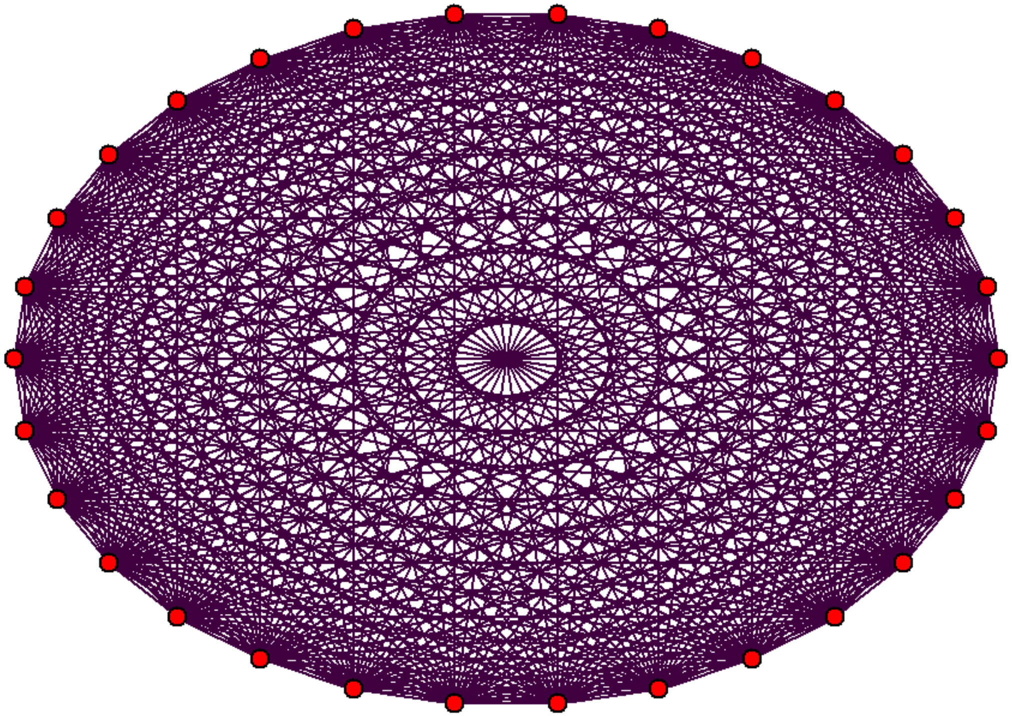
θ = - 3; τ = **0.75**

Start from complete graph

Burn-in 50,000

200,000 simulations

Final graph

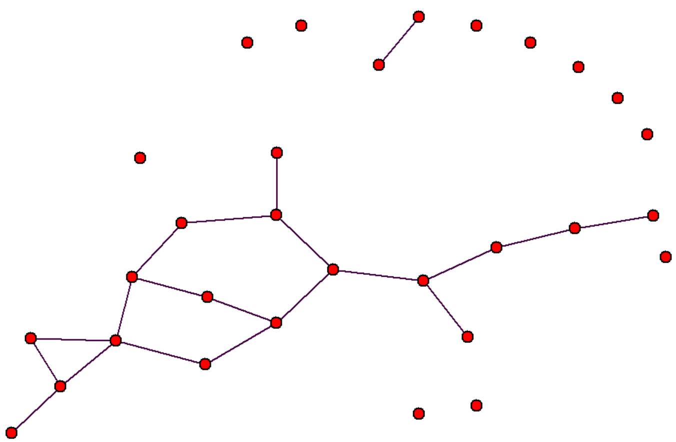


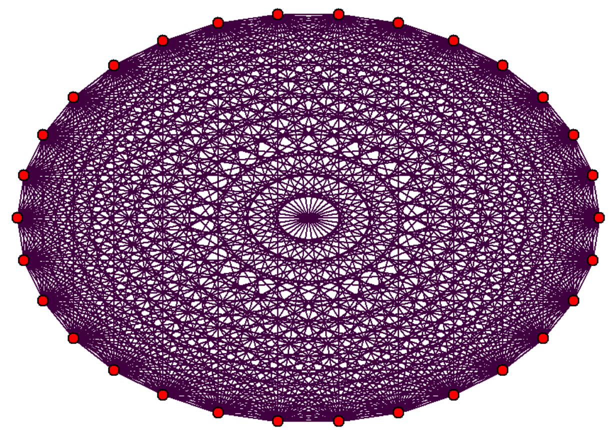
Edge/Triangle model:

θ = - 3; τ = 0.75

SAME Parameters

COMPLETELY DIFFERENT GRAPHS!





Model degeneracy

For certain parameter values, a model may imply that only one or two graphs with non-zero probability.

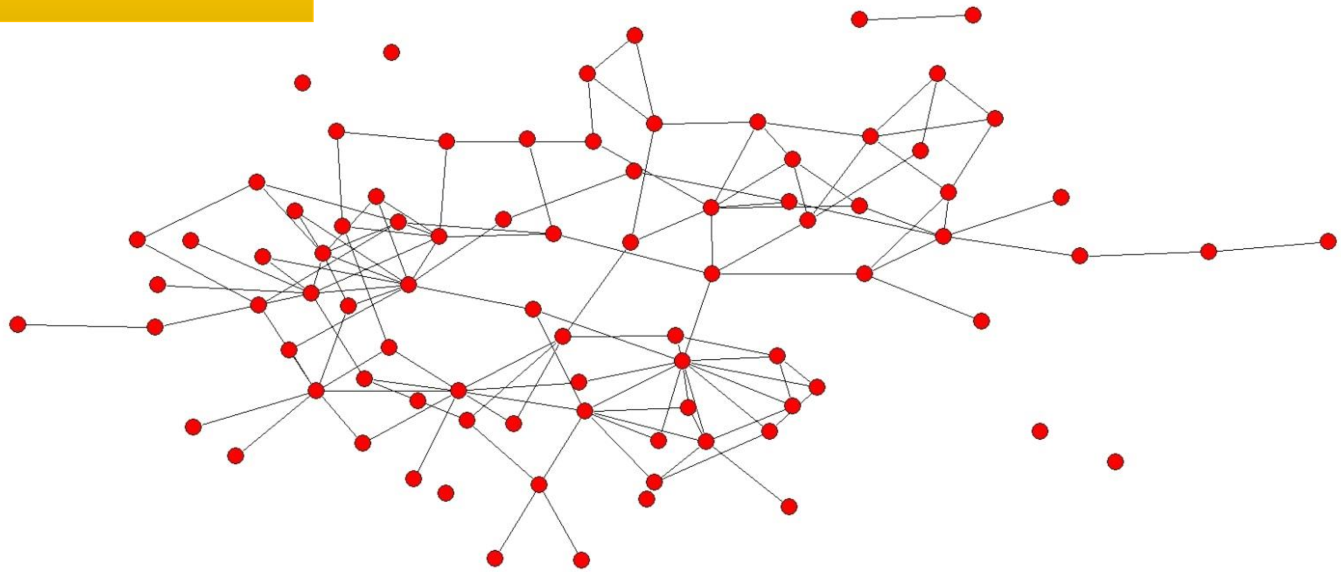
Often such graphs are the empty or full graph

> (or a graph of complete disconnected components).

These *near degenerate* models cannot be estimated.

