Matrix Computations for Discovering Network Structures

Des Higham
School of Mathematics
University of Edinburgh

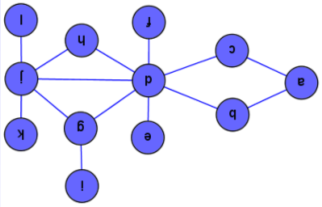
Overview

- Adjacency matrix
- Some interesting structures
- Spectral methods
- Walk-based methods

Motivation

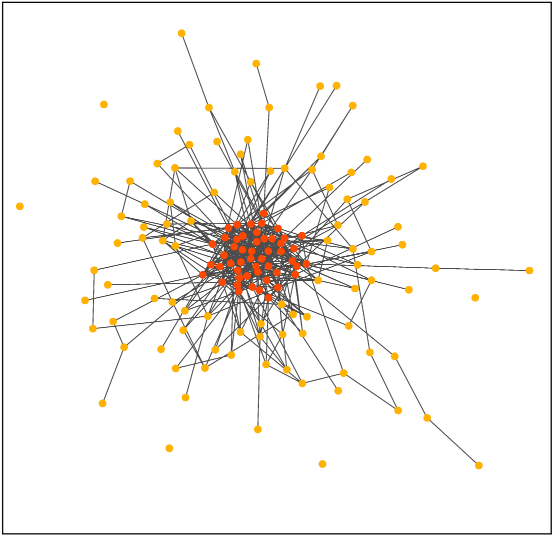
Focus first on simple graphs (undirected, unweighted, connected).

In classical graph theory, a graph is defined through a **list of nodes** and a **list of edges**. We don't care how the nodes are labelled.



In many applications, we want to quantify and visualize **structure**. Here, we may want to label the nodes a certain way (ranking), or place them in $\mathbb{R}^2$ ...

Core-Periphery Visualization



Matrix Viewpoint

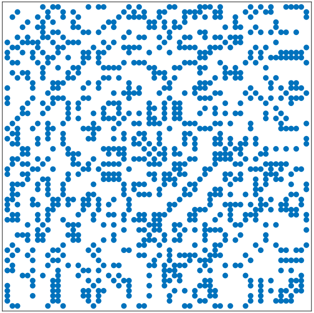
Here, we will focus on representing a graph through its adjacency matrix. For a graph with n nodes labelled $1, 2, \ldots, n$, we have $A \in \mathbb{R}^{n \times n}$ with

$$a_{ij} = 1$$

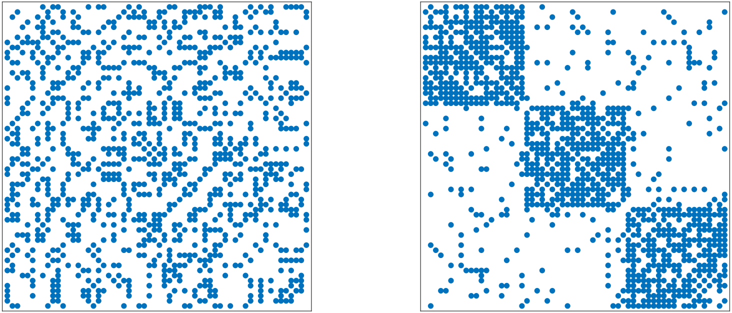
when nodes i and j are connected.

Relabelling the nodes corresponds to symmetrically **permuting** the rows and columns of the matrix.

