Network Science

2019. szeptember 16.

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source" node:

- number of first neighbors?

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source” node:

- number of first neighbors?

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source" node:

- number of first neighbors $\simeq \langle k \rangle$

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source" node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors?

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source” node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source” node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$
- number of third neighbors?

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source" node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$
- number of third neighbors $\simeq \langle k \rangle^3$

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source” node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$
- number of third neighbors $\simeq \langle k \rangle^3$
- etc.

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source" node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$
- number of third neighbors $\simeq \langle k \rangle^3$
- etc.
- $\langle k \rangle^{\langle \ell \rangle} = ?$

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source” node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$
- number of third neighbors $\simeq \langle k \rangle^3$
- etc.
- $\langle k \rangle^{\langle \ell \rangle} \simeq N$

Let us estimate $\langle \ell \rangle$ for a random graph!

We assume:

- N is large,
- the average number of connections per node, $\langle k \rangle$, is small (the graph is sparse).

Randomly chosen „source" node:

- number of first neighbors $\simeq \langle k \rangle$
- number of second neighbors $\simeq \langle k \rangle^2$
- number of third neighbors $\simeq \langle k \rangle^3$
- etc.
- $\langle k \rangle^{\langle \ell \rangle} \simeq N$
- Thus,

$$\langle \ell \rangle \simeq \frac{\ln N}{\ln \langle k \rangle}$$

The small world property

The small world property

- A network has the small world property if $\langle \ell \rangle \sim \ln N$ (at most).

The small world property

REAL NETWORKS HAVE THE SMALL WORLD PROPERTY!

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153, 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook *et al.* 2001a, Pastor-Satorras *et al.* 2001	2
Movie actors	225, 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52, 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b	4
MEDLINE coauthorship	1, 520, 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b	5
SPIRES coauthorship	56, 627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11, 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b	7
Math coauthorship	70, 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.* 2001	8
Neurosci. coauthorship	209, 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.* 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cooccurence	460.902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22, 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.* 2001	15
Power grid	4, 941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

- Almost all random graph models have it as well.
- Example for non small world networks?

The small world property

REAL NETWORKS HAVE THE SMALL WORLD PROPERTY!

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153, 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook *et al.* 2001a, Pastor-Satorras *et al.* 2001	2
Movie actors	225, 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52, 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b	4
MEDLINE coauthorship	1, 520, 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b	5
SPIRES coauthorship	56, 627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11, 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b	7
Math coauthorship	70, 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.* 2001	8
Neurosci. coauthorship	209, 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.* 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cooccurence	460.902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22, 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.* 2001	15
Power grid	4, 941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

- Almost all random graph models have it as well.
- Example for non small world networks?

The small world property

REAL NETWORKS HAVE THE SMALL WORLD PROPERTY!

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153,127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook *et al.* 2001a, Pastor-Satorras *et al.* 2001	2
Movie actors	225,226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52,909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b	4
MEDLINE coauthorship	1,520,251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b	5
SPIRES coauthorship	56,627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11,994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b	7
Math coauthorship	70,975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.* 2001	8
Neurosci. coauthorship	209,293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.* 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cooccurence	460.902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22,311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.* 2001	15
Power grid	4,941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

- Almost all random graph models have it as well.
- Example for non small world networks?

The small world property

REAL NETWORKS HAVE THE SMALL WORLD PROPERTY!

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153, 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook *et al.* 2001a, Pastor-Satorras *et al.* 2001	2
Movie actors	225, 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52, 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b	4
MEDLINE coauthorship	1, 520, 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b	5
SPIRES coauthorship	56, 627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11, 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b	7
Math coauthorship	70, 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.* 2001	8
Neurosci. coauthorship	209, 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.* 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cooccurence	460.902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22, 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.* 2001	15
Power grid	4, 941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

- Almost all random graph models have it as well.
- Example for non small world networks: regular lattices.

The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1^{st}, 2^{nd}, 3^{d}, etc. neighborhoods:

The small world property

Implications

What are the consequences of $\langle\ell\rangle \sim \ln N$?

