Network Science Revision Lecture

Spring 2024

Exam details
Overview of the module
Review exercises

Network Science Exam

- The exam will be on Wednesday 15th May
- There are 4 multi-part questions, each of which counts for 20 marks
- The exam counts towards 70% of your overall mark
- You will be provided with an information sheet; this has been posted on Blackboard
- Review questions & solutions have been posted on Blackboard
- Past papers 2021, 2022, 2023

Examinable material: *You will not be asked to analyze or write any python code or pseudocode. You will not be tested on your understanding of NetworkX. "Network science and climate science", K-means clustering, and the method for spectral clustering/weighted cut size are not examinable (weighted graphs are examinable).*

Overview

- Graph properties and structure
- Random graph models
- Dynamics on graphs: modeling, analysis, and simulation
- Communities & community detection

Graph Properties & Structure

Adjacency matrix

Number of links

Degree distributions

Diameter

Cosine similarity

Weighted Adjacency matrix

$$\overline{k^2} = \frac{1}{N}\sum_{i=1}^{N} k_i^2$$

Global clustering

Jaccard similarity

Clustering coefficient

Node average: $\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i$

Total degree, $K = \sum_{i=1}^{N} k_i$

Node centrality

Degree centrality: the higher the degree, the more important the node

Eigenvector centrality: A node's centrality, x_i, should be proportional to the centrality of its neighbors: $x_i = \alpha \sum_{j=1}^{N} A_{ij} x_j$ or $\mathbf{Ax} = \lambda \mathbf{x}$ with $\lambda = \alpha^{-1}$. We want α^{-1} to be an eigenvalue.

Katz centrality: $x_i = \alpha \sum_{j=1}^{N} A_{ij} x_j + 1$. We now have to solve the linear system, $(\mathbf{I} - \alpha \mathbf{A})\mathbf{x} = \mathbf{z}$ where $\mathbf{z}$ is a N-element column vector of ones, $\mathbf{z} = [1,1,1 \ldots, 1]^T$. We need to ensure that $\det(\mathbf{I} - \alpha \mathbf{A}) \neq 0$, i.e. α^{-1} should not be an eigenvalue of $\mathbf{A}.$

Page-rank centrality: $x_i = \sum_{j=1}^{N} \left(\frac{(1-m) A_{ij} x_j}{\max\left(k_j^{out}, 1\right)} + \frac{m x_j}{N} \right), 0 < m \leq 1$

<u>Or</u> $\mathbf{Gx} = \mathbf{x}$ where $G_{ij} = \frac{A_{ij}(1-m)}{\max\left(k_j^{out}, 1\right)} + \frac{m}{N}$, we want eigenvector corresponding to $\lambda = 1$.

Perron-Frobenius theorem

- **Eigenvector centrality:**
  - For an *undirected connected* graph, there will be exactly one eigenvector where all elements have the same sign, and this eigenvector corresponds to a simple positive eigenvalue of $\mathbf{A}$ (this eigenvalue is $\geq$ in magnitude to all other eigenvalues).
  - Scale this leading eigenvector so that all elements are positive (the magnitude of the scaling is not considered to be important)
  - Then the *eigenvector centrality* of node i is the i^{th} element of the scaled vector.
- **Katz centrality:**
  - Non-negative square matrix $\rightarrow$ there will be a real non-negative eigenvalue, λ_1, with $\lambda_1 \geq \max(|\lambda_i|)\ with\ i \in \{1, 2, \ldots, N\}$. If $\lambda_1 > 0$, we set $\alpha^{-1} > \lambda_1$ and this will guarantee a non-trivial solution of our system.
- **Page-rank centrality:**
  - For a positive square matrix:
    - The matrix will have a positive, real, simple eigenvalue strictly larger in magnitude than all other eigenvalues *and*
    - All elements of the corresponding eigenvector will have the same sign
    - There are no other eigenvectors where all elements have the same sign

Graph Models

- **G_{Np} random graph model:**
  - An individual *realization* of a G_{Np} graph can be constructed via a sequence of $N(N-1)/2$ Bernoulli trials
  - We were interested in computing expectations over the set of graphs produced by the model for a given $N$and p
  - We were able to deduce properties about the structure of the graph as $N \to \infty$ w.h.p.
  - $P(G) = p^L(1-p)^{N'-L}$. $N' = \binom{N}{2}$ is the maximum possible number of links in the graph
  - $\langle k_i \rangle = \sum_{k=0}^{N-1} P(k_i = k)k$
  - $p_k = \binom{N-1}{k} p^k (1-p)^{(N-1-k)}$
- **The configuration model:**
  - Given a degree sequence $\mathrm{d} = (k_1, k_2, \dots, k_N)$, graph is generated by randomly connecting stubs
  - One stub is equally likely to connect to any other stub with probability $1/(K-1)$
  - $\langle l_{ij} \rangle = \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} \langle X_{ab} \rangle = \frac{k_i k_j}{K-1},$
  - Friendship paradox

