4 State Space models

Let us begin with one example.

Example XXIII:
We have seen that the process $Z_t = X_t + X_{t-1}$ of Example 5 is nor Markov. The problem is that, in order to make a good prediction of Z_{t+1} we need X_t but, in the hypotheses of the Markov process, we have to make do with Z_t and that, as we have seen, doesn't give us all the information that a complete history would give us. We can try to "Markovize" the process by retaining, at time t, more information than just Z_t. Since we need X_t in order to predict Z_{t+1}, let us bring it along. Define

$$Q_t = \begin{bmatrix} q_{t,0} \\ q_{t,1} \end{bmatrix} = \begin{bmatrix} Z_t \\ X_t \end{bmatrix} \tag{258}$$

The process Q_t is now Markov with

$$\begin{aligned} P(q_{t+1,0} = m \big| Q_t) &= P_X(m - X_t) = P_X(m - q_{t,1}) \\ P(q_{t+1,1} = m \big| Q_t) &= P_X(m) \end{aligned} \tag{259}$$

We now have, at time t, all the information necessary to do the best possible prediction of Q_{t+1}. The process Z_t can easily be recovered one we have Q_t as

$$Z_t = [1, 0]Q_t \tag{260}$$

We can be more explicit about the way Q_t evolves. We have

$$Q_t = \begin{bmatrix} Z_t \\ X_t \end{bmatrix} = \begin{bmatrix} X_t + X_{t-1} \\ X_t \end{bmatrix} = \begin{bmatrix} X_t + q_{t-1,1} \\ X_t \end{bmatrix} \tag{261}$$

We can recast this all as a recursive equation

$$\begin{aligned} Q_t &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} Q_{t-1} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} X_t \\ Z_t &= [1, 0]Q_t \end{aligned} \tag{262}$$

What we have done here is to separate the evolution of the process in two parts. The vector Q_t is the *state* of the process, and its value at t depends only on its value at $t-1$ plus the valus of a white process. That is, the evolution of the state is a Markov-type process. The process Z_t is the *observation*, the value that we actually observe. (end of example)

In this section, we shall study the generalization of this idea, and analyze some of its consequences. Many natural processes and technical devices can be modeled through a model that, although it is

not Markov, has a finite memory. In a Markov process, $P(m_{t+1}|m_t) = P(m_{t+1}|m_t, \ldots, m_0)$, that is, one value of the process is enough to give us all the information we need. In many cases, things are not quite that simple, but there is a finite n such that

$$P(y_t|y_{t-1}, \ldots, y_{t-n}) = P(y_t|y_{t-1}, \ldots, y_0) \tag{263}$$

(You might wonder why did I suddenly switch from m_t to y_t. Fair enough. This is mainly a matter of customs and habits: the literature on the subject tends to use variables names such as x, y, z, w. Eventually, we shall have to comply with this habit, so we might as well do it now. Besides, m conveys the idea of an integer variable, while the models that we study in this section take typically real values.)

autoregressive model

The standard form of these models is the **autoregressive** model (AR) with finite memory, that is, by a process of the type

$$y(t) = a_1 y(t-1) + \cdots + a_n y(t-n) + w(t) \tag{264}$$

order of a model

where w is a white process and n is the **order** of the model.

state

The model depends on n past samples so it is not, *stricti dictu*, a Markov model. However, since the output y_t depends on a finite and fixed number of past sample, we can modify it by defining a **state** that encodes all the past information that we need for predicting the value at t. If we revrite (264) as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-n+1} \end{bmatrix} = \begin{bmatrix} a_1, & \ldots & \ldots, & a_n \\ 1, & \ldots & \ldots, & 0 \\ 0, & 1, & \ldots, & 0 \\ \vdots & & & \vdots \\ 0 & \ldots, & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-n} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w_t \tag{265}$$

then we can define the vector $x_t = [y_t, \ldots, y_{t-n+1}]$ obtaining a one-step equation

$$\begin{aligned} x_t &= \mathbf{A}x_{t-1} + \mathbf{D}w_t \\ z_t &= [1, 0, \ldots, 0]x_t \stackrel{\triangle}{=} \mathbf{C}x_t \end{aligned} \tag{266}$$

observation

where $\mathbf{A}$ and $\mathbf{B}$ are defined as in (265). In this model we have a one-step evolution of the state, but we mantain the output of (264) by creating an **observation** process z_t. In many cases we assume that the state x_t is hidden, and that the only information we have about the system is the process z_t. In this case, of course, it is quite easy to reconstruct the state from the values z_t: one just has to store the last n outputs, and these will give us the value of the state x_t.

We can generalize the model by transforming the matrices and the vectors, which is our example have a very specific form, into general matrices and vectors. In order to make the model more

general, we assume that we have an input u_t, which we are able to control and that observation z_t is also subject to noise. We thus arrive at the standard **state-space model**:

state-space model

$$\begin{aligned} x_{t+1} &= \mathbf{A}x_t + \mathbf{B}u_t + w_t \\ z_t &= \mathbf{C}x_t + v_t \end{aligned} \tag{267}$$

with $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $z \in \mathbb{R}^q$, $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, $\mathbf{C} \in \mathbb{R}^{q\times n}$; $w_t \in \mathbb{R}^n$ $v_t \in \mathbb{R}^q$ are white processes.

This is not the most general model possible; we could assume that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ depend on time. This model, however, is enough for our purposes here. As a matter of further simplification we shall often assume that $m = q = 1$ (the input and the observations are scalar).

In this model, the state contains all the relevant information about the system, the input u_t allows us to intervene on the state, and the output z_t allows us to observe the effects of changes of state. The noises w_t and v_t are there to remind us that life is tough, that we don't always have all the information we would like to have, and that there are things outside of our control that cause interferences.

Let us for the moment ignore the noises, and assume that we have perfect knowledge of the system. We still don't have direct access to the state. All we can do is to observe z_t trying to deduce what the state might be and use u_t trying to make it do what we want. Because of these limitations, we might not be able to accomplish our goals, that is, we might not be able to know everything about the state simply by observing z_t or to make it do what we want simply by manipulating u_t. These observations lead to two important concepts in system theory: **reachability** and **observability**.

Intuitively, a state x_t is **reachable** (at time t) if we can start from an arbitrary state (say, $x_0 = 0$—the actual value is not terribly important) and devise an input $[u_0, \ldots, u_{t-1}]$ that leads us to that state. More formally, considering a state x_t, we can write

reachability

$$\begin{aligned} x_t &= \mathbf{A}x_{t-1} + \mathbf{B}u_{t-1} \\ &= \mathbf{A}(\mathbf{A}x_{t-2} + \mathbf{B}u_{t-2}) + \mathbf{B}u_{t-1} \\ &= \mathbf{A}^2x_{t-2} + \mathbf{AB}u_{t-2} + \mathbf{B}u_{t-1} \\ &= \mathbf{A}^2(\mathbf{A}x_{t-3} + \mathbf{B}u_{t-3}) + \mathbf{AB}u_{t-2} + \mathbf{B}u_{t-1} \\ &= \mathbf{A}^3x_{t-3} + \mathbf{A}^2\mathbf{B}u_{t-3} + \mathbf{AB}u_{t-2} + \mathbf{B}u_{t-1} \\ &\vdots \\ &= \mathbf{A}^tx_0 + \mathbf{A}^{t-1}\mathbf{B}u_0 + \cdots + \mathbf{AB}u_{t-2} + \mathbf{B}u_{t-1} \end{aligned} \tag{268}$$

Based on this relation, we can give the following definition.