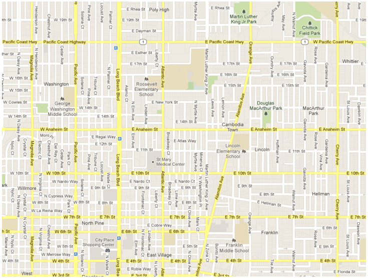
$\rightarrow$ If we consider the „concentric shells" of 1^{st}, 2^{nd}, 3^{d}, etc. neighborhoods:

The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells” of 1$^{\mathrm{st}}$, 2$^{\mathrm{nd}}$, 3$^{\mathrm{d}}$, etc. neighborhoods:

- in a regular lattice, e.g., a city:

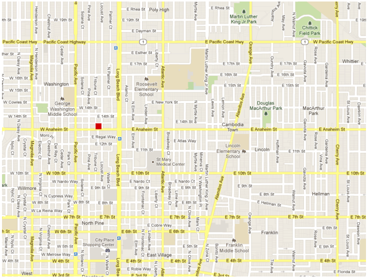


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1$^{\mathrm{st}}$, 2$^{\mathrm{nd}}$, 3$^{\mathrm{d}}$, etc. neighborhoods:

- in a regular lattice, e.g., a city:

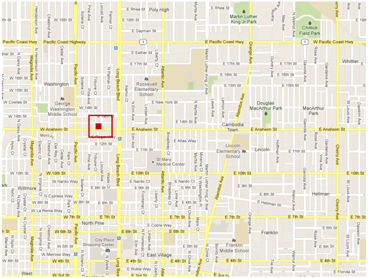


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells” of 1$^{\mathrm{st}}$, 2$^{\mathrm{nd}}$, 3$^{\mathrm{d}}$, etc. neighborhoods:

- in a regular lattice, e.g., a city:

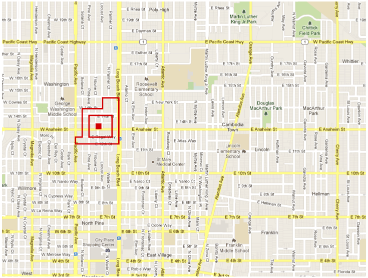


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells" of 1st, 2nd, 3d, etc. neighborhoods:

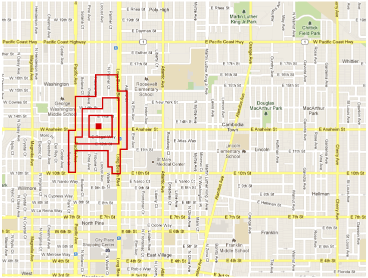
- in a regular lattice, e.g., a city:



What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1$^{\rm st}$, 2$^{\rm nd}$, 3$^{\rm d}$, etc. neighborhoods:

- in a regular lattice, e.g., a city:

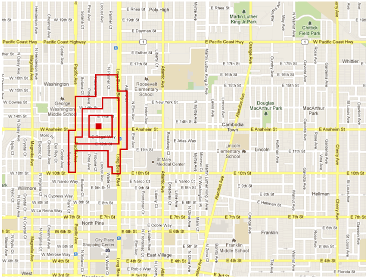


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells" of 1st, 2nd, 3d, etc. neighborhoods:

- in a regular lattice, e.g., a city:



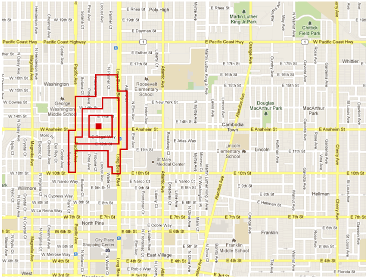
  - number of nodes (buildings, blocks, etc.) grows roughly as $n \sim \ell^d$, (e.g., as $n \sim \ell^2$ in a city)

The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells” of 1st, 2nd, 3d, etc. neighborhoods:

- in a regular lattice, e.g., a city:



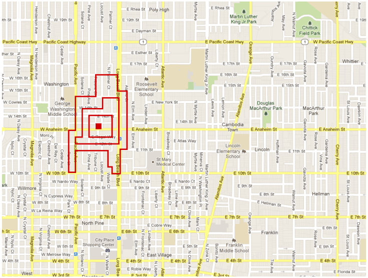
  - number of nodes (buildings, blocks, etc.) grows roughly as $n \sim \ell^d$, (e.g., as $n \sim \ell^2$ in a city)
  - to cover the whole system, we need roughly $\ell \sim N^{1/d}$!

The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1$^{\text{st}}$, 2$^{\text{nd}}$, 3$^{\text{d}}$, etc. neighborhoods:

- in a regular lattice, e.g., a city:

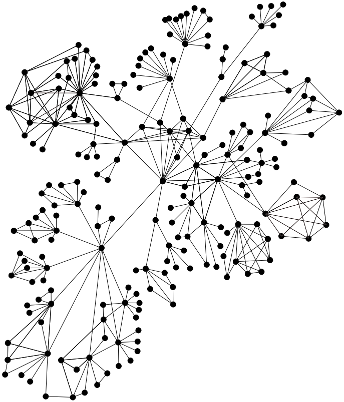


  - number of nodes (buildings, blocks, etc.) grows roughly as $n \sim \ell^d$, (e.g., as $n \sim \ell^2$ in a city)
  - to cover the whole system, we need roughly $\ell \sim N^{1/d}$!
  - (e.g., in a city with 100.000 buildings, $\ell \simeq$ **300** !)

