PART I RECAP

	Erdos-Renyi	Watts-Strogatz	Barabasi-Albert
Path length	V	V	V
Clustering coefficient	V	X	~
Degree distribution	X	X	V

OTHER PREFERENTIAL ATTACHMENT MODELS

There are **many versions** of preferential attachment models

Some are **generalisations**, some provide better models for **specific cases**

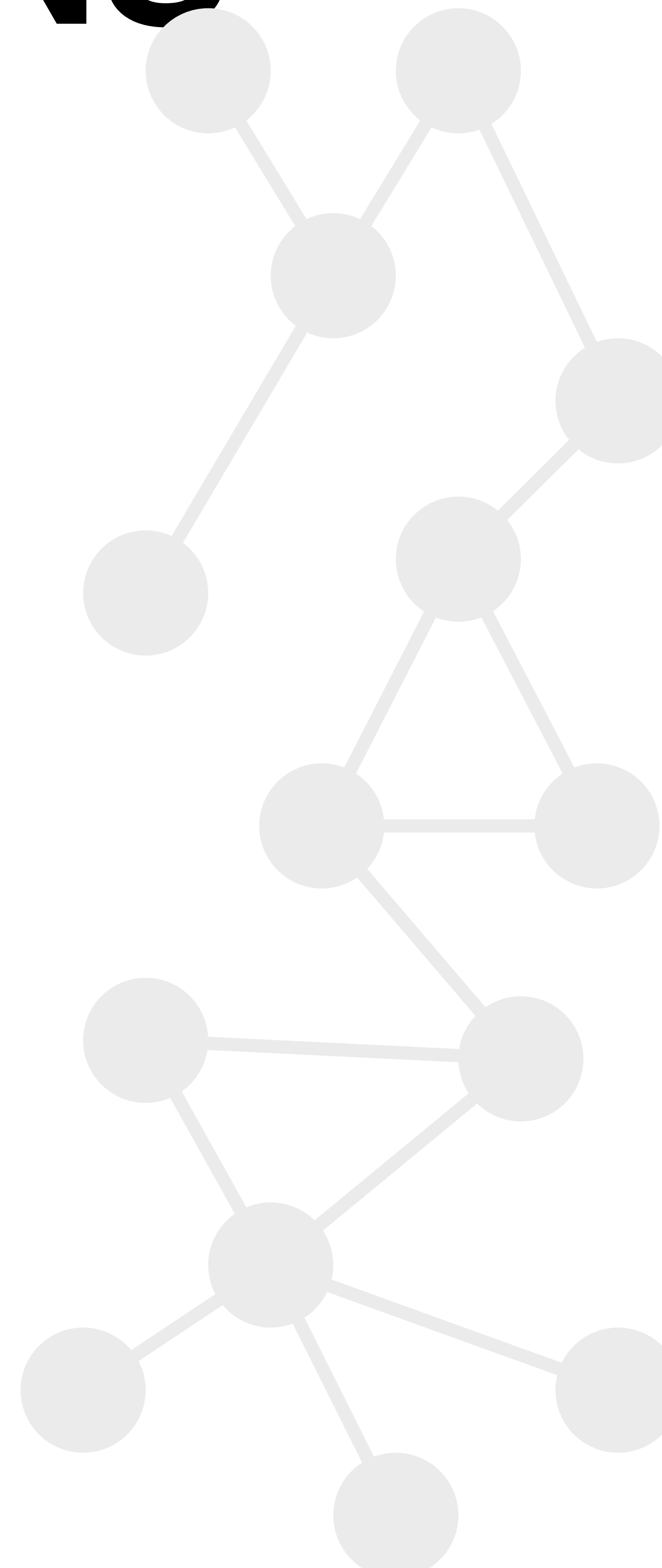
BA LIMITATIONS

FIXED DEGREE **DISTRIBUTION SLOPE**

BA LIMITATIONS

FIXED DEGREE **DISTRIBUTION SLOPE**

OLDEST NODES BECOME HUBS



BA LIMITATIONS

FIXED DEGREE **DISTRIBUTION SLOPE**

OLDEST NODES BECOME HUBS

CLUSTERING LOWER THAN REAL-WORLD NETWORKS

BA LIMITATIONS

FIXED DEGREE **DISTRIBUTION SLOPE**

OLDEST NODES BECOME HUBS

CLUSTERING LOWER THAN REAL-WORLD NETWORKS

NODES/LINKS **CANNOT BE REMOVED**

BA LIMITATIONS

FIXED DEGREE **DISTRIBUTION SLOPE**

OLDEST NODES BECOME HUBS

CLUSTERING LOWER THAN REAL-WORLD NETWORKS

NODES/LINKS **CANNOT BE REMOVED**

SINGLE **CONNECTED COMPONENT**

BA LIMITATIONS

FIXED DEGREE **DISTRIBUTION SLOPE**

OLDEST NODES BECOME HUBS

CLUSTERING LOWER THAN REAL-WORLD NETWORKS

NODES/LINKS **CANNOT BE REMOVED**

SINGLE **CONNECTED COMPONENT**

LIMITING **INITIAL CONDITIONS**

NON-LINEAR PREFERENTIAL ATTACHMENT

$$\Pi_\alpha(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

$\alpha = 1$ Barabási-albert model

$\alpha < 1$ No hubs

$\alpha > 1$ Fewer, larger hubs

ATTRACTIVENESS MODEL

$$\Pi(k_i) = \frac{A + k_i}{\sum_j A + k_j} \qquad \forall A \geq 0$$

ATTRACTIVENESS MODEL

$$\Pi(k_i) = \frac{A + k_i}{\sum_j (A + k_j)} \qquad \forall A \geq 0$$

Slope of degree distribution depends on A

FITNESS MODEL

$$\Pi(k_i) = \frac{\eta_i k_i}{\sum_j (\eta_j k_j)}$$

Fittest, not oldest nodes become hubs

FITNESS MODEL

$$\Pi(k_i) = \frac{\eta_i k_i}{\sum_j (\eta_j k_j)}$$

Fittest, not oldest nodes become hubs

We take η from a distribution $\rho(\eta)$

If $\rho(\eta)$ has a finite support (finite maximum value) there are hubs. Otherwise, there will likely be a single massive hub.