Exercise:
Markov models often have convergence problems

The fishermen's network:
85 nodes



Try to estimate an edge/2-star/3-star/triangle Markov model.
You will find the convergence ratios very bad

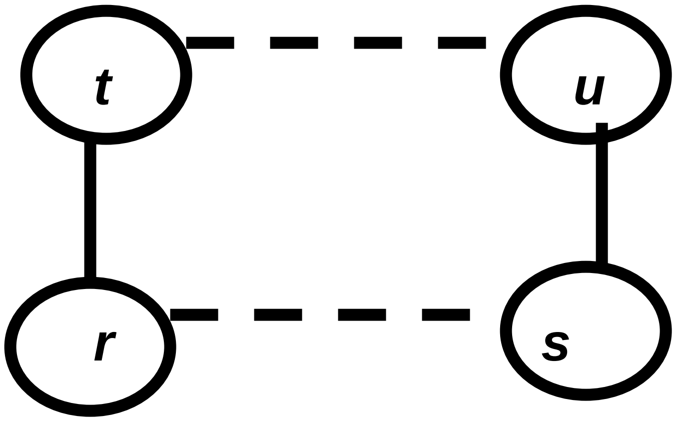
Social circuit dependence

Network ties self-organize within 4-cycles.

(Pattison & Robins, 2002; Snijders et al, 2006)

Two possible network ties are conditionally dependent if they would form a 4-cycle.

Social circuit dependence



Configurations/model parameters for social circuit models

Parameters correspond to configurations of the following types:

Edges

Stars

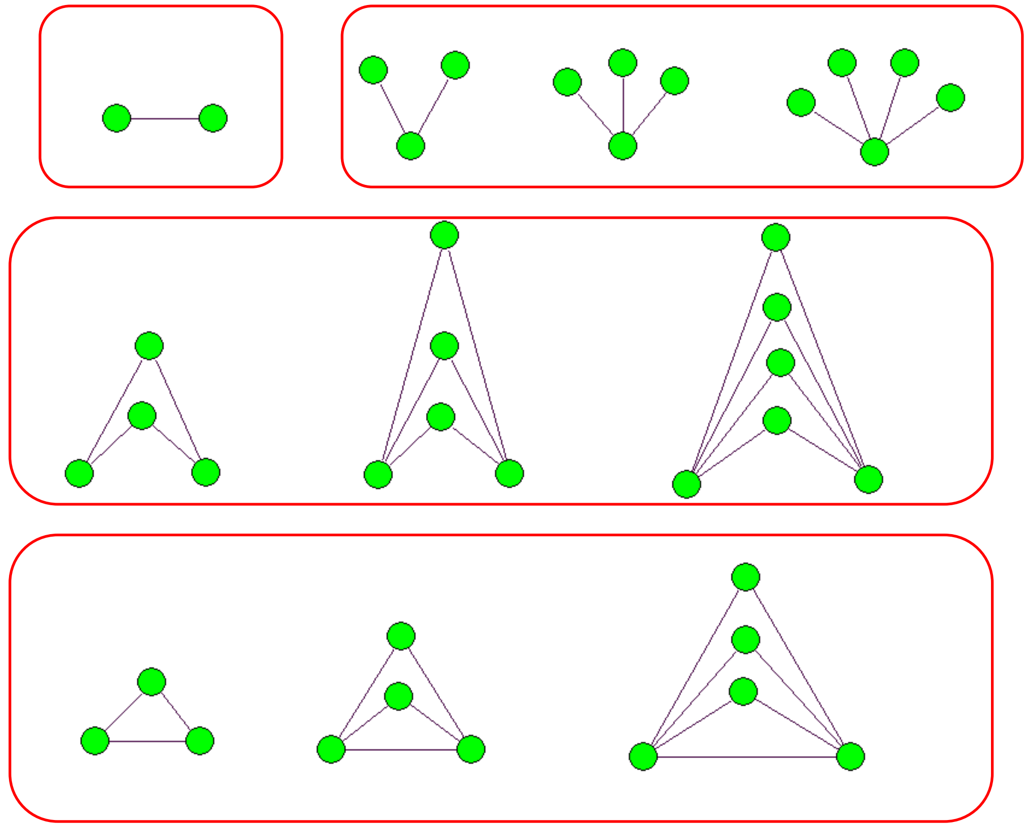
(*k*-stars)

Multiple paths

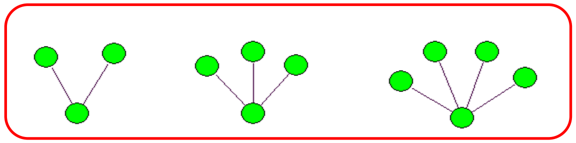
(*k*-2-paths)

Multiple triangles

(triangulation)



Centralization parameter: Alternating *k*-stars



$$z(\mathbf{x}) = S_2 - \frac{S_3}{\lambda} + \frac{S_4}{\lambda^2} - \ldots + (-1)^{n-2}\frac{S_{n-1}}{\lambda^{n-3}}$$

Usually we set λ = 2. Hunter & Handcock (2006) show how to estimate lambda

Interpretation:

Positive parameter indicates centralization through a small number of high degree nodes

core-periphery based on popularity

More dispersed degree distribution

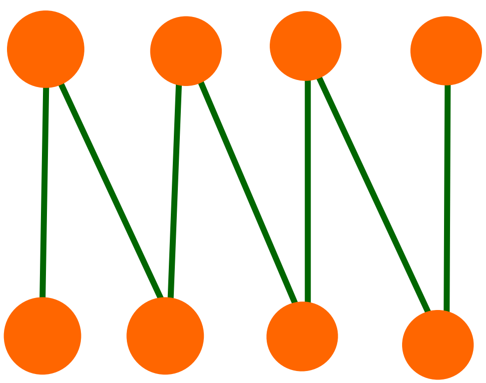
Negative parameter: “truncated” (less dispersed) degree distribution; nodes tend not to have particularly high degrees.

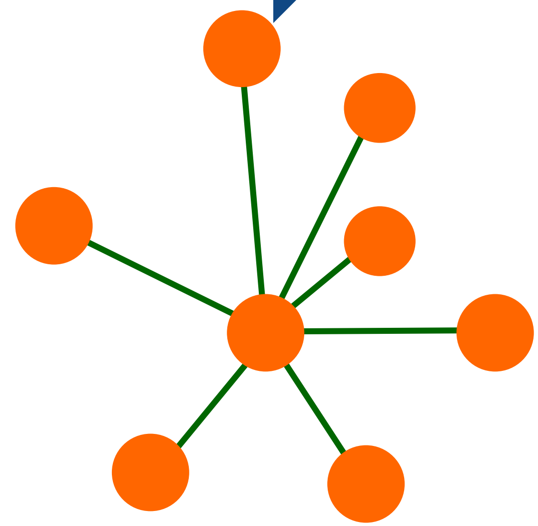
Equivalent to **geometrically weighted degree distribution parameter** (Hunter, 2007)

SWIN BUR NE

Centralization parameter: Alternating *k*-stars

Interpreting the alternating star **parameter**:





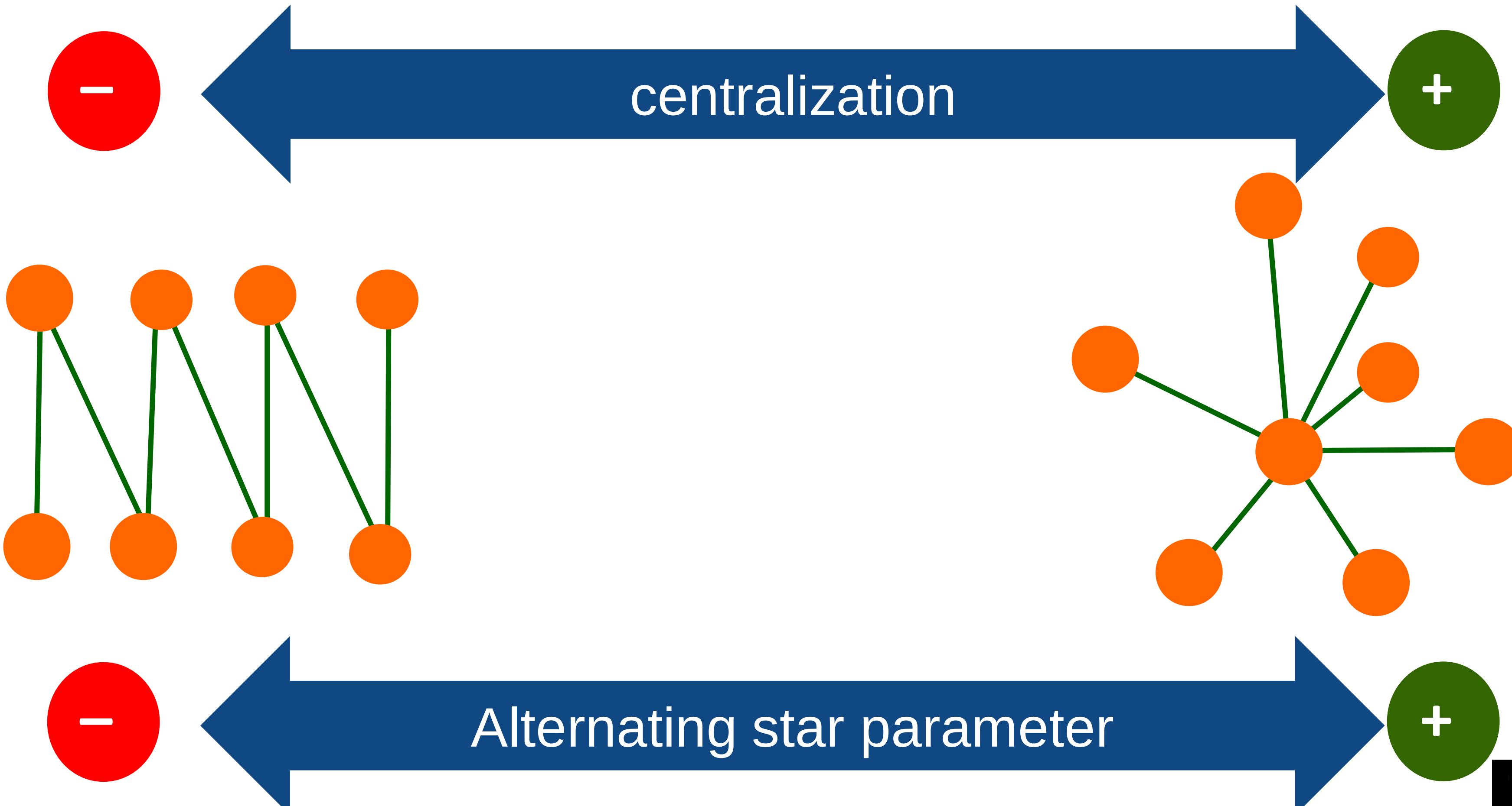
$$Var(x_{i+}) = \sum_i \frac{(x_{i+} - \bar{x})^2}{n - 1} = .21$$

$$Var(x_{i+}) = 4.42$$

variance of degree measure centralization

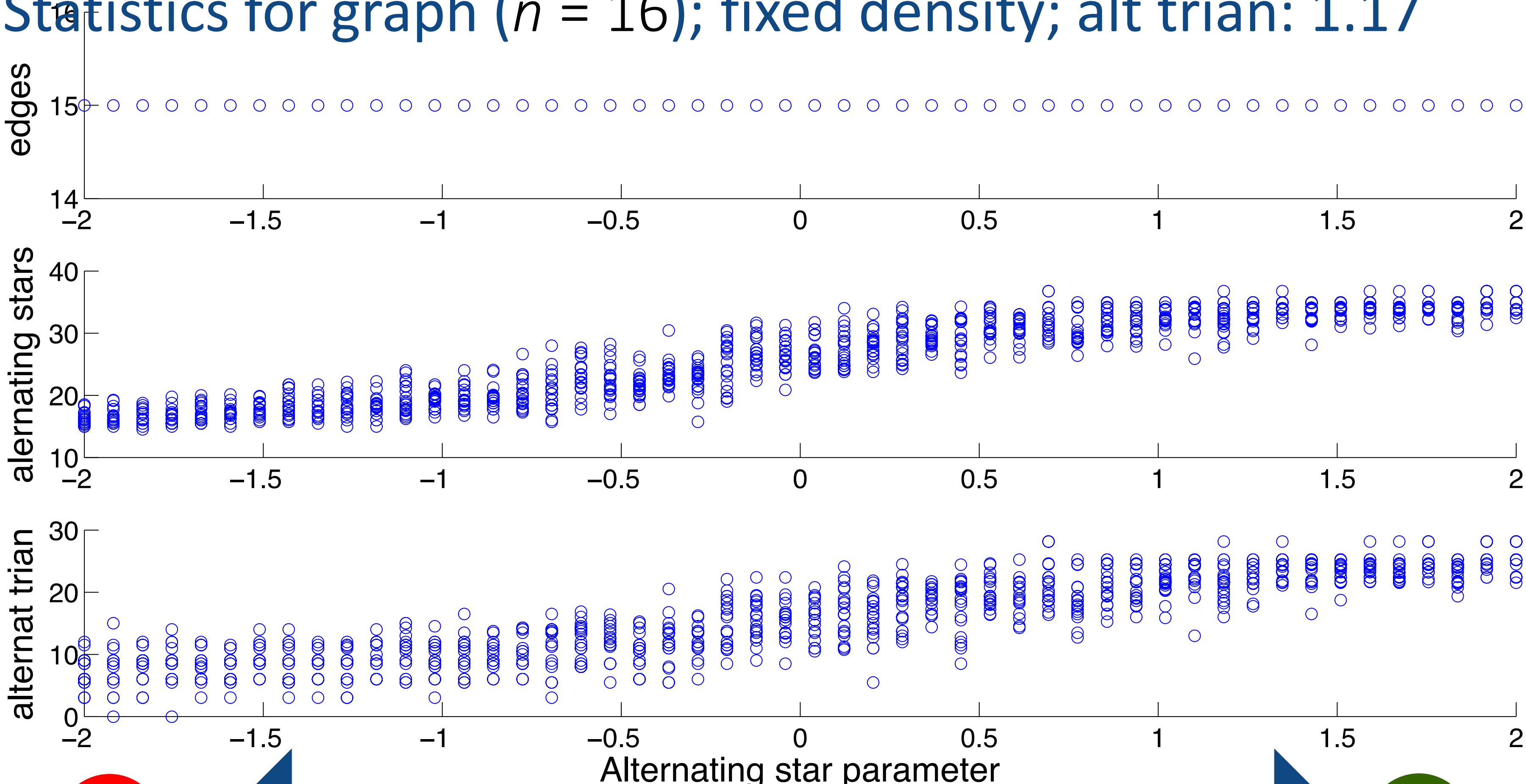
Centralization parameter: Alternating *k*-stars

