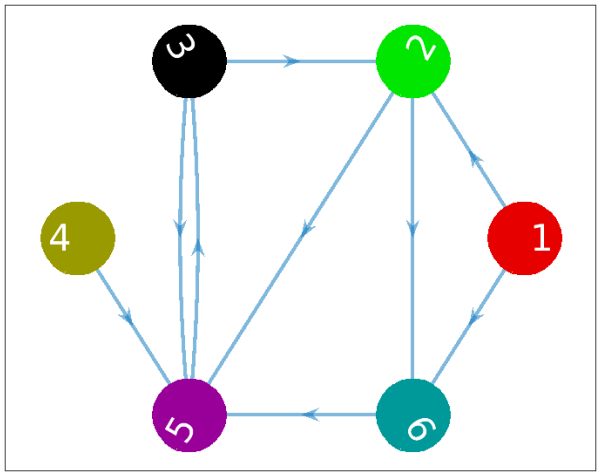
CS319: Scientific Computing (with MATLAB)

(PageRank) and Review of CS319

Niall Madden

Week 12: 28 March 2023



Important: you should read:

- The MATLAB Guide, Chapters 19 and 21: https://doi-org.nuigalway.idm.oclc.org/10.1137/1.9781611974669

This week...

1. 1. Projects (last time!)
2. 3: Directed Graphs
3. 6: PageRank
   - Other Applications
4. Module review
   - Not covered
   - THE END!

1. Projects (last time!)

The slides for this section are at https://www.niallmadden.ie/2223-CS319/2223-CS319-Projects.pdf

3: Directed Graphs

Last week, we learned about the MATLAB class for representing (undirected) **graphs**. Now we'll look at the class for **undirected** graphs, called *digraph*,,

The notes for this section are in the MATLAB live script, *CS319_Week12_PageRank.mlx*.

You can access it in the `Week12` folder in `https://bitbucket.org/niallmadden/2223-cs319/src/main/` or from `https://www.niallmadden.ie/2223-CS319/`

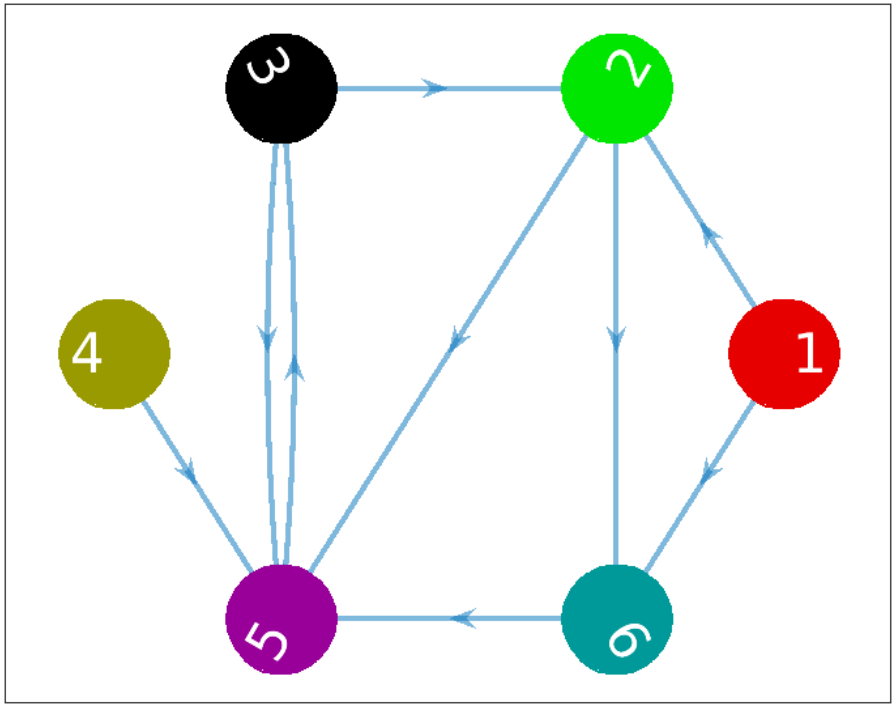
6: PageRank

Google initial break-through in search engine design was derived from their **PageRank** algorithm which gives an objective way of computing the relative importance of web-pages.

