Lecture 12: PageRank, Connectivity, and Small-world Networks

Modelling Complex Systems (Spring 2026)

Jordi-Lluís Figueras

Matematiska Institutionen, Uppsala Universitet

May 5, 2026

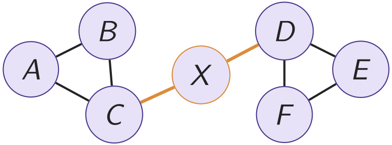
Prelecture discussion

Before Lecture 12, think about a transport or movement network that you know.

- Which roads, bridges, passes, or stations limit movement between regions?
- Which parts would you protect if many people or vehicles had to cross the network?
- What changes if one hidden or newly built route bypasses the bottleneck?
- Can a few long-range links make the whole network easier to cross?

Questions to reflect on:

- What information is lost if we only count the number of incident edges?
- How do shortest distance and route capacity differ?
- When can local structure remain clustered while global distances become short?



two clustered groups joined by a bridge

Lecture goals

In this lecture we will:

- briefly complete the PageRank algorithm with dangling-node correction and teleportation;
- define the Google matrix and explain why Perron-Frobenius gives a unique attracting rank vector;
- shift from ranking nodes to understanding connectivity in networks;
- use Thermopylae as a graph-theoretic example of paths, bottlenecks, cuts, and capacity;
- explain how alternate routes can change shortest paths and network vulnerability;
- define clustering and average path length as structural network statistics;
- introduce the Watts-Strogatz model and the small-world mechanism.

Completing PageRank

Where Lecture 11 left us

In Lecture 11 we introduced the simplified PageRank iteration

$$r_{t+1} = Mr_t.$$

The matrix M sends the rank of each page to the pages it links to.

But the simplified model had two structural problems:

- a page with no outgoing links produces a zero column and rank disappears;
- a directed web graph need not be strongly connected, so the limit may depend on the initial state.

The standard PageRank algorithm fixes both problems by turning the link dynamics into a positive stochastic matrix.

Dangling pages

A dangling page is a page with no outgoing links.

In the random-surfer interpretation, the surfer arrives at such a page but has no hyperlink to follow.

A common correction is:

Dangling-node correction

If a page has no outgoing links, replace its missing outgoing distribution by the uniform distribution over all pages.

Thus a dangling page is treated as if it linked weakly to every page. This restores conservation of total probability.

From the link matrix to a stochastic matrix

Suppose there are N pages. Let S be the corrected transition matrix.

For a non-dangling page j with outgoing degree $L(j) > 0$,

$$S_{ij} = \begin{cases} 1/L(j) & \text{if } j \text{ links to } i, \\ 0 & \text{otherwise.} \end{cases}$$

For a dangling page j,

$$S_{ij} = \frac{1}{N} \qquad \text{for every } i.$$

Then each column of S sums to 1, so S is column-stochastic.

Teleportation

The second correction is teleportation.

At each step, the random surfer:

- with probability d, follows the corrected link dynamics S;
- with probability $1 - d$, jumps to a page chosen from a fixed distribution v.

The parameter d is the damping factor; the classical value is approximately

$$d = 0.85.$$

For uniform teleportation, $v = \frac{1}{N}\mathbf{1}$.

The Google matrix

With our column-stochastic convention, the damped PageRank update is

$$r_{t+1} = G r_t,$$

where

$$G = dS + (1-d)v\mathbf{1}^T.$$

If r_t is a probability vector, then $\mathbf{1}^T r_t = 1$, so

$$G r_t = dS r_t + (1-d)v.$$

For uniform teleportation,

$$G = dS + \frac{1-d}{N}\mathbf{1}\mathbf{1}^T.$$

Why damping fixes convergence

Assume $0 < d < 1$ and $v_i > 0$ for every page i. Then

$$G_{ij} = dS_{ij} + (1-d)v_i \geq (1-d)v_i > 0.$$

So G is a positive column-stochastic matrix.

Perron-Frobenius now gives the qualitative behavior we wanted:

- there is a unique stationary distribution r^*;
- every page has positive rank;
- the iteration $r_{t+1} = Gr_t$ converges to r^* from every initial probability vector.