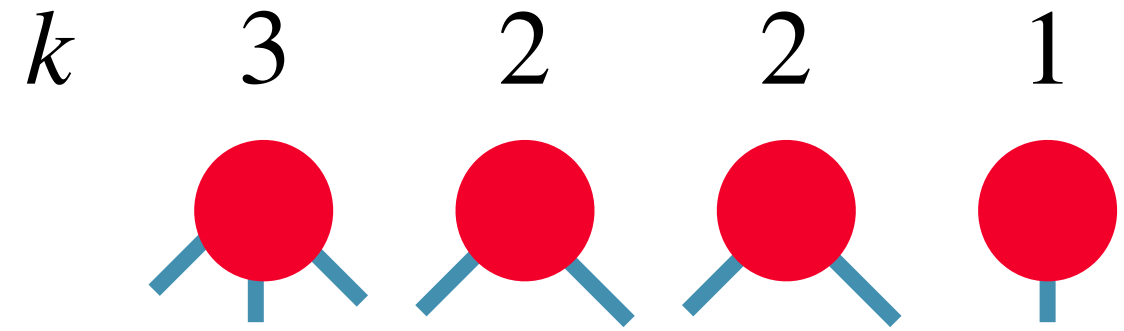
CONFIGURATION MODEL

Takes a **degree sequence** as input

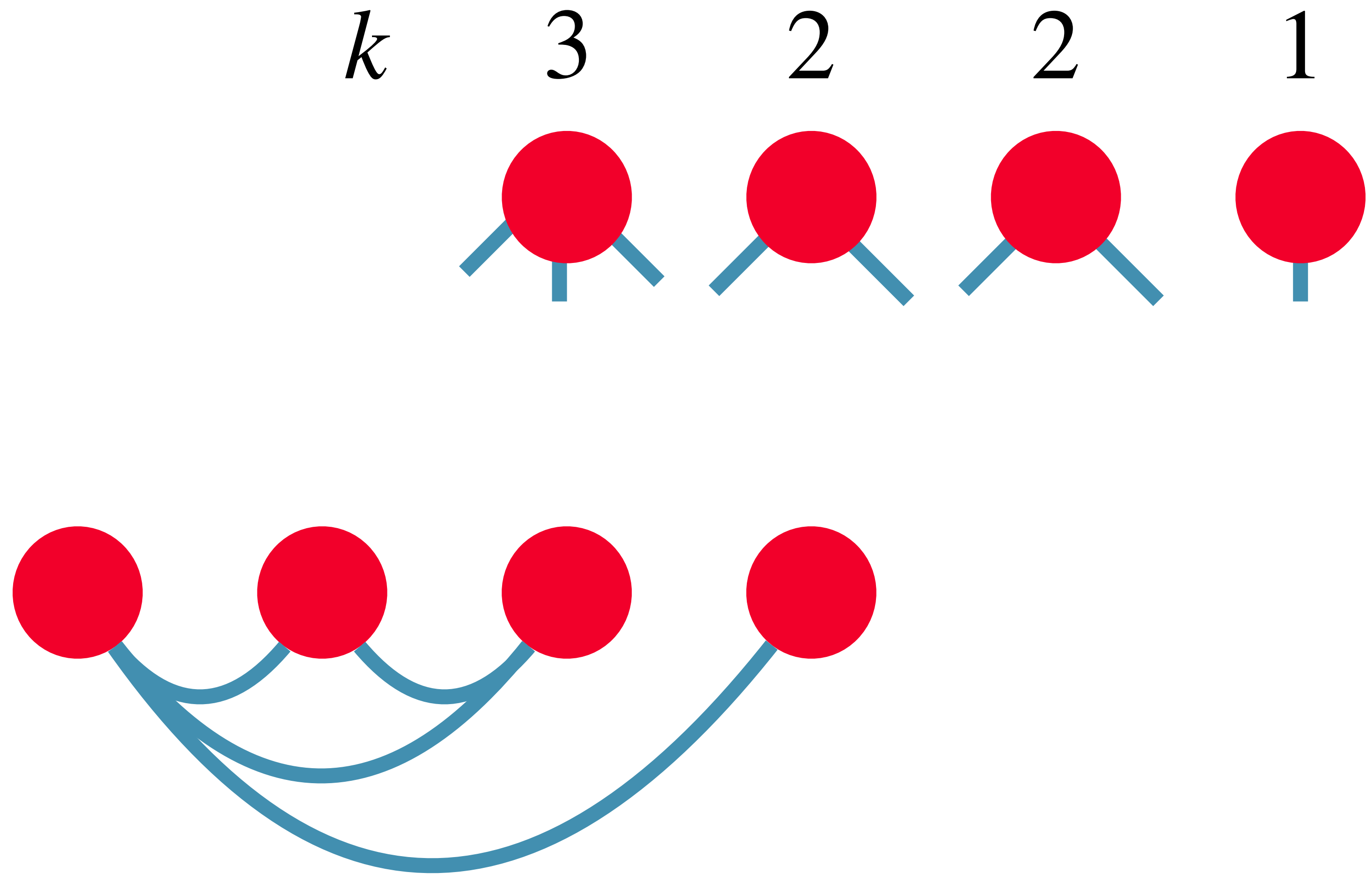
Links are **rewired** randomly, degree sequence preserved

Very important for **benchmarking**!

