Lecture notes on Network Dynamics

Giacomo Como[1] and Fabio Fagnani[2]

June 1, 2021

[1]Department of Mathematical Sciences, Politecnico di Torino, Italy, and Department of Automatic Control, Lund University, Sweden, giacomo.como@polito.it.

[2]Department of Mathematical Sciences, Politecnico di Torino, Italy, fabio.fagnani@polito.it.

Contents

Preface

Chapter 1

Networks as graphs

The fundamental mathematical concept used to model the interconnection pattern of a network is that of a *graph*. This chapter provides a brief introduction to some basic notions and results from graph theory.

We start by emphasizing three key aspects in modeling networks that are captured by the concept of graph considered in these notes.

- We start with a set of nodes $\mathcal{V}$ representing the units that constitute the network (e.g., depending on the application, nodes may represent people, biological entities, computers, sensors, financial entities, etc.).

- Pairwise connections between units are reconded by a set $\mathcal{E}$ of links, whereby each link $e = (i, j)$ is an ordered pair of nodes i and j in $\mathcal{V}$. The presence of a particular link (i, j) in $\mathcal{E}$ may have different interpretations depending on the application: it may indicate that node i influences node j or, conversely, that i observes j, in the sense that i has access to the state of j and gets influenced by it. In other cases, where links are supposed to carry a flow of some physical or virtual commodity, the presence of the link (i, j) means that such commodity can actually flow directly from i to j.

- Sometimes, it proves useful to associate a positive scalar value W_{ij} to each link (i, j) in $\mathcal{E}$, to be referred to as the link's weight, with the aim of quantifying the 'strength' of the connection. Depending on the specific context, the link's weight may measure for instance the strength of a connection in terms of influence between two nodes, or, in the network flow applications, the conductance or the capacity of the link.

In certain applications, links have an intrinsic bidirectional meaning (e.g,, symmetric interaction, friendship, partnership). This corresponds to a situation where two links (i, j) and (j, i) are either both present with the same weight $W_{ij} = W_{ji} > 0$, or both absent (so that $W_{ij} = W_{ji} = 0$). Graphs with this feature will be referred to as *undirected*.

1.1 Weighted directed graphs

Following the previous discussion, we define a *(directed weighted) graph* as a triple

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$$

where:

- $\mathcal{V}$ is the countable (and typically —but not always— finite in these notes) set of *nodes* (also called *vertices*);
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of *links* (also called *edges*);
- $W \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ is the *weight matrix* and has the property that $W_{ij} > 0$ if and only if $(i, j) \in \mathcal{E}$, i.e., if (i, j) is a link.

Throughout, we shall denote by

$$n = |\mathcal{V}|$$

the *order* of the graph. We shall refer to links (i, i) whose head node coincides with its tail node as *self-loops*.

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is referred to as:

- *unweighted* if $W_{ij} \in \{0, 1\}$ for all nodes $i, j \in \mathcal{V}$. In this case the graph is often described by the pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. In this case, the matrix W can be unambigously deduced from the set $\mathcal{E}$ and is referred to as the *adjacency matrix* of $\mathcal{G}$;
- *undirected* if the weight matrix $W = W'$ is symmetric, i.e., when a link (i, j) exists if and only if the link with reversed direction (j, i) exists, and they have the same weight $W_{ij} = W_{ji}$.
- *simple* if it is undirected, unweighted, and the weight matrix W has zero diagonal, equivalently if $\mathcal{G}$ contains no self-loops.

For undirected graphs, one can model undirected links that are not self-loops as unordered pairs $\{i, j\}$, each of which corresponds to the pair of directed links (i, j) and (j, i). Note that, by doing so, to every undirected link there correspond two directed links. When referring generically to a graph, we will implicitly intend it to be weighted and directed, unless it is otherwise specified or clear from the context.

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, we introduce the following notion.

- The *out-neighborhood* and the *in-neighborhood* of a node i in $\mathcal{V}$ are, respectively, the sets

  $$\mathcal{N}_i = \{j \in \mathcal{V} \,|\, (i, j) \in \mathcal{E}\}, \quad \mathcal{N}_i^- = \{j \in \mathcal{V} \,|\, (j, i) \in \mathcal{E}\}$$

  Nodes in $\mathcal{N}_i$ and $\mathcal{N}_i^-$ are referred to, respectively, as *out-neighbors* and *in-neighbors* of node i in $\mathcal{G}$.

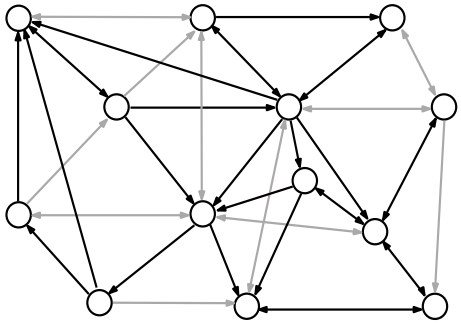


Figure 1.1: A directed weighted graph where links of weight 1 are drawn in grey and links of weight 2 are drawn in black.

- Nodes with no out-neighbors other than possibly themselves are called *sinks*, while nodes with no in-neighbors other than possibly themselves are called *sources*. E.g., the graph in Figure 1.6(a) contains no sources or sinks, while the one in Figure 1.6(b) contains two sinks, node s_0 and node s_1, and no sources.

- The *out-degree* and *in-degree* of a node i are defined, respectively, as

$$w_i = \sum_{j \in \mathcal{V}} W_{ij}\,, \qquad \text{and} \qquad w_i^- = \sum_{j \in \mathcal{V}} W_{ji}\,.$$

  Often, we will use the shorter term degree for out-degree.

- A node i is called *balanced* if $w_i = w_i^-$.

In undirected graphs there is no distinction between out- and in-neighbors, out- and in-neighborhoods, and out- and in-degree. In this case we simply use the terms neighbors, neighborhoods, degrees. Also notice that, in undirected graphs, sinks and sources are simply isolated nodes, and all nodes are balanced.

For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, we use the compact notation

$$w = W\mathbb{1}\,, \qquad w^- = W'\mathbb{1}$$

for the out-degree and in-degree vectors. The term

$$\mathbb{1}'W\mathbb{1} = \sum_{i,j \in \mathcal{V}} W_{ij}$$

represents the *total degree* of the graph while

$$\overline{w} = \frac{1}{n}\mathbb{1}'W\mathbb{1} = \frac{1}{n}\sum_{i,j \in \mathcal{V}} W_{ij}\,,$$

is the *average degree*.

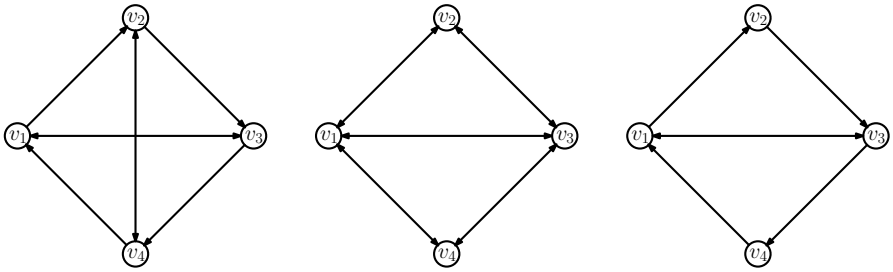


Figure 1.2: Three examples of balanced graphs with four nodes. The graph in (a) is regular but not undirected. The graph in (b) is undirected but not regular. The graph in (c) is neither regular nor undirected.

For simple graphs the *total degree* is necessarily an even number, equal to twice the number of undirected links: this follows from the fact that in the summation

$$\mathbb{1}'W\mathbb{1} = \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}} W_{ij} = \sum_{(i,j)\in\mathcal{E}} 1 = |\mathcal{E}|$$

every undirected link $\{i,j\}$ is counted twice, once as (i,j) and once as (j,i). This property is sometimes referred to as the *hand-shaking lemma*.

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is called

- *balanced* if all its nodes are so, i.e., if $w = w^-$;
- *regular* if all its nodes have the same degree, i.e., if $w = w^- = \overline{w}\mathbb{1}$.

Observe that a regular graph is always balanced but not necessarily undirected (see See Figure 1.2(a).) On the other hand, undirected graphs are always balanced but not necessarily regular (See See Figure 1.2(b)). In fact, it is not hard to come up with examples of balanced graphs that are neither regular nor undirected (see See Figure 1.2(c)). The relative implications between the notions of balanced, regular, and undirected graph are illustrated in Figure 1.3.

Example 1.1. Three unweigthed graphs are depicted in Figure 1.2. Their adjacency matrix W, in-degree vector w^-, and out-degree vector w are respectively given by

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \qquad w^- = W'\mathbb{1} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}, \qquad w = W\mathbb{1} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix},$$

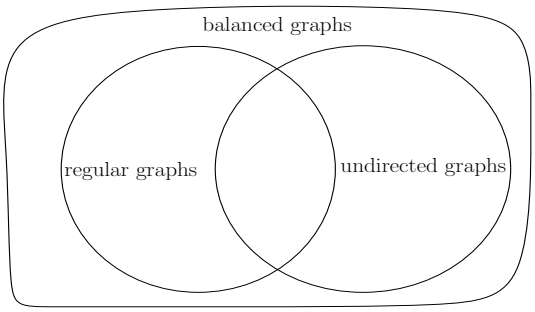


Figure 1.3: Representation of the relative inclusions among the sets of balanced, regular, and undirected graphs.

for the graph in Figure 1.2 (a),

$$W = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \qquad w^- = W'\mathbb{1} = \begin{bmatrix} 3 \\ 2 \\ 3 \\ 2 \end{bmatrix}, \qquad w = W\mathbb{1} = \begin{bmatrix} 3 \\ 2 \\ 3 \\ 2 \end{bmatrix}.$$

for the graph in Figure 1.2 (b), and

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad w^- = W'\mathbb{1} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \qquad w = W\mathbb{1} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 1 \end{bmatrix}.$$

for the graph in Figure 1.2 (c). Observe that all three graphs are balanced since $w = w^-$. In fact, the graph in Figure 1.2 (a) is regular (since $w = w^- = 2\mathbb{1}$) but not undirected (since $W \neq W'$), the graph in Figure 1.2 (b) is undirected ($W = W'$) but not regular (since, e.g., $w_1 = 2 \neq 3 = w_2$), while the graph in Figure 1.2 (c) is neither regular nor undirected.

1.2 Subgraphs

In this section we introduce the notions of subgraph and discuss several notable special subgraphs.

Before proceeding, we introduce the notion of graph isomorphism that proves useful in many applications. Two graphs are referred to as isomorphic if they only differ for a relabeling of the node set. More formally, two graphs $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1, W^{(1)})$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2, W^{(2)})$ are referred to as *isomorphic* if there exists a bijection (i.e., a one-to-one correspondence) $f : \mathcal{V}_1 \to \mathcal{V}_2$ such that: (a) (i, j) in $\mathcal{E}_1$ is a link for $\mathcal{G}_1$ if and only if $(f(i), f(j))$ in $\mathcal{E}_2$ is a link for $\mathcal{G}_2$; and (b) $W^{(1)}_{ij} = W^{(2)}_{f(i)f(j)}$ for every two nodes i and j in $\mathcal{V}_1$. In the these

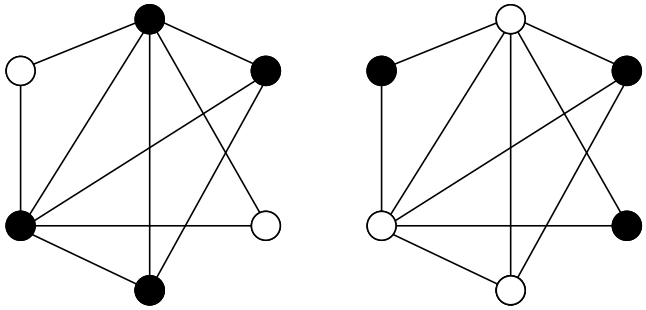


Figure 1.4: In (a) a maximal clique and in (b) a maximal independent set in a simple graph with 6 nodes.

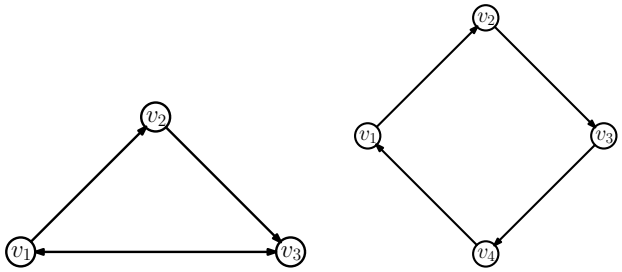
notes, isomorphic graphs will often be identified. Notice that, as consequence, whenever convenient this allows one to identify the node set $\mathcal{V}$ with the set of the first n positive integers $\{1, 2, \ldots, n\}$: in that case the weight matrix W belongs to $\mathbb{R}_+^{n\times n}$.

A *subgraph* of $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is any graph $\mathcal{H} = (\mathcal{U}, \mathcal{F}, Z)$ with node set $\mathcal{U} \subseteq \mathcal{V}$, link set $\mathcal{F} \subseteq \mathcal{E}$, and link weights $Z_{ij} \leq W_{ij}$ for every $i, j \in \mathcal{U}$. There are two types of subgraphs that play an important role in the theory: *induced* subgraphs and *spanning* subgraphs.

Induced graphs are obtained by selecting a node subset $\mathcal{U} \subseteq \mathcal{V}$ in the original graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ and removing from $\mathcal{E}$ all links that stem from or point towards nodes in the complementary subset $\mathcal{V} \setminus \mathcal{U}$. Formally, given $\mathcal{U} \subseteq \mathcal{V}$, the subgraph of $\mathcal{G}$ induced by $\mathcal{U}$ is given by $\mathcal{G}|_{\mathcal{U}} = (\mathcal{U}, \mathcal{F}, W_{|\mathcal{U}\times\mathcal{U}})$, where $\mathcal{F} = \{(i, j) \in \mathcal{E} : i, j \in \mathcal{U}\}$. If $\mathcal{G}|_{\mathcal{U}}$ is the complete graph (possibly plus self-loops), i.e., if (i, j) is a link in $\mathcal{E}$ for every two distinct nodes $i \neq j$ in $\mathcal{U}$, the set $\mathcal{U}$ is called a *clique*. If $\mathcal{G}|_{\mathcal{U}}$ only consists of isolated nodes (i.e., if $\mathcal{F} = \emptyset$), the set $\mathcal{U}$ is called an *independent set*. A *maximal* clique (independent set) is one that is not a subset of any larger clique (respectively, independent set). See Figure 1.4.

A spanning subgraph of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is obtained by removing some links from $\mathcal{G}$ while maintaining the same node set, i.e., a spanning subgraph of $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a graph $\mathcal{H} = (\mathcal{V}, \mathcal{F}, Z)$, where $\mathcal{F} \subseteq \mathcal{E}$ and where $Z_{ij} = W_{ij}$ for all $(i, j) \in \mathcal{F}$ and $Z_{ij} = 0$ for every $(i, j) \notin \mathcal{F}$. (See Figure 1.5.)

A *matching* or *independent link set* in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a subset of links $\mathcal{F} \subseteq \mathcal{E}$ such that every node is either the tail or the head node of at most one link in $\mathcal{M}$, i.e., if no two links in $\mathcal{M}$ share a common (tail or head) node. For a matching $\mathcal{M}$ in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, a node i in $\mathcal{V}$ is referred to as *matched* if it is the tail or head node of at least one link in $\mathcal{F}$ and *unmatched* if it is neither the tail nor the head node of any link in $\mathcal{M}$: equivalently, a node is matched if the sum of its in- and out-degree in the spanning subgraph $\mathcal{H} = (\mathcal{V}, \mathcal{M}, Z)$ is positive, while it is unmatched if such sum is 0.



(a) (b)

Figure 1.5: Two subgraphs of the graph $\mathcal{G}$ in Figure 1.2(a). In (a) the subgraph of $\mathcal{G}$ induced by $\mathcal{U} = \{v_1, v_2, v_3\}$. In (b) a spanning subgraph of $\mathcal{G}$.

A matching is *maximal* if every subset of links $\mathcal{F} \subseteq \mathcal{E}$ strictly containing $\mathcal{M}$ (i.e., such that $\mathcal{M} \subsetneq \mathcal{F}$) is not a matching. A *maximum-cardinality* (*maximum-weight*) matching in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a matching such that there exists no matching in $\mathcal{G}$ with larger cardinality (total link weight). A matching is *perfect* (sometimes also called *complete*) if all nodes are matched. A perfect matching is also a minimum-size link cover. Clearly, a graph can only contain a perfect matching when the graph has an even number of nodes. A *near-perfect* matching is one in which exactly one node is unmatched and all the others are matched. Clearly, a graph can only contain a near-perfect matching if it has an odd number of nodes, and avery near-perfect matching is a maximum-cardinality matching.

1.3 Reachability, connectedness, periodicity

In this section, we introduce the notion of reachability between nodes in a graph. We then show how it can be used to introduce an equivalence relation among the nodes of a graph, whose equivalence classes are known as the connect components of the graph. Finally, we introduce the notion of period of a node and of a graph.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Then:

- A *walk* from node i to node j is a finite string of nodes $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l)$ such that $\gamma_0 = i$, $\gamma_l = j$, and (γ_{h-1}, γ_h) belongs to $\mathcal{E}$ for all $h = 1, \ldots, l$, i.e., there is a link between every two consecutive nodes. Here, l is called the *length* of the walk. By convention, we consider walks of length 0 as going from a node to itself;

- Given two walks $\gamma^{(1)} = (\gamma_0^{(1)}, \gamma_1^{(1)}, \ldots, \gamma_{l_1}^{(1)})$ from $i = \gamma_0^{(1)}$ to $j = \gamma_{l_1}^{(1)}$ and $\gamma^{(2)} = (\gamma_0^{(2)}, \gamma_1^{(2)}, \ldots, \gamma_{l_2}^{(2)})$ from $j = \gamma_0^{(2)}$ to $k = \gamma_{l_2}^{(2)}$, the *concatenation* of $\gamma^{(1)}$ and $\gamma^{(2)}$ is the length-$(l_1 + l_2)$ walk $\gamma = \gamma^{(1)} \circ \gamma^{(2)}$ from $\gamma_0 = i$ to

$\gamma_{l_1+l_2} = k$ given by

$$\gamma = \gamma^{(1)} \circ \gamma^{(2)} = \left(\gamma_0^{(1)}, \gamma_1^{(1)}, \ldots, \gamma_{l_1}^{(1)} = \gamma_0^{(2)}, \gamma_1^{(2)}, \ldots, \gamma_{l_2}^{(2)}\right) ;$$

- A node j is said to be *reachable* from a node i if there exists a walk from i to j;
- A walk $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l)$ such that $\gamma_h \neq \gamma_k$ for all $0 \leq h < k \leq l$, except for possibly $\gamma_0 = \gamma_l$, is called a *path*. I.e., a path is a walk that does not pass through a previously visited node except possibly for ending in the node where it starts from;
- A walk or path $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l)$ starting and ending in the same node $\gamma_0 = \gamma_l$ is referred to as *closed*;
- A closed walk of length $l \geq 1$ is called a *circuit*;
- A closed path of length $l \geq 3$ is called a *cycle*. Equivalently, a cycle is a circuit of length at least 3 visiting distinct nodes except for the initial and the final ones that coincide. Notice that a self-loop (i, i) in $\mathcal{E}$ is a circuit but not a cycle since it has length 1. Similarly, if there exists two nodes $i \neq j$ in $\mathcal{V}$ such that both (i, j) and (j, i) belong to the link set $\mathcal{E}$, then there exists a length-2 circuit (i, j, i), which however is not a cycle;
- If $\mathcal{G}$ does not contain any circuit it is referred to as *circuit-free*, if it does not contain any cycle it is referred to as *acyclic*.[1]

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, a natural notion of *distance* can be introduced: for two nodes i and j in $\mathcal{V}$, the distance $\mathrm{dist}(i, j)$ of j from i is the length of the shortest path from i to j in $\mathcal{G}$, with the convention that $\mathrm{dist}(i, j) = +\infty$ if no such path exists. A path from i to j of minimal length $\mathrm{dist}(i, j)$ is called a *geodesic path*. The *diameter* of $\mathcal{G}$

$$\mathrm{diam}(\mathcal{G}) := \max_{i,j \in \mathcal{V}} \mathrm{dist}(i, j) \tag{1.1}$$

is the maximum distance between any two nodes in $\mathcal{G}$.

A graph $\mathcal{G}$ is called *strongly connected* if given any two nodes i and j, we have that i is reachable from j. Equivalently, $\mathcal{G}$ is strongly connected if and only if its diameter $\mathrm{diam}(\mathcal{G})$ is finite.

In many applications it is useful to have notions of connectedness with respect to a subset of nodes. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ and node subset $\mathcal{U} \subseteq \mathcal{V}$, we say that $\mathcal{U}$ is

[1] Please, note that some of the literature refers to directed circuit-free graphs as "directed acyclic graphs" and uses the acronym DAG for such graphs. We do not follow this convention as it would require to call cycles also length-2 circuits and self-loops, in particular with the result that no nontrivial undirected graph could be considered acyclic. In contrast, with our convention, there clearly are non-trivial acyclic undirected graphs, as an undirected link by itself does not give rise to a cycle.

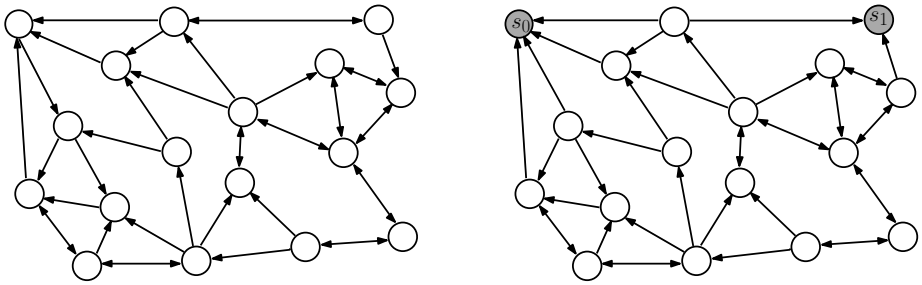


Figure 1.6: Two directed unweighted graphs. The graph in (a) is strongly connected. The graph in (b) is not strongly connected and has in fact two sinks: s_0 and s_1. Note that $\mathcal{S} = \{s_0, s_1\}$ is globally reachable in the graph in (b).

- *trapping* (in $\mathcal{G}$) if for every node i in $\mathcal{U}$ and every walk in $\mathcal{G}$ from i to some j, we have that j necessarily belongs to $\mathcal{U}$
- *globally reachable* (in $\mathcal{G}$) if for every node i in $\mathcal{V} \setminus \mathcal{U}$ there exists a walk from i to some node j in $\mathcal{U}$. Notice that $\mathcal{G}$ is strongly connected if and only if every subset of nodes $\mathcal{U} \subseteq \mathcal{V}$ is globally reachable.

The analysis of the connectedness of a graph can be further refined by considering the so called *connected components* of $\mathcal{G}$ that are the maximal subsets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k$ of the node set $\mathcal{V}$ such that, for every two nodes i and j in the same component $\mathcal{V}_h$, there exists a path from i to j. Notice that the order of a connected component may range from 1 (in case there exists a node i such that there exists no other node $j \neq i$ such that both j is reachable from i and vice versa) to n (when the graph is strongly connected). The splitting in connected components constitutes a *partition* of the node set $\mathcal{V}$, i.e., one has that

$$\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \ldots \cup \mathcal{V}_k\,, \qquad \mathcal{V}_h \cap \mathcal{V}_l = \emptyset\,, \quad h \neq l\,.$$

If the graph is undirected, then there are no links between nodes belonging to different connected components. In contrast, in the general case, there can be links pointing from some nodes in one connected component towards nodes in another connected component. This can be usefully captured by introducing the *condensation graph* $\mathcal{H}_\mathcal{G}$ obtained from $\mathcal{G}$ by collapsing nodes in every connected component of $\mathcal{G}$ into single 'supernodes' of $\mathcal{H}_\mathcal{G}$ and adding a link in $\mathcal{H}_\mathcal{G}$ from a supernode to another one whenever there is a link in the original graph pointing from some node in the connected component corresponding to the first supernode to some node in the connected component corresponding to the second supernode. (See Figure 1.7.) Many interesting properties can easily be read out from the structure of $\mathcal{H}_\mathcal{G}$.

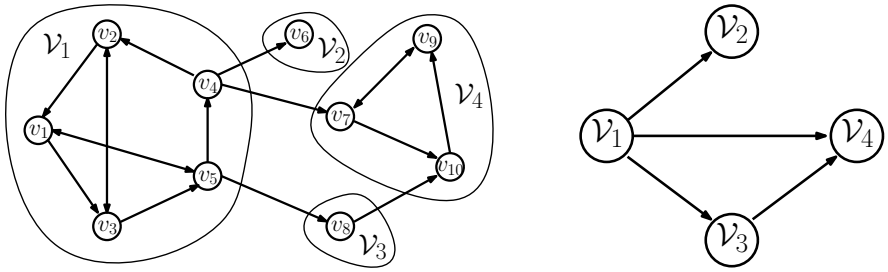


Figure 1.7: (a) A directed graph $\mathcal{G}$ with $n = 10$ nodes and $c_{\mathcal{G}} = 4$ connected components: $\mathcal{V}_1 = \{v_1, v_2, v_3, v_4, v_5\}$, $\mathcal{V}_2 = \{v_6\}$, $\mathcal{V}_3 = \{v_7, v_8, v_9\}$, $\mathcal{V}_4 = \{v_{10}\}$. (b) The condensation graph $\mathcal{H}_{\mathcal{G}}$ has $c_{\mathcal{G}} = 4$ nodes, of which $s_{\mathcal{G}} = 2$ are sinks, namely $\mathcal{V}_2$ and $\mathcal{V}_4$. Observe that, e.g., the node subset $\mathcal{U} = \{v_6, v_7\}$ has nonempty intersection with both $\mathcal{V}_2$ and $\mathcal{V}_4$, hence it is globally reachable.

Remark 1. *The condensation graph $\mathcal{H}_{\mathcal{G}}$ satisfies the following properties:*

- *$\mathcal{H}_{\mathcal{G}}$ is a circuit-free graph;*
- *$\mathcal{G}$ is strongly connected if and only if $\mathcal{H}_{\mathcal{G}}$ consists of a single node;*
- *a subset of nodes $\mathcal{U} \subseteq \mathcal{V}$ is globally reachable in $\mathcal{G}$ if and only if it has nonempty intersection $\mathcal{U} \cap \mathcal{V}_i \neq \emptyset$, $i = 1, \ldots, k$, with each connected component $\mathcal{V}_i$ that corresponds to a sink in the condensation graph $\mathcal{H}_{\mathcal{G}}$;*
- *$\mathcal{G}$ contains a globally reachable node if and only if $\mathcal{H}_{\mathcal{G}}$ has just one sink.*

Proving the statements above is left as an exercise to the reader.

Finally, another important concept related to connectedness of a graph is that of period. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, the *period* $\mathrm{per}_{\mathcal{G}}(i)$ of a node i in $\mathcal{V}$ is defined as the greatest common divisor of the lengths of all circuits starting and ending in i, with the convention that $\mathrm{per}_{\mathcal{G}}(i) = 1$ if there are no circuits starting and ending in i.

Proposition 1.1. Assume that $\mathcal{G}$ is strongly connected. Then, $\mathrm{per}_{\mathcal{G}}(i) = \mathrm{per}_{\mathcal{G}}(j)$ for every two nodes i and j in $\mathcal{V}$.

Proof. Assume that $|\mathcal{V}| \geq 2$ otherwise there is nothing to prove. Given any two nodes $i \neq j$, let γ_{ij} be a path of length l_{ij} connecting i to j. If γ is any length-l closed path starting and ending in node i, one can consider the concatenations $\gamma_{ji} \circ \gamma \circ \gamma_{ij}$ and $\gamma_{ji} \circ \gamma_{ij}$ which lead to two closed paths starting and ending in node j of length, respectively, $l_{ij} + l_{ji} + l$ and $l_{ij} + l_{ji}$. Therefore, $\mathrm{per}_{\mathcal{G}}(j)$ divides both and, thus, it divides also l. Hence, $\mathrm{per}_{\mathcal{G}}(j)$ divides $\mathrm{per}_{\mathcal{G}}(i)$. Considering that i and j have been chosen arbitrarily, this proves the claim □

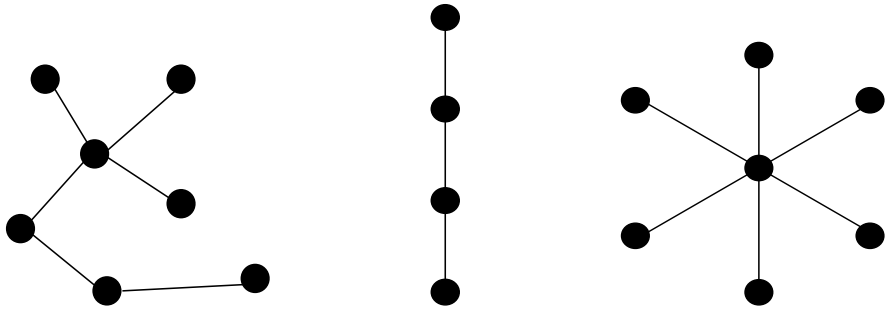


Figure 1.8: Three trees: the one in (b) is a line graph, the one in (c) is a star graph.

The common period of all nodes in a strongly connected graph $\mathcal{G}$ is denoted by $\text{per}_{\mathcal{G}}$ and referred to as the period of $\mathcal{G}$. A strongly connected graph $\mathcal{G}$ is called *aperiodic* if $\text{per}_{\mathcal{G}} = 1$.

Example 1.2. The three graphs in Figure 1.2 are all strongly connected and aperiodic. The graphs in Figure 1.5 are also both strongly connected, however, while the one in Figure 1.5 (a) is also aperiodic —as it contains both a cycle of length 3 and a length-2 circuit, i.e., a bidirectional link— the graph in Figure 1.5 (b) has period 4 hence it is not aperiodic.

1.4 Examples

The following are basic examples of (families of) graphs which frequently recur in the theory and in the applications:

- A *tree* is a simple connected acyclic graph.

  Examples of trees are (see Figure1.8):

  - the *line* graph $L_n = (\{1, \dots, n\}, \mathcal{E})$ where $\mathcal{E} = \{(i, i+1), (i+1, i),\ i = 1, \dots, n-1\}$;
  - the *star* graph $S_n = (\{0, 1, \dots, n\}, \mathcal{E})$ where $\mathcal{E} = \{(0, i), (i, 0),\ i = 1, \dots, n\}$.

Trees share many interesting properties. Some of them are gathered in the result below.

Proposition 1.2. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a simple connected graph with $n = |\mathcal{V}|$ nodes and $m = |\mathcal{E}|/2$ undirected links. Then,

1. $m \geq n - 1$;

2. $\mathcal{G}$ is a tree if and only if $m = n - 1$;

3. If $\mathcal{G}$ is a tree, then there exist at least two and, if $n \geq 3$, at most $n - 1$ nodes of degree one (*leaves*);

Proof. 1. We prove it by induction on n. If $n = 1$, we necessarily have $m = 0$ and thus $m \geq n - 1$ holds true. Suppose now that the statement holds true for any simple connected graph with number of nodes $n < \bar{n}$ where $\bar{n} \geq 2$. Assume that $\mathcal{G}$ is a simple connected graph with $\bar{n}$ nodes. If $\mathcal{G}$ is a tree, we consider any of its edges (i, j) and we remove it together with (j, i). The spanning subgraph thus obtained is composed of two connected components (verify this point) one containing the node i and one containing the node j with, respectively, n_1 and n_2 nodes and m_1 and m_2 undirected edges. By induction, we have that $m_k \geq n_k - 1$ for $k = 1, 2$. Since $m = m_1 + m_2 + 1$ and $n = n_1 + n_2$, summing the two inequalities we immediately obtain that $m \geq n - 1$. If $\mathcal{G}$ is not a tree, we first remove a number of edges till we get a spanning tree with $\tilde{m} < m$ edges and we then proceed as above. Again, we obtain $m > \tilde{m} \geq n - 1$.

2. The 'only if' part can be proven along the same lines as point 1 and we leave it to the reader. As for the 'If' part: If $\mathcal{G}$ is not a tree, then it must have a cycle. Removing an undirected link from this cycle we obtain a subgraph that is still connected but with number of undirected links $m - 1 = n - 2$ and this contradicts point 1.

3. is left as an exercise. □

We now consider other notable examples of simple graphs.

- The *complete* graph K_n is a simple graph consisting of n nodes each connected to every other node.

- The *barbell* graph is obtained by connecting two independent complete graphs with a link between one node in each side;

- The *cycle* graph C_n is a simple connected graph with n nodes all of which have degree 2.

A useful operation to generate new families of graphs is the product.

Given two simple graphs $\mathcal{G}^i = (\mathcal{V}^i, \mathcal{E}^i, W^i)$ for $i = 1, 2$, the *product graph* is formally defined as $\mathcal{G}^1 \times \mathcal{G}^2 = (\mathcal{V}^1 \times \mathcal{V}^2, \mathcal{E}^1 \otimes \mathcal{E}^2, W^1 \otimes W^2)$ where

$$((i^1, i^2), (j^1, j^2)) \in \mathcal{E}^1 \otimes \mathcal{E}^2 \;\Leftrightarrow\; \begin{cases} i^1 = j^1, & (i^2, j^2) \in \mathcal{E}^2 \\ i^2 = j^2, & (i^1, i^1) \in \mathcal{E}^1 \end{cases}$$

and $W^1 \otimes W^2$ is the corresponding adjacency matrix.

Proposition 1.3. Let $\mathcal{G}^i = (\mathcal{V}^i, \mathcal{E}^i, W^i)$ for $i = 1, 2$ be two simple graphs. The following properties hold:

- If $\mathcal{G}^1$ and $\mathcal{G}^2$ are strongly connected, $\mathcal{G}^1 \times \mathcal{G}^2$ is strongly connected;

- $\mathrm{diam}(\mathcal{G}^1 \times \mathcal{G}^2) = \mathrm{diam}(\mathcal{G}^1) + \mathrm{diam}(\mathcal{G}^2)$.

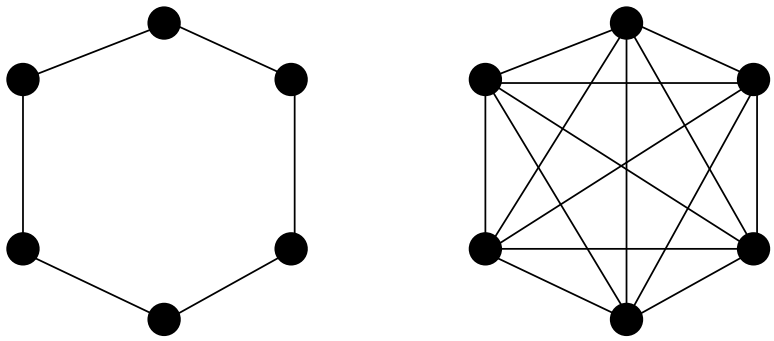


Figure 1.9: In (a) a cycle graph. In (b) a complete graph.

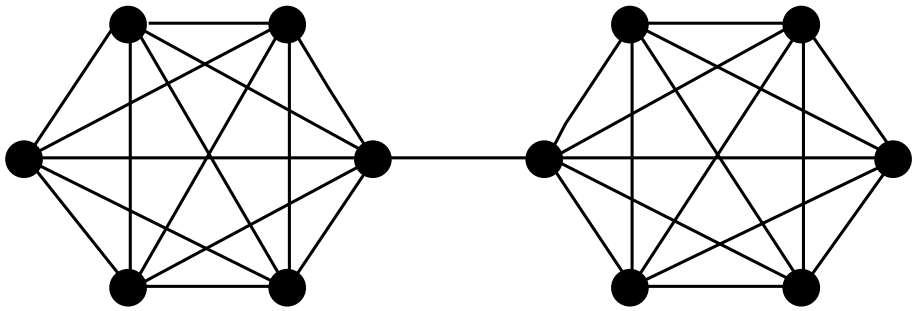


Figure 1.10: A barbell graph.

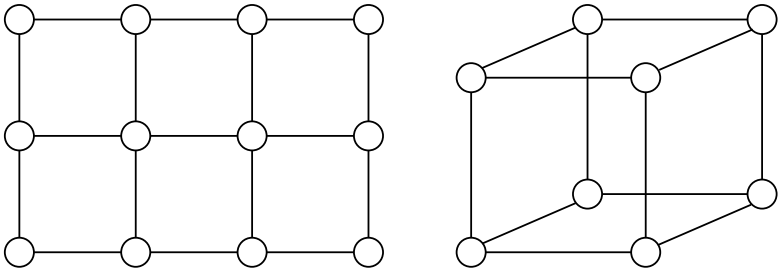


Figure 1.11: In (a) a grid graph. In (b) a cube graph

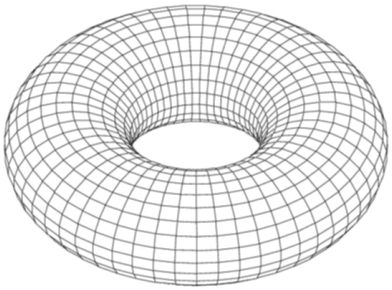


Figure 1.12: A toroidal graph

Given two line graphs L_h and L_k, their product $L_h \times L_k$ is the *grid* graph with $n = h \cdot k$ nodes. A product of three lines of length 2 give rise to the *cube* graph $L_2 \times L_2 \times L_2$ Higher product of k lines of length 2, $L_2^k = L_2 \times L_2 \times \cdots L_2$ is called *hypercube*. The product of the cycles $C_h \times C_k$ is called a *toroidal grid*

1.5 Bipartite graphs and graph colorings

A simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is called *bipartite* if we can split the node set into two nonempty subsets so that there are no links between nodes in the same subset. In other words, $\mathcal{G}$ is bipartite if there exists a partition $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$ such that $W_{ij} = 0$ for all $i, j \in \mathcal{V}_0$ and $W_{ij} = 0$ for all $i, j \in \mathcal{V}_1$. Notice that the bipartite property can be equivalently expressed saying that the two set of nodes $\mathcal{V}_0$ and $\mathcal{V}_1$ are both independent.

Example 1.3. The complete bipartite graph $K_{n,m}$ is a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ where $\mathcal{V}$ is the union of two disjoint sets of nodes $\mathcal{V}_0 = \{i_1, \ldots, i_n\}$ and $\mathcal{V}_1 = \{j_1, \ldots, j_m\}$ and where

$$\mathcal{E} = \{(i_h, j_k), (j_k, i_h),\ h = 1, \ldots n,\ k = 1, \ldots, m\}$$

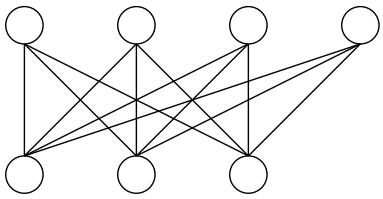


Figure 1.13: The graph $K_{4,3}$

Proposition 1.4. A simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is bipartite if and only if every circuit in $\mathcal{G}$ has even length.

Proof. Observe first that we can assume that $\mathcal{G}$ is connected, as, otherwise, it is sufficient to prove the result on each single connected component.

Suppose that $\mathcal{G}$ is bipartite and let $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$ be the corresponding partition of the node set. Let $\gamma = (\gamma_0, \ldots, \gamma_{l-1}, \gamma_l = \gamma_0)$ be a circuit in $\mathcal{G}$ and assume, without loss of generality that $\gamma_0 \in \mathcal{V}_0$. Then, necessarily, $\gamma_i \in \mathcal{V}_0$ if i is even, while $\gamma_i \in \mathcal{V}_1$ if i is odd. Consequently, l must be even.

Suppose conversely that every circuit in $\mathcal{G}$ has even length. Fix a node $i_0 \in \mathcal{V}$ and let $\mathcal{V}_0$ be the subset of nodes containing i_0 and all nodes that can be reached from i_0 with a path of even length. We now define $\mathcal{V}_1$ as the subset of nodes that can be reached from i_0 with a path of odd length. Since $\mathcal{G}$ is connected we have that $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$. Notice moreover that $\mathcal{V}_0 \cap \mathcal{V}_1 = \emptyset$ otherwise there would exist points reachable from i_0 with a path of even length and another path of odd length and a concatenation of these two would yield a circuit of odd length. We claim that $\mathcal{G}$ is bipartite with respect to such partition. Indeed, if for two nodes $i_1, i_2 \in \mathcal{V}_s$ for $s = 1, 2$ it holds that $(i_1, i_2) \in \mathcal{E}$, then, using the definition of $\mathcal{V}_s$, we would easily construct, by concatenation a circuit of odd length containing the edge (i_1, i_2) that is absurd. Proof is then complete. □

Example 1.4. The following simple graphs are bipartite (verification is left to the reader):

- Trees
- Cycles of even length
- Grids

Given a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ and a set of colors Ω a function $\psi : \mathcal{V} \to \Omega$ is called a *coloring* of $\mathcal{G}$ if neighbor nodes have always different colors, formally if

$$(i, j) \in \mathcal{E} \;\Rightarrow\; \psi(i) \neq \psi(j)$$

The graph $\mathcal{G}$ is called $|k|$-*colorable* if there exists a coloring of $\mathcal{G}$ with a set of colors Ω such that $k = |\Omega|$. The *chromatic number* of $\mathcal{G}$ is the smallest k for which $\mathcal{G}$ is k-colorable.

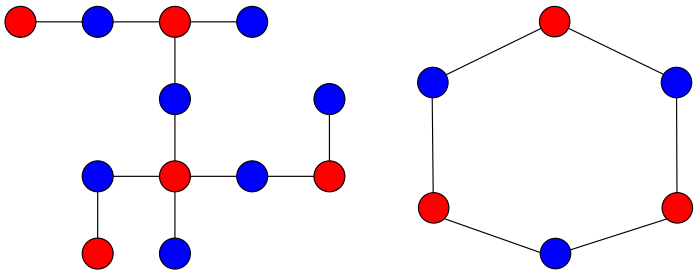


Figure 1.14: 2-coloring for two bipartite graphs.

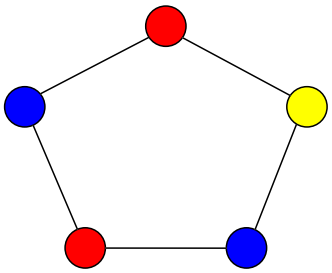


Figure 1.15: Odd length cycles have chromatic number 3.

Notice that a graph is 2-colorable if and only if it is bipartite. Bipartite graphs are thus exactly the graphs having chromatic number equal to 2 (unless the graph has no edge in which case the chromatic number is 1).

1.6 Common graph statistics

We conclude this section by introducing some *statistics* commonly computed for graphs. Two such statistics are the graph diameter defined in (1.1) and the average distance

$$\overline{\mathrm{dist}}(\mathcal{G}) := \frac{1}{n^2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}} \mathrm{dist}(i,j)\,. \tag{1.2}$$

Other widely used graph statistics are the *degree distributions*. Consider an unweighted graph $\mathcal{G}$ (or more in general, a weighted graph where weights are integers), and, for every nonnegative integers h and k, let

$$p_{hk} := \frac{1}{n} |\{i \in \mathcal{V} :\, w_i^- = h,\, w_i = k\}|\,, \qquad p_k := \sum_{l \geq 0} p_{lk}\,, \qquad p_h^- := \sum_{l \geq 0} p_{hl}\,, \tag{1.3}$$

denote the fractions of nodes with, respectively, in-degree h and out-degree k, out-degree k, and in-degree h. The sequences $\{p_k\}_{k \geq 0}$ and $\{p_h^-\}_{h \geq 0}$ are known

as the out- and in-degree distributions, while $\{p_{hk}\}_{h,k\geq 0}$ is the joint in- and out-degree distribution. In balanced graphs —hence in particular in undirected graphs—

$$p_k = p_{kk} = p_k^- \qquad p_{hk} = 0\,, \qquad h \neq k\,,$$

so that there is no real distinction between out-, in-, and joint in- and out-degree distributions. In contrast, such a difference is relevant for general, unbalanced graphs. Observe that the average degree can be computed from the degree distribution as

$$\overline{w} = \frac{1}{n}\mathbb{1}'W\mathbb{1} = \sum_{k\geq 0} p_k k = \sum_{k\geq 0} p_k^- k\,.$$

Hence, even for unbalanced graphs, it does not matter whether the in- or the out-degree distributions are used in the computation of $\overline{w}$. Higher moments —such as the average square degree— can also be computed from the degree distributions, however, unless the graph is balanced it does matter whether the in- or out-degree distributions are used.

Remark 2. *There are other statistics that can be associated with a graph and are relevant in some contexts. E.g., in social networks, an important property is* transitivity*: a friend of my friend is more likely to be my friend than a random person. In a simple graph, transitivity can be measured by means of the* clustering coefficient *defined as*

$$\mathrm{Clust}(\mathcal{G}) = \frac{6\cdot(\#\textit{triangles})}{\#\textit{length-2paths with distinct start and end nodes}}$$

Another important property is homophily*, or* assortative mixing*. This is the property that nodes belonging to the same class tend to be more connected among each other than nodes of different classes. To measure this, assume that the node set is partitioned in k classes $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \ldots \cup \mathcal{V}_k$, and let c_i be the class node i belongs to. Then, one defines the* modularity *of an undirected graph as*

$$mod(\mathcal{G}) = \frac{1}{n\overline{w}} \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}} \left(W_{ij} - \frac{w_i w_j}{n\overline{w}}\right) \delta(c_i, c_j)\,,$$

where $\delta(h,l) = 1$ if $h = l$ and $\delta(h,l) = 0$ if $h \neq l$.

1.7 Problems

1. Prove that a simple connected graph with only nodes of degree one or two is either a line or a ring cycle graph.

2. Prove the statements in Remark 1.

3. Prove the two statements in Proposition 1.3.

4. A *Eulerian* circuit in a simple connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a circuit $\gamma = (\gamma_0, \ldots, \gamma_l)$ that touches all nodes and never uses edges twice, formally

$$\{\gamma_0, \ldots, \gamma_{l-1}\} = \mathcal{V},\ (\gamma_h, \gamma_{h+1}) \neq (\gamma_k, \gamma_{k+1}), (\gamma_{k+1}, \gamma_k) \text{ for } h \neq k$$

   Prove that $\mathcal{G}$ admits a Eulerian circuit if and only if the degree of each node is an even number.

5. Prove that trees are bipartite.

6. Prove that the product of bipartite graphs is bipartite. Deduce that grid graphs are always bipartite.

7. Compute the chromatic number of the following graphs:

   - The complete graph K_n
   - The cycle graph C_n
   - The barbell graph

Chapter 2

Algebraic graph theory and network centrality

One of the key achievements of modern graph theory is the recognition that many graph properties admit an equivalent linear algebraic version. This is the subject of *algebraic graph theory* [21] some of whose basic notions are revised in this chapter. We will then show how these notions can be applied in order to introduce some meaningful measures of node centrality in a network that prove useful in several applications, as will be shown in the next chapters.

2.1 Basics of algebraic graph theory

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, the first natural matrix associated to it is of course the weight matrix W itself. The powers of the weight matrix W encode interesting information on the walks on the graph $\mathcal{G}$, as shown below. First, let us define the *weight* of a length-l walk $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l)$ as the product of its l link weights

$$W_\gamma = \prod_{1 \le h \le l} W_{\gamma_{h-1}\gamma_h} ,$$

with the convention that length-0 walks have unitary weight. In the special case of unweighted graphs all walks have unitary weight. Observe that a walk of length $l = 1$ from node i to node j is nothing but a link and its weight is simply W_{ij}. The following simple but important property states that the (i, j)-th entry of the matrix power W^l coincides with the sum of the weights of length-l walks from i to j.

Proposition 2.1. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Then, for every length $l \ge 0$

and every two nodes $i, j \in \mathcal{V}$,

$$(W^l)_{ij} = \sum_{\substack{\gamma \text{ length-}l \text{ walk} \\ \text{from } i \text{ to } j}} W_\gamma$$

Moreover, if $\mathcal{G}$ is unweighted, the right-hand side coincides the number of length-l walks from i to j.

Proof. By induction on $l \geq 0$. For $l = 0$, the statement is trivial as length-0 walks start and end in the same node and have unitary weight by convention. Then, assume the claim to be true for some l. Observe that the set of all length-$(l+1)$ walks from node i to node j can be partitioned according to the l-th visited node: if $\gamma = (i, \ldots, k, j)$ is a walk of length $(l+1)$, we have that $W_\gamma = W_{\tilde{\gamma}} W_{kj}$ where $\tilde{\gamma}$ is the subwalk of γ from i to k. We can compute

$$\begin{aligned} \sum_{\substack{\gamma \text{ length-}(l+1) \text{ walk} \\ \text{from } i \text{ to } j}} W_\gamma &= \sum_{k \in \mathcal{V}} W_{kj} \sum_{\substack{\gamma \text{ length-}l \text{ walk} \\ \text{from } i \text{ to } k}} W_{\tilde{\gamma}} \\ &= \sum_{k \in \mathcal{V}} (W^l)_{ik} W_{kj} \\ &= (W^{l+1})_{ij} \end{aligned}$$

where the second equality follows from the inductive assumption. □

Proposition 2.1 has the following straightforward consequence.

Corollary 2.1. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Then, for every length $l \geq 0$ and every two nodes i and j in $\mathcal{V}$,

(i) $(W^l)_{ij} > 0$ if and only if there exists a walk of length l from i to j;

(ii) $\mathcal{G}$ is strongly connected if and only if for every two nodes i and j in $\mathcal{V}$, there exists $l > 0$ such that $(W^l)_{ij} > 0$;

(iii) the l-th spectral moment of W, i.e., $\sum_{h=1}^{l} \lambda_h^l$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of W, is real nonnegative.

Proof. Points (i) and (ii) are immediate consequences of Proposition 2.1. For point (iii) it is sufficient to apply the formula for the trace of a matrix

$$\sum_{h=1}^{n} \lambda_h^3 = \operatorname{tr}(W^l) = \sum_{i \in \mathcal{V}} (W^l)_{ii} \,,$$

and then noticing that, by Proposition 2.1, the rightmost side of the above coincides with the total weight of length-l circuits in $\mathcal{G}$. □

In fact, also periodicity has a nice characterization in terms of powers of the weigh matrix W as stated in the following result.

Proposition 2.2. Let $\mathcal{G}$ ba a strongly connected graph. Then, $\mathcal{G}$ is aperiodic if and only if there exists $l > 0$ such that $(W^l)_{ij} > 0$ for every two nodes i and j in $\mathcal{V}$.

Proof. Notice that, if $(W^l)_{ij} > 0$ for every two nodes i and j in $\mathcal{V}$, then we also have that $(W^{l+1})_{ij} > 0$ for every i and j in $\mathcal{V}$. In particular this implies that $\text{per}_{\mathcal{G}}$ divides both l and $l+1$ and is thus equal to 1, so that $\mathcal{G}$ is aperiodic.

Conversely, if $\mathcal{G}$ is aperiodic, then, for every node i in $\mathcal{V}$ we can find circuits $\gamma^1, \gamma^2, \ldots, \gamma^s$ starting and ending in i of relatively prime lengths $l_1, \ldots, l_s$. This implies that we can find integers $a_1, \ldots, a_s$ such that $\sum_i a_i l_i = 1$. Now, let $m = \sum_i |a_i| l_i$. Then, clearly $m \geq 1$ and $\sum_i (|a_i| + a_i) l_i = m + 1$. If we now choose an integer $l > m(m-1)$, we can write it as $l = am + b$ for some integers $a \geq m - 1$ and $b \leq m - 1$. Then, $l = (a - b)m + b(m + 1)$ can be expressed as a non-negative integers combination of m and $m + 1$, hence there exists a length-l walk from i to i in $\mathcal{G}$. Indeed, if we put $m_i = (a - b)(|a_i| + a_i) + b|a_i|$, such a walk can be obtained by the concatenation of m_1 times γ^1, m_2 times γ^2 and so on. From this it is easy to see that we can find $\bar{l}$ such that, for every $l \geq \bar{l}$ and for every node i in $\mathcal{V}$, there is a path of length exactly l from i to i. Namely, $(W^l)_{ii} > 0$ for every $l \geq \bar{l}$ and for every i in $\mathcal{V}$. If we now consider $l = \text{diam}(\mathcal{G}) + \bar{l}$, it is immediate to verify that $(W^l)_{ij} > 0$ for every two nodes i and j in $\mathcal{V}$. □

We conclude this section with stating and proving the following result as a consequence of Propostiion 2.1.

Corollary 2.2. For a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, we have that:

1. (i) $(W^2)_{ii} = w_i$ for every node i in $\mathcal{V}$;
2. (ii) $\sum_{h=1}^n \lambda_h^2 = |\mathcal{E}| = 2\cdot$number of undirected links;
3. (iii) $\sum_{h=1}^n \lambda_h^3 = 6\cdot$number of triangles.

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ eigenvalues of W.

Proof. (i) Since $\mathcal{G}$ is simple, it contains no self-loops. Hence, the only length-2 circuits in $\mathcal{G}$ correspond to walking an undirected link from one end node to the other and back. It follows from Proposition 2.1 that, for every node i in $\mathcal{V}$, the term $(W^2)_{ii}$ coincides with the number of length-2 circuits starting and ending in i. From the previous observation, this is the same as the number of undirected links incident on i. Since the graph is simple, this nothing but the degree w_i of node i.

(ii) From point (i) we have that

$$\sum_{i \in \mathcal{V}} (W^2)_{ii} = \sum_{i \in \mathcal{V}} w_i \,.$$

The claim then follows from the well known formula for the trace of a matrix.

(iii) Proposition 2.1 implies that, for every node i in $\mathcal{V}$, the term $(W^3)_{ii}$ coincides with the number of length-3 circuits starting and ending in i. Arguing as in the proof of point (i), since $\mathcal{G}$ contains no self-loops, the only length-3 circuits in $\mathcal{G}$ are triangles. Hence, $(W^3)_{ii}$ is equal to twice the number of triangles passing through node i, where the factor 2 is due to the fact that every triangle gives rise to two length-3 circuits starting in i, one walked in clockwise and another one walked in courter-clockwise direction. The claim now follows again by applying the formula for the trace of a matrix.

□

2.2 Normalized weight, Laplacian, and node-link incidence matrices

Besides its weight/adjacency matrix, two other matrices commonly associated to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ turn out to be particularly relevant: these are the *normalized weight matrix* P and the *Laplacian* matrix L to be defined below. As we shall see, on the one hand such matrices P and L prove useful in characterizing some fundamental graph-theoretical properties, on the other hand, they are associated to fundamental linear network dynamics to be studied starting from Chapter 6.

In order to define the normalized weight matrix P and the Laplacian L we shall assume that all nodes have positive out-degree, i.e., $w_i > 0$ for all $i \in \mathcal{V}$. This causes no real loss of generality since, if $w_i = 0$ for some node i, we can always modify $\mathcal{G}$ by adding a self-loop on i of some positive weight W_{ii}. Then, let

$$D = \operatorname{diag}(w) \tag{2.1}$$

be a diagonal matrix whose diaginal entries correspond to the out-weight of the corresponding nodes and define the *normalized weight matrix*

$$P = D^{-1}W\,, \tag{2.2}$$

and the *Laplacian matrix*

$$L = D - W\,. \tag{2.3}$$

Observe that (2.2) is equivalent to that

$$P_{ij} = \frac{W_{ij}}{w_i}\,, \qquad \forall i, j \in \mathcal{V}\,, \tag{2.4}$$

i.e., the (i,j)-the entry of the normalized weight matrix P is nothing but the weight of the link (i,j) normalized by the out-weight $w_i = \sum_j W_{ij}$ of its tail node i. Similarly, (2.3) is equivalent to

$$L_{ij} = \begin{cases} -W_{ij} & \text{if} \quad i \neq j \\ \sum_{j \neq i} W_{ik} & \text{if} \quad i = j \end{cases}\,, \qquad \forall i, j \in \mathcal{V}\,, \tag{2.5}$$

i.e., every out-diagonal entry L_{ij} of the Laplacian matrix corresponds to the opposite of the weight W_{ij}, while every diagonal entry L_{ii} coincides with the total weigh of links with tail node i and head node $k \neq i$.

