# Networks - Week 6 - Discrete Random Walks on Networks

Antonio León Villares

November 2023

## Contents

# 1 Formalising Discrete Random Walks

## 1.1 Discrete Random Walks on Networks

- **Given a transition matrix, how can we compute the probability of a random walker arriving at a given vertex?**
  - let $T$ be a **transition matrix**, such that:

    $$T_{ij} = \text{probability that the walker goes from } i \text{ to } j$$

  - the **probability** of the walker reaching the $i$th node in $t$ steps is $p_i(t)$, which is the compoennt of the matrix:

    $$\underline{p}(t) = \underline{p}(0)T^t$$

  - this section focuses on understanding $\underline{p}$ through the **spectral properties** of $T$
- **What is the trnasition matrix for a random walk?**
  - given a **random walk**, where the walker will uniformly randomly pick an adjacent edge:

    $$T_{ij} = \frac{A_{ij}}{d_i}$$

    where $d_i$ denotes the **degree** of the $i$th vertex
- **When is the transition matrix symmetric?**
  - whenever the underlying network is **symmetric** (need an undirected graph) and whenever each vertex has the same degree
  - for $T_{ij} = T_{ji}$ we need that:

    $$A_{ij} = A_{ji} \qquad d_i = d_j$$
- **If the transition matrix is not symmetric, how can we derive its spectral properties?**
  - **symmetric matrices** have very nice, well-studied **spectral properties** (**Perron-Frobenius Theorem**, guarantee on real eigenvalues, etc...)
  - even if $T$ isn't symmetric, we can still study its properties through a **related** matrix which is symmetric:

    $$\tilde{A}_{ij} = \frac{A_{ij}}{\sqrt{d_i d_j}}$$

    (whenever $A$ is a symmetric matrix)
  - notice, the **spectral properties** of $\tilde{A}$ will be essentially equivalent to those of the **normalised Laplacian**

*Let $A$ be an* ***adjacency matrix****. The* ***Laplacian*** *of $A$ is constructed by defining a matrix*

$$D = \operatorname{diag}(d_1, \dots, d_n)$$

*where $d_i$ is the $i$th row sum of $A$. The Laplacian of $A$ is the* ***symmetric*** *matrix:*

$$L = D - A$$

*The* ***normalised Laplacian*** *is given by:*

$$\tilde{L} = I - D^{-1/2} A D^{-1/2}$$

*where since $D$ is a diagonal matrix:*

$$D^{\omega} = \operatorname{diag}(d_1^{\omega}, \dots, d_n^{\omega})$$

where we identify:

$$\tilde{A} = D^{-1/2} A D^{-1/2}$$

- **How can we express $\tilde{A}$ in terms of its spectrum?**
  - since $\tilde{A}$ is **real** and **symmetric**, we can **diagonalise** it:

    $$\tilde{A} = U \Lambda U^T$$

  - if $\{\lambda_i\}_{i \in [1,n]}$ gives a set of **eigenvalues**, and $\{\underline{u}_i\}_{i \in [1,n]}$ gives a set of **eigenvectors**, then the **orthonormal decomposition** of $\tilde{A}$ is:

    $$\tilde{A} = \sum_{i=1}^{n} \lambda \underline{u}_i \underline{u}_i^T$$

- **How can we use the spectrum of $\tilde{A}$ to express $T$?**
  - notice:

    $$T_{ij} = \frac{\sqrt{d_j} \tilde{A}_{ij}}{\sqrt{d_i}}$$

    so in particular, if $D = \operatorname{diag}(d_1, \dots, d_n)$ denotes the diagonal matrix of degrees of $A$:

    $$T = D^{-\frac{1}{2}} \tilde{A} D^{\frac{1}{2}}$$

  - in particular, the **eigenvalues** of $T$ are the same as those for $\tilde{A}$ and:
    - the **right eigenvectors** of $T$ are:

      $$\underline{w}_i^R = D^{-\frac{1}{2}} \underline{u}_i$$

      since:

      $$D^{-\frac{1}{2}} \tilde{A} D^{\frac{1}{2}} (D^{-\frac{1}{2}} \underline{u}_i) = D^{-\frac{1}{2}} (\tilde{A} \underline{u}_i) = \lambda_i D^{-\frac{1}{2}} \underline{u_i}$$