Definition 4.1. *A state x_t is reachable (at time t) if, given $x_0 = 0$, there is a sequence of inputs $u_0, \dots, u_{t-1}$ such that*

$$x_t = [\mathbf{B}|\mathbf{AB}|\cdots|\mathbf{A}^{t-1}\mathbf{B}] \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_0 \end{bmatrix} \tag{269}$$

The matrix

$$\mathcal{R}^+(t) = [\mathbf{B}|\mathbf{AB}|\cdots|\mathbf{A}^{t-1}\mathbf{B}] \tag{270}$$

is called the **reachability matrix** at time t. Let $\chi^+(t)$ be the set of states reachable at time t, then

reachability matrix

$$\chi^+(t) = \mathrm{Im}[\mathcal{R}^+(t)] \tag{271}$$

$\mathcal{R}^t$ is an $n \times t$ matrix, therefore it has rank at most n, which entails

$$\{0\} = \chi^+(0) \subseteq \chi^+(1) \subseteq \cdots \subseteq \chi^+(n) = \chi^+(n+1) = \cdots \tag{272}$$

Set $\mathcal{R}^+ \triangleq \mathcal{R}^+(n)$. If $\mathrm{rank}(\mathcal{R}^+) = n$, then $\chi^+(n) = \mathbb{R}^n$, that is, any state can be reached in at most n step. In this case, we say that the system itself is **reachable**.

reachable system

Example XXIV:
To see an example of unreachable system, divide $x(t) \in \mathbb{R}^n$ into $x'(t) \in \mathbb{R}^p$ and $x''(t) \in \mathbb{R}^m$, with $n = p + m$, and assume that the system has the following structure.

$$\begin{bmatrix} x'_t \\ - \\ x''_t \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{P} & \mathbf{Q} \\ - & - \\ 0 & \mathbf{R} \end{array}\right] \begin{bmatrix} x'_{t-1} \\ - \\ x''_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{T} \\ - \\ 0 \end{bmatrix} u_{t-1} \tag{273}$$

which can be written as

$$\begin{aligned} x'_t &= \mathbf{P}x'_{t-1} + \mathbf{Q}x''_{t-1} + \mathbf{T}u_{t-1} \\ x''_t &= \mathbf{R}x''_{t-1} \end{aligned} \tag{274}$$

The input u acts only on x', and x' has no effect on x'', so any evolution that starts with $x_0 = 0$ will remain with $x''_t = 0$ for all t, independently of the input u. Any state in which $x'' \neq 0$ is unreachable.

To see how this relates to the reachability matrix, consider that

$$\mathbf{A}^2 = \left[\begin{array}{c|c} \mathbf{P} & \mathbf{Q} \\ - & - \\ 0 & \mathbf{R} \end{array}\right] \left[\begin{array}{c|c} \mathbf{P} & \mathbf{Q} \\ - & - \\ 0 & \mathbf{R} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{P}^2 & \mathbf{PQ} + \mathbf{QR} \\ - & - \\ 0 & \mathbf{R}^2 \end{array}\right] \tag{275}$$

Iterating, we can see that all $\mathbf{A}^t$ have the same structure:

$$\mathbf{A}^t = \left[\begin{array}{ccc} \mathbf{P}^t & | & \mathbf{K}_t \\ - & - & - \\ 0 & | & \mathbf{R}^t \end{array}\right] \tag{276}$$

for some $\mathbf{K}_t$. Therefore

$$\mathbf{A}^t\mathbf{B} = \left[\begin{array}{c} \mathbf{P}^t\mathbf{T} \\ - \\ 0 \end{array}\right] \tag{277}$$

and

$$\mathcal{R}^+ = \left[\begin{array}{cccc} \mathbf{T} & \mathbf{PT} & & \mathbf{P}^{n-1}\mathbf{T} \\ - & - & \cdots & - \\ 0 & 0 & & 0 \end{array}\right] \tag{278}$$

which has rank at most p (end of example)

Similar considerations hold for **observability**. In this case, given a state x_0, we let the system evolve freely (setting $u = 0$) for a time t, and by observing the outputs z_t, we want to reconstruct the state x_0. If this is possible, x_0 is **observable**. In this case we can write

observability

$$\begin{aligned} z_0 &= \mathbf{C}x_0 \\ z_1 &= \mathbf{C}x_1 = \mathbf{CA}x_0 \\ &\vdots \\ z_{t-1} &= \mathbf{C}x_{t-1} = \cdots = \mathbf{CA}^{t-1}x_0 \end{aligned} \tag{279}$$

We define, analogously to the previous case, the **observability matrix**

observability matrix

$$\mathcal{O}^+(t) = \left[\begin{array}{c} \mathbf{C} \\ - \\ \mathbf{CA} \\ - \\ \vdots \\ - \\ \mathbf{CA}^{t-1} \end{array}\right] \tag{280}$$

and, as before, we say that a system is **observable** if $\mathcal{O}^+ \triangleq \mathcal{O}^+(n)$, and $\text{rank}(\mathcal{O}^+) = n$.

observable system

Example XXV:

As a parallel to the previous case, the system

$$\begin{bmatrix} x'_t \\ - \\ x''_t \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{P} & \mathbf{Q} \\ - & - \\ 0 & \mathbf{R} \end{array}\right] \begin{bmatrix} x'_{t-1} \\ - \\ x''_{t-1} \end{bmatrix} \qquad z_t = \left[\begin{array}{c|c} \mathbf{T} & 0 \end{array}\right] x_t \tag{281}$$

has $\text{rank}(\mathcal{O}^+) = p$, and is not observable (x'' has no effect, direct or indirect, on z).(end of example)

This kind of systems is used mainly to model physical systems for the purpose of control. This can work only if the system is observable (we can observe what's going on and get the complete picture of the state) and reachable (we can use the inputs to modify the state as we wish). In this situation, lack of observability or reachability is the mark of a poor model, one that doesn't serve the purpose for which it was designed—it is generally a sign that the model has to be re-done. We shall, from now on, assume that the people who do the modeling know their business, so we shall always consider systems that are both observable and reachable.

4.1 The Kalman Filter

So far we have considered an ideal world in which no noise is present, but now we have to come back to earth, where noise and measurement errors are a fact of life. In one common scenario, the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are known and, by observing the output of the system, we need to know the state x_t at all times. In the absence of noise, the matrix $\mathcal{O}^+$ allows us this estimation, albeit with the delay of n steps (we need up to n observations in order to know x_0). If there is noise, the situation is more complicated. Here we present a classical solution to the problem: the **Kalman Filter**. In the following, the input u_t, which is controlled by us and therefore is always known without noise is not essential, so we shall assume, for the moment, that $u_t \equiv 0$. Consider therefore the system

Kalman filter

$$\begin{aligned} x_{t+1} &= \mathbf{A}x_t + w_t \\ z_t &= \mathbf{C}x_t + v_t \end{aligned} \tag{282}$$

where w_t and v_t are Gaussian white noises with zero mean (if the mean is $\mu \neq 0$ we can simply model it as a system with a constant input μ and a zero-mean noise) and covariance

$$\begin{aligned} \mathbf{Q} &= \mathbb{E}[w_t w'_t] \\ \mathbf{R} &= \mathbb{E}[v_t v'_t] \end{aligned} \tag{283}$$

The arguments that we present do not require the noise to be Gaussian (they do require it to be stationary) but the solution that we find is optimal only in the case of Gaussian noise. Our purpose

is to find an estimate $\hat{x}_t$ based on the prior knowledge of the system and the observed output so as to minimize the square error

$$\mathbf{P}_t = \mathbb{E}[e_t e_t'] \stackrel{\triangle}{=} \mathbb{E}[(\hat{x}_t - x_t)(\hat{x}_t - x_t)'] \tag{284}$$

where ew have defined the deviation

$$e_t = \hat{x}_t - x_t \tag{285}$$

between our estimate and the true state. Specifically, we are interested in minimizing the trace $T[\mathbf{P}_t]$, which corresponds to the component-wise mean square error

$$T[\mathbf{P}_t] = \mathbb{E}[\sum_k (\hat{x}_{t,k} - x_{t,k})^2] \tag{286}$$

(here and in the following, we indicate with $x_{t,k}$ the kth component of teh state x_t). Assume that we have a prior estimate of x_t, $\bar{x}_t$. We create $\hat{x}_t$ by correcting for the prediction error between the output predicted through $\bar{x}_t$ and the output that we observed:

$$\hat{x}_t = \bar{x}_t + K_t(z_t - \mathbf{C}\bar{x}_t) \tag{287}$$

K_t is called the **Kalman gain**, and it is the element that we have to determine in order to minimize the error. The stochastic process