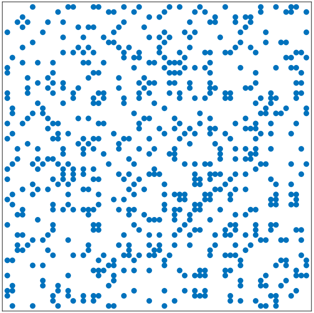
Clusters (communities)



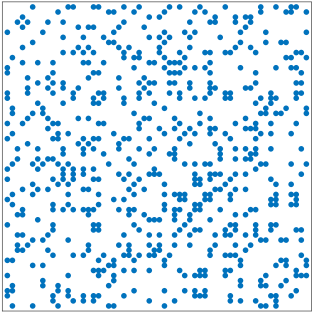
Clusters (communities)

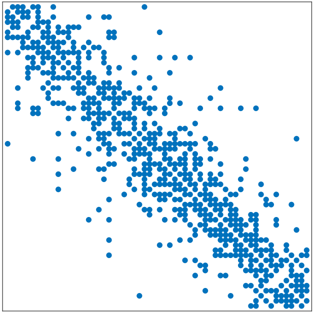


Lattice (small-world)

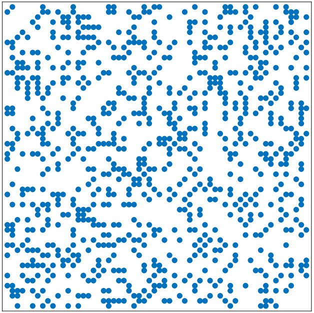


Lattice (small-world)

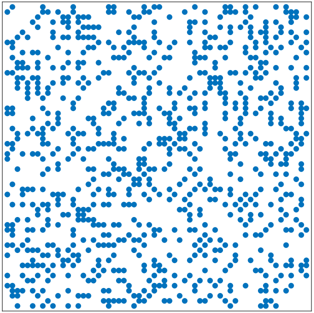


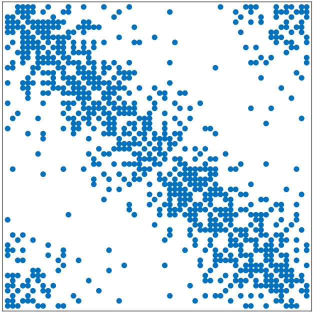


Periodic Lattice



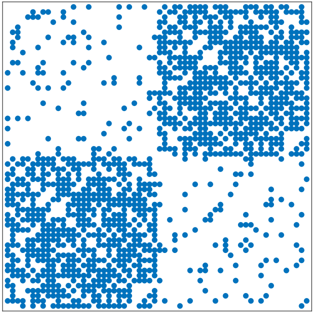
Periodic Lattice





Bipartite

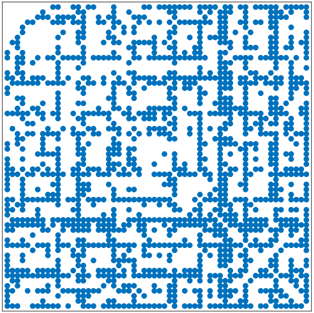
Bipartite



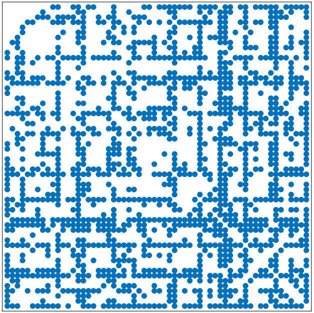
Core-Periphery

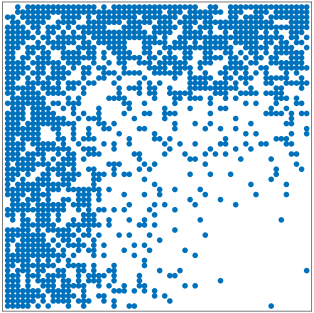
Core-Periphery

Core-Periphery: graded



Core-Periphery: graded





Focus on the Small World Structure

Loosely, a graph, with an appropriate ordering, exhibits a **small world structure** if most of the nonzeros are close to the diagonal (most edges are **short-range**)

This could be quantified through the **two-sum**

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \, (i - j)^2$$

Given a graph, with nodes in arbitrary order, we could look for small-world structure by **reordering the nodes to minimize the two-sum**

More Notation

Undirected, **unweighted** graph with n nodes

Adjacency matrix A:
$a_{ij} = 1$ if nodes i and j are connected, $a_{ij} = 0$ otherwise

Let $D = \mathrm{diag}(d)$, where $d_i = \sum_{j=1}^{n} a_{ij}$ is **degree** of node i