From PageRank to connectivity

PageRank gives one answer to the question “which nodes are important?”

It is appropriate when edges describe directed attention or movement of a random surfer.

Now we change the modelling question.

Connectivity question

Which nodes or edges control movement between different parts of a network?

This requires graph-theoretic language for paths, distances, connected components, cuts, bottlenecks, and capacities.

Networks and Connectivity

Thermopylae and *300* as a graph-theoretic case study

Case study: Thermopylae as a graph

Image: *300* poster, File:300poster.jpg, Wikipedia; copyrighted non-free image, cultural reference only.

The historical setting is useful because it separates the dramatic picture from the mathematical model.

In graph-theoretic language:

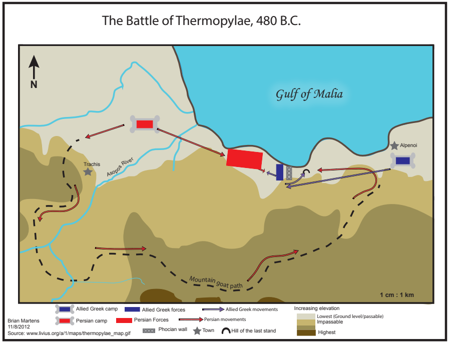
- nodes represent positions or regions: camps, defensive line, rear position, nearby settlements;
- edges represent feasible movement corridors, such as the coastal pass or the mountain path;
- weights may represent travel time, terrain difficulty, or military cost;
- capacities may represent how many soldiers can pass per unit time.

The modelling question is not "who had more soldiers?" but:

Network question

Which nodes or edges control the movement of a large army through the terrain graph?

The pass as a bottleneck



Map: BMartens, *Battle of Thermopylae*, Wikimedia Commons, CC BY-SA 3.0.

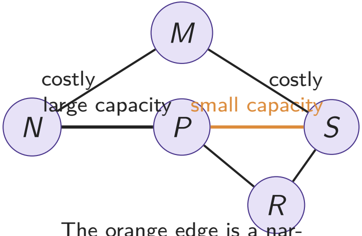
Thermopylae was strategically important because it was a narrow movement corridor.

In network terms:

- the bottleneck is not important because its endpoints have many incident edges;
- it is important because many feasible north–south routes must use the corridor;
- holding the pass removes or strongly constrains a key edge of the movement graph;
- in this model, the pass has high edge betweenness and belongs to a small cut set.

First lesson after PageRank: a low-degree local structure can be globally decisive.

A bottleneck for a large army



The orange edge is a narrow pass: route capacity is limited by the weakest edge.

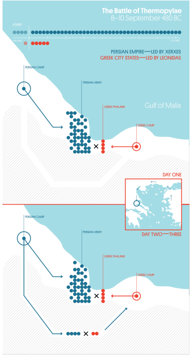
For a large army, distance alone is not enough.

We may use a capacitated graph:

- each edge e has capacity $c(e)$;
- the capacity of a route is limited by its narrowest segment;
- a defensive position tries to reduce the effective flow through a cut;
- the relevant object is a small cut capacity between north and south.

Thermopylae illustrates why a low-degree edge can dominate a transport or movement network.

The hidden path as an alternate route



Scheme: Nicolo Arena/DensityDesign Research Lab,

Wikimedia Commons, CC BY-SA 4.0.

The mountain path changes the graph.

Before the alternate route is used:

- the pass carries most feasible movement between the two sides;
- the Greek position controls a high-betweenness corridor;
- the effective north–south cut has very small capacity.

After the alternate route is available:

- the movement graph gains a second path to the Greek rear;
- shortest weighted routes and edge-betweenness scores change;
- a local bottleneck is bypassed by a shortcut-like edge.

This prepares the later small-world idea: a few nonlocal links can strongly reduce global distances.

Connectivity and Bottlenecks

Paths and weighted distances

A path from node i to node j is a sequence of edges connecting i to j.

In an unweighted graph, the distance $\operatorname{dist}(i,j)$ is the smallest number of edges in such a path.