Example 2.1. For the graph in Figure 1.2 (c), the normalized weight matrix and the Laplacian matrix are, respectively,

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad L = \begin{bmatrix} 2 & -1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 2 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

The normalized weight matrix P and the Laplacian L of a graph have important structural features with remarkable consequences on their properties. First, as far as P is concerned, notice that all its entries are nonnegative: matrices with this property are simply referred to as *nonnegative*. Moreover,

$$P\mathbb{1} = \mathbb{1}\,, \tag{2.6}$$

i.e., every row of P sums up to 1. Non-negative square matrices satisfying property (2.6) are referred to as *stochastic* matrices. Observe that the weight matrix W of a graph is itself a nonnegative matrix, while typically it is not a stochastic matrix. Notice that every stochastic matrix P can be thought of as the normalized weight matrix of a graph. It is sufficient to consider $\mathcal{G}_P = (\mathcal{V}, \mathcal{E}, P)$ where $\mathcal{E} = \{(i,j) : P_{ij} > 0\}$. We shall refer to such $\mathcal{G}_P$ as the *graph associated* with the stochastic matrix P. Sometimes we shall use the following terminology: a stochastic matrix P is referred to as *irreducible* if $\mathcal{G}_P$ is strongly connected, and it is referred to as *aperiodic* if $\mathcal{G}_P$ is so.[1]

On the other hand, when it comes to the Laplacian matrix L, notice that all its off-diagonal entries are nonpositive, so that $-L$ is a so-called *Metzler* matrix.[2] Moreover, observe that

$$L\mathbb{1} = w - w = 0\,, \tag{2.7}$$

i.e., all rows of L have zero sum, so that 0 is an eigenvalue of L. Square matrices L such that $-L$ is Metzler and $L\mathbb{1} = 0$ are generally referred to as *Laplacian* matrices. Analogously to what done for stochastic matrices, we can associate to every Laplacian matrix L the graph $\mathcal{G}_L = \{\mathcal{V}, \mathcal{E}, W\}$, with link set $\mathcal{E} = \{(i,j) : i \neq j \text{ and } L_{ij} < 0\}$ and weight matrix W with $W_{ii} = 0$ for all i and $W_{ij} = -L_{ij}$ for all $i \neq j$. Such $\mathcal{G}_L = \{\mathcal{V}, \mathcal{E}, W\}$ will be referred to as the graph *associated*[3] with the Laplacian matrix L and the latter will be referred to as irreducible if $\mathcal{G}_L$ is strongly connected.

[1] Clearly, $\mathcal{G}_P$ is not the only graph whose normalized weight matrix coincides with P. For every positive diagonal matrix $\overline{D}$, the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \overline{W})$ with link set $\mathcal{E} = \{(i,j) : P_{ij} > 0\}$ and weight matrix $\overline{W} = \overline{D}P$ has normalized weight matrix P.

[2] A *Metzler* matrix is a square matrix whose extradiagonal entries are all nonnegative.

[3] Observe that $\mathcal{G}_L$ is the only graph without self-loops whose Laplacian is L. On the other hand, for any Laplacian matrix L and nonnegative vector s, the graph $\mathcal{G} = (\mathcal{V}, \overline{\mathcal{E}}, \overline{W})$ with link set $\mathcal{E} = \{(i,j) : i \neq j \text{ and } L_{ij} < 0\} \cup \{(i,i) : s_i > 0\}$ and weight matrix $\overline{W}$, with $\overline{W}_{ii} = s_i$ for all i and $\overline{W}_{ij} = -L_{ij}$ for all $i \neq j$, has Laplacian equal to L.

We conclude this section by introducing the *node-link incidence matrix* of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, that is a matrix B in $\mathbb{R}^{\mathcal{V}\times\mathcal{E}}$ whose entries are defined by

$$B_{ie} = \begin{cases} +1 & \text{if} \quad e = (i,j) \text{ for some } j \neq i \\ -1 & \text{if} \quad e = (j,i) \text{ for some } j \neq i \\ 0 & \text{if} \quad e = (i,i) \text{ or } e = (j,k) \text{ for some } j \neq i,\ k \neq i\,. \end{cases} \tag{2.8}$$

In other words, the node-link incidence matrix B has rows labeled by nodes i in $\mathcal{V}$ and columns labeled by links e in $\mathcal{E}$: columns corresponding to a directed link $e = (i,j)$ contain all zeros except —only for links that are not self-loops— for the entry B_{ie} corresponding to the tail node i, which is equal to $+1$, and the entry B_{je} corresponding to the head node j, which is equal to -1. Observe that the link weights are not taken into account in the definition of the node-link incidence matrix B. In fact, it proves convenient to stack up all link weights on the diagonal of a diagonal matrix H in $\mathbb{R}_+^{\mathcal{E}\times\mathcal{E}}$ whose entries are defined by $H_{ee} = W_{ij}$ for every link $e = (i,j)$ in $\mathcal{E}$ and $H_{ee'} = 0$ for every two links $e \neq e'$ in $\mathcal{E}$. The following simple result relates the node-link adjacency matrix to the Laplacian of a graph.

Proposition 2.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Let L be its Laplacian, B its node-link incidence matrix, and H the diagonal matrix containing the link weights. Then,

$$BHB' = \operatorname{diag}(w) - W + \operatorname{diag}(w^-) - W'\,. \tag{2.9}$$

In particular, if $\mathcal{G}$ is undirected, then

$$BHB' = 2L\,. \tag{2.10}$$

2.3 Perron-Frobenius theory

Below we recall some general facts on non-negative matrices for which a deep theory has been pioneered by the German mathematicians *Perron* and *Frobenius* [31, 18, 19, 20]. One of the main results of this theory is that nonnegative (not necessarily symmetric) matrices enjoy the property that their dominant eigenvalue is a nonnegative real number with a corresponding eigenvector with nonnegative entries. Precisely, we have the following result.

Theorem 2.1. Let M in $\mathbb{R}_+^{n\times n}$ be a non-negative square matrix. Then, there exist a nonnegative real eigenvalue $\lambda_M \geq 0$ and non-negative vectors $x \neq 0$ and $y \neq 0$ such that:

(i) $Mx = \lambda_M x,\ \ M'y = \lambda_M y$;

(ii) $\max\{\omega_{\min}, \omega^-_{\min}\} \leq \lambda_M \leq \min\{\omega_{\max}, \omega^-_{\max}\}$, where

$$\omega_{\min} = \min_i \sum_j M_{ij}\,, \qquad \omega^-_{\min} = \min_j \sum_i M_{ij}$$

are the minimum row and, respectively, column sum of M, while

$$\omega_{\max} = \max_i \sum_j M_{ij}\,, \qquad \omega^-_{\max} = \max_j \sum_i M_{ij}$$

are the maximum row and, respectively, column sum of M;

(iii) every eigenvalue λ of M is such that $|\lambda| \leq \lambda_M$.

Proof. Our strategy consists in constructing a nonempty convex compact subset $\mathcal{S} \subseteq \mathbb{R}^n_+$ and a continuous map $f : \mathcal{S} \to \mathcal{S}$, then applying Brower's fixed point theorem to ensure existence of a fixed point $x = f(x)$ and finally showing that such nonnegative vector x is necessarily an eigenvector of M associated to the dominant eigenvalue of M.

Let λ in $\mathbb{C}$ be an eigenvalue of M with maximum absolute value and let v in $\mathbb{C}^n$ be an associated eigenvector. Without loss of generality, we can assume that $\sum_i |v_i| = 1$. Let z in $\mathbb{R}^n_+$ be the vector with entries $z_i = |v_i|$. Then,

$$|\lambda| z_i = |\lambda v_i| = |(Mv)_i| = \left|\sum\nolimits_j M_{ij} v_j\right| \leq \sum\nolimits_j |M_{ij}|\,|v_j| = \sum\nolimits_j |M_{ij} z_j = (Mz)_i\,, \tag{2.11}$$

for every $i = 1, \ldots, n$. It follows that the compact convex set

$$\mathcal{S} = \{y \in \mathbb{R}^n_+ :\ \mathbb{1}'y = 1,\, My \geq |\lambda| y\}$$

is nonempty.

To continue, let us first assume that M has at least one positive entry in each column, equivalently, that $\omega^-_{\min} > 0$. Observe that, in this case

$$\mathbb{1}'My = \sum\nolimits_j y_j \sum\nolimits_i M_{ij} \geq \omega^-_{\min} > 0\,, \qquad \forall y \in \mathcal{S}\,,$$

so that the map $f : \mathcal{S} \to \mathbb{R}^n$ defined by

$$f(y) = \frac{My}{\mathbb{1}'My} \qquad y \in \mathcal{S}\,,$$

is well-defined and continuous and continuous on $\mathcal{S}$. We are now left with showing that the image of the function f is contained in $\mathcal{S}$ itself, i.e, that $f(\mathcal{S}) \subseteq \mathcal{S}$. Towards this goal, consider an arbitrary vector y in $\mathcal{S}$. First observe that, since M and y are both nonnegative, so is y. It also easily verified that $\mathbb{1}'f(y) = \mathbb{1}'My/(\mathbb{1}'My) = 1$. Moreover, since M is nonnegative and $My \geq |\lambda| y$ for every y in $\mathcal{S}$, we have

$$Mf(y) = \frac{M(Mz)}{\mathbb{1}'My} \geq \frac{|\lambda| My}{\mathbb{1}'My} = |\lambda| f(y)\,,$$

so that $f(y)$ belongs to $\mathcal{S}$. We have thus proved that f is continuous on $\mathcal{S}$ and $f(\mathcal{S}) \subseteq \mathcal{S}$.

Then, Brower's fixed point theorem guarantees that $f : \mathcal{S} \to \mathcal{S}$ admits at least one fixed point $x = f(x)$ in $\mathcal{S} \subseteq \mathbb{R}^n_+$. Consider any such fixed point x, let $\lambda_M = \mathbb{1}'Mx$. Observe that

$$\Lambda_M x = \Lambda_M f(x) = Mx\,,$$

so that Λ_M is indeed an eigenvalue of M with associated eigenvector x. Since $|\lambda|$ is the spectral radius and λ_M is an eigenvalue of M, and x is in $\mathcal{S}$, we have that

$$|\lambda| \geq \lambda_M = \mathbb{1}'Mx \geq |\lambda|\mathbb{1}'x = |\lambda|\,,$$

so that necessarily $\lambda_M = |\lambda|$ coincides with the spectral radius. Finally, observe that

$$\lambda_M = \sum\nolimits_i \sum\nolimits_j M_{ij}x_j = \sum\nolimits_j \omega_j x_j \leq \omega^-_{\max}\,,$$

and analogously, $\lambda_M \geq \omega^-_{\min}$.

The case when M contains some all-zero columns can be treated as follows. One starts by removing one all-zero column and the corresponding row from M, thus reducing its dimension by 1, and iterates this process until: either (a) the remaining matrix has dimension 0×0; or (b) the remaining matrix has dimension $k \times k$ with $k \geq 1$, and has no all-zero columns. The former case is equivalent to M being nilpotent, in which case claim (i) follows with $\lambda_M = 0$ and $x = \delta^{(i)}$ where i is the index of one of the all-zero columns of M. In the latter case, one applies the previous argument to the remaining matrix of dimension $k \times k$ and then extends its eigenvector to an eigenvector of M by putting all other $n - k$ entries equal to 0.

To complete the proof, notice that the arguments above can be applied to the transpose matrix M' which has the same eigenvalues as M. □

The eigenvalue λ_M is called the *dominant eigenvalue* of M and the vectors x and y are referred to as the right and, respectively, the left *dominant eigenvectors* of M. An immediate consequence of Theorem 2.1 is the following.

Corollary 2.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Assume that the out-degree of every node i is strictly positive. Then, there exists a positive dominant eigenvalue $\lambda_W > 0$ with associated nonnegative right eigenvector $x = \lambda_W^{-1}Wx$ and left eigenvector $y = \lambda_W^{-1}W'y$.

Proof. From Theorem 2.1 we are only left with proving that the dominant eigenvalue $\lambda_W > 0$. Notice that, since the out-degree w_i of every node i in $\mathcal{V}$ is strictly positive, there are walks in $\mathcal{G}$ of every possible length $l \geq 0$ (starting in any node you can keep on moving to an out-neighbor and never have to stop). By Proposition 2.1, this implies that $W^l \neq 0$ for every $l \geq 0$. Hence, 0 cannot be the only eigenvalue of W (for otherwise there would exist some $\bar{l} > 0$ such that $W^l = 0$ for $l > \bar{l}$). Hence, necessarily the dominant eigenvalue satisfies $\lambda_W > 0$. □

In fact, Theorem 2.1 implies several important properties of stochastic matrices that are gathered in the result below.

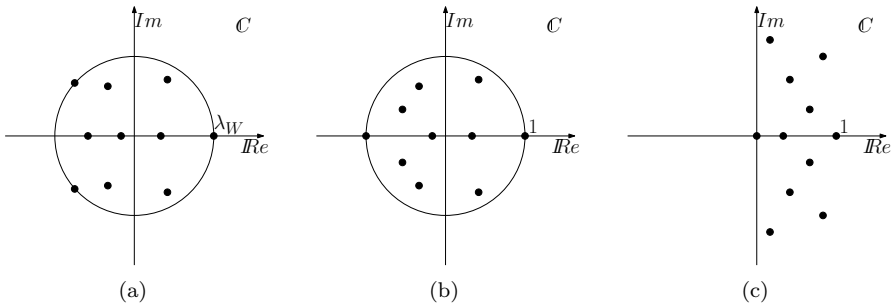


Figure 2.1: The spectra (i.e., sets of eigenvalues) of the weight matrix W (a), the normalized weight matrix P (b), and the Laplacian matrix L (c) of a graph $\mathcal{G}$. The spectrum of W is contained in a disk centered in 0 of radius equal to the dominant eigenvalue λ_W. The spectrum of P is contained in a disk centered in 0 of unitary radius: -1 is an eigenvalue if the graph is bipartite. The spectrum of L includes 0 and all other eigenvalues are contained in the open right semiplane of the complex plane (positive real part).

Proposition 2.4. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph, and let $P = D^{-1}W$ be its normalized weight matrix. Then,

(i) $\lambda_P = 1$;

(ii) there exists a nonnegative vector π such that $\mathbb{1}'\pi = 1$ and $P'\pi = \pi$;

(iii) w is an eigenvector of P' associated to the eigenvalue 1 if and only if $\mathcal{G}$ is balanced;

(iv) $\mathbb{1}$ is an eigenvector of P' associated to the eigenvalue 1 if $\mathcal{G}$ is regular;

(v) if $\mathcal{G}$ is strongly connected, then $\lambda_P = 1$ is geometrically and algebraically simple and there exists a positive vector $y \in \mathbb{R}^n$ such that $P'y = y$;

(vi) if $\mathcal{G}$ is strongly connected and aperiodic, any eigenvalue $\lambda \neq 1$ of P is such that $|\lambda| < 1$;

(vii) if $\mathcal{G}$ is bipartite, then -1 is an eigenvalue of P;

(viii) if $\mathcal{G}$ is undirected, then all eigenvalues of P are real and have algebraic multiplicity equal to their geometric multiplicity.

Proof.

(i) That 1 is an eigenvalue of P, hence of P', is implied by $P\mathbb{1} = \mathbb{1}$, as noted already. To prove that no eigenvalue of P is larger than 1 in absolute value, let y be a nonnegative eigenvector of P' associated to the dominant eigenvalue λ_P. Then, $\lambda_P y = P'y$, and taking the scalar product with the all-one vector gives $\lambda_P \mathbb{1}'y = \mathbb{1}'P'y = \mathbb{1}'y$. Since y is nonnegative and

different from 0, one has that $\mathbb{1}'y > 0$, so that $\lambda_P \mathbb{1}'y = \mathbb{1}'y$ implies that $\lambda_P = 1$.

(ii) It is an immediate consequence of Theorem 2.1. Notice that if $P'y = y$ and y is non-negative, necessarily $y'\mathbb{1} \neq 0$ and it is thus sufficient to consider $\pi = (y'\mathbb{1})^{-1}y$.

(iii) We have $P'w = W'D^{-1}w = W'\mathbb{1} = w^-$, so $P'w = w$ if and only if $w = w^-$.

(iv) If $\mathcal{G}$ is regular, then it is balanced. Then, by point (iii),

$$P'\mathbb{1} = \frac{1}{\overline{w}}P'w = \frac{1}{\overline{w}}w = \mathbb{1}\,.$$

(v) Suppose $x \in \mathbb{C}^{\mathcal{V}}$ is such that $Px = x$. Then, $P|x| \geq |Px| = |x|$. Consider $\mathcal{W} := \operatorname{argmax}_{i\in\mathcal{V}} |x_i|$ and notice that, if $i \in \mathcal{W}$,

$$|x|_i \leq \sum_j P_{ij}|x_j| \leq \sum_j P_{ij}|x_i| = |x_i|$$

This implies that, for every j such that $P_{ij} > 0$, we must have $j \in \mathcal{W}$. Since $\mathcal{G}$ is strongly connected, this easily implies that $\mathcal{W} = \mathcal{V}$ and thus $|x| = a\mathbb{1}$ for some scalar $a > 0$. It also follows that $P|x| = |x|$ and thus $|Px| = P|x|$. For every node i we thus have the identity $|\sum_j P_{ij}x_j| = \sum_j |P_{ij}x_j|$ which implies that there exists a unitary complex number z_i such that $x_j = az_i$ for every $j \in N_i$. Using the identity $Px = x$, we thus get that $x_i = x_j$ for every $j \in N_i$. Using again strong connectedness, it follows that $x = \alpha\mathbb{1}$ for some scalar α. This proves that $\lambda_P = 1$ is geometrically simple. If it was not algebraically simple, in the Jordan canonical form of P, it would necessarily show up a Jordan block of type

$$J = \begin{pmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & \ddots & 1 \\ & & & 1 \end{pmatrix}$$

This leads to a contradiction as the off-diagonal 1's imply that J^t blows up as t grows large while, on the other hand, P^t is a stochastic matrix for all $t \geq 0$, hence in particular it remains bounded as t grows large.

It remains to be shown the existence of an eigenvector $y > 0$ for P' relative to the eigenvalue 1. From Theorem 2.1, we know there exists a non-negative vector $y \neq 0$ such that $P'y = y$. From $y_j = \sum_j P_{ij}y_i$, we have that if $y_j = 0$ we also have that $y_i = 0$ for every $i \in N_j^-$. If there were components with $y_i = 0$, then again strong connectedness would yield $y = 0$ which is a cotradiction.

(vi) Suppose $x \in \mathbb{C}^{\mathcal{V}}$ is such that $Px = \lambda x$ for some $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$. It follows from Proposition 2.2 that there exists $N \in \mathbb{N}$ such that P^N has all non-zero elements. From the relation $P^N x = \lambda^N x$ we obtain $P^N|x| \geq |P^N x| = |x|$ and repeating the same computations done in proving (iv), we obtain that $x = \alpha \mathbb{1}$ for some scalar α. This implies that $\lambda = 1$.

(vii) Suppose that $\mathcal{G}$ is bipartite with respect to the decomposition $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$. Let $x \in \mathbb{R}^{\mathcal{V}}$ be the vector such that $x_i = 1$ for all $i \in \mathcal{V}_0$ and $x_i = -1$ for all $i \in \mathcal{V}_1$. If there are no links between nodes in $\mathcal{V}_0$, then for every $i \in \mathcal{V}_0$ one has $\sum_{j\in\mathcal{V}} P_{ij}x_j = \sum_{j\in\mathcal{V}_1} P_{ij}x_j = -\sum_{j\in\mathcal{V}_1} P_{ij} = -1$. Similarly, if there are no links between nodes in $\mathcal{V}_1$, then for every $i \in \mathcal{V}_1$ one has $\sum_{j\in\mathcal{V}} P_{ij}x_j = \sum_{j\in\mathcal{V}_0} P_{ij}x_j = \sum_{j\in\mathcal{V}_0} P_{ij} = 1$. Hence, $Px = -x$.

(viii) Let M in $\mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ be the nonnegative square matrix with entries

$$M_{ij} = P_{ij}\frac{\sqrt{w_i}}{\sqrt{w_j}}\,, \qquad \forall i, j \in \mathcal{V}\,. \tag{2.12}$$

For every i and j in $\mathcal{V}$, using (2.12), (2.4), the assumption that $\mathcal{G}$ is undirected, and then again (2.4) and (2.12), we get that

$$M_{ji} = P_{ji}\frac{\sqrt{w_j}}{\sqrt{w_i}} = \frac{W_{ji}}{\sqrt{w_j w_i}} = \frac{W_{ij}}{\sqrt{w_j w_i}} = P_{ij}\frac{\sqrt{w_i}}{\sqrt{w_j}} = M_{ij}\,,$$

thus showing that the matrix M is symmetric. As such, M has all real eigenvalues and an orthonormal basis of eigenvectors. Now, notice that, for the diagonal matrix D defined in (2.1), we have that $P = D^{-1/2}MD^{1/2}$. Therefore, P is similar to M, hence in particular it is diagonalizable with the same real eigenvalues as M, each with geometric multiplicity equal to its algebraic multiplicity. □

Some of the properties of the normalized weight matrix P presented in Proposition 2.4 have an analogue for the Laplacian matrix L, as summarized in the following result.

Proposition 2.5. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph, and let L be its Laplacian matrix. Then,

(i) 0 is an eigenvalue of L and L' and all their other eigenvalues have positive real part;

(ii) $\overline{y}$ is in the kernel of L' if and only if $\overline{y} = D^{-1}y$ for some y such that $P'y = y$;

(iii) there exists a non-negative vector $\overline{\pi}$ such that $\mathbb{1}'\overline{\pi} = 1$ and $L'\overline{\pi} = 0$;

(iv) $\mathbb{1}$ is in the kernel of L' if and only if $\mathcal{G}$ is balanced;

(v) if $\mathcal{G}$ is strongly connected, then 0 is algebraically and geometrically simple as an eigenvalue of L and there exists $\overline{y} > 0$ such that $L'\overline{y} = 0$;

(vi) if $\mathcal{G}$ is undirected, then all the eigenvalues of L are real nonnegative and have geometric multiplicity equal to their algebraic multiplicity.

Proof.

(i) The fact that 0 is an eigenvalue of L, and hence of its transpose L', is implied by (2.7). On the other hand, it follows by Gershgorin's circle theorem that every eigenvalue of L lies within the union over all $i = 1, \dots, n$ of the disks centered at $L_{ii} = w_i$ of radius $r_i = \sum_{j \neq i} |L_{ij}| = \sum_{j \neq i} W_{ij} = w_i$. All those disks lie on the right of the imaginary axis in the complex plane and are tangent to the imaginary axis in the origin, from which it can be deduced that all non-zero eigenvalues of L have strictly positive real part.

(ii) We have $L'\overline{y} = (D - W')\overline{y} = (I - P')D\overline{y}$ which is 0 if and only if $y = D\overline{y}$ satisfies $P'y = y$.

(iii) This follows from point (ii) above and Proposition 2.4 (ii).

(iv) This follows from point (ii) above and Proposition 2.4 (iii).

(v) Notice that $x \in \mathbb{C}^{\mathcal{V}}$ satisfies $Lx = 0$ if and only if $Px = x$. Therefore, if $\mathcal{G}$ is strongly connected, then 0 is a simple eigenvalue of L because 1 is a simple eigenvalue of P. The second part of the statement follows from point (ii) above and Proposition 2.4 (v).

(vi) This immediately follows from the fact that the Laplacian matrix L is symmetric when the graph $\mathcal{G}$ is undirected.

□

Besides stochastic matrices and Laplacian matrices, there is another class of non-negative matrices playing an important role in the study of graphs in relation to specific subset of nodes. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph and let $\mathcal{R} \subseteq \mathcal{V}$ be a subset of nodes. Let $P = D^{-1}W$ be the normalized weight matrix of $\mathcal{G}$ and let

$$\underline{P} = P_{|\mathcal{R} \times \mathcal{R}}$$

be its restriction to the subset of nodes $\mathcal{R}$. If $\mathcal{R}$ is a trapping set on $\mathcal{G}$, then, $\underline{P}$ is also stochastic. If instead $\mathcal{R}$ is not trapping, we will have that

$$\underline{P}\mathbb{1} \leq \mathbb{1}\,, \qquad \underline{P}\mathbb{1} \neq \mathbb{1}\,. \tag{2.13}$$

Non-negative square matrices $\underline{P}$ satisfying (2.13) are called *sub-stochastic*. Any sub-stochastic matrix can always be though of as obtained as the restriction of the normalized weight matrix of a graph. Indeed, given a sub-stochastic matrix $\underline{P}$ on the set $\mathcal{R}$, consider $\mathcal{U} = \mathcal{R} \cup \{o\}$ and the graph $\mathcal{G} = (\mathcal{U}, \mathcal{E}, P)$ where $P \in \mathbb{R}^{\mathcal{U} \times \mathcal{U}}$ is the stochastic matrix with entries

$$P_{oo} = 1\,, \quad P_{oj} = 0\,, \quad P_{ij} = \underline{P}_{ij}\,, \quad P_{io} = 1 - \sum_{j \in \mathcal{R}} \underline{P}_{ij}\,, \quad \forall i, j \in \mathcal{R}\,,$$

and the link set is $\mathcal{E} = \{(i,j) \,|\, P_{ij} > 0\}$. Such matrix P is called the *stochastic completion* of $\underline{P}$. The following is a key property of sub-stochastic matrices.

Proposition 2.6. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph and $\mathcal{R} \subseteq \mathcal{V}$ a proper subset of nodes such that $\mathcal{U} = \mathcal{V} \setminus \mathcal{R}$ is globally reachable. Then, the non-negative matrix $\underline{P} = P_{|\mathcal{R}\times\mathcal{R}}$ is sub-stochastic with $\lambda_{\underline{P}} < 1$.

Proof. The fact that the matrix $\underline{P}$ is sub-stochastic immediately follows from the fact that since $\mathcal{U}$ is nonempty and globally reachable, $\mathcal{R}$ cannot be trapping.

We now prove the second assertion. Let $y \neq 0$ be a non-negative vector such that $\underline{P}'y = \lambda_{\underline{P}} y$ and let $\mathcal{S} \subseteq \mathcal{R}$ be the support of y, i.e., $\mathcal{S}$ is the set of nodes i such that $y_i > 0$. Since $\mathcal{U}$ is globally reachable, $\mathcal{S}$ is not trapping, so that

$$\min_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}} \underline{P}_{ij} < 1\,.$$

Then, we have that

$$\lambda_{\underline{P}} \sum_{i\in\mathcal{S}} y_i = \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{V}} \underline{P}_{ji} y_j = \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}} \underline{P}_{ji} y_j < \sum_{j\in\mathcal{S}} y_j\,,$$

which implies that $\lambda_{\underline{P}} < 1$. □

We can now establish the following result showing an interesting link between the condensation graph and the spectral properties of P.

Proposition 2.7. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a finite graph and let $s_{\mathcal{G}}$ be the number of sinks in the condensation graph $\mathcal{H}_{\mathcal{G}}$. Then, the algebraic and geometric multiplicities of 1 as an eigenvalue of the normalized weight matrix P coincide with $s_{\mathcal{G}}$ and so do the algebraic and geometric multiplicities of 0 as an eigenvalue of L.

Proof. Let $\mathcal{V}_1, \ldots, \mathcal{V}_k$ be the connected components of $\mathcal{G}$, ordered in such a way that, if in the condensation graph $\mathcal{H}_{\mathcal{G}}$ the node corresponding to $\mathcal{V}_i$ is connected to the node corresponding to $\mathcal{V}_j$, then $1 \leq i \leq j \leq n$. Upon possibly a relabeling, we may assume that the nodes in $\mathcal{V}$ are ordered in such a way that, for every $1 \leq i \leq j \leq n$, every node in $\mathcal{V}_i$ comes before every node in $\mathcal{V}_j$. Then, the normalized weight matrix P has the following block-triangular structure:

$$P = \begin{pmatrix} P^{(1,1)} & \cdots & P^{(1,k)} \\ & \ddots & \vdots \\ 0 & & P^{(k,k)} \end{pmatrix}.$$

Notice now that a connected component $\mathcal{V}_i$, for $1 \leq i \leq k$, corresponds to a sink in the condensation graph $\mathcal{H}_{\mathcal{G}}$ if and only if the matrix $P^{(i,i)}$ is stochastic and irreducible while $P^{(i,j)} = 0$ for all $i < j \leq k$. On the other hand, a connected component $\mathcal{V}_i$ does not correspond to a sink in $\mathcal{H}_{\mathcal{G}}$ if and only if $P^{(i,i)}$ is sub-stochastic and $P^{(i,j)} \neq 0$ for some $i < j \leq k$. Considering that the

algebraic and geometric multiplicity of the eigenvalue 1 in P is the sum of the corresponding multiplicities of the eigenvalue 1 in the various diagonal blocks $P^{(i,i)}$, for $1 \leq i \leq k$, the first claim immediately follows from Propositions 2.4 and 2.6. The second claim follows arguing as in the proof of Proposition 2.5 (v). □

We end this section by introducing some further terminology that will prove useful in the next chapters. Recall that Proposition 2.4(ii) and Proposition 2.5(iii) state that the transpose matrices P' and L' of the normalized weight matrix and of the Laplacian matrix of a graph $\mathcal{G}$ admit non-negative eigenvectors related to their eigenvalue 1 and 0, respectively. Considering the important role they play in many applications, we give them a specific name: non-negative vectors y and $\overline{y}$ such that

$$P'y = y \tag{2.14}$$

and

$$L'\overline{y} = 0 \tag{2.15}$$

are referred to as, respectively, *invariant distributions* and *Laplace invariant distributions* of the graph $\mathcal{G}$. Moreover, (Laplace) invariant distributions $\pi = P'\pi$ such that $\mathbb{1}'\pi = 1$ are referred to as *(Laplace) invariant probability distributions* or, more briefly, as *(Laplace) invariant probabilities* of the graph $\mathcal{G}$.

The following result provides a characterization of the set of (Laplace) invariant probability distributions.

Proposition 2.8. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Then,

(i) Any convex combinations of (Laplace) invariant probability distributions of $\mathcal{G}$ is a (Laplace) invariant probability distribution of $\mathcal{G}$;

(ii) For every sink in the condensation graph $\mathcal{H}_{\mathcal{G}}$ of $\mathcal{G}$, there exists a (Laplace) invariant probability distribution supported on the connected component of $\mathcal{G}$ corresponding to such sink of $\mathcal{H}_{\mathcal{G}}$. Such (Laplace) invariant probability distributions are referred to as *extremal*;

(iii) Every (Laplace) invariant probability distribution can be obtained as a convex combination of the extremal (Laplace) invariant probability distributions.

Proof. Point (i) is straightforward. On the other hand, since an invariant probability distribution is nothing but a nonnegative eigenvector of P' associated to its eigenvalue 1 normalized in such a way that its entries sum up to 1, items (ii), (iii) immediately follow from Propositions 2.4 and 2.7. Analogous arguments can be applied to Laplace invariant probability distributions. □

By combining Propositions 2.8 and 2.7 we get the following result.

Corollary 2.4. Let $\mathcal{G}$ be a graph with $s_{\mathcal{G}} = 1$. Then, $\mathcal{G}$ has a unique invariant probability distribution π and a unique Laplace invariant probability

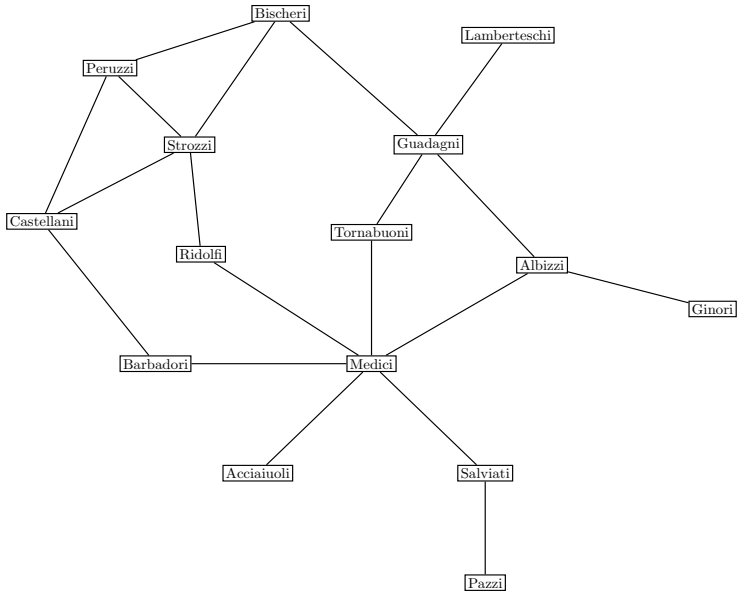


Figure 2.2: Marriages among prominent Florentine families in the 15th century from Padgett and Ansell [30].

distribution $\overline{\pi}$. Moreover, both π and $\overline{\pi}$ have support coinciding with the connected component of $\mathcal{G}$ that corresponds to the sink of $\mathcal{H}_{\mathcal{G}}$. In particular, if $\mathcal{G}$ is strongly connected, then $\pi > 0$ and $\overline{\pi} > 0$.

2.4 Network centrality

We now want to study measures that capture the importance of a node's position in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$. These are referred to as *centrality* measures and there is a vast literature on them. We will not be exhaustive here, but rather focus on few key concepts that are naturally connected to the dynamical models that we will study in the next chapters.

The simplest notion is the *degree centrality* whereby the importance of a node i is simply by its degree. Of course, in non-balanced networks one should decide whether the in-degree w_i^- (number of links pointing to node i) or the out-degree (number of links originating from node i) is to be used. Both choices may be useful in appropriate contexts. E.g., for a Twitter account, the number of its followers (i.e., its in-degree) may be a first relevant measure of its influence. Similarly, the number of citations received can be thought of as a measure of the impact of a research article. Both number of followers in Twitter and number of citations in a citation network are exemples of in-degree centrality measures.

A natural extension of the (in-)degree centrality is the eigenvector centrality.

The key idea is that connections from other nodes that have themselves high centrality should contribute more to the centrality of a node than connections from nodes that have low centrality. Formally, we would like the centrality y_i of node i to be proportional to the sum of the centralities of the in-neighbors j of i, i.e.,

$$z_i \propto \sum_j W_{ji} z_j \,.$$

If we denote the proportionality constant by $1/\lambda > 0$, and let $z \in \mathbb{R}^n$ be the vector of node centralities, we can rewrite this as

$$\lambda z = W' z \,. \tag{2.16}$$

This says that y is an eigenvector of W' of eigenvalue $\lambda > 0$. If we choose $\lambda = \lambda_W$ the dominant eigenvalue of W, it follows from Corollary 2.3 that W' admits a corresponding non-negative eigenvector $z = \lambda_W^{-1} W' z$. If we assume $\mathcal{G}$ to be strongly connected and we impose the normalization $z'\mathbb{1} = 1$, then z is unique and called the *eigenvector centrality* of $\mathcal{G}$.

One drawback of the eigenvector centrality notion is that nodes contribute to the centrality of all their out-neighbors irrespective of their out-degree. To overcome this issue, one can proceed by normalizing by the out-degree so that

$$z_i \propto \sum_j \frac{W_{ji}}{w_j} z_j \,.$$

This corresponds to replacing the adjacency matrix W by its normalized version, the stochastic matrix $P = D^{-1}W$ where $D = \operatorname{diag}(w)$. Considering that the dominant eigenvalue for P is 1, this leads to the equation

$$z = P' z \,, \tag{2.17}$$

i.e., z is an invariant distribution. In particular, if assume the normalization $\mathbb{1}'z = 1$, then we get that $z = \pi$ is an invariant probability. Again, if $\mathcal{G}$ is strongly connected, then Proposition 2.4 (v) implies that such π is unique. This centrality measure is called the *invariant distribution centrality* of $\mathcal{G}$.

Both the eigenvector centrality and the invariant distribution centrality suffer from the limitation that nodes can increase the centrality of a given node arbitrarily, by adding a self-loop on this node of very large weight. In the limit as the weight of a self-loop grows large, the ratio between the centrality of this node and the total centrality of all other nodes grows unbounded, even without loosing connectivity. Even if self-loops are not allowed, one can easily take two nodes and add an undirected link between them of larger and larger weight: in the limit as the weight of this undirected link grows large, the ratio between the sum of the two nodes' centralities and the centralities of all other nodes grows to infinity. This drawback is overcome by modifying the notion of centrality by allowing nodes to get some centrality, independently of their in-neighbors. Formally, let us choose some parameter $\beta \in (0, 1]$, and a nonnegative vector μ

to be thought of some intrinsic centrality. The standard choice is $\mu = \mathbb{1}$, so that all nodes have identical intrinsic centrality. Then, we can define the *Katz centrality* vector [25] as the solution $z^{(\beta)}$ of

$$z^{(\beta)} = \left(\frac{1-\beta}{\lambda_W}\right) W' z^{(\beta)} + \beta\mu\,, \tag{2.18}$$

where λ_W is the dominant eigenvalue of W'. Note that, for every $0 < \beta \leq 1$, the dominant eigenvalue of $\lambda_W^{-1}(1-\beta)W'$ is smaller than 1, so that the matrix $(I - \lambda_W^{-1}(1-\beta)W')$ is invertible. This implies that the Katz centrality vector is well defined, unique, and can be represented as

$$z^{(\beta)} = \left(I - \lambda_W^{-1}(1-\beta)W'\right)^{-1} \beta\mu\,.$$

In the limit as β converges to 0, the right-hand side of (2.18) converges to the one of (2.16), so that the eigenvector centrality notion is recovered. On the other hand, for $\beta = 1$, we simply get $z^{(\beta)} = \mu$, i.e., the Katz centrality vector does not depend on the graph structure. Hence, we should think of β as a parameter that measures the weight the intrinsic centrality relative to the network topology in the definition of the Katz centrality.

If the normalized adjacency matrix P is used instead of the non-normalized one W, then the solution of

$$z^{(\beta)} = (1-\beta)\, P' z^{(\beta)} + \beta\mu\,, \tag{2.19}$$

is referred to as the *Bonacich* centrality [8]. A version of this centrality is also known as the *P*ageRank centrality vector, as introduced by Brin and Page [10] to measure the relative importance of webpages in the WWW, where typical values of β used are about 0.15. For general choices of μ, the vector $z^{(\beta)}$ solution to (2.19) is referred to as the personalized PageRank [24], and is used in context-sensitive searches. Observe that the dominant eigenvalue of $(1-\beta)\, P'$ is equal to $1-\beta$, so that $(I - (1-\beta)P')$ is invertible. Using the expansion of the geometric series, the Pagerank centrality vector can be expressed as

$$z^{(\beta)} = \beta \sum_{k\geq 0} (1-\beta)^k (P')^k \mu = \beta\mu + \beta(1-\beta)P'\mu + \beta(1-\beta)^2(P')^2\mu + \dots \tag{2.20}$$

Equation (2.20) shows how the PageRank centrality $\pi_i^{(\beta)}$ of a node i can be expressed as a convex combination of its own *a priori* centrality μ_i and of the terms $(\mu' P'^k)_i$ that depend on the centralities of the other nodes within distance k from i. The form of the corresponding weight coefficients $\beta(1-\beta)^k$ show that the way nodes influence the determination of the Page-Rank centrality $\pi_i^{(\beta)}$ of a node i, is exponentially decreasing with respect to the distance. A similar interpretation also applies to Katz's centrality.

2.4.1 Network centrality in production networks

TO BE WRITTEN

Chapter 3

Network Connectivity and Flows

3.1 Connectivity and Menger's Theorem

In this section and in the following ones, it will be sometimes convenient to work with a notion of network structure that is somewhat broader than the notion of directed weighted and that allows to possibly have multiple link connecting the same pair pair of nodes. Specifically we will consider *directed multigraphs* as quadruples $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$ where $\mathcal{V}$ and $\mathcal{E}$ are two sets, representing as usual nodes and links respectively. $\theta : \mathcal{E} \to \mathcal{V}$ and $\kappa : \mathcal{E} \to \mathcal{V}$ are two functions with the interpretation that if $e \in \mathcal{E}$, $\theta(e)$ and $\kappa(e)$ are two distinct nodes representing the start and the termination of link e. Note that in this setting a link can no longer be interpreted as simply unordered pair of nodes as there can be more links insisting on the same pair of nodes. Two links $e_1, e_2 \in \mathcal{E}$ such that $\theta(e_1) = \theta(e_2)$ and $\kappa(e_1) = \kappa(e_2)$ are called *parallel*. Instead if $\theta(e_1) = \kappa(e_2)$ and $\kappa(e_1) = \theta(e_2)$, they are called *opposite*.

To every unweighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, as defined in Chapter 1 a multi-graph can be naturally associated by simply considering $\tilde{\mathcal{G}} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$ having the same set of vertices and links of the original $\mathcal{G}$ and start and terminal functions defined by

$$\theta(i, j) = i, \quad \kappa(i, j) = j$$

From now on we will identify the two object $\mathcal{G}$ and $\tilde{\mathcal{G}}$ and tend to use the same notation for both.

For such multi-graphs it is possible to consider essentially all concepts introduced in Chapters 1 and 2. In particular, we will frequently use the node-link incidence matrix $B \in \mathbb{R}^{\mathcal{V} \times \mathcal{E}}$ here defined as

$$B_{\theta(e)e} = +1\,, \quad B_{\kappa(e)e} = -1\,, \quad B_{ke} = 0\,, \quad \forall k \in \mathcal{V} \setminus \{\theta(e), \kappa(e)\}\,,$$

for every link e in $\mathcal{E}$ that is not a self-loop (i.e., such that $\theta(e) \neq \kappa(e)$), and B_{ie} for all i in $\mathcal{V}$ if e is a self-loop ($\theta(e) = \kappa(e)$). In this new context, walks are

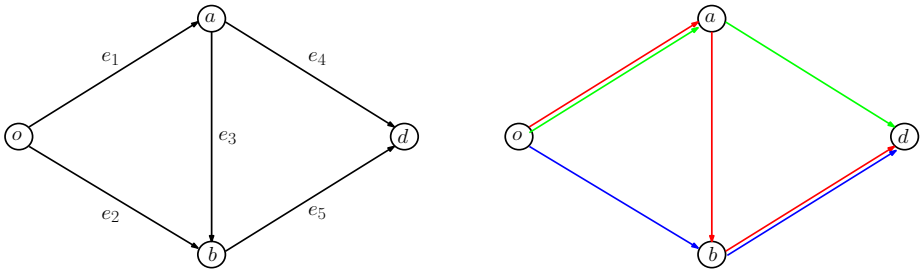


Figure 3.1: On the left, a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V} = \{o, a, b, d\}$ and link set $\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$. On the right, the three distinct o-d paths are colored: $\gamma^{(1)} = (o, a, d)$ (green); $\gamma^{(2)} = (o, a, b, d)$ (red); $\gamma^{(3)} = (o, b, d)$ (blue). The node-link incidence matrix B and the link-path incidence matrix A are reported in equation (3.10). Note that paths $\gamma^{(1)}$ and $\gamma^{(3)}$ are both node- and link-independent, while $\gamma^{(2)}$ is neither node- nor link-independent from either $\gamma^{(1)}$ or $\gamma^{(3)}$. The node-connectivity and the link-connectivity of node o to node d then satisfy $c_{\text{node}}(o, d) = c_{\text{link}}(o, d) = 2$.

defined as sequences of links $\gamma = (e_1, e_2, \ldots, e_l)$ such that $\kappa(e_i) = \theta(e_{i+1})$ for every $i = 1, \ldots, l-1$. A walk γ is said to start in $\theta(e_1)$ and terminate (or end) in $\kappa(e_l)$ and l is its length. If e is one of the link of the walk γ, we write, with slight abuse of notation, $e \in \gamma$. A walk $\gamma = (e_1, e_2, \ldots, e_l)$ such that $\theta(e_h) \neq \theta(e_k)$ for all $h \neq k$ in $\{1, \ldots, l\}$, or equivalently $\kappa(e_h) \neq \kappa(e_k)$ for all $h \neq k$ in $\{1, \ldots, l\}$, is referred to as a path. A walk $\gamma = (e_1, e_2, \ldots, e_l)$ such that $\theta(e_1) = \kappa(e_l)$ is referred to as a circuit. A path $\gamma = (e_1, e_2, \ldots, e_l)$ such that $\theta(e_1) = \kappa(e_l)$ and $l \geq 3$ is referred to as a cycle.

Consider a multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$ and two nodes $i, j \in \mathcal{V}$. We denote by $\Gamma_{(i,j)}$ the set of paths in the graph $\mathcal{G}$ starting in i and ending in j (i-j *paths*). Clearly, the set $\Gamma_{(i,j)}$ is nonempty if and only if j is reachable from i. We would like to relate the degree of connectivity of a graph to the size of these sets, however, we have to take into account the fact that paths in $\Gamma_{(i,j)}$ may share nodes or links.

The next definition introduces two notions of independence in the set $\Gamma_{(i,j)}$ and corresponding notions of connectivity.

Definition 3.1. Two i-j paths $\gamma^{(t)} = (e_1^{(t)}, \ldots, e_{l_t}^{(t)}) \in \Gamma_{(i,j)}$, for $t = 1, 2$, are

- *node-independent* if they share no intermediate node, i.e., if, defined the set of intermediate nodes of the two paths as

$$I^{(t)} = \{\theta(e_h^{(t)}),\ h = 2, \ldots, l_t\} \cup \{\kappa(e_h^{(t)}),\ h = 1, \ldots, l_t - 1\}$$

  we have that $I^{(1)} \cap I^{(2)} = \emptyset$;

- *link-independent* if they share no link, i.e., if $e_h^{(1)} \neq e_k^{(2)}$ for every $h = 1, \ldots, l_1$ and $k = 1, \ldots, l_2$.

The *node-connectivity* $c_{\text{node}}(i,j)$ and the *link-connectivity* $c_{\text{link}}(i,j)$ of the pair of nodes (i,j) are the maximum number of node-independent and, respectively, link-independent i-j paths.

We can now define notions of connectivity for the entire graph.

Definition 3.2. Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$ The *node-connectivity* and *link-connectivity* of $\mathcal{G}$ are defined, respectively, as

$$c_{\text{node}}(\mathcal{G}) = \min_{i \neq j \in \mathcal{V}} c_{\text{node}}(i,j)\,, \qquad c_{\text{link}}(\mathcal{G}) = \min_{i \neq j \in \mathcal{V}} c_{\text{link}}(i,j)\,.$$

Example 3.1. In the graph of Figure 3.1, of the three o-d paths ($\gamma^{(1)}$, $\gamma^{(2)}$, and $\gamma^{(3)}$) paths $\gamma^{(1)}$ and $\gamma^{(3)}$ are both node- and link-independent, while $\gamma^{(2)}$ is neither node- nor link-independent from either $\gamma^{(1)}$ or $\gamma^{(3)}$. The node-connectivity and the link-connectivity of node o to node d then satisfy $c_{\text{node}}(o,d) = c_{\text{link}}(o,d) = 2$. (Note that the graph $\mathcal{G}$ in Figure 3.1 is not connected so that $c_{\text{node}}(\mathcal{G}) = c_{\text{link}}(\mathcal{G}) = 0$.)

The following result, known as Menger's Theorem [7, Ch. 3,Th. 5], relates the node and link connectivities to the minimum number of nodes and links that have to be removed from a graph in order to disconnect two nodes.

Proposition 3.1 (Menger's Theorem)**.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$ be a graph and $i \neq j$ two distinct nodes. Then, minimum number of nodes (respectively, links) that has to be removed from $\mathcal{G}$ in order for j not to be reachable from i equals $c_{\text{node}}(i,j)$ (resp. $c_{\text{link}}(i,j)$). Hence, the minimum number of nodes (resp. links) that have to be removed from $\mathcal{G}$ in oder to disconnect it equals $c_{\text{node}}(\mathcal{G})$ (resp. $c_{\text{link}}(\mathcal{G})$).

We do not prove Menger's theorem immediately, but rather show how it can be interpreted as a special case of a more general result, the *max-flow min-cut* theorem that will be discussed in the following.

3.2 Network flows

The purpose of this section is to make the reader acquainted with the notion of network flows that is useful in many applications and will recur frequently in the next chapters.

We now introduce the notion of network flows. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$, an *exogenous net flow* on $\mathcal{G}$ is a vector $\nu \in \mathbb{R}^{\mathcal{V}}$ satisfying the constraint

$$\sum_i \nu_i = 0\,. \tag{3.1}$$

The positive part $[\nu_i]_+ = \max\{0, \nu_i\}$ and the negative part $[\nu_i]_- = \max\{0, -\nu_i\}$ of the net flow in a node i are to be interpreted as the *exogenous inflow* in and, respectively, the *external outflows* from i. The constraint (3.1) is then equivalent

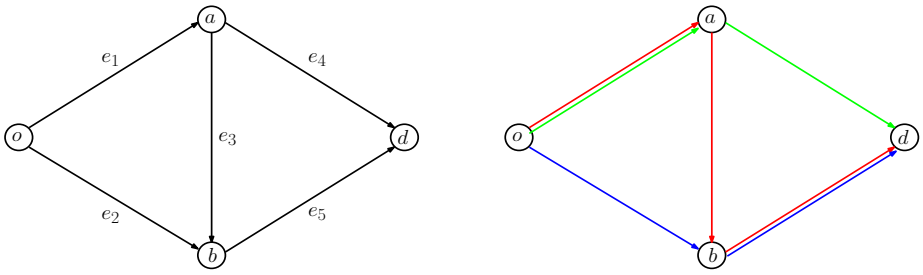


Figure 3.2: On the left, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$ with node set $\mathcal{V} = \{o, a, b, d\}$ and link set $\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$. On the right, the three distinct o-d paths are colored: $\gamma^{(1)} = (o, a, d)$ (green); $\gamma^{(2)} = (o, a, b, d)$ (red); $\gamma^{(3)} = (o, b, d)$ (blue). The node-link incidence matrix B and the link-path incidence matrix A are reported in equations (3.9) and (3.10), respectively.

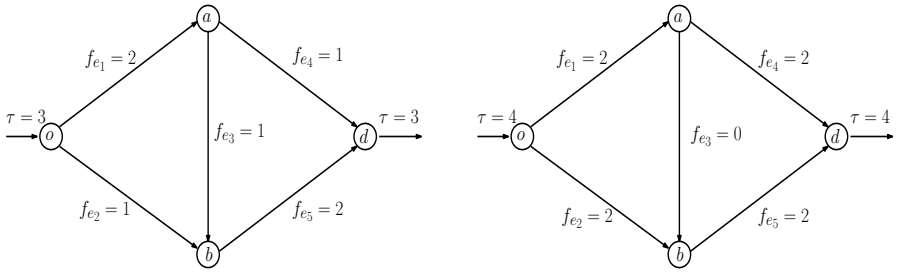


Figure 3.3: Two o-d flows for the network in Figure 3.2 that are both feasible with respect to the link capacities specified in Figure 3.5. While both satisfy the flow balance (3.11) as well as the capacity constraints (**??**), the flow on the left has a lower throughput ($\upsilon = 3$) than the one on the right ($\upsilon = 4$).

to requiring that the total exogenous inflow matches the total external outflow. We shall refer to

$$\upsilon = \frac{1}{2}\sum_i |\nu_i| = \sum_i [\nu_i]_+ = \sum_i [\nu_i]_- \tag{3.2}$$

as the *throughput*, i.e., the total flow that goes through the network. Depending on the applications, nodes i such that $\nu_i > 0$ are called *sources*, *origins*, or *generators*, whereas nodes i such that $\nu_i < 0$ are called *sinks*, *destinations*, or *loads*.

Given a graph $\mathcal{G}$ and a vector of exogenous net flows $\nu \in \mathbb{R}^{\mathcal{V}}$ satisfying (3.1), a *network flow* is a nonnegative vector $f \in \mathbb{R}_+^{\mathcal{E}}$ whose entries f_e satisfy the flow balance equations

$$\nu_i + \sum_{e\in\mathcal{E}\,|\,\kappa(e)=i} f_e = \sum_{e\in\mathcal{E}\,|\,\theta(e)=i} f_e\,, \qquad i \in \mathcal{V}\,. \tag{3.3}$$

The term f_e represents the flow on the link $e \in \mathcal{E}$. Equation (3.3) above states that the total inflow in a node i —resulting from both possible exogenous inflow $[\nu_i]_+$ and flows f_e from incoming links e — equals the total outflow from i — resulting from both possible external outflow $[\nu_i]_-$ and flows f_e towards outgoing links e. The flow balance equations (3.3) can be rewritten more compactly in terms of the node-link incidence matrix as

$$Bf = \nu\,. \tag{3.4}$$

Network flows with a single origin o and a single destination d, i.e., nonnegative vectors $f \in \mathbb{R}_+^{\mathcal{E}}$ such that

$$Bf = \upsilon(\delta^{(o)} - \delta^{(d)}) \tag{3.5}$$

for some throughput value υ, will be referred to as *o-d flows*.

We now show how *o-d* flows can be constructed easily using *o-d* paths and cycles. To this aim, it is useful to introduce some more notation. First, let Δ be the set of all closed paths. Notice that this includes self-loops, length-2 circuits that are not the concatenation of a self-loop, and cycles in $\mathcal{G}$. In fact, shall identify closed paths modulo their start node, so that, e.g., (e_1, e_2, e_3), (e_2, e_3, e_1) and (e_3, e_1, e_2) will be considered as the same length-3 cycle (directed triangle). With a slight abuse of notation, in the rest of this chapter we shall refer to an element γ of Δ as a cycle.[1] Then, let C in $\{0,1\}^{\mathcal{E}\times\Delta}$ be the link-cycle incidence matrix defined by

$$C_{e\gamma} = \begin{cases} 1 & \text{if} \quad e \text{ is along } \gamma \\ 0 & \text{if} \quad e \text{ is not along } \gamma\,. \end{cases}$$

Also, for two nodes $o \neq d$ in $\mathcal{V}$ such that d is reachable from o, let the *link-path incidence matrix* $A^{(o,d)}$ in $\{0,1\}^{\mathcal{E}\times\Gamma_{(o,d)}}$ be defined by

$$A^{(o,d)}_{e\gamma} = \begin{cases} 1 & \text{if} \quad e \text{ is along } \gamma \\ 0 & \text{if} \quad e \text{ is not along } \gamma\,. \end{cases} \tag{3.6}$$

[1] In other terms we are considering seflloops as length-1 cycles and length-2 circuits as length-2 cycles.

Then, we have the following simple result.

Lemma 3.1. Consider a multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$ and let B and C be its node-link incidence matrix and its link-cycle incidence matrix, respectively. Then,

$$BC = 0\,. \tag{3.7}$$

Moreover, for two nodes $o \neq d$ in $\mathcal{V}$, let $A^{(o,d)}$ be the link-path incidence matrix. Then,

$$BA^{(o,d)} = (\delta^{(o)} - \delta^{(d)})\mathbb{1}'\,. \tag{3.8}$$

Proof. First, let $\gamma = (e_1, \ldots, e_l)$ in Δ be a length-l cycle. Let us use the convention $e_0 = e_l$. Then, if a node i in $\mathcal{V}$ is along γ, then there exist a unique $j_i \in \{1, \ldots, l\}$ such that $i = \theta(e_{j_i}) = \kappa(e_{j_i - 1})$ so that $B_{ie_{j_i-1}} = 1 = -B_{ie_{j_i}}$. It follows that

$$(BC)_{i\gamma} = \sum_{e \in \mathcal{E}} B_{ie} C_{e\gamma} = \sum_{j=1}^{l} B_{ie_j} = B_{ie_{j_i-1}} - B_{ie_{j_i}} = 0\,.$$

On the other hand, if a node i in $\mathcal{V}$ is not along γ, then $B_{ie_j} = 0$ for $1 \leq j \leq l$, so that

$$(BC)_{i\gamma} = \sum_{e \in \mathcal{E}} B_{ie} C_{e\gamma} = \sum_{j=1}^{l} B_{ie_j} = 0\,.$$

This proves (3.7)

Now, let $\gamma = (e_1, \ldots, e_l)$ in $\Gamma_{o,d}$ be a length-l o-d path. For $i = o$, we have $\theta(e_1) = o$, so that $B_{oe_1} = 1$ and $B_{oe_j} = 0$ for $1 < j \leq l$. Then,

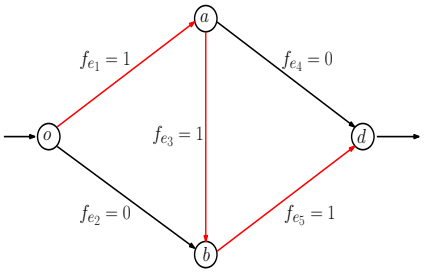
$$(BA^{(o,d)})_{o\gamma} = \sum_{e \in \mathcal{E}} B_{oe} A^{(o,d)}_{e\gamma} = \sum_{j=1}^{l} B_{oe_j} = 1\,.$$

Reasoning analogously, for $i = d$, we have $\kappa(e_l) = d$, so that $B_{de_l} = -1$ and $B_{de_j} = 0$ for $1 \leq j < l$, so that $(BA^{(o,d)})_{d\gamma} = -1$. On the other hand, for every node i in $\mathcal{V} \setminus \{o, d\}$, we have $B_{ie_j} = -1$ for some $j = 1. \ldots, l-1$ if and only if $B_{ie_{j+1}} = +1$. This implies that $(BA^{(o,d)})_{i\gamma} = 0$, thus completing the proof of (3.8). □

In the graph of Figure 3.2, we have

$$\mathcal{V} = \{o, a, b, d\},\ \mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$$

There are three different o-d paths: $\gamma^{(1)} = (o, a, d)$ (colored in green); $\gamma^{(2)} = (o, a, b, d)$ (red); $\gamma^{(3)} = (o, b, d)$ (blue). The link-path incidence matrix and the



Example 3.2.