Kalman gain

$$i_t = z_t - \mathbf{C}\bar{x}_t \tag{288}$$

is called the **innovation process**: it represents the part of the observable behavior of the system which is "new" respect to our prediction: the unpredictable and unexpected. Inserting the second of (282) into (287), we have

innovation process

$$\hat{x}_t = \bar{x}_t + K_t(\mathbf{C}x_t + v_t - \mathbf{C}\bar{x}_t) \tag{289}$$

and

$$\begin{aligned} e_t &= x_t - \hat{x}_t \\ &= x_t - \bar{x}_t - K_t\big(\mathbf{C}(x_t - \bar{x}_t) + v_t\big) \\ &= (\mathbf{I} - K_t\mathbf{C})(x_t - \bar{x}_t) - K_t v_t \end{aligned} \tag{290}$$

The error $(x_t - \bar{x}_t)$ does not depend on the observations at time t, therefore it is uncorrelated with v_t:

$$\mathbb{E}[(x_t - \bar{x}_t)v_t] = 0 \tag{291}$$

We have therefore

$$\begin{aligned} \mathbf{P}_t &= \mathbb{E}[e_t e_t'] \\ &= (\mathbf{I} - K_t\mathbf{C})\mathbb{E}[(x_t - \bar{x}_t)(x_t - \bar{x}_t)'](\mathbf{I} - K_t\mathbf{C})' + K_t\mathbb{E}[v_t v_t']K_t' \end{aligned} \tag{292}$$

The value

$$\bar{\mathbf{P}}_t = \mathbb{E}[(x_t - \bar{x}_t)(x_t - \bar{x}_t)'] \tag{293}$$

is the prior estimate of $\mathbf{P}$. From this and the second of (283) we have

$$\mathbf{P}_t = (\mathbf{I} - K_t\mathbf{C})\bar{\mathbf{P}}_t(\mathbf{I} - K_t\mathbf{C})' + K_t\mathbf{R}K_t' \tag{294}$$

Expanding $\mathbf{P}_t$ we have

$$\begin{aligned}\mathbf{P}_t &= \bar{\mathbf{P}}_t(\mathbf{I}-K_t\mathbf{C})' - K_t\mathbf{C}\bar{\mathbf{P}}_t(\mathbf{I}-K_t\mathbf{C})' + K_t\mathbf{R}K_t' \\ &= \bar{\mathbf{P}}_t - \bar{\mathbf{P}}_t(K_t\mathbf{C})' - K_t\mathbf{C}\bar{\mathbf{P}}_t + K_t\mathbf{C}\bar{\mathbf{P}}_t(K_t\mathbf{C})' + K_t\mathbf{R}K_t' \\ &= \bar{\mathbf{P}}_t - \bar{\mathbf{P}}_t(K_t\mathbf{C})' - K_t\mathbf{C}\bar{\mathbf{P}}_t + K_t(\mathbf{C}\bar{\mathbf{P}}_t\mathbf{C}'+\mathbf{R})K_t'\end{aligned} \tag{295}$$

The trace is invariant to transposition ($T[\mathbf{A}] = T[\mathbf{A}']$), so we have

$$T[\mathbf{P}_t] = T[\bar{\mathbf{P}}_t] = 2T[K_t\mathbf{C}\bar{\mathbf{P}}_t] + T[K_t(\mathbf{C}\bar{\mathbf{P}}_t\mathbf{C}'+\mathbf{R})K_t'] \tag{296}$$

We minimize it by setting its first derivative to zero:

$$\frac{d}{dK_t}T[\mathbf{P}_t] = -2(\mathbf{C}\bar{\mathbf{P}}_t) + 2K_t(\mathbf{C}\bar{\mathbf{P}}_t\mathbf{C}'+\mathbf{R}) \tag{297}$$

which gives

$$K_t = \bar{\mathbf{P}}_t\mathbf{C}'(\mathbf{C}\bar{\mathbf{P}}_t\mathbf{C}'+\mathbf{R})^{-1} \tag{298}$$

With this Kalman gain, we can compute the error variance

$$\begin{aligned}\mathbf{P}_t &= \bar{\mathbf{P}}_t - K_t\mathbf{C}\bar{\mathbf{P}}_t - \bar{\mathbf{P}}_t(K_t\mathbf{C})' + K_t(\mathbf{C}\bar{\mathbf{P}}_t\mathbf{C}'+\mathbf{R})K_t' \\ &= \bar{\mathbf{P}}_t - K_t\mathbf{C}\bar{\mathbf{P}}_t - \bar{\mathbf{P}}_t(K_t\mathbf{C})' + \bar{\mathbf{P}}_t\mathbf{C}'K_t' \\ &= (\mathbf{I}-K_t\mathbf{C})\bar{\mathbf{P}}_t\end{aligned} \tag{299}$$

With these equations, we can update our estimate of $\bar{x}_t$ and our error $\mathbf{P}_t$ once we have computed the priors $\bar{x}_t$ and $\bar{\mathbf{P}}_t$. We get the priors from our estimated at the previous time:

$$\begin{aligned}\bar{x}_t &= \mathbf{A}\hat{x}_{t-1} \\ \bar{\mathbf{P}}_t &= \mathbb{E}[e_te_t'] = \mathbb{E}\big[(\mathbf{A}e_{t-1}+w_{t-1})(\mathbf{A}e_{t-1}+w_{t-1})'\big] \\ &\overset{(*)}{=} \mathbb{E}\big[\mathbf{A}e_{t-1}(\mathbf{A}e_{t-1})'\big] + \mathbb{E}\big[w_{t-1}w_{t-1}'\big] \\ &= \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}' + \mathbf{Q}\end{aligned} \tag{300}$$

where the equality (*) is because e_t and w_t are uncorrelated. So, suppose that we have our priors $\bar{x}_t$ and $\bar{\mathbf{P}}_t$. One step of the Kalman filter consists of the following:

Compute the Kalman Gain	$K_t = \bar{\mathbf{P}}_t\mathbf{C}'(\mathbf{C}\bar{\mathbf{P}}_t\mathbf{C}'+\mathbf{R})^{-1}$
Update the estimate	$\hat{x}_t = \bar{x}_t + K_t(z_t - \mathbf{C}\bar{x}_t)$
Update the covariance	$\mathbf{P}_t = (\mathbf{I}-K_t\mathbf{C})\bar{\mathbf{P}}_t$
Compute the priors	$\bar{x}_{t+1} = \mathbf{A}\bar{x}_t$
	$\bar{\mathbf{P}}_{t+1} = \mathbf{A}\mathbf{P}_t\mathbf{A}' + \mathbf{Q}$

5 Random Walks on Graphs

In this section, we consider a specific case of discrete-time Markov process with a finite state-space of considerable importance in applications: *random walks on graphs*. In this case, the state-space is the set of nodes of the graph, and the process is in state i is the walker is positioned at node i of the graph. We shall consider and study different types of walk,characterized by different transition probabilities $P(i|j)$. We begin with the simplest one, in which a walker jumps at random from a node to one of its neighbors.

Let $G = (V, E)$ be an undirected graph with $V = \{1, \dots, n\}$ and $E \subseteq V \times V$, with $|E| = 2m$.[6]The graph can be defined in terms of the adjacency matrix $\mathbf{A}$ with elements

$$a_{uv} = \begin{cases} 1 & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \tag{301}$$

Note that, since the graph is undirected, $\mathbf{A}$ is symmetric ($\mathbf{A} = \mathbf{A}'$). We indicate with d_u the degree of node u:

$$d_u = |\{(u, v) \in E\}| = \sum_v a_{uv} \tag{302}$$

At time t, the walk is at a node u. At time $t + 1$, the walker picks one of its neighbors, v, at random with uniform probability and moves to it. The walk is a Markov process, since the probability of moving to v at $t + 1$ depends only on the fact that at time t the walk was at node u, not on how it got there. The transition probability is

$$P(v|u) = q(u \to v) = \begin{cases} \dfrac{1}{d_u} & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \tag{303}$$

Note that the walk always moves so if the graph has no self loops ($a_{u,u} = 0$) we have $P(a|a) = 0$. Define the matrix

$$\mathbf{D} = \text{diag}(d_1, \dots, d_n) \tag{304}$$

and

$$\begin{aligned} \mathbf{M} &= \mathbf{D}^{-1}\mathbf{A} \triangleq \left\{m_{ab}\right\}_{a,b=1}^{n} \\ \mathbf{W} &= \mathbf{A}\mathbf{D}^{-1} = \mathbf{M}' \triangleq \left\{w_{ab}\right\}_{a,b=1}^{n} \end{aligned} \tag{305}$$

[6]The "2" in $2m$ is due to the fact that G is undirected. The definition of E as a set of pairs of nodes implies that each edge in E is directed. In order to have an undirected graph, if $(u, v) \in E$, then $(v, u) \in E$, which means that the size of E is always even and equal to twice the number of undirected edges.