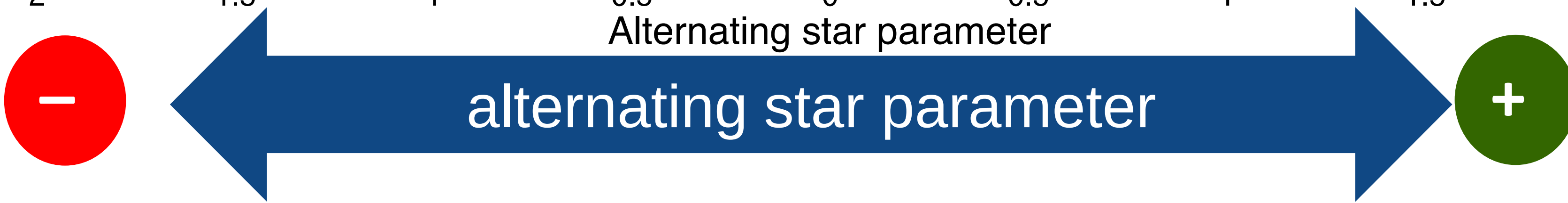
Interpreting the alternating star **parameter**:



Centralization parameter: Alternating *k*-stars

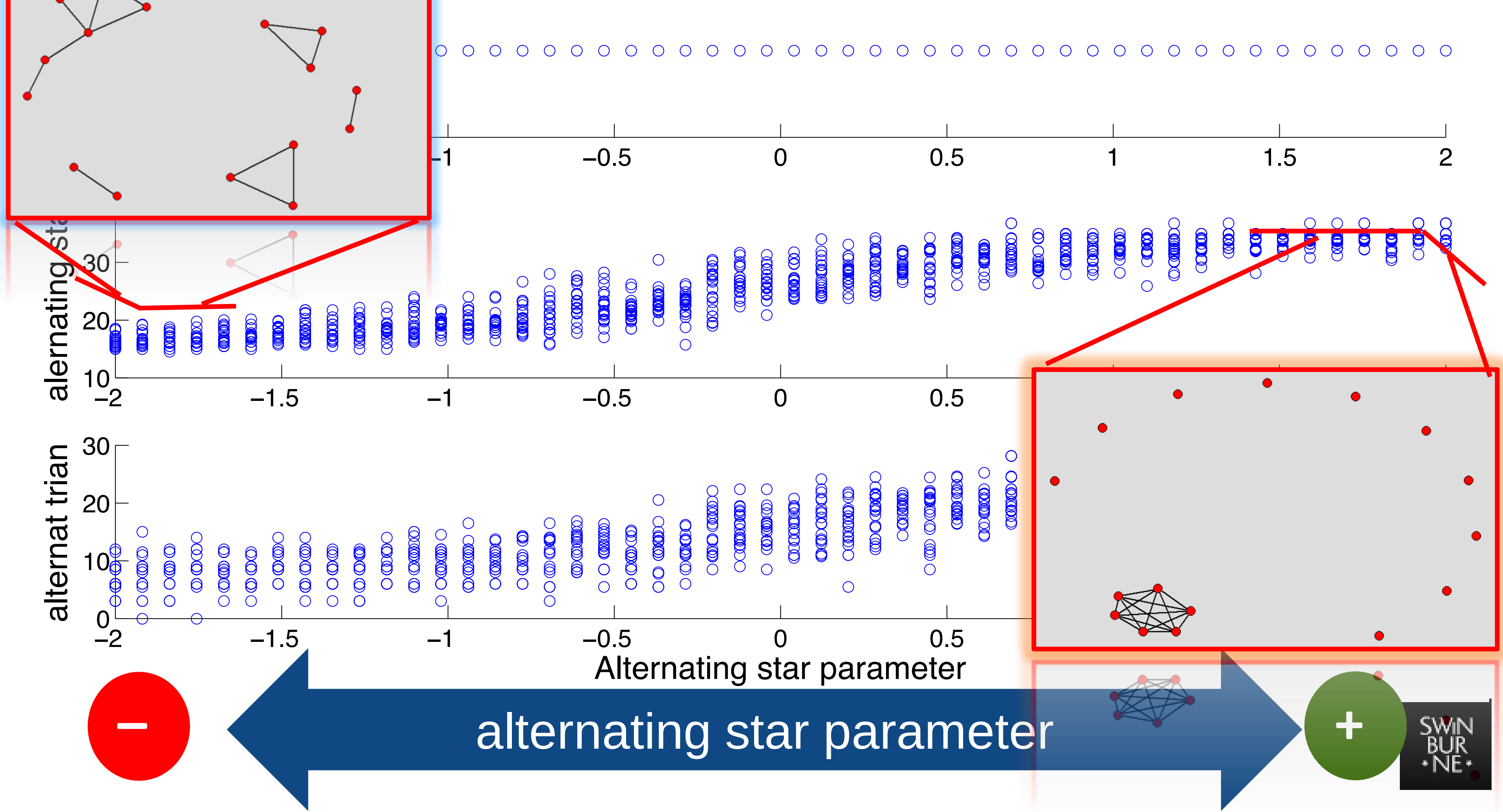
Statistics for graph (n = 16); fixed density; alt trian: 1.17





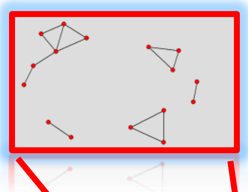
Centralization parameter: Alternating *k*-stars

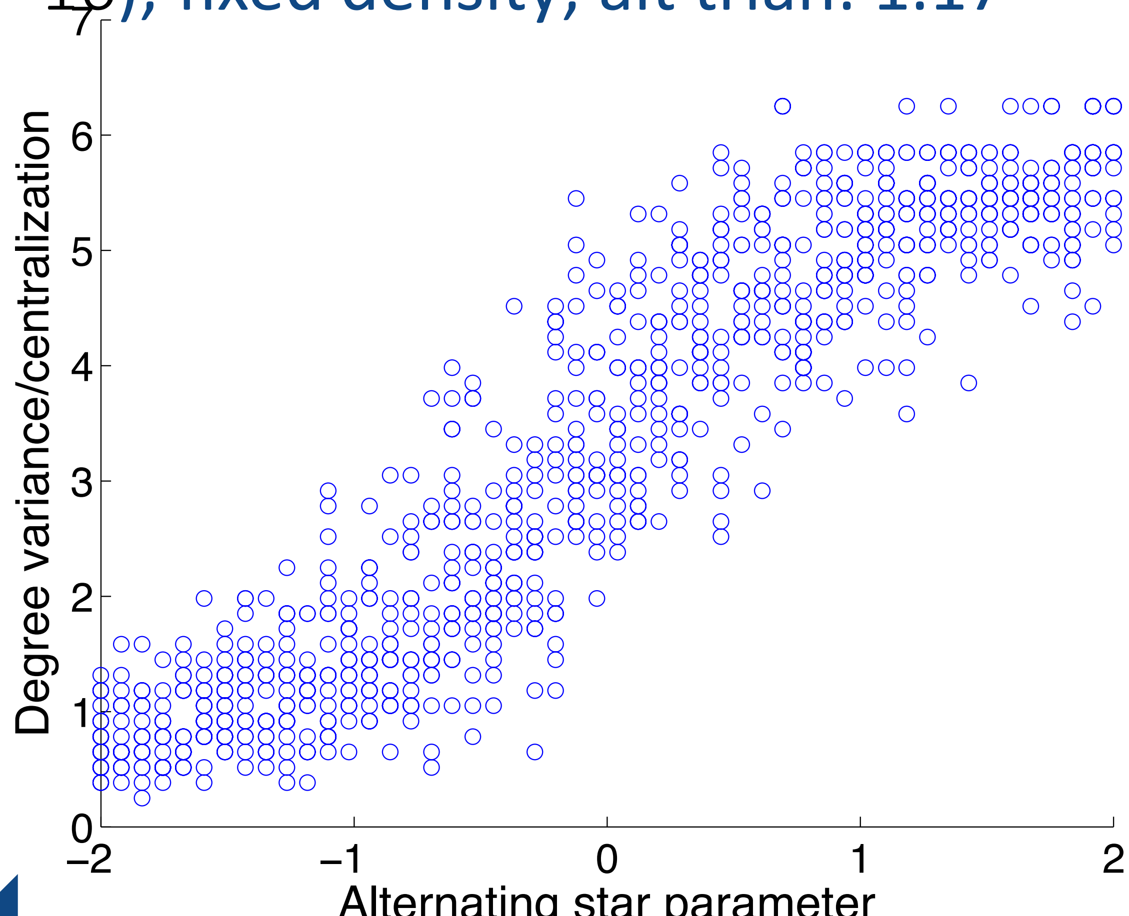
Statistics for graph (n = 16); fixed density; alt trian: 1.17