    - similarly, the **left eigenvectors** of $T$ are:

      $$\underline{w}_i^L = D^{\frac{1}{2}} \underline{u}_i$$

  - this gives us the **eigendecomposition**

$$T = \sum_{i=1}^{n} \lambda_i \underline{w}_i^R \left(\underline{w}_i^L\right)^T$$

- **Using the eigendecomposition of the transition, how may we express the walker probability matrix $\underline{p}(t)$ (known as teh stationary density)?**
  - using the **eigendecomposition**, we have that:

$$T^t = \sum_{i=1}^{n} \lambda_i^t \underline{w}_i^R \left(\underline{w}_i^L\right)^T$$

  - hence, if $\underline{p}(t)$ is a **row vector**:

$$\underline{p}(t) = \underline{p}(0)T^t = \sum_{i=1}^{n} \lambda_i^t \left(\underline{w}_i^L\right)^T \left(\underline{p}(0)\underline{w}_i^R\right)$$

- **In practice, how is the stationary density computed?**
  - via a **power method**
  - that is, start with a **random vector** $\underline{v}_0$
  - then define:

$$\underline{v}_{i+1} = T\underline{v}_i$$

## 1.2 Proposition: Spectral Information of the Transition Matrix

*Let $T$ be the **transition matrix** given by:*

$$T_{ij} = \frac{A_{ij}}{d_i}$$

*Then:*

1. *All **eigenvalues** $\lambda_i$ of $T$ lie in the interval $[-1, 1]$*
2. *The **left eigenvector** $(\underline{w}_j^L)^T$ corresponding to the **eigenvalue***

   $$\lambda_j = 1$$

   *gives the **stationary density** of the **transition matrix**; that is we have:*

   $$\underline{(w_j^L)}^T = \underline{p}^*$$

   *whereby:*

   $$\lim_{t \to \infty} \underline{p}(t) = \underline{p}^* \qquad \text{and} \qquad p^* = p^* T$$

3. *The **right eigenvector** $\underline{w}_j^R$ corresponding to $\lambda_j = 1$ is such that:*

   $$\underline{w}_j^R \propto \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

4. *The **eigenvalue** $\lambda_j$ is the **largest magnitude eigenvalue**, and all elements of $\underline{w}_j^R, \underline{w}_j^L$ are **positive**.*

---

*Proof.* We make extensive use of the Perron-Frobenius Theorem, since we assume that $T$ is irreducible (i.e that there is always a path between any 2 vertices $i, j$, or that $\exists n : T^n = 0$).

*Let $A$ be a $n \times n$ matrix, such that $A$:*

- *is **irreducible***
- ***non-negative***
- *has **spectral radius** $\rho(A) = r > 0$*

*Then:*

1. *r is an **eigenvalue** of A (called the **Perron-Frobenius** eigenvalue)*
2. *r is **simple**. In particular:*
    - *r has **algebraic multiplicity** 1 (it is not a repeated eigenvalue)*
    - *r has **geometric multiplicity** 1 (both right and left eigenspaces are one-dimensional - this is because geometric multiplicity is bounded by algebraic multiplicity*
3. *A has left/right **eigenvectors** with **eigenvalue** r, and whose components are **all positive***
4. *the only **eigenvectors** whose components are **all positive** are those associated to r*
5. *r is **bounded** above/below by the maximum and minimum **row** sums of A (and also the **column** sums):*

$$\min_{i\in[1,n]} \sum_{j=1}^{n} A_{ij} \leq r \leq \underset{i\in[1,n]}{max} \sum_{j=1}^{n} A_{ij}$$