Then[7]

$$q(a \to b) = m_{ab} = w_{ba} \tag{306}$$

Our main instrument for the study of random walks is the *master equation*: the probability of being at node u at time $t+1$ is the sum over all nodes k of the probability of being in k at time t and then jumping from k to u:

$$P_u(t+1) = \sum_{k=1}^{n} P_k(t) q(k \to u) = \sum_{k=1}^{n} w_{uk} P_k(t) \tag{307}$$

Defining the probability vector

$$P(t) = \big[p_1(t), \ldots, p_n(t)\big]' \tag{308}$$

we have the **master equation**

master equation

$$P(t+1) = \mathbf{W} P(t) \tag{309}$$

If P_0 is the initial probability distribution of the walker on the graph (viz. $P_0(u)$ is the probability of starting the walk at node u), then we can iterate (309) obtaining

$$P(t) = \mathbf{W}^t P_0 \tag{310}$$

Since w_{uv} is the probability of moving from v to u, it is $w_{uv} \geq 0$ (and, for the same reason, $m_{uv} \geq 0$). Moreover, if at t we are in v, at $t+1$ we are certainly somewhere that we can reach from v. This entails

$$\sum_u w_{uv} = 1 \tag{311}$$

for all v, and

$$\sum_v m_{uv} = 1 \tag{312}$$

for all u.

5.1 Other types of walk

The random walk we just introduced is the most "classical" and the most studied. Before we continue, we shall here introduce other types of random walks that we shall study, with more or less details, in the following.

In the radom walk that we have just considered, at any step the walker moves from one node to another. The **lazy random walk** is a walk in which, with probability 1/2, the walker doesn't move

lazy random walk

[7] In the literature, we find both matrices used to describe random walks. Both have their pros and cons. $\mathbf{M}$ is often more intuitive since the indices of its elements go in the same order as the movement: m_{ab} is the probability of moving from a to b. On the other hand, $\mathbf{W}$ generates linear equation in the customary form, using column vectors rather than its transposes. Here, we use the matrix $\mathbf{W}$, which generates the standard form of the equations, and the matrix $\mathbf{M}$ or the notation $q(a \to b)$ whenver it is convenient.

and remains in the same node where it was. Its master equation is:

$$\begin{aligned} P_u(t+1) &= \frac{1}{2}P_u(t) + \frac{1}{2}\sum_{k=1}^{n} P_k(t) q(k \to u) \\ &= \frac{1}{2}P_u(t) + \frac{1}{2}\sum_{k=1}^{n} w_{uk} P_k(t) \end{aligned} \tag{313}$$

That is

$$P(t+1) = \widehat{\mathbf{W}} P(t) \tag{314}$$

with

$$\widehat{\mathbf{W}} = \frac{1}{2}(\mathbf{I} + \mathbf{W}) \tag{315}$$

Pagerank walk

In the **PageRank walk**, the walker moves from one node to a random neighbor with probability $(1-\alpha)$ while, with probability α, it jumps to a random node of the graph ($0 \leq \alpha \leq 1$); its master equation is

$$P_u(t+1) = \frac{\alpha}{n} + (1-\alpha)\sum_{k=1}^{n} w_{uk} P_k(t) \tag{316}$$

that is

$$P(t+1) = \frac{\alpha}{n}\mathbf{1} + (1-\alpha)\mathbf{W}P(t) \tag{317}$$

(here and in the following, $\mathbf{1}$ is the vector with all components equal to 1).

personal page rank

A variation of this walk is the so-called **personal page rank**: a node u is given and, with probability α, the walk jumps back to u. If we define the vectors of the natural basis of $\mathbb{R}^n$ as

$$e_u = \big[\underbrace{0, \ldots, 1}_{u}, 0, \ldots, 0\big]' \tag{318}$$

then the master equation of the personal page rank walk is

$$P(t+1) = \frac{\alpha}{n} e_u + (1-\alpha)\mathbf{W}P(t) \tag{319}$$

long range random walk

Finally, **long range random walks** are the analogous of Levy walks in the plane []. The master equation is still formally the same as (307) but, unlike the standard walk, we admit the possibility of jumping at nodes at a distance δ from the current one, although the probability of doing so decreases with the distance. If δ_{ab} is the graph distance between nodes a and b, then

$$w_{ab} \sim \delta_{ab}^{-\alpha} \tag{320}$$

for some parameter α. Normalizing, we have

$$w_{ab} = \frac{\delta_{ab}^{-\alpha}}{\sum_{u \neq a} \delta_{au}^{-\alpha}} \tag{321}$$

We shall, for the time being, limit out consideration to the standard and the lazy walks.

5.2 The stationary probability

Each walk is, of course, different, and evolves in time according to its own random choice of the neighbor on which to move next. We are not interested in studying the walks one by one, but rather in studying their statistical properties as an ensemble. One thing we are especially interested in is the **stationary probability**: the probability of being in this or that node when the walk time becomes very large, that is:

stationary probability
π

$$\pi = \lim_{t \to \infty} P(t) \tag{322}$$

The interest in this distribution derives mainly from it being insesitive to the initial conditions: if the walk is not cahotic (and, in this case, it is not), no matter from which node we start, we shall end up with the same stationary probability. The stationary probability satisfies the equation $\mathbf{W}\pi = \pi$ and we can check that

$$\begin{aligned} \pi &= [\pi_1, \ldots, \pi_n]' \\ \pi_u &= \frac{d_u}{2m} \end{aligned} \tag{323}$$

is an equilibrium:

$$\begin{aligned} {}_u &= \sum_k w_{uk}\pi_k = \sum_k [\mathbf{A}'\mathbf{D}^{-1}]_{uk}\pi_k \\ &= \sum_k \mathbf{A}_{ku}\mathbf{D}_{kk}^{-1}\pi_k \\ &= \sum_k \mathbf{A}_{ku}\frac{1}{2m} \\ &= \frac{d_u}{2m} = \pi_u \end{aligned} \tag{324}$$

Note that $\mathbf{W}\pi = \pi$ entails that π is an eigenvector of $\mathbf{W}$ with eigenvalue $\lambda = 1$. Let $\mathbf{1}_n = [1, 1, \ldots, 1]' \in \mathbb{R}^n$. Because of (311), we have $\mathbf{1}_n'\mathbf{W} = \mathbf{1}_n'$, that is, $\mathbf{1}_n'$ is a right eigenvector of $\mathbf{W}$ with engenvalue $\lambda = 1$ or, equivalently, $\mathbf{1}_n$ is an eigenvector of $\mathbf{W}' = \mathbf{M}$ with $\lambda = 1$:

$$\mathbf{W}'\mathbf{1}_n = \mathbf{M}\mathbf{1}_n = \mathbf{1}_n \tag{325}$$

The vector π is not just an equilibrium point of the iteration: it is the only one.

Lemma 5.1. *For each vector p, if $\sum_u p_u = 1$ and $\mathbf{W}p = p$, then $p = \pi$.*

Proof. Given p, let u be any node that maximizes

$$\frac{p_u}{d_u} \tag{326}$$

Since $\mathbf{W}p = p$, we have

$$\begin{aligned} \big[\mathbf{W}p\big]_u &= \sum_v w_{uv} p_v \\ &= \sum_{v:(u,v)\in E} \frac{p_v}{d_v} \\ &\overset{\dagger}{\leq} \sum_{v:(u,v)\in E} \frac{p_u}{d_u} \\ &= p_u \end{aligned} \tag{327}$$

The only way to achieve the equality is if all the inequalities in (†) are in fact equalities, that is

$$\frac{p_v}{d_v} = \frac{p_u}{d_u} \tag{328}$$

for all neighbors of u. We can extend (371) to the whole graph as follows.