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1st, 2nd, 3^{d}, etc. neighborhoods:

- in a random network, e.g., social network:

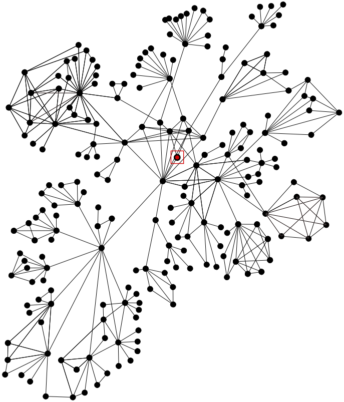


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1$^{\text{st}}$, 2$^{\text{nd}}$, 3$^{\text{d}}$, etc. neighborhoods:

- in a random network, e.g., social network:

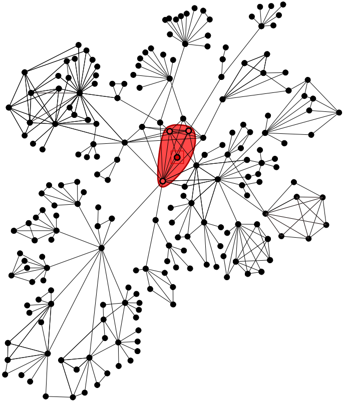


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells” of 1^{st}, 2^{nd}, 3^{d}, etc. neighborhoods:

- in a random network, e.g., social network:



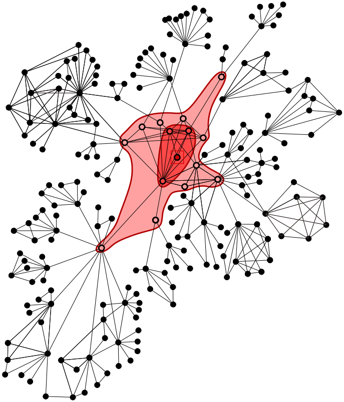
The small world property

Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells" of 1$^{\mathrm{st}}$, 2$^{\mathrm{nd}}$, 3$^{\mathrm{d}}$, etc. neighborhoods:

- in a random network, e.g., social network:



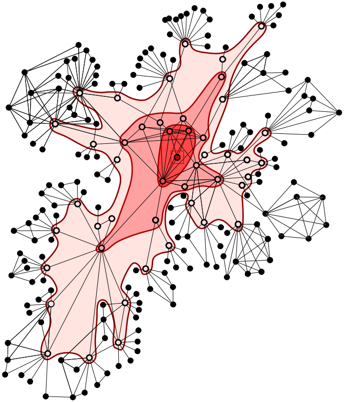
The small world property

Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

$\rightarrow$ If we consider the „concentric shells" of 1^{st}, 2^{nd}, 3^{d}, etc. neighborhoods:

- in a random network, e.g., social network:

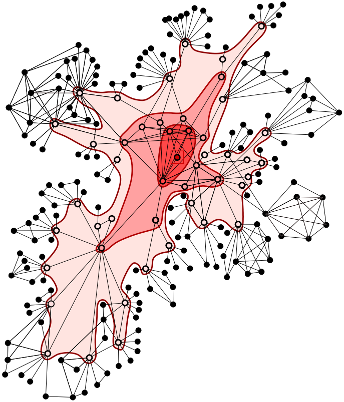


The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells" of 1st, 2nd, 3^{d}, etc. neighborhoods:

- in a random network, e.g., social network:



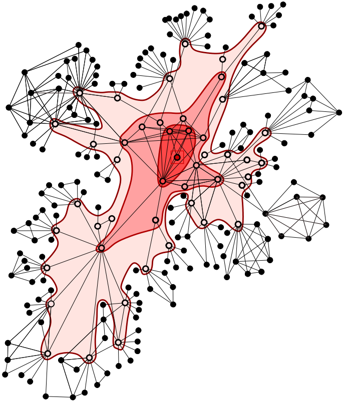
  - number of nodes grows **exponentially** as $n \sim \langle k \rangle^{\ell}$!

The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells" of 1st, 2nd, 3d, etc. neighborhoods:

- in a random network, e.g., social network:



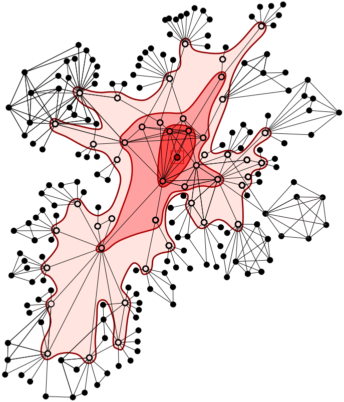
- number of nodes grows **exponentially** as $n \sim \langle k \rangle^{\ell}$!
- to cover the whole system, **only a couple of neighborhoods** are needed,

The small world property
Implications

What are the consequences of $\langle \ell \rangle \sim \ln N$?