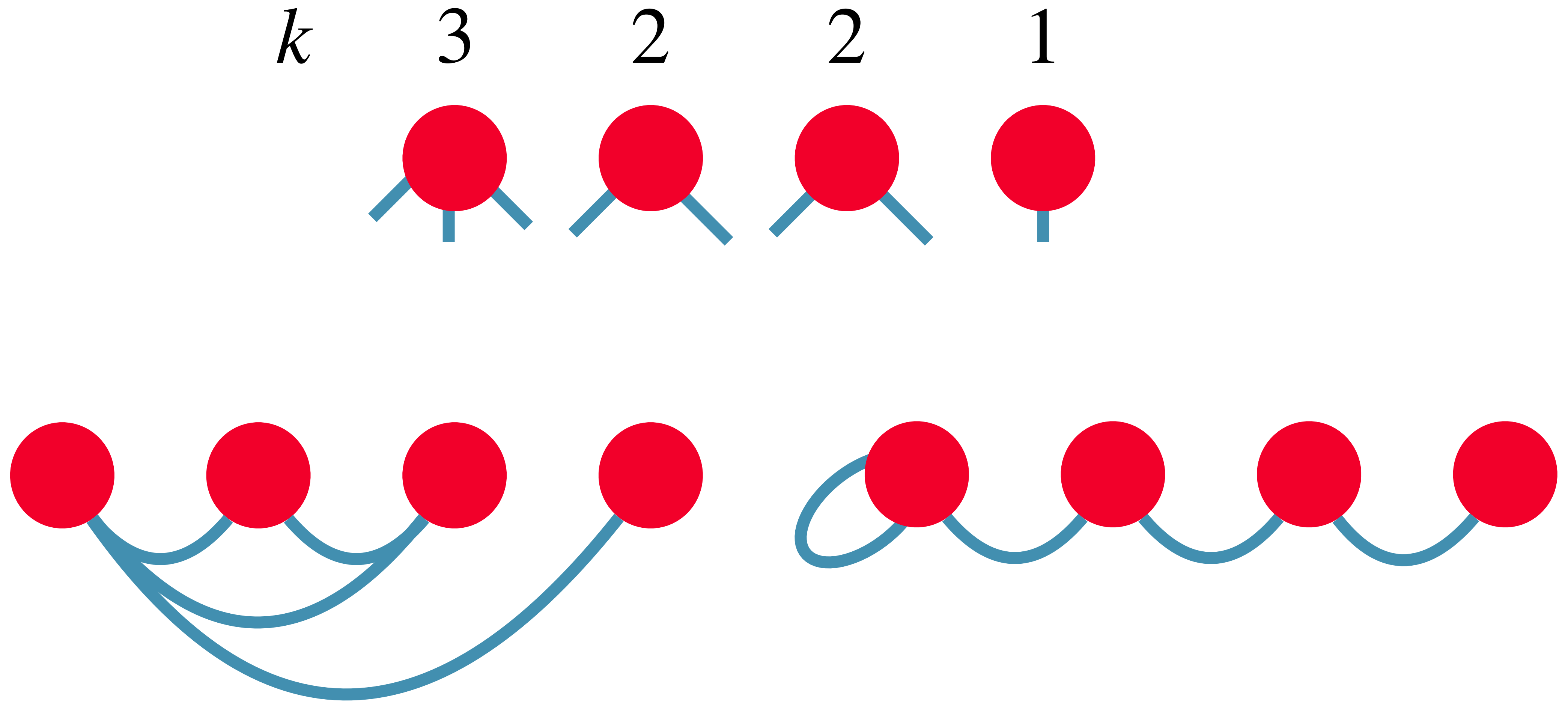
CONFIGURATION MODEL



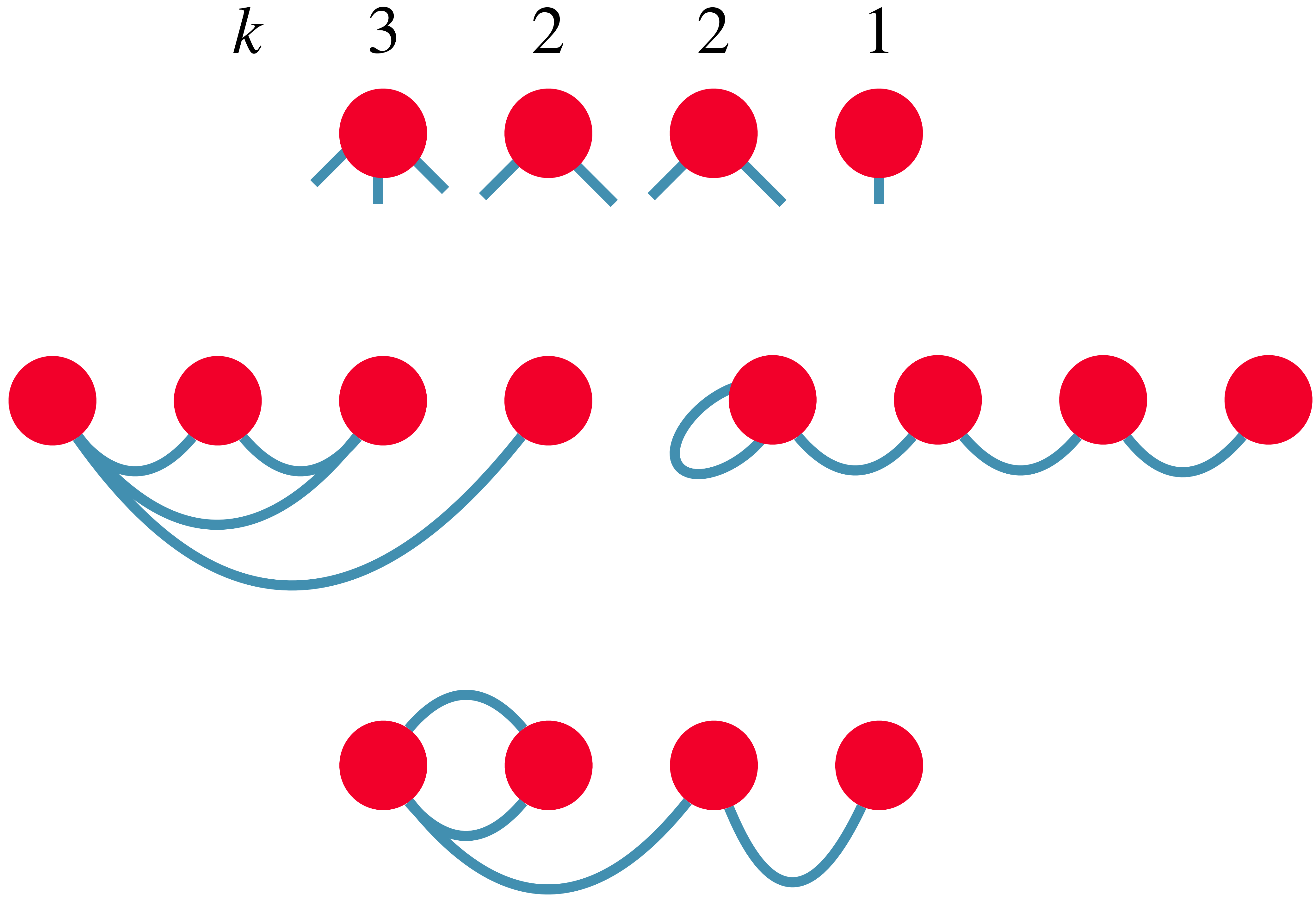
CONFIGURATION MODEL



CONFIGURATION MODEL



CONFIGURATION MODEL



EXERCISE

INTERBANK NETWORK

Soramäki, Bech, Arnold, Beyeler - The topology of interbank payment flows - Physica A (2007)

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. \$tr = \$trillion, \$mn = \$million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