$\mathcal{P}_n$ denotes the set of all **permutations** of $\{1, 2, 3, \ldots, n\}$

$x \in \mathcal{P}_n$ denotes a particular permutation vector:
node i gets relabelled x_i

We can look for small world structure by tackling

$$\min_{x \in \mathcal{P}_n} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} (x_i - x_j)^2$$

Now manipulate that objective function…

Two-sum problem

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 =$$

Two-sum problem

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i^2 - 2x_i x_j + x_j^2)$$

Two-sum problem

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i^2 - 2x_i x_j + x_j^2)$$

$$= 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

Two-sum problem

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i^2 - 2x_i x_j + x_j^2)$$

$$= 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

$$= 2\sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} a_{ij} - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

Two-sum problem

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i^2 - 2x_i x_j + x_j^2) \\ &= 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\ &= 2\sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} a_{ij} - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\ &= 2\sum_{i=1}^{n} x_i^2 d_i - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j\end{aligned}$$

Two-sum problem

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i^2 - 2x_i x_j + x_j^2) \\
&= 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\
&= 2\sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} a_{ij} - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\
&= 2\sum_{i=1}^{n} x_i^2 d_i - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\
&= 2x^T D x - 2x^T A x
\end{aligned}$$

Two-sum problem

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i - x_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x_i^2 - 2x_ix_j + x_j^2) \\
&= 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j \\
&= 2\sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} a_{ij} - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j \\
&= 2\sum_{i=1}^{n} x_i^2 d_i - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j \\
&= 2x^T Dx - 2x^T Ax
\end{aligned}$$

So we want to minimize $x^T\,(D - A)\,x$

Two-sum problem ct'd

$$\min_{x \in \mathcal{P}_n} \quad x^T (D - A) x$$

Now

- **relax** to $x \in \mathbb{R}^n$
- **eliminate the trivial solutions** $x = \mathbf{0}$ and $x = \mathbf{1}$
- **normalize** with $\| x \|_2 = 1$

This **quadratic form** is minimized by the **Fiedler vector** $x = v^{[2]}$; the eigenvector corresponding to the **second smallest** eigenvector of $D - A$

We can recover a permutation using the size of the entries in x

Fiedler vector

Also used for **clustering** (arrived at by similar arguments)

The relaxation process **can be justified rigorously to some extent**

An alternative is based on the **normalized Laplacian** $I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$

We can also justify reordering with the Fiedler vector via a **random graph argument**: **Grindrod's trick** (2002). . .

Random Graph Viewpoint

Suppose we have a **range-dependent random graph** where every possible edge $i \leftrightarrow j$ appears with independent probability $f(|i-j|)$

Note that this concept involves a specific ordering of the nodes

Suppose we are given a graph, which we suspect to be a sample from a range-dependent random graph, **with the nodes in arbitrary order**

We may then ask: what is the most likely ordering of the nodes in the original model?

For any reordering $x \in \mathcal{P}_n$, the likelihood of seeing our graph is

$$\prod_{i,j:a_{ij}=1} f(|x_i - x_j|) \times \prod_{i,j:a_{ij}=0} (1 - f(|x_i - x_j|))$$

Equivalence

Lemma Two-sum reordering is the same as maximum likelihood reordering with the range-dependency

$$f(k) = \frac{e^{-k^2}}{1+e^{-k^2}} \qquad [\text{Note}: f(k)/(1-f(k)) = e^{-k^2}.]$$

Proof Rewrite likelihood as

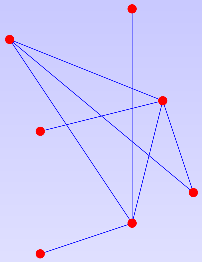
$$\prod_{i,j:a_{ij}=1} \frac{f(|x_i - x_j|)}{1 - f(|x_i - x_j|)} \times \prod_{i,j} (1 - f(|x_i - x_j|))$$

Problem reduces to maximizing

$$\prod_{i,j:a_{ij}=1} e^{-(i-j)^2}$$

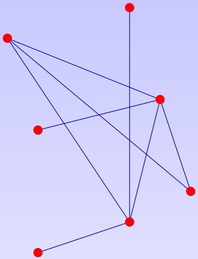
Take logs $\mapsto$ minimizing the two-sum!