Figure 3.4: The $\gamma^{(2)}$-th column of the link-path incidence matrix A for the network in Figure 3.2 is a unitary o-d flow.

node-link incidence matrices are

$$A^{(o,d)} = \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \gamma^{(1)} & \gamma^{(2)} & \gamma^{(3)} \\ \end{matrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \, , \end{matrix} \tag{3.9}$$

$$B = \begin{matrix} \\ o \\ a \\ b \\ d \, , \end{matrix} \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ \end{matrix} \begin{bmatrix} +1 & +1 & 0 & 0 & 0 \\ -1 & 0 & +1 & +1 & 0 \\ 0 & -1 & -1 & 0 & +1 \\ 0 & 0 & 0 & -1 & -1 \end{bmatrix} \tag{3.10}$$

and their product gives

$$BA^{(o,d)} = \begin{matrix} \\ o \\ a \\ b \\ d \, . \end{matrix} \begin{matrix} \gamma^{(1)} & \gamma^{(2)} & \gamma^{(3)} \\ \end{matrix} \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

An o-d flow is a nonnegative vector $f = (f_{e_1}, f_{e_2}, f_{e_3}, f_{e_4}, f_{e_5})$ satisfying

$$\upsilon = f_{e_1} + f_{e_2} \, , \qquad f_{e_1} = f_{e_3} + f_{e_4} \, , \qquad f_{e_2} + f_{e_3} = f_{e_5} \, , \qquad f_{e_4} + f_{e_5} = \upsilon \, . \tag{3.11}$$

Two o-d flows are displayed in Figure 3.3.

The following result, known as the Flow Decomposition Theorem states that every assignment of flows to both o-d paths and cycles in the multigraph induces

a unique network flow f on the links and that, conversely, for every network flow f on the links, there exists a —possibly (and typically) non-unique— assignment of flows to both o-d paths and cycles in the multigraph that induces f. This is a very useful result that we will use later on in this chapter when proving the max-flow min-cut theorem in order to construct a feasible flow as well as in the next chapter when dealing with network flow optimization.

Theorem 3.1 (Flow Decomposition Theorem)**.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$ be a multigraph with link-cycle incidence matrix C. Let $o \neq d$ be two nodes such that d is reachable from o in $\mathcal{G}$ and let $A^{(o,d)}$ be the link-path incidence matrix. Then, for every two vectors z in $\mathbb{R}_+^{\Gamma_{(o,d)}}$ and w in $\mathbb{R}_+^{\Delta}$,

$$f = A^{(o,d)} z + C w \,, \tag{3.12}$$

is an o-d flow vector of throughput $\upsilon = \mathbb{1}'z$. Conversely, for every o-d flow vector of throughput υ, there exist vectors z in $\mathbb{R}_+^{\Gamma_{(o,d)}}$ and w in $\mathbb{R}_+^{\Delta}$ such that (3.12) holds true.

Proof. Let f be as in (3.12). Clearly, since the matrices $A^{(o,d)}$ and C and the vectors z and w are all nonnegative, so is f. On the other hand, it follows from Lemma 3.1 that

$$Bf = B(A^{(o,d)} z + Cw) = BA^{(o,d)} z = (\delta^{(o)} - \delta^{(d)}) \mathbb{1}' z = (\delta^{(o)} - \delta^{(d)}) \upsilon \,.$$

thus proving that the flow balance equation (3.5) is satisfied.

For the converse implication, see [2, Theorem 2.1]. □

3.3 Capacity and max-flow min-cut theorem

In this section, we study flows that satisfy capacity constraints on the links. Precisely, given a multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$ we assign to every link $e \in \mathcal{E}$ a capacity $c_e > 0$ that represents the maximum flow that is allowed through it. (See, e.g., Figure 3.5.) We denote the vector of all link capacities by $c \in \mathbb{R}^{\mathcal{E}}$.

A natural question is then to characterize the maximum throughput υ from a given node o to another node d that can be achieved by a flow vector f without violating the link capacity constraints, e.g. $f_e \leq c_e$ for every $e \in \mathcal{E}$.

Example 3.3. For the network in Figure 3.1, two different flows satisfying the capacity constraints are displayed in Figure 3.3: the one on the left has a lower throughput ($\upsilon = 3$) than the one on the right ($\upsilon = 4$).

Formally, this is the optimization problem that we want to consider. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$, a capacity vector $c \in \mathbb{R}^{\mathcal{E}}$, $c > 0$, and two distinct nodes $o, d \in \mathcal{V}$, the *maximum flow problem* is

$$\upsilon^*_{o,d} = \max_{\substack{\upsilon \geq 0 \\ 0 \leq f \leq c \\ Bf = \upsilon(\delta^{(o)} - \delta^{(d)})}} \upsilon \tag{3.13}$$

In the optimization problem (3.13), $f \in \mathbb{R}_+^{\mathcal{E}}$ is an $|\mathcal{E}|$-dimensional network flow vector that, besides the conservation of mass constraint $Bf = v(\delta^{(o)} - \delta^{(d)})$, it also satisfies the capacity constraints $0 \leq f \leq c$. We will refer to such flows as *feasible flows*. The objective function is the corresponding throughput $v \geq 0$, i.e., the total flow that goes through the network entering in node o and leaving from node d. Observe that (3.13) is a linear program (both objective function and constraints are all linear functions of the variables). Also, note that the set of feasible flows in (3.13) is never empty as it contains the null flow $f = 0$ with throughput $v = 0$. We will refer to the quantity $v_{o,d}^*$ as the *maximum throughput* from o to d.

How do we guarantee that a flow vector achieves the maximum throughput from an origin node o to a destination node d? The max-flow min-cut theorem allows us to answer this question by relating the value of the optimal throughput $v_{o,d}^*$ to some geometrical properties of the graph $\mathcal{G}$.

In particular, we first introduce the notion of an *o-d cut*: this is simply a partition of the node set $\mathcal{V}$ in two subsets, $\mathcal{U}$ and $\mathcal{U}^c$, such that the origin node o belongs to $\mathcal{U}$ and the destination node d belongs to the complementary set $\mathcal{U}^c$. We can visualize this by dividing the network into two parts by a separating line such that the origin remains on one side (say the left one) and the destination on the other side (say the right one). For convenience, we will identify an *o-d* cut with the set $\mathcal{U}$ that contains the origin. Then, we define the capacity of the cut as the aggregate capacity of the links crossing it from the left to the right. In formulas, the *capacity* of an *o-d* cut $\mathcal{U}$ is given by

$$c_{\mathcal{U}} := \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}^c} c_{ij} \,. \tag{3.14}$$

Example 3.4. The four *o-d* cuts for the network in Figure 3.5 are $\mathcal{U}_1 = \{o\}$, $\mathcal{U}_2 = \{o, a\}$, $\mathcal{U}_3 = \{o, b\}$, and $\mathcal{U}_4 = \{o, a, b\}$ and have capacities $c_{\mathcal{U}_1} = 4$, $c_{\mathcal{U}_2} = 9$, $c_{\mathcal{U}_3} = 4$, and $c_{\mathcal{U}_4} = 6$, respectively. (See Figure 3.5.)

The max-flow min-cut theorem tells us that the maximum throughput from o to d coincides with the minimum capacity among all *o-d* cuts:

$$c_{o,d}^* = \min_{\substack{\mathcal{U} \subseteq \mathcal{V} \\ o \in \mathcal{U},\, d \notin \mathcal{U}}} c_{\mathcal{U}} \,. \tag{3.15}$$

We will refer to $c_{o,d}^*$ defined above as the *min-cut capacity* of the network, and to a cut $\mathcal{U}$ with minimal capacity $c_{\mathcal{U}} = c_{o,d}^*$ as a *bottleneck*.

Theorem 3.2 (Max-flow min-cut theorem [16, 15])**.** Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$, a capacity vector $c \in \mathbb{R}^{\mathcal{E}}$, $c > 0$, and two distinct nodes $o, d \in \mathcal{V}$. Then,

$$v_{o,d}^* = c_{o,d}^* \,, \tag{3.16}$$

Moreover, if the link capacities are all integer-valued, then a feasible maximum throughput flow can be constructed such that the flow on every link is an integer value.

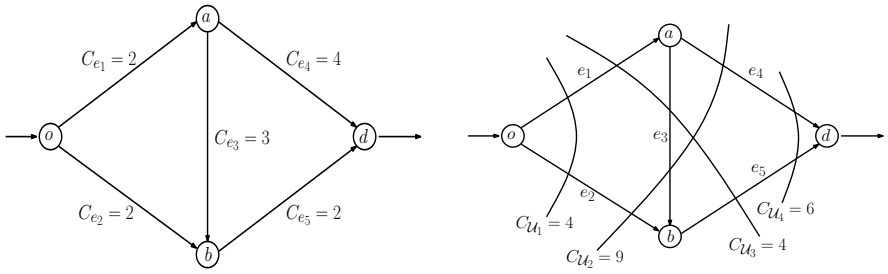


Figure 3.5: The same graph as in Figure 3.1 with capacities c_e assigned to its links, so that the capacity vector reads $c = (2, 2, 3, 4, 2)$. There are four possible o-d cuts in the network above $\mathcal{U}_1 = \{o\}$, $\mathcal{U}_2 = \{o, a\}$, $\mathcal{U}_3 = \{o, b\}$, and $\mathcal{U}_4 = \{o, a, b\}$. The capacities of the cuts are given, respectively, by $c_{\mathcal{U}_1} = c_{e_1} + c_{e_2} = 4$, $c_{\mathcal{U}_2} = c_{e_2} + c_{e_3} + c_{e_4} = 9$, $c_{\mathcal{U}_3} = c_{e_1} + c_{e_5} = 4$, and $c_{\mathcal{U}_4} = c_{e_4} + c_{e_5} = 6$. The min-cut capacity is thus $c^*_{o,d} = 4$ and the minimal capacity cuts are $\mathcal{U}_1$ and $\mathcal{U}_3$.

Proof. The proof of Theorem 3.2 can naturally be divided in two parts: (i) an first part that consists in showing that $v^*_{o,d} \leq c^*_{o,d}$, i.e., that no feasible flow can have throughput larger than the min-cut capacity; and (ii) a second part that consists in showing that $v^*_{o,d} \geq c^*_{o,d}$, i.e., there exists a feasible flow with throughput equal to the min-cut capacity.

(i) We first introduce a notation. Given two subsets of nodes $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$, we define

$$\mathcal{E}_{\mathcal{A}\to\mathcal{B}} = \{e \in \mathcal{E} : \ \theta(e) \in \mathcal{A}\,, \ \kappa(e) \in \mathcal{B}\}$$

for the set of links directed from some node in $\mathcal{A}$ to some node in $\mathcal{B}$. We now show that for every feasible flow f from o to d with throughput v, and for every o-d cut $\mathcal{U}$, the following relation holds

$$\sum_{e\in\mathcal{E}_{\mathcal{U}\to\mathcal{U}^c}} f_e = v + \sum_{e\in\mathcal{E}_{\mathcal{U}^c\to\mathcal{U}}} f_e\,, \tag{3.17}$$

Equation (3.17) states the intuitive fact that the aggregate outflow from $\mathcal{U}$ equals the sum of the inflow v in the origin o plus the aggregate flow on the links directed from the rest of the node set $\mathcal{U}^c$ towards $\mathcal{U}$. To formally verify (3.17), we sum the rows in $\mathcal{U}$ of relation (3.5) and we obtain (3.17)

$$\begin{aligned} v &= \sum_{i\in\mathcal{U}} \sum\nolimits_{e\in\mathcal{E}} B_{ie} f_e = \sum_{i\in\mathcal{U}} \left[\sum_{\substack{e\,\in\,\mathcal{E}\\ \theta(e)\,=\,i}} f_e - \sum_{\substack{e\,\in\,\mathcal{E}\\ \kappa(e)\,=\,i}} f_e \right] \\ &= \sum_{e\in\mathcal{E}_{\mathcal{U}\to\mathcal{U}}} f_e + \sum_{e\in\mathcal{E}_{\mathcal{U}\to\mathcal{U}^c}} f_e - \sum_{e\in\mathcal{E}_{\mathcal{U}\to\mathcal{U}}} f_e - \sum_{e\in\mathcal{E}_{\mathcal{U}^c\to\mathcal{U}}} f_e \\ &= \sum_{e\in\mathcal{E}_{\mathcal{U}\to\mathcal{U}^c}} f_e - \sum_{e\in\mathcal{E}_{\mathcal{U}^c\to\mathcal{U}}} f_e \end{aligned} \tag{3.18}$$

Applying (3.17) with an o-d cut $\mathcal{U}$ of minimal capacity and using the feasibility

constraints $0 \leq f_e \leq c_e$ for every link $e \in \mathcal{E}$, we finally obtain that

$$c^*_{o,d} = \sum_{e \in \mathcal{E}_{\mathcal{U} \to \mathcal{U}^c}} c_e \geq \sum_{e \in \mathcal{E}_{\mathcal{U} \to \mathcal{U}^c}} f_e = \upsilon + \sum_{e \in \mathcal{E}_{\mathcal{U}^c \to \mathcal{U}}} f_e \geq \upsilon \,.$$

Since this holds true for every feasible flow, we have proven that

$$\upsilon^*_{o,d} \leq c^*_{o,d} \,.$$

(ii) In order to prove the converse inequality, $\upsilon^*_{o,d} \geq c^*_{o,d}$, we construct a feasible flow f from o to d with throughput υ equal to the min-cut capacity $c^*_{o,d}$. This is done through a celebrated iterative algorithm due to Ford and Fulkerson [16, 2] and it is illustrated below. The state of the algorithm at the t-th iteration, where $t \geq 0$, consists of a flow vector $f^{(t)} \in \mathbb{R}^{\mathcal{E}}_+$ of throughput $\upsilon^{(t)}$ from o to d. As a function of the flow vector $f^{(t)}$ we define a *residual capacity* vector

$$c^{(t)} = c - f^{(t)} \tag{3.19}$$

and a *residual graph* $\mathcal{G}^{(t)} = (\mathcal{V}, \mathcal{E}^{(t)}, s, t)$ whose node set $\mathcal{V}$ is the same as in the original graph $\mathcal{G}$, and whose link set $\mathcal{E}^{(t)}$ is defined as follows:

$$e \in \mathcal{E}^{(t)} \quad \text{if} \quad e \in \mathcal{E} \text{ and } c^{(t)}_e > 0, \text{ or } \exists \bar{e} \in \mathcal{E} \text{ opposite to } e, \text{ and } f^{(t)}_{\bar{e}} > 0 \,. \tag{3.20}$$

The initial flow and throughput are set to zero, i.e.,

$$f^{(0)} = 0 \,, \qquad \upsilon^{(0)} = 0 \,,$$

so that

$$c^{(0)} = c \,, \qquad \mathcal{G}^{(0)} = \mathcal{G} \,,$$

At every step $t \geq 0$, the algorithm proceeds as follows. Define

$$\mathcal{U}^{(t)} = \left\{ i \in \mathcal{V} \,:\, i \text{ reachable from } o \text{ in } \mathcal{G}^{(t)} \right\} .$$

- If $d \notin \mathcal{U}^{(t)}$, i.e., if node d is not reachable from node o in $\mathcal{G}^{(t)}$, then the algorithm halts and returns the flow $f^{(t)}$.
- If $d \in \mathcal{U}^{(t)}$, then choose any o-d path

  $$\gamma = (e_1, \ldots, e_l)$$

  in $\mathcal{G}^{(t)}$. For each $i = 1, \ldots, l$ define

  $$\bar{f}^{(t)}_{e_i} = \max_{\substack{\bar{e} \,\in\, \mathcal{E} \\ \text{oppos. to } e_i}} f^{(t)}_{\bar{e}}$$

  and put

  $$\varepsilon^{(t)} := \min_{1 \leq i \leq l} \max \left\{ c^{(t)}_{e_i}, \bar{f}^{(t)}_{e_i} \right\} . \tag{3.21}$$

Update the flow $f^{(t)}$ to a flow $f^{(t+1)}$ of throughput $\upsilon^{(t+1)} = \upsilon^{(t)} + \varepsilon^{(t)}$ as follows

$$f^{(t+1)} = f^{(t)} + \varepsilon^{(t)} \sum_{1 \le i \le l} \chi_i^{(t)}, \qquad \chi_i^{(t)} = \begin{cases} \delta^{e_i} & \text{if} \quad \varepsilon^{(t)} > \bar{f}_{e_i}^{(t)} \\ -\delta^{\bar{e}_i} & \text{if} \quad \varepsilon^{(t)} \le \bar{f}_{e_i}^{(t)} \end{cases} \tag{3.22}$$

where $\bar{e}_i$ is any link in $\mathcal{E}$ opposite to e_i that maximizes the flow, namely $f_{\bar{e}_i}^{(t)} = \bar{f}_{e_i}^{(t)}$.

The vector $f^{(t)}$ constructed in this way is a feasible flow for every $t \ge 0$. This follows from the following inductive argument. Notice first that $f^{(0)}$ is feasible. Assume now now that $f^{(t)}$ is feasible. Notice that $(\varepsilon^{(t)} \sum_{1 \le i \le l} \chi_i^{(t)})_e = 0$ if $e \neq e_i, \bar{e}_i$ for $i = 1, \dots, l$. This yields

$$0 \le f_e^{(t+1)} \le c_e$$

if $e \neq e_i, \bar{e}_i$. Given i, if $\varepsilon^{(t)} > \bar{f}_{e_i}^{(t)}$, necessarily we have that $\varepsilon^{(t)} \le c_{e_i}^{(t)}$ and thus

$$0 \le f_{e_i}^{(t)} \le f_{e_i}^{(t+1)} \le f_{e_i}^{(t)} + c_{e_i}^{(t)} \le c_{e_i}, \quad 0 \le f_{\bar{e}_i}^{(t+1)} = f_{\bar{e}_i}^{(t)} \le c_{\bar{e}_i}$$

In the case instead when $\varepsilon^{(t)} \le \bar{f}_{e_i}^{(t)}$,

$$c_{\bar{e}_i} \ge f_{\bar{e}_i}^{(t)} \ge f_{\bar{e}_i}^{(t+1)} \ge f_{\bar{e}_i}^{(t)} - \varepsilon^{(t)} \ge 0, \quad 0 \le f_{e_i}^{(t+1)} = f_{e_i}^{(t)} \le c_{e_i}$$

Note now that, for all $1 \le i \le l$, it holds that $B\chi_i^{(t)} = \delta^{(\theta(e_i))} - \delta^{(\kappa(e_i))}$. Hence, by (3.22),

$$\begin{array}{rcl} Bf^{(t+1)} & = & Bf^{(t)} + \varepsilon^{(t)} \sum_{1 \le i \le l} \left(\delta^{(\theta(e_i))} - \delta^{(\kappa(e_i))} \right) \\ & = & Bf^{(t)} + \varepsilon^{(t)} \left(\delta^{(o)} - \delta^{(d)} \right) \\ & = & (\upsilon^{(t)} + \varepsilon^{(t)}) \left(\delta^{(o)} - \delta^{(d)} \right) \end{array} .$$

This proves that $f^{(t+1)}$ is a feasible $o-d$ flow of throughput $\upsilon^{(t+1)} = \upsilon^{(t)} + \varepsilon^{(t)}$

The final step is to prove that, if the algorithm stops at time t^*, then $\mathcal{U}^{(t^*)}$ is an o-d cut such that $\upsilon^{(t^*)} = c_{\mathcal{U}_{t^*}} \ge c^*_{o,d}$ Indeed, that $\mathcal{U}^{(t^*)}$ is an o-d cut follows from the fact that while $o \in \mathcal{U}_t$ for all t, the definition of t^* yields $d \notin \mathcal{U}_{t^*}$. It follows from (3.20) that for all $e \in \mathcal{E}_{\mathcal{U}^{(t^*)} \to \mathcal{U}^{(t^*)\,c}}$ it holds $c_e^{(t)} = 0$ or, equivalently, $f_e^{(t)} = c_e$. Instead, if $e \in \mathcal{E}_{\mathcal{U}^{(t^*)\,c} \to \mathcal{U}^{(t^*)}}$, then $f_e^{(t)} = 0$. Applying formula (3.17) with $\mathcal{U} = \mathcal{U}^{(t^*)}$ we finally obtain that

$$c_{\mathcal{U}_{t^*}} = \sum_{e \in \mathcal{E}_{\mathcal{U}^{(t^*)} \to \mathcal{U}^{(t^*)\,c}}} c_e = \upsilon^{(t^*)}$$

This completes the proof Finally, a straightforward check shows that, if the link capacities are all positive integers, then the flow built as above has integer entries. □

Remark 3. *Before proving Theorem 3.2, we observe that it can be equivalently rephrased as follows: the minimum total capacity that a hypothetical adversary needs to remove from the network in order to make node d not reachable from node o coincides with the min-cut capacity* $c^*_{o,d}$ *as defined in* (3.15)*. To see that this interpretation is indeed equivalent, first note that if such hypothetical adversary removed all the links in a minimum capacity o-d cut (thus removing a total capacity of* $c^*_{o,d}$*), then the min-cut capacity of the resulting graph would be* 0 *and Theorem 3.2 would imply that there exists no feasible flow from node o to node d in such resulting graph but the all-zero one, hence d would not be reachable from o. On the other hand, if a total capacity of less than* $c^*_{o,d}$ *were removed in any possible way by such adversary, the min-cut capacity of the resulting graph would still be positive so that Theorem 3.2 would imply that there exists a feasible flow with positive throughput from node o to node d.*

Proof of Menger's theorem, Proposition 3.1: We consider first the link connectivity case. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s, t)$ and two distinct nodes $o, d \in \mathcal{V}$, we consider the capacity vector c where $c_e = 1$ for every $e \in \mathcal{E}$. From Remark 3, it follows that, in this case, $c^*_{o,d}$ can be interpreted as the minimum number of links that a hypothetical adversary needs to remove from the network in order to make node d not reachable from node o. On the other hand, given any integer-valued feasible $o-d$-flow f, the corresponding support $\{e \in \mathcal{E} \,|\, f_e = 1\}$ is (Problem 2.2) necessarily the union of a set of link-disjoint paths from o to d and the throughput is exactly the number of such paths. Therefore $c_{\text{link}}(o, d)$ is equal the maximum possible throughput $v^*_{o,d}$from o to d. Result then follows from Theorem 3.2.

□

Example 3.5. Let us apply the Ford-Fulkerson algorithm in order to construct a feasible flow vector f of maximal throughput v for the capacitated graph of Figure 3.5. Clearly, we start with

$$f^{(0)} = (0,0,0,0,0)\,, \qquad c^{(0)} = (2,2,3,4,2)\,, \qquad \mathcal{E}_0 = \mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}\,,$$

and $\mathcal{U}_0 = \mathcal{V} = \{o, a, b, d\}$. We then have to choose $\gamma^{(0)}$ to be one of the three possible o-d paths. If we choose $\gamma^{(0)} = (o, a, d)$, we then get

$$\varepsilon_0 = \min\{c_{e_1}, c_{e_4}\} = 2\,, \qquad f^{(1)} = f^{(0)} + 2(\delta^{(e_1)} + \delta^{(e_4)}) = (2,0,0,2,0)\,,$$

so that, with $e_6 = (a, o)$ and $e_9 = (d, a)$, we have

$$c^{(1)} = c - f^{(1)} = (0,2,3,2,2)\,, \qquad \mathcal{E}_1 = \{e_2, e_3, e_4, e_5, e_6, e_9\}\,, \qquad \mathcal{U}_1 = \{o, a, b, d\}\,.$$

Then, in the next step we have to take $\gamma^{(1)} = (o, b, d)$ as this is the only o-d path in the residual graph $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$, so that

$$\varepsilon_1 = \min\{c^{(1)}_{e_2}, c^{(1)}_{e_5}\} = 2\,, \qquad f^{(2)} = f^{(1)} + 2(\delta^{(e_2)} + \delta^{(e_5)}) = (2,2,0,2,2)\,,$$

$$c^{(2)} = c - f^{(2)} = (0,0,3,2,0)\,, \qquad \mathcal{E}_2 = \{e_3, e_4, e_6, e_7, e_9, e_{10}\}\,, \qquad \mathcal{U}_2 = \{o\}\,, \tag{3.23}$$

where $e_7 = (b, o)$ and $e_{10} = (d, b)$. At this point, the algorithm halts returning the feasible flow $f = (2, 2, 0, 2, 2)$ of maximal throughput 4.

It is interesting to check what would have happened, had a different choice of the initial o-d path been made. If one chooses $\gamma^{(0)} = (o, a, b, d)$, then

$$\varepsilon_0 = \min\{c_{e_1}, c_{e_3}, c_{e_5}\} = 2\,, \qquad f^{(1)} = f^{(0)} + 2(\delta^{(e_1)} + \delta^{(e_3)} + \delta^{(e_5)}) = (2,0,2,0,2)\,,$$

so that, with $e_6 = (a, o)$, $e_8 = (b, a)$ and $e_{10} = (d, b)$, we have

$$c^{(1)} = c - f^{(1)} = (0,2,1,4,0)\,, \qquad \mathcal{E}_1 = \{e_2, e_3, e_4, e_6, e_8, e_{10}\}\,, \qquad \mathcal{U}_1 = \mathcal{V}\,.$$

In the next step, we then have to take $\gamma^{(1)} = (o, b, a, d)$, as this is the only o-d path in the residual graph $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$, so that

$$\varepsilon_1 = \min\{c^{(1)}_{e_2}, f^{(1)}_{e_3}, c^{(1)}_{e_4}\} = 2\,, \qquad f^{(2)} = f^{(1)} + 2(\delta^{(e_2)} - \delta^{(e_3)} + \delta^{(e_4)}) = (2,2,0,2,2)\,,$$

and (3.23) holds true, so that the algorithm halts returning the feasible flow $f = (2, 2, 0, 2, 2)$ of maximal throughput 4, exactly as in the other case.

3.4 Matchings

Recall from Section 1.2 that a matching in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a subset of links $\mathcal{F} \subseteq \mathcal{E}$ such that every node is either the tail or the head node of at most one link in $\mathcal{M}$, i.e., if no two links in $\mathcal{M}$ share a common (tail or head) node. In this section, we focus on matchings in bipartite graphs.

Consider a simple bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, whereby the node set can be partitioned as $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$ with $\mathcal{V}_0 \cap \mathcal{V}_1 = \emptyset$ and no link e connecting two nodes that are either both in $\mathcal{V}_0$ or both on $\mathcal{V}_1$. For $h = 0, 1$, we shall refer to a matching $\mathcal{M}$ in $\mathcal{G}$ as $\mathcal{V}_h$-perfect (or $\mathcal{V}_h$-saturating) if every node i in $\mathcal{V}_h$ is matched in $\mathcal{M}$. The following classical result is known as Hall's Marriage Theorem.

Theorem 3.3. For a simple bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{V}_0 \subseteq \mathcal{V}$, there exists a $\mathcal{V}_0$-perfect matching in $\mathcal{G}$ if and only if

$$|\mathcal{U}| \leq |\mathcal{N}_\mathcal{U}|\,, \qquad \forall \mathcal{U} \subseteq \mathcal{V}_0 \tag{3.24}$$

where $\mathcal{N}_\mathcal{U} = \bigcup_{i \in \mathcal{U}} \mathcal{N}_i$ is the neighborhood of $\mathcal{U}$ in $\mathcal{G}$.

Proof. Let $\mathcal{V}_1 = \mathcal{V} \setminus \mathcal{V}_0$ and consider a capacitated directed multi-graph $\overline{\mathcal{G}} = (\overline{\mathcal{V}}, \mathcal{F}, c)$ with node set $\overline{\mathcal{V}} = \{o\} \cup \mathcal{V} \cup \{d\}$ and link set $\mathcal{F}$ of cardinality $|\mathcal{F}| = |\mathcal{V}| + |\mathcal{E}|/2$ such that

- for every node i in $\mathcal{V}_0$ there exists a link e in $\mathcal{F}$ with $\theta(e) = o$ and $\kappa(e) = i$ and capacity $c_e = 1$;

- for every node i in $\mathcal{V}_1$ there exists a link e in $\mathcal{F}$ with $\theta(e) = j$ and $\kappa(e) = d$ and capacity $c_e = 1$;
- for every link (i, j) in $\mathcal{E}$ from node i in $\mathcal{V}_0$ to node j in $\mathcal{V}_1$, there exists a link e in $\mathcal{F}$ with $\theta(e) = i$ and $\kappa(e) = j$ and capacity $c_e = +\infty$.

Now, we shall compute the min-cut capacity $c^*_{o,d}$ from node o to node d in $\overline{\mathcal{G}}$. For this, consider an arbitrary o-d cut $\{o\} \cup \mathcal{S}$. Let $\mathcal{U} = \mathcal{U} \cup \mathcal{V}_0$ and let $\mathcal{N}^+_{\mathcal{U}} = \bigcup_{i \in \mathcal{V}_0} \{j \in \mathcal{V}_1 : (i, j) \in \mathcal{E}\}$ be its out-neighborhood. Observe that

$$c_{\mathcal{S}} = \begin{cases} +\infty & \text{if} \quad \mathcal{N}^+_{\mathcal{U}} \not\subseteq \mathcal{S} \\ |\mathcal{V}_0 \setminus \mathcal{U}| + |\mathcal{S} \cap \mathcal{V}_1| \geq |\mathcal{V}_0| - |\mathcal{U}| + |\mathcal{N}^+_{\mathcal{U}}| & \text{if} \quad \mathcal{N}^+_{\mathcal{U}} \subseteq \mathcal{S} \end{cases}$$

It follows that the min-cut capacity from o to d in $\overline{\mathcal{G}}$ is

$$c^*_{o,d} = \min_{\substack{\mathcal{S} \\ o-d \text{ cut}}} c_{\mathcal{S}} = \min_{\mathcal{U} \subseteq \mathcal{V}_0} \{|\mathcal{V}_0| - |\mathcal{U}| + |\mathcal{N}^+_{\mathcal{U}}|\} = |\mathcal{V}_0| + \min_{\mathcal{U} \subseteq \mathcal{V}_0} \{|\mathcal{N}^+_{\mathcal{U}}| - |\mathcal{U}|\}. \quad (3.25)$$

It then follows that $c^*_{o,d} \leq |\mathcal{V}_0|$ with equality if and only if (3.24) holds true.

At this point, it is sufficient to observe that matchings in $\mathcal{G}$ are in one-to-one correspondence with feasible integer o-d flows in the capacitated multigraph $\overline{\mathcal{G}}$. Then, Theorem 3.2 and the discussion above imply that the cardinality of a maximum matching is given by $c^*_{o,d}$ as in (3.25), so that there exists a $\mathcal{V}_0$-perfect matching if and only if (3.24) holds true. □

Chapter 4

Network flow optimization

In this chapter, we study network flow optimization. More specifically, we focus on convex separable network flow optimization problems whereby, for a given multigraph, a vector of exogenous net flows, and an assignment of convex increasing scalar cost functions for each link, we are interested in minimizing the aggregate cost over all possible network flows. As we shall see, this problem boils down to the minimization of a convex separable function with linear equality constraints corresponding to mass conservation at the nodes and linear inequality constraints corresponding to nonnegativity of the flow variables.

First, we show how to use convex optimization techniques to get insight into the problem and the structure of optimal flows. In particular, we introduce Lagrange multipliers and show their role in deriving necessary and sufficient conditions for optimality as a well as in the sensitivity analysis of the optimal cost with respect to the exogenous flows.

Then, we illustrate how these results prove useful in a number of key special cases including dissipative flow networks (electrical resistors, water, and gas) and traffic networks. In particular, for electrical networks, we show how these analysis can be used to prove the variational characterization.

On the other hand, for traffic networks, we introduce the notions of system user optimality, study optimal traffic assignments, Wardrop equilibria and discuss the Braess paradox, the notion of price of anarchy, and the use of (marginal cost) tolls to align the user optimal flow to the system optimal one.

4.1 Convex separable network flow optimization

We shall study network flows on a directed multi-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$, with node-link incidence matrix B. Recall from Chapter 3 that, given a zero-sum vector ν in $\mathbb{R}^{\mathcal{V}}$ of exogenous net flows, a network flow on $\mathcal{G}$ is a vector f in $\mathbb{R}^{\mathcal{E}}$ that satisfies the non-negativity and conservation of mass constraints

$$f \geq 0\,, \qquad Bf = \nu\,. \tag{4.1}$$

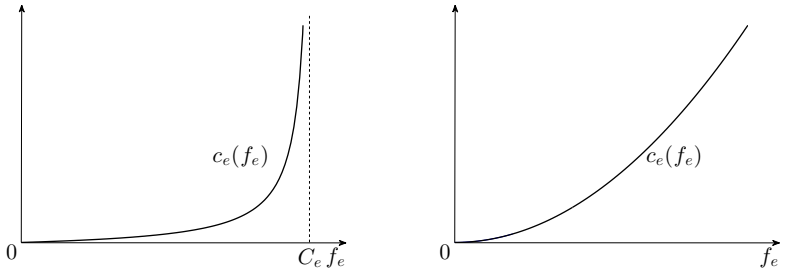


Figure 4.1: Two link flow cost functions $\psi_e(f_e)$ with finite (left) and infinite (right) flow capacities c_e.

We will be concerned with the problem of selecting a network flow f^* from the set of vectors satisfying (4.1) that minimizes a separable convex cost function. Specifically, we will consider link cost functions $\psi_e(f_e)$, for e in $\mathcal{E}$, describing the cost $\psi_e(f_e)$ incurred when a flow $f_e \geq 0$ passes through link e. We will make the following assumptions on such cost functions. (See Figure 4.1.)

Assumption 1. *For every link e in $\mathcal{E}$, the cost function*

$$\psi_e : [0, +\infty) \to [0, +\infty]$$

is such that $\psi_e(0) = 0$, non-decreasing, continuously differentiable and convex in the interval $[0, c_e)$ where

$$c_e = \sup\{x \geq 0 : \ \psi_e(x) < +\infty\}$$

is referred to as the flow capacity *of link e.*

Some considerations about Assumption 1 are in order. First, the requirements that $\psi_e(0) = 0$, i.e., there is no cost for sending no flow, and that $\psi_e(x)$ is non-decreasing, i.e., the more the flow the higher the cost incurred, are natural. Second, allowing for $\psi_e(x)$ to possibly take value $+\infty$ allows us to treat at once both the case of infinite link flow capacity, i.e., when $c_e = +\infty$ and $\psi_e(x) < +\infty$ for every $x \geq 0$, and of finite link flow capacity, i.e., when $c_e < +\infty$ and $\psi_e(x) = +\infty$ for every $x \geq c_e$. Third, convexity of the cost function $\psi_e(x)$ is equivalent to that its derivative $\psi'_e(x)$ be non-decreasing in x. Such derivative $\psi'_e(x)$ can be interpreted as the marginal cost, i.e., the infinitesimal change of the cost incurred when an infinitesimal unit of flow is added to the quantity x already on the link. Hence, convexity of the cost function is equivalent to that the marginal cost $\psi'_e(x)$ is non is nondecreasing in the flow x on the link e, which is also quite a natural and common assumption. Finally, the assumption of differentiability is adopted since, on the one hand it simplifies the derivations and exposition, and on the other hand is satisfied by all the main examples we will deal with. It is worth nonetheless noting that this assumption can be relaxed

at the expense of the introduction of further technicalities, see, e.g., [4][Chapter 9].

We will then be concerned with the following convex semparable network flow optimization problem [2, 4, 38]

$$M(\nu) := \inf_{\substack{f \in \mathbb{R}_+^{\mathcal{E}} \\ Bf = \nu}} \sum_{e \in \mathcal{E}} \psi_e(f_e)\,. \tag{4.2}$$

When $M(\nu) = +\infty$ we shall refer to the network flow optimization problem (4.2) as *unfeasible*, when $M(\nu) < +\infty$, we will refer to it as *feasible*.

The following result provides necessary and sufficient conditions for feasibility in terms of cut capacities of the network.

Proposition 4.1. The network flow optimization problem (4.2) on a multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$, with flow capacities c_e for every link e in $\mathcal{E}$, is feasible if and only if

$$\sum_{i \in \mathcal{U}} \nu_i < \sum_{\substack{e \in \mathcal{E}: \\ \theta(e) \in \mathcal{U} \\ \kappa(e) \in \mathcal{V} \setminus \mathcal{U}}} c_e\,, \tag{4.3}$$

for every subset of nodes $\mathcal{U} \subseteq \mathcal{V}$ such that $\sum_{i \in \mathcal{U}} \nu_i > 0$.

Proof. If problem (4.2) is feasible, then there exists a flow vector f in $\mathbb{R}_+^{\mathcal{E}}$ such that $Bf = \nu$ and $f_e < c_e$ for every link e in $\mathcal{E}$. For every $\mathcal{U} \subseteq \mathcal{V}$, conservation of mass and flow nonnegativity imply that

$$\sum_{i \in \mathcal{U}} \nu_i = \sum_{i \in \mathcal{U}} (Bf)_i = \sum_{\substack{e \in \mathcal{E}: \\ \theta(e) \in \mathcal{U} \\ \kappa(e) \in \mathcal{V} \setminus \mathcal{U}}} f_e - \sum_{\substack{e \in \mathcal{E}: \\ \kappa(e) \in \mathcal{U} \\ \theta(e) \in \mathcal{V} \setminus \mathcal{U}}} f_e \leq \sum_{\substack{e \in \mathcal{E}: \\ \theta(e) \in \mathcal{U} \\ \kappa(e) \in \mathcal{V} \setminus \mathcal{U}}} f_e\,.$$

If $\sum_{i \in \mathcal{U}} \nu_i > 0$, then the summation in the rightmost side of the above cannot be empty and each of its addends is strictly less than the link capacity c_e, thus implying (4.3).

Proving sufficiency of (4.3) entails a careful generalization of the max-flow min-cut theorem to single-commodity network flows with multiple origins and destinations. For a proof, see, e.g., [11]. □

4.2 Lagrangian techniques and duality

In this section, we shall prove necessary and sufficient conditions for optimal network flows, i.e., for solutions of (4.2). In order to proceed, we introduce the Lagrangian of the optimization problem (4.2) that is the function

$$L : \mathbb{R}_+^{\mathcal{E}} \times \mathbb{R}^{\mathcal{V}} \times \mathbb{R}^{\mathcal{V}} \to \mathbb{R}\,,$$

$$L(f,\lambda,\nu)=\sum_{e\in\mathcal{E}}\psi_e(f_e)+\sum_{i\in\mathcal{V}}\lambda_i\left(\sum_{\substack{e\in\mathcal{E}:\\ \kappa(e)=i}}f_e-\sum_{\substack{e\in\mathcal{E}:\\ \theta(e)=i}}f_e+\nu_i\right). \tag{4.4}$$

In the expression above, f in $\mathbb{R}_+^{\mathcal{E}}$ is a nonnegative vector whose entries f_e represent the flow on the different links e in $\mathcal{E}$, ν in $\mathbb{R}^{\mathcal{V}}$ is the exogenous flow vector, while λ in $\mathbb{R}^{\mathcal{V}}$ is a vector whose entries λ_i are the Lagrange multipliers associated with the conservation of mass equation in the nodes node i. Observe that, if f in $\mathbb{R}_+^{\mathcal{E}}$ is a flow vector, i.e., if it satisfies the flow balance constraints $Bf=\nu$, then, $L(f,\lambda,\nu)=\sum_e\psi_e(f_e)$. Notice that we have not introduced Lagrange multipliers for the nonnegativity constraints, as we will keep treating such constraints explicitly.

We also define the *dual function* $D:\mathbb{R}^{\mathcal{V}}\times\mathbb{R}^{\mathcal{V}}\to\mathbb{R}$

$$D(\lambda,\nu):=\inf_{f\in\mathbb{R}_+^{\mathcal{E}}}L(f,\lambda,\nu)\,. \tag{4.5}$$

The *dual problem* of the optimization problem (4.2) (which is referred to as the *primal problem*) is then

$$M^*(\nu)=\sup_{\lambda\in\mathbb{R}^{\mathcal{V}}}D(\lambda,\nu)\,. \tag{4.6}$$

Observe that by simply rearranging terms in the righthand side of (4.4), we get the equivalent representation

$$L(f,\lambda,\nu)=\sum_{e\in\mathcal{E}}\big(\psi_e(f_e)-f_e(\lambda_{\theta(e)}-\lambda_{\kappa(e)})\big)+\sum_{i\in\mathcal{V}}\lambda_i\cdot\nu_i\,. \tag{4.7}$$

What makes equation (4.7) particularly interesting is the fact that its righthand side is a separable function of the flows f_e on the different links e in $\mathcal{E}$. As a result, for every arbitrary assignment of the multipliers λ_i in the nodes i in $\mathcal{V}$, the minimization in the right-hand side of (4.5) gets decoupled in the following single-dimensional maximization problems:

$$\psi_e^*(y_e)=\sup_{f_e\geq 0}\left\{y_ef_e-\psi_e(f_e)\right\},\qquad y_e=\lambda_{\theta(e)}-\lambda_{\kappa(e)}\,,\quad e\in\mathcal{E}\,. \tag{4.8}$$

The quantity in the curly brackets in (4.8) may be interpreted as the net profit that a hypothetic link operator would make by accepting a flow f_e for which he charges a fee of y_e per unit and that has an aggregate cost $\psi_e(f_e)$ for him to transport, so that $\mathrm{cot}_e^*(y_e)$ represents the supremum of all profits that such operator could possibly make. Notice that the dual cost functions $\psi_e^*(y_e)$ are convex regardless from whether the original cost functions $\psi_e(f_e)$ are, as they are defined in (4.8) as the supremum of linear functions. On the other hand, the dual function defined in (4.5) is concave as it can be rewritten as

$$D(\lambda,\nu)=\sum_{i\in\mathcal{V}}\lambda_i\cdot\nu_i-\sum_{e\in\mathcal{E}}\psi_e^*(\lambda_{\theta(e)}-\lambda_{\kappa(e)})\,.$$

The following simple result provides necessary and sufficient conditions for the flow f_e on a link e to minimize the quantity between the parentheses in the righthand side of (4.7).

Lemma 4.1. For every link e in $\mathcal{E}$ with associated cost function satisfying Assumption 1 and every pair of Lagrange multiplies $(\lambda_{\theta(e)}, \lambda_{\kappa(e)})$ in $\mathbb{R}^2$,

$$f_e^* \in \operatorname*{argmin}_{0 \le f_e < c_e} \left\{ \psi_e(f_e) - (\lambda_{\theta(e)} - \lambda_{\kappa(e)}) f_e \right\} , \tag{4.9}$$

if and only if

$$\psi_e'(f_e^*) \ge \lambda_{\theta(e)} - \lambda_{\kappa(e)} , \qquad f_e^* \left(\psi_e'(f_e^*) - (\lambda_{\theta(e)} - \lambda_{\kappa(e)}) \right) = 0 . \tag{4.10}$$

Moreover, if the cost $\psi_e(x)$ is strictly convex on the interval $[0, c_e)$, and

$$\lambda_{\theta(e)} - \lambda_{\kappa(e)} < \sup\{\psi_e'(x) : x \in [0, c_e)\} ,$$

then the minimizer in (4.9) is unique and given by

$$f_e^* = \begin{cases} 0 & \text{if} \quad \lambda_{\theta(e)} - \lambda_{\kappa(e)} \le \psi_e'(0) \\ (\psi_e')^{-1}(\lambda_{\theta(e)} - \lambda_{\kappa(e)}) & \text{if} \quad \lambda_{\theta(e)} - \lambda_{\kappa(e)} > \psi_e'(0) . \end{cases} \tag{4.11}$$

Proof. Let $a := \lambda_{\theta(e)} - \lambda_{\kappa(e)}$ and consider the function $g(x) = \psi_e(x) - ax$ for $0 \le x < c_e$. Clearly, $g_e(x)$ is convex on the interval $[0, c_e)$. Hence, its minimum value is achieved: (i) in an internal point x^* in $(0, c_e)$ if and only if $g'(x^*) = \psi_e'(x^*) - a = 0$; (ii) in the left extreme of the interval $x^* = 0$ if and only if $g_e'(0) = \psi_e'(0) - a \ge 0$. By combining (i) and (ii) , one gets that $x^* \in \operatorname{argmin}\{g(x) : x \in [0, c_e)\}$ if and only if $\psi_e'(x^*) \ge a$ and $x^*(\psi_e'(x^*) - a) = 0$, thus proving the first part of the claim. For the second part, let $\overline{a} = \sup\{\psi_e'(x) : x \in [0, c_e)\}$. Then, observe that when the cost function $\psi_e(x)$ is strictly convex, its derivative $\psi_e'(x)$ is strictly increasing, hence invertible in the interval $[\psi_e'(0), \overline{a})$. Moreover, in this case, the function $g(x)$ is also strictly convex and its unique minimizer x^* is equal to 0 whenever $a \le \psi_e'(0)$ and equal to $(\psi_e')^{-1}(a)$ for every a in $(\psi_e'(0), \overline{a})$. □

Observe that (4.10) is equivalent to that $\psi_e'(f_e^*) \ge \lambda_{\theta(e)} - \lambda_{\kappa(e)}$ and $\psi_e'(f_e^*) = \lambda_{\theta(e)} - \lambda_{\kappa(e)}$ whenever $f_e^* > 0$. We shall refer to (4.10) as the *complementary slackness* conditions for optimality. As suggested by Lemma 4.1, given link costs $\psi_e(f_e)$ satisfying Assumption 1 and a vector of Lagrange multipliers λ in $\mathbb{R}^{\mathcal{V}}$, it is quite straightforward to solve the complementary slackness solutions separately on each link and find a nonnegative vector $f^*(\lambda)$ that satisfies (4.10) on every link e in $\mathcal{E}$. However, in general such vector $f^*(\lambda)$ will not be a feasible flow vector as typically $Bf^*(\lambda) \ne \nu$. The following result shows that, if one is able to find a vector of Lagrange multipliers λ^* in $\mathbb{R}^{\mathcal{V}}$ and a corresponding nonnegative vector $f^*(\lambda^*)$ in $\mathbb{R}_+^{\mathcal{E}}$ that together satisfy both the complementary slackness conditions and the flow balance constraints $Bf^*(\lambda^*) = \nu$, then $f^*(\lambda^*)$ would be an optimal solution of the convex separable network flow optimization problem (4.2).

Proposition 4.2. A feasible flow vector f^* in $\mathbb{R}_+^{\mathcal{E}}$ such that $Bf^* = \nu$ and a vector λ^* in $\mathbb{R}^{\mathcal{V}}$ of Lagrange multipliers satisfy the complementary slackness conditions (4.10) on every link e in $\mathcal{E}$ if and only if f^* is an optimal solution of (4.2), λ^* is an optimal solution of (4.6), and $M(\nu) = M^*(\nu)$.

Proof. We shall prove the "only if" part of the statement and refer to [4][Proposition 9.2] for the converse implication. First, observe that, for every feasible flow vector f in $\mathbb{R}_+^{\mathcal{E}}$ such that $Bf = \nu$ and every vector λ in $\mathbb{R}^{\mathcal{V}}$ of Lagrange multipliers, we have

$$D(\lambda, \nu) = \inf_{\tilde{f} \in \mathbb{R}_+^{\mathcal{E}}} L(\tilde{f}, \lambda, \nu) \leq L(f, \lambda, \nu) = \sum_{e \in \mathcal{E}} \psi_e(f_e)\,. \tag{4.12}$$

Now, if f^* in $\mathbb{R}_+^{\mathcal{E}}$ is a feasible flow vector and λ^* in $\mathbb{R}^{\mathcal{V}}$ a vector of Lagrange multipliers that together satisfy the complementary slackness conditions (4.10) on every link e in $\mathcal{E}$, we have that

$$\begin{aligned} D(\lambda^*, \nu) &= \inf_{f \in \mathbb{R}_+^{\mathcal{E}}} L(f, \lambda^*, \nu) \\ &= \sum_{e \in \mathcal{E}} \min_{f_e \in \mathbb{R}_+^{\mathcal{E}}} \Big(\psi_e(f_e) - f_e(\lambda^*_{\theta(e)} - \lambda^*_{\kappa(e)}) \Big) + \sum_{i \in \mathcal{V}} \lambda_i^* \cdot \nu_i \\ &= \sum_{e \in \mathcal{E}} \Big(\psi_e(f_e^*) - f_e^*(\lambda^*_{\theta(e)} - \lambda^*_{\kappa(e)}) \Big) + \sum_{i \in \mathcal{V}} \lambda_i^* \cdot \nu_i \\ &= L(f^*, \lambda^*, \nu) \\ &= \sum_{e \in \mathcal{E}} \psi_e(f_e^*)\,, \end{aligned} \tag{4.13}$$

where the first equality follows from the definition (4.5) of the dual function, the second and the forth ones from (4.7), the third one from Lemma 4.1 and the fact that f^* and λ^* together satisfy the complementary slackness conditions, and the last one from the fact that a feasible flow vector f^* satisfies the flow balance constraints $Bf^* = \nu$. Then, (4.13) together with (4.12) with $\lambda = \lambda^*$ imply that

$$\sum_{e \in \mathcal{E}} \psi_e(f_e^*) = D(\lambda^*, \nu) \leq \sum_{e \in \mathcal{E}} \psi_e(f_e)\,, \qquad \forall f \in \mathbb{R}_+^{\mathcal{E}} \text{ s.t. } Bf = \nu\,,$$

so that the feasible flow f^* is necessarily an optimal solution of (4.2). On the other hand, equation (4.13) together with (4.12) with $f = f^*$ imply that

$$D(\lambda^*, \nu) = \sum_{e \in \mathcal{E}} \psi_e(f_e^*) \geq D(\lambda, \nu)\,, \qquad \forall \lambda \in \mathbb{R}^{\mathcal{V}}\,,$$

so that λ^* is an optimal solution of (4.6). Then, (4.13) also implies that

$$M(\nu) = \sum_{e \in \mathcal{E}} \psi_e(f_e^*) = D(\lambda^*, \nu) = M^*(\nu)\,,$$

i.e., the optimal primal value and the optimal dual value are the same. □

Proposition 4.2 implies that, if one is able to find a vector of Lagrange multipliers λ^* such that the corresponding vector $f^*(\lambda^*)$ solving the complementary slackness conditions together with λ^* is a feasible flow vector, then the vector $f^*(\lambda^*)$ is guaranteed to be an optimal solution of the convex separable network flow optimization problem (4.2). E.g., in the case of strictly convex link costs, this can be used by imposing the conservation of mass constraints $Bf^*(\lambda) = \nu$ to the λ-dependent vector with entries as in (4.11) thus obtaining a system of n (in general nonlinear) equations in n unkwnowns λ_i, at most $n-1$ of which are linearly independent: in fact, observe also that a solution λ^* remains so if any constant vector is added to it. Alternatively, one may seek the vector of Lagrange multipliers λ^* as the solution the dual optimization problem (4.6), which turns out to be an unconstrained convex optimization (once again, observe that adding a constant vector to λ does not change the value of $D(\lambda)$).

In fact, it turns out that the optimal values of the Lagrange multipliers λ^* can be interpreted as the *marginal costs*, i.e., the change rates of the minimal cost $M(\nu)$ corresponding to changes of the exogenous net flow vector ν. This is formalized in the last point of the following result which also gathers the conclusions achieved thus far in this section.

Theorem 4.1. Consider the network flow optimization problem (4.2) on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with cost functions satisfying Assumption 1. Then, for every vector of exogenous net flows ν in $\mathbb{R}^{\mathcal{V}}$ satisfying (4.3):

(i) the dual function (4.5) achieves a maximum value that coincides with the optimal cost, i.e.,

$$\max_{\lambda \in \mathbb{R}^{\mathcal{V}}} D(\lambda, \nu) = \min_{\substack{f \in \mathbb{R}_+^{\mathcal{E}} \\ Bf = \nu}} \sum_{e \in \mathcal{E}} \psi_e(f_e)\,, \tag{4.14}$$

(ii) a flow vector $f^* \in \mathbb{R}_+^{\mathcal{E}}$ and a vector $\lambda^* \in \mathbb{R}^{\mathcal{V}}$ are optimal solutions of the righthand and, respectively, the lefthand side of (4.14) if and only if they satisfy (4.10);

(iii) if the optimal cost $M(\nu)$ is differentiable in ν, then

$$\frac{\partial}{\partial \nu_i} M(\nu) - \frac{\partial}{\partial \nu_j} M(\nu) = \lambda_i^* - \lambda_j^*\,. \tag{4.15}$$

Proof. (i) It follows from Proposition 4.1 that, if (4.3) is satisfied, then there exists a feasibke flow vector $\overline{f}$ in $\mathbb{R}_+^{\mathcal{E}}$, such that $B\overline{f} = \nu$ and $\overline{c} = \sum_{e \in \mathcal{E}} \psi_e(\overline{f}_e) < +\infty$, so that the set

$$\mathcal{K} = \left\{ f \in \mathbb{R}_+^{\mathcal{E}} : Bf = \nu, \sum_{e \in \mathcal{E}} \psi_e(f_e) \leq \overline{c} \right\}$$

is nonempty. In fact, since the cost functions $\psi_e(f_e)$ are convex and continuous in $[0, c_e)$, the set $\mathcal{K}$ is convex and compact, so that there exists an optimal

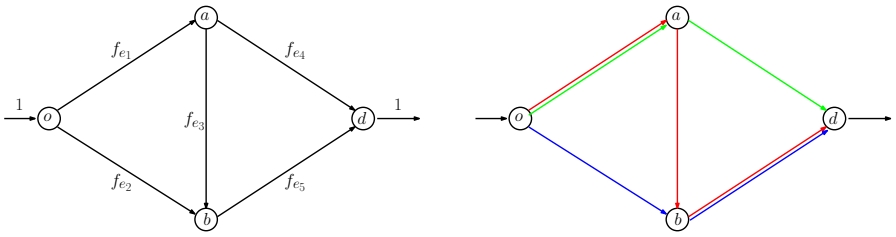


Figure 4.2: A network with node set $\mathcal{V} = \{o, a, b, d\}$ and link set $\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$. There are three distinct o-d paths: $p^{(1)} = (o, a, d)$ (green); $p^{(2)} = (o, b, d)$ (blue); and $p^{(3)} = (o, a, b, d)$ (red). Hence $\Gamma_{o,d} = \{p^{(1)}, p^{(2)}, p^{(3)}\}$. The node-link incidence matrix B and the link-path incidence matrix A are reported in equation (3.10).

solution f^* in $\mathcal{K}$ of optimal control problem (4.2). Then, [4, Proposition 9.3] guarantees that there exists a vector of Lagrange multipliers λ^* in $\mathbb{R}^{\mathcal{V}}$ that satisfies (4.10) together with f^*. Proposition 4.2 then implies point (ii).