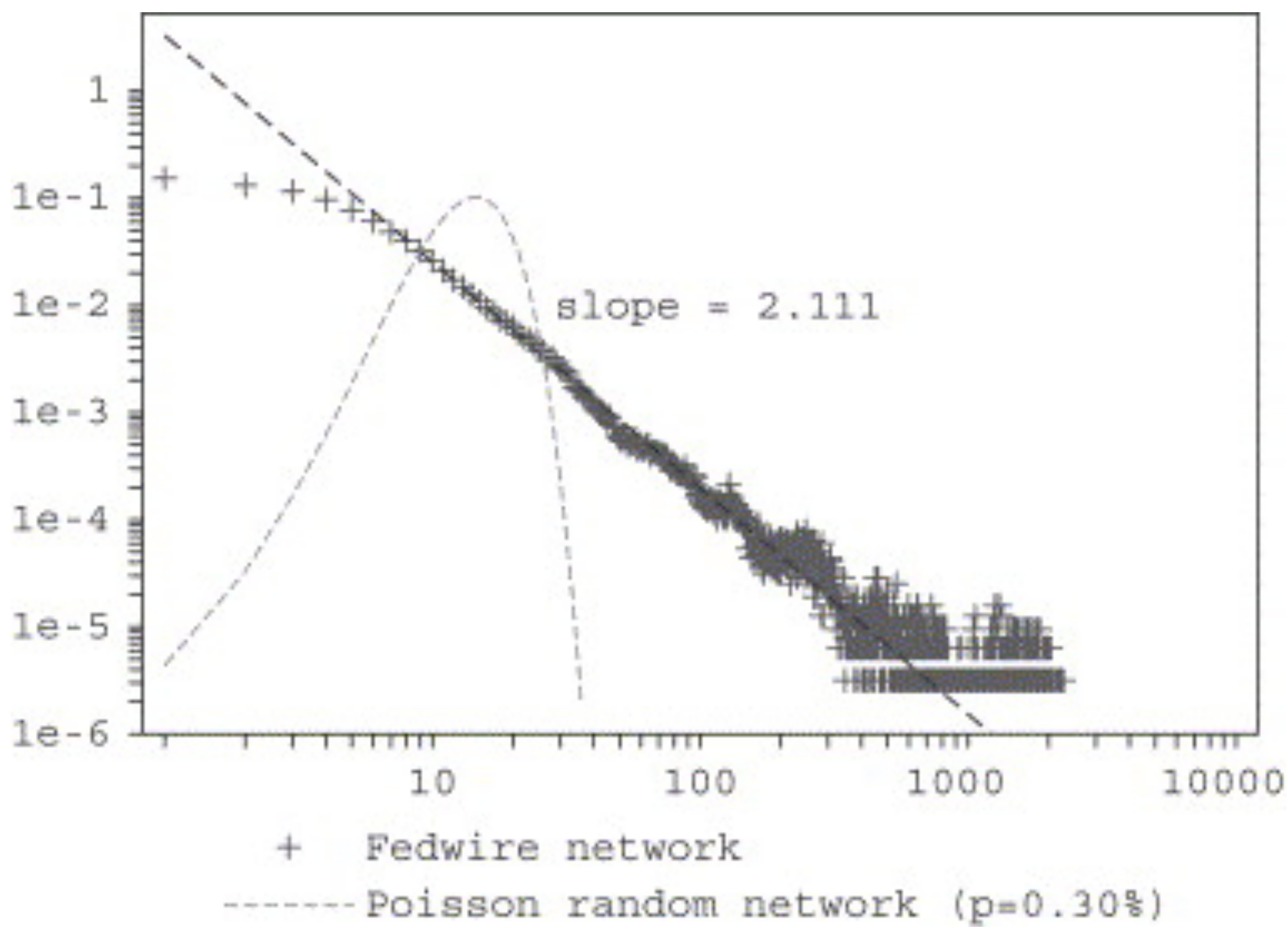
TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Value ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\text{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\text{MLE}}$	2.15	2.15	2.15	2.18	0.01

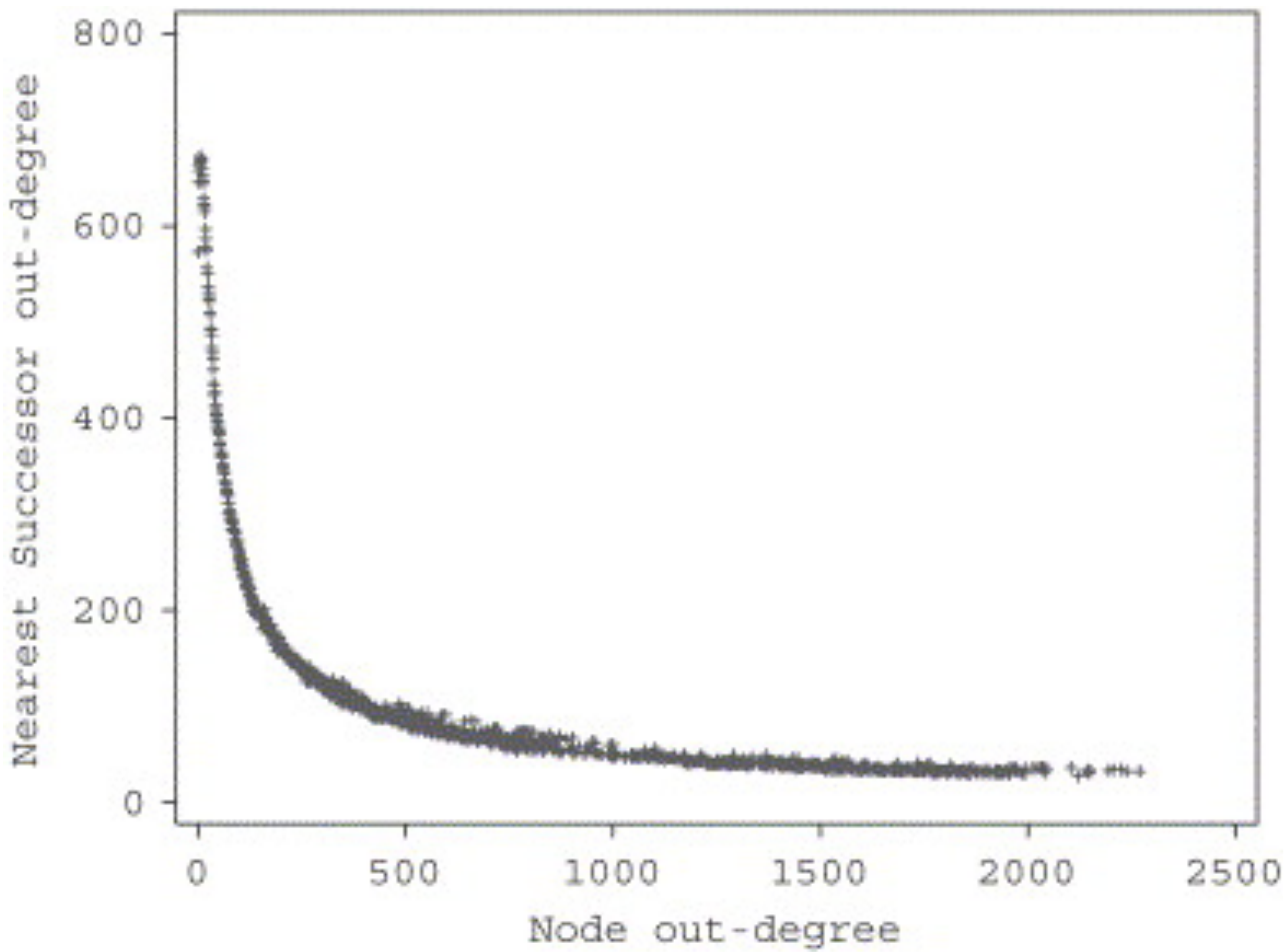
TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