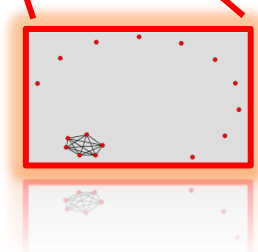


Centralization parameter: Alternating *k*-stars

Graphs (n = 16); fixed density; alt trian: 1.17







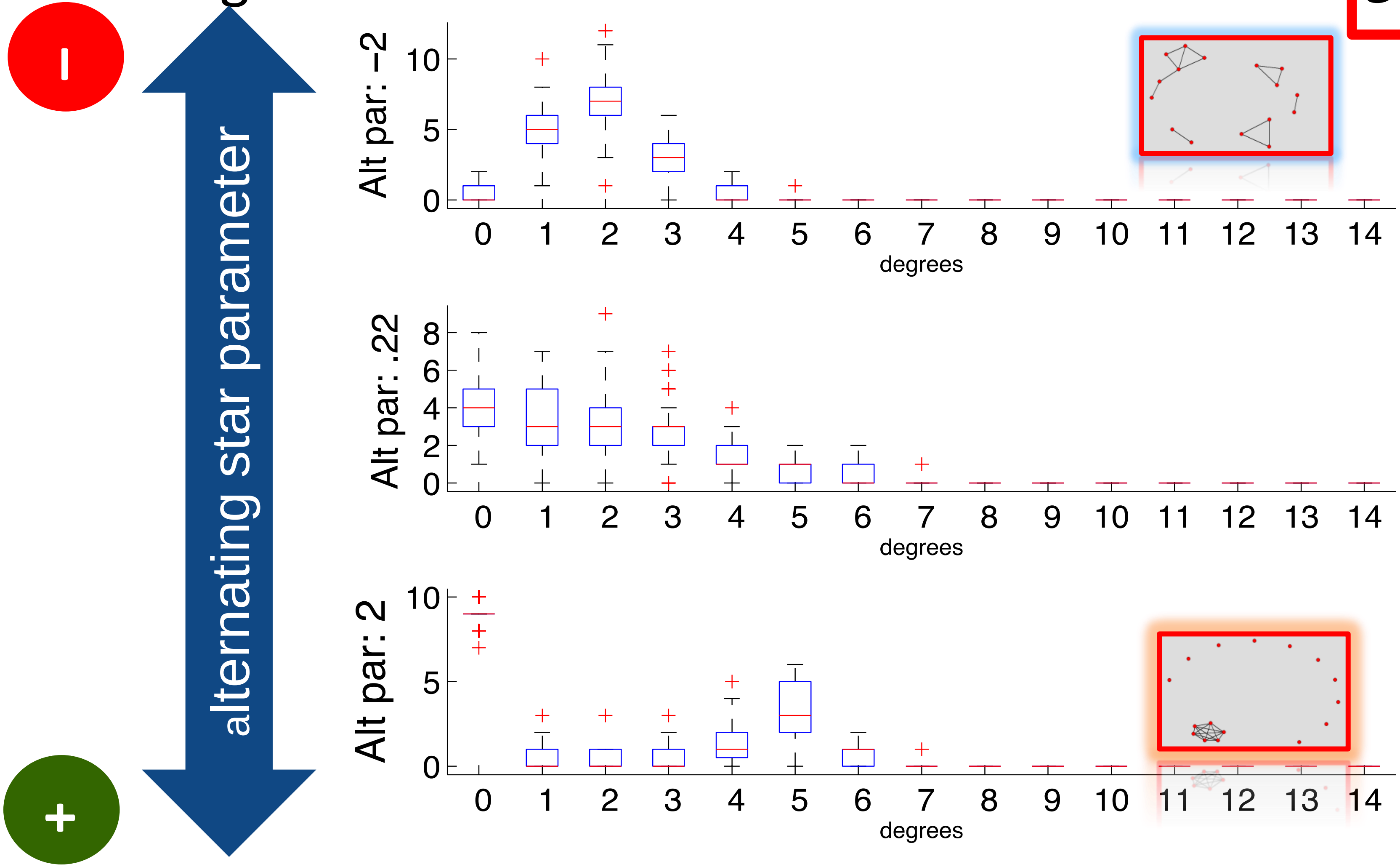
−

alternating star parameter

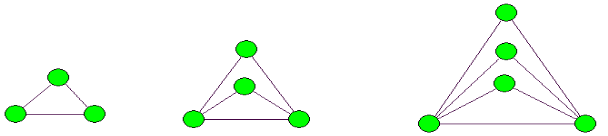
+

Centralization parameter: Alternating *k*-stars

The degree distribution for different values of: s_{AKS}



Network Closure parameter: Alternating *k*-triangles



$$u(\mathbf{y}) = T_1 - \frac{T_2}{\lambda} + \frac{T_3}{\lambda^2} - \ldots + (-1)^{n-2}\frac{T_{n-2}}{\lambda^{n-3}}$$

Interpretation:

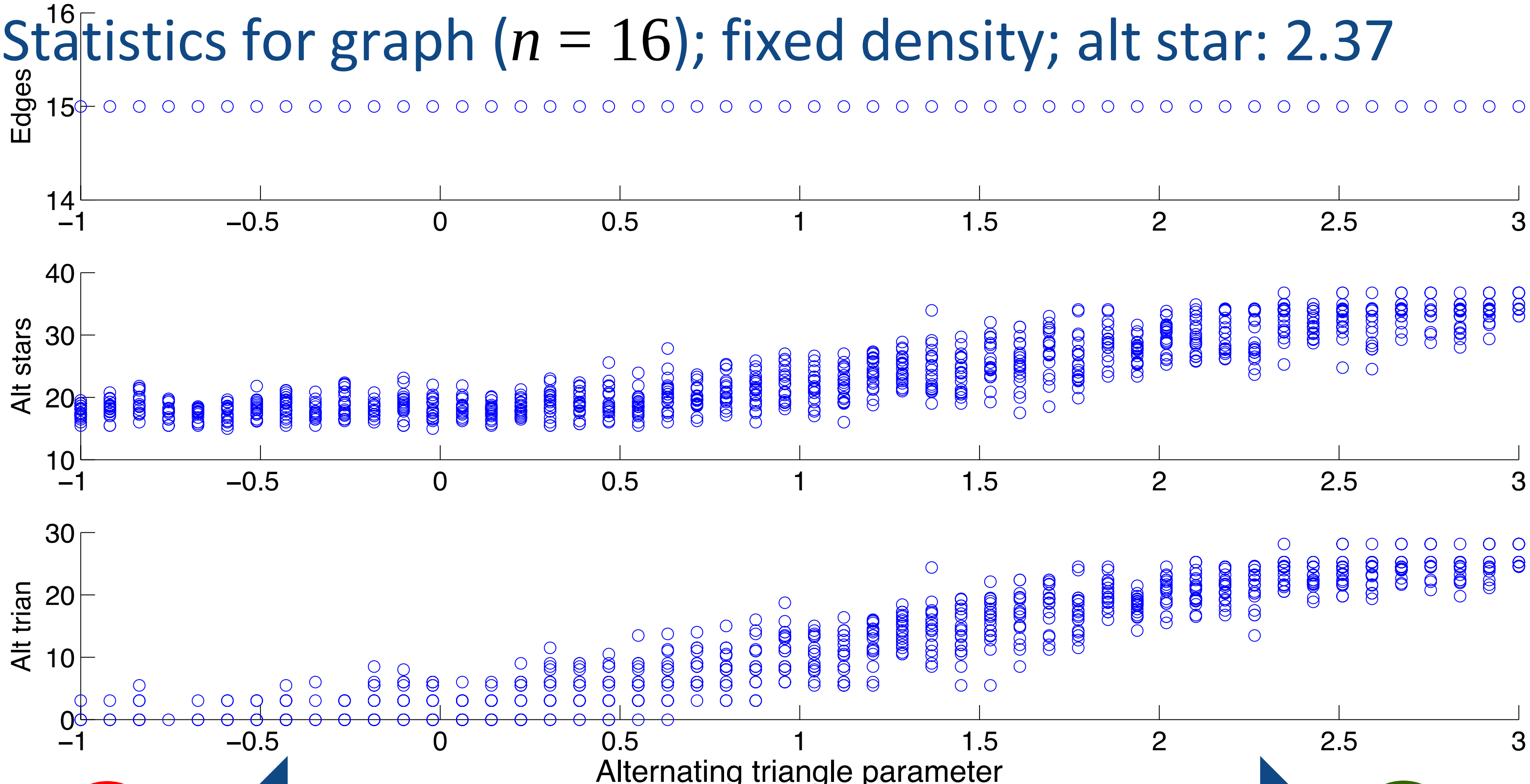
a. Positive parameter suggests triangles tend to "clump" together in denser regions of the network.

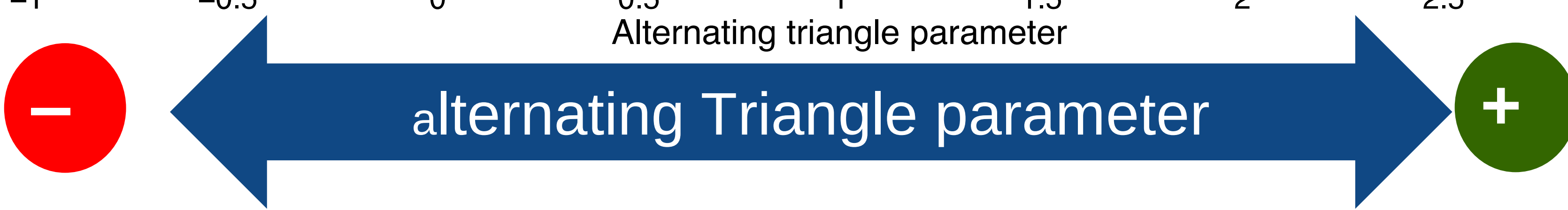
b. Models the ***edgewise shared partner*** **distribution**:

For each pair of tied nodes, how many partners do they share? (Hunter, 2007)

Network Closure parameter: Alternating *k*-triangles

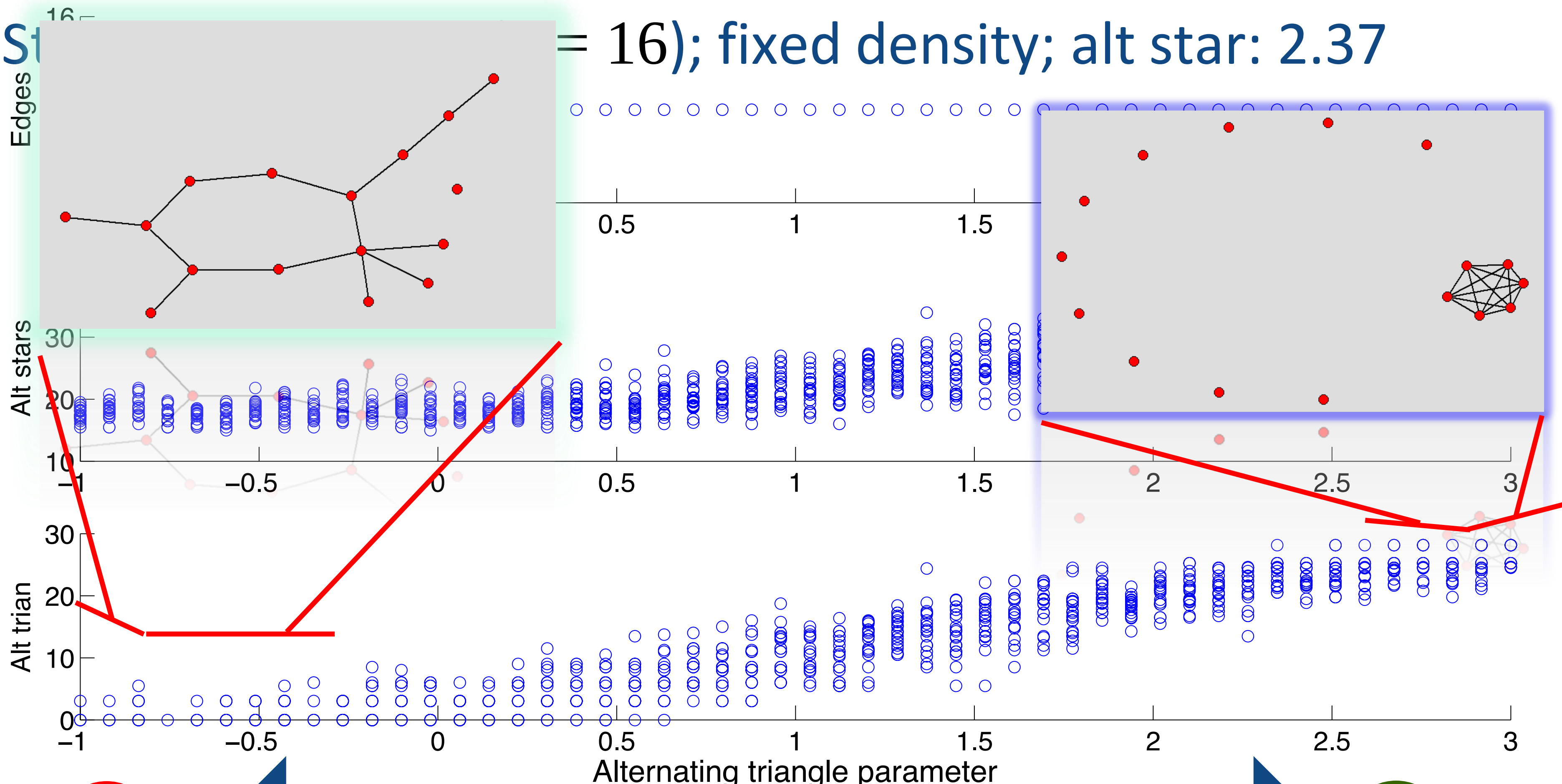
Statistics for graph ($n = 16$); fixed density; alt star: 2.37