(ii) This follows from Proposition 4.2 .

(iii) It follows from (4.14) that $M(\nu)$ is convex in ν, as it is the maximum, over all possible values of the Lagrange multipliers λ, of the dual function $D(\lambda, \nu)$ that is linear in ν. Moreover, (4.14) implies that for every $\tilde{\nu}$ such that $\mathbb{1}'\tilde{\nu} = 0$,

$$M(\tilde{\nu}) = \max_{\lambda} D(\lambda, \tilde{\nu}) \geq D(\lambda^*, \tilde{\nu}) = D(\lambda^*, \nu) + \sum_i \lambda_i^* (\tilde{\nu}_i - \nu_i) = M(\nu) + \sum_i \lambda_i^* (\tilde{\nu}_i - \nu_i) \,.$$

Then, for every two nodes i and j, taking $\tilde{\nu} = \nu + \varepsilon(\delta^{(i)} - \delta^{(j)})$ in the above, with $\varepsilon > 0$ and $\varepsilon < 0$, respectively, implies that

$$\liminf_{\varepsilon \downarrow 0} \frac{M\left(\nu + \varepsilon(\delta^{(i)} - \delta^{(j)})\right) - M(\nu)}{\varepsilon} \geq \lambda_i^* - \lambda_j^* \geq \limsup_{\varepsilon \uparrow 0} \frac{M\left(\nu + \varepsilon(\delta^{(i)} - \delta^{(j)})\right) - M(\nu)}{\varepsilon} \,.$$

If $M(\nu)$ is differentiable in ν, then the two limits on the leftmost and the rightmost sides of the above coincide and are both equal to

$$\frac{\partial}{\partial \nu_i} M(\nu) - \frac{\partial}{\partial \nu_j} M(\nu) \,,$$

which implies (4.15). □

4.3 Shortest path and optimal transport

For a multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$, let $o \neq d$ be two nodes such that d is reachable from o and let $\Gamma_{o,d}$ be the set of o-d paths. Let each link e in $\mathcal{E}$ be assigned a positive weight l_e representing its physical length and a linear cost function

$$\psi_e(f_e) = l_e f_e \,. \tag{4.16}$$

Observe that the linear costs (4.16) do not account for congestion effects: the cost for a unit of flow is a constant l_e.

Now, consider the network flow optimization problem (4.2) where the exogenous net-flow vector is $\nu = \delta^{(o)} - \delta^{(d)}$, i.e., where we constrain ourselves to unitary flows from o to d. Clearly, the problem is always feasible since the link capacities are all infinite and we have assumed node d to be reachable from node o. Then, Theorem 4.1 (i) implies that a unitary o-d flow f^* is optimal if and only if there exist multipliers λ_i, for i in $\mathcal{V}$, such that

$$l_e \begin{cases} = \lambda_{\theta(e)} - \lambda_{\kappa(e)} & \text{if} \quad f_e^* > 0 \\ \geq \lambda_{\theta(e)} - \lambda_{\kappa(e)} & \text{if} \quad f_e^* = 0\,, \end{cases} \qquad \forall e \in \mathcal{E}\,.$$

This implies that, for every o-d path $\gamma = (e_1, e_2, \ldots, e_k)$,

$$\sum_{j=1}^{k} l_{e_j} \geq \sum_{j=1}^{k} (\lambda_{\theta(e_j)} - \lambda_{\kappa(e_j)}) = \lambda_o - \lambda_d\,, \tag{4.17}$$

with equality whenever $f_{e_j}^* > 0$ for every $1 \leq j \leq k$, where the second equality in (4.17) follows from the fact that $\theta(e_1) = o$, $\kappa(e_l) = d$, while $\kappa(e_{j-1}) = \theta(e_j)$ for every $1 \leq j \leq l$, so that all these terms cancel out. Equation (4.17) implies that every optimal flow f^* is supported only on the links along minimum length o-d paths. In particular, if the shortest o-d path is unique, then using the conservation of mass equation we find that, as expected, $f_e^* = 1$ on all links e along such path, and $f_e = 0$ on all other links.

The optimal transport problem is a generalization of the shortest path problem with linear costs (4.16), and arbitrary zero-sum net-flow ν in $\mathbb{R}^{\mathcal{V}}$.

4.4 System-Optimum Traffic Assignment

In this and in the following sections we shall be concerned with optimal traffic assignments in transportation networks. In this context, links e in $\mathcal{E}$ represent (portions of) roads and nodes represent junctions, while the cost function is defined in terms of a delay function $\tau_e(f_e)$ that returns the delay encountered by any user traversing that link as a function of the flow on that link. We shall assume that such delay functions satisfy the following assumption. (See Figure 4.3.)

Assumption 2. *For every link e in $\mathcal{E}$, the delay function*

$$\tau_e : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$$

is non-decreasing, twice continuously differentiable in the interval $[0, c_e)$, and such that $\tau_e(f_e) = +\infty$ for $f_e > c_e$ and

$$2\tau_e'(f_e) + f_e \tau_e''(f_e) \geq 0\,, \qquad 0 \leq f_e < c_e\,. \tag{4.18}$$

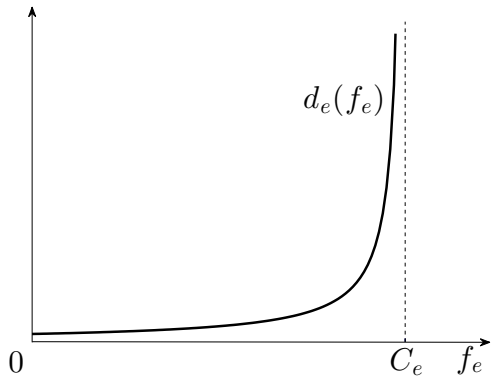


Figure 4.3: The delay function in (4.19).

Similarly to Assumption 1 for the cost functions, Assumption 2 allows us to treat both the case of finite and infinite link flow capacities. Observe that assuming that the delay is never negative is very natural as is assuming that the delay is non-decreasing in the flow: congestion can only slow down traffic, never speed it up. On the other hand, the assumption of twice continuous differentiability is done for simplicity and could be relaxed at the cost of more involved technicalities. Finally, the rationale for (4.18) (which immediately follows if, e.g., the delay functions are convex) will become apparent soon. A special case of delay function satisfying Assumption 2 is

$$\tau_e(f_e) = \frac{l_e}{1 - f_e/c_e}\,, \qquad f_e \in [0, c_e)\,, \tag{4.19}$$

where $l_e = \tau_e(0) > 0$ represents the delay on link e when it is empty, and $\tau_e(f_e) = +\infty$ for $f_e \geq c_e$. (See Figure 4.3.) In fact, the delay function in the form (4.19) can be given a microscopic justification in terms of an M/M/1 queue with service rate c_e and arrival rate f_e for which the expected sojourn time equals $1/(c_e - f_e)$, which coincides with the righthand side of (4.19) when $l_e = 1/c_e$.

Let delay functions τ_e satisfying Assumption 2 be assigned for every link of a multigraph describing the topology of a transportation network. The System-Optimum Traffic Assignment Problem (SO-TAP) consists in the network flow optimization problem (4.2) with link cost functions

$$\psi_e(f_e) = f_e \cdot \tau_e(f_e)\,, \qquad f_e \geq 0\,, \qquad e \in \mathcal{E}\,. \tag{4.20}$$

The interpretation is that if the flow on link e is f_e, then the delay is $\tau_e(f_e)$, and the cost represents the product of the delay times the flow, i.e., the total delay. Observe that

$$\psi_e'(f_e) = \tau_e(f_e) + f_e\tau_e'(f_e)\,, \qquad \psi_e''(f_e) = 2\tau_e'(f_e) + f_e\tau_e''(f_e)\,,$$

so that Assumption 2 on the delay functions $\tau_e(f_e)$ implies that the link costs (4.20) satisfy Assumption 1.

4.5 User optimal traffic assignment

In this section, we shall study the User-Optimum Traffic Assignment Problem (UO-TAP). As in Section 4.4, we shall interpret the nodes and links of a multigraph $\mathcal{G}$ as representing junctions and (portions of) roads, respectively, in a transportation network. We shall assume that each link e in $\mathcal{E}$ is equipped with delay functions $\tau_e(f_e)$ returning the delay experienced by any user traversing link e when the total flow through it is f_e. Such delay functions are assumed to be continuous and non-decreasing. Then, the UO-TAP consists in the network flow optimization problem (4.2) with link cost functions

$$\psi_e(f_e) = \int_0^{f_e} \tau_e(s)\mathrm{d}s\,, \qquad f_e \geq 0\,, \qquad e \in \mathcal{E}\,. \tag{4.21}$$

Observe that monotonicity and convexity of the link cost functions (4.21) follow simply from the fact that the delay functions are nonnegative-valued and non-decreasing, while differentiability of τ_e and (4.18) are not needed in this case. The rationale for the form of these cost functions will become apparent soon.

The idea is to model flows resulting not from a centralized optimization, but rather as the outcome of selfish behaviors of drivers. Such selfish behavior is modeled by assuming that drivers choose their route so as to minimize the delay they experience along it. This is formalized by the notion of Wardrop equilibrium that we introduce below.

First, given an origin destination pair $o \neq d$ in $\mathcal{V}$ such that d is reachable from o, recall that $\Gamma_{o,d}$ stands for the set of all o-d paths while the link-path incidence matrix is denoted as $A^{(o,d)}$. Now consider a nonnegative vector z in $\mathbb{R}^{\Gamma_{o,d}}$ such that $\mathbb{1}'z = \upsilon$, whose entries z_p represent the aggregate flow along the o-d path p and recall that, by the FLow Decomposition Theorem (Theorem 3.1),

$$f = A^{(o,d)}z$$

satisfies

$$f \geq 0\,, \qquad Bf = BA^{(o,d)}z = \upsilon(\delta^{(o)} - \delta^{(p)})\,,$$

i.e., it is a network flow from o to d of throughput υ.

Definition 4.1 (Wardrop equilibrium and price of anarchy [37, 33])**.** For a given throughput $\upsilon > 0$, a *Wardrop equilibrium* is a flow vector

$$f^{(0)} = A^{(o,d)}z$$

where z in $\mathbb{R}^{\Gamma_{o,d}}$ is such that $z \geq 0$, $\mathbb{1}'z = \upsilon$, and for every path p in $\Gamma_{o,d}$

$$z_p > 0 \quad \Longrightarrow \quad \underbrace{\sum_{e\in\mathcal{E}} A_{ep}^{(o,d)} \tau_e(f_e^{(0)})}_{\substack{\text{total delay} \\ \text{on path } p}} \quad \leq \quad \underbrace{\sum_{e\in\mathcal{E}} A_{eq}^{(o,d)} \tau_e(f_e^{(0)})}_{\substack{\text{total delay} \\ \text{on path } q}}, \qquad \forall q \in \Gamma_{o,d}\,. \tag{4.22}$$

The *price of anarchy* associated to a Wardrop equilibrium $f^{(0)}$ is

$$\text{PoA}(0) = \frac{\displaystyle\sum_{e\in\mathcal{E}} f_e^{(0)} \tau_e(f_e^{(0)})}{\displaystyle\min_{\substack{f\geq 0 \\ Bf = \upsilon(\delta^{(o)} - \delta^{(d)})}} \sum_{e\in\mathcal{E}} f_e \tau_e(f_e)}\,, \tag{4.23}$$

i.e., it is the ratio between the total delay at the Wardrop equilibrium and the minimum possible total delay.

The definition above can be interpreted as follows: a Wardrop equilibrium is a network flow vector $f^{(0)}$ that is associated with an o-d path distribution z such that, if $z_p > 0$ for some o-d path p, i.e., if some drivers choose p as their route from o to d, then the total delay associated with this path $\sum_{e\in\mathcal{E}} A_{ep}^{(o,d)} \tau_e(f_e^{(0)})$ cannot be worse than the total delay $\sum_{e\in\mathcal{E}} A_{eq}^{(o,d)} \tau_e(f_e^{(0)})$ associated with any other o-d path q. This should be interpreted as the result of a rational, and selfish, behavior of drivers: none of them would choose a sub-optimal route (i.e., an o-d path with larger delay) if a better route is available.

Example 4.1. Consider the network in Figure 4.2, having four nodes ($\mathcal{V} = \{o, a, b, d\}$), five links ($\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$), and three different o-d paths: $p^{(1)} = (o, a, d)$ (colored in green); $p^{(2)} = (o, b, d)$ (blue); $p^{(3)} = (o, a, b, d)$ (red). Let the throughput be $\upsilon = 1$ and the delay functions be given by

$$\tau_{e_1}(f_{e_1}) = f_{e_1}\,, \quad \tau_{e_2}(f_{e_2}) = 1\,, \quad \tau_{e_3}(f_{e_3}) = 0\,, \quad \tau_{e_4}(f_{e_4}) = 1 \quad \tau_{e_5}(f_{e_5}) = f_{e_5}\,.$$

Let $z = (z_{p^{(1)}}, z_{p^{(2)}}, z_{p^{(3)}})$, such that $z \geq 0$ and $z'\mathbb{1} = 1$, be a vector of flow assignments to o-d paths. Observe that $f = A^{(o,d)} z$ has entries

$$f_{e_1} = z_{p^{(1)}} + z_{p^{(3)}}\,, \qquad f_{e_2} = z_{p^{(2)}}\,, \qquad f_{e_3} = z_{p^{(3)}}\,, \qquad f_{e_4} = z_{p^{(1)}}\,, \qquad f_{e_5} = z_{p^{(2)}} + z_{p^{(3)}}\,.$$

The associated delays on the different links are then

$$\tau_{e_1}(f_{e_1}) = z_{p^{(1)}} + z_{p^{(3)}}\,, \quad \tau_{e_2}(f_{e_2}) = 1\,, \quad \tau_{e_3}(f_{e_2}) = 0\,, \quad \tau_{e_4}(f_{e_4}) = 1\,, \quad \tau_{e_5}(f_{e_5}) = z_{p^{(2)}} + z_{p^{(3)}}\,,$$

so that

$$\begin{aligned} \text{total delay on path } p^{(1)} &= \tau_{e_1}(f_{e_1}) + \tau_{e_4}(f_{e_4}) = z_{p^{(1)}} + z_{p^{(3)}} + 1 \\ \text{total delay on path } p^{(2)} &= \tau_{e_2}(f_{e_2}) + \tau_{e_5}(f_{e_5}) = z_{p^{(2)}} + z_{p^{(3)}} + 1 \end{aligned}$$

$$\begin{matrix}\text{total delay}\\ \text{on path } p^{(3)}\end{matrix} = \tau_{e_1}(f_{e_1}) + \tau_{e_3}(f_{e_3}) + \tau_{e_5}(f_{e_5}) = z_{p^{(1)}} + z_{p^{(2)}} + 2z_{p^{(3)}}\,.$$

Now assume that $z_{p^{(1)}} > 0$. Then, $z_{p^{(2)}} + z_{p^{(3)}} = 1 - z_{p^{(1)}} < 1$ so that

$$\begin{matrix}\text{total delay}\\ \text{on path } p^{(1)}\end{matrix} = z_{p^{(1)}} + z_{p^{(3)}} + 1 > z_{p^{(1)}} + z_{p^{(2)}} + 2z_{p^{(3)}} = \begin{matrix}\text{total delay}\\ \text{on path } p^{(3)}\end{matrix}\,.$$

This implies that, at Wardrop equilibrium, necessarily $z_{p^{(1)}} = 0$. Similarly, if $z_{p^{(2)}} > 0$, then $z_{p^{(1)}} + z_{p^{(3)}} = 1 - z_{p^{(2)}} < 1$ so that

$$\begin{matrix}\text{total delay}\\ \text{on path } p^{(2)}\end{matrix} = z_{p^{(2)}} + z_{p^{(3)}} + 1 > z_{p^{(1)}} + z_{p^{(2)}} + 2z_{p^{(3)}} = \begin{matrix}\text{total delay}\\ \text{on path } p^{(3)}\end{matrix}\,.$$

Hence, at Wardrop equilibrium, necessarily $z_{p^{(2)}} = 0$. The only possibility remaining is that all flow in on path $p^{(3)}$, i.e., $z_{p^{(1)}} = z_{p^{(2)}} = 0$ and $z_{p^{(3)}} = 1$. In this case, one gets that

$$\begin{matrix}\text{total delay}\\ \text{on path } p^{(1)}\end{matrix} = \begin{matrix}\text{total delay}\\ \text{on path } p^{(2)}\end{matrix} = \begin{matrix}\text{total delay}\\ \text{on path } p^{(3)}\end{matrix} = 2\,,$$

so that $f^{(0)} = A^{(o,d)}z$ with entries

$$f_{e_1} = 1\,, \qquad f_{e_2} = 0\,, \qquad f_{e_3} = 1\,, \qquad f_{e_4} = 0\,, \qquad f_{e_5} = 1$$

is a Wardrop equilibrium flow. Observe that, at such Wardrop equilibrium, the total delay experienced by all users is 2. This contrasts the total delay that every user would incur in case of an socially optimal assignment $z_{p^{(1)}} = z_{p^{(2)}} = 1/2$, $z_{p^{(3)}} = 0$, with corresponding links flows $f^*_{e_1} = f^*_{e_2} = f^*_{e_4} = f^*_{e_5} = 1/2$, $f^*_{e_3} = 0$, whereby every user would have incurred a total delay of 1.5. Thus the *price of anarchy* [33] in this example is

$$\text{PoA}(0) = \frac{\begin{matrix}\text{total delay at}\\ \text{user optimum}\end{matrix}}{\begin{matrix}\text{total delay at}\\ \text{social optimum}\end{matrix}} = \frac{\displaystyle\sum_{e\in\mathcal{E}} f_e^{(0)}\tau_e(f_e^{(0)})}{\displaystyle\sum_{e\in\mathcal{E}} f_e^{*}\tau_e(f_e^{*})} = \frac{2}{1.5} = \frac{4}{3}\,.$$

Observe that, if the link e_3 had been removed from the network, the system optimum network flow for the resulting network would be exactly the same f^* as with link e_3 present, since it does not use link e_3. On the other hand, removing link e_3 would change the Wardrop equilibrium flow drastically, making it coincide with the system optimimum network flow f^* and thus reducing the price of anarchy to 1. This counterintuitive phenomenon that the removal of a link can improve the system performance at the Wardrop equilibrium is known as the Braess paradox [9]. On the other hand, observe that the removal of a link can never improve the system performance at the system optimum. □

We wish to emphasize once more that the definition of a Wardrop equilibrium captures the rational and selfish behavior of drivers that minimize their own

delay. One way to influence such behavior is to introduce tolls $\omega_e \geq 0$ that are charged to users traversing link e in $\mathcal{E}$ and to assume that the cost perceived by any user for traversing link e is

$$\omega_e + \tau_e(f_e)\,, \tag{4.24}$$

i.e., the sum of the toll charged and the delay incurred on that link.[1] It is then natural to define the Wardrop equilibrium $f^{(\omega)}$ associated with the vector $\omega = (\omega_e)_{e\in\mathcal{E}}$ of link tolls generalizing Definition 4.1 as follows.

Definition 4.2 (Wardrop equilibrium with tolls)**.** For a given throughput $\upsilon > 0$, and a vector of tolls $\omega \in \mathbb{R}^{\mathcal{E}}_+$, a Wardrop equilibrium is a flow vector

$$f^{(\omega)} = A^{(o,d)} z$$

where $z \in \mathbb{R}^{\Gamma_{o,d}}$ is such that $z \geq 0$, $\mathbb{1}'z = \upsilon$, and for every path $p \in \Gamma_{o,d}$

$$z_p > 0 \qquad \Longrightarrow \qquad \underbrace{\sum_{e\in\mathcal{E}} A^{(o,d)}_{ep}\left(\tau_e(f^{(\omega)}_e) + \omega_e\right)}_{\substack{\text{total perceived cost}\\ \text{on path } p}} \quad \leq \quad \underbrace{\sum_{e\in\mathcal{E}} A^{(o,d)}_{eq}\left(\tau_e(f^{(\omega)}_e) + \omega_e\right)}_{\substack{\text{total perceived cost}\\ \text{on path } q}}, \qquad \forall q \in \Gamma_{o,d}\,. \tag{4.25}$$

The *price of anarchy* associated to a Wardrop equilibrium $f^{(\omega)}$ is

$$\mathrm{PoA}(\omega) = \frac{\displaystyle\sum_{e\in\mathcal{E}} f^{(\omega)}_e \tau_e(f^{(\omega)}_e)}{\displaystyle\min_{\substack{f\geq 0\\ Bf = \upsilon(\delta^{(o)} - \delta^{(d)})}} \sum_{e\in\mathcal{E}} f_e\tau_e(f_e)}\,. \tag{4.26}$$

While the Wardrop equilibrium (both with or without tolls) captures a user-perspective notion of optimization, it turns out that we can always interpret it as an optimal network flow, provided that we consider proper costs on the links. Specifically, let

$$D_e(f_e) := \int_0^{f_e} \tau_e(s)\mathrm{d}s\,, \qquad e \in \mathcal{E}\,, \tag{4.27}$$

be a primitive of the delay function on link e. Observe that, unless the delays are constant $\tau_e(f_e) = l_e$, the primitive of the delay functions (4.27) differ from the costs (**??**) we defined in Example **??** for the SO-TAP.

Theorem 4.2. For a multigraph $\mathcal{G} = (\mathcal{G}, \mathcal{E}, \theta, \kappa)$, let $o \neq d$ be two nodes such that d is reachable from o. Let each link e be equipped with a nondecreasing differentiable delay function τ_e, such that every cycle in $\mathcal{G}$ contains a link e with $\tau_e(0) > 0$. Let ω in $(0, c^*_{o,d})$ be a feasible throughput and let $\omega \in \mathbb{R}^{\mathcal{E}}_+$ be a vector

[1] Note that we are implicitly assuming that delay and toll are measured in the same unit, something that is realistic provided that we have properly rescaled the toll.

of link tolls. Then, an o-d flow $f^{(\omega)}$ is a Wardrop equilibrium if and only if it is a solution of the network flow optimization problem

$$f^{(\omega)} \in \operatorname*{argmin}_{\substack{f \geq 0 \\ Bf = v(\delta^{(o)} - \delta^{(d)})}} \sum_{e \in \mathcal{E}} \left(D_e(f_e) + \omega_e f_e \right) . \tag{4.28}$$

Moreover, if f^* is the solution of a network flow optimization problem (4.2) with strictly convex increasing link costs $\psi_e(f_e)$, and

$$\omega_e^* = \psi_e'(f_e^*) - \tau_e(f_e^*) , \qquad e \in \mathcal{E} , \tag{4.29}$$

then the corresponding Wardrop equilibrium coincides with the system optimum flow, i.e.,

$$f^{(\omega^*)} = f^* .$$

Proof. First, let f° be an optimal solution of the minimization problem in (4.28). By the flow decomposition theorem, we can write f° as a combination of flows over paths in $\Gamma_{o,d}$ and over cycles. However, since every cycle in $\mathcal{G}$ contains a link e with $\tau_e(0) > 0$, every optimal solution of (4.28) cannot involve flow over cycles, so that necessarily $f^\circ = A^{(o,d)} z^\circ$ for some $z^\circ \in \mathbb{R}_+^{\Gamma_{o,d}}$ such that $\mathbb{1}' z^\circ = v$.

Now, observe that the complementary slackness conditions (4.10) imply that every optimal solution f° of (4.28) satisfies

$$\omega_{(i,j)} + \tau_{(i,j)}(f^\circ_{(i,j)}) \geq \lambda_i - \lambda_j , \qquad f^\circ_{(i,j)} \left(\omega_{(i,j)} + \tau_e(i,j)(f^\circ_{(i,j)}) - \lambda_i + \lambda_j \right) = 0$$

on every link $(i,j) \in \mathcal{E}$. It follows that, if $f^\circ = A^{(o,d)} z^\circ$ is an optimal solution of (4.28), then

$$\sum_{(i,j) \in \mathcal{E}} A_{(i,j),q} \left(\omega_{(i,j)} + \tau_e(i,j)(f^\circ_{(i,j)}) \right) \geq \sum_{(i,j) \in \mathcal{E}} A_{(i,j),q} (\lambda_i - \lambda_j) = \lambda_o - \lambda_d ,$$

for every path $q \in \Gamma_{o,d}$, whereas, for every path $p \in \Gamma_{o,d}$ such that $z_p^\circ > 0$, one has that $f^\circ_{(i,j)} \geq z_p^\circ > 0$ and hence $\omega_{(i,j)} + \tau_e(i,j)(f^\circ_{(i,j)}) = \lambda_i - \lambda_j$ for all (i,j) such that $A_{(i,j),p} = 1$, so that

$$\sum_{(i,j) \in \mathcal{E}} A_{(i,j),p} \left(\omega_{(i,j)} + \tau_e(i,j)(f^\circ_{(i,j)}) \right) = \sum_{(i,j) \in \mathcal{E}} A_{(i,j),p} (\lambda_i - \lambda_j) = \lambda_o - \lambda_d .$$

It then follows that every $f^\circ = A^{(o,d)} z^\circ$ satisfies (4.25), hence it is a Wardrop equilibrium. □

Corollary 4.1. Under the assumption of Theorem 4.2, a Wardrop equilibrium flow $f^{(\omega)}$ always exists. Moreover, if the delay functions $\tau_e(f_e)$ are strictly increasing $f^{(\omega)}$ is unique.

Proof. By Theorem 4.2, $f^{(\omega)}$ is an optimal network flow for the feasible problem (4.28), hence it exists. When $\tau_e(f_e)$ are strictly increasing, the cost in (4.28) is strictly convex, impliying uniqueness of the optimal network flow f^ω. □

Corollary 4.2. Let the assumptions of Theorem 4.2 be satisfied and additionally let the delay functions τ_e satisfy Assumption 2. Let f^* be a solution of the SO-TAP with link costs (4.20), and let the link tolls be chosen as

$$\omega_e^* = f_e^* \tau_e'(f_e^*)\,, \qquad e \in \mathcal{E}\,.$$

Then, the Wardrop equilibrium flow $f^{(\omega^*)}$ coincides with the system optimum flow f^*.

Proof. It follows from Theorem 4.2 upon observing that, with link costs (4.20), one has $\psi_e'(f_e^*) - \tau_e(f_e^*) = f_e^* \tau_e'(f_e^*)$. □

Theorem 4.2 states that the Wardrop equilibrium can indeed be seen as the optimal network flow provided that we consider the perceived costs defined in (4.27) instead of the ones defined by (4.20) that we used to define the SO-TAP. Hence, by selfishly and rationally choosing a route that minimizes their own delay, drivers do end up minimizing some global objective: the price of anarchy is then implied by the fact that the global objective $\sum_{e\in\mathcal{E}} \left(\int_0^{f_e} \tau_e(s)\mathrm{d}s + \omega_e f_e \right)$ implicitly optimized by the drivers does not coincide in general with the social objective considered in Section 4.4, i.e., the average delay $\sum_{e\in\mathcal{E}} \psi_e(f_e) = \sum_{e\in\mathcal{E}} f_e \tau_e(f_e)$.

However, Corollary 4.2 implies that, with the right choice of link tolls $\omega_e^* = \psi_e'(f_e^*) - \tau_e(f_e^*) = f_e^* \tau_e'(f_e^*)$, the Wardrop equilibrium gets realigned to the system optimum network flow f^* and thus the price of anarchy is reduced to 1. This specific choice of tolls are known as *marginal cost tolls*. Their economic interpretation is that they make users "internalize their negative externality": the idea is that the tolls are computed by charging the difference between the marginal social cost ψ_e' caused by the addition of an infinitesimal user on link e and the delay τ_e, that is the only cost perceived by the user in the absence of any tolls. In fact, the idea of marginal cost pricing predates the notion of Wardrop equilibrium and is attributed to the economist Pigou [**?**].

4.6 Power Dissipation and Electrical Networks

In this section, we focus on another special class of optimal network flow problems that show up in many physical applications. Throughout this section, we shall consider undirected multigraphs, i,e., multigraphs $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \theta, \kappa)$ such that there exists a bijection $e \mapsto \overleftarrow{e}$ of the link set $\mathcal{E}$ in itself such that $\theta(\overleftarrow{e}) = \kappa(e)$ and $\kappa(\overleftarrow{e}) = \theta(e)$ for every link e in $\mathcal{E}$. We shall then consider link costs in the form

$$\psi_e(f_e) = \frac{R_e^\alpha}{\alpha+1} f_e^{\alpha+1}\,, \qquad e \in \mathcal{E}\,,$$

where $\alpha > 0$ and $R_e = R_{\overleftarrow{e}} > 0$ is the resistance of both link e and its reverse link $\overleftarrow{e}$. Special cases are $\alpha = 1$ for direct current (DC) electrical networks; $\alpha = 2$ for gas networks; and $\alpha = 1/1.85 \simeq 0.54$ for water networks. In all those cases, the solution of the optimal flow problem (4.2) corresponds to the flow

observed in nature, i.e., the physical laws governing the flow can be modeled as minimizing the quantity

$$\sum_{e\in\mathcal{E}} \frac{R_e^\alpha}{\alpha+1} f_e^{\alpha+1}\,, \tag{4.30}$$

under the mass conservation and boundary flow constraints (4.1). The quantity (4.30) can typically be interpreted as a power dissipation (due to the flow resistances R_e), so that the solution of (4.2) is a minimal dissipation flow.

For every link e in $\mathcal{E}$ and $f_e \geq 0$, we have $\psi_e'(f_e) = R_e^\alpha f_e^\alpha$ and the complementary slackness condition (4.11) reduces to

$$f_e^* = \begin{cases} \dfrac{\big(\lambda_{\theta(e)} - \lambda_{\kappa(e)}\big)^{1/\alpha}}{R_e} & \text{if} \quad \lambda_{\theta(e)} > \lambda_{\kappa(e)} \\ 0 & \text{if} \quad \lambda_{\theta(e)} \leq \lambda_{\kappa(e)}\,, \end{cases}$$

i.e.,

$$f_e^* = \frac{1}{R_e}\left[\lambda_{\theta(e)} - \lambda_{\kappa(e)}\right]_+^{1/\alpha}\,.$$

We shall now focus on the special case $\alpha = 1$, i.e., on electrical networks, where the above reads

$$f_e^* = \frac{1}{R_e}\left[\lambda_{\theta(e)} - \lambda_{\kappa(e)}\right]_+\,. \tag{4.31}$$

In this case, it is convenient to define the weights

$$W_{ij} = \sum_{\substack{e\in\mathcal{E}:\\ \theta(e)=i\\ \kappa(e)=j}} \frac{1}{R_e}\,, \qquad W_{ii} = 0\,, \qquad i,j\in\mathcal{V}\,,$$

to be interpreted as the conductance between nodes i and j. We shall then consider the undirected graph with node set $\mathcal{V}$ obtained by merging parallel links and weight matrix W that with a slight abuse of notation will be denoted by $\mathcal{G}$ as well. With this notation, the net flow from node i to node j satisfies

$$z_{ij}^* = W_{ij}(\lambda_i - \lambda_j)\,. \tag{4.32}$$

The above can be understood in physical terms as the famous Ohm's law upon interpreting z_{ij}^* as the net electrical current from i to j and the multipliers λ_i and λ_j as the voltages in the different nodes. The dual function is

$$D(\lambda,\nu) = L(f^*(\lambda),\lambda,\nu) = \sum_{e\in\mathcal{E}} \psi_e^*(\lambda_{\theta(\overline{e})} - \lambda_{\kappa(\overline{e})}) + \sum_i \lambda_i\nu_i\,,$$

where

$$\psi_e^*(y_e) = \frac{1}{R_e}\left(\frac{1}{2}[y_e]_+^2 - [y_e]_+ y_e\right)\,.$$

Now, notice that, for $y_e = -y_{\overline{e}}$, either $[y_e]_+ = y_e$ and $[y_{\overline{e}}]_+ = [-y_e]_+ = 0$ or $[y_e]_+ = 0$ and $[y_{\overline{e}}]_+ = [-y_e]_+ = -y_e$. Then, since $R_e = R_{\overline{e}}$, we have that

$$\psi_e^*(\lambda_{\theta(e)} - \lambda_{\kappa(e)}) + \psi_{\overline{e}}^*(\lambda_{\kappa(\overline{e})} - \lambda_{\theta(\overline{e})}) = -\frac{1}{2R_e}\left(\lambda_{\theta(e)} - \lambda_{\kappa(e)}\right)^2 .$$

It follows that

$$D(\lambda,\nu) = \sum_{e\in\mathcal{E}} \psi_e^*(\lambda_{\theta(e)} - \lambda_{\kappa(e)}) + \sum_i \lambda_i\nu_i = -\sum_{\{i,j\}} \frac{W_{ij}}{2}(\lambda_i - \lambda_j)^2 + \sum_i \lambda_i\nu_i ,$$

where the summation is meant to run over all unordered pairs of nodes and, in fact, over all undirected links in the graph $\mathcal{G}$ (since $W_{ij} = 0$ whenever node i and node j are not linked). It follows that

$$M^*(\nu) = \max_{\lambda\in\mathbb{R}^{\mathcal{V}}} D(\lambda,\nu) = -\min_{\lambda\in\mathbb{R}^{\mathcal{V}}} \left\{ \sum_{\{i,j\}} \frac{W_{ij}}{2}(\lambda_i - \lambda_j)^2 - \sum_i \lambda_i\nu_i \right\} . \tag{4.33}$$

Now, consider the case when $\nu = \delta^{(h)} - \delta^{(k)}$ for two nodes $h \neq k$ and let

$$\lambda^* = \operatorname*{argmin}_{\lambda\in\mathbb{R}^{\mathcal{V}}} \left\{ \sum_{\{i,j\}} \frac{W_{ij}}{2}(\lambda_i - \lambda_j)^2 - \lambda_h + \lambda_k \right\} . \tag{4.34}$$

Then define the *effective resistance* between node h and node k as

$$R_{hk}^{\mathcal{G}} = \lambda_h^* - \lambda_k^* . \tag{4.35}$$

We have the following result, known as the Thompson principle.

Theorem 4.3. Consider an undirected multigraph $\mathcal{G} = (\mathcal{V},\mathcal{E},\theta,\kappa)$ with resistances $R_e = R_{\overline{e}}$ for every $e \in \mathcal{E}$. Then, the effective resistance between any pair of nodes satisfies

$$R_{hk}^{\mathcal{G}} = \min_{\substack{f\in\mathbb{R}_+^{\mathcal{V}}:\\ Bf=\delta^{(h)}-\delta^{(k)}}} \sum_{e\in\mathcal{E}} R_e f_e^2 , \qquad h,k \in \mathcal{V} . \tag{4.36}$$

Proof. It follows from Proposition 4.1, (4.33), (4.34), (4.31), and (4.35) that

$$\begin{aligned}
-M(\delta^{(h)} - \delta^{(k)}) &= -M^*(\delta^{(h)} - \delta^{(k)}) \\
&= \min_{\lambda\in\mathbb{R}^{\mathcal{V}}} \left\{ \sum_{\{i,j\}} \frac{W_{ij}}{2}(\lambda_i - \lambda_j)^2 + \lambda_k - \lambda_h \right\} \\
&= \sum_{\{i,j\}} \frac{W_{ij}}{2}(\lambda_i^* - \lambda_j^*)^2 + \lambda_k^* - \lambda_h^* \\
&= \frac{1}{2}\sum_{e\in\mathcal{E}} R_e (f_e^*)^2 - R_{hk}^{\mathcal{G}} \\
&= M(\delta^{(h)} - \delta^{(k)}) - R_{hk}^{\mathcal{G}} ,
\end{aligned}$$

from which it follows that

$$R^{\mathcal{G}}_{hk} = 2M(\delta^{(h)} - \delta^{(k)}) = \min_{\substack{f \in \mathbb{R}^{\mathcal{V}}_{+}: \\ Bf = \delta^{(h)} - \delta^{(k)}}} \sum_{e \in \mathcal{E}} R_e f_e^2 \,,$$

thus proving the claim. □

Chapter 5

Reversibility and Electrical Networks

5.1 Reversible stochastic matrices

The computation of the (normalized) left dominant eigenvector of a stochastic matrix plays a crucial role in many applications. One context is that of centralities considered in the previous section, as it corresponds to the Bonacich centrality. More applications will be discussed in Chapter 6. In general, if the matrix P is of large dimension, this can be a major computational problem. There is an important class of stochastic matrices for which such computation is straightforward, it is that of reversible matrices.

A stochastic matrix P is said to be *reversible* with respect to a non negative vector y if

$$y_i P_{ij} = y_j P_{ji} \qquad \forall i, j \in \mathcal{V}\,. \tag{5.1}$$

Observe that equation (5.1) implies that

$$\sum_{j\in\mathcal{V}} P_{ji} y_j = \sum_{j\in\mathcal{V}} y_i P_{ij} = y_i\,, \qquad i \in \mathcal{V}\,, \tag{5.2}$$

hence $y = P'y$, in other words y is a left dominant eigenvector. If P is also irreducible, we know it possesses just one normalized left dominant eigenvector π (see (ii) of Theorem 2.4) that, consequently, must coincide with a normalization of y. In particular we also have that $\pi_i P_{ij} = \pi_j P_{ji}$ for all i, j: namely, if P is irreducible and reversible, it is reversible with respect to its unique normalized left dominant eigenvector π. The following example is in fact a characterization of the reversibility property.

Example 5.1. If $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is an undirected graph, then it is immediate to verify that $P = D^{-1}W$, where $D = \mathrm{diag}\,(w)$ and $w = W\mathbb{1}$, is reversible with respect to the vector w. Conversely, for a stochastic matrix P to be reversible with respect to a non negative vector y it means that $W = \mathrm{diag}\,(y)P$ is a symmetric

nonnegative matrix, so that the weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is undirected. Notice that a stochastic matrix P may be reversible but not symmetric (this is always the case if the graph $\mathcal{G}$ is not regular). Given such a reversible non symmetric P, if one considers the associated graph $\mathcal{G}_P = (\mathcal{V}, \mathcal{E}, P)$, we clearly have that $\mathcal{G}_P$ is not undirected. In other terms, also graphs that are not undirected may give rise to reversible stochastic matrices.

Reversible stochastic matrices have other relevant properties. One of these is the fact that they are diagonalizable. For an irreducible reversible stochastic matrix P with normalized left dominant eigenvector π, define the matrix

$$M := \Pi^{1/2} P \Pi^{-1/2}, \qquad \Pi := \operatorname{diag}(\pi).$$

It is immediate to check that M is symmetric. Then, M (and hence P since they are similar matrices) has real eigenvalues

$$1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1.$$

The quantity $1-\lambda_2$ is called the *spectral gap* of P and has remarkable connections with geometric properties of the underlying graph as illustrated in the next section.

5.2 Spectral gap and network conductance

In this section we analyze the spectral gap of reversible stochastic matrices and show its connection with geometric properties of the underlying graph.

The first step is to establish useful *variational characterization* of the spectral gap. For two vectors $x, y \in \mathbb{R}^n$, we define the *Dirichlet form*

$$\mathcal{E}(x, y) := \frac{1}{2} \sum_{i,j \in \mathcal{V}} \pi_i P_{ij} (x_i - x_j)(y_i - y_j). \tag{5.3}$$

When choosing $x = y$ the above reduces to the following quadratic form

$$\mathcal{E}(x, x) := \frac{1}{2} \sum_{i,j \in \mathcal{V}} \pi_i P_{ij} (x_i - x_j)^2.$$

Then, the following result holds true:

Proposition 5.1 (Variational characterization of the spectral gap)**.** Let P be an irreducible reversible stochastic matrix. Let π be its invariant probability distribution and let λ_2 be its second-largest eigenvalue. Then, the spectral gap of P satisfies

$$1 - \lambda_2 = \min \left\{ \frac{\mathcal{E}(x, x)}{\sum_i \pi_i x_i^2} : x \neq 0, \, \pi' x = 0 \right\}, \tag{5.4}$$

where $\mathcal{E}$ is the Dirichlet form defined in (5.3).

Proof. Notice that

$$\begin{aligned}
\mathcal{E}(x,x): &= \tfrac{1}{2}\sum_{i,j\in\mathcal{V}}\pi_i P_{ij}(x_i-x_j)^2\\
&= \tfrac{1}{2}\sum_{i,j\in\mathcal{V}}[\pi_i P_{ij}x_i(x_i-x_j)-\pi_i P_{ij}x_j(x_i-x_j)]\\
&= \tfrac{1}{2}\sum_{i,j\in\mathcal{V}}[\pi_i P_{ij}x_i(x_i-x_j)-\pi_j P_{ji}x_i(x_j-x_i)]\\
&= \tfrac{1}{2}\sum_{i,j\in\mathcal{V}}[\pi_i P_{ij}x_i(x_i-x_j)-\pi_i P_{ij}x_i(x_j-x_i)]\\
&= \sum_{i\in\mathcal{V}}\pi_i x_i^2-\sum_{i,j\in\mathcal{V}}\pi_i P_{ij}x_i x_j\\
&= x'\Pi(I-P)x
\end{aligned}$$

where the third equality follows by a simple exchange of index names in the second formula and the fourth equality is the reversibility. Hence,

$$\begin{aligned}
\min\left\{\tfrac{\mathcal{E}(x,x)}{\sum_i \pi_i x_i^2}:\, x\neq 0,\, \pi'x=0\right\} &= \min\{\tfrac{x'\Pi(I-P)x}{x'\Pi x}:\, x\neq 0,\, \pi'x=0\}\\
&= \min\{\tfrac{y'\Pi^{1/2}(I-P)\Pi^{-1/2}y}{y'y}:\, y\neq 0,\, \pi'^{1/2}y=0\}\\
&= \min\{\tfrac{y'(I-M)y}{y'y}:\, y\neq 0,\, \pi'^{1/2}y=0\}
\end{aligned}$$

where the second equality simply comes from the change of variable $y=\Pi^{1/2}x$. Notice now that, if $\pi'^{1/2}y=0$,

$$y'(I-M)y=\sum_{1\leq k\leq n}(1-\lambda_k)[y'z_{(k)}]^2=\sum_{2\leq k\leq n}(1-\lambda_k)[y'z_{(k)}]^2\geq(1-\lambda_2)y'y$$

Choose now $z_{(2)}$ to be an eigenvector of M of eigenvalue λ_2 such that $||z_{(2)}||_2=1$. Then,

$$z'_{(2)}(I-M)z_{(2)}=(1-\lambda_2)$$

This proves the result. □

Through the variational characterization it is possible to relate the spectral gap of a reversible matrix to the geometry of the associated graph.

Consider a strongly connected undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E},W)$ with associated stochastic matrix P and relative normalized left dominant eigenvector $\pi=(w'\mathbb{1})^{-1}w$. Given a subset $\mathcal{U}\subseteq\mathcal{V}$, put $w_{\mathcal{U}}:=\sum_{i\in\mathcal{U}}w_i$ and $\pi_{\mathcal{U}}:=\sum_{i\in\mathcal{U}}\pi_i$. The *bottleneck ratio* of U is defined as

$$\Phi(\mathcal{U}):=\frac{\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{V}\setminus\mathcal{U}}W_{ij}}{w_{\mathcal{U}}}=\frac{\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{V}\setminus\mathcal{U}}\pi_i P_{ij}}{\pi_{\mathcal{U}}}$$

This constant measures the way a subset $\mathcal{U}$ is connected to the remaining part of the graph. Note that $\Phi(\mathcal{U})=0$ iff $\mathcal{U}$ is a trapping set. The *bottleneck ratio* of a $\mathcal{G}$ (also known as *conductance*, *isoperimetric constant* or *Cheeger constant*), is defined as

$$\Phi=\min_{\substack{\mathcal{U}\subseteq\mathcal{V}:\\ 0<w_{\mathcal{U}}\leq\frac{1}{2}\mathbb{1}'w}}\Phi(\mathcal{U})$$

where $w_{\mathcal{U}} := \sum_{i\in\mathcal{U}} w_i$ stands for the total degree of nodes in $\mathcal{U}$. Note that $\Phi = 0$ iff the graph is disconnected.

For a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ this simplifies as

$$\Phi = \min_{\substack{\mathcal{U}\subseteq\mathcal{V}:\\ 0<w_{\mathcal{U}}\leq \frac{1}{2}\mathbb{1}'w}} \frac{|\partial_{\mathcal{U}}|}{w_{\mathcal{U}}},$$

where $\partial_{\mathcal{U}} = \{(i,j) \in \mathcal{E} : i \in \mathcal{U},\, j \in \mathcal{V}\setminus\mathcal{U}\}$ denotes its boundary, i.e., the set of links connecting $\mathcal{U}$ to its complementary set $\mathcal{V}\setminus\mathcal{U}$.

Example 5.2. In the complete graph with n nodes and without self-loops, the bottleneck ratio of a subset $\mathcal{U}$ of nodes of cardinality $|\mathcal{U}| = k$ is $k(n-k)/(k(n-1)) = (n-k)/(n-1)$ so that the bottleneck Φ is slightly more than $1/2$ (precisely, $\Phi = \frac{1}{2}(1 + \frac{1}{n-1})$ for even n and $\Phi = \frac{1}{2}(1 + \frac{2}{n-1})$ for odd n).

On the other hand, for the barbell graph consisting of two complete graphs with $n/2$ nodes each interconnected by a single link, $\Phi = (n^2/4 - n/2 + 1)^{-1}$ converges to 0 as n grows large.

Theorem 5.1 (Cheeger's inequality)**.** The spectral gap of a stochastic matrix P reversible with respect to an invariant probability π satisfies

$$\frac{1}{2}\Phi^2 \leq 1 - \lambda_2 \leq 2\Phi\,.$$

Proof. We prove the right-most inequality. Given a subset $\mathcal{U} \subseteq \mathcal{V}$ such that $\pi(\mathcal{U}) \leq 1/2$, consider the vector $x^{\mathcal{U}}$ defined as follows: $x^{\mathcal{U}}_i = \pi(\mathcal{U}^c)$ if $i \in \mathcal{U}$, while $x^{\mathcal{U}}_i = -\pi(\mathcal{U})$ if $i \in \mathcal{U}^c$. We now compute the Dirichlet form $\mathcal{E}$ over such a vector $x^{\mathcal{U}}$:

$$\begin{aligned}
\mathcal{E}(x^{\mathcal{U}}, x^{\mathcal{U}}): &= \tfrac{1}{2}\textstyle\sum_{i,j\in\mathcal{V}} \pi_i P_{ij}(x^{\mathcal{U}}_i - x^{\mathcal{U}}_j)^2 \\
&= \textstyle\sum_{i\in\mathcal{U}, j\in\mathcal{U}^c} \pi_i P_{ij}(x^{\mathcal{U}}_i - x^{\mathcal{U}}_j)^2 \\
&= \textstyle\sum_{i\in\mathcal{U}, j\in\mathcal{U}^c} \pi_i P_{ij}(\pi(\mathcal{U}^c) + \pi(\mathcal{U}))^2 \\
&= \textstyle\sum_{i\in\mathcal{U}, j\in\mathcal{U}^c} \pi_i P_{ij}
\end{aligned}$$

Moreover,

$$\sum_i \pi_i (x^{\mathcal{U}}_i)^2 = \sum_{i\in\mathcal{U}} \pi_i \pi(\mathcal{U}^c)^2 + \sum_{i\in\mathcal{U}^c} \pi_i \pi(\mathcal{U})^2 = \pi(\mathcal{U})\pi(\mathcal{U}^c)$$

From the variational inequality (5.4) we obtain

$$1 - \lambda_2 \leq \frac{\mathcal{E}(x^{\mathcal{U}}, x^{\mathcal{U}})}{\sum_i \pi_i (x^{\mathcal{U}}_i)^2} = \frac{\sum_{i\in\mathcal{U}, j\in\mathcal{U}^c} \pi_i P_{ij}}{\pi(\mathcal{U})\pi(\mathcal{U}^c)} \leq 2\Phi(\mathcal{U})$$

Since this holds true for every subset $\mathcal{U}$ such that $\pi(\mathcal{U}) \leq 1/2$, thesis follows.

The other inequality is much more involved, and we refer to [26, Theorem 13.14] for a proof. □

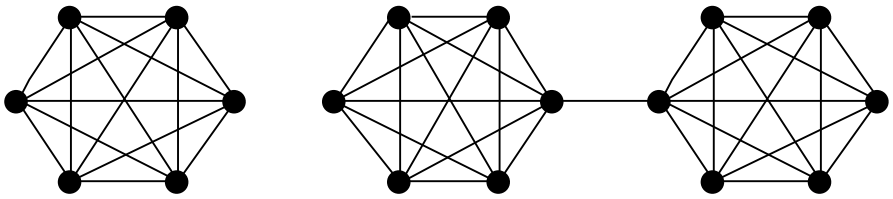


Figure 5.1: Left: a complete graph (with no self-loops), whose conductance is $\Phi = \min_{0<k\leq n/2} \frac{n-k}{n-1} \geq 1/2$. Right: a barbell graph whose conductance is $\Phi = (n^2/4 - n/2 + 1)^{-1}$.

5.3 Electrical networks and effective resistances

??

Undirected graphs are also a basic natural model for electrical networks. Specifically, a strongly connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with no self-loops can be interpreted as a (purely resistor) electrical network where $W_{ij} = W_{ji}$ is the *electrical conductance* of the undirected link $\{i, j\}$. The term $1/W_{ij}$ is consequently to be interpreted as the *electrical resistance* of $\{i, j\}$.

We consider an exogenous net flow ν on $\mathcal{G}$ (with the usual requirement that $\sum_i \nu_i = 0$) interpreted as an exogenous flow of currents.

The corresponding *current flow* $\phi \in \mathbb{R}^{\mathcal{E}}$ and *voltage vector* $x \in \mathbb{R}^{\mathcal{V}}$ are defined as the solution of the equations

$$\sum_j \phi_{ij} - \sum_j \phi_{ji} = \nu_i \tag{5.5}$$

$$\phi_{ij} = W_{ij}[x_i - x_j]_+ \tag{5.6}$$

The first one is the usual flow balance equation that every network flow determined by the exogenous net flow ν has to satisfy. In the electrical context, it takes the name of *Kirchoff's law*. The second equation is instead specific of the electrical interpretation and is known as the *Ohm's law*: it states that current flows along edges always from the higher voltage node to the lower one and flow is proportional to the voltage difference with a constant of proportionality given by the electrical conductance of the edge.

Below, we formally prove the existence of a current flow ϕ and of a voltage vector x satisfying equations (5.5) and (5.6) and we give a description in terms of algebraic objects related to the graph $\mathcal{G}$.

First, if we substitute the expression for ϕ_{ij} in (5.6) inside (5.5), we obtain

$$\begin{array}{rcl} \nu_i & = & \sum_j W_{ij}[x_i - x_j]_+ - \sum_j W_{ji}[x_j - x_i]_+ \\ & = & \sum_j W_{ij}([x_i - x_j]_+ - [x_j - x_i]_+) \\ & = & \sum_j W_{ij}(x_i - x_j) = w_i x_i - \sum_j W_{ij} x_j \end{array}$$

In other terms, recalling that $L = D - W$ (where $D = \mathrm{diag}\,(W\mathbb{1})$) is the Laplacian matrix associated to the graph, this can be rewritten as

$$Lx = \nu \tag{5.7}$$

We have thus shown that the voltage function x necessarily has to satisfy relation (5.7). Actually, equation (5.7) is equivalent to the relations (5.5) and (5.6) in the following sense. If x is a vector that satisfies (5.7) and we set the network flow ϕ by imposing that, for every edge $(i,j) \in \mathcal{E}$, it holds $\phi_{ij} = W_{ij}[x_i - x_j]_+$, then ϕ satisfies (5.5). We next study the solutions of (5.7).

Notice that L in this case is symmetric and can thus be decomposed as

$$L = \sum_{k\geq 2} \overline{\lambda}_k \overline{z}_{(k)} \overline{z}'_{(k)}$$

where $0 = \overline{\lambda}_1 < \overline{\lambda}_2 \leq \overline{\lambda}_3 \leq \ldots \overline{\lambda}_n$ are the real nonnegative eigenvalues of L and $\overline{z}_{(k)}$ the corresponding normalized eigenvectors with $\overline{z}_{(1)} = n^{-1/2}\mathbb{1}$. We now define the *Green matrix* of the graph $\mathcal{G}$ as

$$Z = \sum_{k\geq 2} \frac{1}{\overline{\lambda}_k} \overline{z}_{(k)} \overline{z}'_{(k)} \tag{5.8}$$

The following relations follow directly from the definition

$$ZL = LZ = I - n^{-1}\mathbb{1}\mathbb{1}', \quad Z\mathbb{1} = 0 \tag{5.9}$$

We can now state the following result:

Proposition 5.2. Consider a strongly connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with no self-loops and an exogenous net flow ν on $\mathcal{G}$. The following facts hold true:

1. x is a solution of (5.7 if and only if

   $$x = Z\nu + c\mathbb{1}$$

   where c is any constant.

2. All the solutions of the relations (5.5) and (5.6) are given by any vector $x = Z\nu + c\mathbb{1}$ and the unique network flow ϕ determined by (5.6).

Proof. 1. If $Lx = \nu$, it follows from (5.9) that

$$Z\nu = ZLx = x - n^{-1}\mathbb{1}\mathbb{1}'x$$

If we put $c = n^{-1}\mathbb{1}'x$, we have that $x = Z\nu + c\mathbb{1}$. Conversely, if x has the form $x = Z\nu + c\mathbb{1}$ for some constant c, using again the relations (5.9) and the fact that $\mathbb{1}'\nu = 0$, we have that

$$Lx = LZ\nu + cL\mathbb{1} = 0$$

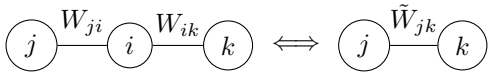
2. follows from 1. and previous considerations.

Remark 4. *The computation of voltages and low currents in an electrical network specified by a graph $\mathcal{G}$ and an exogenous net flow ν can sometimes be simplified using the reduction rules specified below. Their validity is a direct consequence of the Kirchoff's and Ohm's laws.*

1. Series law: *Whenever we have a node i with just two neighbors j, k and for which $\nu_i = 0$, we can remove it from the network and replace it with an edge between j and k having weight $\tilde{W}_{jk}$ satisfying*

$$\tilde{W}_{jk}^{-1} = W_{ji}^{-1} + W_{ik}^{-1}$$

In other words, the resistance of the new edge is the sum of the resistances of the two old edges. The new electrical network is equivalent to the original one in the sense that the flow current and the voltage through all edges and nodes except i and the edges $\{j, i\}$ and $\{i, k\}$ are the same while the flow current through the new edge $\{j, k\}$ is the same than along the original $\{j, i\}$ and $\{i, k\}$.



2. Parallel law: *Whenever we have two parallel edges connecting two nodes i and j having conductance, respectively, W'_{ij} and W''_{ij}, we can replace it with just one edge having new conductance $\tilde{W}_{ij}$ given by*

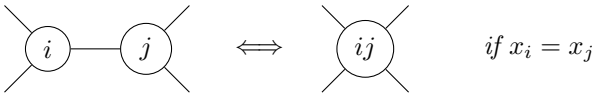
$$\tilde{W}_{ij} = W'_{ij} + W''_{ij}$$

The new electrical network is equivalent to the original one in the sense that the flow current and the voltage through all edges and nodes except the modified edges are the same in the two networks.

Since we are here dealing with undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ and not multigraphs, the reader may wonder on the applicability of the parallel law as parallel edges in principle should not be present. Notice however that parallel edges could show up as the result of applying the series law. While we have preferred not to use explicitly the more general formalism of multigraphs, the presence of parallel edges is possible and can be dealt with using the parallel law.

3. *Glueing: If two nodes connected by an edge are known to be at the same voltage (and thus the flow of current through the edge be equal to 0), we*

can glue the two nodes together and removing the connecting edge. The new electrical network is equivalent to the original one in the sense that the flow current and the voltage through all edges and nodes remain the same.



$$\text{if } x_i = x_j$$

In certain contexts, it is natural to assign the voltage in certain specific nodes of the network instead of the exogenous currents. We here discuss the feasibility of the problem and how to compute in this case the voltage in the remaining node.