EXERCISE



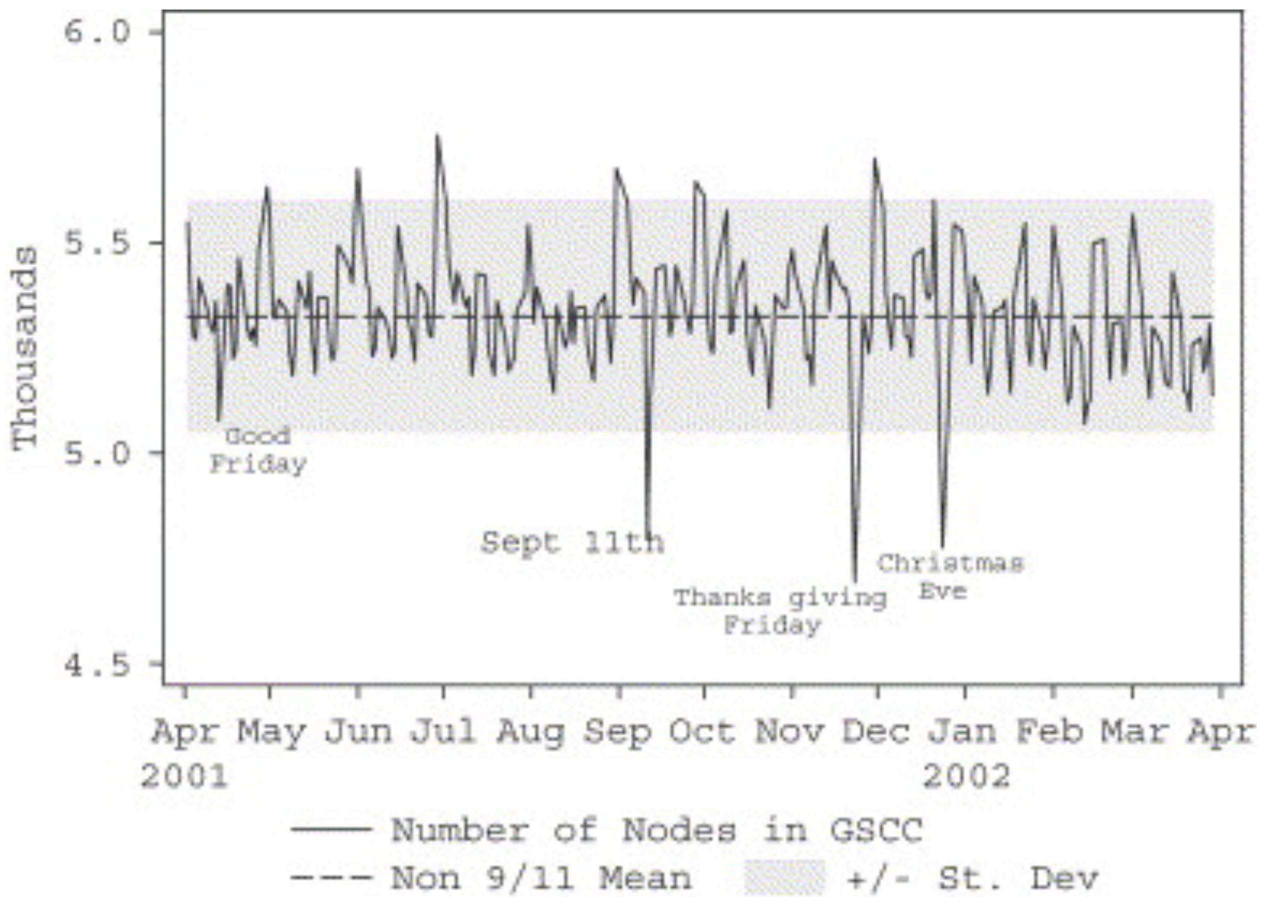
OUT-DEGREE DISTRIBUTION

EXERCISE



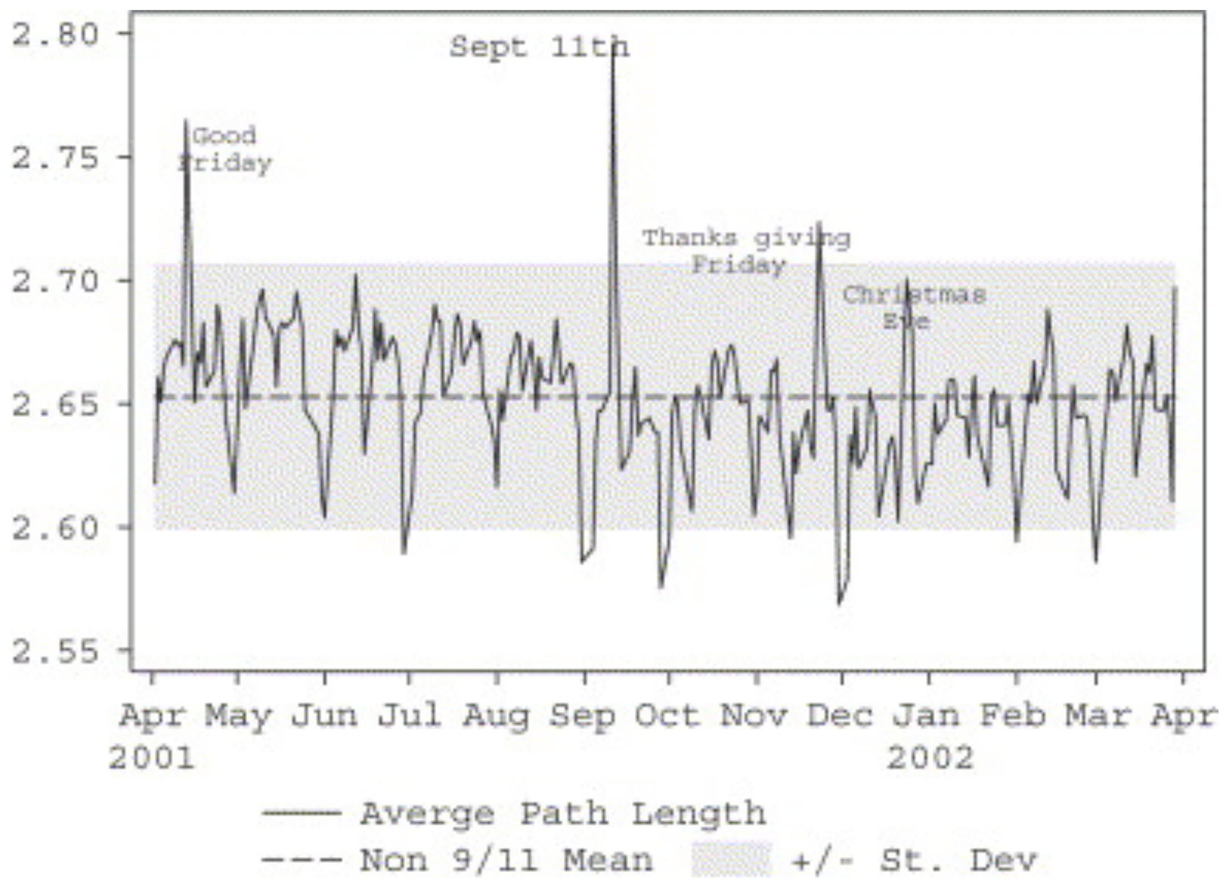
K-NEAREST NEIGHBOURS

EXERCISE



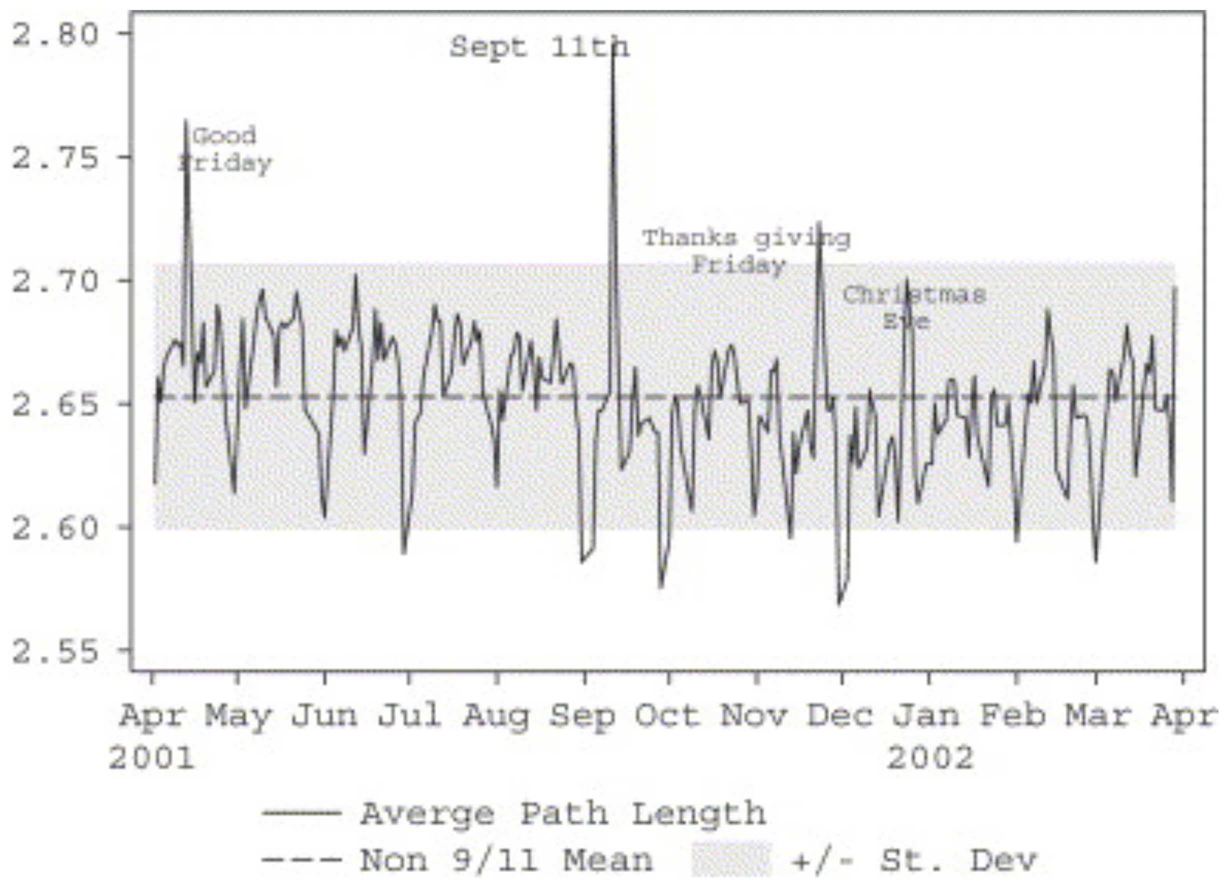