①

By Perron-Frobenius (part 5), the Perron-Frobenius Eigenvalue of $T$ is bounded by row/column sums of $T$. In particular:

$$r \leq \sum_{j=1}^{n} \frac{A_{ij}}{d_i} = \frac{1}{d_i} \sum_{j=1}^{n} \frac{A_{ij}}{=} \frac{d_i}{d_i} = 1$$

where we've used the definition of degree as:

$$\sum_{j=1}^{n} \frac{A_{ij}}{=} d_i$$

In particular, since the eigenvalues of $T$ are all real (as they are eigenvalues of a real symmetric matrix $\tilde{A}$), and the absolute value of all eigenvalues is bounded by the spectral radius (which is given by the Perron-Frobenius Eigenvalue), it follows that any eigenvalue of $T$ satisfies:

$$\lambda_i \in [-1, 1]$$

②

By definition of the stationary density:

*The* ***non-negative stationary density*** *is a vector:*

$$\underline{p}^* = (p_1^*, \dots, p_n^*)$$

*where:*

$$p_i^* = \lim_{t\to\infty} p_i(t)$$

*and:*

$$\underline{p}^* = \underline{p}^* T$$

it corresponds to the left eigenvector of the transition matrix with eigenvalue 1. This is precisely $(\underline{w}_j^L)^T$.

③

If $\underline{w}_j^R$ is the right eigenvector of $T$ with eigenvalue 1, we have that:

$$T\underline{w}_j^R = \underline{w}_j^R$$

Notice also that if

$$\underline{v} = c\begin{pmatrix} 1 \\ \vdots 1 \end{pmatrix}$$

Then:

$$T(c\underline{v}) = c\underline{v}$$

since the row sums of $T$ are 1 (as each row element is a probability). Hence, $c\underline{v}$ is a right eigenvector of $T$ with eigenvalue 1, so in particular:

$$\forall c \in \mathbb{R}, \qquad \underline{w}_j^R = c\underline{v}$$

④

This is properties 1 and 3 of the Perron-Frobenius Theorem.

□

## 1.3 The Cheeger Constant and the Spectral Gap

### 1.3.1 The Spectral Grap

- **How can we approximate the stationary desnity for large $t$?**
  - by the **Perron-Frobenius Theorem**, the **eigenvalues** of $T$ are bounded on:

    $$\lambda_i \in [-1, 1]$$

  - in particular, since the **stationary density** is given by:

    $$\underline{p}(t) = \sum_{i=1}^{n} \lambda_i^t \left(\underline{w}_i^L\right)^T (\underline{p}(0)\underline{w}_i^R)$$

    as $t \to \infty$, most eigenvalues will tend to 0
  - thus, if $\lambda_{max} = 1$, and $\lambda_2$ is the second largest eigenvalue (in magnitude), it follows that:

    $$\underline{p}(t) \approx \left(\underline{w}_{max}^L\right)^T (\underline{p}(0)\underline{w}_{max}^R) + \lambda_2^t \left(\underline{w}_2^L\right)^T (\underline{p}(0)\underline{w}_2^R)$$

- **What is the spectral gap?**
  - the value:
$$\gamma = \lambda_{max} - |\lambda_2| = 1 - |\lambda_2|$$
- **What does a large spectral gap imply?**
  - if $\gamma$ is large, the term:
$$\lambda_2^t \left(\underline{w}_2^L\right)^T \left(\underline{p}(0)\underline{w}_2^R\right)$$
    goes to 0 much quicker (since $|\lambda_2|$ is small)
  - in particular, $\lambda_2$ controls the "relaxation time" (i.e the time required for the stationary density to stabilise into a constant)

### 1.3.2 Definition: The Cheeger Inequality

*The Cheeger Inequality gives useful bounds on $|\lambda_2|$, and thus informs us on the relaxation time.*

*Let $S$ be a **vertex subset** of a **network**, and let $\overline{S}$ be its **complement**. The **Cheeger Constant** (also known as **conductance**) is defined by:*

$$h = \min_S \left\{ \frac{\text{of edges connecting } S \text{ and } \overline{S}}{\min\{|S|, |\overline{S}|\}} \right\}$$

