Structure and dynamics of complex networks

February 18, 2020

DISTANCE AND PATHS

Distance

Distance

We need to define a path first...

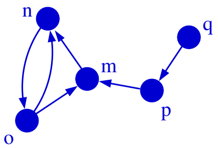
Pathology

Path

- A sequence of nodes in which each node is adjacent to the next one.

$\rightarrow$ it may intersect itself,

$\rightarrow$ in directed networks it must follow the direction of the links.

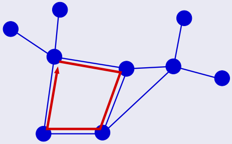


E.g., a legitimate path:
$p \rightarrow m \rightarrow n \rightarrow o \rightarrow m \rightarrow n$

Pathology
Special paths

Cycle

A path with the same start and end node



Self avoiding path

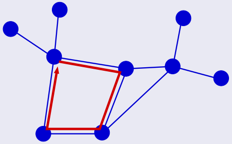
A path that does not intersect itself.

Pathology

Special paths

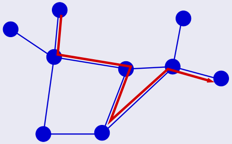
Cycle

A path with the same start and end node



Self avoiding path

A path that does not intersect itself.

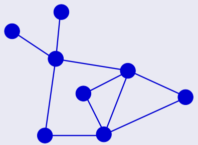


Pathology

Exotic paths

Eulerian path

A path traversing each link exactly once.



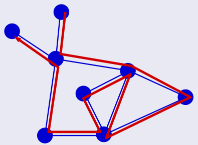
(E.g., the problem of the bridges of Königsberg.)

Hamiltonian path

A path visiting each node exactly once.

Eulerian path

A path traversing each link exactly once.



(E.g., the problem of the bridges of Königsberg.)

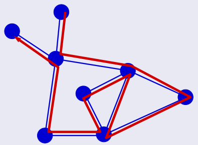
Hamiltonian path

A path visiting each node exactly once.

Pathology
Exotic paths

Eulerian path

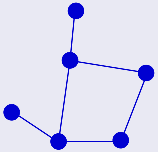
A path traversing each link exactly once.



(E.g., the problem of the bridges of Königsberg.)

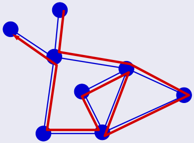
Hamiltonian path

A path visiting each node exactly once.



Eulerian path

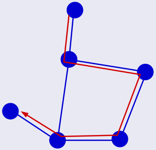
A path traversing each link exactly once.



(E.g., the problem of the bridges of Königsberg.)

Hamiltonian path

A path visiting each node exactly once.



Shortest path

Distance (shortest path length)

- ℓ_{ij}: minimum number of steps it takes to reach j from i following the links.
- Properties:
  - $\ell_{ii} = 0$
  - In undirected networks $\ell_{ij} = \ell_{ji}$.
  - If we cannot reach j from i, then $\ell_{ij} = \infty$.
- weighted networks?

Shortest path

Distance (shortest path length)

- ℓ_{ij}: minimum number of steps it takes to reach j from i following the links.
- Properties:
  - $\ell_{ii} = 0$
  - In undirected networks $\ell_{ij} = \ell_{ji}$.
  - If we cannot reach j from i, then $\ell_{ij} = \infty$.
- weighted networks?

Shortest path

Distance (shortest path length)

- ℓ_{ij}: minimum number of steps it takes to reach j from i following the links.
- Properties:
  - $\ell_{ii} = 0$
  - In undirected networks $\ell_{ij} = \ell_{ji}$.
  - If we cannot reach j from i, then $\ell_{ij} = \infty$.
- weighted networks: minimum weight path.

Shortest path

Average distance (average shortest path length)

- For a given node i:

$$\langle \ell_i \rangle = \frac{1}{N-1} \sum_j \ell_{ij},$$

for the whole network

$$\langle \ell \rangle = \frac{2}{N(N-1)} \sum_{i<j} \ell_{ij}$$

- $\langle \ell \rangle$ is also called as the average shortest path length between nodes.

Shortest path

Average distance (average shortest path length)

- For a given node i:

$$\langle \ell_i \rangle = \frac{1}{N-1} \sum_j \ell_{ij},$$

  for the whole network

$$\langle \ell \rangle = \frac{2}{N(N-1)} \sum_{i<j} \ell_{ij}$$

- $\langle \ell \rangle$ is also called as the average shortest path length between nodes.