Let $S \subseteq V$ be the set of nodes for which (371) holds, and assume by contradiction that there is a node $k \in V \backslash S$. Since the graph is connected, it contains a path from u to k. So there must be an edge (a, b) on this path with $a \in S$ and $b \notin S$. Repeating the same argument above we get

$$\frac{p_a}{d_a} = \frac{p_b}{d_b} \tag{329}$$

which contradicts $a \in S$ and $b \notin S$. From the fact that (371) holds for all nodes, we get $p_u \sim d_u$ and normalizing so that $\sum_u p_u = 1$, we obtain $p = \pi$. □

Note that at equilibrium

$$\pi_u q(u \to v) = \pi_v q(v \to u) = \frac{1}{2m} \tag{330}$$

that is, we move from u to v just as often as we move from v to u. Moreover, $\pi_u q(u \to v) = 1/2m$ is a constant independent of u and v, which means that we move on any edge, in any direction, with equal probability.

Remark 1: as we have seen, a random walk is a Markov process in which the nodes are the states and $q(u \to v)$ is the transition probability from state u to state v. In this case the property (330)

means that the Markov process is time reversible.
(end of remark)

A slight extension of these considerations will be useful in the following. Let $P_{ab}(t)$ the probability that, beginning the walk at node a, we end up in b at time t. Then, iterating the master equation (307), we have

$$P_{ab}(t) = \sum_{j_1,\dots,j_{t-1}} q(a \to j_1) q(j_1 \to j_2) \cdots q(j_{t-1} \to b) \tag{331}$$

From (303) we have

$$q(u \to v) = \frac{d_v}{d_u} q(v \to u) \tag{332}$$

Replacing this in all the factors of (331) and simplifying we have

$$\begin{aligned} P_{ab}(t) &= \sum_{j_1,\dots,j_{t-1}} \frac{d_{j_1}}{d_a} q(j_1 \to a) \frac{d_{j_2}}{d_{j_1}} q(j_2 \to j_1) \cdots \frac{d_b}{d_{j_{t-1}}} q(b \to j_{t-1}) \\ &= \frac{d_b}{d_a} \sum_{j_1,\dots,j_{t-1}} q(j_1 \to a) q(j_2 \to j_1) \cdots q(b \to j_{t-1}) \\ &= \frac{d_b}{d_a} P_{ba}(t) \end{aligned} \tag{333}$$

that is

$$d_a P_{ab}(t) = d_b P_{ba}(t) \tag{334}$$

Note that this implies $d_u \pi_v = d_v \pi_u$, from which we can derive $\pi_u = d_u / 2m$.

5.3 Spectral analysis

One powerful instrument for the study if random walks is specral analysis, as it allows to express the powers $\mathbf{W}^t$ that appears in the iterated master equation as an equation in the powers of the eigenvectors of the matrix $\mathbf{W}$.

eigenvalues, eigenvectors

We begin by reminding a few general facts about **eigenvalues and eigenvectors**. For each $n \times n$ symmetric matrix $\mathbf{A}$ there is an orthonormal basis of $\mathbb{R}^n$ $\{v_1, \dots, v_n\}$ such that, for each v_i, $\mathbf{A}v_i = \lambda_i v_i$. The λ_i are the *eigenvalues* of $\mathbf{A}$ and the v_i are the corresponding *eigenvectors*. The v_i form an orthonormal basis of $\mathbb{R}^n$ so, for every $x \in \mathbb{R}^n$, it is

$$x = \sum_{i=1}^{n} (v_i' x) v_i \tag{335}$$

Also

$$\mathbf{A}x = \mathbf{A} \cdot \sum_{i=1}^{n} (v_i' x) v_i = \sum_{i=1}^{n} (v_i' x)(\mathbf{A}v_i) = \sum_{i=1}^{n} (v_i' x) \lambda_i v_i \tag{336}$$

and

$$\mathbf{A}^k x = \sum_{i=1}^{n} (v_i' x) \lambda_i^k v_i \tag{337}$$

* * *

Coming back to our random walks, we notice that, unfortunately, the matrix $\mathbf{W}$ is not symmetric. We can however work out a transformation of $\mathbf{W}$ into a symmetric matrix that will give us useful information. From the definition of $\mathbf{D}$ we have

$$\begin{aligned} \mathbf{D}^{\frac{1}{2}} &= \mathrm{diag}(\sqrt{d_1}, \ldots, \sqrt{d_n}) \\ \mathbf{D}^{-\frac{1}{2}} &= \mathrm{diag}(\frac{1}{\sqrt{d_1}}, \ldots, \frac{1}{\sqrt{d_n}}) \end{aligned} \tag{338}$$

Define the matrix

$$\mathbf{\Xi} \stackrel{\triangle}{=} \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}} (\mathbf{A}\mathbf{D}^{-1}) \mathbf{D}^{\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \tag{339}$$

which is symmetric, since $\mathbf{A}$ is symmetric. For the lazy walk, we define

$$\widehat{\mathbf{\Xi}} = \frac{1}{2}(\mathbf{I} + \mathbf{\Xi}) = \mathbf{D}^{-\frac{1}{2}} \widehat{\mathbf{W}} \mathbf{D}^{\frac{1}{2}} \tag{340}$$

The important property of $\mathbf{\Xi}$, from the point of view of our analysis, is that is symmetric and it has the same eigenvalues as $\mathbf{W}$: if v is an eigenvector of $\mathbf{\Xi}$ with eigenvalue λ, we have

$$\lambda v = \mathbf{\Xi} v = (\mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{\frac{1}{2}}) v \tag{341}$$

therefore

$$\lambda (\mathbf{D}^{\frac{1}{2}} v) = \mathbf{W} (\mathbf{D}^{\frac{1}{2}} v) \tag{342}$$

So, if v_i is an eigenvector of $\mathbf{\Xi}$ with eigenvalue λ_i, then $\mathbf{D}^{\frac{1}{2}} v$ is an eigenvector of $\mathbf{W}$ with the same eigenvalue λ_i. Note that this entails that the eigenvalues of $\mathbf{W}$ are all real.

With the eigenvector decomposition given by $\mathbf{\Xi}$, we can compute a closed form for $\mathbf{W}^k x$ for any vector x:

$$\begin{aligned} \mathbf{W}^k x &= (\mathbf{D}^{\frac{1}{2}} \mathbf{\Xi} \mathbf{D}^{-\frac{1}{2}})^k x \\ &= \mathbf{D}^{\frac{1}{2}} \mathbf{\Xi}^k \mathbf{D}^{-\frac{1}{2}} x \\ &= \sum_{i=1}^{n} \lambda_i^k \mathbf{D}^{\frac{1}{2}} v_i (v_i' \mathbf{D}^{-\frac{1}{2}} x) \end{aligned} \tag{343}$$

Equation (343) shows how the eigenvalues determine the behavior of the various components of $\mathbf{W}^k x$ as $k \to \infty$. If $|\lambda_i| > 1$, the component in the direction $\mathbf{D}^{\frac{1}{2}} v_i$ will grow without bounds; if $|\lambda_i| < 1$, the corresponding component will shrink to zero; if $\lambda_i = -1$, the corresponding component will oscillate back and forth wothout settling to any value; finally, if $\lambda_i = 1$ the corresponding component will remain constant. The probabilistic interpretation requires that no component grow without bounds so, if we had $|\lambda_k| > 1$ for some k, we would be in trouble. Fortunately, as we shall see shortly, this is not the case.

Remark 2: Before we analyze the eigenvalues of $\mathbf{W}$, we shall check that (343) implies that $\mathbf{W}^t \pi = \pi$. Apart from checking that we are indeed on the right track, this "sanity check" will give us a couple of useful equations that we shall need in the following.