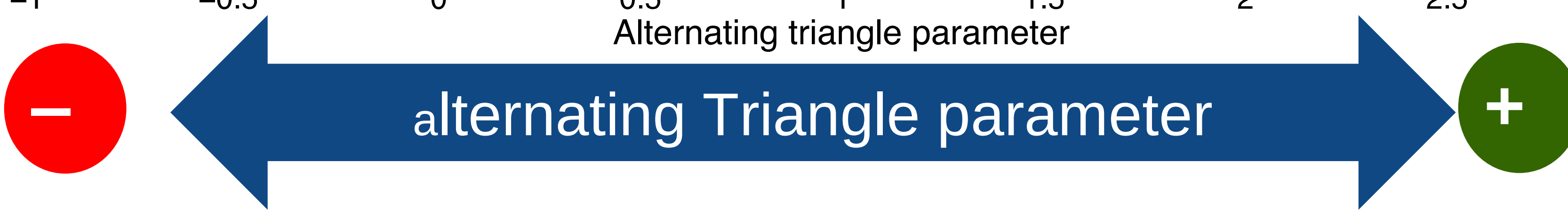




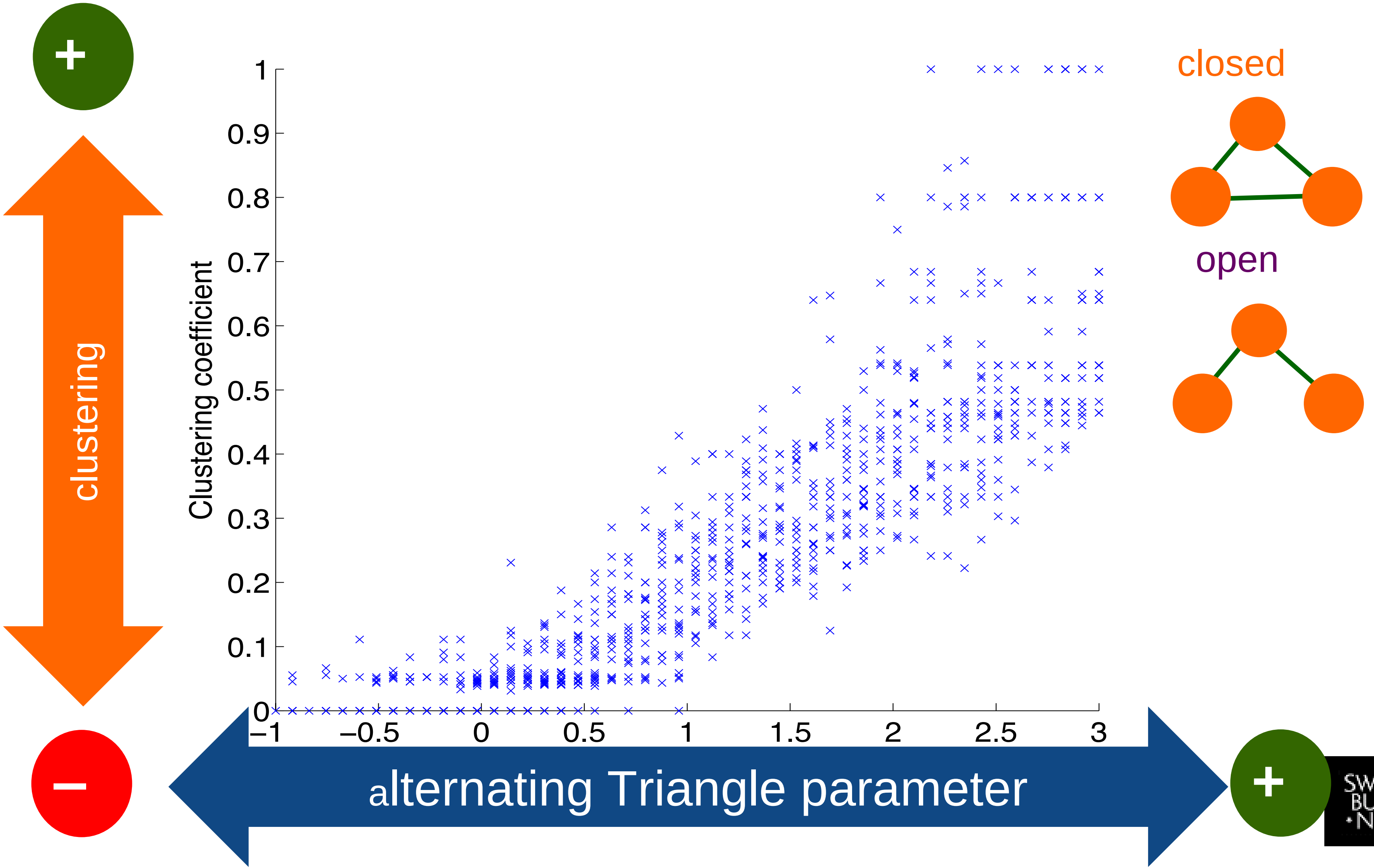
Network Closure parameter: Alternating *k*-triangles

St… = 16); fixed density; alt star: 2.37

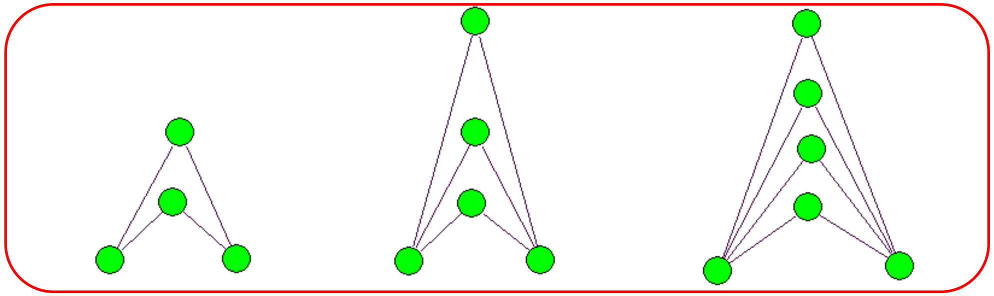




Network Closure parameter: Alternating *k*-triangles



Connectivity Parameter: Alternating *k*-2paths



Interpretation:

a. Localized structural equivalence
b. With the AKT parameter in the model, this indicates presence of structural holes.

c. Models the ***pairwise shared partner* distribution**:
For each pair of nodes (tied or not), how many partners do they share? (Hunter, 2007)

Social circuit dependence

Larger network configurations emerge:

Parameters for **degree sequences**, **denser regions of triangulation**, **multiple connectivity**.