*The **Cheeger Inequality** is:*

$$\frac{h^2}{2} < 1 - |\lambda_2| \leq 2h$$

*Intuitively, this makes sense. If $h$ is small, this implies that there are a small number of edges connecting $S, \overline{S}$. This can be interpreted as the graph being split into 2 communities (namely $S, \overline{S}$) with a few edges connecting them. In such a scenario, the probability that a random walker moves between communitieis is low, which menas that so too the spectral gap will be small, so in particular the relaxation time will be longer.*

# 2 Applications of Random Walks

## 2.1 The PageRank Algorithm

### 2.1.1 Motivating PageRank

- **What is the purpose of the PageRank algorithm?**
  - define a **centrality measure** on **directed networks**
  - **PageRank** is defined such that a node $v_i$ is "important" if:
    - it has many **incoming edges**
    - the **incoming edges** originate from **important nodes**

    - ∗ if $v_j$ has an **outgoing edge** to $v_i$, the **out-degree** of $v_j$ is small
  - in short, $v_i$ has to be connected to a lot of important nodes, which themselves don't share many connections with other nodes
  - these requirements were used to **rank** webpages, based on the **hyperlinks** which connected the websites

- **What is the PageRank of a given node?**
  - the corresponding **stationary density** of said node, given a **discrete-time random walk** through the **directed network**

- **Why can't we immediately derive a stationary density $\underline{p}^*$ in the case of PageRank (i.e in some random directed network)?**
  - the **stationary density** exists and is **unique** if and only if the network is **strongly connected**
  - otherwise:
    - ∗ many possible $\underline{p}^*$ can exist (i.e if the network contains multiple absorbing states, whereby $T_{ii} = 1$)
    - ∗ certain $p_i^*$ will be 0 (i.e they will be considered as irrelevant)
  - generally, **real world networks** (and especially the Internet) don't satisfy **strong connectedness**

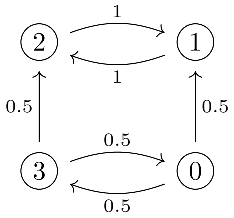

Figure 1: Example of the **spider trap** problem, whereby 2 nodes will get $p^* = 0$. This is because once states $1, 2$ are entered, they will indefinitely cycle amongst themselves. In particular, this is an example of a **periodic Markov Chain**, and it yields $\underline{p}^* = \langle 0, 0.5, 0.5, 0 \rangle$; here, two states take away all the "ìmportance" of the other two states.

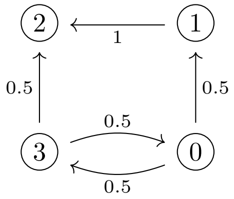

Figure 2: Example of the **dead ends** problem, whereby node 2 is an absorbing state. In such a case, the algorithm will fail to converge and no unique stationary density vector will be found. This is because once state 2 is reached, it can never be escaped; thus, different $\underline{p}^*$ can be found through **power iteration** methods, depending on the initial conditions of the process.

- **How can lack of strong connectivity be tackled?**
  - **PageRank** can "ignore" the lack of **strong connectivity** by assuming **node teleportation**: from a given website, you can randomly "jump" to any other website (i.e you don't use a hyperlink to find a new website, you just search it up)
  - this reasonable assumption then allows us to find a**unique stationary density**

### 2.1.2 Definition: Random Walk with Teleportation

*Suppose we have a* ***directed network*** *traversed by a* ***random walker*** *who's dynamics are defined by the* ***transition matrix****:*

$$T_{ij} = \frac{A_{ij}}{d_i^{out}}$$

*Then,* ***random walks with teleportation*** *are defined by:*

$$\underline{p}(t+1) = \alpha \underline{p}(t) T + (1-\alpha)\underline{b}$$

*or componentwise:*

$$p_i(t+1) = \alpha \sum_{j=1}^{n} p_j(t) T_{ji} + (1-\alpha) b_i$$

*where:*

- $\alpha \in (0,1)$ *is the* ***damping factor****, which is the probability that at any given time-step the walker follows links (and* $1-\alpha$ *is thus the probability that it teleports)*
- $\underline{b}$ *is a* ***preference row vector*** *satisfying:*

  $$\sum_{i=1}^{n} b_i = 1$$

  *and such that* $b_i \in [0,1]$ *is the probability of randomly teleporting to the* $i$*th vertex*