We start with an observation. Consider an exogenous net flow ν and put

$$\mathcal{S} = \{i : \nu_i \neq 0\}, \quad \mathcal{R} = \mathcal{V} \setminus \mathcal{S}$$

If x is a corresponding voltage, we know it satisfies the relation $Lx = \nu$. This implies that $(Lx)_i = 0 \ \forall i \in \mathcal{R}$. A vector x that satisfies this property is called *harmonic* on $\mathcal{R}$. The next result shows that it is always possible, given a subvector of real values relative to a subset of nodes of $\mathcal{G}$, to extend it to a vector that is harmonic on the complementary set. Precisely we have the following.

Proposition 5.3 (Harmonic extensions)**.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be an undirected strongly connected graph and $\mathcal{S} \subseteq \mathcal{V}$ be a non empty subset of it. Put $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$. Fix $u \in \mathbb{R}^{\mathcal{S}}$. Then,

1. There exists a unique $x \in \mathbb{R}^{\mathcal{V}}$ such that

$$\begin{aligned} (Lx)_i &= 0 \qquad i \in \mathcal{R}\,, \\ x_s &= u_s \qquad s \in \mathcal{S}\,. \end{aligned} \tag{5.10}$$

   In other terms, x is harmonic on $\mathcal{R}$ and coincide with u on $\mathcal{S}$

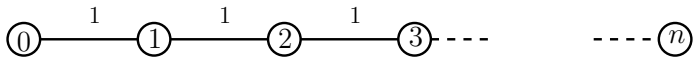
2. There exists a unique exogenous net flow ν supported on $\mathcal{S}$ such that x is a voltage corresponding to it.

Proof. 1. will be proven later as a consequence of a result on averaging dynamics.

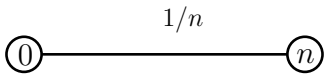
2. Let x be the solution of (5.10) for a given u and put $\nu = Lx$. Notice that $\mathbb{1}'\nu = \mathbb{1}'Lx = 0$. This implies that ν is a exogenous net flow. Moreover, the fact that x is harmonic on $\mathcal{R}$ exactly says that $\nu_i = 0$ for every $i \notin \mathcal{S}$. This completes the proof. □

Expression (5.10) is sometimes referred to as the discrete Laplace equation on $\mathcal{G}$ with boundary conditions on $\mathcal{S}$, also known as the *Dirichlet problem*. x is also called the *harmonic extension* of u

Example 5.3. Consider the simple line graph



We assign the voltage in the two leaves nodes: $u_0 = 0$, $u_n = 1$. In order to find the harmonic extension x, we make use of the electrical reduction techniques illustrated above. Since we know that x is a voltage relative to a exogenous net flow supported on $\mathcal{S} = \{0, n\}$, an iterative use of the series law implies that the above grid can be reduced to



Ohm's law yields

$$\phi_{n0} = W_{0n}(x_n - x_0) = \frac{1}{n}$$

Kirchoff's law now implies

$$\phi_{i+1i} = \frac{1}{n}, \quad x_{i+1} - x_i = \frac{\phi_{i+1i}}{W_{i+1i}} = \frac{1}{n}$$

Finally, we get for every $k = 0, \ldots, n$

$$x_k = \sum_{i=0}^{k-1} (x_{i+1} - x_i) = \frac{k}{n}$$

Chapter 6

Distributed averaging and linear flow dynamics

In this chapter, we study two deterministic linear network dynamics: distributed averaging and linear flow dynamics.

6.1 Linear distributed averaging

Consider an influence network described as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, where the node set $\mathcal{V}$ represents a population of agents, the link set $\mathcal{E}$ represents interactions among agents, and the entries W_{ij} of the weight matrix quantify the strength of the influence that agent j has on agent i. Let P be the normalized weight matrix of $\mathcal{G}$. Notice that our convention is that the direction of the influence is the opposite of the one of the links, i.e., the presence of a link (i, j) indicates that agent i 'observes' agent j and gets influenced by her/him. Assume that every agent i has a state $x_i(t) \in \mathbb{R}$ that is updated at discrete time steps $t = 0, 1, \ldots$ in response to the current states of her/his out-neighbors in $\mathcal{G}$ according to the following linear averaging rule

$$x_i(t+1) = \sum_{j \in \mathcal{V}} P_{ij} x_j(t)\,, \qquad i \in \mathcal{V}\,. \tag{6.1}$$

Equation (6.1) prescribes that the new state $x_i(t+1)$ of agent i coincides with the weighted average of the current opinions $x_j(t)$ of her/his out-neighbors $j \in \mathcal{N}_i$, each weighted proportionally to the (i, j) link's weight W_{ij}. Observe that, by allowing for the presence of self-loops in the graph $\mathcal{G}$, i.e., nonzero diagonal entries of the weight matrix W, we can account for some inertia in the update rule (6.1). The update rule (6.1) can be rewritten more compactly as

$$x(t+1) = Px(t)\,, \qquad t = 0, 1, \ldots \tag{6.2}$$

where $x(t) = (x_i(t))_{i \in \mathcal{V}}$ is the vector of all agents' opinions. We shall refer to (6.2) as the *linear averaging* dynamics on $\mathcal{G}$. As we shall discuss more in detail

later on in this chapter, the linear averaging dynamics has several applications in many different fields, from social science —where (6.2) is known as the French-DeGroot opinion dynamics model [17, 23, 12]— to sensor networks —where (6.2) is used as a distributed estimation algorithm.

What are the possible asymptotic outcomes of the DeGroot opinion dynamics model? Observe that an equilibrium vector x for the dynamics (6.2) satisfies

$$x = Px\,, \tag{6.3}$$

i.e., it is an eigenvector of the stochastic matrix P corresponding to its leading eigenvalue 1. Since P is a stochastic matrix, we know that $P\mathbb{1} = \mathbb{1}$, i.e., any vector $x = \alpha\mathbb{1}$ for some $\alpha \in \mathbb{R}$ is a solution of (6.3). Such vectors $x = \alpha\mathbb{1}$ are called *consensus* vectors, since all their entries are equal to the same *consensus value* α. Hence, we have found out that every consensus vector is an equilibrium for the DeGroot opinion dynamics (6.2). Moreover, it follows from Proposition 2.4 that, if $\mathcal{G}$ is strongly connected, and more in general if $s_{\mathcal{G}} = 1$, then the only equilibria of (6.2) are the consensus vectors. More in general, the aforementioned results imply that the space of equilibria of (6.2) has dimension equal to the number of sinks $s_{\mathcal{G}}$ of the condensation graph of $\mathcal{G}$.

Let us first focus on the case when $s_{\mathcal{G}} = 1$, hence in particular when $\mathcal{G}$ is strongly connected, and address the following questions: does the opinion dynamics (6.2) always converge to a consensus $x = \alpha\mathbb{1}$? If so, how does the consensus value α depend on the initial opinion vector $x(0)$? We start by answering the second question. Recall that, if $\mathcal{G}$ is strongly connected, and more in general if $s_{\mathcal{G}} = 1$, then there exists a unique invariant probability distribution

$$\pi = P'\pi\,.$$

Now observe that, by taking the scalar product of π with both sides of equation (6.2), we get

$$\pi' x(t+1) = \pi' P x(t) = (P'\pi)' x(t) = \pi' x(t)\,, \qquad t = 0, 1, \dots\,,$$

which shows that the weighted average $\pi' x(t)$ is invariant for the dynamics (6.2). This in particular implies that

$$\pi' x(t) = \pi' x(0)\,, \qquad t = 0, 1, \dots\,.$$

Taking the limit as t grows large in the equation above, and recalling that, since $\mathcal{G}$ is connected, the only equilibria of (6.2) are consensus vectors, we get that, if the opinion vector $x(t)$ actually converges, its limit must be a consensus vector $x = \alpha\mathbb{1}$ with consensus value

$$\alpha = \pi' x(0) = \sum_{i \in \mathcal{V}} \pi_i x_i(0)\,. \tag{6.4}$$

Equation (6.4) states that the consensus value is the weighted average of the agents' initial opinions, the weight assigned to each agent i's initial opinion

$x_i(0)$ being equal to her/his invariant probability π_i. It is particularly insightful that the asymptotic consensus value coincides with the weighted average of the initial opinions with weights given by the node invariant probabilities. This gives the invariant probability distribution π of the graph $\mathcal{G}$ significance also in a dynamic context. In particular, recall that, when the graph is balanced —hence in particular when it is undirected— the invariant probability distribution coincides with the normalized degree vectors: $\pi_i = w_i/(n\overline{w})$. For general unbalanced networks, such an explicit expression is not available.

We still have to answer our first question, namely whether the opinion vector $x(t)$ actually converges. In fact, strong connectedness of the graph $\mathcal{G}$ by itself is not sufficient to guarantee this as is clear from the following example.

Example 6.1. Consider a network with two nodes, no self loops, and one undirected link of weight 1 connecting them, so that $P = W = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then, unless already started at a consensus, the dynamics (6.2) will keep on oscillating with both nodes copying one the other's opinion and never ending up agreeing with each other or approaching some equilibrium.

In fact, the situation in Example 6.1 is rather pathological, and can be ruled out provided that, in addition to being connected, the graph $\mathcal{G}$ is *aperiodic*, i.e., the maximum common divisor of all its cycle lengths equals 1. This additional assumption is quite a weak one: e.g., it is satisfied whenever at least one node has a self-loop, i.e., whenever at least one diagonal entry of the weight matrix W is strictly positive (in this case the maximum common divisor is clearly 1). In many applications this is a very natural assumption. In fact, in some cases it is customary to to consider the *lazy* version of a stochastic matrix P as $P_{(\text{lazy})} = \frac{1}{2}(P + I)$.

When the graph $\mathcal{G}$ is such that $s_{\mathcal{G}} = 1$ and aperiodic, we have the following result implying convergence of the linear distributed averaging dynamics.

Proposition 6.1. Let $\mathcal{G}$ be a graph with normalized weight matrix P. If $s_{\mathcal{G}} = 1$, and the connected component $\mathcal{U}$ corresponding to the unique sink of the condensation graph is aperiodic in $\mathcal{G}$, then

$$\lim_{t \to +\infty} P^t = \mathbb{1}\pi' ,$$

where $\pi = P'\pi$ is the unique invariant probability distribution of $\mathcal{G}$.

Proof. Consider first the case when $\mathcal{G}$ is strongly connected and aperiodic. Put $M = P - \mathbb{1}\pi'$ and notice first of all that $M'\pi = 0$. Consider now any other eigenvalue $\lambda \neq 0$ of M' with corresponding eigenvector y. Since $M\mathbb{1} = 0$, it follows that

$$0 = \mathbb{1}'M'y = \lambda \mathbb{1}'y \;\Rightarrow\; \mathbb{1}'y = 0$$

Therefore, $P'y = M'y + \pi\mathbb{1}'y = \lambda y$. Since $\mathcal{G}$ is strongly connected, it follows from (v) of Proposition 2.4 that, necessarily, $\lambda \neq 1$ and, since $\mathcal{G}$ is also aperiodic, it it follows from (vi) of Proposition 2.4 that $|\lambda| < 1$. We have thus shown that

all eigenvalues λ of M' (and thus also of M) are such that $|\lambda| < 1$. This immediately implies that $M^t \to 0$ for $t \to +\infty$.[1] Since $M^t = P^t - \pi \mathbb{1}'$, this yields the result.

We now consider the general case. Let $\mathcal{U}$ be the connected component in $\mathcal{G}$ corresponding to the sink of its condensation graph. Upon ordering nodes in $\mathcal{V}$ so that nodes in $\mathcal{V} \setminus \mathcal{U}$ come first, the matrix P takes the following structure

$$P = \begin{pmatrix} Q & R \\ 0 & S \end{pmatrix}$$

where, due to the assumptions on the $\mathcal{G}$, the sub-stochastic matrix Q and the stochastic matrix S are, respectively, out-connected and irreducible and aperiodic. A similar structure will be inherited by its powers

$$P^t = \begin{pmatrix} Q(t) & R(t) \\ 0 & S(t) \end{pmatrix},$$

where $Q(t)$, $R(t)$, and $S(t)$ satisfy $Q(0) = R(0) = S(0)$ and

$$Q(t+1) = Q(t)Q\,, \qquad R(t+1) = Q(t)R + R(t)S\,, \qquad S(t+1) = S(t)S\,. \tag{6.5}$$

Solving the recursions (6.5) above gives

$$Q(t) = Q^t\,, \qquad R(t) = \sum_{1 \le k \le t} Q^{t-k} R S^{k-1}\,, \qquad S(t) = S^t$$

Since S is irreducible and aperiodic, it follows from the previous part that S^t converges to $\mathbb{1}\pi'$ as t grows large. On the other hand, since Q is sub-stochastic and out-connected, Proposition 2.6 implies that Q has spectral radius strictly smaller than 1, so that Q^t converges to 0 as t exponentially fast as t grows large. The fact that Q is exponentially stable and S^t is convergent to $\mathbb{1}\pi'$ imply that, as t grows large, $R(t)$ converges to some matrix $R(\infty)$. Taking the limit as t grows large in both sides of the second recursion in (6.5), we get that $R(\infty)$ must satisfy the relation $R(\infty) = R(\infty)S$. This says that the rows of $R(\infty)$ are invariant distributions of S and thus, by Proposition 2.8, they are all multiples of π. On the other hand, since P^t is stochastic, one has that $\mathbb{1} = Q(t)\mathbb{1} + R(t)\mathbb{1} \stackrel{t\to\infty}{\longrightarrow} R(\infty)\mathbb{1}$, so that $R(\infty)\mathbb{1} = \mathbb{1}$. It follows that $R(t) \stackrel{t\to\infty}{\longrightarrow} R(\infty) = \mathbb{1}\pi'$, so that $P^t \stackrel{t\to\infty}{\longrightarrow} \mathbb{1}\pi'$, thus completing the proof. □

The following corollary is a straightforward consequence of previous result. It asserts that if the condensation graph of $\mathcal{G}_P$ contains a singular sink component that is aperiodic (in particular, if $\mathcal{G}_P$ is strongly connected and aperiodic), the averaging dynamics converges to a consensus point while the flow dynamics converges to a rescaled version of the unique invariant probability.

[1] This can be seen for instance by writing M in its Jordan form.

Corollary 6.1. Let $\mathcal{G}$ be a graph such that $s_{\mathcal{G}} = 1$ and the connected component of $\mathcal{G}$ corresponding to the unique sink of its condensation graph is aperiodic. Let P be the normalized weight matrix of $\mathcal{G}$ and let $\pi = P'\pi$ be its unique invariant probability. Then, the discrete-time averaging dynamics (6.2) satisfies

$$\lim_{t\to+\infty} x(t) = \alpha \mathbb{1}\,, \qquad \alpha = \pi' x(0)\,.$$

We conclude this section by discussing two important applications of the linear distributed averaging dynamics.

Social learning and the wisdom of crowds

In the context of *social learning* [22], the averaging dynamics is given an interpretation of opinion dynamics. There is an underlying state of the world represented by a scalar parameter θ, and every agent $i \in \mathcal{V}$ observes a noisy version of it. Such noisy observation coinciding with the agent's initial opinion, so that we can write

$$x_i(0) = \theta + \xi_i\,, \qquad i \in \mathcal{V}\,,$$

where the ξ_i are zero-mean random variables representing the noise in the agents' observations of the state of the world. Through the dynamics $x(t+1) = Px(t)$ (that in this context is known as DeGroot dynamics), a social aggregation of information takes place. This produces a final estimate $\overline{x}$ of the state or the world θ that is common to all individuals. Observe that, since all the noises ξ have zero mean, both the expectation of every $x_i(0)$ and that of $\overline{x}$ equal θ, i.e., they are all unbiased estimators of θ. When $\overline{x}$ is a better estimate of θ than any single node's initial measurement $x_i(t)$, one talks about the *wisdom of crowd* [35].

A typical assumption is that the noises ξ_i are independent with the same variance σ^2. In this case, the variance of the asymptotic consensus value $\overline{x}$ satisfies

$$\sigma^2_{\overline{x}} = \sigma^2 \sum_i \pi_i^2\,.$$

Observe that, since $\pi_i < 1$ for all i, one has that $\sum_i \pi_i^2 < \sum_i \pi_i = 1$, so that in this case $\sigma^2_{\overline{x}} < \sigma^2$, which we can interpret as saying that the crowd is wiser than any single individual. In social networks such that there exists some constant C independent of the network order n such that $\pi_i \leq C/n$ for every agent $i \in \mathcal{V}$ —that are referred to as *democratic* societies— one gets that $\sigma^2_{\overline{x}} \leq C\sigma^2/n$, i.e., the variance of the asymptotic consensus value is bounded by a term inversely proportional to the network order.

Distributed computation of averages [36, 5, 6]

Let the node set $\mathcal{V}$ describe a set of sensors that are deployed in some region in order to collect measurements of some quantity of interest (for example, the temperature). Assume that these sensors have limited communication and computation capabilities that allow each of them to exchange information only

with those other sensors that are close enough in space. Let the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ describe the pattern of vicinity among the sensors i and j so that there is an undirected link between node i and node j if they can communicate to each other (possibly using link weights decreasing with distance). Then, one can design a distributed algorithm for computing the average of the sensor's measurements based on the averaging dynamics.

More precisely, let $x_i(0)$ be the measurement of each node $i \in \mathcal{V}$, and $x(0) \in \mathbb{R}^n$ be the vector of all these measurements. We are interested in designing an iterative distributed algorithm that allows the nodes to compute $x^* = \frac{1}{n}\sum_{i\in\mathcal{V}} x_i(0) = \frac{1}{n}\mathbb{1}'x(0)$. Here distributed means that every node updates its state based on information that it received from its neighbors in $\mathcal{G}$ only.

A first attempt consists in designing the distributed computation algorithm by letting $x_i(t+1) = \sum_j P_{ij}x_j(t)$ for all $i \in \mathcal{V}$, where $P_{ij} = W_{ij}/w_i$ are the entries of the normalized weight matrix of the communication graph $\mathcal{G}$. More compactly, we can write $x(t+1) = Px(t)$. In this case, the previous results imply that we get convergence to a consensus vector $\lim_{t\to\infty} x(t) = \overline{x}\mathbb{1}$ with consensus value $\overline{x} = \sum_{i\in\mathcal{V}} \pi_i x_i(0)$. If the communication graph $\mathcal{G}$ is regular, then $\pi_i = 1/n$ for all i, so that we have succeeded. However, if $\mathcal{G}$ is not regular, what our algorithms computes $\overline{x}$ is not the arithmetic average $\frac{1}{n}\sum_{i\in\mathcal{V}} x_i(0)$ as desired, but rather a weighted version of it, $\sum_{i\in\mathcal{V}} \pi_i x_i(0)$. How can we overcome this problem?

Observe that the communication graph is undirected, hence in particular balanced. Recall that, for balanced graphs, the centrality vector is proportional to the degree vector, specifically, we have $\pi_i = w_i/(n\overline{w})$ for every node i. It is reasonable to assume that every sensor knows its degree w_i. Then, one idea is to normalize the initial condition as

$$x'_i(0) = \frac{x_i(0)}{w_i}\,, \qquad i \in \mathcal{V}\,,$$

and then run the consensus algorithm starting from it. Compactly, we get

$$x'(t+1) = Px'(t)\,. \tag{6.6}$$

As a consequence of the previous results, we get

$$\lim_{t\to+\infty} x'_i(t) = \sum_j \pi_j x'_j(0) = \sum_j \frac{w_j}{\overline{w}n}\frac{x_j(0)}{w_j} = \frac{1}{\overline{w}}\left(\frac{1}{n}\sum_j x_j(0)\right)\,. \tag{6.7}$$

This means that, with the iteration $x'(t+1) = Px'(t)$, what the nodes compute is the arithmetic average of the initial measurements divided by the average degree $\overline{w}$. If every node knew the average degree, we would be done! However, in practice this might not be the case: while every sensor knows its degree (i.e., the number of sensors it can communicate to), it might not know the degree of all other sensors in order to compute their average. The way to get around this is to run another consensus algorithm in parallel to (6.6) that allows nodes to

compute the average degree in a distributed way. This can be done by setting an initial condition

$$x_i''(0) = \frac{1}{w_i}\,, \qquad i \in \mathcal{V}\,,$$

and then running the consensus algorithm

$$x''(t+1) = Px''(t)\,. \tag{6.8}$$

As a consequence of the previous results, we get

$$\lim_{t\to+\infty} x_i''(t) = \sum_j \pi_j x_j''(0) = \sum_j \frac{w_j}{\overline{w}n}\frac{1}{w_j} = \frac{1}{\overline{w}}\,. \tag{6.9}$$

By combining (6.7) and (6.9) we get

$$\lim_{t\to+\infty} \frac{x_i'(t)}{x_i''(t)} = \frac{x^*/\overline{w}}{1/\overline{w}}\,. \tag{6.10}$$

6.2 Linear network flow dynamics

In this section, we introduce the linear network flow dynamics. These are models of transport phenomena derived from mass balance equations in physical networks and find applications in several diverse fields including infrastructure networks, epidemiology, ecology, pharmacokinetics, and more. In some of the literature, such linear network flow dynamics are referred to as *compartmental systems* [**?**].

Consider a physical network described as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with normalized weight matrix P, where the node set $\mathcal{V}$ represent cells or compartments containing a homogeneous mass of some matter, the link set $\mathcal{E}$ represents physical constraints, so that the the matter above can flow directly from cell i to cell j, and the entries P_{ij} of the normalized weight matrix P (that in traffic applications is often referred to as the *routing* matrix) represent the fraction of outflow from cell i that goes directly to cell j. For every $i \in \mathcal{V}$, let $y_i(t)$ represent the quantity of mass in cell i at time $t \geq 0$. Then, the law of mass balance —or mass conservation— prescribes that

$$y_i(t+1) = y_i(t) + \sum_j f_{ji}(t) - \sum_j f_{ij}(t)\,, \tag{6.11}$$

where the variables $f_{ij}(t) \geq 0$ represent the flow of mass from cell i to cell j at time t and are such that, whenever $(i,j) \notin \mathcal{E}$, $f_{ij}(t) = 0$ for all $t \geq 0$. Discrete-time linear network flow dynamics are then obtained by combining the law of mass conservation (6.11) with the linear dependance

$$f_{ij}(t) = P_{ij} y_i(t) \tag{6.12}$$

of the flow variables on the mass variables. Observe that, since $\sum_j P_{ij} = 1$ for all $i \in \mathcal{V}$, (6.11) and (6.12) imply that

$$y_i(t+1) = \sum_j P_{ji} y_j(t)\,, \qquad i \in \mathcal{V}\,, \quad t \geq 0\,, \tag{6.13}$$

which may be rewritten more compactly as

$$y(t+1) = P'y(t)\,, \tag{6.14}$$

where the the state vector $y(t) \in \mathbb{R}_+^{\mathcal{V}}$ is obtained by stacking up all cells' mass variables $y_i(t)$. We shall refer to (6.14) as the discrete time linear network flow dynamics on $\mathcal{G}$. Observe that (6.14) has the physically meaningful property of preserving non-negativity: if $y(0) \geq 0$ entrywise, then so is $y(t)$ for all $t \geq 0$.

Linear network flow dynamics (6.14) can be interpreted as the the dual of the linear averaging dynamics (6.2). In particular, equilibria

$$y = P'y$$

of the network flow dynamics (6.14) are invariant distributions of $\mathcal{G}$. It follows from Corollary 2.4 that, whenever $s_{\mathcal{G}} = 1$, such invariant distributions are unique up to a multiplicative constant. In that case then every equilibrium of the linear flow dynamics (6.14) is in the form $y = \beta\pi$, where $\pi = P'\pi$ is the invariant probability distribution of $\mathcal{G}$. Moreover, by taking the scalar product of both sides of (6.14) with the all-one vector $\mathbb{1}$, one gets $\mathbb{1}'y(t+1) = \mathbb{1}'P'y(t) = \mathbb{1}'y(t)$, so that an induction argument implies that

$$\mathbb{1}'y(t) = \mathbb{1}'y(0)\,, \qquad t \geq 0\,, \tag{6.15}$$

i.e., the total mass in the systems is preserved by the linear network flow dynamics (6.14). By combining these arguments with Propostion 6.1, we get the following result that can be considered as an analogue of Corollary 6.2 for linear network flow dynamics.

Corollary 6.2. Let $\mathcal{G}$ be a graph such that $s_{\mathcal{G}} = 1$ and the connected component of $\mathcal{G}$ corresponding to the unique sink of its condensation graph is aperiodic. Let P be the normalized weight matrix of $\mathcal{G}$ and let $\pi = P'\pi$ be its unique invariant probability. Then, the discrete-time linear network flow dynamics (6.14) satisfy

$$\lim_{t \to +\infty} y(t) = \beta\pi\,, \qquad \beta = \mathbb{1}'y(0)\,.$$

6.3 Averaging and flow dynamics with inputs

The distributed averaging and linear flow dynamics introduced in the previous sections admit interesting extensions to situations where also exogenous inputs are present. In this section, we shall describe such models with exogenous inputs. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph and let $\mathcal{S} \subseteq \mathcal{V}$ be a nonempty subset of nodes.

We shall modify the distributed averaging and linear flow dynamics in such a way that the states of nodes in $\mathcal{S}$ are no longer updated according to (6.1) and (6.13), respectively, but rather they are put equal to some exogenous inputs.

Formally, let $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$ be the set of regular nodes and let us partition the normalized weight matrix P in its $\mathcal{S} \times \mathcal{S}$, $\mathcal{S} \times \mathcal{R}$, $\mathcal{R} \times \mathcal{S}$, and $\mathcal{R} \times \mathcal{R}$ blocks so that

$$P = \begin{array}{c} \begin{array}{cc} \mathcal{R} & \mathcal{S} \end{array} \\ \left[\begin{array}{cc} Q & E \\ F & G \end{array} \right] \begin{array}{c} \mathcal{R} \\ \mathcal{S} \end{array} \end{array}.$$

Moreover, partition the state vectors of the distributed averaging and linear flow dynamics respectively as

$$x(t) = \left[\begin{array}{c} \underline{x}(t) \\ u(t) \end{array} \right] \begin{array}{c} \mathcal{R} \\ \mathcal{S} \end{array}, \qquad y(t) = \left[\begin{array}{c} \underline{y}(t) \\ v(t) \end{array} \right] \begin{array}{c} \mathcal{R} \\ \mathcal{S} \end{array},$$

where $u(t) \in \mathbb{R}^{\mathcal{S}}$ and $v(t) \in \mathbb{R}_+^{\mathcal{S}}$ are exogenous input vectors. Then, the distributed linear averaging on $\mathcal{G}$ with input set $\mathcal{S}$ can be written as

$$\underline{x}(t+1) = Q\underline{x}(t) + Eu(t)\,, \tag{6.16}$$

while the linear network flow dynamics on $\mathcal{G}$ with input set $\mathcal{S}$ can be written as

$$\underline{y}(t+1) = Q'\underline{y}(t) + F'v(t)\,. \tag{6.17}$$

The averaging dynamics model with inputs (6.16) has been considered in the context of rendez-vous dynamics where nodes in $\mathcal{S}$ are interpreted as leader units whose motion can be externally controlled. A similar interpretation as leader nodes is possible in opinion dynamics. In this case, of particular interest is the case when $u(t) = u$, namely when the opinions of the input nodes do not change in time. In this case we talk about *stubborn* nodes.

On the other hand, the linear flow dynamics with inputs (6.17) are used to model transport dynamics in physical networks. In particular,

$$\lambda(t) = F'v(t)$$

can be interpreted as the exogenous inflow vector, while the entries Q_{ij} of the substochastic matrix Q —that is is often referred as the routing matrix— represent the fractions of outflow from node i that is routed to node j, so that

$$f_{ij}(t) = Q_{ij}y_i(t)\,, \qquad \mu_i(t) = (1 - \sum\nolimits_j Q_{ij})y_i(t)$$

are respectively the flow from node i to node j and the exogenous outflow from node i that leaves the network directly. The quantity $\mathbb{1}'(I - Q')\underline{y}(t) = \mathbb{1}'\mu(t)$ is thus the total outflow at time t. Notice that

$$\mathbb{1}'y(t+1) - \mathbb{1}'y(t) = \mathbb{1}'F'v(t) + \mathbb{1}'Q'y(t) - \mathbb{1}'y(t) = \mathbb{1}'\lambda(t) - \mathbb{1}'\mu(t)$$

that says that the variation of mass in $\mathcal{R}$ is equal to the difference between the total inflow and the total outflow. This extends the property of conservation of mass we had seen for the autonomous flow dynamics.

The following result guarantees convergence for both the distributed averaging and the linear flow dynamics with constant inputs provided that the set $\mathcal{S}$ is globally reachable.

Proposition 6.2. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph and $\mathcal{S} \subseteq \mathcal{V}$ a nonempty globally reachable subset of nodes. Let $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$, and let Q be the $\mathcal{R} \times \mathcal{R}$ block of the normalized weight matrix P. Then, the matrix $I - Q$ is invertible with nonnegative inverse matrix

$$H = (I - Q)^{-1} = \sum_{k \geq 0} Q^k \,.$$

Moreover, for every constant input vectors $u \in \mathbb{R}^{\mathcal{S}}$ and $v \in \mathbb{R}_+^{\mathcal{S}}$, the distributed averaging and linear flow dynamics with inputs (6.16) and (6.17) satisfy

$$\lim_{t \to +\infty} \underline{x}(t) = HEu, \qquad \lim_{t \to +\infty} \underline{y}(t) = H'F'v \tag{6.18}$$

for every initial state vectors $\underline{x}(0) \in \mathbb{R}^{\mathcal{R}}$, $\underline{y}(0) \in \mathbb{R}_+^{\mathcal{R}}$.

Proof. It follows from Proposition 2.6 that Q is asymptotically stable. Result follows by taking the limit $t \to +\infty$ in the closed formulas

$$\underline{x}(t) = Q^t \underline{x}(0) + \sum_{h=0}^{t-1} Q^{t-h-1} Eu, \quad \underline{y}(t) = Q^t \underline{y}(0) + \sum_{h=0}^{t-1} Q'^{t-h-1} F'v \,.$$

□

Corollary 6.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph and $\mathcal{S} \subseteq \mathcal{V}$ a nonempty globally reachable subset of nodes. Then, for every initial state $y(0) \in \mathbb{R}_+^{\mathcal{V}}$ that is supported on $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$ there exists a positive constant α such that the linear flow dynamics (6.17) with input $v = 0$ satisfies

$$\sum_{i \in \mathcal{R}} y_i(t) \leq \exp(-\alpha t + 1) \sum_{i \in \mathcal{R}} y_i(0) \,.$$

Proof. The claim simply follows from Proposition 6.2 and the fact that the dynamics are linear. More specifically, Proposition 6.2 in the special case $v = 0$ implies that $\underline{y}(t) \overset{t \to +\infty}{\longrightarrow} 0$. Let $\bar{t} := \min\{t \geq 0 : \sum_{i \in \mathcal{R}} y_i(t) \leq \frac{1}{e} \sum_{i \in \mathcal{R}} y_i(0)\}$. The, by linearity, one gets that $\sum_{i \in \mathcal{R}} y_i((k+1)\bar{t}) \leq \sum_{i \in \mathcal{R}} y_i(k\bar{t})$ for every $k \geq 0$. By induction, this implies that $\sum_{i \in \mathcal{R}} y_i(k\bar{t}) \leq \sum_{i \in \mathcal{R}} y_i(0) \exp(-k)$. On the other hand, notice that, since the input $v = 0$, we have $\sum_{i \in \mathcal{R}} y_i(t+1) \leq \sum_{i \in \mathcal{R}} y_i(k\bar{t})$ for every $t \geq 0$. Hence, we get that

$$\sum_{i \in \mathcal{R}} y_i(t) \leq \sum_{i \in \mathcal{R}} y_i(\lfloor t/\bar{t} \rfloor \bar{t}) \leq \exp(-\lfloor t/\bar{t} \rfloor) \sum_{i \in \mathcal{R}} y_i(0) \leq \exp(-t/\bar{t} + 1) \sum_{i \in \mathcal{R}} y_i(0) \,.$$

Hence, the claim follows with $\alpha = 1/\bar{t}$. □

Averaging dynamics with stubborn nodes

In this paragraph we focus on the averaging dynamics with stubborn nodes

$$\underline{x}(t+1) = Q\underline{x}(t) + Eu$$

on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, where, for a nonempty globally reachable set of stubborn nodes $\mathcal{S} \subseteq \mathcal{V}$, $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$, we have $Q = P_{|\mathcal{R}\times\mathcal{R}}$, $E = P_{|\mathcal{R}\times\mathcal{S}}$, and $u \in \mathbb{R}^{\mathcal{S}}$ is the vector of inputs. These averaging dynamics with inputs have become popular as opinion dynamics models whereby the stubborn nodes play the role of media or leaders not changing opinion (at least in a certain time range) [1]. We now make further considerations on the first equation in (6.18) describing the asymptotic states of the regular nodes. Notice first of all that, contrarily to the usual average consensus, the initial states of the regular nodes do not play any role in the asymptotics that instead only depends on the fixed state of the stubborn nodes. If we put

$$x = \begin{pmatrix} HEu \\ u \end{pmatrix} \in \mathbb{R}^{\mathcal{V}}, \qquad H = (I - Q)^{-1} \tag{6.19}$$

it is easy to see that such vector x can be equivalently expressed as the only vector satisfying the equations

$$\begin{aligned} &\sum_{j\in\mathcal{V}} L_{ij}x_j = 0 \qquad i \in \mathcal{R}\,, \\ &x_s = u_s \qquad s \in \mathcal{S}\,. \end{aligned} \tag{6.20}$$

where $L = D - W$ is the Laplacian of the graph $\mathcal{G}(\Sigma)$. To see the connection, it is sufficient to first multiply x in (6.19) by $I - P$ and then multiply each component x_i for $i \in \mathcal{R}$ by the outdegree w_i.

Remark 5. *Notice how this is the same exact problem considered in (5.10) in the context of electrical networks. There u was the assignment of the voltage on a subset of nodes and the solution x was the voltage in the remaining nodes. In that context we were considering undirected networks, while this assumption is not needed here. This connection has two remarkable consequences:*

- *Item 1. of Proposition 5.3 that had remained unproven and was concerned on the existence of a solution to (6.20) is now an immediate consequence of the above characterization. Indeed when $\mathcal{G}$ is strongly connected, $\mathcal{S}$ is for sure globally reachable (as long as it is non-empty) and the solution x is then simply the asymptotic value of the corresponding average dynamics with stubborn nodes considered above*

- *In case when the $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is undirected, the electrical interpretation can be fruitfully used to simplify the computation of the vector asymptotic opinion vector x, using the reduction techniques illustrated in Remark 4.*

We now make some further comments on the characterization of the asymptotic opinion vector. An important point is that, when there are at least two stubborn nodes with different opinion, typically the solution of (6.20) is not constant over $\mathcal{R}$, i.e., different regular agents reach a different equilibrium opinion, depending on their position in the network relative to the stubborn agents. Thus a consensus is not reached, and a condition of *disagreement* among the regular agents persists.

The opinion of the singular regular agents also admits the following explicit characterization. First observe that, since $Q\mathbb{1} + E\mathbb{1} = \mathbb{1}$, one has $HE = (I - Q)^{-1}E\mathbb{1} = \mathbb{1}$. Since H and E are nonnegative matrices, this shows that rows of HE are probability vectors. Then, we have

$$x_i = \sum_{s\in\mathcal{S}} (HE)_{is} u_s \,, \qquad i \in \mathcal{R} \,, \tag{6.21}$$

$$(HE)_{is} = \sum_{k\geq 0} \sum_{\substack{i_1,\dots,i_{k-1}\in\mathcal{R} \\ i_0=i, i_k=s}} \prod_{1\leq h\leq k} P_{i_{h-1}i_h} \,, \qquad i \in \mathcal{R} \,,\ s \in \mathcal{S} \,, \tag{6.22}$$

i.e., the asymptotic opinion x_i of a regular agent i is a convex combination of the opinions x_s of the stubborn agents, each weighted by a coefficient $(HE)_{is}$. Such coefficient is expressed in terms of the normalized weight $\prod_{1\leq h\leq k} P_{i_{h-1}i_h}$ of all walks $i = i_0, i_1, \dots, i_k = s$ in the graph $\mathcal{G}$ that start in the regular node i and terminate in the stubborn nodes s, without ever passing through any stubborn node in any intermediate step. Each length-k walk starting in node i and terminating in the stubborn node s is taken into account with a weight equal to the product $P_{ii_1}P_{i_1i_2}\dots P_{i_{k-1}s}$ of the normalized weights of the successive links. Observe that, since the normalized weight of a link is smaller than 1, longer paths have smaller weight than shorter ones. Indeed, equation (6.21) captures the intuition that, if a regular agent i is closer in the social network to one stubborn agent s than to the other ones, so that there are more shorter-length higher-weight paths connecting i to s than to any other stubborn node in $\mathcal{S}$, then the equilibrium opinion x_i tends to be biased towards opinion x_s. On the other hand, regular agents which are more equidistant from all the stubborn agents get influenced by them in a more balanced way. This is illustrated in the following example.

Example 6.2. Consider the simple graph consisting of a line L_{n+1} of length n where the two leaf nodes 0 and n are stubborn agents with opinion, respectively, $x_0 = 0$ and $x_n = 1$. Then, the solution of (6.20) was already obtained in Example 5.3 in the context of the electrical networks and is given by

$$x_i = \frac{i}{n} \,, \qquad i = 0, 1, \dots, n \,,$$

i.e., the equilibrium opinions of the regular agents linearly interpolate between the extreme values held by the stubborn agents.

Distributed computation of the Page-Rank centrality

In Section 2.4 we have introduced the so-called page-rank centrality vector of a graph $\mathcal{G}$ as the solution z of the equation

$$z = (1-\beta)\, P' z + \beta \nu \,, \tag{6.23}$$

where P is the normalized weight matrix of $\mathcal{G}$, $\beta \in (0,1)$ is a parameter, and ν is a nonnegative vector. An explicit form of z as the sum of a convergent series was proposed in (2.20). In fact, it proves convenient to think z in (6.23) as the limit of the linear network flow dynamics model with exogenous inputs

$$\underline{y}(t+1) = Q' \underline{y}(t) + \lambda \,, \tag{6.24}$$

where $Q = (1-\beta)P$ and $\lambda = \beta\nu$. Equation (6.24) can be interpreted as the linear network flow dynamics in a graph $\overline{\mathcal{G}} = (\overline{\mathcal{V}}, \overline{\mathcal{E}}, \overline{W})$ where $\overline{\mathcal{V}} = \mathcal{V} \cup \{s\}$, $\overline{\mathcal{E}} = \mathcal{E} \cup \bigcup_{i \in \mathcal{V}} \{((s,i)), (i,s)\}$,

$$\overline{W} = \begin{bmatrix} (1-\beta)W & \beta w \\ \beta \nu' & 0 \end{bmatrix},$$

the input set $\mathcal{S} = \{s\}$ and the input $v = 1$. Clearly, $\mathcal{S}$ is globally reachable in $\overline{\mathcal{G}}$, so that, regardless of the initial state $\overline{y}(0)$, the linear network flow dynamics (6.24) satisfies

$$\lim_{t \to +\infty} \underline{y}(t) = z \,. \tag{6.25}$$

In fact, (6.24) can be though of as an iterative distributed algorithm for the computation of the PageRank vector z.

6.4 Continuous-time linear averaging and network flow dynamics

The linear distributed averaging (6.2) and network flow dynamics (6.14) on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ have an analogue in continuous time. Such analogue is given by the linear ordinary differential equations

$$\dot{x} = -Lx, \qquad \dot{y} = -L'y \tag{6.26}$$

where L is the Laplacian of $\mathcal{G}$.

Convergence results analogous to those presented in Sections 6.1 and 6.2 respectively hold true for the the continuous time averaging and flow dynamics on strongly connected graphs, and more in general for graphs with $s_{\mathcal{G}} = 1$. The main differences with respect to the discrete time case are that: (i) the assumption of aperiodicity is no longer required and connectedness alone is sufficient to guarantee convergence; and (ii) the role played by the invariant probability distribution π in discrete time is played by the Laplace invariant probability distribution $\overline{\pi}$. This is summarized in the following result.

Proposition 6.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph and let L be its Laplacian. If $s_{\mathcal{G}} = 1$, then the continuous time distributed averaging and network flow dynamics (6.26) are such that

$$\lim_{t \to +\infty} x(t) = \alpha \mathbb{1}\,, \qquad \alpha = \overline{\pi}' x(0) \tag{6.27}$$

$$\lim_{t \to +\infty} y(t) = \beta \overline{\pi}\,, \qquad \beta = \mathbb{1}' y(0)\,, \tag{6.28}$$

where $\overline{\pi}$ is the unique Laplace invariant probability distribution of $\mathcal{G}$.

In fact, also the linear distributed averaging and network flow dynamics with inputs can be considered in continuous time. Let $\mathcal{S} \subseteq \mathcal{V}$ be a nonempty subset of nodes in $\mathcal{G}$, let $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$, and partition the Laplacian matrix in blocks as

$$L = \begin{array}{c} \begin{array}{cc} \mathcal{R} & \mathcal{S} \end{array} \\ \left[\begin{array}{cc} K & M \\ N & O \end{array} \right] \begin{array}{c} \mathcal{R} \\ \mathcal{S}\,. \end{array} \end{array}$$

Then, for exogenous inputs $u(t) \in \mathbb{R}^{\mathcal{S}}$ and $v(t) \in \mathbb{R}^{\mathcal{S}}_+$, we consider the dynamics

$$\dot{\underline{x}} = -K\underline{x} - Mu\,, \qquad \dot{\underline{y}} = -K'\underline{y} - N'v\,. \tag{6.29}$$

Whenever $\mathcal{S}$ is globally reachable in $\mathcal{G}$, the dynamics above are convergent as stated in the following result.

Proposition 6.4. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a graph. Let $\mathcal{S} \subseteq \mathcal{V}$ a nonempty globally reachable subset of nodes, let $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$, and let K be the $\mathcal{R} \times \mathcal{R}$ block of the Laplacian L. Then, the matrix K is invertible with nonnegative inverse matrix. Moreover, for every constant input vectors $u \in \mathbb{R}^{\mathcal{S}}$ and $v \in \mathbb{R}^{\mathcal{S}}_+$, the continuous time distributed averaging and linear flow dynamics with inputs (6.29) satisfy

$$\lim_{t \to +\infty} \underline{x}(t) = K^{-1}Mu, \qquad \lim_{t \to +\infty} \underline{y}(t) = (K')^{-1}N'v \tag{6.30}$$

for every initial state vectors $\underline{x}(0) \in \mathbb{R}^{\mathcal{R}}$, $\underline{y}(0) \in \mathbb{R}^{\mathcal{R}}_+$.

6.5 Speed of convergence for reversible networks

A crucial issue regarding the dynamical systems introduced in the previous sections is the estimation of their convergence time. Both in applications where such systems are used to model social behaviors or when they are instead used as algorithms, it is clearly a fundamental point to understand how convergence is influenced by the network topology and how it scales with it. This will be the content of this section and the following one. First we will consider the special case when the underlying stochastic matrix is time-reversible, while the more general case will be considered in the next section.

Before presenting the results, it ius useful to make some preliminary comments on the way convergence can be evaluated. We recall below some basic definitions on norms of vectors and matrices. The l_1-, l_2-, and l_∞-norms of a vector $x \in \mathbb{R}^n$ are defined as

$$||x||_1 = \sum_i |x_i|\,, \qquad ||x||_2 = \left(\sum_i x_i^2\right)^{1/2}, \qquad ||x||_\infty = \max_i |x_i|\,,$$

respectively, and the corresponding induced norms for a matrix M are

$$\begin{aligned} ||M||_1 &= \sup_{x\neq 0} \frac{||Mx||_1}{||x||_1} &&= \max_j \sum_i |M_{ij}| \\ ||M||_2 &= \sup_{x\neq 0} \frac{||Mx||_2}{||x||_2} &&= \sigma_{\max}(M) \\ ||M||_\infty &= \sup_{x\neq 0} \frac{||Mx||_\infty}{||x||_\infty} &&= \max_i \sum_j |M_{ij}|\,, \end{aligned} \tag{6.31}$$

where $\sigma_{\max}(M) = \max_i \sqrt{\lambda_i(M'M)}$ is the largest singular value of M, i.e., the square root of the largest eigenvalue of the symmetric positive semidefinite matrix $M'M$.

It is well known that all norms are equivalent in finite-dimensional spaces. For instance we have that $||x||_2 \leq ||x||_1 \leq n^{1/2}||x||_2$ and, consequently, $||M||_i \leq n^{1/2}||M||_j$ for $i, j \in \{1, 2\}$. However, coefficients can play an important role on our applications where we often want to consider the large scale limit, namely the behavior with respect to $n \to +\infty$. For this reason, it is sometime useful to obtain bounds in different norms.

For a time-reversible stochastic matrix P, the rate of convergence of the dynamics in (6.2) can be efficiently studied exploiting the diagonalizability of P. Recall indeed (see Section 5.1) that, denoted by π the invariant probability distribution of P and with $\Pi := \mathrm{diag}\,(\pi)$, the matrix $M := \Pi^{1/2} P \Pi^{-1/2}$ is symmetric. Consequently, M and P have real eigenvalues denoted $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1$.

The following result provides an estimate of the rate of convergence of the sequence P^t to the limit described in Proposition 6.1 in two different matrix norms.

Theorem 6.1. The following estimation holds

$$||P^t - \mathbb{1}\pi'||_2 \leq \sqrt{\frac{\pi^*}{\pi_*}}\lambda^t\,, \tag{6.32}$$

$$||P^t - \mathbb{1}\pi'||_\infty = ||(P')^t - \pi\mathbb{1}||_1 \leq \frac{\lambda^t}{\pi_*}\,, \tag{6.33}$$

where

$$\lambda := \max\{\lambda_2, |\lambda_n|\}\,, \qquad \pi_* = \min_i \pi_i\,, \qquad \pi^* = \max_i \pi_i\,,$$

are, respectively, the second dominant eigenvalue of P, the minimum, and the maximum entry of π.

Proof. For $1 \leq k \leq n$, let $z_{(k)}$ be the eigenvector of M corresponding to eigenvalue λ_k. Since M is symmetric, such eigenvectors can be chosen to form an orthonormal basis of $\mathbb{R}^n$, so that M admits the representation

$$M = \sum_{1 \leq k \leq n} \lambda_k z_{(k)} z'_{(k)} \,.$$

Now, recall that $\lambda_1 = 1$ and observe that $M\Pi^{1/2}\mathbb{1} = \Pi^{1/2}P\mathbb{1} = \Pi^{1/2}\mathbb{1}$, so that the corresponding eigenvector can be chosen equal to

$$z_{(1)} = \Pi^{1/2}\mathbb{1} \,.$$

It follows that, for all $t \geq 0$,

$$\Pi^{1/2}P^t\Pi^{-1/2} = M^t = \sum_{1 \leq k \leq n} \lambda_k^t z_{(k)} z'_{(k)} = \Pi^{1/2}\mathbb{1}\mathbb{1}'\Pi^{1/2} + \sum_{2 \leq k \leq n} \lambda_k^t z_{(k)} z'_{(k)} \,,$$

so that

$$P^t - \mathbb{1}\pi' = \Pi^{-1/2} N \Pi^{1/2} \,, \qquad N = \sum_{2 \leq k \leq n} \lambda_k^t z_{(k)} z'_{(k)} \,. \tag{6.34}$$

It follows from (6.34) that

$$||P^t - \mathbb{1}\pi'||_2 \leq ||\Pi^{1/2}||_2 ||N||_2 \Pi^{-1/2} = \sqrt{\frac{\pi^*}{\pi_*}}\lambda \,,$$

thus proving (6.32).

On the other hand, for $1 \leq i \leq n$, let $e_{(i)}$ be the vector whose i-th entry equals 1 and all whose other entries equal 0. By equating the (i,j)-th entry of the leftmost and rightmost side of (6.34), one gets

$$(P^t)_{ij} - \pi_j = \sqrt{\frac{\pi_j}{\pi_i}} \sum_{k=2}^{n} \lambda_k^t e'_{(i)} z_{(k)} z'_{(k)} e_{(j)} \,,$$

so that

$$\begin{aligned}
||P^t - \mathbb{1}\pi'||_\infty &= \max_i \sum_j |(P^t)_{ij} - \pi_j| \\
&= \max_i \sum_j \sqrt{\frac{\pi_j}{\pi_i}} \left| \sum_{k=2}^{n} \lambda_k^t e'_{(i)} z_{(k)} z'_{(k)} e_{(j)} \right| \\
&\leq \lambda^t \sum_j \frac{\pi_j}{\pi_*} \left(\sum_{k=2}^{n} (e'_{(i)} z_{(k)})^2 \right)^{1/2} \left(\sum_{k=2}^{n} (z'_{(k)} e_{(j)})^2 \right)^{1/2} \\
&\leq \frac{\lambda^t}{\pi_*} \sum_j \pi_j \\
&= \frac{\lambda^t}{\pi_*} \,,
\end{aligned}$$

where the first inequality follows from Cauchy-Schwartz, and the second one from the fact that $\sum_{k=1}^{n}(e'_{(i)}z_{(k)})^2 = ||e_{(i)}||_2^2 = 1$, since $\{z_{(k)}\}_{1\leq k\leq n}$ is an orthonormal basis of $\mathbb{R}^n$.

□

Observe that Theorem 6.1 guarantees that, for time-reversible stochastic matrices, convergence of P^t to the limit matrix $\pi\mathbb{1}'$ is exponentially fast in time, provided that $\lambda_2 < 1$ and $\lambda_n > -1$. Now, $\lambda_2 < 1$ is equivalent, by (v) of Proposition 2.4, to irreducibility of P, i.e., strong connectedness of the graph, which is our standing assumption. On the other hand, the condition $\lambda_n > -1$ is equivalent, by (v) of Proposition 2.4, to the fact that P is aperiodic. Recall that aperiodicity is necessary for global asymptotic convergence as shown in Example **??**.

If we consider $P_{(\text{lazy})} = \frac{1}{2}(P + I)$, we can observe that that the invariant probability of $P_{(\text{lazy})}$ coincides with the one of P, while its eigenvalues are given by $\lambda_k^{(\text{lazy})} = (1 + \lambda_k)/2$ where λ_k are the eigenvalues of P. Notice in particular that $\lambda_n^{(\text{lazy})} = (1+\lambda_n)/2 \in [0, \lambda_2^{(\text{lazy})})$. As a consequence, for $P_{(\text{lazy})}$, the bounds (6.32) and (6.33) hold true with $\lambda = \lambda_2^{(\text{lazy})}$.

The quantity $1 - \lambda_2$ is called the *spectral gap* of P, while its inverse

$$\tau_{\text{rel}} := \frac{1}{1 - \lambda_2}$$

is called the *relaxation time* of P. The reason for this is the following: we have that $\lambda_2^t \leq \epsilon$ iff $t \geq \ln \epsilon^{-1} / \ln \lambda_2^{-1}$ and thus

$$\inf\{t \,:\, \lambda_2^t \leq \epsilon\} = \left\lceil \frac{\ln \epsilon^{-1}}{\ln \lambda_2^{-1}} \right\rceil$$

Standard calculus considerations yield $1/\ln \lambda_2^{-1} \leq \tau_{\text{rel}}$ so that

$$\inf\{t \,:\, \lambda_2^t \leq \epsilon\} \leq (\log \epsilon^{-1})\tau_{\text{rel}}$$

This shows that τ_{rel} is an estimation of the time needed for the exponential term λ^t to contract of a factor $1/e$. This estimation is particularly accurate when λ_2 is close to 1, since we know that $1/\ln \lambda_2^{-1} \sim \tau_{\text{rel}}$ for $\lambda_2 \to 1$. This happens in many situations when we consider large scale graphs (n large).

It follows from Theorem 5.1 that the relaxation time can be estimated in terms of the bottleneck ratio:

$$\frac{1}{2\Phi} \leq \tau_{\text{rel}} \leq \frac{2}{\Phi^2}$$

As expected, a small bottleneck ratio yields a slow convergence time of the underlying average and flow dynamics.

Chapter 7

Markov chains and random walks

In this chapter we introduce a first example of stochastic network dynamics: random walks. Random walks are Markov chains whose state space is identified with the node set of a graph: the links of this graph in turn represent the possible state transitions. As we shall see, every Markov chain with finite state space can be thought of as a random walk on a properly defined (directed, weighted) graph, and this chapter shall emphasize the relationships between the geometrical properties of the graph and properties of the associated random walk. On the other hand, many stochastic network dynamics to be studied in the next chapters can be modeled as interacting systems, i.e., Markov chains whose finite state space is best interpreted as the direct product of local state spaces of the nodes of an interaction graph.The general results for Markov chains presented in this chapter will then be instrumental for the study of such interacting systems in the next chapters.

Let us start with a simple example before introducing the general framework. Consider a gambler betting on the binary outcome of a sequence of independent events, all having probability of success p. Assume that the gambler starts with a budget of $k > 0$ euros, bets one euro each time, and quits the first time she either runs out of money or has accumulated a capital of n euros, where $n > k$ is a pre-defined threshold. We can then visualize the dynamics of the gambler's capital as a random walk on the line graph depicted in Figure 7.1: at each time the capital is increased or decreased by one euro —correspondingly the walk proceeds one hop right of left— with probability p and $q = 1 - p$, respectively, until the first time either level 0 or level n is reached. Natural questions we may be interested in concern what the probability is that the gambler quits with 0 or n euros —equivalently, what the probability is that the random walk terminates in state 0 or n— as a function of k and n, or how long it will take her to quit in the average. Answers to these questions will be presented in Section 7.3.

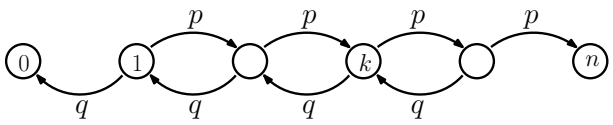


Figure 7.1: At each time step the gambler bets one euro. She wins with probability p and loses with probability q. The gambler's initial capital is $k > 0$ euros, and the gambler quits when she either gets broke (corresponding to a capital of 0 euros) or reaches a pre-defined capital target of $n > k$ euros. The gambler's capital can thus be represented as a random walk on the graph above. What is the probability that the gambler eventually quits having achieved her target of n euros as opposed to getting broke? What is the expected time it takes her to quit (either broke or with n euros)? Note that specifying the out-going links of nodes 0 and n is not relevant to answer these questions.

7.1 Some basic notions

The gambler's example can be generalized as follows. Given a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, imagine a particle moving from one of its nodes to another: each time the particle randomly chooses the next node to move to among the out-neighbors of the current node with probability proportional the weight of the link pointing from the current node to next one. (See Figure 7.2.)

Formally, what we are describing is a discrete-time stochastic process $X(t)$, $t = 0, 1, \ldots$, with state space $\mathcal{X}$ coinciding with the node set $\mathcal{V}$, and such that, for any states i and j in $\mathcal{X}$

$$\mathbb{P}\big(X(t+1) = j | X(0) = i_0, X(1) = i_1, \ldots, X(t-1) = i_{t-1}, X(t) = i_t\big) = P_{i_t j}\,, \tag{7.1}$$

where $P = D^{-1}W$ is the normalized weight matrix of the graph $\mathcal{G}$. Equation (7.1) states that the future state $X(t+1) = j$ is independent from the past —i.e., the trajectory $X(0) = i_0, X(1) = i_1, \ldots, X(t-1) = i_{t-1}$ followed by the particle before time t— given the present state $X(t) = i_t$, and that the probability of moving from node i_t to node j equals $P_{i_t j}$. The fact that the dependance of the future state on the history is limited to the present state is known as the *Markov property*, and the process $X(t)$ is usually referred to as a discrete-time *Markov chain* with *transition probability matrix* P.