→ If we consider the „concentric shells” of 1st, 2nd, 3d, etc. neighborhoods:

- in a random network, e.g., social network:



- number of nodes grows **exponentially** as $n \sim \langle k \rangle^{\ell}$!
- to cover the whole system, **only a couple of neighborhoods** are needed,
- (e.g., in a network with $N = 100.000$ and $\langle k \rangle = 5$ we need only $\ell \simeq$ **7**!)

Node centralities

Node centralities

Closeness

- One of the widely used centrality measures is called „closeness”. How would you define it?

Basic network characteristics

Distance and paths

The small world property

Centralities

Components

Closeness

- The closeness or closeness centrality of node i is usually defined as

$$C_c(i) \equiv \frac{1}{\langle \ell_i \rangle},$$

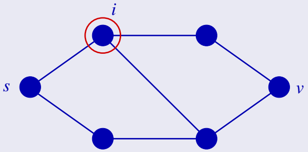
  where the nodes unreachable from i are left out of the average.
- With this definition a node „closer” to the rest of the network obtains a higher C_c value.
- Nodes „closer” to the rest of the network are intuitively central.

- Another important centrality measure is called „betweenness”. How would you define it?

Betweenness

- The betweenness of a node or link is equal to the number of shortest paths (between all possible pairs of nodes) passing through the given node or link.
- If multiple shortest paths are possible between a given pair of nodes, they are given equal weights adding up to one:

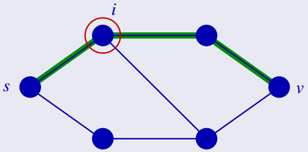
$$b_i \equiv \sum_{s \neq i, v \neq i} \frac{\sigma_{sv}(i)}{\sigma_{sv}}.$$



Betweenness

- The betweenness of a node or link is equal to the number of shortest paths (between all possible pairs of nodes) passing through the given node or link.
- If multiple shortest paths are possible between a given pair of nodes, they are given equal weights adding up to one:

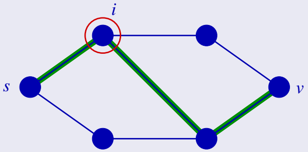
$$b_i \equiv \sum_{s \neq i, v \neq i} \frac{\sigma_{sv}(i)}{\sigma_{sv}}.$$



Betweenness

- The betweenness of a node or link is equal to the number of shortest paths (between all possible pairs of nodes) passing through the given node or link.
- If multiple shortest paths are possible between a given pair of nodes, they are given equal weights adding up to one:

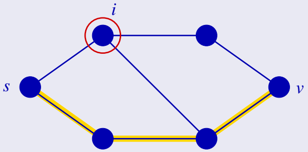
$$b_i \equiv \sum_{s \neq i, v \neq i} \frac{\sigma_{sv}(i)}{\sigma_{sv}}.$$



Betweenness

- The betweenness of a node or link is equal to the number of shortest paths (between all possible pairs of nodes) passing through the given node or link.
- If multiple shortest paths are possible between a given pair of nodes, they are given equal weights adding up to one:

$$b_i \equiv \sum_{s\neq i, v\neq i} \frac{\sigma_{sv}(i)}{\sigma_{sv}}.$$



PageRank and eigenvector centrality

PageRank and eigenvector centrality

Eigenvector centrality:

- The eigenvalue problem of the adjacency matrix:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

- The eigenvectors have the same number of components (elements) as the number of nodes in the system...

→ We can take the **largest eigenvalue** λ_1, and treat the components of the corresponding eigenvector $\mathbf{v}_1$ as the value of a centrality measure, associated to the corresponding node in the network.

PageRank and eigenvector centrality

Eigenvector centrality:

- The eigenvalue problem of the adjacency matrix:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

- The eigenvectors have the same number of components (elements) as the number of nodes in the system...

→ We can take the **largest eigenvalue** λ_1, and treat the components of the corresponding eigenvector $\mathbf{v}_1$ as the value of a centrality measure, associated to the corresponding node in the network.

PageRank and eigenvector centrality

Eigenvector centrality:

- The eigenvalue problem of the adjacency matrix:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

- The eigenvectors have the same number of components (elements) as the number of nodes in the system...

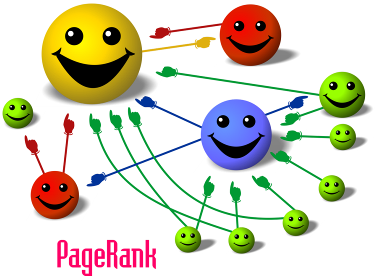
→ We can take the **largest eigenvalue** λ_1, and treat the components of the corresponding eigenvector $\mathbf{v}_1$ as the value of a centrality measure, associated to the corresponding node in the network.