The Power of Matrices



$(A^2)_{ij} = \sum_{p=1}^{n} a_{ip} a_{pj}$, **num. paths of length two** from i to j

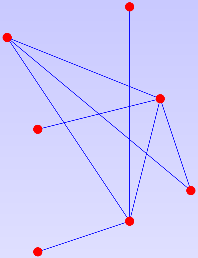
The Power of Matrices



$(A^2)_{ij} = \sum_{p=1}^{n} a_{ip} a_{pj}$, **num. paths of length two** from i to j

$(A^k)_{ij}$ gives **num. walks of length** k from i to j

The Power of Matrices

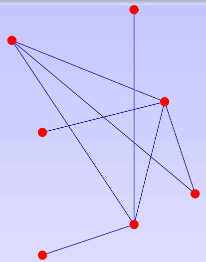


$(A^2)_{ij} = \sum_{p=1}^{n} a_{ip}a_{pj}$, **num. paths of length two** from i to j

$(A^k)_{ij}$ gives **num. walks of length k** from i to j

Also true when the edges are **directed**

The Power of Matrices, ct'd



So **matrix functions** such as

$$\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots + \frac{A^k}{k!} + \cdots$$

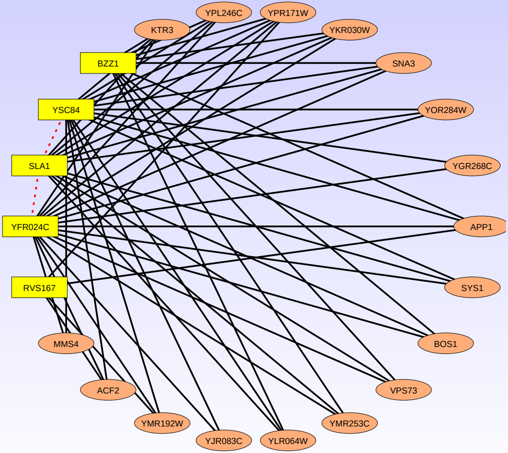
and

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots + \alpha^k A^k + \cdots$$

can reveal useful structure

Bipartivity

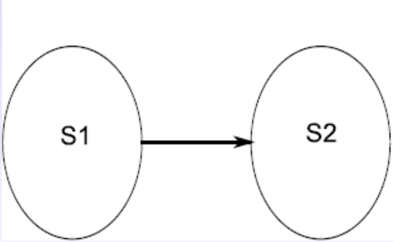
Bipartite groups are associated with **clusters** in A^2:

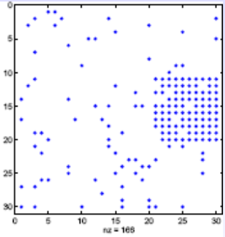


Approximate Directed Bipartivity

Suppose ∃ distinct subsets of nodes S_1 and S_2 such that

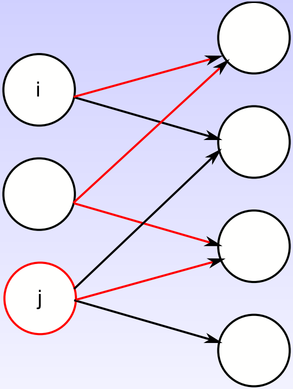
- S_1 has few internal links
- S_2 has few internal links
- there are many $S_1 \mapsto S_2$ links
- few other links involve S_1 or S_2





Alternating Walks

An **alternating walk** is a traversal that successively follows links in the forwards and reverse directions:



This motivates . . .

$$f(A) = I - A + \frac{AA^T}{2!} - \frac{AA^TA}{3} + \frac{AA^TAA^T}{4!} - \dots$$