These models fit much better.

A dependence assumption that captures emergence may be necessary to model real social networks.

(Robins, Snijders, Wang, Handcock & Pattison, 2007)

ERGM - goodness of fit test

Peng Wang

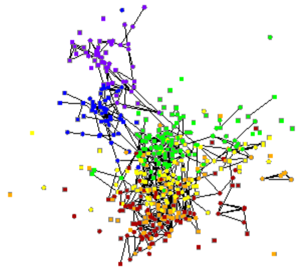
SWINBURNE UNIVERSITY OF TECHNOLOGY

Model goodness of fit (GOF) test

- Estimate parameters for a model
- Using these parameter values, simulate a distribution of graphs consistent with the model
- From a sample of these graphs, collect graph statistics: degrees, closure, clique structure, geodesics, etc
- Compare the observed data with the collected statistics:
  - If the data is not extreme (e.g. $|t| < 2.0$), then the model plausibly explains that feature of the data
  - For parameters in the model, we want the data to be central in the distribution (say, $|t| < 0.2$), else model may not have converged.

Goodness of fit

DATA



MODEL

MODEL PARAMETER ESTIMATES

SIMULATE A DISTRIBUTION OF GRAPHS

INSPECT THE GRAPH STATISTIC FROM DATA

TAKE A SAMPLE OF SIMULATED GRAPHS AND INSPECT DISTRIBUTION OF ANY GRAPH STATISTIC OF INTEREST

IS THE DATA EXTREME IN THE DISTRIBUTION?

YES: THIS MODEL CANNOT EXPLAIN THIS STATISTIC

NO: AS FAR AS THIS GRAPH FEATURE IS CONCERNED THE DATA COULD HAVE COME THIS MODEL

GOODNESS OF FIT

Goodness of fit (GOF) in PNet

- After estimating the model, select the GOF tab
- Choose the number of nodes and network file as usual
- Under Select Parameters, select the Markov parameters, ASA, ATA, A2P
  - You can select other parameters if you wish
- On the main window, click ‘Update!’. This uses the parameter values from the last estimation.
- Then ‘Start!’ the simulation

SWIN BUR NE

Goodness of fit (GOF) in MPNet

- The criteria:
  - If it is a fitted effect in the model (i.e. a parameter used in estimation), then we want a small t-value
    - Preferably less than 0.1, or at least less than 0.2 or 0.3
  - Otherwise we should be a sceptical about whether the model has converged and go back to estimation
  - If it is NOT a fitted effect, then an absolute t-value of less than 2 suggests the model plausibly recaptures that feature of the data.

Fitting models and GOF

- Fit social circuit models to:
  - Fishermen network
  - Communication network
- In each case, run GOF to see which aspects of the data the model does not explain well.

Fisherman - Estimation

effects	estimates	stderr	t-ratio	
edge	-3.346527	0.52166	0.08211	*
AS(2.00)	-0.263609	0.21143	0.08706	
AT(2.00)	0.799427	0.12803	0.09738	*
A2P(2.00)	0.044543	0.03452	0.09187	

NB: ERGM are stochastic models so your estimates may differ slightly from those shown above

Fisherman – GOF

effects	observed	mean	stddev	t-ratio
edge	130	128.562	22.018	0.065
2-star	503	480.605	205.762	0.109
3-star	836	783.046	698.444	0.076
4-star	1205	1319.167	2645.48	-0.043
5-star	1417	2546.05	11976.57	-0.094
triangles	30	28.372	14.689	0.111
4-clique	0	0	0	-0.065
5-clique	0	0	0	-1.#IO
6-clique	0	0	0	-0.065
7-clique	0	0	0	-1.#IO
Isolates	7	6.919	2.885	0.028
Triangle2	34	28.399	33.536	0.167
3Path	1712	1942.658	1420.647	-0.162
4Cycle	47	37.394	42.132	0.228
1ET	375	370.433	343.249	0.013
2ET	1036	1105.067	1913.339	-0.036

Fisherman – GOF (cont…..)

effects	observed	mean	stddev	t-ratio
0ET	139	110.357	50.081	0.572
ETNT	108	71.589	29.849	1.22
2_3Star	85	89.082	172.424	-0.024
1_2Triangle	13	14.172	6.231	-0.188
1_2ET	-143	-182.101	649.879	0.06
AS(2.00)	268.528	263.503	74.381	0.068
AS(2.00)	268.528	263.503	74.381	0.068
AT(2.00)	75.75	72.642	31.869	0.098
AT(2.00)	75.75	72.642	31.869	0.098
A2P(2.00)	460.125	446.354	172.968	0.08
AC(2.00)	0	0	0	-0.065
AET(2.00)	138	139.231	83.566	-0.01
Std Dev degree dist	2.367	2.186	0.444	0.409
Skew degree dist	1.368	1.057	0.464	0.67
Global Clustering	0.179	0.173	0.028	0.23
Mean Local Clustering	0.2	0.153	0.04	1.166
Variance Local Clustering	0.089	0.062	0.019	1.474

Estimation - Communication

effects	**estimates**	**stderr**	**t-ratio**
edge	-1.46765	0.96369	-0.05845
AS(2.00)	-0.39465	0.46553	-0.05517
AT(2.00)	1.099287	0.20639	-0.0392 *
A2P(2.00)	-0.17176	0.11687	-0.04866

GOF - Communication

effects	observed	mean	stddev	t-ratio
edge	71	71.152	9.585	-0.016
2-star	281	281.51	87.268	-0.006
3-star	369	385.126	212.052	-0.076
4-star	329	404.187	361.69	-0.208
5-star	200	342.658	492.06	-0.29
triangles	35	34.831	12.04	0.014
4-clique	0	0	0	0.016
5-clique	0	0	0	-0.016
6-clique	0	0	0	0.016
7-clique	0	0	0	-0.016
Isolates	0	1.467	1.237	-1.186
Triangle2	70	69.561	48.771	0.009
3Path	1049	1187.508	609.73	-0.227
4Cycle	63	62.571	43.248	0.01
1ET	360	398.111	236.625	-0.161
2ET	559	789.538	755.131	-0.305
0ET	48	41.405	15.928	0.414
ETNT	33	20.021	10.616	1.223
2_3Star	97	88.947	28.728	0.263
1_2Triangle	0	0.051	13.496	-0.004
1_2ET	81	3.342	158.055	0.488
AS(2.00)	158.188	158.703	34.02	-0.015
AS(2.00)	158.188	158.703	34.02	-0.015
AT(2.00)	75.813	76.055	19.771	-0.012
AT(2.00)	75.813	76.055	19.771	-0.012
A2P(2.00)	227.688	228.673	58.388	-0.017
AC(2.00)	0	0	0	0.016
AET(2.00)	176	174.956	71.712	0.013
Std Dev degree dist	2.165	2.067	0.356	0.275
Skew degree dist	0.3	0.372	0.318	-0.228
Global Clustering	0.374	0.37	0.046	0.087
Mean Local Clustering	0.372	0.372	0.063	-0.006
Variance Local Clustering	0.117	0.094	0.022	1.028