In a weighted graph, an edge weight may represent time, cost, terrain difficulty, or risk:

$$\operatorname{dist}(i,j) = \min_{\gamma: i \to j} \sum_{e \in \gamma} w(e).$$

For a connected graph, the average path length is

$$\ell(G) = \frac{1}{N(N-1)} \sum_{i \neq j} \operatorname{dist}(i,j),$$

and the diameter is $\operatorname{diam}(G) = \max_{i,j} \operatorname{dist}(i,j)$.

Components, cuts, and bottlenecks

In an undirected graph, a connected component is a maximal set of nodes that can reach each other by paths.

A cut set is a set of nodes or edges whose removal separates part of the graph.

Important special cases:

- a cut edge or bridge is an edge whose removal disconnects the graph;
- a cut vertex is a node whose removal disconnects the graph;
- a bottleneck may not disconnect the graph, but it strongly limits movement between regions.

Thermopylae is a bottleneck example: the terrain graph is not just sparse; movement between north and south is concentrated through a narrow corridor.

Capacity and large-scale movement

For a large army, a route is not only a path. It also has a limited capacity.

Let $c(e)$ be the capacity of edge e. If U is one side of a network cut, its cut capacity is

$$C(U, U^c) = \sum_{e \in \delta(U)} c(e),$$

where $\delta(U)$ is the set of edges crossing from U to U^c.

The movement between two regions is constrained by small-capacity cuts.

This explains why a narrow pass can be strategically decisive even if it has low degree.

Bottleneck importance

One way to quantify bottleneck importance is edge betweenness.

Let σ_{st} be the number of shortest paths from s to t, and let $\sigma_{st}(e)$ be the number of those paths that use edge e.

The edge betweenness of e is

$$C_B(e) = \sum_{s<t} \frac{\sigma_{st}(e)}{\sigma_{st}}.$$

Edges with high edge betweenness are bridges, passes, corridors, or routes used by many shortest paths.

This is a global notion: it cannot be read off from the local degree alone.

Directed connectivity

Directed graphs require more care.

A directed graph is strongly connected if every node can reach every other node by a directed path.

A strongly connected component is a maximal set of nodes with this mutual reachability property.

This connects back to PageRank:

- closed directed components can trap probability mass;
- reducibility creates multiple possible long-time outcomes;
- teleportation removes this obstruction by allowing jumps between all nodes.

Local clustering

Networks often contain triangles: neighbors of a node may also be connected to each other.

Let node i have degree k_i. Among its neighbors there are at most

$$\frac{k_i(k_i-1)}{2}$$

possible edges.

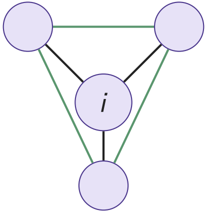
If e_i of these edges are present, the local clustering coefficient is

$$C_i = \frac{2e_i}{k_i(k_i-1)}.$$

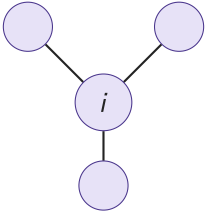
High clustering means that local neighborhoods are tightly interconnected.

Clustering: visual interpretation

High local clustering



Low local clustering



Clustering measures whether local neighborhoods look like cohesive groups rather than independent contacts.

Small-world Networks

The small-world phenomenon

The Thermopylae example showed how an alternate route can change global reachability.

The Watts-Strogatz model abstracts this mechanism: mostly local connectivity plus a few shortcuts.

Many real networks have two properties at the same time:

- high clustering: neighbors of a node are often connected to each other;
- short average path length: most nodes are separated by only a few steps.

This combination is called the small-world property.

Regular, random, and small-world

Network type	Clustering	Typical distances
Regular lattice	high	long
Random graph	low	short
Small-world network	high	short

The surprising regime is the third one.

A few long-range shortcuts can drastically reduce $\ell(G)$ while most local triangles remain intact.

Watts-Strogatz construction

The Watts-Strogatz model gives a simple interpolation between a regular ring lattice and a random graph.

Start with N nodes on a ring, each connected to k nearby neighbors. Then rewire each local edge with probability p, avoiding self-loops and duplicate edges.

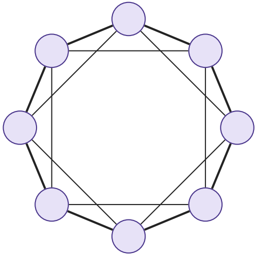
The parameter p controls the amount of randomness:

- $p = 0$: regular ring lattice;
- small $p > 0$: a few shortcuts;
- $p = 1$: highly randomized network.