NODES IN GSCC

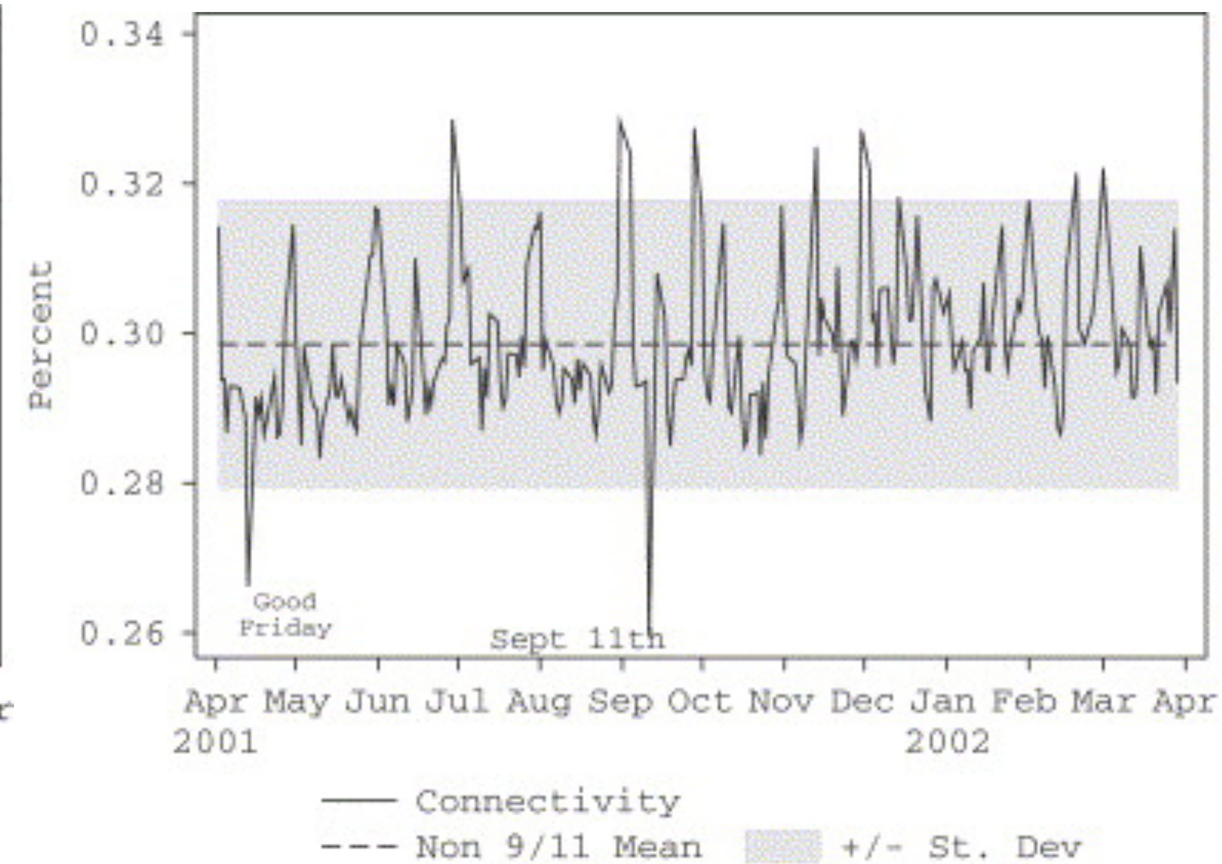
EXERCISE



AVERAGE PATH LENGTH

EXERCISE

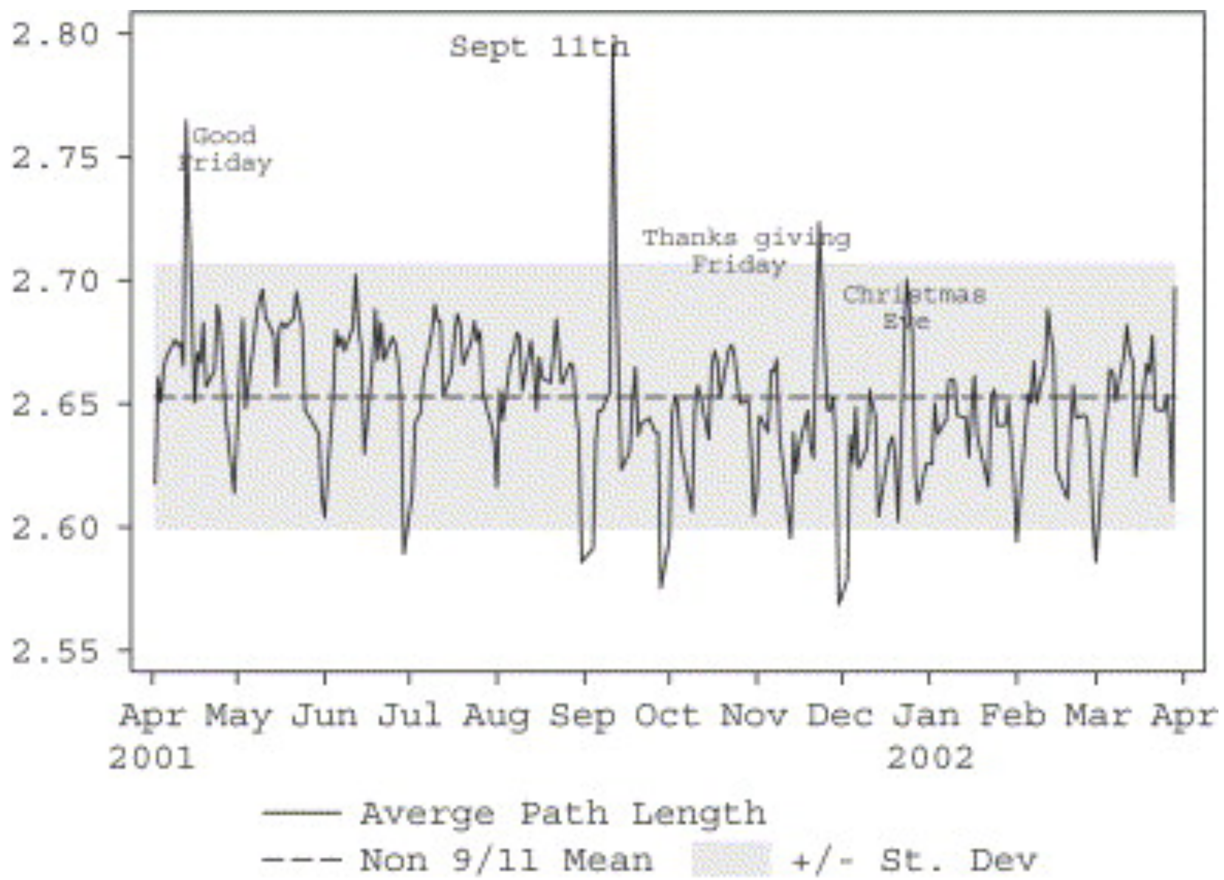


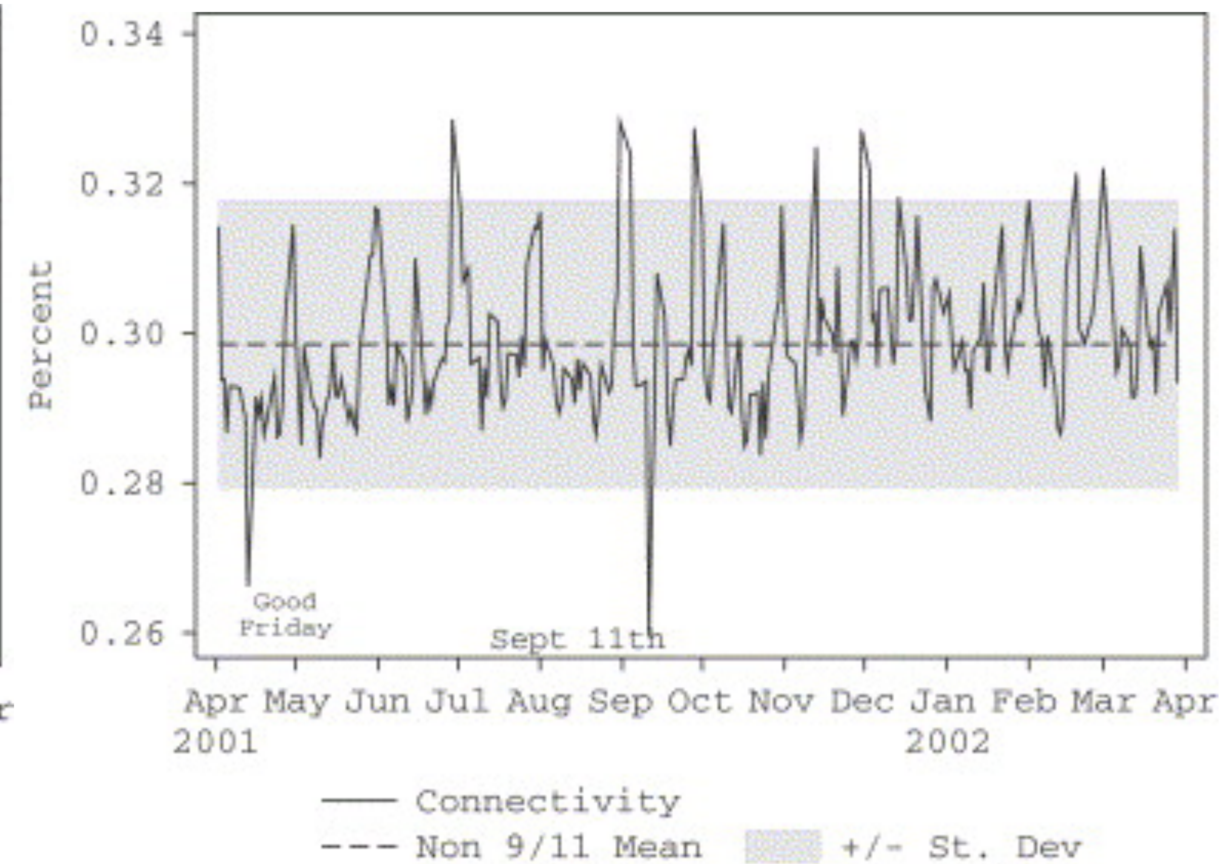