PageRank and eigenvector centrality

What is PageRank?

PageRank and eigenvector centrality

The basic concept of PageRank:



The importance of a node is affected by:

- the number of in-neighbors,
- the importance of the in-neighbors.

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly:

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly:

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly:

$$\underbrace{r_i(t+1)}_{\text{PageRank of } i} = \underbrace{\sum_{j \in M(i)}}_{\text{neighs. of } i} \frac{r_j(t)}{k_j}.$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly:

$$\underbrace{r_i(t+1)}_{\text{PageRank of } i} = \underbrace{\sum_{j\in M(i)}}_{\text{neighs. of } i} \frac{r_j(t)}{k_j}.$$

$\rightarrow$ $\sum_{i=1}^{N} r_i$ is conserved!

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly:

$$\underbrace{r_i(t+1)}_{\text{PageRank of } i} = \underbrace{\sum_{j\in M(i)}}_{\text{neighs. of } i} \frac{r_j(t)}{k_j}.$$

$\rightarrow$ $\sum_{i=1}^{N} r_i$ is conserved!

$\rightarrow$ For simplicity we can assume $\sum_{i=1}^{N} r_i = 1$.

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly:

$$\underbrace{r_i(t+1)}_{\text{PageRank of } i} = \sum_{\underbrace{j\in M(i)}_{\text{neighs. of } i}} \frac{r_j(t)}{k_j}.$$

$\rightarrow$ $\sum_{i=1}^{N} r_i$ is conserved!

$\rightarrow$ For simplicity we can assume $\sum_{i=1}^{N} r_i = 1$.

- **Damping**: with probability p_d we „click” at random instead of following the links:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d) \sum_{j \in M(i)} \frac{r_j(t)}{k_j}.$$

Practical algorithm

- Initially distribute r_i evenly, i.e., $r_i = 1/N$.
- Iterate according to the rule above, and r_i will converge soon.

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d) \sum_{j \in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?

Steady state means $r_i(t+1) = r_i(t) = r_i$.

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d) \sum_{j \in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?
Steady state means $r_i(t+1) = r_i(t) = r_i$.
Let us rewrite the steady state equation in a matrix form:

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix} = \frac{p_d}{N} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + ?$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?
Steady state means $r_i(t+1) = r_i(t) = r_i$.
Let us rewrite the steady state equation in a matrix form:

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix} = \frac{p_d}{N}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + (1-p_d)\begin{pmatrix} \frac{A_{11}}{k_1} & \frac{A_{21}}{k_2} & \cdots & \frac{A_{N1}}{k_N} \\ \frac{A_{12}}{k_1} & \frac{A_{22}}{k_2} & \cdots & \frac{A_{N2}}{k_N} \\ \vdots & & \ddots & \vdots \\ \frac{A_{1N}}{k_1} & \frac{A_{2N}}{k_2} & \cdots & \frac{A_{NN}}{k_N} \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}$$

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?
Steady state means $r_i(t+1) = r_i(t) = r_i$.
Let us rewrite the steady state equation in a matrix form:

$$\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}}_{\mathbf{r}} = \frac{p_d}{N} \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbf{1}} + (1-p_d) \underbrace{\begin{pmatrix} \frac{A_{11}}{k_1} & \frac{A_{21}}{k_2} & \cdots & \frac{A_{N1}}{k_N} \\ \frac{A_{12}}{k_1} & \frac{A_{22}}{k_2} & \cdots & \frac{A_{N2}}{k_N} \\ \vdots & & \ddots & \vdots \\ \frac{A_{1N}}{k_1} & \frac{A_{2N}}{k_2} & \cdots & \frac{A_{NN}}{k_N} \end{pmatrix}}_{\mathbf{U}=\left[\mathbf{K}^{-1}\mathbf{A}\right]^T} \underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}}_{\mathbf{r}}$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?

Steady state means $r_i(t+1) = r_i(t) = r_i$.

Let us rewrite the steady state equation in a matrix form:

$$\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}}_{\mathbf{r}} = \frac{p_d}{N}\underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbf{1}} + (1-p_d)\underbrace{\begin{pmatrix} \frac{A_{11}}{k_1} & \frac{A_{21}}{k_2} & \cdots & \frac{A_{N1}}{k_N} \\ \frac{A_{12}}{k_1} & \frac{A_{22}}{k_2} & \cdots & \frac{A_{N2}}{k_N} \\ \vdots & & \ddots & \vdots \\ \frac{A_{1N}}{k_1} & \frac{A_{2N}}{k_2} & \cdots & \frac{A_{NN}}{k_N} \end{pmatrix}}_{\mathbf{U}=[\mathbf{K}^{-1}\mathbf{A}]^T}\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}}_{\mathbf{r}}$$