---

- **What value is typically used for the damping factor?**
  - a well-agreed value is $\alpha = 0.85$

#### 2.1.3 Proposition: PageRank Stationary State

*The* ***PageRank Stationary State*** *can be calculated via:*

$$\underline{p}^* = (1-\alpha)\underline{b}(I-\alpha T)^{-1}$$

*or componentwise:*

$$p_i^* = b_i + \sum_{l=1}^{\infty}\sum_{j=1}^{n} b_j(T_{ji} - T_{ji}^{l-1})$$

---

*Notice, here we assume that $\alpha$ is well chosen, such that the geometric series form of $(I-\alpha T)^{-1}$ is well-defined, like in Katz Centrality*

## 2.2 Epidemic Modelling

### 2.2.1 The SIS Model

#### 2.2.1.1 Definition: SIS Model

*Consider a* ***clique*** *of n vertices. Let:*

- *$S(t)$ be the* ***proportion*** *of nodes which are* ***susceptible*** *(can be infected) at time t*
- *$I(t)$ be the* ***proportion*** *of nodes which are* ***infected*** *(can infect) at time t*

*The* ***dynamics*** *of the* ***SIS Model*** *(susceptible-infected-susceptible) are defined by a system of ODEs:*

$$\frac{dS}{dt} = -\beta I(t)S(t) + \mu I(t)$$
$$\frac{dI}{dt} = \beta I(t)S(t) - \mu I(t)$$

*where:*

- *$\beta$ is the* ***infection rate*** *(the rate at which a* ***suceptible node*** *contracts an infection, and thus becomes an* ***infected node****)*
- *$\mu$ is the* ***recovery rate*** *(the rate at which an* ***infected node*** *recovers from the infection, and thus becomes a* ***susceptible node***

#### 2.2.1.2 Explaining the SIS Model

- **What is the probability of infection on some small time frame $\Delta t$?**
  - we can approximate this probability via:
    $$\beta \Delta t$$
- **What is the intuition behind the ODEs defining the SIS model?**
  - for the **susceptible nodes**:
    - if $S$ increases, then we expect the rate of change of $S$ to decrease (since more susceptible ndoes are available to infect)
    - increasing $I$ will both **decrease** the number of susceptibles (since more nodes can now infect at a rate $\beta$) and **increase** the number of susceptibles (since more nodes can now recover at a rate $\mu$)
    - thus, if $\beta >> \mu$, the number of susceptibles will generally decrease, but if $\mu >> \beta$, the number of susceptibles will generally increase
  - similar arguments work for the **infected nodes**, but the other way round
  - in particular, we have that:
    $$\frac{d}{dt}(S+I) = 0 \implies \forall t \geq 0, S+I = 1$$
    (since $S(0) + I(0) = 1$)
- **What is equilibrium like for the SIS model?**
  - let $S^*, I^*$ be equilibrium values (that is, points at which $S'(t) = I'(t) = 0$.
  - then, assuming that $I^* \neq 0$:
    $$\begin{aligned} 0 &= -\beta I^* S^* + \mu I^* \qquad 0 = \beta I^* S^* - \mu I^* \\ \implies \beta S^* &= \mu \\ \implies S^* &= \frac{\mu}{\beta} \end{aligned}$$
    and since $S + I = 0$ always, we must then have that:
    $$S^* = \frac{\mu}{\beta} \qquad I^* = 1 - \frac{\mu}{\beta}$$
  - otherwise, we will have that $I^* = 0$ (i.e at equilibrium there are no infections
- **What is the epidemic threshold of the SIS model?**
  - the **epdiemic threshold** $\varepsilon$ is a value such that if $\varepsilon > 1$, an **epidemic** is possible; if $\varepsilon < 1$, then sustaining the infections becomes impossible
  - in the **SIS model**, we have that:
    $$\varepsilon = \frac{\beta}{\mu}$$
    since if:
    $$\frac{\beta}{\mu} > 1$$
    then $S^*, I^*$ remain in the range $(0, 1)$ and we have epidemic activity
  - otherwise, we see that $S^* > 1, I^* < 0$ and the epidemic is unsustainable