Calculating shortest path length

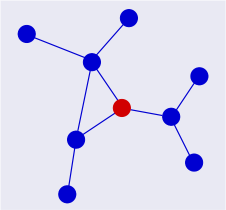
Calculating shortest path length

- breadth first search,
- depth first search,
- Dijkstra's algorithm

Breadth first search

Breadth first search

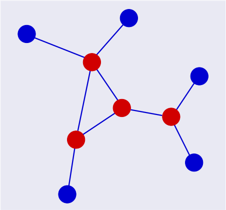
- Start from the root node, explore all of its neighbors.
- Next, explore all unvisited neighbors of the first neighbors.
- And so on, continue until no more unvisited nodes can be reached.



Breadth first search

Breadth first search

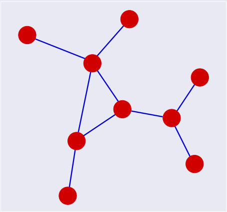
- Start from the root node, explore all of its neighbors.
- Next, explore all unvisited neighbors of the first neighbors.
- And so on, continue until no more unvisited nodes can be reached.



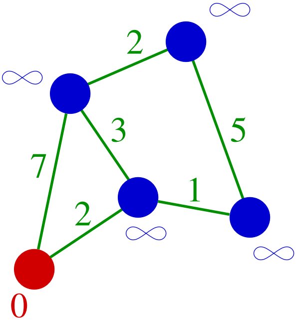
Breadth first search

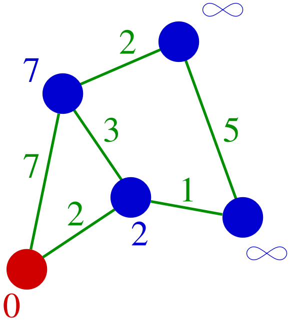
Breadth first search

- Start from the root node, explore all of its neighbors.
- Next, explore all unvisited neighbors of the first neighbors.
- And so on, continue until no more unvisited nodes can be reached.



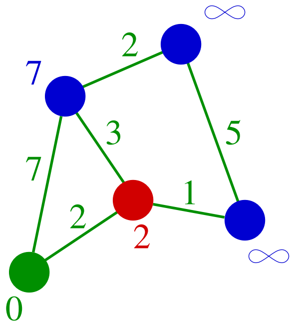
Dijkstra's algorithm
Illustration



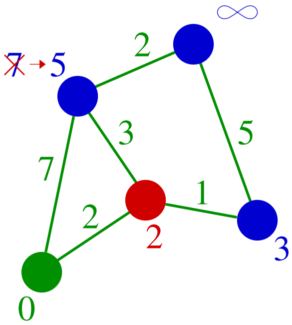


∞
2
7
3
5
7
1
2
2
∞
0

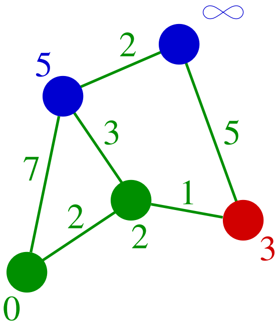
Dijkstra's algorithm
Illustration



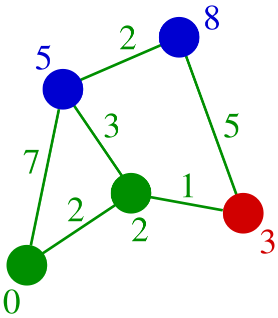
Dijkstra's algorithm
Illustration



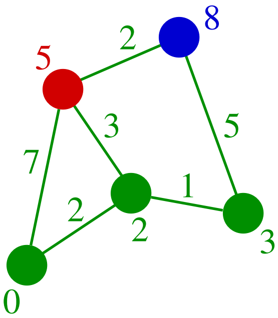
Dijkstra’s algorithm
Illustration



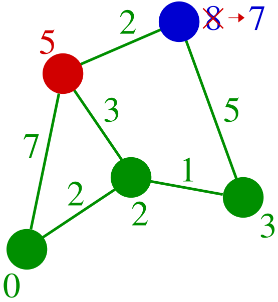
Dijkstra's algorithm
Illustration



Dijkstra's algorithm
Illustration

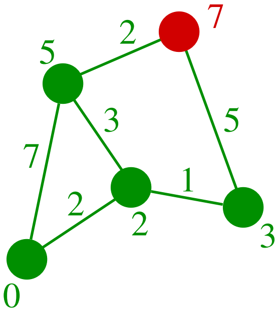


Dijkstra's algorithm
Illustration

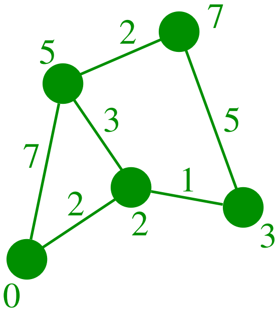


Dijkstra's algorithm

Illustration



Dijkstra's algorithm
Illustration



Dijkstra’s algorithm

Dijkstra’s algorithm

This algorithm calculates the minimum weight paths to all nodes from a given source node:

- Assign to every node a weight: set it to zero for the source node and to infinity for all other nodes.
- Mark all nodes as unvisited. Set source node as current.
- For current node, calculate the tentative weight of its unvisited neighbors by adding the link weight to the weight of the current node. If this weight is less than the previously recorded weight on the neighbor, (infinity in the beginning), overwrite the weight.
- When done with considering all neighbors of the current node, mark it as visited. A visited node will not be checked ever again; its weight is already minimal.
- Set the unvisited node with the smallest weight to be the next current node. If no more unvisited nodes are left, finish.
- The complexity of the algorithm for sparse graphs can be estimated to be $\mathcal{O}(M \log(N))$.

Diameter

Diameter

Diameter

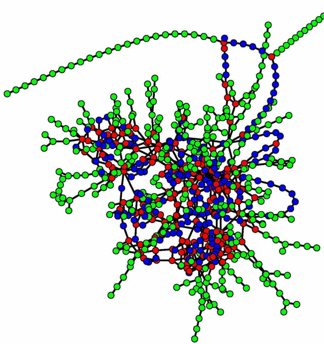
- The diameter of a network is given by the maximum distance between any pair of nodes,

$$D = \max_{ij} \ell_{ij}.$$

Diameter
Drawback

Diameter
Drawback

Usually the average shortest path length is more descriptive:



The world is small...

The world is small...

The network of acquaintances:

Although the system size is of the magnitude of $\mathcal{O}(10^7)$, the average distance $\langle \ell \rangle$ between people is only $\mathcal{O}(10)$....

The world is small...

The network of acquaintances:

Although the system size is of the magnitude of $\mathcal{O}(10^7)$, the average distance $\langle \ell \rangle$ between people is only $\mathcal{O}(10)$....

How do we know?

The world is small...

The network of acquaintances:

Although the system size is of the magnitude of $\mathcal{O}(10^7)$, the average distance $\langle \ell \rangle$ between people is only $\mathcal{O}(10)$....

How do we know?

$\longrightarrow$ The Milgram experiment.

Frigyes Karinthy

The six degrees of separation was first suggested by **Frigyes Karinthy**, a Hungarian writer in the 1920s...

(1887-1938)

Chains (1929)
"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but to the best of my knowledge a good friend of mine. So I could easily ask him to send a telegram via the general director telling Ford that he should talk to the manager and have the worker in the shop quickly hammer together a car for me, as I happen to need one."

The Milgram experiment (1967)

HOW TO TAKE PART IN THIS STUDY

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.
2. DETACH ONE POSTCARD. FILL IT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.
3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.
4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.

The Milgram experiment

Results of the experiment:

- Surprisingly a part of the letters did reach the target, and the average number of steps needed was around 6.
- Thus, a rough estimate of $\langle \ell \rangle$ for people in the USA is around 6, which is several orders of magnitude smaller than the size of the population...
- Indeed, the world is small.

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.

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 $\langle \ell \rangle \sim \ln N$?

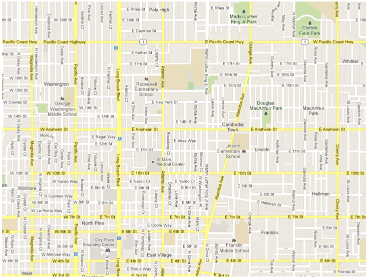
$\rightarrow$ If we consider the “concentric shells” of 1$^{\text{st}}$, 2$^{\text{nd}}$, 3$^{\text{d}}$, etc. neighborhoods:

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$^{\text{st}}$, 2$^{\text{nd}}$, 3$^{\text{d}}$, etc. neighborhoods:

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

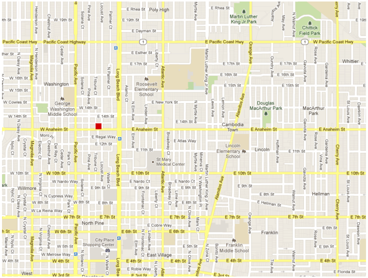


The small world property
Implications

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:

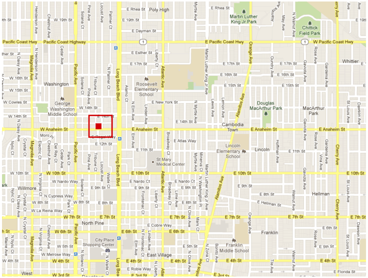


The small world property
Implications

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:

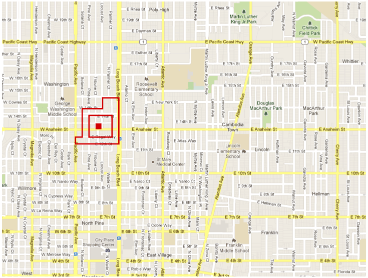


The small world property
Implications

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

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

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

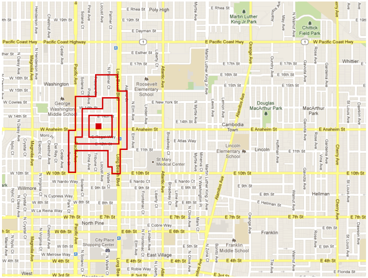


The small world property
Implications

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

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

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

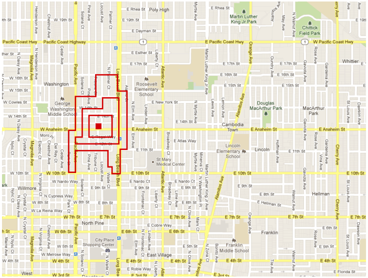


The small world property
Implications

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:



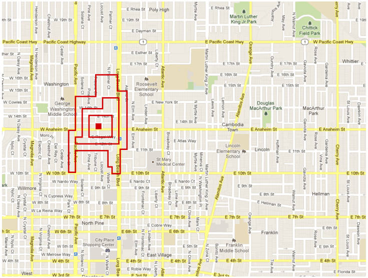
  - number of nodes (buildings, blocks, etc.) grows roughly as $n \sim \ell^d$, (e.g., as $n \sim \ell^2$ in a city)

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

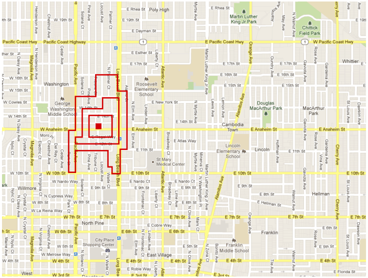
Basic network characteristics

Distance and paths
Pathology
Distance
The small world property
Centralities

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:

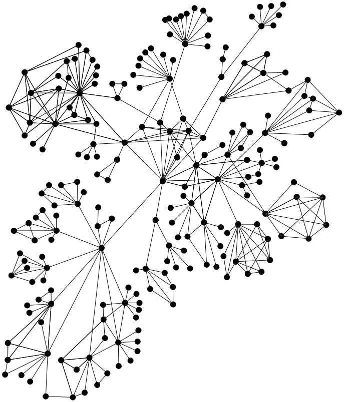


- number of nodes (buildings, blocks, etc.) grows roughly as $n \sim \ell^d$, (e.g., as $n \sim \ell^2$ in a city)
- to cover the whole system, we need roughly $\ell \sim N^{1/d}$!
- (e.g., in a city with 100.000 buildings, $\ell \simeq$ **300** !)

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

→ 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:



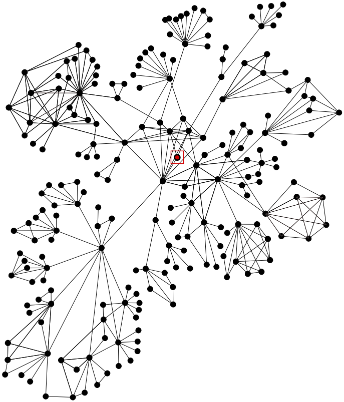
The small world property

Implications

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

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

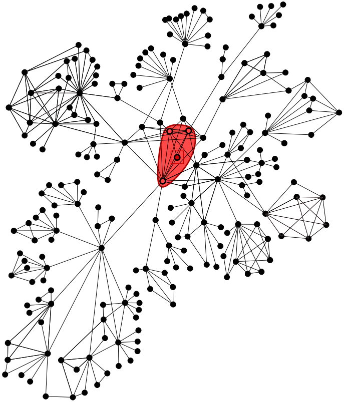
- in a random network, e.g., social network:



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:



Basic network characteristics

Distance and paths

Pathology

Distance

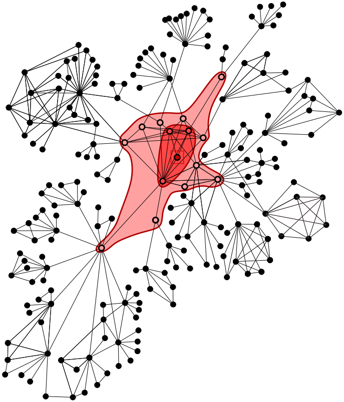
The small world property

Centralities

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

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

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

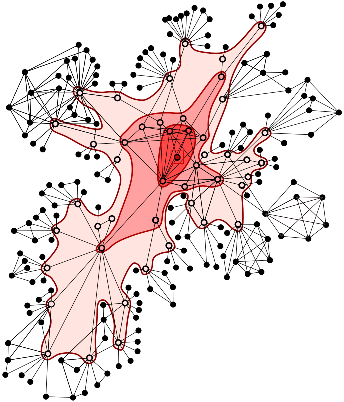


The small world property
Implications

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

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

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

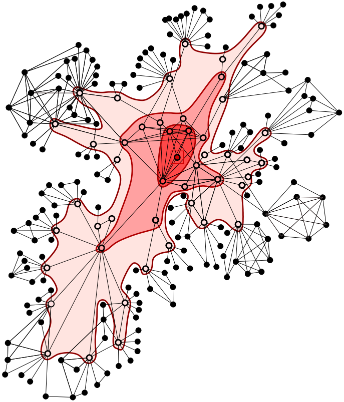


The small world property
Implications

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

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

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



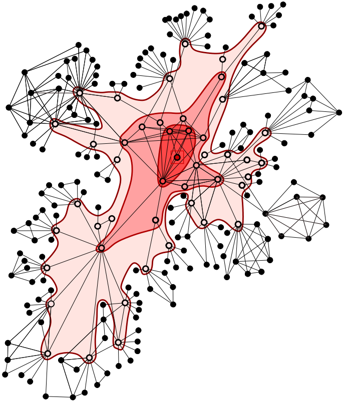
  - number of nodes grows **exponentially** as $n \sim \langle k \rangle^{\ell}$!

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:

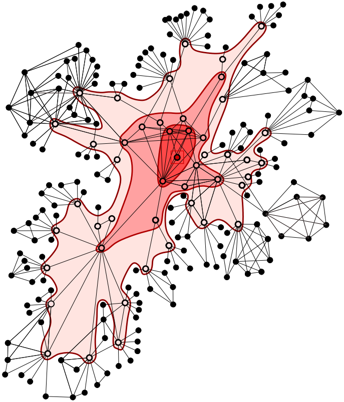


- number of nodes grows **exponentially** as $n \sim \langle k \rangle^{\ell}$!
- to cover the whole system, **only a couple of neighborhoods** are needed,

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

→ 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:



- 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

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

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.

Node centralities

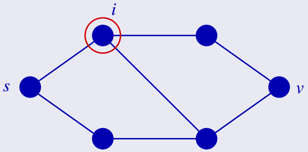
Betweenness

- 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:

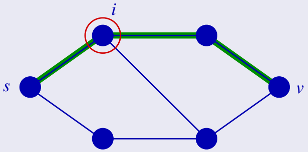
$$b_i \equiv \sum_{s \neq i, v \neq i} \frac{\sigma_{sv}(i)}{\sigma_{sv}}.$$



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:

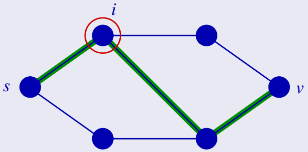
$$b_i \equiv \sum_{s \neq i, v \neq i} \frac{\sigma_{sv}(i)}{\sigma_{sv}}.$$



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:

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



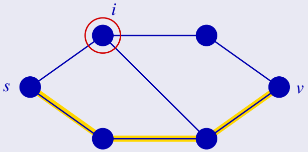
Basic network characteristics

Distance and paths
Pathology
Distance
The small world property
Centralities

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

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...

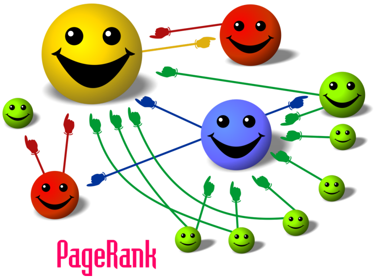
→ We can take the **largest eigenvalue** λ_1, and treat the components of the corresponding eigenvector $\mathbf{v}_1$ as the value of a centrality measure, associated to the corresponding node in the network.

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:

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

→ $\sum_{i=1}^{N} r_i$ is conserved!

→ 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 $r_i(t+1) = r_i(t) = r_i$.

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:

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

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:

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

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!