In fact, a discrete-time Markov chain $X(t)$ with finite state space $\mathcal{X}$ can be associated to any pair of a *stochastic matrix* P, i.e., a nonnegative square matrix whose entries are labeled by the corresponding elements of $\mathcal{X}$ and whose rows all sum up to 1, and an *initial probability distribution* $\pi(0)$, i.e., a probability vector with entries corresponding to the elements of $\mathcal{X}$ such that

$$\mathbb{P}(X(0) = i) = \pi_i(0)\,, \qquad i \in \mathcal{X}\,. \tag{7.2}$$

On the other hand, notice that any stochastic matrix P can always be thought of as the normalized weight matrix of some (in general, weighted and

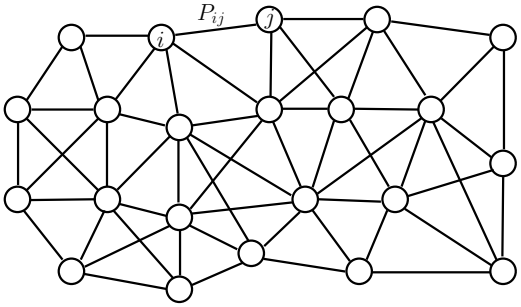


Figure 7.2: A random walk on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a Markov chain with transition probabilities $P_{ij} = W_{ij}/w_i$, where $w_i = \sum_j W_{ij}$ is node i's out-degree. When the graph is undirected (as is the case in the one displayed above), the normalized weight vector is a stationary probability distribution and the Markov chain is reversible with respect to it.

directed) graph. E.g., it is sufficient to consider the graph $\mathcal{G}_P = (\mathcal{V}, \mathcal{E}, P)$ whose node set $\mathcal{V} = \mathcal{X}$ coincides with the state space, whose link set $\mathcal{E}$ consists of those pairs $(i, j) \in \mathcal{V} \times \mathcal{V}$ for which $P_{ij} > 0$, and whose weight matrix coincide with P. The graph $\mathcal{G}_P$ will be referred to as the *normal graph* associated to P. Throughout this chapter, when using some graph-theoretic terminology for a Markov chain, we shall implicitly intend it as referred to the normal graph $\mathcal{G}_P$ associated to its transition probability matrix P. For example, we shall say that a subset of states $\mathcal{S} \subseteq \mathcal{X}$ is *reachable* from a state $i \in \mathcal{X}$ if there exists a path in $\mathcal{G}_P$ starting from i and ending in some $j \in \mathcal{S}$.

Clearly, the normal graph $\mathcal{G}_P$ is not the only graph whose normalized weight matrix coincides with a given stochastic matrix P, as any graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ where $W = DP$ for some positive diagonal matrix D also has P as normalized weight matrix. In the literature, a Markov chain $X(t)$ with transition probability matrix P coinciding with the normalized weight matrix of a graph $\mathcal{G}$, is often referred to as the *standard random walk* on $\mathcal{G}$. In these notes, we will use the two terms Markov chain and random walk interchangeably, depending on whether or not we want to emphasize the role of the underlying graph. In the case when $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is unweighted (e.g. W is the $0-1$ adjacency matrix of $(\mathcal{V}, \mathcal{E})$), the corresponding random walk $X(t)$ is referred to as the *simple random walk* on $\mathcal{G}$.

In order to determine the probability distribution of the trajectories of a Markov chain $X(t)$, one needs to specify both a transition probability matrix P and an initial probability distribution to be denoted by $\pi(0)$. Then, the Markov chain's trajectory satisfies

$$\mathbb{P}(X(0) = i_0, X(1) = i_1, \dots X(t) = i_t) = \pi_{i_0}(0) \prod_{1 \le s \le t} P_{i_{s-1} i_s}\,, \qquad t \ge 0\,. \quad (7.3)$$

From equation (7.3), for every time $t \ge 0$, one can derive recursive formulas for

the marginal probability distribution $\pi(t)$ of $X(t)$, whose entries

$$\pi_i(t) := \mathbb{P}(X(t) = i)\,, \qquad i \in \mathcal{X}\,,$$

coincide with the probability that the Markov chain is in a specific node i at time t, as well as for the t-step transition probability matrix $P(t)$, whose entries

$$P_{ij}(t) = \mathbb{P}(X(t) = j | X(0) = i)\,, \qquad i, j \in \mathcal{X}\,,$$

specify the conditional probability that the chain is in node j at time t given that it started in node i at time 0. Such recursive equations read as

$$\pi(t+1) = P'\pi(t) \tag{7.4}$$

and, respectively,

$$P(t+1) = P(t)P\,, \qquad P(t+1) = PP(t)\,. \tag{7.5}$$

In particular, the two recursions in (7.5) are known as the discrete-time *Kolmogorov backward equation* and *Kolmogorov forward equation*, respectively. The recursive equations (7.4) and (7.5) imply that

$$\pi'(t) = \pi'(0)P^t\,, \qquad P(t) = P^t\,, \qquad t \geq 0\,.$$

Often, it is useful to consider deterministic initial conditions $X(0) = i$ for some $i \in \mathcal{X}$, so that $\pi(0) = \delta^{(i)}$ is a delta distribution concentrated on the node i. For the probability induced by the process $X(t)$ for such an initial condition we use the notation:

$$\mathbb{P}_i(\,\cdot\,) = \mathbb{P}(\,\cdot\,|X(0) = i)\,.$$

Similarly, we use the notation

$$\mathbb{E}_i[\,\cdot\,] = \mathbb{E}[\,\cdot\,|X(0) = i]$$

for the corresponding conditional expected value.

7.2 Stationarity, convergence, and ergodicity

Let P be a stochastic matrix and let

$$\pi = P'\pi \tag{7.6}$$

be an invariant probability distribution for P. If the initial state of the Markov chain $X(0)$ has probability distribution $\pi(0) = \pi$, then equations (7.4) and (7.6) imply that

$$\pi(t) = \pi\,, \qquad t \geq 0\,,$$

i.e., each $X(t)$ has the same marginal probability distribution π as $X(0)$. That justifies the term invariant probability distribution when referring to π: if the

initial distribution of the Markov chain is π, it does not matter how long we wait, the marginal probability distribution at any time t will always be π.

Since the dynamics of the marginal probabilities (7.4) coincide with the linear network flow dynamics (6.14), we can apply all convergence results considered in Chapter 6. In particular, a straightforward consequence of Corollary 6.2 is the following.

Corollary 7.1 (Convergence in probability)**.** Let P be an irreducible and aperiodic stochastic matrix and $\pi = P'\pi$ be its unique stationary probability vector. Let $\pi(t)$, for $t = 0, 1, \ldots$, be the probability distribution vector of a Markov chain with transition probability matrix P. Then,

$$\lim_{t \to +\infty} \pi(t) = \pi \,,$$

for every initial probability distribution $\pi(0)$.

Corollary 7.1 states that, if a Markov chain $X(t)$ has irreducible and aperiodic transition probability matrix P, then the marginal probability distribution of $X(t)$ converges to the stationary probability vector π as t grows large. This means that, if we wait longer and longer, the probability of observing the Markov chain $X(t)$ in a given subset $\mathcal{S} \subseteq \mathcal{X}$ of the state space approaches the aggregate stationary probability of $\mathcal{S}$

$$\pi_{\mathcal{S}} := \sum_{s \in \mathcal{S}} \pi_s \,,$$

in the limit as t grows large.

The following result —known as the *ergodic theorem*— is in some respects more useful in applications. Ergodic theorems capture the idea that, under proper assumptions, *time averages equal space averages*. In particular, the following result guarantees that the empirical frequency of visits of a Markov chain $X(t)$ with irreducible transition probability matrix P (note that aperiodicity is not required in this case) in any given state i converges to the stationary probability π_i of that state.

Theorem 7.1 (Ergodic theorem)**.** Let $X(t)$, $t = 0, 1, \ldots$, be a Markov chain with state space $\mathcal{X}$, whose transition probability matrix P is irreducible. Let $\pi = P'\pi$ be the invariant probability distribution of P. Then, for any arbitrary initial probability distribution,

$$\lim_{t \to +\infty} \frac{1}{t} \sum_{h=0}^{t-1} f(X(h)) = \sum_{i \in \mathcal{X}} \pi_i f(i) \qquad \forall f : \mathcal{X} \to \mathbb{R} \,, \tag{7.7}$$

with probability 1.

The Ergodic Theorem finds very relevant applications in Markov Chain Monte Carlo (MCMC) statistical methods. The central idea is that, in many relevant applications, the state space $\mathcal{X}$ is very large and one is interested in computing weighted averages $\sum_{i \in \mathcal{X}} \pi_i f(i)$ of observable functions $f : \mathcal{X} \to \mathbb{R}$

with respect to a probability vector π that is hard or practically impossible to compute explicitly. However, it could be much simpler to simulate an irreducible Markov chain with stationary distribution π. Then, one can use the empirical time average $\frac{1}{t}\sum_{h=0}^{t-1} f(X(h))$ to approximate the spatial average $\sum_{i\in\mathcal{X}} \pi_i f(i)$ and Theorem 7.1 guarantees that this is an asymptotically exact approximation, with probability one. Of course, the practical use of the MCMC method relies on how fast the convergence in (7.7) is.

7.3 Hitting times and absorbing probabilities

Consider a Markov chain $X(t)$ with transition probability matrix P on a finite state space $\mathcal{X}$. How long does it take for $X(t)$ to move from a state i to another state j? And how long does it take to get back to its starting state i? Of course, answers to these questions depend on the specific walk that $X(t)$ follows when starting from node i. Since this walk is a random one, the time to go from a node to another one or to get back to itself are in general random variables.

Specifically, let the *hitting time* and the *return time* on a given node $j \in \mathcal{X}$ be defined by

$$T_j := \inf\{t \geq 0 :\, X(t) = j\}\,, \qquad \text{and} \qquad T_j^+ := \inf\{t \geq 1 :\, X(t) = j\}\,,$$

respectively, where we use the convention that the infimum of an empty set is $+\infty$. In plain words, the hitting time T_j and the return time T_j^+ are defined as the first time $t \geq 0$ and, respectively, the first time $t \geq 1$, that $X(t)$ hits j. Observe that:

- if $X(0) = i \neq j$, then $T_j = T_j^+$;
- if $X(0) = j$, then $0 = T_j < T_j^+$, hence the terminology 'return time'.

More in general, for a subset of states $\mathcal{S} \subseteq \mathcal{X}$, we can define the hitting time and the return time on $\mathcal{S}$ as

$$T_{\mathcal{S}} := \inf\{t \geq 0 :\, X(t) \in \mathcal{S}\} = \min_{s\in\mathcal{S}}\{T_s\}$$

and

$$T_{\mathcal{S}}^+ := \inf\{t \geq 1 :\, X(t) \in \mathcal{S}\} = \min_{s\in\mathcal{S}}\{T_s^+\}\,,$$

respectively. The following result asserts finiteness (both with probability one and in expectation) of the hitting times on every set of states that is reachable from the rest of the states. Precisely,

Proposition 7.1. Let $X(t)$ be a Markov chain with finite state space $\mathcal{X}$, transition probability matrix P, and initial state $i \in \mathcal{X}$. Let $\mathcal{S} \subseteq \mathcal{X}$ be a subset of states and $T_{\mathcal{S}}$ be the corresponding the hitting time. Then,

(i) for every state i such that $\mathcal{S}$ is not reachable from i

$$\mathbb{P}_i(T_{\mathcal{S}} = +\infty) = 1\,,$$

(ii) if $\mathcal{S}$ is globally reachable, then

$$\mathbb{P}_i(T_{\mathcal{S}} < +\infty) = 1\,, \qquad \mathbb{E}_i[T_{\mathcal{S}}] < +\infty\,,$$

for every $i \in \mathcal{X}$.

Proof. Point (i) is immediate, hence we focus on the proof of (ii). Observe that, if $i \in \mathcal{S}$, then the claim is immediate, as $\mathbb{P}_i(T_{\mathcal{S}} = 0) = 1$ and $\mathbb{E}_i[T_{\mathcal{S}}] = 0$. Hence, let us focus on the case where $\mathcal{S}$ is globally reachable and $i \in \mathcal{R}$, where $\mathcal{R} = \mathcal{V} \setminus \mathcal{S}$. Let $\tilde{X}(t)$ be a Markov chain with the same initial state i as $X(t)$ and transition probabilities

$$\tilde{P}_{ij} = \begin{cases} P_{ij} & \text{if} \quad i \in \mathcal{X} \setminus \mathcal{S} \text{ and } j \in \mathcal{X} \\ 1 & \text{if} \quad i \in \mathcal{S} \text{ and } j = i \\ 0 & \text{if} \quad i \in \mathcal{S} \text{ and } j \in \mathcal{X} \setminus \{i\}\,. \end{cases}$$

Observe that, provided that the two chains start in the same state $X(0) = \tilde{X}(0) = i$, they first hit the set $\mathcal{S}$ at the same time $T_{\mathcal{S}} = \tilde{T}_{\mathcal{S}}$. Now, notice that the transition probability matrix of the chain $\tilde{X}(t)$ can be written as

$$\tilde{P} = \begin{bmatrix} Q & E \\ 0 & I \end{bmatrix}.$$

Since the set $\mathcal{S}$ is reachable from the initial state $\tilde{X}(0) = X(0) = i$, Corollary 6.3 implies that the marginal probability vector $\tilde{\pi}(t)$ of $\tilde{X}(t)$ satisfies

$$\sum_{j \in \mathcal{X} \setminus \mathcal{S}} \tilde{\pi}_j(t) \leq \exp(-\alpha t + 1)\,, \qquad t \geq 0\,,$$

for some positive constant α that depends on Q but not on t. On the other hand, since the set $\mathcal{S}$ is absorbing for the chain $\tilde{X}(t)$, one gets that $\tilde{T}_{\mathcal{S}} > t$ if and only if $\tilde{X}(t) \notin \mathcal{S}$. It follows that the expected hitting time on the set $\mathcal{S}$ can be upper-bounded as

$$\begin{aligned} \mathbb{E}_i[T_{\mathcal{S}}] &= \mathbb{E}_i[\tilde{T}_{\mathcal{S}}] \\ &= \sum_{t \geq 0} \mathbb{P}_i\left(\tilde{T}_{\mathcal{S}} > t\right) \\ &= \sum_{t \geq 0} \mathbb{P}_i\left(\tilde{X}(t) \notin \mathcal{S}\right) \\ &= \sum_{t \geq 0} \sum_{j \in \mathcal{X} \setminus \mathcal{S}} \tilde{\pi}_j(t) \\ &\leq \sum_{t \geq 0} \exp(-\alpha t + 1) \\ &= \frac{\exp(1)}{1 - \exp(-\alpha)}\,, \end{aligned}$$

hence it is finite. Then, necessarily $\mathbb{P}_i(\tilde{T}_{\mathcal{S}} < +\infty) = 1$. The claim now follows from the previous observation that the hitting time on $\mathcal{S}$ for the chain $X(t)$ coincides with the one for the chain $\tilde{X}(t)$, i.e., $T_{\mathcal{S}} = \tilde{T}_{\mathcal{S}}$. □

While the hitting and return time are random variables, their expectations are of course deterministic quantities, that typically depend on the initial state of the Markov chain. In particular, the expected hitting times on a given subset of states that is reachable from any other state admit the following characterization as solutions of a certain linear system of equations.

Proposition 7.2. Let $X(t)$ be a Markov chain with finite state space $\mathcal{X}$, transition probability matrix P, and initial state $i \in \mathcal{X}$. Let $\mathcal{S} \subseteq \mathcal{X}$ be a subset of states that is reachable from every state i in $\mathcal{X}$. Then, the expected hitting times $\{\mathbb{E}_i[T_{\mathcal{S}}]\}_{i\in\mathcal{X}}$ are the unique solution of the following linear system:

$$\begin{aligned} \mathbb{E}_i[T_{\mathcal{S}}] &= 0 \quad \text{if} \quad i \in \mathcal{S} \\ \mathbb{E}_i[T_{\mathcal{S}}] &= 1 + \sum_{j\in\mathcal{X}} P_{ij}\mathbb{E}_j[T_{\mathcal{S}}] \quad \text{if} \quad i \notin \mathcal{S}\,. \end{aligned} \tag{7.8}$$

Proof. Proposition 7.1 implies that the $\mathbb{E}_i[T_{\mathcal{S}}]$'s are all finite. The fact that they satisfy (7.8) can be shown by conditioning on the state visited at time 1:

$$\mathbb{E}_i[T_{\mathcal{S}}] = \sum_{j\in\mathcal{X}} \mathbb{E}_i[T_{\mathcal{S}} \,|\, X(1) = j]\mathbb{P}_i(X(1) = j) = 1 + \sum_{j\in\mathcal{X}} \mathbb{E}_j[T_{\mathcal{S}} \,|\, X(1) = j]P_{ij}$$

On the other hand, if we put $\mathcal{R} = \mathcal{X}\backslash\mathcal{S}$, we know from Chapter 2 that the $\mathcal{R}\times\mathcal{R}$ block Q of the matrix P has spectral radius strictly smaller than 1, so that the linear system $x = Qx + \mathbb{1}$ has a unique solution $x = (I - Q)^{-1}\mathbb{1}$. This implies that (7.8) uniquely characterizes the expected hitting times on a set $\mathcal{S}$. □

The idea in the proof of Proposition 7.2 can be readily applied also to expected return times as a linear function of the expected hitting times, as illustrated in the following result.

Corollary 7.2. Let $X(t)$ be a Markov chain with finite state space $\mathcal{X}$ and transition probability matrix P. Then, for any state $i \in \mathcal{X}$, the expected return times satisfy

$$\mathbb{E}_i[T_i^+] = 1 + \sum_{j\in\mathcal{X}} P_{ij}\mathbb{E}_j[T_i]\,. \tag{7.9}$$

In particular, $\mathbb{E}_i[T_i^+]$ is finite if and only if state i is reachable from every state $j \neq i$ such that $P_{ij} > 0$.

Proof. Equation (7.9) readily follows from conditioning on the state visited by the Markov chain at time 1 after starting in $X(0) = i$. Clearly, $\mathbb{E}_i[T_i] = 0$, so that the summation on the right-hand side of equation (7.9) can always be restricted to those states $j \neq i$ such that $P_{ij} > 0$. Then, (7.9) and Proposition 7.1 imply that $\mathbb{E}_i[T_i^+]$ is finite if and only if state i is reachable from every state j such $P_{ij} > 0$. □

For simple graph structures, Proposition 7.2 can be readily applied to get explicit formulas for the expected hitting times, as illustrated in the following example.

Example 7.1. In the gambler problem discussed in the beginning of Section 7.1, one has $\mathcal{X} = \{0, 1, \ldots, n\}$, $P_{i,i+1} = P_{i,i-1} = 1/2$ for all $1 \leq i < n$. Then, for $\mathcal{S} = \{0, n\}$, equation (7.8) reads

$$\mathbb{E}_0[T_{\mathcal{S}}] = \mathbb{E}_n[T_{\mathcal{S}}] = 0\,, \qquad \mathbb{E}_i[T_{\mathcal{S}}] = 1 + \frac{1}{2}\mathbb{E}_{i-1}[T_{\mathcal{S}}] + \frac{1}{2}\mathbb{E}_{i+1}[T_{\mathcal{S}}]\,, \qquad 1 \leq i < n\,,$$

and the solution can be verified to be

$$\mathbb{E}_k[T_{\mathcal{S}}] = k(n-k)\,, \qquad k = 0, 1, \ldots, n\,.$$

Hence, the expected time it takes the gambler to quit betting by either going bankrupt or leaving with a fortune of n euros depends on the initial capital k and it is the highest when k is equally distant from 0 and n, and the lowest when k is the closest to either 0 or n.

Hitting times and return times are examples of the so-called stopping times. In general, a nonnegative-integer-valued random variable S is said to be a *stopping time* for the Markov chain $X(t)$ if, for every time $t \geq 0$, the indicator of the event $\{S \leq t\}$ is deterministic function of the states visited by the process up to time t, namely it is a function of $\{X(h)\}_{0 \leq h \leq t}$. The symbol $X(S)$ denotes the state of the Markov chain at the random time S. We have the following general result that turns out to be extremely useful in many applications.

Proposition 7.3. Let $X(t)$ be a Markov chain over a finite state space $\mathcal{X}$ with irreducible transition probability matrix P, invariant probability π, and initial state $i \in \mathcal{X}$. Let S be a stopping time for $X(t)$ such that

$$S > 0\,, \qquad X(S) = i\,, \qquad \mathbb{E}_i[S] < +\infty\,.$$

Then,

$$\mathbb{E}_i[\#\{\text{visits to } j \text{ before time } S\}] = \pi_j \mathbb{E}_i[S]\,, \qquad \forall j \in \mathcal{X}\,.$$

Proof. Put $\rho_j = \mathbb{E}_i[\#\{\text{visits to } j \text{ before time } S\}]$. Then,

$$\begin{aligned}
\rho_k &= \mathbb{E}_i[\sum_{t=0}^{S-1} \mathbb{1}_{X(t)=k}] \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_i(X(t)=k, S>t) \\
&= \sum_{t=0}^{+\infty} \sum_{s=t+2}^{+\infty} \mathbb{P}_i(X(t+1)=k, S=s) + \sum_{s=1}^{+\infty} \mathbb{P}_i(X(0)=k, S=s) \\
&= \sum_{t=0}^{+\infty} \sum_{s=t+2}^{+\infty} \mathbb{P}_i(X(t+1)=k, S=s) + \sum_{t=0}^{+\infty} \mathbb{P}_i(X(0)=k, S=t+1) \\
&= \sum_{t=0}^{+\infty} \sum_{s=t+2}^{+\infty} \mathbb{P}_i(X(t+1)=k, S=s) + \sum_{t=0}^{+\infty} \mathbb{P}_i(X(t+1)=k, S=t+1) \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_i(X(t+1)=k, S>t) \\
&= \sum_{t=0}^{+\infty} \sum_{j\in\mathcal{X}} \mathbb{P}_i(X(t)=j, X(t+1)=k, S>t) \\
&= \sum_{t=0}^{+\infty} \sum_{j\in\mathcal{X}} \mathbb{P}_i(X(t)=j, S>t) P_{jk} \\
&= \sum_{j\in\mathcal{X}} \rho_j P_{jk}
\end{aligned}$$

where the fifth equality follows from the fact that $X(S) = i$ and the eighth equality is implied by the Markov property. By assembling the values ρ_j's into a vector ρ, we thus have that $\rho = P'\rho$ is an eigenvector of P' associated to its leading eigenvalue 1. Since P is irreducible, necessarily $\rho = c\pi$ where $c = \mathbb{1}'\rho = \mathbb{E}_i[S]$. This proves the result. □

Corollary 7.3. Let $X(t)$ be a Markov chain over a finite state space $\mathcal{X}$ with irreducible transition probability matrix P, invariant probability π, and initial state $i \in \mathcal{X}$. Then,

$$\mathbb{E}_i[T_i^+] = \frac{1}{\pi_i}\,.$$

Proof. Consider the stopping time $S = T_i^+$. Then,

$$\mathbb{E}_i[\text{number of visits to } i \text{ before time } S] = 1$$

Applying Proposition 7.3, we get the result. □

When the set $\mathcal{S}$ consists of more than one element, we may be interested not only in the expected hitting time on $\mathcal{S}$, but also on where the set $\mathcal{S}$ is first hit. E.g., in the gambler's case we would like to know the probability that she goes bankrupt before accumulating a fortune of n (i.e., the probability that $T_0 < T_n$), or vice versa. This idea is captured by the notion of *absorbing probability*. Given a subset of states $\mathcal{S} \subseteq \mathcal{X}$, for every initial state $i \in \mathcal{X}$ we define the absorbing probability in $s \in \mathcal{S}$ as

$$\Gamma_{i,s} = \mathbb{P}_i(X(T_{\mathcal{S}}) = s) = \mathbb{P}_i(T_{\mathcal{S}} = T_s)\,.$$

In plain words, $\Gamma_{i,s}$ represents the probability that, when started from node i, the Markov chain $X(t)$ hits node $s \in \mathcal{S}$ before any other node $j \in \mathcal{S} \setminus \{s\}$. Similarly to the hitting times, also the absorbing probabilities admit an algebraic characterization which can be proven by the same argument used in the proof of Proposition 7.2.

Proposition 7.4. Let $X(t)$ be a Markov chain over a finite state space $\mathcal{X}$ with transition probability matrix P. Let $S \subseteq \mathcal{X}$ be a subset of states reachable from every node in $\mathcal{X}$. Then, for all $s \in \mathcal{S}$, the absorbing probabilities $\{\Gamma_{i,s}\}_{i\mathcal{X}}$ is the only family of values satisfying the relations:

$$\begin{aligned} &\Gamma_{i,s} = \sum_{j\in\mathcal{X}} P_{ij}\Gamma_{j,s} \quad && \text{if} \quad i \notin \mathcal{S} \\ &\Gamma_{s,s} = 1 && \text{if} \quad i = s \\ &\Gamma_{i,s} = 0 && \text{if} \quad i \in \mathcal{S}\,,\ i \neq s\,. \end{aligned} \tag{7.10}$$

Remark: Confronting (7.10) with the characterization of the asymptotic opinion of an average model with stubborn agents, we can immediately conclude that $\Gamma_{i,s}$ can be interpreted as the asymptotic opinion of agent i when nodes in $\mathcal{S}$ are stubborn with $u_s = 1$ and $u_{s'} = 0$ for every $s' \in S \setminus \{s\}$. On the other hand, notice that $\Gamma_{i,s}$ is the same as the weight coefficient appearing in the expression (6.21).

Example 7.2. In the gambler's ruin problem we find that $\Gamma_{i,n} = 1 - \Gamma_{i,0}$ satisfy

$$\Gamma_{0,n} = 0\,, \qquad \Gamma_{n,n} = 1\,, \qquad \Gamma_{i,n} = \frac{1}{2}\Gamma_{i-1,n} + \frac{1}{2}\Gamma_{i+1,n}\,, \qquad 1 \leq i < n\,,$$

whose solution can be easily checked to be

$$\Gamma_{k,n} = \frac{k}{n}\,, \qquad 0 \leq k \leq n\,.$$

I.e., the probability that the gambler succeeds in accumulating a fortune of n euros before going bankrupt is a linear function of the initial capital k that interpolates between $\Gamma_{0,n} = 0$ and $\Gamma_{n,n} = 1$.

7.4 Time-reversible and birth-and-death chains

In this section we consider *time-reversible* Markov chains, i.e., Markov chains with reversible transition probability matrix P. Recall from a Section 5.1 that a stochastic matrix P is said to be *reversible* with respect to a probability distribution π if it satifies the detailed balance equation

$$\pi_i P_{ij} = \pi_j P_{ji} \qquad \forall i, j \in \mathcal{X}\,. \tag{7.11}$$

Also, recall that summing up both sides of (7.11) gives

$$\pi_i = \sum_j \pi_i P_{ij} = \sum_j \pi_j P_{ji} \qquad \forall i \in \mathcal{X}\,,$$

so that, if P is reversible with respect to a probability distribution π, then necessarily $\pi = P'\pi$ is a stationary probability distribution.

Consider a Markov chain with transition probability matrix P reversible with respect to a stationary distribution π, and assume that the initial state $X(0)$ has distribution π. Then, we have that

$$\mathbb{P}(X(0) = i, X(1) = j) = \pi_i P_{ij} = \pi_j P_{ji} = \mathbb{P}(X(1) = i, X(0) = j)\,, \qquad i, j \in \mathcal{X}\,,$$

i.e., the pair $(X(0), X(1))$ has the same joint probability distribution as the reversed pair $(X(1), X(0))$. The argument above can be generalized to show that

$$(X(0), X(1), \ldots, X(t-1), X(t)) \stackrel{\text{dist}}{=} (X(t), X(t-1), \ldots, X(1), X(0))$$

for all $t \geq 0$, where $\stackrel{\text{dist}}{=}$ stands for equality in distribution. I.e., reversing the arrow of time does not change the probability distribution of a stationary reversible Markov chain. This property is known as *time-reversibility* of the chain $X(t)$. Discrete-time simple random walks over undirected graphs are time-reversible Markov chains (see Example 5.1). Another fundamental example is provided by birth-and-death chains discussed in the following. More examples of reversible Markov chains will be discussed in Chapter **??**

A *birth-and-death* chain is a Markov chain with state space $\mathcal{X} = \{0, 1, \ldots, n\}$ and a transition probability matrix

$$P_{ij} = \begin{cases} p_i & \text{if} \quad 0 \leq i < n\,,\ j = i+1 \\ q_i & \text{if} \quad 1 \leq i \leq n\,,\ j = i-1 \\ r_i & \text{if} \quad 0 \leq i \leq n\,,\ j = i \\ 0 & \text{if} \quad 0 \leq i, j \leq n\,,\ |j-i| \geq 2\,, \end{cases}$$

where $p, q, r \in \mathbb{R}_+^{\mathcal{X}}$ are nonnegative vectors such that

$$p + q + r = \mathbb{1}\,, \qquad p_n = q_0 = 0\,.$$

In other terms, at every time step $t \geq 0$, the state $X(t) = i$ of a birth-and-death chain can only increase by 1 (with conditional probability p_i), decrease by 1 (with conditional probability q_i), or remain the same (with conditional probability $r_i = 1 - p_i - q_i$). (See Figure 7.3.) Throughout, we shall assume that

$$p_i > 0\,, \quad 0 \leq i < n\,, \qquad q_i > 0\,, \quad 1 \leq j \leq n\,,$$

so that the birth-and-death chain is irreducible. What makes birth-and-death chains particularly useful is that for them many quantities of interest such as the stationary probability distribution π, the expected hitting times, and absorbing probabilities can be computed explicitly.

Every birth-and-death chain is reversible with respect to its stationary probability distribution π. Indeed, we can prove reversibility and explicitly compute the stationary probability distribution π all at once, by solving the linear system

$$\pi_k p_k = \pi_{k+1} q_{k+1}\,, \qquad k = 0, 1, \ldots, n-1\,, \qquad \sum_{0 \leq k \leq n} \pi_k = 1\,, \tag{7.12}$$

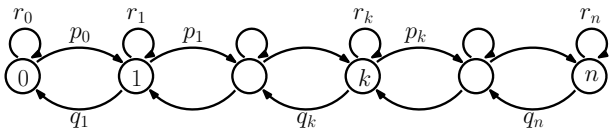


Figure 7.3: A birth-and-death chain. The state space is $\{0, 1, \ldots, n\}$. From every state $1 \leq k < n$, the chain increases by 1 with probability p_k, decreases by 1 with probability q_k, and stays put with probability r_k. At the boundary point $k = 0$ ($k = n$), the chain can only stay put or increase (respectively, decrease). Observe that the gambler's capital chain in Figure 7.1 is a special case of birth-and-death chain.

that entails the detailed balance equations and normalization. Rearranging terms in the leftmost equation in (7.12) gives

$$\pi_{k+1} = \frac{p_k}{q_{k+1}} \pi_k \,, \qquad k = 0, 1, \ldots, n-1 \,,$$

whose solution is

$$\pi_k = \pi_0 \prod_{j=1}^{k} \frac{p_{j-1}}{q_j} \,, \qquad k = 0, 1, \ldots, n \,,$$

with the usual convention that an empty product is equal to 1. In order to find π_0 we normalize as per the rightmost equation in (7.12), thus finding

$$\pi_k = \frac{\displaystyle\prod_{j=1}^{k} \frac{p_{j-1}}{q_j}}{\displaystyle\sum_{h=0}^{n} \prod_{j=1}^{h} \frac{p_{j-1}}{q_j}} \,, \qquad 0 \leq k \leq n \,. \tag{7.13}$$

In the special case when all the birth probabilities $p_k = p$ are the same and so are the death probabilities $q_k = q$, equation (7.13) for the invariant probability vector of the birth-and-death chain reduces to

$$\pi_k = \frac{\rho^k}{1 + \rho + \rho^2 + \ldots \rho^n} \,, \qquad \rho = \frac{p}{q} \,, \qquad k = 0, 1, \ldots, n \,.$$

When $\rho = 1$, i.e., when the birth and death probabilities $p = q$ coincide, then the above is the uniform distribution on the set $\mathcal{X} = \{0, 1, \ldots, n\}$. On the other hand, for $\rho < 1$, i.e., when the death probability q is larger than the birth one p, then the stationary probability distribution is the highest at node 0 and is exponentially decreasing with the distance from it. Symmetrically, for $\rho = \frac{p}{q} > 1$, the stationary probability distribution is the highest at node n and is exponentially decreasing with the distance from it.

Absorbing probabilities and expected hitting times on the set $\mathcal{S} = \{0, n\}$ can also be determined, albeit with slightly more involved arguments. Let us start with the absorbing probabilities $\Gamma_{i,n} = 1 - \Gamma_{i,0}$. By applying (7.10) to the birth-and-death chain, we get

$$\Gamma_{0,n} = 0\,, \qquad \Gamma_{1,n} = 1\,, \qquad \Gamma_{k,n} = q_k\Gamma_{k-1,n} + r_k\Gamma_{k,n} + p_k\Gamma_{k+1,n},\,, \quad 1 \le k < n\,.$$

By rearranging terms in the above, recalling that $q_k + p_k + r_k = 1$, and defining $\alpha_k = \Gamma_{k,n} - \Gamma_{k-1,n}$, we get

$$\alpha_{k+1} = \frac{q_k}{p_k}\alpha_k\,, \qquad k = 1, \dots, n-1\,.$$

The recursion above has solution

$$\alpha_k = \alpha_1 \prod_{1\le j<k} \frac{q_j}{p_j}\,, \quad k = 1, \dots, n-1\,, \qquad \alpha_1 = \left(\sum_{1\le k\le n} \prod_{1\le j<k} \frac{q_j}{p_j} \right)^{-1}\,,$$

where the value of α_1 is determined by imposing the constraint

$$\sum_{1\le k\le n} \alpha_k = \Gamma_{n,n} - \Gamma_{0,n} = 1\,.$$

Then, we have

$$\Gamma_{k,n} = 1 - \Gamma_{k,0} = \sum_{1\le j\le k} \alpha_k = \frac{\displaystyle\sum_{1\le h\le k} \prod_{1\le j<h} \frac{q_j}{p_j}}{\displaystyle\sum_{1\le h\le n} \prod_{1\le j<h} \frac{q_j}{p_j}}\,, \qquad k = 0, 1, \dots, n\,. \tag{7.14}$$

In the special case when $p_k/q_k = \rho$ does not depend on $k = 1, 2, \dots, n-1$, equation (7.14) reduces to

$$\Gamma_{k,n} = 1 - \Gamma_{k,0} = \sum_{1\le j\le k} \alpha_k = \frac{\displaystyle\sum_{1\le h\le k} \rho^{1-h}}{\displaystyle\sum_{1\le h\le n} \rho^{1-h}}\,, \qquad k = 0, 1, \dots, n\,. \tag{7.15}$$

When $\rho \neq 1$, the above reduces to

$$\Gamma_{k,n} = 1 - \Gamma_{k,0} = \frac{1 - 1/\rho^k}{1 - 1/\rho^n}\,. \tag{7.16}$$

On the other hand, when $\rho = 1$, we get

$$\Gamma_{k,n} = 1 - \Gamma_{k,0} = \frac{k}{n}\,, \qquad k = 0, 1, \dots, n\,,$$

that is consistent with what found for the gambler's ruin problem in Example 7.2.

We finally compute the expected hitting times $\tau_k = \mathbb{E}_k[T_{\mathcal{S}}]$. By applying (7.8) to the birth-and-death chain, we get

$$\tau_0 = \tau_n = 0\,, \qquad \tau_k = 1 + q_k\tau_{k-1} + r_k\tau_k + p_k\tau_{k+1}\,, \quad 1 \le k \le n-1\,.$$

By rearranging terms in the above, recalling that $q_k + p_k + r_k = 1$, and defining $\beta_k = \tau_k - \tau_{k-1}$, we get

$$\beta_{k+1} = \frac{q_k}{p_k}\beta_k - \frac{1}{p_k}\,, \qquad k = 1, 2, \ldots, n-1\,,$$

The recursion above has solution

$$\beta_k = \beta_1 \prod_{j=1}^{k-1} \frac{q_j}{p_j} - \sum_{h=1}^{k-1} \frac{1}{p_h} \prod_{j=1}^{h-1} \frac{q_j}{p_j}\,, \qquad k = 1, 2, \ldots, n\,, \tag{7.17}$$

Observe that, for $1 \le k \le n$,

$$\sum_{i=1}^{k} \beta_i = \sum_{i=1}^{k} (\tau_i - \tau_{i-1}) = \tau_k - \tau_0 = \tau_k\,.$$

Hence, in particular, $\beta_1 = \tau_1$ and substituting (7.17) in the above gives

$$\tau_k = \sum_{i=1}^{k} \beta_i = \tau_1 \sum_{i=1}^{k} \prod_{j=1}^{i-1} \frac{q_j}{p_j} - \sum_{i=1}^{k} \sum_{h=1}^{i-1} \frac{1}{p_h} \prod_{j=1}^{h-1} \frac{q_j}{p_j}\,, \qquad k = 1, 2, \ldots, n\,, \tag{7.18}$$

To determine τ_1, we notice that $\sum_{k=1}^{n} \beta_k = \tau_n - \tau_0 = 0$, so that

$$\tau_1 = \left(\sum_{k=1}^{n} \prod_{j=1}^{k-1} \frac{q_j}{p_j} \right)^{-1} \sum_{k=1}^{n} \sum_{h=1}^{k-1} \frac{1}{p_h} \prod_{j=1}^{h-1} \frac{q_j}{p_j}\,. \tag{7.19}$$

In the special case when $p_k = q_k = \gamma$ for $1 \le k < n$, equation (7.19) reduces to

$$\tau_1 = \frac{1}{\gamma n} \sum_{k=1}^{n} (k-1) = \frac{n-1}{2\gamma}\,,$$

so that (7.18) gives

$$\tau_k = k\tau_1 - \frac{1}{\gamma} \sum_{i=1}^{k} (i-1) = \frac{k(n-k)}{2\gamma}\,, \qquad 0 \le k \le n\,,$$

which is consistent with what found in Example 7.1 for the gambler's ruin problem. On the other hand, when $p_k = \gamma$ and $q_k = \gamma/\rho$ for some $\rho > 0$ and

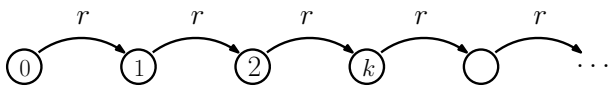


Figure 7.4: In a rate-r Poisson arrival process, the time between any two consecutive particle arrivals is an independent rate-r exponential random variable.

$1 \leq k < n$, equation (7.19) reduces to

$$\tau_1 = \frac{\sum_{k=1}^{n}\sum_{h=1}^{k-1}\rho^{1-h}}{\gamma\sum_{h=1}^{n}\rho^{1-h}} = \frac{\sum_{k=1}^{n}(1-\rho^{1-k})}{\gamma(1-\rho^{-n})} = \frac{n}{\gamma(1-\rho^{-n})} + \frac{\rho}{\gamma(1-\rho)}\,,$$

so that, as n grows large, $\tau_1 \asymp n/\gamma$ if $\rho > 1$, whereas $\tau_1 \to \rho/(\gamma(1-\rho))$ if $\rho < 1$.

7.5 Continuous-time Markov chains

In many applications, it proves convenient to consider continuous-time Markov chains instead of the discrete-time chains we have discussed so far. In order to define continuous-time Markov chains we need to first introduce *Poisson processes* defined as follows: for a given $r > 0$, let $S_1, S_2, \ldots$ be a sequence of independent random variables with rate-r exponential distribution

$$\mathbb{P}(S_i \geq t) = e^{-rt}\,, \qquad t \geq 0\,,\ i = 1, 2, \ldots\,,$$

each with expected value

$$\mathbb{E}[S_i] = \int_0^{+\infty} \mathbb{P}(S_i \geq t)\mathrm{d}t = \int_0^{+\infty} e^{-rt}\mathrm{d}t = 1/r\,.$$

The key property of the exponential distribution is that it is memoryless:

$$\mathbb{P}(S_i \geq t+s | S_i \geq t) = \frac{\mathbb{P}(S_i \geq t+s)}{\mathbb{P}(S_i \geq t)} = \frac{e^{-r(t+s)}}{e^{-rt}} = e^{-rs} = \mathbb{P}(S_i \geq s)\,.$$

I.e., it does not matter how long we have waited already, given that the event has not occurred by time t, the time we still have to wait for it to occur has conditional distribution that is a rate-r exponential one. Now, consider the sequence

$$T_0 = 0\,, \qquad T_k = \sum_{1 \leq j \leq k} S_j\,, \qquad k = 1, 2, \ldots\,, \tag{7.20}$$

which we will refer to as a rate-r *Poisson clock*. In other words, a Poisson clock is characterized by the property that the time elapsed between any two of

its consecutive ticks is an independent random variable with rate-r exponential distribution. We can then define the Poisson process

$$N_t := \sup\{k \geq 0 : T_k \leq t\}\,, \qquad t \geq 0\,, \tag{7.21}$$

i.e., N_t stands for the number of ticks of the Poisson clock occurred by time t. Observe that we can think of the Poisson process $N_t = X(t)$ as a random walk on the infinite set of nonnegative integers: that starts in 0 at time $t = 0$, increases by one unit at every tick of the Poisson clock, and never decreases.

More in general, given a finite state space $\mathcal{X}$ and a square nonnegative matrix $\Lambda = (\Lambda_{ij})_{i,j\in\mathcal{X}} \neq 0$ with entries indexed by elements of $\mathcal{X}$ and diagonal entries $\Lambda_{ii} = 0$, we define a continuous-time Markov chain $X(t)$ with transition rate matrix Λ in the following way. Let

$$\omega_i = \sum_{j\neq i} \Lambda_{ij}\,, \quad i \in \mathcal{X}\,, \qquad \omega_* = \max_{i\in\mathcal{X}}\{\omega_i\}\,,$$

and consider a rate-ω_* Poisson clock $T_0 \leq T_1 \leq \ldots$, the associated Poisson process N_t defined as in (7.21), and an independent discrete-time Markov chain $U(k)$, for $k = 0, 1, \ldots$, with transition probabilities

$$\overline{P}_{ij} = \frac{\Lambda_{ij}}{\omega_*}\,, \qquad i \neq j\,, \qquad \overline{P}_{ii} = 1 - \sum_{j\neq i} \overline{P}_{ij}\,. \tag{7.22}$$

Then, define

$$X(t) = U(N_t)\,, \qquad t \geq 0\,. \tag{7.23}$$

The discrete-time Markov chain $U(k)$ is called the *jump chain* associated to the continuous-time Markov chain $X(t)$, as it describes the sequence of moves of $X(t)$, while the Poisson clock $T_0 \leq T_1 \leq \ldots$ describes the times of these moves.

In fact, we can define a continuous-time Markov chain in at least two other equivalent ways that might be more convenient in some cases. The first one is: given that $X(t) = i$, the chain $X(t)$ waits a rate-ω_i exponential time in node i before jumping to a another state j chosen with conditional probabilities

$$P_{ij} = \frac{\Lambda_{ij}}{\omega_i}\,, \qquad i, j \in \mathcal{X}\,.$$

Once in j, it waits a rate-ω_j exponential time before jumping to a new state k chosen with probability P_{jk}, and so on. The other equivalent definition consists in equipping every link $(i, j) \in \mathcal{E}$ by an independent rate-Λ_{ij} Poisson clock: when the (i, j)-th clock ticks, a 'gate' on link (i, j) is open and, if the chain is currently in state i, it jumps to state j, otherwise it stays put.

It can be proven that the probability distribution $\overline{\pi}(t)$ of the continuous-time Markov chain $X(t)$ with transition rate matrix Λ, i.e.,

$$\overline{\pi}_i(t) = \mathbb{P}(X(t) = i)\,, \qquad i \in \mathcal{X}\,,$$

satisfies the linear system of differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{\pi}(t) = -L'\overline{\pi}(t)\,, \tag{7.24}$$

where $L = \mathrm{diag}\,(\omega) - \Lambda$. Analogously to the discrete-time case, a stationary probability vector $\overline{\pi}$ is a vector satisfying the equations

$$L'\overline{\pi} = 0\,, \qquad \mathbb{1}'\overline{\pi} = 1\,.$$

The left-most equation above states that $\overline{\pi}$ is invariant for the dynamics (7.24), the second one states that its entries sum up to 1, consistently with it being a probability vector. We will refer to $\overline{\pi}$ as the stationary probability vector of the continuous-time Markov chain with transition rate matrix Λ: if the initial probability distribution $\pi(0) = \overline{\pi}$ is an invariant probability vector, then, as a solution of (7.24), it is not going to change in time, i.e., $\pi(t) = \overline{\pi}$ for all $t \geq 0$. Observe that such stationary probability distribution $\overline{\pi}$ is an invariant probability vector for the jump chain with transition probability matrix $\overline{P}$ defined in (7.22), i.e., it holds true that $\overline{\pi} = \overline{P}'\overline{\pi}$. Indeed,

$$\overline{P}'\overline{\pi} = \frac{1}{\omega^*}\left(\Lambda' + \mathrm{diag}\,(\omega_*\mathbb{1}) - \mathrm{diag}\,(\omega)\right)\overline{\pi} = (-\frac{1}{\omega_*}L' + I)\overline{\pi} = \overline{\pi}$$

In contrast, notice that $\overline{\pi}$ does not, in general, coincide with the stationary probability distribution $\pi = P'\pi$ of the discrete-time Markov chain with transition probability P. In fact, it can easily be verified that π and $\overline{\pi}$ are related by the identity

$$\overline{\pi}_i = \frac{\pi_i/\omega_i}{\sum_j \pi_j/\omega_j}\,, \qquad i \in \mathcal{X}\,.$$

Therefore, they coincide if and only if $\omega_i = \omega_*$ for all $i \in \mathcal{X}$.

A continuous-time Markov chain with transition rates $(\Lambda_{ij})_{i,j\in\mathcal{X}}$ is said to be reversible with respect to a probability distribution $\overline{\pi}$ if

$$\overline{\pi}_i\Lambda_{ij} = \overline{\pi}_j\Lambda_{ji}\,, \qquad i,j \in \mathcal{X}\,.$$

This can be easily verified to be equivalent to that the associated jump chain be reversible as a discrete-time Markov chain with transition probability matrix $\overline{P}$ defined as above. An analogous theory to the one for discrete-time Markov chains applies to convergence of irreducible continuous-time Markov chains, as summarized in the following result.

Theorem 7.2. Let $X(t)$ be a continuous-time Markov chain with finite state space $\mathcal{X}$ and transition rates Λ_{ij} for $i \neq j \in \mathcal{X}$. Let Λ be the nonnegative square matrix with out-diagonal entries equal to the transition rates and diagonal entries $\Lambda_{ii} = 0$. Let $\mathcal{G}_\Lambda = (\mathcal{X}, \mathcal{E}, \Lambda)$ be the graph with node set $\mathcal{X}$, and weight matrix Λ, and let $\omega = \Lambda\mathbb{1}$, $P = \mathrm{diag}\,(\omega)^{-1}\Lambda$, and $L = \mathrm{diag}\,(\omega) - \Lambda$ be its degree vector, normalized weight matrix, and Laplacian matrix, respectively. If $\mathcal{G}_\Lambda$ is strongly connected, then

(i) there exists a unique Laplace-invariant probability distribution $\overline{\pi}$ (i.e., a nonnegative vector $\overline{\pi}$ such that $L'\overline{\pi} = 0$, $\mathbb{1}'\overline{\pi} = 1$);

(ii) for every initial probability distribution $\mathbb{P}(X(0) = i) = \overline{\pi}_i(0)$, for $i \in \mathcal{X}$, the time-t marginal probability distribution $\overline{\pi}_i(t) = \mathbb{P}(X(t) = i)$ satisfies

$$\lim_{t \to +\infty} \overline{\pi}(t) = \overline{\pi}\,;$$

(iii) for every given observable function $f : \mathcal{X} \to \mathbb{R}$, one has that

$$\lim_{t \to +\infty} \frac{1}{t} \int_0^t f(X(s))\mathrm{d}s = \sum_{i \in \mathcal{X}} \overline{\pi}_i f(i)\,,$$

with probability 1;

(iv) the expected return times satisfy

$$\mathbb{E}_i[\overline{T}_i^+] = \frac{1}{\omega_i \overline{\pi}_i}\,, \qquad i \in \mathcal{X}\,,$$

where $\overline{T}_i^+ = \inf\{t \geq 0 :\ X(t) = i \text{ and } X(s) \neq i \text{ for some } s \in (0,t)\}$.

On the other hand, if $\mathcal{S} \subseteq \mathcal{X}$ is a subset of states that is globally reachable in $\mathcal{G}_\Lambda$, then

(v) the expected hitting times

$$\overline{\tau}_i^{\mathcal{S}} = \mathbb{E}_i[T_{\mathcal{S}}]\,, \qquad i \in \mathcal{X}\,,$$

are the unique solutions of

$$\overline{\tau}_s^{\mathcal{S}} = 0\,, \qquad s \in \mathcal{S}\,, \qquad\qquad \overline{\tau}_i^{\mathcal{S}} = \frac{1}{\omega_i} + \sum_{j \in \mathcal{X}} P_{ij} \overline{\tau}_j^{\mathcal{S}}\,, \qquad i \in \mathcal{X} \setminus \mathcal{S}\,;$$

(vi) the absorbing probabilities

$$\Gamma_{i,j} = \mathbb{P}_i(T_{\mathcal{S}} = T_j)\,, \qquad i \in \mathcal{X}\,, j \in \mathcal{S}\,,$$

are the unique solutions of (7.10).

We now study an important example.

Example 7.3. Consider a continuous-time Markov chain $X(t)$ with state space $\mathcal{X} = \{0, 1, \ldots, n-1, n\}$ that increases at rate $\lambda_i > 0$ if in state $i = 0, 1, \ldots n-1$, and decreases at rate $\mu_i > 0$ if in state $i = 1, 2, \ldots, n$. We can think of it as a continuous-time random walk on the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ where the link set is $\mathcal{E} = \{(0,1), (1,2), \ldots, (n-1,n), (n,n-1), \ldots, (2,1), (1,0)\}$ and the link weights are $W_{i,i+1} = \lambda_i$ for $i = 0, 1, \ldots n-1$ and $W_{i,i-1} = \mu_i$ for $i = 1, 2, \ldots n$. By

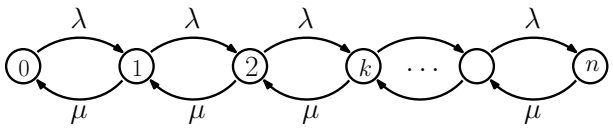


Figure 7.5: A continuous-time birth-and-death process with rate-λ arrival rates and rate-μ departure rates that are both state-independent. The stationary probability distribution is given in (7.26), while the expected occupancy and sojourn time in stationarity are given by (7.27) and (7.28), respectively.

imposing the detailed balance $\overline{\pi}_i \lambda_i = \overline{\pi}_{i+1} \mu_{j+1}$ for $i = 0, 1, \ldots, n-1$, we get $\overline{\pi}_i = \overline{\pi}_0 \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}$. Upon normalizing, we get that the vector $\overline{\pi}$ with entries

$$\overline{\pi}_i = \frac{1}{\zeta} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}\,, \qquad \zeta = \sum_{k=0}^{n} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}\,, \qquad i = 0, 1, \ldots, n\,, \tag{7.25}$$

is Laplace invariant and the associated Markov chain is reversible with respect to it. This is the continuous-time analogue of the discrete-time birth-and-death chain introduced in Section 7.4. We will interpret increments of the state as exogenous arrivals and decrements as departures, so that state represents the number of particles currently in the system.

Two special cases are worth being considered in detail. The first one is when the arrival rate $\lambda_i = \lambda$ is independent from the state and so is the departure rate $\mu_i = \mu$. (See Figure 7.5) This can be interpreted as a queue where particles arrive at constant rate λ, provided that the queue is not full (corresponding to state n) and leave one at a time at rate μ. In this case, the invariant probability distribution reads

$$\overline{\pi}_i = \frac{\rho^i}{1 + \rho + \ldots + \rho^n}\,, \qquad i = 0, 1, \ldots, n\,, \tag{7.26}$$

where

$$\rho = \frac{\lambda}{\mu}\,.$$

Hence, in stationarity, the probability that a newly arriving particle is not admitted is given by

$$\overline{\pi}_n = \frac{\rho^n}{1 + \rho + \ldots + \rho^n}\,.$$

while the expected number of particles in the queue is

$$\mathbb{E}[X] = \sum_{i=0}^{n} i \cdot \overline{\pi}_i = \frac{\sum_{i=0}^{n} i\rho^i}{1 + \rho + \ldots + \rho^n}\,. \tag{7.27}$$

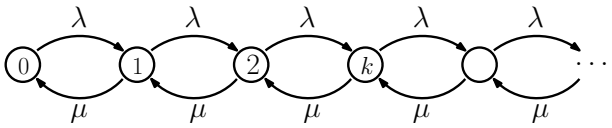


Figure 7.6: An MM1 queue with rate-λ Poisson arrivals and rate-μ Poisson departures. The chain is stable if and only if $\rho = \lambda/\mu < 1$ and in that case the stationary distribution is $\overline{\pi}_i = (1-\rho)\rho^i$ for $i \geq 0$ and, in stationarity, the expected occupancy is $\mathbb{E}[V] = \rho/(1-\rho)$ and the expected sojourn time is $\mathbb{E}[T] = (\mu(1-\rho))^{-1} = (\mu-\lambda)^{-1}$.

If $\lambda < \mu$, so that $\rho < 1$, then π_n is exponentially decreasing to 0 as the order n grows large, while

$$\mathbb{E}[X] = \frac{\rho}{1-\rho} \cdot \frac{1-(n+1)\rho^n + n\rho^{n+1}}{1-\rho^{n+1}} \overset{n\to\infty}{\longrightarrow} \frac{\rho}{1-\rho}\,.$$

In contrast, if $\lambda > \mu$, so that $\rho > 1$, such probability π_n approaches 1 and $\mathbb{E}[X]$ blows up as n grows large. Finally, for $\lambda = \mu$, so that $\rho = 1$, $\pi_n = 1/(n+1)$ and $\mathbb{E}[X] = n/2$. On the other hand, observe that, if we assume a first-in first-out (FIFO) rule, so that the particle to leave is the one that had entered the earliest (and hence stayed the longest) then the permanence time of a newly arriving particle in the queue depends on the number of particles inside the queue at time of arrival (i.e., the state of the Markov chain) and is exponentially distributed with rate $\mu/(i+1)$ (hence expected value $(i+1)/\mu$) conditioned on the state at time of arrival being i. Then, in stationarity, the expected sojourn time of an arriving particle in the queue is

$$\mathbb{E}[T] = \frac{1+\mathbb{E}[X]}{\mu}\,. \tag{7.28}$$

For $\rho < 1$, such expected sojourn time satisfies

$$\mathbb{E}[T] = \frac{1}{\mu(1-\rho)} \cdot \frac{1-(n+1)\rho^n + n\rho^{n+1}}{1-\rho^{n+1}} \overset{n\to\infty}{\longrightarrow} \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}\,.$$

In fact, if we let $n = +\infty$, the Markov chain studied here is known as the $M/M/1$ queue, representing the queue length in a system having a single server, where arrivals are determined by a rate-λ Poisson process and job service times have a rate-μ exponential distribution (see Figure 7.6). While for $\rho \geq 1$, such $M/M/1$ queue is unstable, for $\rho < 1$, i.e., $\lambda < \mu$, it is stable and admits a geometric stationary probability distribution

$$\overline{\pi}_i = (1-\rho)\rho^i\,, \qquad i \geq 0\,,$$

with expected value and variance given by

$$\mathbb{E}[X] = \frac{\rho}{1-\rho}\,, \qquad \mathrm{Var}[X] = \frac{\rho}{(1-\rho)^2}\,,$$

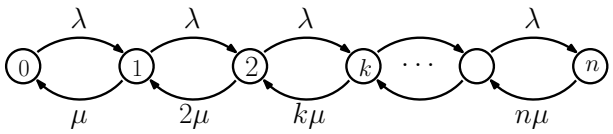


Figure 7.7: A continuous-time birth-and-death process with rate-λ state-independent arrival rates and state-dependent proportional departure rates $i\mu$. The probability to find the system full in stationarity is given by Erlang's formula (7.29).

respectively. The sojourn time in stationarity with a FIFO rule, in this case, turns out to be exponentially distributed with rate $\mu - \lambda$ and hence expected value

$$\mathbb{E}[T] = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}\,.$$

Remarkably, the formula above for the expected sojourn time remains valid independently of the chosen service rule (i.e., not only FIFO), a result that can be recognized as a special case of Little's law.