- **Does the epidemic threshold make sense?**
  - suppose that
$$\frac{\beta}{\mu} > 1$$
    such that the **epidemic** holds
  - if $\mu$ is **large** (i.e high recovery rate) or $\beta$ is **small** (i.e low infection rate), we see that:
$$S^* \to 1 \qquad I^* \to 0$$
  - conversely, for **small** $\mu$ or **large** $\beta$:
$$S^* \to 0 \qquad I^* \to 1$$

### 2.2.2 The SIR Model

#### 2.2.2.1 Definition: SIR Model

*Consider a* ***clique*** *of n vertices. Let:*

- *$S(t)$ be the* ***proportion*** *of nodes which are* ***susceptible*** *(can be infected) at time t*
- *$I(t)$ be the* ***proportion*** *of nodes which are* ***infected*** *(can infect) at time t*
- *$R(t)$ be the* ***proportion*** *of nodes which are* ***recovered*** *(were infected, recovered and can no longer infect or be infected) at time t*

*The* ***dynamics*** *of the* ***SIR Model*** *(susceptible-infected-recovered) are defined by a system of ODEs:*

$$\begin{aligned}\frac{dS}{dt} &= -\beta I(t)S(t) + \mu I(t)\\ \frac{dI}{dt} &= \beta I(t)S(t) - \mu I(t)\\ \frac{dR}{dt} &= \mu I(t)\end{aligned}$$

*where:*

- *$\beta$ is the* ***infection rate*** *(the rate at which a* ***suceptible node*** *contracts an infection, and thus becomes an* ***infected node****)*
- *$\mu$ is the* ***recovery rate*** *(the rate at which an* ***infected node*** *recovers from the infection, and thus becomes a* ***susceptible node***

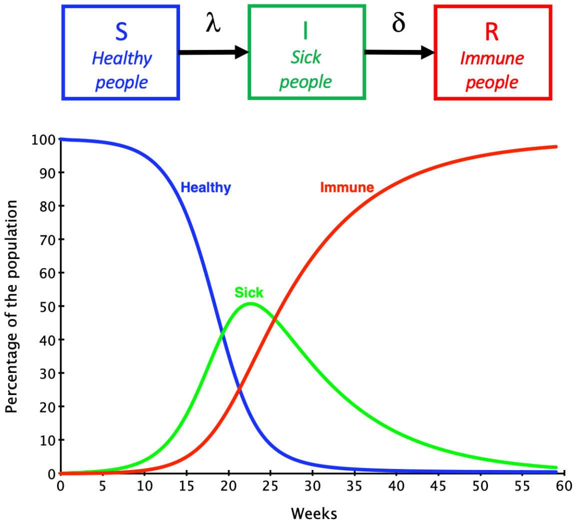

Figure 3: Dynamics of the **SIR Model**. Notice, the **infections** always reach a peak of contagion. The idea behind **epidemic management** is to spread the peak as much as possible, to avoid strain on medical services.

#### 2.2.2.2 Explaining the SIR Model

- **What is the key dynamic difference between the SIS and SIR models?**
  - in the **SIR model**, the infection eventually **dies out**, even if $\beta$ is large
  - the final state solely consists of **susceptible** and **recovered** nodes
- **When deos the peak observed above occur?**
  - if
    $$\frac{dI}{dt} > 0$$
    then there will be a **perceptible** infectious outbreak
  - otherwise, the rate of infection won't lead to outbreaks on a visible scale