Since π is an eigenvector of $\mathbf{W}$ with eigenvalue 1, there has to be a corresponding eigenvector of $\mathbf{\Xi}$, also with eigenvalue 1. Call v_1 this eigenvector which, because of (342), is

$$v_1 = \frac{\mathbf{D}^{\frac{1}{2}} \pi}{\|\mathbf{D}^{\frac{1}{2}} \pi\|} \tag{344}$$

(we divide by the norm because we require the eigenvectors to be of unit norm). Let us check what the norm in the denominator is. We have $\pi_a = d_a / 2m$, so

$$\left[\mathbf{D}^{-\frac{1}{2}} \pi\right]_a = \frac{\sqrt{d_a}}{2m} \tag{345}$$

and

$$\|\mathbf{D}^{\frac{1}{2}} \pi\| = \frac{1}{2m} \sqrt{\sum_a (\sqrt{d_a})^2} = \frac{1}{2m} \sqrt{\sum_a d_a} = \frac{1}{\sqrt{2m}} \tag{346}$$

This gives us

$$v_i = \sqrt{2m} \mathbf{D}^{\frac{1}{2}} \pi \tag{347}$$

that is, component-wise

$$v_{1,a} = \sqrt{\frac{d_a}{2m}} = \sqrt{\pi_a} \tag{348}$$

The basis is orthonormal, and $\mathbf{D}^{-\frac{1}{2}} \pi$ lies in the direction of v_1, so

$$v_i' \mathbf{D}^{-\frac{1}{2}} \pi = 0 \qquad i \geq 2 \tag{349}$$

and

$$v_1' \mathbf{D}^{-\frac{1}{2}} \pi = \|\mathbf{D}^{-\frac{1}{2}} \pi\| = \frac{1}{\sqrt{2m}} \tag{350}$$

so, applying (343), of which only the term $i = 1$ is non-zero,

$$\mathbf{W}^k \pi = \lambda_1^k \mathbf{D}^{\frac{1}{2}} v_1 \frac{1}{\sqrt{2m}} = \mathbf{D}^{\frac{1}{2}} \frac{\mathbf{D}^{-\frac{1}{2}}}{1/\sqrt{2m}} \frac{1}{\sqrt{2m}} = \pi \tag{351}$$

(end of remark)

Now that we have made our check, we can study the eigenvalues of $\mathbf{W}$ and, through them, the convergence of the walk to the stationary solution.

Theorem 5.1. *Let* $\mathbf{W}$ *be the walk matrix of a connected graph. Then all the eigenvalues of* $\mathbf{W}$ *are in* $[-1, 1]$*, and the eigenvalue* $\lambda = 1$ *has mutiplicity 1.*

Proof. Lemma 5.1, in which we proved that the steady state solution is unique, implies that $\lambda = 1$ is an eigenvalue of multiplicity one.

Let v be an eigenvector with eigenvalue λ, and let a be a vertex such that, for all $b \in V$,

$$\frac{|v_a|}{d_a} \geq \frac{|v_b|}{d_b} \tag{352}$$

We have

$$\lambda v_a = \left[\mathbf{W}v\right]_a = \sum_b w_{ab} v_b = \sum_{b:(a,b)\in E} \frac{v_b}{d_b} \tag{353}$$

therefore

$$\begin{aligned} |\lambda||v_a| &= \left| \sum_{b:(a,b)\in E} \frac{v_b}{d_b} \right| \\ &\leq \sum_{b:(a,b)\in E} \frac{|v_b|}{d_b} \\ &\leq \sum_{b:(a,b)\in E} \frac{|v_a|}{d_a} \\ &= |v_a| \end{aligned} \tag{354}$$

so that $|\lambda| \leq 1$. □

A similar, but nicer (for convergence) property holds for the lazy walk matrix $\widehat{\mathbf{W}}$.

Corollary 5.1. *All eigenvalues of* $\widehat{\mathbf{W}}$ *are in* $[0, 1]$*, and the eigenvalue* $\lambda = 1$ *has multiplicity 1.*

Proof. Since $\widehat{\mathbf{W}} = \frac{1}{2}\mathbf{I} + \frac{1}{2}\mathbf{W}$, $\widehat{\mathbf{W}}$ has the same eigenvectors as $\mathbf{W}$, with

$$\widehat{\mathbf{W}} v_i = \frac{1}{2}\mathbf{I} v_i + \frac{1}{2}\mathbf{W} v_i = \frac{1}{2}(v_i + \lambda_i v_i) \tag{355}$$

so, if λ_i is an eigenvalue of $\mathbf{W}$, $(1+\lambda_i)/2$ is an eigenvalue of $\widehat{\mathbf{W}}$ □

In the following we shall consider that the eigenvalues are ordered, so that

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1 \tag{356}$$

The inequality $\lambda_n \geq -1$ deserves some consideration. The Master equation guarantees that if p_0 is a probability ($\sum p_{0,i} = 1$), $\mathbf{W}^t p_0$ also is, which entails that no component of $\mathbf{W}^t p_0$ diverges, that is, $|\lambda_i| \leq 1$ for all i.

It is easy to see that if G is bipartite, $\mathbf{W}^t p_0$ does not converge. Let $V = S \cup S'$, with $S \cap S' = \emptyset$, and let all the edges be between a node of S and a node of S'. If p_0 is concentrated in S, then at all t odd, the walker will be in S' and at all t even it will be in S. So, $\mathbf{W}^t p_0$ will oscillate between S and S' without converging. This corresponds to the presence of $\lambda_n = -1$. On the other hand, if G is not bipartite, this will not happen, and standard arguments show that $\lambda_n > -1$, that is, $\mathbf{W}^t p_0$ will converge to a steady state. This is shown in the following theorem and its corollary

Theorem 5.2. $\mathbf{M}$ *has an eigenvalue* $\lambda = -1$ *iff the graph is bipartite.*

Proof. Suppose first that the graph is bipartite, with the two sets of nodes that constitute its parts containing n and m nodes. Label the nodes so that the first set contains the nodes $1, \ldots, n$ and the second the nodes $n+1, \ldots, n+m$. This entails that the adjacency matrix has a structure

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{B} \\ \mathbf{B}' & 0 \end{bmatrix} \tag{357}$$

with $B \in \{0,1\}^{n \times m}$. and

$$\mathbf{M} = \begin{bmatrix} 0 & \mathbf{M}_1 \\ \mathbf{M}_2 & 0 \end{bmatrix} \tag{358}$$

with $\mathbf{M}_1 \in \mathbb{R}^{n \times m}$ and $\mathbf{M}_2 \in \mathbb{R}^{m \times n}$. From (312) and the structure of $\mathbf{M}$, we have

$$\begin{aligned} \mathbf{M}_1 \mathbf{1}_m &= \mathbf{1}_n \\ \mathbf{M}_2 \mathbf{1}_n &= \mathbf{1}_m \end{aligned} \tag{359}$$

Consider now the vector

$$\mathbf{b} = \begin{bmatrix} \mathbf{1}_n \\ -\mathbf{1}_m \end{bmatrix} \tag{360}$$

We have

$$\mathbf{M}\mathbf{b} = \begin{bmatrix} 0 & \mathbf{M}_1 \\ \mathbf{M}_2 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1}_n \\ -\mathbf{1}_m \end{bmatrix} = \begin{bmatrix} -\mathbf{M}_1 \mathbf{1}_m \\ \mathbf{M}_2 \mathbf{1}_n \end{bmatrix} = \begin{bmatrix} -\mathbf{1}_n \\ \mathbf{1}_m \end{bmatrix} = -\mathbf{b} \tag{361}$$

Therefore $\mathbf{b}b$ is an eigenvector with eigenvalue -1.

* * *

Suppose now that $\mathbf{M}$ has en eigenvalue -1. Let b be the corresponding eigenvector, and write

$$\mathbf{c} = \mathbf{W}\,\mathbf{b} \tag{362}$$

Clearly $c_i = -b_i$.

We first show by contradiction that all the components of $\mathbf{b}$ have the same absolute value. Let b_h be the element of $\mathbf{b}$ with the smallest absolute value, b_k the one with the largest, and that

$$|b_h| < |b_k| \tag{363}$$

Then

$$|c_k| = \left|\sum_j m_{kj} b_j\right| \leq \sum_j m_{kj}|b_j| \overset{(\dagger)}{<} |b_k| \sum_j m_{mj} \overset{\ddagger}{=} |b_k| \tag{364}$$

where (†) is strict because of (363) and (‡) is due to (312). This contradicts $c_k = -b_k$.