AVERAGE PATH LENGTH

DENSITY

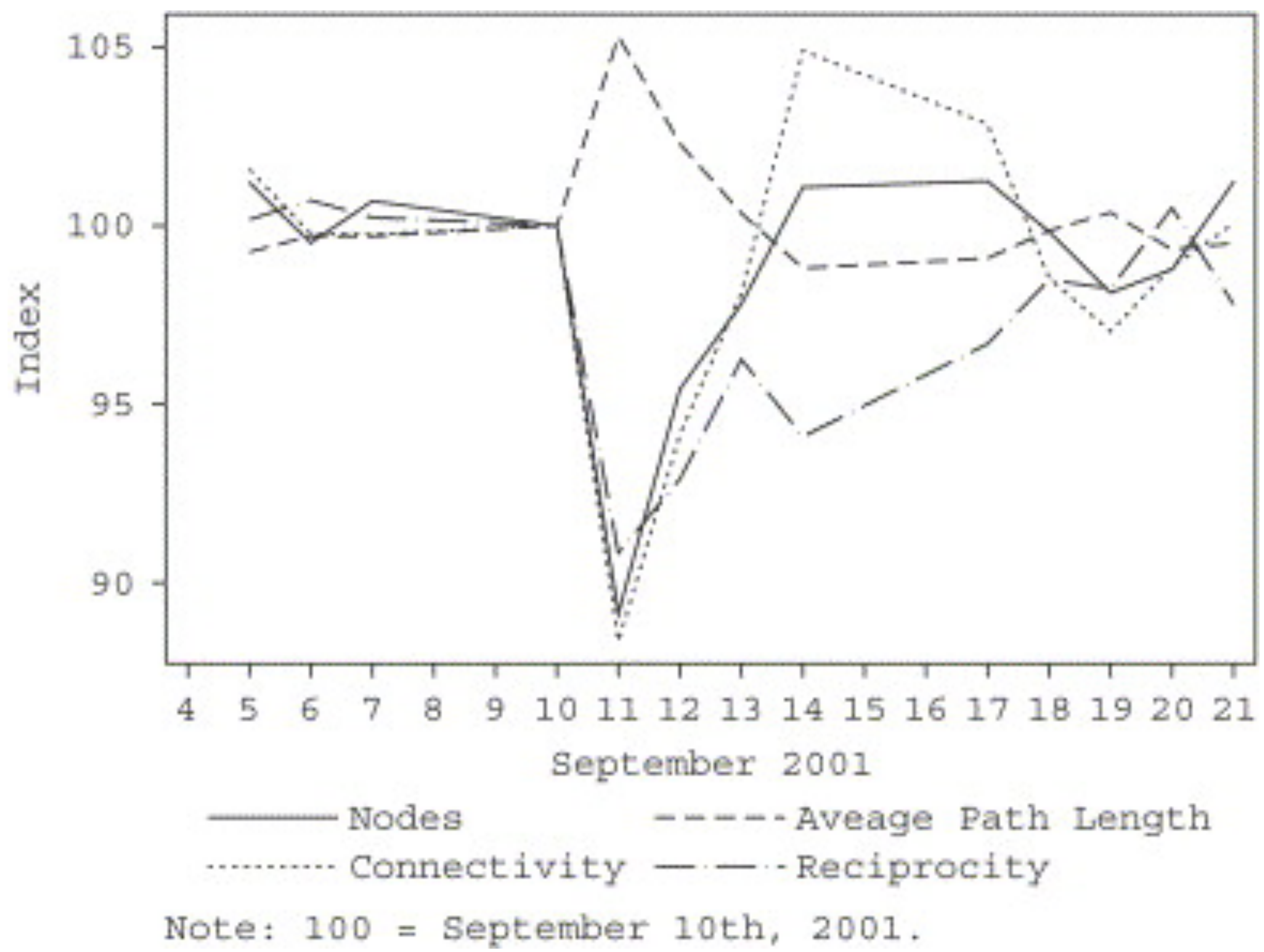
EXERCISE





AVERAGE PATH LENGTH **DENSITY**

EXERCISE



NORMALISED QUANTITIES - SEPTEMBER 2001

SUMMARY

We can **generate networks** with algorithms

Most real-world networks are compatible with one of these models

We can **compare the properties** of a real-world network with that of a model to understand how they are formed