A second case is when the arrival rate $\lambda_i = \lambda$ is independent from the state, while the departure rate $\mu_i = i\mu$ is proportional to the state i (see Figure 7.7). This can be interpreted as the particles arriving at constant rate (with exception of when the queue is in the full state n and arrivals cannot be accommodated), and waiting an independent rate-μ exponential time before leaving. In this case, the stationary probability distribution has entries

$$\overline{\pi}_i = \frac{\rho^i/i!}{\sum_{j=0}^{n} \rho^j/j!}\,, \qquad i = 0, 1, \ldots, n\,.$$

In stationarity, the probability of finding the queue full is then

$$\overline{\pi}_n = \frac{\rho^n/n!}{\sum_{j=0}^{n} \rho^j/j!}\,, \tag{7.29}$$

which is know as Erlang's formula. As n grows large, even if $\rho \geq 1$, the factorial term $n!$ dominates the exponential one ρ^n, so that $\pi_n \overset{n\to\infty}{\longrightarrow} 0$. Observe that in this case, an arriving particle that does not find the system full has to wait a rate-μ exponential time independently from the current occupancy of the system, in contrast to the case where the departure rate was independent from the state.

Chapter 8

Epidemics and pairwise interacting network systems

8.1 Introduction and main examples

In this chapter, we study *pairwise interacting network systems*. In such systems, a finite population of agents, identified with nodes of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, update their state as a result of the superposition of independent mutations and pairwise interactions with their neighbors. We shall focus on the case when the state of each node belongs to a finite space $\mathcal{A}$ and the dynamical process is Markovian. Specifically, let us refer to $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$ as the *configuration space*. Then, pairwise interacting network systems are continuous-time Markov chains $X(t)$ with state space $\mathcal{X}$ and a particular structure as illustrated in the following.

The entries $X_i(t)$ of the *system configuration* $X(t)$ represent the states of the single nodes. Independent state mutations are described in terms of *mutation kernels*, i.e., stochastic matrices

$$\psi^{(i)} \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}, \qquad i \in \mathcal{V}, \tag{8.1}$$

whose entries $\psi^{(i)}_{a,b}$ represent the conditional probability that, when a node i gets activated, it spontaneously mutates its state into b given that its current state is a. On the other hand, pairwise interactions are described by *pairwise interaction kernels*, i.e., stochastic matrices

$$\varphi^{(i,j)}(c) \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}, \qquad (i,j) \in \mathcal{E}, \qquad c \in \mathcal{A}, \tag{8.2}$$

whose entries $\varphi^{(i,j)}(c)$ represent the conditional probability that, when a link (i,j) gets activated, node i changes its state into b given that it is currently in state a and node j is in state j.

We shall assume that spontaneous node mutations occur at the ticking of independent rate-1 Poisson clocks: if node i's clock ticks at some time t, its current state mutates from $X_i(t^-) = a$ into $X_i(t) = b$ with conditional probability

$\psi_{ab}^{(i)}$. On the other hand, pairwise interactions associated to links (i,j) occur at the ticking of independent Poisson clocks of rate βW_{ij}, where $\beta > 0$ is the *interaction frequency parameter*: if link (i,j)'s clock ticks at some time $t \geq 0$, and node j's current state is $X_j(t) = c$, node i's state changes from $X_i(t^-) = a$ to $X_i(t) = b$ with conditional probability $\varphi_{a,b}^{(i,j)}(c)$.

In summary, pairwise interacting network systems are continuous-time Markov chains on the configuration space $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$, with transition rates

$$\Lambda_{x,y} = \begin{cases} \psi_{x_i,y_i}^{(i)} + \beta \sum_j W_{ij} \varphi_{x_i y_i}^{(i,j)}(x_j) & \text{if} \quad x_i \neq y_i \text{ and } x_{-i} = y_{-i} \\ 0 & \text{if} \quad x \text{ and } y \text{ differ in more than 1 entry}\,, \end{cases} \tag{8.3}$$

for every x and y in $\mathcal{X}$ with $x \neq y$. Observe that the system is fully described by the specification of: (i) the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$; (ii) the mutation kernels (8.1); (iii) the pairwise interaction kernels (8.2); (iv) the interaction frequency parameter β; and (v) an initial distribution over the configuration space $\mathcal{X}$.

In many applications, the mutation and the interaction kernels are the same for all units. In this case we use the notation ψ, $\varphi(\cdot)$, dropping the superscripts that refer to the specific nodes or links. We refer to these pairwise interacting network systems as *homogeneous*. The simplest case is that of the homogeneous pairwise interacting network systems with binary state space $\mathcal{A} = \{0,1\}$ that are described by the graph, the interaction frequency parameter β, three 2×2 stochastic matrices ψ, $\varphi(0)$, and $\varphi(1)$, and an initial configuration distribution.

We now present a number of classical homogeneous pairwise interacting network system models taken from epidemiology and opinion dynamics. Models with heterogeneous units will be considered later on. The weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is given once and for all, as well the interaction frequency parameter β. We recall that all nodes get activated independently at rate 1 while links (i,j) get activated indepedently at rate βW_{ij}.

Example 8.1. [SI epidemics] In the SI (Susceptible-Infected) epidemics, nodes are in one of two possible states: state 0, to be referred to as *susceptible*, or state 1, to referred to as *infected*. Hence, $\mathcal{A} = \{0,1\}$. There are no spontaneous mutations. On the other hand, infected nodes j infect susceptible in-neighbors i independently at rate βW_{ij}. Hence, the mutation kernels are trivially $\psi^{(i)} = I$ for every node i, whereas the pairwise interaction kernels are

$$\varphi^{(i,j)}(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \varphi^{(i,j)}(1) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \qquad (i,j) \in \mathcal{E}\,. \tag{8.4}$$

Note that, in the SI epidemics, infected nodes can never become susceptible.

Example 8.2. [SIS epidemics] In the SIS (Susceptible-Infected-Susceptible) epidemics, nodes have the same binary state space $\mathcal{A} = \{0,1\}$ as in the SI epidemics, with state 0 referred to as *susceptible*, and state 1 as *infected*. Susceptible nodes can become infected via contacts with infected neighbor nodes in the same way as in the SI epidemics, i.e., infected nodes j infect susceptible in-neighbors i independently at rate βW_{ij}. On the other hand, infected nodes

spontaneously recover and return susceptible independently at a constant rate 1. Therefore, in the SIS epidemics, the pairwise interaction kernels are as in (8.4), while the mutation kernels are

$$\psi^{(i)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \qquad i \in \mathcal{V}. \tag{8.5}$$

Example 8.3. [SIR epidemics] In the SIR (Susceptible-Infected-Recovered) epidemics, nodes can be in one of three possible states: susceptible (0), infected (1), or *recovered* (2), so that $\mathcal{A} = \{0, 1, 2\}$. Exactly as in the SI and in the SIS epidemics, also in the SIR epidemics infected nodes j infect susceptible in-neighbors i independently at rate βW_{ij}. On the other hand, infected nodes spontaneously recover independently at a constant rate 1, similarly to the SIS epidemics, with the crucial difference that recovered nodes are not susceptible anymore but rather remain in state 2 ever after. Therefore, in the SIR epidemics the mutation kernels are

$$\psi^{(i)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad i \in \mathcal{V}, \tag{8.6}$$

and the pairwise interaction kernels are

$$\varphi^{(i,j)}(0) = \varphi^{(i,j)}(2) = I, \qquad \varphi^{(i,j)}(1) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (i,j) \in \mathcal{E}, \tag{8.7}$$

Example 8.4 (Voter model)**.** In the *voter model* [], the nodes' states belong to a finite set $\mathcal{A}$ (often, but not always, $\mathcal{A}$ is binary). There are no spontaneous mutations so that the mutation kernels are trivially $\psi^{(i)} = I$ for every node i. On the other hand, nodes i copy their out-neighbors j at independent rate-βW_{ij} Poisson times. Hence, the pairwise interaction kernels are matrices with a single all-one column (corresponding to the state of the observed neighbor) and all other columns equal to zero, i.e.,

$$\varphi^{(i,j)}(c) = \mathbb{1}(\delta^{(c)})^T, \qquad (i,j) \in \mathcal{E}, \quad c \in \mathcal{A}, \tag{8.8}$$

where we recall that $\delta^{(c)}$ is a vector with all zeros except for a 1 in entry c.

Example 8.5 (Evolutionary dynamics)**.** In the *evolutionary graph dynamics* introduced in [27] the states belong to a finite set of species $\mathcal{A}$ (often, but not always, $\mathcal{A}$ is binary). To every specie $a \in \mathcal{A}$ a *fitness parameter* $f_a \in (0, 1]$ is associated. Analogously to the voter model, there are no spontaneous mutations so that the mutation kernels are trivially $\psi^{(i)} = I$ for every node i. On the other hand, the specie $X_i(t)$ present in node i at time t gets replaced by the specie $X_j(t)$ currently present in node j at rate $W_{ij} f_{X_j(t)}$. We can fit this in our pairwise interacting network systems by defining the pairwise interaction kernels as

$$\varphi^{(i,j)}(c) = (1 - f_c)I + f_c \mathbb{1}(\delta^{(c)})^T, \qquad (i,j) \in \mathcal{E}, \quad c \in \mathcal{A}. \tag{8.9}$$

Observe that, if all species have unitary fitness, i.e., if $f_a = 1$ for all $a \in \mathcal{A}$, then the evolutionary dynamics coincide with the voter model.

8.2 Hitting times and absorbing probabilities

All the five examples of pairwise interacting network dynamics above presented in Section 8.1 are reducible Markov chains on the configuration space and have at least one absorbing state. In fact, we can characterize their respective absorbing states as a function of the graph structure as follows.

Proposition 8.1. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a strongly connected graph of order $|\mathcal{V}| = n$ and let $\beta > 0$ be an interaction frequency parameter. Then,

(i) **SI:** the only absorbing configurations for the SI epidemics are $x = \mathbb{1}$ (the all-infected configuration) and $x = 0$ (the all-susceptible configuration). Moreover, the all-susceptible configuration is not reachable by any other configuration $x \neq 0$ while the all-infected configuration is reachable from every configuration $x \neq 0$ with at least one infected node;

(ii) **SIS:** the only absorbing configuration for the SIS epidemics is the all-susceptible configuration $x = 0$, which is reachable from every configuration $x \in \{0, 1\}^{\mathcal{V}}$;

(iii) **SIR:** the absorbing configurations for the SIR epidemics are the 2^n configurations $x \in \{0, 1, 2\}^{\mathcal{V}}$ with no infected nodes, i.e., such that $x_i \in \{0, 2\}$ for every node $i \in \mathcal{V}$. Moreover, the set of such absorbing configurations is globally reachable in $\mathcal{X}$;

(iv) **Voter Model/Evolutionary Dynamics:** the only absorbing states for the voter model and the evolutionary dynamics are the consensus configurations $x = c\mathbb{1}$ for $c \in \mathcal{A}$. Moreover, each consensus configuration $x = c\mathbb{1}$ is reachable from any other configuration y such that $y_i = c$ for some node $i \in \mathcal{V}$.

Proof. To be completed □

In all the examples considered above, the set of absorbing states is always non empty and globally reachable. As a consequence, the asymptotics of such systems are described by Proposition 7.1 ensuring that the process $X(t)$ will almost surely be absorbed in one of the absorbing states in finite time. At this point there are mainly two issues worth of interest:

- the computation (or estimation) of the absorbing time of the process. It is particularly interesting to understand how such absorbing time depends on the topology of the network and how it scales with respect to its size. As we will see, there are situations where this absorbing time is so large that the absorbing event is not observed in real applications. In this case, the transient behavior of the system becomes the main point of interest;

- the computation of the absorbing probabilities when more than one absorbing states are present. This is the case, for instance, of the SIR epidemics, the voter and the evolutionary dynamics models. In the SIR model such absorbing probabilities will give us the distribution of the final number of nodes that are affected by the infection and then recover. In the voter model and in the evolutionary dynamics, these will give us the probability that consensus is reached on a given state and, respectively the fixation probability for a specie.

Homogeneous pairwise interacting network system with binary state space share some peculiarities that allow a certain simplification in their analysis. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, for every binary configuration $x \in \{0,1\}^{\mathcal{V}}$, we define

$$\eta(x) = \mathbb{1}'x = \sum_{i \in \mathcal{V}} x_i \,, \quad \zeta_+(x) = \sum_{i,j \in \mathcal{V}} W_{ij}(1-x_i)x_j, \quad \zeta_-(x) = \sum_{i,j \in \mathcal{V}} W_{ij}x_i(1-x_j) \,. \tag{8.10}$$

The quantity $\eta(x)$ simply coincides with the total number of nodes in state 1 in configuration x (these may represent, e.g., the number of infected nodes in the SI or SIS epidemic models). On the other hand, the quantity $\zeta_+(x)$ coincides with the aggregate weight of the *active in-boundary* of the configuration x, i.e., the set of links pointing from nodes in state 0 to nodes in state 1, and, symmetrically, the quantity $\zeta_-(x)$ coincides with the aggregate weight of the *active out-boundary* of the configuration x, i.e., the set of links connecting nodes in state 1 with nodes in state 0. Notice that, if the graph $\mathcal{G}$ is undirected, then the in- and out-boundary of any configuration x have the same aggregate weight, so that $\zeta_+(x = \zeta_-(x)$.

Now let us consider a pairwise interactive network system $X(t)$ on a graph $\mathcal{G}$ with binary state space $\mathcal{A} = \{0,1\}$ and homogeneous mutation kernel ψ and interaction kernels $\varphi(0)$ and $\varphi(1)$. Let us further assume that

$$\varphi_{ab}(a) = 0 \,, \qquad a \neq b \in \{0,1\} \,. \tag{8.11}$$

In fact, property (8.11) amounts to say that pairwise interactions between nodes with the same state have no effect on the dynamics of the pairwise interacting network system. This clearly holds true in the examples of pairwise interactive systems with binary state space that we have discussed earlier on, i.e., SI and SIS epidemics, binary voter model, and evolutionary dynamics. There are two auxiliary processes that play a crucial role in this case. Through the functions defined in (8.10), we introduce the auxiliary processes

$$N(t) := \eta(X(t)) \,, \qquad B_+(t) := \zeta_+(X(t)) \,, \qquad B_-(t) := \zeta_-(X(t)) \,.$$

Observe that the process $N(t)$ takes values in the set $\{0, 1, \ldots, n\}$, that is much smaller that the configuration space $\mathcal{X} = \{0,1\}^{\mathcal{V}}$. Moreover, the only transitions for the process $N(t)$ correspond to an increment or a decrement of a single unit. Furthermore, one has that $X(t) = 0$ if and only if $N(t) = 0$ and $X(t) = \mathbb{1}$ if and

only if $N(t) = n$, so that in particular hitting times for $X(t)$ in the all-zero and all-one configurations coincide with the hitting times for $N(t)$ in states 0 and n, respectively. However, the process $N(t)$ is in general non-Markovian, since the rate at which the global number of nodes in state 1 increases or decreases in general depends on the current configuration $X(t)$ and not just on $N(t) = \eta(X(t))$. In fact, given the current configuration $X(t) = x$, the increase and decrease rates for the total number of 1's $N(t)$ are given by

$$\begin{aligned} \sum_{\substack{i\in\mathcal{V}:\\ x_i=0}} \Lambda_{x,x+e_i} &= \sum_{i\in\mathcal{V}} (1-x_i)\left(\psi_{0,1} + \sum_{j\in\mathcal{V}} x_j \beta W_{ij}\varphi_{01}(1)\right) \\ &= (n-\eta(x))\psi_{0,1} + \zeta_+(x)\beta\varphi_{01}(1) \end{aligned} \tag{8.12}$$

and

$$\begin{aligned} \sum_{\substack{i\in\mathcal{V}:\\ x_i=1}} \Lambda_{x,x-e_i} &= \sum_{i\in\mathcal{V}} x_i \left(\psi_{1,0} + \sum_{j\in\mathcal{V}} (1-x_j)\beta W_{ij}\varphi_{10}(0)\right) \\ &= \eta(x)\psi_{1,0} + \zeta_-(x)\beta\varphi_{10}(0)\,, \end{aligned} \tag{8.13}$$

respectively. Notice how these two quantities are not simply function of the quantity $\eta(x)$, but also on $\zeta_+(x)$ and $\zeta_-(x)$. This implies that the process $N(t)$ is not Markovian in general. There are however cases where $\zeta(x)$ is actually a function of $\eta(x)$. In these cases $N(t)$ is Markovian and in fact it is a continuous-time birth-and-death chain whose birth and death rates are given by (8.12) and (8.13), respectively. In this case the computation of the absorbing time and the absorbing probabilities prove particularly simple. A general example in this sense is when the underlying graph is complete: in this case $\zeta(x) = \eta(x)(n-\eta(x))$. Other examples will be considered below. There are other situations where $\zeta(x)$, even if not reducible to a function of $\eta(x)$, can be bounded by quantities depending on $\eta(x)$ only, making it possible to obtain bounds on the expected hitting times and absorbing probabilities in the all-zero and all-one configurations.

Below we present a number of basic examples where these ideas are fully exploited. If S is the set of absorbing states and T_S is the hitting time of S, we put $\tau_k = \mathbb{E}[T_S \,|\, N(0) = k]$.

Example 8.6 (Absorbing time for SI)**.** Below we compute the mean absorbing time for the SI model in three different topologies: the star, the cycle, and the complete graph. We recall that the only absorbing state is the all-infected configuration $\mathbb{1}$.

- **Star graph** Let us consider the SI epidemics on a star graph with one center node $i = 1$ and $n-1$ leaf nodes $i = 2, 3, \ldots, n$. Let the initial configuration $X(0) = x$ be such that the center node is infected ($x_1 = 1$) and all leaf nodes susceptible ($x_i = 0$ for $2 \le i \le n$). Clearly, the center node remains infected so that $X_1(t) = 1$ for every $t \ge 0$. Notice now

that, for every configuration x such that $x_1 = 1$ and $x \neq \mathbb{1}$, we have that $\zeta(x) = n - \eta(x)$ so that (restricting to such initial condition) $N(t)$ is a birth-and-death chain with increase and decrease rates from state $k \in \{1, 2, \dots, n\}$ given by $\lambda_k = (n-k)\beta$ and $\mu_k = 0$ respectively. It follows from point (v) of Theorem 7.2 that the expected hitting times $\tau_k = \mathbb{E}_k[T_n]$ for this chain satisfy the recursions $\tau_k = \lambda_k^{-1} + \tau_{k+1}$ for $k = 1, \dots, n-1$ and $\tau_n = 0$ so that

$$\tau_k = \sum_{h=k}^{n-1} \lambda_h^{-1} = \beta^{-1} \sum_{j=1}^{n-k} \frac{1}{j}$$

In particular, $\tau_1 = \beta^{-1} H_{n-1}$. where $H_{n-1} = \sum_{k=1}^{n-1} \frac{1}{k}$ is the $(n-1)$-th harmonic number, satisfying $\log n \leq H_{n-1} < \log n + 1$ for $n \geq 1$. Hence, for the SI epidemics on the star graph where the center node is the only one initially infected, we get

$$\tau_1 = \mathbb{E}\left[T_{\mathbb{1}} | X(0) = e_1\right] \asymp \frac{1}{\beta} \log n \,. \tag{8.14}$$

- **Ring graph** Let us consider the SI epidemics on a cycle graph with initial configuration $X(0) = x$ having a single node infected and all the others susceptible. We observe that, at any time $t \geq 0$, the set of infected nodes will always be connected (formally, this can be easily proved by induction). Notice that for any such configuration $x \neq \mathbb{1}$, we have that $\zeta(x) = 2$ so that, again, $\eta(x)$ is a birth and death chain with increasing and decreasing rates from the state $k \in \{1, 2, \dots, n-1\}$ given by

$$\lambda_k = 2\beta, \ \mu_k = 0$$

Hence, for the SI epidemics on the cycle graph where only one node is initially infected, we get

$$\tau_1 = \mathbb{E}\left[T_{\mathbb{1}} | N(0) = 1\right] = \frac{1}{2\beta}(n-1) \,. \tag{8.15}$$

- **Complete graph**

Finally, we consider the SI epidemics on a complete graph with n nodes where the initial configuration $X(0) = x$ has only one node infected. In this case we know that $\eta(x)$ is a birth and death chain with increasing and decreasing rates from the state $k \in \{1, 2, \dots, n-1\}$ given by

$$\lambda_k = k(n-k)\beta, \ \mu_k = 0$$

Therefore

$$\tau_1 = \mathbb{E}\left[T_{\mathbb{1}} | N(0) = 1\right] = \sum_{k=1}^{n-1} \frac{1}{\beta k(n-k)} = \frac{1}{\beta n} \sum_{k=1}^{n-1} \left(\frac{1}{k} + \frac{1}{n-k} \right) = \frac{2}{\beta} \frac{H_{n-1}}{n} \asymp \frac{2}{\beta} \frac{\log n}{n} \,. \tag{8.16}$$

In particular, the expected time for complete infection for the SI epidemics on the complete graph is decreasing in the network size n and vanishes as n grows large. This comes as the result of a dual effect of the network size in the complete graph: on the one hand, as n grows large, the number of nodes to infect grows linearly in n, on the other hand, when a nontrivial fraction of the population is infected, the size of the active boundary grows quadratically in n, thus dominating the former effect.

Example 8.7 (Extinction time for the SIS epidemics on the complete graph)**.** Let us consider the SIS epidemics on a complete graph with n nodes. We know from general considerations that $N(t) = \sum_i X_i(t)$ is a birth-and-death chain with transition rates

$$\Lambda_{k,k-1} = k\,, \qquad \Lambda_{k,k+1} = \beta k(n-k)\,, \qquad k = 0, \ldots, n\,.$$

The mean extinction times τ_k satisfy the linear equations

$$\begin{aligned} \tau_0 &= 0 \\ \tau_k &= \frac{1}{k+\beta k(n-k)} + \frac{k\tau_{k-1}}{k+\beta k(n-k)} + \frac{\beta k(n-k)\tau_{k+1}}{k+\beta k(n-k)}\,, \qquad 1 \le k < n \\ \tau_n &= \frac{1}{n} + \tau_{n-1}\,. \end{aligned} \tag{8.17}$$

To solve the above, it proves convenient to introduce the new variables

$$y_l = \tau_{n-l+1} - \tau_{n-l}\,, \qquad l = 1, \ldots, n\,.$$

Then, (8.17) implies the recursion

$$y_1 = \frac{1}{n}\,, \qquad y_{l+1} = \frac{1}{n-l} + \beta l y_l\,, \qquad l = 1, \ldots, n\,,$$

whose solution is

$$y_l = \sum_{k=0}^{l-1} \frac{(l-1)!}{k!} \cdot \frac{\beta^{l-1-k}}{n-k}\,, \qquad l = 1, \ldots, n\,.$$

From the above, we find in particular that

$$\tau_1 = y_n = \sum_{k=0}^{n-1} \frac{(n-1)!}{k!} \cdot \frac{\beta^{n-1-k}}{n-k}\,, \qquad \tau_n = \sum_{l=1}^{n} y_l = \sum_{l=1}^{n} \sum_{k=0}^{l-1} \frac{(l-1)!}{k!} \cdot \frac{\beta^{l-1-k}}{n-k}\,.$$

Observe that, regardless of the the value of β,

$$\tau_n \ge \sum_{l=1}^{n} \frac{1}{n-l+1} = H_n \ge \log n\,,$$

where H_n is the n-th harmonic number. Hence, the extinction time when starting from an all-infected network is at least logarithmic in network size n.

In fact, let

$$\alpha = n\beta\,. \tag{8.18}$$

Then, if $\alpha < 1$, we have that

$$\tau_n = \sum_{k=0}^{n-1} \frac{1}{n-k} \sum_{j=0}^{n-1-k} \frac{(j+k)!}{k!} \left(\frac{\alpha}{n}\right)^j \leq \sum_{k=0}^{n-1} \frac{1}{n-k} \sum_{j=0}^{n-1-k} \alpha^j \leq \frac{H_n}{1-\alpha} \asymp \frac{\log n}{1-\alpha}\,,$$

i.e., the expected extinction time from any initial configuration with a positive number of infected nodes (including a fully infected network) grows logarithmically in n.

On the other hand, if $\alpha > 1$, then, for every $\gamma \in [1/\alpha, 1]$,

$$\tau_1 \geq \frac{(n-1)!}{\lfloor n\gamma \rfloor!} \cdot \frac{\beta^{n-1-\lfloor n\gamma \rfloor}}{n - \lfloor n\gamma \rfloor} \geq \frac{(n-1)(n-2)\dots(\lfloor n\gamma \rfloor + 1)\alpha^{n-1-\lfloor n\gamma \rfloor}}{n^{n-1-\lfloor n\gamma \rfloor}(n - \lfloor n\gamma \rfloor)} \geq \frac{(\alpha\gamma)^{(1-\gamma)n}}{\gamma\alpha n}\,,$$

so that, for every $k = 1, \dots, n$,

$$\tau_k \geq \tau_1 \geq \frac{e^{g(\alpha)n}}{\alpha n}\,, \qquad g(\alpha) = \max_{\frac{1}{\alpha} \leq \gamma \leq 1} (1-\gamma)\log(\alpha\gamma) > 0$$

i.e., the expected extinction time from any initial configuration with a positive number of infected nodes (even just a single one!) grows exponentially fast in n.

Example 8.8 (Absorbing probabilities for the binary voter model)**.** Let us consider the voter model on a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with $\mathcal{A} = \{0, 1\}$. By Proposition 8.1, the configuration $X(t)$ will reach a consensus in finite time. The computation of the mean absorbing time is analytically feasible only in particular cases like the complete graph using techniques similar than in the SIS epidemics. Here we instead concentrate on the concentration of the absorbing probabilities, namely the probabilities that $X(t)$ will be absorbed by 0 or by $\mathbb{1}$, respectively. Surprisingly enough this computation can be carried on analytically on any graph topology. Indeed, since in this case $\varphi_{01}(1) = \varphi_{10}(0) = 1$, it follows from (**??**) that the rate at which the number of 1's can increase or decrease by 1 in any configuration $x \neq 0, \mathbb{1}$, are always equal. This implies that, even if in general $N(t)$ is not a Markovian process, the underlying jump chain

$$M(h) = N(S_h^+)\,, \qquad h \geq 0\,,$$

where $0 = S_0 \leq S_1 \leq S_2 \leq \dots$ are the random times when the value of $N(t)$ changes, is a birth and death chain identical to the gambler's model already analyzed. Hence, the probability that the final consensus value is 1 for the voter model on a connected undirected graph of size n is the same as the probability that the gambler's fortune reaches value n before value 0, as computed in Example 7.2:

$$\mathbb{P}\left(\lim_{t \to +\infty} X(t) = \mathbb{1} | X(0)\right) = \frac{1}{n} N(0) = \frac{1}{n} \sum_i X_i(0)\,. \tag{8.19}$$

Example 8.9 (Fixation probability for evolutionary dynamics on undirected graph)**.** Let us consider the evolutionary dynamics on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with two species only, a native specie 0 with fitness f_0 and a mutant specie 1 with fitness f_1, so that $\mathcal{A} = \{0, 1\}$, and denote by $\rho = f_1/f_0$ the relative fitness of the mutant specie with respect to the native one. We are interested in computing the *fixation probability*, i.e., the probability that only mutants survive while the native specie eventually gets extinct, when starting with an initial configuration comprising a positive but typically small number of mutants. We proceed as in Example 8.8 and we notice that in this case $\varphi_{01}(1) = \rho\varphi_{10}(0) = 1$. As a consequence, the ratio between the increasing and the decrease rates in (8.13) is also ρ in any configuration $x \neq 0, \mathbb{1}$. This implies that, in this case, the jump chain associated to $N(t)$

$$M(h) = N(S_h^+)\,, \qquad h \geq 0\,,$$

where $0 = S_0 \leq S_1 \leq S_2 \leq \dots$ are the random times when the value of $N(t)$ changes, is a birth and death chain iwith increasing and decreasing probabilities equal to, respectively, $\rho/(1+\rho)$ and $1/(1+\rho)$. Therefore, the fixation probability for the evolutionary dynamics on a connected undirected graph of size n coincides with the probability that the birth-and-death chain with transition probabilities as above reaches state n before state 0 that was computed in (7.16):

$$\mathbb{P}\left(\lim_{t\to+\infty} X(t) = \mathbb{1} | X(0)\right) = \frac{1-\rho^{-N(0)}}{1-\rho^{-n}} = \frac{1-\rho^{-\sum_i X_i(0)}}{1-\rho^{-n}}\,. \tag{8.20}$$

We end this part by analyzing a non-binary example.

Example 8.10 (Final number of recovered nodes for the SIR epidemics on the star graph)**.** Let us consider the SIR epidemics on a star graph with one center node $i = 1$ and $n-1$ leaf nodes $i = 2, 3, \dots, n$. Let the initial configuration $X(0) = x$ be such that the center node infected ($x_1 = 1$) and all leaf nodes susceptible ($x_i = 1$ for $2 \leq i \leq n$). Our goal is to characterize the distribution and the expected value of the the number N of eventually recovered nodes. Trivially,

$$\mathbb{P}(N \geq 1) = 1\,, \tag{8.21}$$

as the initially infected center node eventually recovers with probability 1. Now, observe that all possible infections occur between time zero and the recovery time T of the center node. Once such center node is recovered, no more infections are possible, and those other nodes that have been infected by that time will eventually recover, whereas the nodes that have not been infected by then will remain susceptible. Notice that, conditioned on the recovery time for the center node being $T = t$, each of the other nodes gets infected independently with probability $1 - e^{-\beta t}$. Hence, for $i = 2, 3, \dots, n$ the probability that an arbitrary leaf node gets infected and will eventually end up recovered is

$$\mathbb{P}\left(\lim_{t\to+\infty} X_i(t) = 2\right) = \int_0^{+\infty} (1 - e^{-\beta t}) e^{-t} \mathrm{d}t = \frac{\beta}{1+\beta}\,, \qquad i = 2, 3, \dots, n\,.$$

It follows that the expected number of eventually recovered node is

$$\mathbb{E}[N] = 1 + \sum_{2 \le i \le n} \mathbb{P}\left(\lim_{t \to +\infty} X_i(t) = 2 \right) = 1 + (n-1)\frac{\beta}{1+\beta} = \frac{1+n\beta}{1+\beta}\,. \quad (8.22)$$

On the other hand, finding the probability distribution of N is less straightforward. Indeed, albeit conditional independent given the recovery time of the center node, the infection events for the $n-1$ leaf nodes are not simply independent. In fact, intuition suggests that infections are positively correlated: given that a certain leaf gets infected, the conditional distribution of the recovery time for the center node is no longer rate-1 exponential, but rather tends to put more mass on higher values. In fact, we can obtain obtain the probability distribution of the total number of recovered nodes as follows. First, observe that the probability that no leaf node gets infected by the time the center node recovers, i.e., the probability that $N = 1$, satisfies

$$\mathbb{P}(N=1) = \int_0^{+\infty} \mathbb{P}(N=1|T=t)e^{-t}\mathrm{d}t = \int_0^{+\infty} e^{-(n-1)\beta t}e^{-t}\mathrm{d}t = \frac{1}{1+(n-1)\beta}\,.$$

The argument above can be generalized upon noticing that, for $k \ge 1$, $N \ge k$ if and only if the $(k-1)$-th of the $n-1$ independent links Poisson clocks happens to tick before the recovery time T of the center node. Then, using the memoryless property of the exponential distributions, one has that, conditioned on $N \ge k$, the time for the k-th link Poisson clock to tick and the residual recovery time $T - t$ are independent exponential random variables of rate $(n-k)$ and 1, respectively. Therefore,

$$\mathbb{P}(N=k|N \ge k) = \int_0^{+\infty} e^{-(n-k)\beta t}e^{-t}\mathrm{d}t = \frac{1}{1+(n-k)\beta}\,,$$

$$\mathbb{P}(N \ge k+1|N \ge k) = \frac{(n-k)\beta}{1+(n-k)\beta}\,.$$

Using (8.21) and the above, one finds that

$$\mathbb{P}(N \ge k) = \mathbb{P}(N \ge 1) \prod_{1 \le j \le k-1} \mathbb{P}(N \ge j+1|N \ge j) = \prod_{1 \le j \le k-1} \frac{(n-j)\beta}{1+(n-j)\beta}\,, \quad (8.23)$$

$$\mathbb{P}(N=k) = \mathbb{P}(N=k|N \ge k)\mathbb{P}(N \ge k) = \frac{\prod_{1 \le j \le k}(n-j)\beta}{\prod_{1 \le j \le k-1}(1+(n-j)\beta)}\,, \quad (8.24)$$

for every $k \ge 1$.

Chapter 9

Games, networks, and learning

In this chapter we study network dynamics in the context of game theory. We first present the basic elements of classical game theory and we then introduce two fundamental examples of Markov processes describing evolutive models in a game theoretic framework.

9.1 Basic elements of Game Theory

We consider games in strategic form. There is a finite set of *players* $\mathcal{V}$ and a set of *actions* $\mathcal{A}$. While for the most part of these notes $\mathcal{A}$ is assumed to be finite (and this assumption is implicit in all results), we also present few examples where $\mathcal{A}$ is the set of real numbers. The assignment of an action to each player is described by a vector $x \in \mathcal{A}^{\mathcal{V}}$ that is called an *action profile* or *configuration*. Throughout, we shall denote by

$$\mathcal{X} = \mathcal{A}^{\mathcal{V}}$$

the *configuration space*. Each player $i \in \mathcal{V}$ is equipped with a *utility* function (a.k.a. *reward* or *payoff* function)

$$u_i : \mathcal{X} \to \mathbb{R}$$

that associates with every action profile x in $\mathcal{X}$ the utility $u_i(x)$ that player i gets when each player j is playing action $x_j \in \mathcal{A}$. We will often use the notation

$$x_{-i} = x_{|\mathcal{V}\setminus\{i\}}$$

for the vector obtained from the action profile x by removing its i-th entry, and, with a slight abuse of notation, write

$$u_i(x_i, x_{-i}) = u_i(x) \tag{9.1}$$

for the utility received by player i when she chooses to play action x_i, and the rest of the players choose to play x_{-i}. The triple $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ will be referred to as a *(strategic form) game.*

Remark 6. *In the literature, games are typically presented in a more general setting, allowing for the possibility that players have different action sets $\mathcal{A}_i$ for $i \in \mathcal{V}$. As such greater generality is not necessary in order to get into the basics of game theory, for the sake of notation compactness we prefer to stick to our less general case.*

Every player i is to be interpreted as a rational agent choosing her action x_i from the action set $\mathcal{A}$ so as to maximize her own utility $u_i(x_i, x_{-i})$. In consideration of the fact that this utility depends not only on player i's action x_i but also on the actions of the rest of the players' actions x_{-i}, it is natural to introduce the (set-valued) *best response* (BR) function

$$\mathcal{B}_i(x_{-i}) = \operatorname*{argmax}_{x_i\in\mathcal{A}} u_i(x_i, x_{-i})\,.$$

Assuming that player i knows what the rest of the players' actions are and that these are not changing, choosing an action in $\mathcal{B}_i(x_{-i})$ is for her the rational choice as it makes her utility as large as possible. Of course, when $\mathcal{A}$ is not finite, $\mathcal{B}_i(x_{-i})$ could as well be an empty set.

Definition 9.1 (Pure strategy Nash equilibrium)**.** A *(pure strategy) Nash equilibrium* (NE) for the game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ is an action configuration $x^* \in \mathcal{X}$ such that

$$x_i^* \in \mathcal{B}_i(x_{-i}^*)\,, \qquad i \in \mathcal{V}\,. \tag{9.2}$$

The Nash equilibrium is called *strict* if, moreover, $|\mathcal{B}_i(x_{-i}^*)| = 1$ for every i.

The interpretation of a Nash equilibrium is the following: it is an action profile such that no player has any incentive to *unilaterally* deviate from her current action, as the utility she is getting with the current action is the best possible given the current actions chosen by the other players. Note the emphasis on 'unilaterally': it is not at all guaranteed that coordinated deviations of more than one player from their actions in a Nash equilibrium could not lead to a higher utility for these players. As we shall see, there are games with multiple NE and games which instead have none. We denote by $\mathcal{N}$ the set of NE of a game.

9.2 Two-player, congestion, and network games

In this section, we present some among the most popular examples of games. First, we discuss two-player games and present several examples. Then, we introduce the class of congestion games. Finally, we discuss an important class of network games.

	-1	+1
-1	a,a	d,c
+1	c,d	b,b

Figure 9.1: Payoff matrix of a symmetric two-player game with action set $\mathcal{A} = \{-1, +1\}$.

9.2.1 Two-player games

The simplest examples of games are those with just two players $\mathcal{V} = \{1, 2\}$, that are referred to as *two-player games*. A two-player game is simply characterized by two utility functions $u_i(r, s)$, for $i = 1, 2$, with the understanding that r is the action played by player i and s the action played by his opponent. An important special case is when

$$u_1(r, s) = u_2(r, s) = \varphi(r, s)\,, \qquad r, s \in \mathcal{A}\,, \tag{9.3}$$

which amounts to say that the role of the two players is exchangeable. A two-player game satisfying (9.3) is referred to as *symmetric*. Symmetric two-player games are characterized by a single utility function $\varphi(r, s)$ that is the utility that a player choosing action r gets when the other player plays action s.

When the action set $\mathcal{A}$ is finite, utilities of a two-player game are typically represented by a table, referred to as the *payoff matrix*, whose rows correspond to the actions of player 1, whose columns represent the actions of player 2, and where the (r, s)-th entry displays the pair $u_1(r, s), u_2(s, r)$ of the utility values that player 1 and player 2 receive when they play action r and s respectively. two-player games with binary action space, i.e., when $|\mathcal{A}| = 2$, are typically referred to as 2×2-*games*. The payoff matrix of a symmetric 2×2-game is represented as in Figure 9.1.

Below we report some basic popular examples of two-player games.

Example 9.1 (Prisoner's dilemma)**.** The prisoner's dilemma is a symmetric 2×2-game with payoff matrix as in Figure 9.1 whose entries satisfy:

$$c < a < b < d\,.$$

The classical interpretation is that two members of a criminal gang are arrested and imprisoned. Each prisoner is in solitary confinement with no means of speaking to or exchanging messages with the other. The prosecutors do not have enough evidence to convict the pair on the principal charge but have evidence to convict them to $-b$ years in prison on a lesser charge. Simultaneously, the prosecutors offer each prisoner a bargain. Each prisoner is given the opportunity either to: confess the other by testifying that he committed the main crime (action -1), or to remain silent (action $+1$). If one prisoner betrays the other and the other stays silent, the betrayer is freed of both the minor and major charges (corresponding to a payoff d) and the one who remains silent get

sentenced to $-c$ years in prison. If both prisoners betray each other, they both get sentenced to $-a$ years in prison each.

It is not hard to verify that the prisoner dilemma admits a unique (pure strategy) Nash equilibrium $(-1,-1)$, i.e., where the prisoners betray each other. Indeed, a stronger claim is true: action -1 (betraying) *dominates* action 1 (remaining silent), i.e., irrespective of what the other prisoner does betraying guarantees a better payoff. In other terms, the best response is

$$\mathcal{B}_1(-1) = \mathcal{B}_1(+1) = \mathcal{B}_2(-1) = \mathcal{B}_2(+1) = -1\,.$$

Observe that, on the other hand, if the prisoners could coordinate and remain silent (so that the action profile is $(+1,+1)$) then they would both get a better payoff $b > a$ than the one they get at the Nash equilibrium $(-1,-1)$. □

Example 9.2 (Coordination game)**.** A *coordination game* is a symmetric 2×2-game with payoff matrix as in Figure 9.1 whose entries satisfy:

$$a > c \qquad b > d\,.$$

The inequalities above imply that the best response for each player is to copy the action of the other player, i.e.,

$$\mathcal{B}_1(-1) = \mathcal{B}_2(-1) = -1\,, \qquad \mathcal{B}_1(+1) = \mathcal{B}_2(+1) = +1,\,.$$

Thus, a coordination game admits two Nash equilibria: $(-1,-1)$ and $(+1,+1)$. Coordination games are fundamental models of interactions with a *positive externality*: by choosing an action, every player possibly increases the utility of other players choosing the same action. Note that the two Nash equilibria of a coordination game need not be equally good for the players: in particular, if $a > b$, then the equilibrium $(-1,-1)$ is referred to as *payoff dominant*.

Example 9.3 (Anti-coordination game)**.** An *anti-coordination game* is a symmetric 2×2-game with payoff matrix as in Figure 9.1 whose entries satisfy:

$$a < c \qquad b < d\,.$$

Anti-coordination games admit two Nash equilibria $(+1,-1)$, and $(-1,+1)$. The most popular anti-coordination game is the game of Chicken (also known as Hawk-Dove game) where, additionally,

$$a > d\,.$$

Example 9.4 (Discoordination game)**.** A *discoordination game* is a 2×2-game with payoff matrix

	-1	+1
-1	a,b	c,d
+1	c,d	a,b

whose entries satisfy

$$a > c \qquad d > b$$

The most popular example of a discoordination game is the *matching penny game* where $a = d = 1$ and $b = c = -1$. Matching penny is a *zero-sum game*, since $a + b = c + d = 0$. Discoordination games are not symmetric and admit no Nash equilibria.

Example 9.5 (Rock-Scissor-Paper)**.** The Rock-Scissor-Paper game is a two-player symmetric game with action set $\mathcal{A} = \{R, S, P\}$ and payoff matrix

	R	S	P
R	0,0	1,-1	-1,1
S	-1,1	0,0	1,-1
P	1,-1	-1,1	0,0

It is a zero-sum game and has no pure strategy Nash equilibria.

We now provide an example of a game with continuous strategy space.

Example 9.6 (Cournot competition)**.** This is a two-player game where the players are two firms producing a homogeneous good for the same market. The action of a firm i is a quantity, $x_i \in \mathcal{A} = [0, +\infty)$ representing the amount of good it produces. The utility for each firm is its total revenue minus its total cost,

$$u_i(x_1, x_2) = x_i p(x_1 + x_2) - c x_i \,, \qquad i = 1, 2 \,,$$

where $p(q)$ is the price of the good (as a function of the total quantity), and c is unit cost (same for both firms).

Assume for simplicity that $p(q) = \max\{0, 2 - q\}$. Then, the best response is given by

$$\mathcal{B}_i(x_{-i}) = \begin{cases} 1 - c/2 - x_{-i}/2 & \text{if} \quad x_{-i} \leq 2 - c \\ 0 & \text{if} \quad x_{-i} > 2 - c \,. \end{cases}$$

If $c < 2$ (the case when $c \geq 2$ is of no interest as 0 will be in this case a dominant strategy for both players), a simple computation shows that the only NE is given by the configuration

$$x_1 = x_2 = \frac{2 - c}{3}$$

□

9.2.2 Congestion games

We now present an important family of games known as *congestion games*. Applications are ubiquitous and, in particular, they include the network traffic flow models considered in Chapter 4. In congestion games, the number of players is finite but otherwise arbitrary.

The key idea of a congestion game is that the actions chosen by the players are to be interpreted as (subsets of) shared resources and that the utility associated to an action only depends on the total number of players using the same resource(s). In most applications this dependence is non-increasing: by choosing a resource, every player decreases the utility (equivalently, increases the cost) of all players choosing the same action. For this reason, congestion games provide fundamental economic models of interactions with *negative externalities*.

The following is the simplest instance of a congestion game.

Example 9.7 (Congestion 1)**.** Consider a set of players $\mathcal{V}$, a set of actions $\mathcal{A}$ and, for every $a \in \mathcal{A}$, a function $d_a : \mathbb{R}_+ \to \mathbb{R}$, where $d_a(n_a)$ is to be interpreted as the cost of using resource a when n_a players are using it. Given any configuration $x \in \mathcal{X}$ and any resource $a \in \mathcal{A}$, define

$$n_a(x) = |\{i \in \mathcal{V} : x_i = a\}| \tag{9.4}$$

as the number of players choosing resource a in configuration x. Finally, define the utility of a player i as

$$u_i(x) = -d_{x_i}(n_{x_i}(x)),$$

i.e., as the opposite of the cost associated to the chosen resource.

The following is a more elaborated version of a congestion game, where multiple resources are used by players: it adapts well to the case when resources are edges in a graph and an action is a path between two points in a graph.

Example 9.8 (Congestion 2)**.** Consider a set of players $\mathcal{V}$, a finite set of resources $\mathcal{E}$ (e.g., links in a transportation network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$) and congestion costs $d_e : \mathbb{R}_+ \to \mathbb{R}_+$, where $d_e(f_e)$ is the congestion cost associated with resource e when f_e players use it. Every action $a \in \mathcal{A}$ is a (non-empty) subset of $\mathcal{E}$ (e.g., a given o-d path in $\mathcal{G}$). The resource-strategy incidence matrix $A \in \mathbb{R}^{\mathcal{E}\times\mathcal{A}}$ has entries $A_{ea} = 1$ if resource e is used by action a and $A_{ea} = 0$ otherwise. With every configuration $x \in \mathcal{A}^{\mathcal{V}}$, we associate the vector $n(x) \in \mathbb{R}^{\mathcal{A}}$ as in (9.4). The vector $f = An(x)$ has dimension equal to the number of resources $\mathcal{E}$ and entries $f_e = \sum_{a\in\mathcal{A}} A_{ea} n_a(x)$ corresponding to the number of players using the different resources. Finally, the utility of every player $i \in \mathcal{I}$ is

$$u_i(x_i, x_{-i}) = -\sum_{e\in\mathcal{E}} A_{ex_i} d_e(f_e) = -\sum_{e\in\mathcal{E}} A_{ex_i} d_e((An(x))_e) \tag{9.5}$$

In the case when $\mathcal{E} = \mathcal{A}$ and A is the identity matrix, we recover the case treated in Example 9.7. □

9.3 Network games

In this subsection we consider a class of games that is particularly relevant in interconnected systems, which we call *network games*. These are games where

the players are represented as nodes of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and their utilities depend only on their own and their out-neighbors' actions. Formally a network game over the graph $\mathcal{G}$ (or more briefly a $\mathcal{G}$-game) is any triple $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ whose utility functions satisfy the following property: for any player $i \in \mathcal{V}$ and configurations $x, y \in \mathcal{A}^{\mathcal{V}}$ such that $x_j = y_j$ for every $j \in N_i \cup \{i\}$ it holds

$$u_i(x) = u_i(y)$$

Of course, every game is a network game with respect to the complete graph.

In these notes, we mostly focus on network games constructed in the following canonical way. Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with no self-loop and assume that, for every two neighboring nodes i, j it is defined a two-player game. The corresponding utilities of i and j are, respectively, denoted by $\varphi^{(i,j)} : \mathcal{A}\times\mathcal{A} \to \mathbb{R}$ and $\varphi^{(j,i)} : \mathcal{A}\times\mathcal{A} \to \mathbb{R}$ and are called *interaction utilities*. According to the usual terminology used in two-player games $\varphi^{(i,j)}(a, b)$ is the utility obtained by i when he plays action a and its opponent j plays the action b.

We now define the *network game* $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ over $\mathcal{G}$ by setting the utility of every player $i \in \mathcal{V}$ as simply the weighted sum of the utilities of the various two-player games that i is playing with his neighbors:

$$u_i(x) := \sum_j W_{ij}\varphi^{(i,j)}(x_i, x_j) \tag{9.6}$$

The choice of having the graph undirected was solely for the sake of interpreting the interaction as a two-player game. The definition above adapts to any directed graph as long we have the interaction utility $\varphi^{(i,j)}$ defined on each edge $(i, j) \in \mathcal{E}$.

9.3.1 Network games with binary actions

Below we present some fundamental examples of network games when the action set is binary, say $\mathcal{A} = \{-1, 1\}$. At first we assume that the interaction utility function is the same on every edge: $\varphi^{(i,j)} = \varphi$ for every $(i, j) \in \mathcal{E}$ (this says, in the case when the graph is undirected, that the game among any pair of neighboring nodes is symmetric).

We start with two specific popular examples.

Example 9.9. [Majority game] Consider an unweighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ and the interaction utility function

$$\varphi(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

In this case

$$u_i(x) := \sum_j W_{ij}\varphi(x_i, x_j) = |\{j \in N_i \,|\, x_j = x_i\}| \tag{9.7}$$

The utility of a player is simply given by the number of out-neighbors playing its same action. it follows that, for the player $i \in \mathcal{V}$, the best response is given by

$$\mathcal{B}_i(x_{-i}) = \begin{cases} \{+1\} & \text{if} \quad |\{j \in N_i \,|\, x_j = +1\}| > |N_i|/2 \\ \{-1\} & \text{if} \quad |\{j \in N_i \,|\, x_j = +1\}| < |N_i|/2 \\ \{\pm 1\} & \text{if} \quad |\{j \in N_i \,|\, x_j = +1\} = |N_i|/2\,. \end{cases}$$

The best response function for a player thus coincides with the action played by the majority of its out-neighbors: this is the reason for the name majority game.

Observe that the pure configurations $-\mathbb{1}$ and $+\mathbb{1}$ are always (strict) Nash equilibria for any graph $\mathcal{G}$ possibly not undirected.

In the case when the graph $\mathcal{G}$ is complete these are the only Nash equilibria. This can be seen as follows. Given $x \in \mathcal{A}^{\mathcal{V}}$ denote by $n^+(x)$ and $n^-(x)$ the number of components of x equal to $+1$ and -1, respectively. Clearly, $n^+(x) = n^-(x) = n = \mathcal{V}|$. The conditions for x being Nash are in this case:

$$\begin{aligned} x_i = 1 \;\Rightarrow \quad & \sum_{j \in\ N_i} x_j = n^+(x) - 1 - n^-(x) \geq 0 \\ x_i = -1 \;\Rightarrow \quad & \sum_{j \in\ N_i} x_j = n^+(x) - n^-(x) + 1 \leq 0 \end{aligned}$$

If there was a Nash equilibrium x different from both configurations $+\mathbb{1}$ and $-\mathbb{1}$, we would then obtain

$$1 \leq n^+(x) - n^-(x) \leq -1$$

that is clearly absurd.

When $\mathcal{G}$ is not the complete graph, these are not necessarily the only Nash equilibria.

E.g., Figure 9.2 exhibits all the Nash equilibria of the network coordination game on a ring graph of order 4. Besides the two Nash equilibria $-\mathbb{1}$ and $+\mathbb{1}$, there are in this case other four Nash equilibria (observe that none of them is strict).

Example 9.10. [Minority game] This is a sort of opposite of the majority game: utility of a player is here the number of out-neighbors playing the opposite action. Formally, we consider again an unweighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with the interaction utility function

$$\varphi(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

In this case

$$u_i(x) := \sum_j W_{ij} \varphi(x_i, x_j) = |\{j \in N_i \,|\, x_j \neq x_i\}| \tag{9.8}$$

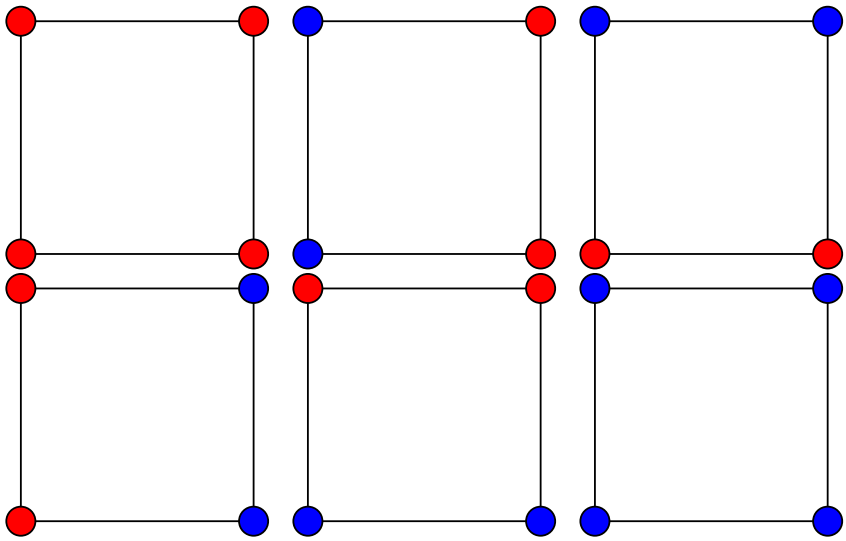


Figure 9.2: The majority coordination game on a ring graph with 4 nodes has 6 Nash equilibria.

For the player $i \in \mathcal{V}$, the best response is now given by

$$\mathcal{B}_i(x_{-i}) = \begin{cases} \{+1\} & \text{if} \quad |\{j \in N_i \,|\, x_j = +1\}| < |N_i|/2 \\ \{-1\} & \text{if} \quad |\{j \in N_i \,|\, x_j = +1\}| > |N_i|/2 \\ \{\pm 1\} & \text{if} \quad |\{j \in N_i \,|\, x_j = +1\} = |N_i|/2\,. \end{cases}$$

The best response function for a player thus coincides with the action played by the minority of its out-neighbors: this is the reason for the name minority game.

In spite of the apparent structural similarity with the majority game, this games exhibits quite different properties. In particular, differently from the majority game, there is not in this case a configuration that is always a Nash equilibrium, independently on the particular graph. For the 4-cycle C_4, Nash equilibria are shown in Figure 9.3.

For this game, Nash equilibria are not guaranteed to exist in general. A simple example is for instance the directed 3-cycle $\mathcal{G} = (\{1,2,3\}, \{(1,2),(2,3),(3,1)\})$. We will see below that NE always exist when the graph is undirected.

The two examples above are special cases of binary network games obtained by aggregating two-player games all having the same utility function φ on every edge. It is worth to describe the general case. Let

$$\varphi(-1,-1) = a\,, \quad \varphi(1,1) = b\,, \quad \varphi(1,-1) = c\,, \quad \varphi(-1,1) = d\,,$$

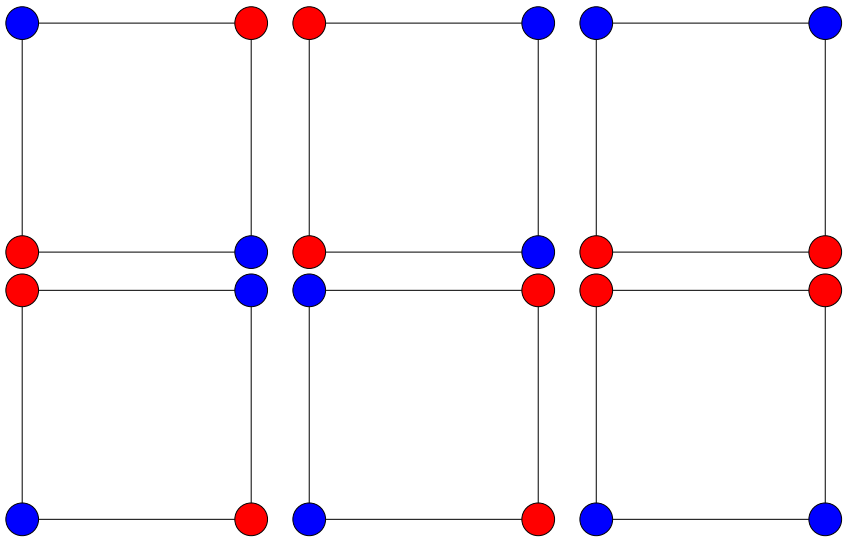


Figure 9.3: The minority game on a ring graph with 4 nodes has 2 Nash equilibria.

as in Figure 9.1. The utilities of the corresponding network game can be expressed as follows. Given a configuration $x \in \mathcal{A}^{\mathcal{V}}$ and a player $i \in \mathcal{V}$, denote

$$w_i^+(x) = \sum_{j \in \mathcal{V} x_j = +1} W_{ij}, \quad w_i^-(x) = \sum_{j \in \mathcal{V} x_j = -1} W_{ij}$$

Recalling that $w = W\mathbb{1}$ is the out-degree vector of $\mathcal{G}$, we clearly have that $w_i^+(x) + w_i^-(x) = w_i$ for every $i \in \mathcal{V}$. Utilities are given by

$$u_i(x_i, x_{-i}) = \sum_{j \in \mathcal{V}} W_{ij}\varphi(x_i, x_j) = \begin{cases} bw_i^+(x) + cw_i^-(x) & \text{if} \quad x_i = +1 \\ \\ dw_i^+(x) + aw_i^-(x) & \text{if} \quad x_i = -1\,, \end{cases} \tag{9.9}$$

The best response function is then given by

$$\mathcal{B}_i(x_i) = \begin{cases} \{+1\} & \text{if} \quad (b-d)w_i^+(x) + (c-a)w_i^-(x) > 0 \\ \{-1\} & \text{if} \quad (b-d)w_i^+(x) + (c-a)w_i^-(x) < 0 \\ \{\pm 1\} & \text{if} \quad (b-d)w_i^+(x) + (c-a)w_i^-(x) = 0\,, \end{cases}$$

Notice that

$$(b-d)w_i^+(x) + (c-a)w_i^-(x) = (a+b-c-d)w_i^+(x) + (c-a)w_i$$

This shows that the best response is in the form of a *threshold function* of the quantity $w_i^+(x)$ (or equivalently $w_i^-(x)$) with respect to the degree w_i. It is useful to introduce the following two constants:

$$\alpha = a - b - c + d\,, \qquad \beta = a + b - c - d$$

When $\alpha > 0$ ($\alpha < 0$), action $+1$ (respectively, -1) is said to be *risk-dominant*, since it is the best response when the total weights of neighbors choosing action $+1$ and -1 are exactly the same. Otherwise, if $\alpha = 0$, the two actions are said to be risk-neutral: this was the case of the majority and minority games described before. On the other hand, the constant β reflects the nature of the interaction: if φ is the utility function of a coordination game we have $a > c$ and $b > d$, so that $\beta > 0$; if instead φ is the utility function of an anti-coordination game we have $a < c$ and $b < d$, so that $\beta < 0$. We will refer to these two cases as to, respectively, a network coordination and anti-coordination game.