Since all the elements have the same absolute value, we can scale the eigenvector to have them all unitary, that is, to have $\mathbf{b} \in \{-1, 1\}^{n+m}$, and we can re-label the nodes of the graph, thereby shifting the rows and columns of $\mathbf{M}$ to that all the positive 1s are the the first n positions of $\mathbf{b}$, and the negative -1s are in the last positions, that is, we can, without loss of generality, consider that $\mathbf{b}$ is in the form (360). Consider not b_i with $i \leq n$. Then $c_i = -b_i = -1$ leads to

$$c_i = \sum_j m_{ij} b_j = \sum_{j=1}^{n} m_{ij} - \sum_{j=n+1}^{n+m} m_{ij} = -b_i = -1 \tag{365}$$

Since $m_{ij} \geq 0$ and (312) holds, the only way to obtain -1 is to have $m_{ij} = 0$ for $j \leq n$ and $\sum_{j=n+1}^{n+m} m_{ij} = 1$. The same argument can be done for all the $i = 1, \ldots, n$. A similar argument for $i > n$ shows that in this case we must have $m_{ij} = 0$ for $j > n$.

Putting the two together we have the condition

$$m_{ij} = 0 \quad \text{if} \quad (i \leq n \text{ and } j \leq m) \text{ or } (i > n \text{ and } j > n) \tag{366}$$

This implies that $\mathbf{M}$ has the structure (358), which entails that the graph is bipartite. □

Since $\mathbf{W} = \mathbf{M}'$ and the two have the same eigenvalues, we have the following corollary:

Theorem 5.3. $\mathbf{W}$ *has an eigenvalue* $\lambda = -1$ *iff the graph is bipartite.*

The lazy walk doesn't have the same problem with bipartite graphs: in that case all the eigenvalues are positive, and the walk will always converge.

5.4 Convergence to π

We shall use the eigenvalues of $\mathbf{W}$ to study the convergence of the walk to π. As before, v_i are the eigenvectors of $\mathbf{\Xi}$, defined as in (339), forming an orthonormal basis; v_1 is the eigenvector corresponding to $\lambda_1 = 1$. Let p_0 be the initial probability distribution of the walk. The eigenvectors form a basis, so we can write

$$\begin{aligned} \mathbf{D}^{-\frac{1}{2}} p_0 &= \sum_i \alpha_i v_i \\ \alpha_i &= v_i' \mathbf{D}^{-\frac{1}{2}} p_0 \end{aligned} \tag{367}$$

Note that, since the v_i are orthonormal

$$\sum_i \alpha_i^2 = \|\mathbf{D}^{-\frac{1}{2}} p_0\|^2 \tag{368}$$

We are interested especially in α_1. From (347), we have

$$\begin{aligned} \alpha_1 &= v_1' \mathbf{D}^{-\frac{1}{2}} p_0 \\ &= \sqrt{2m} \pi' \mathbf{D}^{-\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} p_0 \\ &= \sqrt{2m} \frac{1}{2m} \pi' \mathbf{D}^{-1} p_0 \\ &\stackrel{\dagger}{=} \frac{1}{\sqrt{2m}} \mathbf{1}' p_0 \\ &\stackrel{\ddagger}{=} \frac{1}{\sqrt{2m}} \end{aligned} \tag{369}$$

Equality (†) derives from

$$\pi' \mathbf{D}^{-1} = \Big[\frac{d_1}{2m} \cdot \frac{1}{d_1}, \ldots, \frac{d_n}{2m} \cdot \frac{1}{d_n} \Big] = \frac{1}{2m} \mathbf{1}' \tag{370}$$

while (‡) from the fact that $\mathbf{1}' p_0 = \sum p_{0,i} = 1$, since p_0 is a probability.

We then have, from (343)

$$\begin{aligned}
\mathbf{W}^t p_0 &= \sum_{i=1}^{n} \lambda_i^t \mathbf{D}^{\frac{1}{2}} v_i \big(v_i' \mathbf{D}^{-\frac{1}{2}} p_0\big) \\
&= \sum_{i=1}^{n} \lambda_i^t \mathbf{D}^{\frac{1}{2}} \alpha_i v_i \\
&= \mathbf{D}^{\frac{1}{2}} \alpha_1 v_i + \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i \\
&= \mathbf{D}^{\frac{1}{2}} \frac{1}{\sqrt{2m}} \sqrt{2m} \mathbf{D}^{-\frac{1}{2}} \pi + \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i \\
&= \pi + \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t v_i' \mathbf{D}^{-\frac{1}{2}} p_0 v_i \\
&= \pi + \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t (v_i' p_0) v_i
\end{aligned} \tag{371}$$

This gives us the convergence in terms of the eigenvalues $\lambda_2, \ldots, \lambda_n$. We can say something more about the convergence of the lazy random walk, for which it is $\lambda_i \in [0, 1]$.

Theorem 5.4. *Consider a lazy random walk on a connected graph with a walk matrix $\widehat{\mathbf{W}}$. For any initial probability p_0, we have*

$$\|\widehat{\mathbf{W}}^t p_0 - \pi\| \leq \sqrt{\frac{\max\{d_a\}}{\min\{d_a\}}} \lambda_2^t \tag{372}$$

(where the max and min are taken over all the nodes of the graph). If the walk starts at a vertex a, then for any other vertex b we have

$$\|\big[\widehat{\mathbf{W}}^t p_0\big]_b - \pi_b\| \leq \sqrt{\frac{d_b}{d_a}} \lambda_2^2 \tag{373}$$

Proof. Note that the largest entry of $\mathbf{D}^{-frac12}$ is $1/\sqrt{\min\{d_a\}}$, so

$$\|\mathbf{D}^{-\frac{1}{2}} p_0\|^2 \leq \frac{1}{\sqrt{\min\{d_a\}}} \tag{374}$$

and, consequently

$$\sum_i \alpha_i^2 = \|\mathbf{D}^{-\frac{1}{2}} p_0\|^2 \leq \frac{1}{\sqrt{\min\{d_a\}}} \tag{375}$$

Consider again (371):

$$\widehat{\mathbf{W}}^t p_0 = \pi + \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i \tag{376}$$

Consider the second term of the right-hand side:

$$\begin{aligned} \|\sum_{i=2}^{n} \lambda_i^t \alpha_i v_i\|^2 &= \sum_{i=2}^{n} \left(\lambda_i^t \alpha_i v_i\right)^2 \\ &\leq \lambda_2^{2t} \sum_{i=2}^{n} \alpha_i^2 \\ &\leq \lambda_2^{2t} \frac{1}{\min\{d_a\}} \end{aligned} \tag{377}$$

The largest entry of $\mathbf{D}^{\frac{1}{2}}$ is $\sqrt{\max\{d_a\}}$, therefore

$$\begin{aligned} \|\widehat{\mathbf{W}}^t p_0 - \pi\|^2 &= \|\mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i\|^2 \\ &\leq \|\mathbf{D}^{\frac{1}{2}}\|^2 \lambda_2^{2t} \frac{1}{\min\{d_a\}} \\ &\leq \lambda_2^{2t} \frac{\max\{d_a\}}{\min\{d_a\}} \end{aligned} \tag{378}$$

As to the second point suppose we start at node a ($p_0 = e_a$), then

$$\|\mathbf{D}^{-\frac{1}{2}} p_0\| = \frac{1}{\sqrt{d_a}} \tag{379}$$

Let e_b be the unit vector in direction b, then

$$\left[\widehat{\mathbf{W}}^t p_0\right]_b = e_b' \widehat{\mathbf{W}}^t p_0 = e_b' \pi + e_b' \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i \tag{380}$$

We can upper-bound the last term as we did in the first part:

$$\|e_b' \mathbf{D}^{\frac{1}{2}} \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i\| = \|\sqrt{d_b} e_b' \sum_{i=2}^{n} \lambda_i^t \alpha_i v_i\| \leq \sqrt{d_b} \|\sum_{i=2}^{n} \lambda_i^t \alpha_i v_i\| \leq \lambda_2^t \sqrt{\frac{d_b}{d_a}} \tag{381}$$

□

5.5 Transit times

In addition to the speed of convergence to the steady state, we are interested in the speed of moving around in the graph. That is, if we begin our walk at a node a, after how long, in average, will we

reach node b? This time is known as the MFPT, the *Mean First Passage Time* (we are interested, of course, in the *first* time we reach b: if there is some important information there, then it is when we collected it); we indicate this quantity as $\langle T_{ab} \rangle$. We shall also be interested in its average value over the stationary distribution π:

$$\langle T \rangle = \sum_{a,b=1,b\neq a}^{n} \langle T_{ab} \rangle \pi_b \tag{382}$$

The value is independent of a, and is a characteristic of the structure of the graph.