- **Barabasi- Albert model:**
  - Graph changes at each time step
  - *Preferential attachment:* Probability of a new link attaching to an existing node is linearly proportional to that node's degree

$$\rho_i(t+1|G_a(t)) = \frac{k_i(t|G_a(t))}{\sum_{i=1}^{N(t)} k_i(t)}$$

  - Aim to find $p_k(t+1)$ given $p_k(t)$. We can then obtain an interpretable result when $t \to \infty$ and look for stationary distributions

Laplacian Matrix

- $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the graph ***Laplacian matrix***, where $D_{ii} = k_i$
- ***Weighted Laplacian matrix:*** L$= (\widehat{\mathbf{D}} - \mathbf{W})$, where $\widehat{D}$ is a diagonal matrix with $\widehat{D}_{ii} = \sum_{j=1}^{N} W_{ij}$
- ***Normalised weighted Laplacian Matrix*** : $\hat{\mathbf{L}} = \widehat{\mathbf{D}}^{-\frac{1}{2}}(\widehat{\mathbf{D}} - \mathbf{W})\widehat{\mathbf{D}}^{-\frac{1}{2}}$ where $\widehat{D}$ is a diagonal matrix with $\widehat{D}_{ii} = \sum_{j=1}^{N} W_{ij}$

- For undirected networks, both A, W and L are symmetric, so let's quickly review a few useful properties of symmetric matrices:

  - All eigenvalues are real, and eigenvectors can be chosen to be real
  - The eigenvectors of the matrix form a basis for $\mathbb{R}^N$ (even if some eigenvalues are repeated)
  - A square matrix is orthogonally diagonalizable if and only if it is symmetric

- Orthogonal diagonalizability means that a symmetric matrix B can be diagonalized as, $B = V\Lambda V^T$ where:
  - V is an orthogonal N x N matrix whose i-th column is v_i , and $v_i^T v_j = 1$ if $i = j$ and $v_i^T v_j = 0$ if $i \neq j$ (the eigenvectors are orthonormal, and $V^T V = \mathrm{I}$)
  - Λ is a diagonal matrix where $\Lambda_{ii} = \lambda_i$

Dynamics on Graphs

- Diffusion on graphs: $\frac{dn_i}{dt} = -\alpha \sum_{j=1}^{N} L_{ij} n_j$
    - Modelling particles moving on graph, using Ficks law
    - This is a system of linear constant-coefficient ODEs - *i.e.* it's an eigenvalue problem

- Synchronisation on graphs: simple linear *coupling* between linked nodes: $\frac{dn_i}{dt} = \alpha \sum_{j=1}^{N} A_{ij}(n_j - n_i)$ or more compactly: $\frac{d\mathbf{n}}{dt} = -\alpha \mathbf{L}\mathbf{n}$

- Random walks on graphs: directly modeling particles moving from one node to another *randomly*

- Epidemics on Networks:
    - $\frac{d\langle x_i \rangle}{dt} = \beta \sum_{j=1}^{N} A_{ij} \left\langle (1 - x_i) x_j \right\rangle$ (network SI model)
    - $\frac{d\langle x_i \rangle}{dt} = \beta \langle 1 - x_i \rangle \sum_{j=1}^{N} A_{ij} \left\langle x_j \right\rangle$ (naïve network SI model)
    - Degree based approximation $\frac{d\phi_k}{dt} = k\beta(1 - \phi_k) \sum_{k'=1}^{k'_{max}} \theta(k, k') \phi_{k'-1}$
    - Pair approximations

Community Detection

Modularity:

- The *modularity* of a set of nodes, S_a , is defined as: $M_a = \frac{1}{2L} \sum_{i \in S_a} \sum_{j \in S_a} \left(A_{ij} - \frac{k_i k_j}{2L} \right)$
- For q disjoint sets: $M = M_1 + M_2 + M_3 + M_4 + \cdots + M_q$
- Spectral modularity maximization method to assign sets

Laplacian graph partitioning: break a connected graph into two groups of nodes where the number of links crossing from one group to the other (the *cut size, c*) is minimized

$$c = \frac{1}{4} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} (1 - s_i s_j)$$

Review Questions

Continued on visualiser