The same eq. in vector notation:

$$\mathbf{r} = \frac{p_d}{N}\mathbf{1} + (1-p_d)\mathbf{U}\cdot\mathbf{r}$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d) \sum_{j \in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?

$$\mathbf{r} = \frac{p_d}{N}\mathbf{1} + (1-p_d)\mathbf{U} \cdot \mathbf{r}$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d)\sum_{j\in M(i)} \frac{r_j(t)}{k_j}.$$

→ What is the steady state distribution of r_i ?

$$\mathbf{r} = \frac{p_d}{N}\mathbf{1} + (1-p_d)\mathbf{U}\cdot\mathbf{r}$$

$$\updownarrow$$

Eigenvector centrality v_i is the i-th component of the leading eigenvector $\mathbf{v}$ of $\mathbf{A}$, fulfilling

$$\mathbf{A}\cdot\mathbf{v} = \lambda\mathbf{v}.$$

PageRank and eigenvector centrality

How to calculate PageRank?

- Let us assume an iterative process, where everybody is distributing its current PageRank r among its neighbors evenly, with damping:

$$r_i(t+1) = \frac{p_d}{N} + (1-p_d) \sum_{j \in M(i)} \frac{r_j(t)}{k_j}.$$

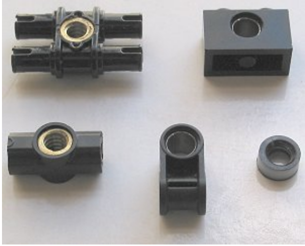
$$\mathbf{r} = \frac{p_d}{N}\mathbf{1} + (1-p_d)\mathbf{U} \cdot \mathbf{r}$$

$$\updownarrow$$

Eigenvector centrality v_i is the i-th component of the leading eigenvector $\mathbf{v}$ of $\mathbf{A}$, fulfilling

$$\mathbf{A} \cdot \mathbf{v} = \lambda \mathbf{v}.$$

$\rightarrow$ Thus, PageRank is a variation of eigenvector centrality!



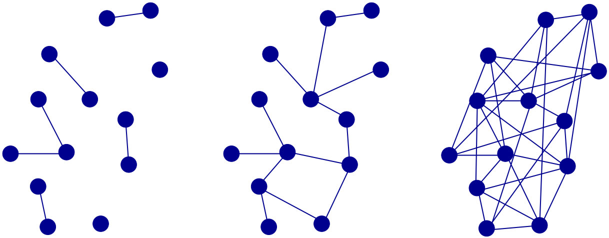
COMPONENTS

Components

Undirected case

Component

A component in an undirected network corresponds to a maximal set of nodes in which a path exists between any pair of nodes.



$n_c = 7$ $\quad$ $n_c = 1$ $\quad$ $n_c = 1$

Giant component

- Most networks we encounter contain a **giant** component.

→ What is a „giant component"?

Giant component

Giant component

A network (graph) with system size $N \to \infty$ contains a giant component if the relative size of this component remains finite, (larger than zero):

$$\lim_{N\to\infty} \frac{S_1}{N} > 0$$

Components

Directed networks

How to generalize the concept of components for the directed case?

Components
Directed networks

Strongly connected component

A strongly connected component is a maximal set of nodes in which a directed path exists between any pair of nodes.

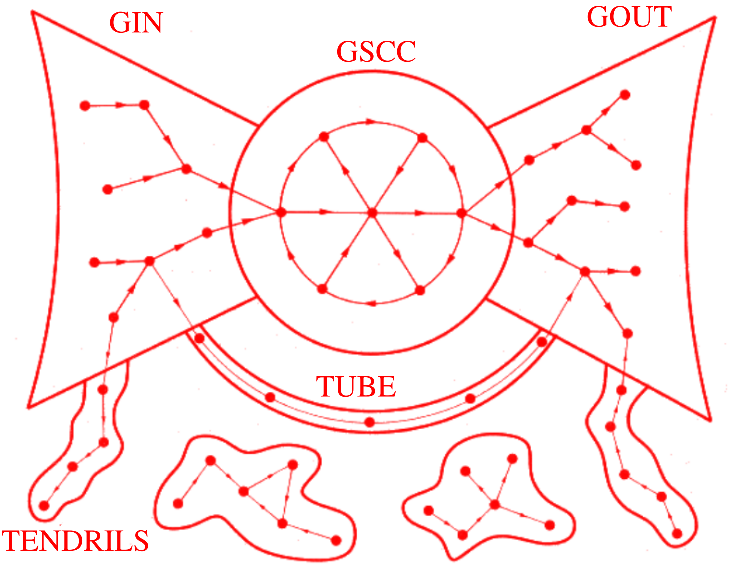
Weakly connected component

A weakly connected component is a maximal set of nodes in which an undirected path exists between any pair of nodes.