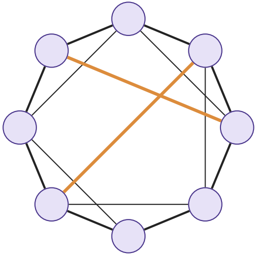
The small-world regime occurs for intermediate small values of p.

A few shortcuts change distances

Ring-like network



With shortcuts



Shortcuts act as global bridges. Like the alternate path at Thermopylae, they can reduce average path length long before they destroy local clustering.

Why small worlds matter for dynamics

Small-world structure affects many processes on networks.

Examples:

- epidemics can spread rapidly once shortcuts connect distant local clusters;
- information can travel quickly without every node having many contacts;
- synchronization may be accelerated by long-range coupling;
- failures or cascades may jump between communities through bridge links.

This is the bridge to the next topic: spreading processes on networks.

NetworkX Examples

What is NetworkX?

`NetworkX` is a Python package for constructing, analyzing, and simulating graphs.

For this lecture, it lets us:

- build directed and undirected graphs;
- compute components, shortest paths, average path length, clustering, and diameter;
- test how bottlenecks and shortcuts affect connectivity;
- generate Watts-Strogatz small-world graphs;
- compare the mathematical definitions with concrete numerical examples.

The library does not replace the definitions. It helps us experiment with them.

On Debian/Ubuntu systems, use a virtual environment before installing packages.

Installing NetworkX locally

If `networkx` is not available, create a local virtual environment from the course root:

```
python3 -m venv .venv
source .venv/bin/activate
python3 -m pip install networkx
```

Then run the scripts while the environment is active, for example:

```
python3 smallWorldWattsStrogatz.py
```

This avoids modifying the system-managed Python installation.

Scripts for this lecture

There are two example files for this lecture.

- `pagerankDampingExample.py`: PageRank on the four-page graph from Lecture 11, including dangling-node correction and damping.
- `smallWorldWattsStrogatz.py`: varies the rewiring probability in the Watts-Strogatz model and prints clustering and path-length statistics.

Typical usage:

```
python3 smallWorldWattsStrogatz.py
```

Classroom experiment

Run the small-world script with different rewiring probabilities.

Observe three quantities:

- average clustering coefficient;
- average shortest path length;
- diameter of the largest connected component.

Guiding question:

Small-world transition

How much random rewiring is needed before path lengths become short, and how much clustering remains at that point?

This is a numerical version of the Watts-Strogatz insight.

Exercise

Connect the Thermopylae example with the small-world model.

Work through the following steps:

1. Draw a simplified terrain graph with north, pass, south, mountain path, and rear position.
2. Mark one small cut set that limits movement of a large army.
3. Add the alternate mountain route and predict what changes in shortest paths and edge betweenness.
4. Run the Watts-Strogatz script and compare this shortcut effect with increasing the rewiring probability p.

The point is to see how one alternate route can change global connectivity without changing most local neighborhoods.

What to remember

- Damped PageRank uses a positive stochastic matrix, so Perron-Frobenius gives a unique attracting rank vector.
- Connectivity questions ask which paths, cuts, and bottlenecks control movement through a graph.
- Thermopylae illustrates why low-degree edges can be strategically decisive when they carry many routes or have small capacity.
- Alternate routes can change shortest paths, edge betweenness, and vulnerability.
- Clustering measures local cohesion; average path length measures global reachability.
- Watts-Strogatz networks show how a few shortcuts can produce high clustering and short global distances.

Suggested bibliography

For students who want to read more:

- D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, 393, 440–442, 1998.
- M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, 2010.
- A.-L. Barabasi, *Network Science*, Cambridge University Press, 2016.
- A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond*, Princeton University Press, 2006.
- NetworkX documentation: `https://networkx.org/documentation/`.

Acknowledgment

These slides were prepared with the assistance of OpenAI Codex / ChatGPT 5.5 as a drafting and editing tool.
All mathematical, pedagogical, and course-specific content remains under the responsibility of the lecturer.