The basic idea is this: **the importance of a web-page is the probability that you are looking at it at any given time**.

6: PageRank

To see how this works, consider the following example, which was made in the live script.



6: PageRank

The method:

(1) Form the **adjacency matrix**, $A = (a_{ij})_{i=1}^{N}$, for the network:

$$a_{i,j} = \begin{cases} 1 & \text{if the graph has an edge from Node } i \text{ to Node } j; \\ 0 & \text{otherwise.} \end{cases}$$

(2) Make the associated **Markov matrix**, $S = (s_{ij})_{i=1}^{N}$, where $S_{i,j}$ is the *proportion* of vertices in A which start at i and go to j. (That is, divide the entries in row i by the sum of the entries in that row). If there are no entries in a given row of A, set the corresponding entries of S to $1/N$.

(3) Choose a "damping" value σ, e.g., $\sigma = 0.85$.

(4) Set the matrix G to be $\left(\sigma S + (1-\sigma)/N\right)^T$.

(5) Finally, find a "fixed point" vector x: should be $Gx = x$. This can be done with the **Power Method**. (Yes, it's an eigenvector...).

6: PageRank

From our example earlier, the first few results are:

Iteration	0	1	2	3
u	$\begin{pmatrix} 0.1667 \\ 0.1667 \\ 0.1667 \\ 0.1667 \\ 0.1667 \\ 0.1667 \end{pmatrix}$	$\begin{pmatrix} 0.0250 \\ 0.1667 \\ 0.1667 \\ 0.0250 \\ 0.4500 \\ 0.1667 \end{pmatrix}$	$\begin{pmatrix} 0.0250 \\ 0.1065 \\ 0.4075 \\ 0.0250 \\ 0.3296 \\ 0.1065 \end{pmatrix}$	$\begin{pmatrix} 0.0250 \\ 0.1746 \\ 0.3084 \\ 0.0250 \\ 0.3612 \\ 0.1059 \end{pmatrix}$

These days, PageRank is only a minor part of what the Google Search Engine does. However, the methodology has been applied to lots of different application domains including

- Biological sciences and genomics,
- Election systems (at least in theory)
- Ranking Sports Teams (very effective if you want a complete ranking of all teams in an event that features group stages followed by knock-out.
- Ranking academic papers based on their citations,
- Etc, etc.

Reminder of CS319 Assessment

The final grade for CS319 is be based on

- **Four lab assignments** (40%).
- Class-test (20%)
- The project (40%)

This module does not have an end-of-semester exam.

- Resit of Class Test today at 16.00. Run on a no-risk basis.
- Deadline for Project Code and Report: 5pm Thu, 6 April.

Module review

The topics we have covered (not necessarily in order) are:

(a) MATLAB Basic I/O, including reading to and writing from files

(b) Flow of control and looping (`if`, `for`, `while`, etc.)

(c) Fundamental data-types (`int32`, `double`, `char`, `logical`, ...).

(d) strings

(e) Computer representation of numbers (underflow, overflow, machine epsilon, ...)

(f) Vectors and matrices; vector indexing.

Module review

(g) Functions: inline and "file" functions.

(h) Optimisation and bisection;

(i) Random number generation.

(j) Sorting algorithms, and their complexities;

(k) Data fitting with polynomials, and piecewise polynomials. Least squares. Convergence.

(l) Solving linear systems by the Jacobi and Gauss-Seidel methods, and using backslash.

Module review

(m) Sparse matrix representation, especially triplet and CCS; Matrix-vector multiplication

(n) Direct and iterative solvers, and prconditioning.

(o) Graphs, Digraphs, and adjacency matrices;

(p) PageRank and the Power Method;

(q) Classes, and operator overloading.

Some topics that would have been nice...

1. Non-polynomial data-fitting (optimisation)
2. More on linear solvers, especially
   - more preconditioners
   - why Krylov methods work
   - LU and Cholesky decomposition.
3. More on eigenvalue problems, especially for sparse matrics.
4. Differential equations.
5. Montecarlo methods.
6. Parallel and distributed computing.
7. The App Builder.

Module review

THE END!

I hope you have enjoyed CS319, and have learned something: I have!

Thank you for your participation, enthusiasm, forbearance, and willing to help each other.