We expect $f(A)_{ij}$ to take

- large positive values when $i, j \in S_1$ and
- large negative values when $i \in S_2$ and $j \in S_1$

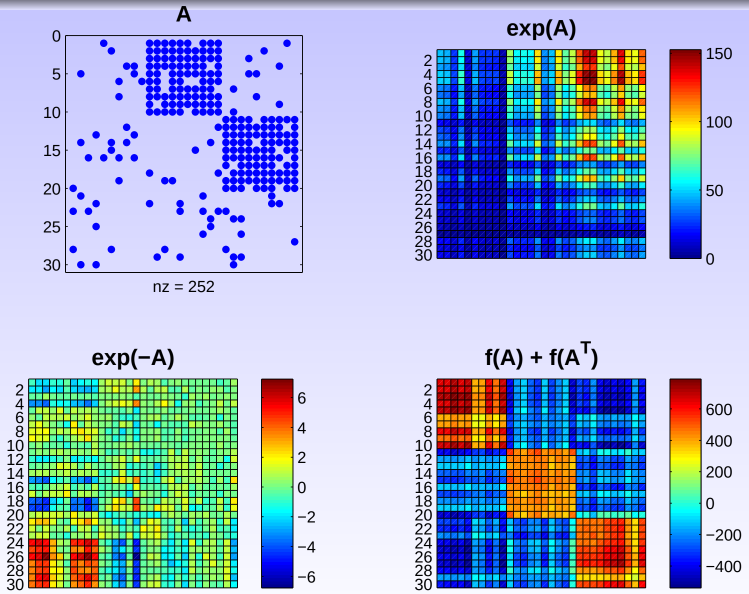
Similar comments apply to $f(A^T)$

So $f(A) + f(A^T)$ should have

- positive values for $S_1 \leftrightarrow S_1$ and $S_2 \leftrightarrow S_2$
- negative values for $S_1 \leftrightarrow S_2$

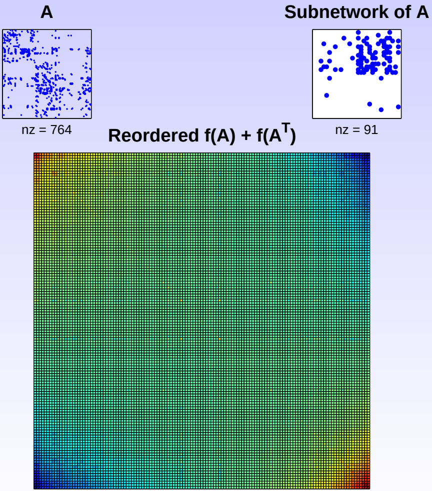
Also, $f(A) + f(A^T)$ is symmetric and amenable to standard clustering techniques

Example



C. Elegans neural data

Automates the computations of Durbin (PhD thesis Cambridge, 1987):



Summary: Networks as Matrices

Two general techniques:

- **spectral information**: optimization and random graph motivations
- **matrix functions**: combinbatoric motivation

Some References

Spectral Graph Theory, Fan R. K. Chung, 1997, CBMS Regional Conference Series in Mathematics Volume: 92

A First Course in Network Theory, Ernesto Estrada and Philip A. Knight, OUP, 2015

Unravelling small world networks, D. J. Higham, J. Computational and Applied Math., 158, 2003, 61-74.

A lock-and-key model for protein-protein interactions, J. L. Morrison, R. Breitling, D. J. Higham and D. R. Gilbert, Bioinformatics, 22, 2006, 2012-2019.

Spectral clustering and its use in bioinformatics, D. J. Higham, G. Kalna, and M. Kibble, J. Computational and Applied Math., 204, 2007, 25-37.

Mapping directed networks, J. J. Crofts, E. Estrada, D. J. Higham and A. Taylor, Electronic Transactions on Numerical Analysis, 37, 2010, 337-350.

Communicability and multipartite structures in complex networks at negative absolute temperatures, E. Estrada, D. J. Higham and N. Hatano, Physical Review E, 78, 2008, 026102.