In order to determine $\langle T_{ab} \rangle$ we begin rewriting the master equation (307). Let $F_{ab}(t)$ be the first passage probability, that is, the probability that, starting at a, we arrive at b *for the first time* at time t. Then:

$$P_{ab}(t) = \delta_{t,0}\delta_{a,b} + \sum_{\tau=0}^{t} P_{bb}(t-\tau) F_b b(\tau) \tag{383}$$

that is, either we start at b, or we arrive at b for the first time at τ (for all possible τ) then go from b back to b in time $t-\tau$. The summation is a convolution for positive indices between P and F so, taking the Laplace transform, we have

$$\tilde{P}_{ab}(s) = \delta_{a,b} + \tilde{P_{bb}}(s)\tilde{F_{ab}}(s) \tag{384}$$

from which we get

$$\tilde{F_{ab}}(s) = \frac{1}{\tilde{P}_{bb}(s)} \big(\tilde{P}_{ab}(s) - \delta_{ab} \big) \tag{385}$$

The average time $\langle T_{ab} \rangle$ is the average of F_{ab} and, for the properties of the Laplace transform:

$$\langle T_{ab} \rangle = \sum_{t=0}^{\infty} t F_{ab}(t) = -\tilde{F}'_{ab}(0) \tag{386}$$

Define the moments

$$R_{ab}^{(n)} = \sum_{t=0}^{\infty} t^n \big[P_{ab}(t) - \pi_b \big] \tag{387}$$

with this definition, we can find an expansion of $\tilde{P}_{ab}(s)$:

$$\begin{aligned}
\tilde{P}_{ab}(s) &\triangleq \sum_{t=0}^{\infty} e^{-st} P_{ab}(t) \\
&= \pi_b \sum_{t=0}^{\infty} e^{-st} + \sum_{t=0}^{\infty} e^{-st} \big[P_{ab}(t) - \pi_b\big] \\
&= \frac{\pi_b}{1-e^{-s}} + \sum_{t=0}^{\infty} \Big(\sum_{n=0}^{\infty} (-1)^n \frac{s^n t^n}{n!} \Big) \big[P_{ab}(t) - \pi_b\big] \quad (*) \\
&= \frac{\pi_b}{1-e^{-s}} + \sum_{n=0}^{\infty} (-1)^n \frac{s^n}{n!} \sum_{t=0}^{\infty} t^n \big[P_{ab}(t) - \pi_b\big] \\
&= \frac{\pi_b}{1-e^{-s}} + \sum_{n=0}^{\infty} (-1)^n \frac{s^n}{n!} R_{ab}^{(n)} \\
&\triangleq \frac{\pi_b}{1-e^{-s}} + Q_{ab}(s)
\end{aligned} \tag{388}$$

where, in (*) we have used the expansion

$$e^{-st} = \sum_{n=0}^{\infty} (-1)^n \frac{s^n t^n}{n!} \tag{389}$$

and we have defined $Q_{ab}(s)$ for convenience; note that $Q_{ab}(0) = R_{ab}^{(0)}$. From (385) we have

$$\begin{aligned}
\tilde{F}_{ab}(s) &= \frac{1}{\frac{\pi_b}{1-e^{-s}} + Q_{ab}(s)} \Big[\frac{\pi_b}{1-e^{-s}} + Q_{ab}(s) - \delta_{a,b} \Big] \\
&= \frac{1}{\pi_b + (1-e^{-s}) Q_{ab}(s)} \Big[\pi_b + (1-e^{-s})(Q_{ab}(s) - \delta_{a,b}) \Big]
\end{aligned} \tag{390}$$

Computing the derivative in 0, we have

$$\begin{aligned}
\tilde{F}'_{ab}(0) &= \frac{1}{\pi_b} \big[Q_{ab}(0) - \delta_{a,b} \big] + \frac{1}{\pi_b} Q_{bb}(0) \\
&= \frac{1}{\pi_b} \big[R_{ab}^{(0)} - R_{bb}^{(0)} - \delta_{ab} \big]
\end{aligned} \tag{391}$$

that is

$$\langle T_{ab} \rangle = \frac{1}{\pi_b} \big[R_{bb}^{(0)} - R_{ab}^{(0)} + \delta_{ab} \big] \tag{392}$$

Two of these terms are interesting:

$$\langle T_{aa} \rangle = \frac{1}{\pi_a} \tag{393}$$

is the mean return time: the average time between two consecutive visits to the node a, and

$$\tau_a = \frac{R_{aa}^{(0)}}{\pi_a} \tag{394}$$

is the average time necessary to reach a from a random node in the graph. Its inverse $C_a = \tau_a^{-1}$ is called the *centrality* of node a, and plays an important role in the analysis of information diffusion in graphs.

In order to compute these quantities, we need to compute $P_{ab}(t)$, namely, the probability that, starting at a, we are in b at time t. Considering (371) with $p_0 = e_a$, we have

$$P_a(t) = \pi + \mathbf{D}^{\frac{1}{2}} \sum_{k=2}^{n} \lambda_k^t \big(v_k' \mathbf{D}^{-\frac{1}{2}} e_a\big) v_k = \pi + \mathbf{D}^{\frac{1}{2}} \sum_{k=2}^{n} \lambda_k^t \frac{1}{\sqrt{d_a}} \big(v_k' e_a\big) v_k \tag{395}$$

that is,

$$\begin{aligned} P_{ab}(t) = e_b' P_a(t) &= \pi_b + e_b' \mathbf{D}^{\frac{1}{2}} \frac{1}{\sqrt{d_a}} \sum_{k=2}^{n} \lambda_k^t \big(v_k' e_a\big) v_k \\ &= \pi_b + \sqrt{\frac{d_b}{d_a}} \sum_{k=2}^{n} \lambda_k^t \big(v_k' e_a\big)(e_b' v_k) \\ &= \pi_b + \sqrt{\frac{d_b}{d_a}} \sum_{k=2}^{n} \lambda_k^t v_{k,a} v_{k,b} \end{aligned} \tag{396}$$

From this we have

$$\begin{aligned} R_{ab}^{(0)} &= \sum_{t=0}^{\infty} \Big[P_{ab}(t) - \pi_b \Big] \\ &= \sqrt{\frac{d_b}{d_a}} \sum_{k=2}^{n} \Big(\sum_{t=0}^{\infty} \lambda_k^t \Big) v_{k,a} v_{k,b} \\ &= \sqrt{\frac{d_b}{d_a}} \sum_{k=2}^{n} \frac{1}{1-\lambda_k} v_{k,a} v_{k,b} \end{aligned} \tag{397}$$

From (348) we have $v_{1,a} = \sqrt{\pi_a}$, that is

$$\pi_a = v_{1,a}^2 = (e_a' v_1)(v_1' e_a) \tag{398}$$

therefore

$$\tau_a = \frac{R_{aa}^{(0)}}{\pi_a} = \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \frac{(v_k' e_a)(e_a' v_k}{(v_1' e_a)(e_a' v_1} = \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \frac{v_{k,a}^2}{v_{1,a}^2} \tag{399}$$

$$\langle T_{ab} \rangle_{a \neq b} = \frac{2m}{d_b} \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(v_{k,b}^2 - \sqrt{\frac{d_b}{d_a}} v_{k,a} v_{k,b} \right) 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{v_{k,b}^2}{\sqrt{d_b}} - \frac{v_{k,a} v_{k,b}}{d_a d_b} \right) \tag{400}$$

and

$$\langle T \rangle = \sum_{m=0}^{n} R_{mm}^{(0)} = \sum_{m=0}^{n} \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \frac{(v_k' e_m)(e_m' v_k}{(v_1' e_m)(e_m' v_1} = \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \sum_{m=0}^{n} v_{k,m}^2 = \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \tag{401}$$