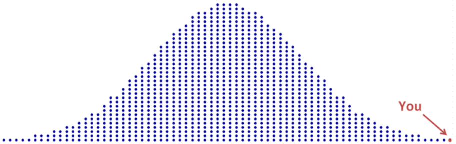
Components of the Internet

How does the structure of a large directed network like the Internet looks like on the large scale from the point of view of components?

Components of the Internet



Advanced network characteristics



DEGREE DISTRIBUTION

Degree distribution

Degree distribution

Degree distribution $\longrightarrow$ **probability distribution of the node degrees**.

Degree distribution

Degree distribution $\longrightarrow$ **probability distribution of the node degrees**.

Degree distribution

- The degree distribution of a network, $p(k)$ is equal to the probability that a randomly chosen node has a degree k.

Degree distribution

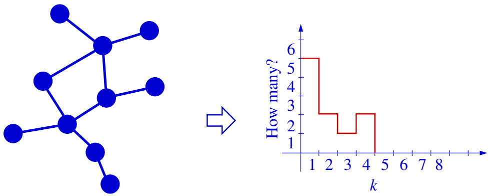
- For a finite network with N nodes,

$$p(k) = \frac{N_k}{N},$$

  where N_k denotes the number of nodes with degree k.
- Thus, $p(k)$ is simply the fraction of nodes with degree k.

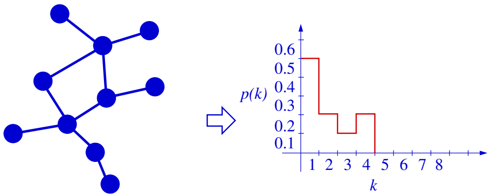
Calculating $p(k)$

- The calculation of the degree distribution is very similar to the construction of a histogram.
- We count for each degree value k how many nodes have that degree,
- and we divide it by the total number of nodes.



Calculating $p(k)$

- The calculation of the degree distribution is very similar to the construction of a histogram.
- We count for each degree value k how many nodes have that degree,
- and we divide it by the total number of nodes.



$p(k)$ in the Erdős-Rényi model

- What is the degree distribution of an Erdős-Rényi graph where N nodes are linked independently with probability p?

$p(k)$ in the Erdős-Rényi model

- What is the degree distribution of an Erdős-Rényi graph where N nodes are linked independently with probability p?
  - The number of edges between i and j:

$$\begin{aligned} \mathcal{P}(e_{ij} = 1) &= p \\ \mathcal{P}(e_{ij} = 0) &= 1 - p \end{aligned}$$

(This is the Bernoulli distribution.)

$p(k)$ in the Erdős-Rényi model

- What is the degree distribution of an Erdős-Rényi graph where N nodes are linked independently with probability p?
  - The number of edges between i and j:

$$\begin{aligned} \mathcal{P}(e_{ij} = 1) &= p \\ \mathcal{P}(e_{ij} = 0) &= 1 - p \end{aligned}$$

  (This is the Bernoulli distribution.)

  - A given node can choose from $N - 1$ possible neighbors, which are attached independently,

$\rightarrow$ $p(k)$ follows a binomial distribution:

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

$p(k)$ in the Erdős-Rényi model

- We usually neglect the -1:

$$p(k) = \binom{N}{k} p^k (1-p)^{N-k},$$

- and approximate in the large N limit the binomial distribution by the Poisson distribution:

$$\begin{aligned} p(k) &= \binom{N}{k} p^k (1-p)^{N-k} \\ &\downarrow \\ p(k) &\simeq \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}, \qquad \langle k \rangle = Np \end{aligned}$$

$p(k)$ in the Erdős-Rényi model

- We usually neglect the -1:

$$p(k) = \binom{N}{k} p^k (1-p)^{N-k},$$

- and approximate in the large N limit the binomial distribution by the Poisson distribution:

$$\begin{aligned} p(k) &= \binom{N}{k} p^k (1-p)^{N-k} \\ &\downarrow \\ p(k) &\simeq \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}, \qquad \langle k \rangle = Np \end{aligned}$$

$p(k)$ in the Erdős-Rényi model

From binomial to Poisson distribution

$$
\begin{aligned}
p(k) &= \frac{N(N-1)\cdots(N-k+1)}{k!}\frac{\langle k\rangle^k}{N^k}\left(1-\frac{\langle k\rangle}{N}\right)^N\left(1-\frac{\langle k\rangle}{N}\right)^{-k} \\
&= \frac{\langle k\rangle^k}{k!}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^N}_{\simeq e^{-\langle k\rangle}}\underbrace{\frac{N(N-1)\cdots(N-k+1)}{N^k}}_{\simeq 1}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^{-k}}_{\simeq 1}
\end{aligned}
$$

The last two factors converge to 1:

$$
\lim_{N\to\infty}\frac{N(N-1)\cdots(N-k+1)}{N^k} = 1,
$$

$$
\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{-k} = 1,
$$

whereas

$$
\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{N} = e^{-\langle k\rangle}
$$

Thus,

$$
p(k) \simeq \frac{\langle k\rangle^k}{k!}e^{-\langle k\rangle}.
$$

$p(k)$ in the Erdős-Rényi model

From binomial to Poisson distribution

$$
\begin{aligned}
p(k) &= \frac{N(N-1)\cdots(N-k+1)}{k!}\frac{\langle k\rangle^k}{N^k}\left(1-\frac{\langle k\rangle}{N}\right)^N\left(1-\frac{\langle k\rangle}{N}\right)^{-k} \\
&= \frac{\langle k\rangle^k}{k!}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^N}_{\simeq e^{-\langle k\rangle}}\underbrace{\frac{N(N-1)\cdots(N-k+1)}{N^k}}_{\simeq 1}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^{-k}}_{\simeq 1}
\end{aligned}
$$

The last two factors converge to 1:

$$
\lim_{N\to\infty}\frac{N(N-1)\cdots(N-k+1)}{N^k} = 1,
$$

$$
\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{-k} = 1,
$$

whereas

$$
\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{N} = e^{-\langle k\rangle}
$$

Thus,

$$
p(k) \simeq \frac{\langle k\rangle^k}{k!}e^{-\langle k\rangle}.
$$

$p(k)$ in the Erdős-Rényi model

From binomial to Poisson distribution

$$
\begin{aligned}
p(k) &= \frac{N(N-1)\cdots(N-k+1)}{k!}\frac{\langle k\rangle^k}{N^k}\left(1-\frac{\langle k\rangle}{N}\right)^N\left(1-\frac{\langle k\rangle}{N}\right)^{-k} \\
&= \frac{\langle k\rangle^k}{k!}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^N}_{\simeq e^{-\langle k\rangle}}\underbrace{\frac{N(N-1)\cdots(N-k+1)}{N^k}}_{\simeq 1}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^{-k}}_{\simeq 1}
\end{aligned}
$$

The last two factors converge to 1:

$$
\lim_{N\to\infty}\frac{N(N-1)\cdots(N-k+1)}{N^k} = 1,
$$

$$
\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{-k} = 1,
$$

whereas

$$
\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{N} = e^{-\langle k\rangle}
$$

Thus,

$$
p(k) \simeq \frac{\langle k\rangle^k}{k!}e^{-\langle k\rangle}.
$$

$p(k)$ in the Erdős-Rényi model

From binomial to Poisson distribution

$$p(k) = \frac{N(N-1)\cdots(N-k+1)}{k!}\frac{\langle k\rangle^k}{N^k}\left(1-\frac{\langle k\rangle}{N}\right)^N\left(1-\frac{\langle k\rangle}{N}\right)^{-k}$$

$$= \frac{\langle k\rangle^k}{k!}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^N}_{\simeq e^{-\langle k\rangle}}\underbrace{\frac{N(N-1)\cdots(N-k+1)}{N^k}}_{\simeq 1}\underbrace{\left(1-\frac{\langle k\rangle}{N}\right)^{-k}}_{\simeq 1}$$

The last two factors converge to 1:

$$\lim_{N\to\infty}\frac{N(N-1)\cdots(N-k+1)}{N^k} = 1,$$

$$\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^{-k} = 1,$$

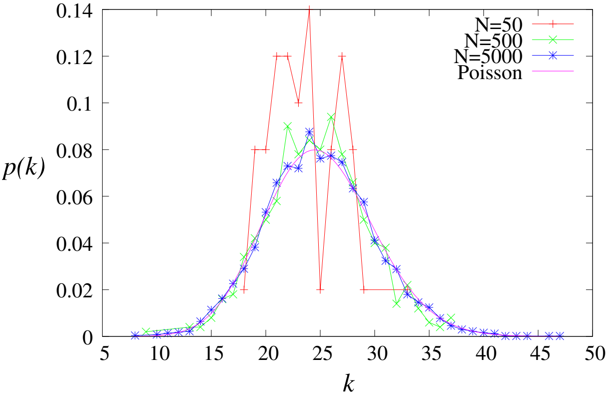
whereas

$$\lim_{N\to\infty}\left(1-\frac{\langle k\rangle}{N}\right)^N = e^{-\langle k\rangle}$$

Thus,

$$p(k) \simeq \frac{\langle k\rangle^k}{k!}e^{-\langle k\rangle}.$$

$p(k)$ in the Erdős-Rényi model



$p(k)$ in the Erdős-Rényi model