In the case of a network coordination game, the best response can be rewritten as

$$\mathcal{B}_i(x_i) = \begin{cases} \{+1\} & \text{if} \quad w_i^+(x) > q w_i \\ \{-1\} & \text{if} \quad w_i^+(x) < q w_i \\ \{\pm 1\} & \text{if} \quad w_i^+(x) = q w_i\,, \end{cases}$$

where $q = 1/2 + \alpha/2\beta$ is called *coordination threshold*. Clearly, as in the majority game, the two pure configurations $-\mathbb{1}$ and $+\mathbb{1}$ are Nash equilibria. To investigate the existence of possible other Nash equilibria, it is useful to introduce the following concept: given a subset $S \subseteq \mathcal{V}$ of nodes, we say that S is *r-cohesive* (where $r \in [0,1]$) if for every $i \in S$ the total weight of the out-links towards nodes in S is at least a fraction r of the total weight of node i. Precisely, we have that for every $i \in S$,

$$\frac{\sum_{j \in S} W_{ij}}{w_i} \geq r$$

Clearly, if S is r-cohesive, it is also r'-cohesive for every $r' \leq r$. The following result is an immediate consequence of previous considerations (see Exercise 1).

Proposition 9.1. Consider a binary network coordination game over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ having coordination threshold q. Given $S \subseteq \mathcal{V}$, the configuration $x = \mathbb{1}_S - \mathbb{1}_{\mathcal{V} \setminus S}$ is a Nash equilibrium of the game iff S is q-cohesive and $\mathcal{V} \setminus S$ is (1-q)-cohesive.

Example 9.11. Consider the undirected graph $\mathcal{G}$ obtained by the union of two complete graphs K_3 and K_4 with an extra edge connecting one node in K_3 with one node in K_4. Then , if we denote by S the set of the three nodes in K_3 we have that S is 2/3-cohesive while S^c is 3/4-cohesive. If we consider a coordination game on $\mathcal{G}$ having coordination threshold q, it follows from Proposition 9.1 that the configuration $x = \mathbb{1}_S - \mathbb{1}_{\mathcal{V} \setminus S}$ is a Nash equilibrium if and only if both conditions $q \leq 2/3$ and $1 - q \leq 3/4$ are satisfied, namely if $1/4 < q < 2/3$.

In case instead of a network anti-coordination game, the best response takes the form

$$\mathcal{B}_i(x_i) = \begin{cases} \{+1\} & \text{if} \quad w_i^+(x) < qw_i \\ \{-1\} & \text{if} \quad w_i^+(x) > qw_i \\ \{\pm 1\} & \text{if} \quad w_i^+(x) = qw_i\,, \end{cases}$$

where q is the same quantity defined above. This generalizes the minority game (where $q = 1/2$) introduced above.

Network coordination and anti-coordination games are instances of two classes of network games quite popular in the socio-economic literature. The first one is that of games of *strategic complements* where the relative utility of a player to taking an action versus not is increasing in the set of neighbors who take the action. The second one is that of games of *strategic substitutes* where the relative utility of a player to taking an action versus not is instead decreasing in the set of neighbors who take the action. Such concepts encompass the case of a binary action set and extends to situations where the action set is any ordered set. Typical applications of the games of strategic complements is in the modeling of technology adoptions or social behaviors like fashion, smoking and so on. Instead games of strategic substitutes show up for instance in buying items (like books) that can be shared, information gathering, and local public good provision. An interesting example that is different from the anti-coordination games considered before is the following:

Example 9.12. ['Best shot' public goods game] We fix an undirected graph $\mathcal{G}$ and we consider a binary network game with the following interpretation: action $+1$ means that a certain good has been bought or some information has been acquired with respect to the null state -1. Taking the action 1 is costly and if in the neighborhood of a player there is already someone taking it, that player is better off not taking it. However, if in the neighborhood of a player nobody is taking action $+1$, then for the player it will be convenient to take it. This can be formalized defining the utility of player i as

$$u_i(x_i, x_{-i}) = \begin{cases} 1 - c & \text{if } x_i = +1 \\ 1 & \text{if } x_i = -1,\ x_j = 1 \text{ for some } j \in N_i \\ 0 & \text{if } x_i = -1,\ x_j = -1 \text{ for each } j \in N_i \end{cases} \tag{9.10}$$

We finally present an interesting example of a network game presenting heterogeneity at the level of network interaction.

Example 9.13. [Mixed majority/minority game] In this example we fix an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, we take as action set $\mathcal{A} = \{-1, +1\}$, and we imagine that on certain edges the interaction utility is that of a coordination game while in others of an anti-coordination game. Precisely, suppose there is a function $\xi : \mathcal{E} \to \{+, -\}$ such that $\xi(i, j) = \xi(j, i)$ and assume that

$$\phi^{(i,j)}(a, b) = \begin{cases} \frac{1+ab}{2} & \text{if } \xi(i, j) = + \\ \frac{1-ab}{2} & \text{if } \xi(i, j) = - \end{cases}$$

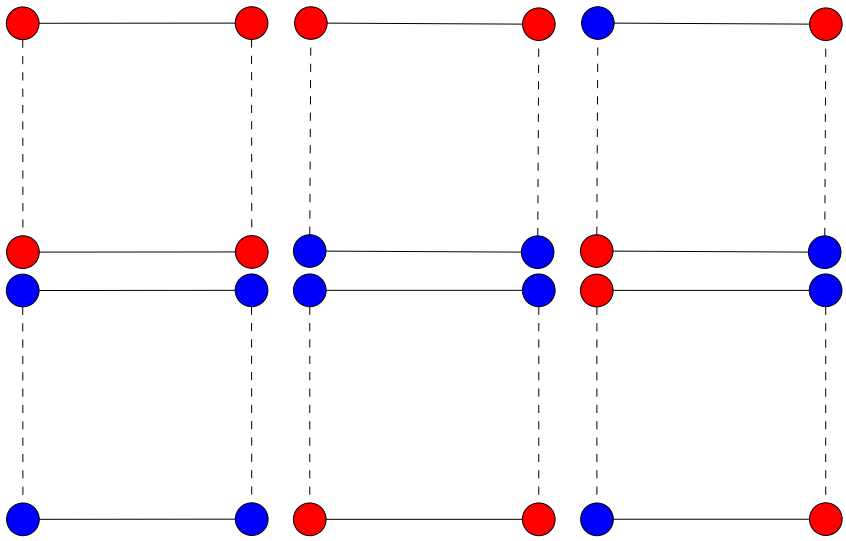


Figure 9.4: The NE on a ring graph with 4 nodes with two anti-coordination edges.

We then consider the network game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ having utilities as in (9.6). They can be written as

$$u_i(x) := \sum_{j:\xi(i,j)=+} W_{ij}\frac{1+x_ix_j}{2} + \sum_{j:\xi(i,j)=-} W_{ij}\frac{1-x_ix_j}{2} \tag{9.11}$$

Fig. 9.4 exhibits all the NE for the cycle graph C_4 having two interaction of coordination type and two (the dashed ones) of anti-coordination type.

9.3.2 Examples of non-binary network games

We now present a couple of examples where $\mathcal{A}$ is no longer binary. In both cases the network game is constructed by a two-player game that is the same for every edge of the graph.

Example 9.14 (Graph coloring)**.** The graph coloring problem concerns assigning to each node of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a color chosen from a finite set $\mathcal{A}$ so that no two neighbors have the same color. When it exists, such assignment is called a *k-coloring* (where $k = |\mathcal{A}|$), and the corresponding graph $\mathcal{G}$ is said to be k-colorable. We can cast this into a network game framework as follows. We fix any distance function d on the set of nodes $\mathcal{V}$ and we take this as the interaction utility function on the edges $\phi(a, b) = d(a, b)$. We obtain the network game where the utility of node i is

$$u_i(x_i, x_{-i}) = \sum_{j\in\mathcal{V}} W_{ij}d(x_i, x_j)\,. \tag{9.12}$$

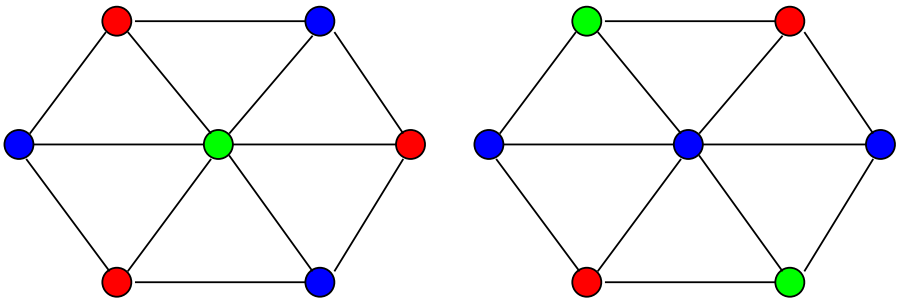


Figure 9.5: Two Nash equilibria for the coloring game with three colors $\mathcal{A} = \{\text{green}, \text{red}, \text{blue}\}$. Observe that the one on the left is a perfect coloring (no two neighbors have the same color), while the one on the right is not.

with W the adjacency matrix $\mathcal{G}$. A special case is when d is the Hamming distance of $\mathcal{A}$ that is:

$$d_H(a,b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

In this case, $u_i(x_i, x_{-i})$ coincide with number of colors conflicts of node i with its neighbors. Notice that when there are only two colors $k = |\mathcal{A}| = 2$, the game is strictly related to the network anti-coordination game as the utility functions of the two games simply differ by a constant (if $d = d_H$) and, as a consequence, the best response functions and Nash equilibria coincide.

For an arbitrary number of colors, it is not hard to verify that every perfect coloring (if any exists) is a Nash equilibrium. On the other hand, perfect colorings are not the only Nash equilibria (see, e.g., Figure 9.5). □

We now present an example of a network game with set of action $\mathcal{A} = \mathbb{R}$.

Example 9.15 (Continuous network coordination game)**.** In this example $\mathcal{A} = \mathbb{R}$ and we consider an interaction utility function on each edge

$$\varphi(a,b) = -(a-b)^2 .$$

For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, we then consider the associated network game with quadratic utility functions

$$u_i(x_i, x_{-i}) = \sum_{j \in \mathcal{V}} W_{ij} \varphi(x_i, x_j) = -\sum_{j \in \mathcal{V}} W_{ij} (x_i - x_j)^2 .$$

Note that, for every player $i \in \mathcal{V}$, $u_i(x_i, x_{-i})$ is concave in x_i and it holds

$$\frac{\partial}{\partial x_i} u_i(x_i, x_{-i}) = 2 \sum_{j \in \mathcal{V}} W_{ij} (x_j - x_i) .$$

This implies that the maximum always exists in x_i and the best-response function is given by

$$b_i(x_{-i}) = \operatorname*{argmin}_{x_i \in \mathbb{R}} u_i(x_i, x_{-i}) = \sum_{j \in \mathcal{V}} \frac{W_{ij}}{w_i} x_j = \sum_{j \in \mathcal{V}} P_{ij} x_j \,,$$

where $w_i = \sum_{j \in \mathcal{V}} W_{ij}$ is the out-degree of node i, and P is the normalized weight matrix of $\mathcal{G}$. Hence, the best-response function for a player i coincides with the weighted average of her out-neighbors' values. Notice now that Nash equilibria are those configurations x satisfying the condition

$$x = Px$$

Therefore, it follows from Proposition 2.7 that the set of Nash equilibria is an $s_{\mathcal{G}}$-dimensional linear subspace of $\mathbb{R}^n$ (where $s_{\mathcal{G}}$ is the number of sinks in the condensation graph of $\mathcal{G}$). In particular, if $\mathcal{G}$ is strongly connected, then the set of Nash equilibria coincides with the one of all consensus vectors.

9.4 Potential games

There is a special class of games that is both relevant in many applications and enjoys special properties. This is the class of *potential games* [28]. This section introduces the definition and some fundamental properties of potential games and discusses some relevant examples.

Definition 9.2. A game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i \in \mathcal{V}})$ is called a *potential game* if there exists a function $\Phi : \mathcal{X} \to \mathbb{R}$ (referred to as the *potential function* of the game) such that for any two configurations $x, y \in \mathcal{X}$, and player $i \in \mathcal{V}$,

$$x_{-i} = y_{-i} \qquad \Longrightarrow \qquad u_i(y) - u_i(x) = \Phi(y) - \Phi(x)\,. \tag{9.13}$$

The game is called an *ordinal potential game* if there exists a function (referred to as *ordinal potential function*) such that for any two configurations $x, y \in \mathcal{X}$, and player $i \in \mathcal{V}$

$$x_{-i} = y_{-i} \qquad \Longrightarrow \qquad \operatorname{sgn}(u_i(y) - u_i(x)) = \operatorname{sgn}(\Phi(y) - \Phi(x))\,. \tag{9.14}$$

Definition 9.2 states that in a potential game, for every action profile x, the utility variation $u_i(y_i, x_{-i}) - u_i(x_i, x_{-i})$ incurred by player i when switching from action x_i to action y_i is the same as the corresponding variation $\Phi(y_i, x_{-i}) - \Phi(x_i, x_{-i})$ in the potential function. For an ordinal potential game, such requirement is weakened to that the utility variation incurred by player i when switching from action x_i to action y_i is only required to have the same sign as the corresponding variation in the ordinal potential function.

(Ordinal) potential games enjoy many important structural properties. In particular, every ordinal potential game admits a (not necessarily unique) Nash

equilibrium, corresponding to a global maximum of $\Phi(x)$ as per the following result.[1]

Proposition 9.2. Consider a finite ordinal potential game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ with potential function $\Phi(x)$. Then, the set $\mathcal{N}$ of its Nash equilibria is non empty and contains the set

$$\mathcal{X}^* := \operatorname*{argmax}_{x\in\mathcal{X}} \Phi(x)$$

of global maximizers of the potential function.

Proof. Let $x^* \in \operatorname{argmax}_{x\in\mathcal{X}} \Phi(x)$ be a global maximizer of the potential function over the configuration space $\mathcal{X}$. Let $i \in \mathcal{V}$ be an arbitrary player, let$a \in \mathcal{A}$ be an arbitrary action, and let $y \in \mathcal{X}$ be the the configuration such that $y_{-i} = x^*_{-i}$ and $y_i = a$. Then, (9.14) implies that $\operatorname{sgn}(u_i(y)-u_i(x^*)) = \operatorname{sgn}(\Phi(y)-\Phi(x^*)) = -1$, so that

$$u_i(a, x^*_{-i}) = u_i(y) \leq u_i(x^*) = u_i(x^*_i, x^*_{-i})\,.$$

From the arbitrariness of the action $a \in \mathcal{A}$, the above implies that $x^*_i \in \mathcal{B}_i(x^*_{-i})$ is a best response. Since the player $i \in \mathcal{V}$ is arbitrary, this proves that x^* is a Nash equilibrium. □

Observe that Proposition 9.2 holds true for ordinal potential games, hence, *a fortiori*, for potential games. Below we gather a few examples of potential games.

Example 9.16. Every congestion game as in Example 9.8 is a potential game [32] with potential

$$\Phi(x) = \sum_{e\in\mathcal{E}} \sum_{l=1}^{(An(x))_e} d_e(l)\,. \tag{9.15}$$

Example 9.17. Every symmetric 2×2 game with payoff matrix as in Figure 9.1 is a potential game with potential Φ defined by $\Phi(-1,-1) = a-c$, $\Phi(+1,+1) = b-d$, and $\Phi(-1,+1) = \Phi(+1,-1) = 0$.

	-1	1
-1	a,a	d,c
1	c,d	b,b

Φ	-1	1
-1	a-c	0
1	0	b-d

This includes, as special cases, coordination and anti-coordination games over undirected graphs.

Discoordination games presented above are not potential games, as they do not admit Nash equilibria. On the other hand, also symmetric two-player games with more than 2 actions per player, are not necessarily potential games. An example in this sense is given by the Rock-Scissor-Paper game presented in Example 9.5 and that also does not admit any Nash equilibrium.

For symmetric games, the existence of a potential can be analyzed in a quite general manner. The following result, simple but a bit subtle, holds.

[1] Note that a potential game could admit multiple pure strategy Nash equilibria corresponding, e.g, to local maxima of the potential function.

Proposition 9.3. Consider a symmetric two-player games with set of actions $\mathcal{A} = \{1, \dots, l\}$ and utility function $\phi \in \mathbb{R}^{l \times l}$. The game is potential if and only there exist $\Phi \in \mathbb{R}^{l \times l}$ symmetric and $R \in \mathbb{R}^{l \times l}$ that is column constant (i.e. $R(a, b) = R(\tilde{a}, b)$ for all $a, \tilde{a}, b \in \mathcal{A}$) such that

$$\phi = \Phi + R \tag{9.16}$$

Moreover, in this case, Φ is a potential for ϕ.

Proof. For a a symmetric two-player game having utility function ϕ, the condition for a matrix $\Phi \in \mathbb{R}^{l \times l}$ to be a potential, are

$$\Phi(a, b) - \Phi(\tilde{a}, b) = \phi(a, b) - \phi(\tilde{a}, b) = \Phi(b, a) - \Phi(b, \tilde{a})$$

for all $a, \tilde{a}, b \in \mathcal{A}$. Indeed, the first equality is saying that variation of the utility of player 1 is equal to the variation of the potential, while the second equality establishes the same thing for player 2. Notice now immediately that, taking $\tilde{a} = b$, we immediately obtain that Φ is necessarily symmetric. Henceforth, if a potential exists, this constitutes a symmetric matrix Φ. Notice that given a symmetric Φ the condition for being a potential is now simply

$$\phi(a, b) - \phi(\tilde{a}, b) = \Phi(a, b) - \Phi(\tilde{a}, b) \tag{9.17}$$

We now come to the core of the proof. If decomposition (9.16) holds for a symmetric Φ and a column constant matrix R, then it can easily be checked that Φ is a potential:

$$\phi(a, b) - \phi(\tilde{a}, b) = \Phi(a, b) + R(a, b) - \Phi(\tilde{a}, b) - R(\tilde{a}, b) = \Phi(a, b) - \Phi(\tilde{a}, b)$$

This implies, given the previous considerations, that Φ is a potential for ϕ.

Conversely, if Φ is a symmetric potential for ϕ, it is sufficient to put $R(a, b) = \phi(a, b) - \Phi(a, b)$ and, using (9.17), we can easily verify that R is indeed column constant. We have thus found a decomposition in the style of (9.16). □

We now investigate on the existence of a potential for network games. The following result shows that if the graph is undirected and the interaction utilities along all edges are those of symmetric two-player potential games, then the corresponding network game is potential.

Proposition 9.4. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be an undirected graph and assume that we have interaction utilities $\varphi^{(i,j)}$ such that for every pair of edges (i, j) and (j, i), it holds $\varphi^{(i,j)}(a, b) = \varphi^{(j,i)}(a, b)$ for every a and b. Assume moreover that the two-player symmetric game having utility $\varphi^{(i,j)} = \varphi^{(j,i)}$ is potential with symmetric potential function $\phi^{\{i,j\}}$ Then, the network game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i \in \mathcal{V}})$ having utilities (9.6) is a potential game with potential

$$\Phi(x) = \frac{1}{2} \sum_{i,j \in \mathcal{V}} W_{ij} \phi^{\{i,j\}}(x_i, x_j). \tag{9.18}$$

Proof. Notice first of all that it is not important to know who has to be considered the first and the second player between i and j in using the potential function $\phi^{\{i,j\}}$ since it is a symmetric function.

Let $x, y \in \mathcal{X}$ be two configurations such that $x_{-i} = y_{-i}$ for some player $i \in \mathcal{V}$. Then, for every $j, k \in \mathcal{V}$, the potentials of the two-player symmetric games satisfy

$$\phi^{\{k,j\}}(y_k, y_j) - \phi^{\{k,j\}}(x_k, x_j) = \begin{cases} 0 & \text{if} \quad j \neq i \text{ and } k \neq i \\ \phi^{\{i,j\}}(y_i, y_j) - \phi^{\{i,j\}}(x_i, x_j) & \text{if} \quad k = i \\ \phi^{\{k,i\}}(y_k, y_i) - \phi^{\{k,i\}}(x_k, x_i) & \text{if} \quad j = i\,. \end{cases} \tag{9.19}$$

Then,

$$\begin{aligned} \Phi(y) - \Phi(x) &= \frac{1}{2} \sum_j \sum_k W_{kj} \left(\phi^{\{k,j\}}(y_k, y_j) - \phi^{\{k,j\}}(x_k, x_j) \right) \\ &= \frac{1}{2} \sum_j W_{ij} \left(\phi^{\{i,j\}}(y_i, y_j) - \phi^{\{i,j\}}(x_i, x_j) \right) + \frac{1}{2} \sum_k W_{ki} \left(\phi^{\{k,i\}}(y_k, y_i) - \phi^{\{k,i\}}(x_k, x_i) \right) \\ &= \frac{1}{2} \sum_j W_{ij} \left(\phi^{\{i,j\}}(y_i, y_j) - \phi^{\{i,j\}}(x_i, x_j) \right) + \frac{1}{2} \sum_k W_{ik} \left(\phi^{\{i,k\}}(y_i, y_k) - \phi^{\{i,k\}}(x_i, x_k) \right) \\ &= \sum_j W_{ij} \left(\phi^{\{i,j\}}(y_i, y_j) - \phi^{\{i,j\}}(x_i, x_j) \right) \\ &= \sum_j W_{ij} \left(\varphi^{(i,j)}(y_i, y_j) - \varphi^{(i,j)}(x_i, x_j) \right) \\ &= u_i(y) - u_i(x)\,, \end{aligned}$$

where the first identity is due to (9.18), the second one is implied by (9.19), the third one follows from the symmetry of the potentials $\phi^{\{i,j\}}$'s and the assumption that the graph is undirected so that $W' = W$, the fifth one by the fact that $\phi^{\{i,j\}}$ is a potential of the two-player symmetric game with utilities $\varphi^{(i,j)}$, and the sixth and final one is due to the definition (9.6) of the network game's utilities. This proves that (9.13) holds true, so that the network game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i \in \mathcal{V}})$ with utilities (9.6) is a potential game with potential function (9.18). □

For network games satisfying the assumption of Proposition 9.4, existence of Nash equilibria is guaranteed by Proposition 9.2. However, in general, an explicit characterization of all the Nash equilibria of such network games is not computationally feasible in complex large-scale networks.

Example 9.9. [Majority game —continued] The majority game on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with utilities given by (9.7) is a potential game with potential function

$$\Phi(x) = \frac{1}{2} \sum_{i,j} W_{ij} \frac{1 + x_i x_j}{2}\,. \tag{9.20}$$

Observe that, if $\mathcal{G}$ is a simple graph, the potential (9.20) coincides with the number of undirected links connecting two players playing the same action. If

$\mathcal{G}$ is connected, then the potential function $\Phi(x)$ has exactly two maximum points, the pure configurations $-\mathbb{1}$ and $+\mathbb{1}$ (see Problem 1 at the end). As discussed earlier on, when $\mathcal{G}$ is not the complete graph, there might be other Nash equilibria.

Example 9.10. [Minority game —continued] The minority game on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ with utilities described in (9.8) is a potential game with potential function

$$\Phi(x) = \frac{1}{2} \sum_{i,j} W_{ij} \frac{1 - x_i x_j}{2} \,. \tag{9.21}$$

In particular, if $\mathcal{G}$ is a simple graph, then the potential (9.21) coincides the number of undirected links connecting nodes with opposite actions.

For the case of the 4-cycle the Nash equilibria are illustrated in Fig. 9.3. The two Nash equilibria in the left-hand side column coincide with the maxima of the potential whose value is equal to 4. Notice that this can be generalized as follows. When the graph $\mathcal{G}$ is bipartite with respect to the partition $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$, the maxima of the potential always coincide with the two configurations that are constant on each subset $\mathcal{V}_1$ and $\mathcal{V}_2$ and take opposite values in the two subsets. In general the network game possess other Nash equilibria.

Example 9.11 (Graph coloring —continued)**.** The network coloring game on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a potential game with potential function

$$\Phi(x) = -\frac{1}{2} \sum_{i,j} W_{ij} d(x_i, x_j) \,. \tag{9.22}$$

In particular, if $\mathcal{G}$ is a simple graph and $d = d_H$ the Hamming distance, then the potential (9.22) coincides with the number of undirected links in $\mathcal{G}$ connecting nodes with the same color. Perfect colorings, when they exist, are clearly maximum points of this potential, hence they are Nash equilibria. More in general, maximum points of the potential (9.22) are given by configurations with minimal constraint violations. Note that, depending on the structure of the graph $\mathcal{G}$, maximum points of (9.22) might not be the only Nash equilibria of the network coloring game.

Example 9.12. [Mixed majority/minority game —continued] Given an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, the game with mixed coordination and anti-coordination interaction utilities as in (9.11) is a potential game with potential function

$$\Phi(x) = \frac{1}{2} \sum_{i,j:\xi(i,j)=+} W_{ij} \frac{1 + x_i x_j}{2} + \frac{1}{2} \sum_{i,j:\xi(i,j)=-} W_{ij} \frac{1 - x_i x_j}{2} \,. \tag{9.23}$$

Being a potential game, it always possess Nash equilibria. For the case of the 4-cycle considered in Fig. 9.4 the maxima of the potential are the two Nash equilibria in the central column.

Example 9.13 (Continuous network coordination game —continued)**.** A continuous network coordination game on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is a potential game with quadratic potential function

$$\Phi(x) = \frac{1}{2} \sum_{i,j} W_{ij}(x_i - x_j)^2 \,. \tag{9.24}$$

9.5 Best response dynamics

In this section, we introduce a first game-theoretic learning process, the *best response dynamics*. We shall focus on the study of the asynchronous best response dynamics, where players in a strategic form game get randomly activated one at a time and switch to a best response action. We will show that, in potential games, the best response dynamics converge to the set of Nash equilibria with probability one in finite time.

Consider a strategic-form game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$. The continuous-time asynchronous *best response* dynamics is a Markov chain $X(t)$ with state space $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$, where every player $i \in \mathcal{V}$ is equipped with an independent rate-1 Poisson clock. When her clock ticks at time t, player i updates her action to some y_i chosen from the action set $\mathcal{A}$ with conditional probability distribution that is uniform over the best response set

$$\mathcal{B}_i(X_{-i}(t)) = \operatorname*{argmax}_{x_i\in\mathcal{A}} \{u_i(x_i, X_{-i}(t))\} \,.$$

In particular, when the best response is unique, player i updates her action to such best response action. Hence, the continuous-time asynchronous best response dynamics is a continuous-time Markov chain $X(t)$ with state space coinciding with the configuration space $\mathcal{X}$ of the game and transition rate matrix Λ as follows: $\Lambda_{xy} = 0$ for every two configurations $x, y \in \mathcal{X}$ that differ in more than one entry, and

$$\Lambda_{xy} = \begin{cases} |\mathcal{B}_i(x_{-i})|^{-1} & \text{if} \quad y_i \in \mathcal{B}_i(x_{-i}) \\ 0 & \text{if} \quad y_i \notin \mathcal{B}_i(x_{-i}) \end{cases} \tag{9.25}$$

for every two configurations $x, y \in \mathcal{X}$ differing in entry i only, i.e., such that $x_i \neq y_i$ and $x_{-i} = y_{-i}$.

If the game is potential, then with probability one the best response dynamics converges to the set of Nash equilibria in finite time, as stated below.

Proposition 9.5. Let $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ be a finite strategic-form ordinal potential game. Let $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$ be the action configuration space and $\mathcal{N} \subseteq \mathcal{X}$ be the set of Nash equilibria. Then, the best response dynamics $X(t)$ is such that, for every distribution of the initial configuration $X(0)$ on $\mathcal{X}$, there exists a random time $T \geq 0$, that is finite with probability one and such that $X(t) \in \mathcal{N}$ for every $t \geq T$.

Proof. Let Λ be the transition rate matrix of the best response dynamics and let $\mathcal{G}_\Lambda$ be the graph with node set $\mathcal{X}$ and weight matrix Λ. We prove the result by showing that any sink connected component in $\mathcal{G}_\Lambda$ is made of NE and then using Proposition 7.1.

Recall that, given two configurations $x \neq y \in \mathcal{X}$, $\Lambda_{xy} > 0$ if and only if there exists some $i \in \mathcal{V}$ such that

$$x_{-i} = y_{-i}\,, \qquad y_i \in \mathcal{B}_i(x_{-i})\,. \tag{9.26}$$

Now, let Φ be the potential of the game. Then, for any such pair of configurations, we have that

$$\operatorname{sgn}(\Phi(y) - \Phi(x)) = \operatorname{sgn}(u_i(y) - u_i(x)) = \operatorname{sgn}(u_i(y_i, x_{-i}) - u_i(x_i, x_{-i})) \geq 0\,,$$

where the second equality and the final inequality follow from (9.26). This implies that on any walk $(x^1, \dots, x^s)$ on $\mathcal{G}_\Lambda$, the potential is also non-decreasing

$$\Phi(x^1) \leq \Phi(x^2) \leq \cdots \leq \Phi(x^s)\,.$$

Consider now a subset $\mathcal{Y} \subseteq \mathcal{X}$ forming a sink connected component in $\mathcal{G}_\Lambda$ and assume by contradiction that $y \in \mathcal{Y}$ is not a NE. Then there must exists $y' \in \mathcal{Y}$ such that $\Lambda_{yy'} > 0$ and $\Phi(y') > \Phi(y)$. Since the potential is always non-decreasing along any walk of $\mathcal{G}_\Lambda$, this says that there can not be a walk from y' to y. This is a contradiction as we were assuming nodes in $\mathcal{Y}$ to form a connected component. □

The proof of the result above actually gives further information on the asymptotics of the BR dynamics. Convergence is insured to a particular subset of NE, that is the one consisting of the nodes of all the sink connected components of the underlying transition graph of the Markov process. In the sequel we denote such subset as $\mathcal{N}_\infty$: it is a trapping set and actually the largest trapping set inside $\mathcal{N}$. We call $\mathcal{N}_\infty$ the set of *recurrent* Nash equilibria. Notice that in those special cases when $\mathcal{N}_\infty$ coincides with $\operatorname{argmax} \Phi(\mathbf{x})$, we actually have that the BR dynamics converges to the subset of Nash equilibria consisting of the maxima of the potential Φ.

We now present some simple examples illustrating the behavior of the best response dynamics. We will refer to the best response dynamics for the majority and minority games as, respectively, the *majority dynamics* and *minority dynamics*.

Example 9.14. Consider the majority dynamics in the cyclic network C_4. There are six Nash equilibria as shown in Figure 9.2. The set of recurrentl Nash equilibria $\mathcal{N}_\infty$ coincides with $\operatorname{argmax} \Phi(\mathbf{x})$, namely consists of the two consensus configurations $\mathcal{N}_\infty = \{\pm \mathbb{1}\}$. It then follows from the proof of Proposition 9.5 that in this case the best response dynamics converges to $\operatorname{argmax} \Phi(\mathbf{x})$ with probability 1.

Example 9.15. Consider the majority dynamics in the barbell network $\mathcal{G}$ consisting of two complete graphs K_n with $n \geq 3$. It is clear that, besides the two

consensus configurations, there are other two Nash equilibria in $\mathcal{N}_\infty$: these are the two configuration presenting opposite signs on the two complete graphs. As a consequence, in this case, the best response dynamics does not necessarily converge to $\operatorname{argmax} \Phi(\mathbf{x})$.

9.6 Noisy best response dynamics

In this section, we introduce the noisy best response dynamics. We then analyze its behavior in potential games, showing that in that case the noisy best response dynamics is a reversible Markov chain whose stationary distribution can be explicitly computed.

As in the previous section, we start with a strategic-form game $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$. The continuous-time asynchronous *noisy best response* dynamics is a Markov chain $X(t)$ with state space $\mathcal{X} = \mathcal{A}^\mathcal{V}$, where every player $i \in \mathcal{V}$ is equipped with an independent rate-1 Poisson clock and if her clock ticks at time t, player i updates her action to some y_i chosen from the action set $\mathcal{A}$ with conditional probability

$$\frac{e^{\eta u_i(y_i, X_{-i}(t))}}{\sum_{a\in\mathcal{A}} e^{\eta u_i(a, X_{-i}(t))}}\,, \tag{9.27}$$

where $\eta > 0$ is a given parameter whose inverse $1/\eta$ is to be thought as a measure of noise. The interpretation of (9.27) is that the probability with which a new action y_i is adopted by player i is increasing with the associated utility $u_i(y_i, X_{-i}(t))$; as $\eta \to 0$, the dependance on this utility vanishes and (9.27) converges to a uniform probability distribution on the action set $\mathcal{A}$; on the other hand, as $\eta \to +\infty$, (9.27) converges to a uniform probability on the best-response set $\mathcal{B}_i(X_{-i}(t))$, so that the noisy best response dynamics reduces to the best response dynamics.

Precisely, the transition rate matrix Λ of the noisy best response dynamics is the following: $\Lambda_{xy} = 0$ if $x, y \in \mathcal{X}$ differ in more than one entry, and

$$\Lambda_{xy} = \frac{e^{\eta u_i(y_i, x_{-i})}}{\sum_{a\in\mathcal{A}} e^{\eta u_i(a, x_{-i})}} \tag{9.28}$$

if x and y differ only in entry i (i.e., if $x_i \neq y_i$ and $x_{-i} = y_{-i}$).

Theorem 9.1. Suppose that $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ is a potential game with potential function Φ. Then, for every $\eta > 0$, the noisy best response dynamics is an irreducible reversible Markov chain with invariant probability distribution

$$\pi_x = \frac{e^{\eta\Phi(x)}}{Z_\eta}\,, \qquad Z_\eta = \sum_{y\in\mathcal{X}} e^{\eta\Phi(y)} \tag{9.29}$$

for every configuration $x \in \mathcal{X}$.

Proof. Irreducibility simply follows from the fact given two action configurations $x, y \in \mathcal{X}$, the MC has a non zero probability to pass from x to y by exchanging one component at a time. Let π be the unique stationary probability distribution of the MC.

Notice now that if x and y differ in exactly one entry i (i.e., if $x_i \neq y_i$ and $x_{-i} = y_{-i}$) then (9.28) and (9.13) imply that

$$\frac{\Lambda_{xy}}{\Lambda_{yx}} = \frac{e^{\eta u_i(y_i, x_{-i})}}{e^{\eta u_i(x_i, x_{-i})}} = e^{\eta(u_i(y_i, x_{-i}) - u_i(x_i, x_{-i}))} = e^{\eta(\Phi(y_i, x_{-i}) - \Phi(x_i, x_{-i}))} = \frac{e^{\eta \Phi(y_i, x_{-i})}}{e^{\eta \Phi(x_i, x_{-i})}} .$$

Hence, we have shown that, for every $x, y \in \mathcal{X}$ that differ in exactly one component, it holds

$$e^{\eta \Phi(x)} \Lambda_{xy} = e^{\eta \Phi(y)} \Lambda_{yx} , \qquad x, y \in \mathcal{X} , \tag{9.30}$$

Clearly, (9.30) also trivially holds if $x = y$ or if x and y differ in more than one component (as in this case $\Lambda_{xy} = \Lambda_{yx} = 0$). This says that the MC is reversible and that its unique invariant probability π must be given by (9.29). □

In the statistical physics jargon, the noisy best response dynamics for an exact potential game is referred to as the *Glauber dynamics*, the stationary probability distribution π defined in (9.29) is referred to as the *Gibbs distribution*, Z_η as the *partition function*, and $1/\eta$ is often interpreted as the 'temperature' of the system. Observe that the Gibbs distribution π defined in (9.29) has the following property: as $\eta \to +\infty$, i.e., in the vanishing noise limit, π converges to a uniform probability over the set of configurations that are the global maximizers of the potential function, i.e., $\operatorname{argmax}\{\Phi(x) : x \in \mathcal{A}^n\}$. Recall that every such global maximizer $x^* \in \operatorname{argmax}\{\Phi(x) : x \in \mathcal{A}^n\}$ of the potential function (or the global maximizer, if it is unique), is necessarily a NE of the game as per Proposition 9.2. On the other hand, as we discussed earlier on, the game might admit additional NE that are not global maximizers of its potential function. For such games, the noisy best response dynamics operates a selection among the NE and concentrates, in the small noise limit, most of the mass of its Gibbs distribution in the subset of those Nash equilibria that correspond to the set of potential maximizers $\operatorname{argmax}\{\Phi(x) : x \in \mathcal{A}^n\}$.

Then, an application of the Ergodic Theorem of Chapter 3 implies that, in long run, the noisy best response Markov chain will have spent almost surely almost all of its time in (or close to for positive noise, i.e., large but finite η) in the set $\operatorname{argmax}\{\Phi(x) : x \in \mathcal{A}^n\}$ of game's potential maximizers. This is at the foundations of the so-called simulated annealing algorithms that are particular Markov Chain Monte Carlo algorithms based on the Glauber dynamics in exact potential games.

9.7 Problems

1. Prove Proposition 9.1.

2. Consider a symmetric two-player games with set of actions $\mathcal{A} = \{1, \ldots, l\}$ and utility function $\phi \in \mathbb{R}^{l \times l}$. Such a game (and the matrix ϕ) is called

passive if $\phi(r,s) = \phi(r',s)$ for every $r, r', s \in \mathcal{A}$, namely if the utility of a player is independent on its own action. Two symmetric two-player games $\phi_1, \phi_2 \in \mathbb{R}^{l \times l}$ are said to be equivalent if there exists a passive game ϕ such that $\phi_1 - \phi_2 = \phi$.

(a) Show that two equivalent symmetric games admit the same set of Nash equilibria;

(b) Determine all Nash equilibria of a passive game.

3. Consider the minority game introduced in Example 9.10 for the case when the underlying graph is the complete graph K_n on n vertices. Given $x \in \mathcal{A}^n$ denote by $n^+(x)$ and $n^-(x)$ the number of components of x equal to $+1$ and -1, respectively.

(a) Prove that $x \in \mathcal{A}^n$ is Nash if and only if $-1 \leq n^+(x) - n^-(x) \leq +1$;

(b) Classify all Nash equilibria, analyzing separately, the two cases when n is even and when it is odd;

(c) Relatively to the previous item, specify which of the Nash equilibria found are strict and which are maxima of the potential.

4. Consider the majority game presented in Example 9.9 for the case when the underlying graph is the cycle graph C_n. Determine all Nash equilibria, specify which are strict and which are maxima of the potential.

5. Show that an undirected unweighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is 2-colorable if and only if it is bipartite.

6. Consider the minority game introduced in Example 9.10 for the case when the underlying graph is the cycle graph C_n.

(a) Determine all Nash equilibria in the cases $n = 3$ and $n = 4$, specifying which are strict and which are maxima of the potential;

(b) Generalize the previous point to a cycle graph C_n with a generic n.

7. Study the Nash equilibria for, respectively, the minority and majority games in the case when the underlying graph is the n-star.

8. Consider the majority game presented in Example 9.9 over a graph $\mathcal{G}$ that is undirected and simple.

(a) Prove that if $\mathcal{G}$ is connected, the maxima of the potential Φ are exactly $-\mathbb{1}$ and $+\mathbb{1}$;

(b) Determine all the maxima of the potential Φ for a general $\mathcal{G}$;

9. Consider the best response dynamics $X(t)$ associated with the Rock-Scissor-Paper Example 9.5.

(a) Write down the graph associated with the transition matrix of $X(t)$;

(b) Describe the asymptotic behavior of $X(t)$.

10. Consider the majority dynamics over the n-cycle C_n. Show that

    (a) the only absorbing states are the pure configurations $-\mathbb{1}$ and $+\mathbb{1}$;

    (b) from any Nash equilibrium of the corresponding coordination game, there is a non zero probability path to one of the pure configurations $-\mathbb{1}$ and $+\mathbb{1}$.

    (c) Conclude that the best response dynamics, in this case, converges to one of the pure configurations $-\mathbb{1}$ and $+\mathbb{1}$ in finite time, with probability 1.

11. Consider the majority dynamics over a graph $\mathcal{G}$ that is undirected and simple. Show that, if the degree of each node is odd, the set of Nash equilibria coincides with the set of absorbing states of the Markov chain.

12. Prove that for the network game in Example 9.12, the configuration $x = \mathbb{1}_S - \mathbb{1}_{S^c}$ is a NE if and only if S is a maximal independent set.

13. Consider a game in strategic form $(\mathcal{V}, \mathcal{A}, \{u_i\}_{i\in\mathcal{V}})$ such that $\mathcal{A}$ is an interval of $\mathbb{R}$ and the utility functions are of class C^1 on $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$. Assume that there exists a C^1 function $\Psi : \mathcal{X} \to \mathbb{R}$ such that for every player i, it holds

$$\frac{\partial u_i}{\partial x_i} = \frac{\partial \Psi}{\partial x_i} \quad \forall x \in \mathcal{X}$$

    (a) Prove that Ψ is a potential for the game in the sense of Definition 9.2.

    (b) Show that the continuous coordination game presented in Example 9.13 is a potential game exhibiting an explicit potential function.

Chapter 10

Random graphs

In this chapter, we provide a brief introduction to some popular models of random graphs.

10.1 The Galton-Watson branching process

The first random graph model that we study is a (potentially infinite) random tree, generated as follows: we start from a root node, then add a random number of neighbors to it, then add a random number of new neighbors to each of its neighbors, and so on. The key assumption we will make is that the number of new neighbors (we will refer to them as *offsprings*) of every node is an independent random variable with a given probability distribution. This random graph process is called the (Galton-Watson) *branching process*. It is illustrated in Figure 10.1.

Let $\{p_k\}_{k\geq 0}$ be a probability distribution over the nonnegative integers, and let

$$\mu := \sum_{k\geq 0} k p_k$$

be its expected value. Consider a double-indexed sequence of independent and identically distributed random variables $\{\xi_i^t\}_{i\geq 1, t\geq 0}$ with distribution $\{p_k\}$, i.e., such that

$$\mathbb{P}(\xi_i^t = k) = p_k\,, \qquad i \geq 1\,,\ t \geq 0\,,\ k \geq 0\,.$$

The branching process is then defined by the recursion

$$X_0 = 1\,, \qquad X_{t+1} = \sum_{i=1}^{X_t} \xi_i^t\,, \qquad t \geq 0\,.$$

The interpretation is that one starts with one individual ($X_0 = 1$) and the process continues with every individual giving birth to an independent and identically distributed number of individuals, each with distribution $\{p_k\}$. Then, t indexes the generation and X_t counts the number of individuals in the t-th

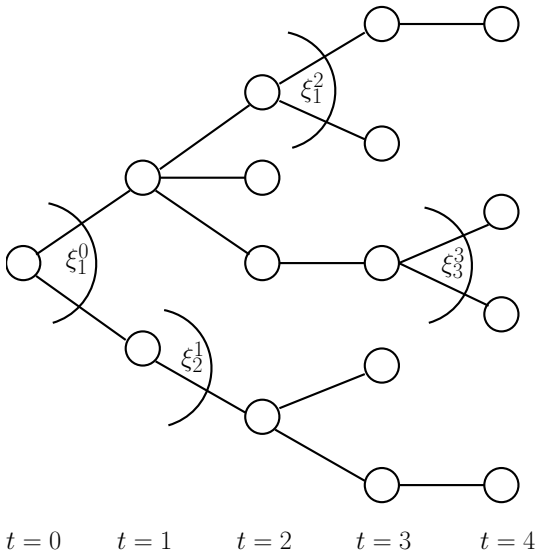


Figure 10.1: The branching process. The i-th member of the t-th generation has an independent and identically distributed number of offsprings ξ_i^t.

generation, while the index i runs over the different individuals in the same generation and ξ_i^t stands for the number of offsprings of the i-th indexed individual in the t-th generation.

We denote by

$$T_{\text{ext}} := \inf\{t \geq 0 : \, X_t = 0\}$$

the (possibly infinite) extinction time, and let

$$\theta_{\text{ext}} := \mathbb{P}(T_{\text{ext}} < +\infty)$$

be the extinction probability. In plain words, the extinction time T_{ext} is the smallest time t such that there are no offsprings of generation t ($X_t = 0$), so that the population is extinct, and θ_{ext} is the probability extinction occurs in finite time. Let

$$\Phi(\theta) := \sum_{k \geq 0} p_k \theta^k \,, \qquad \theta \in [0,1]$$

[1] be the moment generating function of the offspring distribution $\{p_k\}$. Observe that $\Phi(\theta)$ maps the interval $[0,1]$ in itself. See Figure 10.2. The following result characterizes θ_{ext} as the smallest fixed point of $\Phi(\theta)$ in $[0,1]$.

Lemma 10.1. The extinction probability θ_{ext} of the Galton-Watson branching process with offspring distribution $\{p_k\}$ satisfies

$$\theta_{\text{ext}} = \min\{\theta \in [0,1] : \Phi(\theta) = \theta\} \,,$$

[1] Throughout, we always adopt the convention that $0^0 = 1$ and, equivalently, $0 \log 0 = 0$.

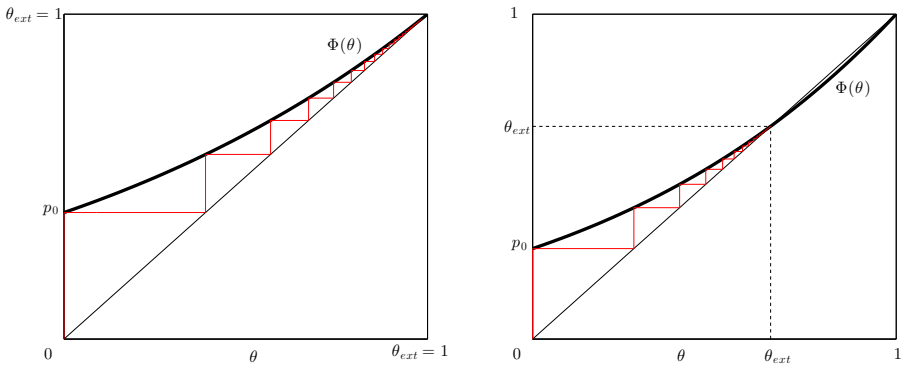


Figure 10.2: Moment generating function $\Phi(\theta)$ for a subcritical (a) and a supercritical (b) branching processes. In both figures, the chosen offspring distribution is the binomial one $p_k = \binom{n}{k}\gamma^k(1-\gamma)^{n-k}$ for $0 \leq k \leq n$, and $p_k = 0$ for $k > n$. The parameter n is chosen equal to 10 in both figures, while $\gamma = 0.09$ in the left figure, so that the expected number of offsprings $\mu = 0.9 < 1$, and $\gamma = 0.12$, so that $\mu = 1.2 > 1$ in the right figure. The iterations $\theta(k+1) = \Phi(\theta(k))$ started at $\theta(0) = 0$ are plotted as well, in red. As predicted by Theorem 10.1, the extinction probability satisfies $\theta_{ext} = 1$ in the subcritical case (left), and $\theta_{ext} < 1$ in the supercritical case (right).

where $\Phi(\theta)$ is the moment generating function of $\{p_k\}$.

Proof. For $t \geq 0$, let $\theta_t := \mathbb{P}(T_{\text{ext}} \leq t)$. Since the events $\{T_{\text{ext}} \leq t\}$ form a nondecreasing sequence, in the sense that

$$\{T_{\text{ext}} \leq t\} \subseteq \{T_{\text{ext}} \leq t+1\}\,, \qquad t \geq 0\,,$$

one has that the sequence $\{\theta_t\}_{t\geq 0}$ is nondecreasing and convergent to

$$\lim_{t\to+\infty} \theta_t = \lim_{t\to+\infty} \mathbb{P}(T_{\text{ext}} \leq t) = \mathbb{P}(T_{\text{ext}} < +\infty) = \theta_{\text{ext}}\,.$$

On the other hand, using the independence and identical distribution property of $\{\xi_i^t\}$, one finds that

$$\begin{aligned}
\theta_{t+1} &= \mathbb{P}(T_{\text{ext}} \leq t+1) \\
&= \sum_{k\geq 0} \mathbb{P}(T_{\text{ext}} \leq t+1, \xi_1^0 = k) \\
&= \sum_{k\geq 0} \mathbb{P}(T_{\text{ext}} \leq t+1 | \xi_1^0 = k)\mathbb{P}(\xi_1^0 = k) \\
&= p_0 + \sum_{k\geq 1} p_k \mathbb{P}(T_{\text{ext}} \leq t)^k \\
&= \sum_{k\geq 0} p_k \theta_t^k\,,
\end{aligned}$$

i.e.,

$$\theta_{t+1} = \Phi(\theta_t)\,, \qquad t \geq 0\,. \tag{10.1}$$

Moreover, $\theta_0 = 0$, since the process is started with $X_0 = 1$. Therefore, θ_{ext} coincides with the limit of the recursion (10.1) with initial condition $\theta_0 = 0$. Since the sequence $\{\theta_t\}$ is nondecreasing, and $\Phi(\theta)$ is continuous, this limit must coincide with the smallest fixed point of $\Phi(\theta)$ in $[0,1]$. □

Now, let us make the following observations on the moment generating function. First, $\Phi(\theta)$ is non-decreasing, $\Phi(0) = p_0$, and $\Phi(1) = 1$. Moreover, $\Phi(\theta)$ is convex and $\Phi'(1) = \mu$. These properties, together with Lemma 10.1, allow one to prove the following fundamental result relating positivity of the survival probability of a branching process to the expected number of offsprings μ, as illustrated in Figure 10.3.

Theorem 10.1. Consider the Galton-Watson branching process with offspring distribution $\{p_k\}$. Let $\mu = \sum_{k\geq 0} k p_k$ be the expected number of offsprings of each individual. Then,

subcritical regime if $\mu < 1$, then $\theta_{\text{ext}} = 1$;

critical regime if $\mu = 1$ and $p_1 < 1$, then $\theta_{\text{ext}} = 1$; if $p_1 = 1$, then $\theta_{\text{ext}} = 0$;

supercritical regime if $\mu > 1$, then $\theta_{\text{ext}} < 1$.

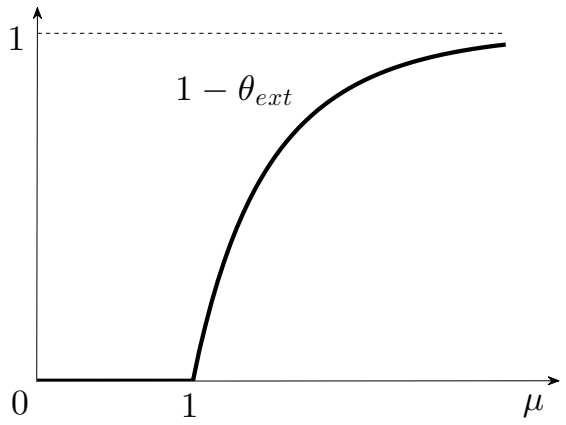


Figure 10.3: Survival probability $1-\theta_{ext}$ as a function of the expected number of offspring's μ for a branching process with binomial offspring distribution of parameters $n = 10$ and $\gamma \in [0, 0.3]$ (cf. Example 10.2). Note the first-order phase transition at the critical value $\mu = 1$: first order refers to the fact that it is the first derivative of the displayed to have a discontinuity.

Proof. Lemma 10.1 characterizes θ_{ext} as the smallest fixed point of the moment generating function $\Phi(\theta)$ in $[0, 1]$, i.e., the smallest zero of $\Psi(\theta) := \Phi(\theta) - \theta$. From the observations we made, we have that the function $\Psi(\theta)$ is convex on the interval $[0, 1]$. Moreover, $\Psi(0) = p_0$, $\Psi(1) = 0$ and $\Psi'(1) = \mu - 1$.

It follows that, if $\mu < 1$, then $\Psi(\theta)$ is strictly decreasing in $[0, 1]$ (since $\Psi'(\theta) \leq \Psi'(1) = \mu - 1$), so it is strictly positive for $0 \leq \theta < 1$, and then its only zero is $\theta_{\text{ext}} = 1$. This proves the subcritical case.

In the critical regime, if $\mu = 1$ and $p_1 < 1$, then necessarily $p_0 > 0$ and $p_k > 0$ for at least one $k \geq 2$. Then, $\Psi(\theta)$ is strictly convex, so that $\Phi'(\theta) < \Phi'(1) = 1$ for all $\theta \in [0, 1)$. Hence, $\Psi(\theta)$ is strictly decreasing in $[0, 1]$ and one can argue as above. On the other hand, if $p_1 = 1$, then $\Psi(\theta) = 0$ for all $\theta \in [0, 1]$, so that $\theta_{\text{ext}} = 0$.

Finally, in the supercritical regime, one has that $\Psi'(1) = \mu - 1 > 0$. Since $\Psi(1) = 0$, this implies that $\Psi(\theta) < 0$ in a left neighborhood of 1. On the other hand, $\Phi(0) = p_0 \geq 0$, so there must exist some $\theta^* \in [0, 1)$ such that $\Phi(\theta^*) = 0$. The smallest such θ^* coincides then with the extinction probability θ_{ext}. □

We now present a few examples of branching processes with explicit offspring distribution.

Example 10.1 (Branching process with Poisson offspring distribution)**.** Consider a Poisson distribution with parameter $\lambda > 0$, i.e.,

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k \geq 0\,.$$

Then, the expected value is $\mu = \sum_{k\geq 0} kp_k = \lambda$, while the moment generating function is

$$\Phi(\theta) = \sum_{k\geq 0} p_k \theta^k = e^{-\lambda} \sum_{k\geq 0} \frac{(\theta\lambda)^k}{k!} = e^{(\theta-1)\lambda}\,.$$

Theorem 10.1 then implies that the extinction probability $\theta_{\text{ext}} = 1$ for $\lambda \leq 1$, so that the survival probability is $1 - \theta_{\text{ext}} = 0$, whereas, for $\lambda > 1$, the survival probability $1 - \theta_{\text{ext}}$ coincides with the unique solution x in $(0, 1)$ of the equation

$$x = e^{\lambda(x-1)}\,.$$

Example 10.2 (Branching process with binomial offspring distribution)**.** Consider a binomial distribution with parameters $n \geq 1$ and $\gamma \in [0, 1]$, i.e.,

$$p_k = \binom{n}{k} \gamma^k (1-\gamma)^{n-k}\,, \quad 0 \leq k \leq n\,, \qquad p_k = 0\,, \quad k > n\,.$$

Then, the expected value is

$$\mu = \sum_{k\geq 0} kp_k = \gamma n\,,$$

while the moment generating function is

$$\begin{aligned}
\Phi(\theta) &= \sum_{k\geq 0} p_k \theta^k \\
&= \sum_{0\leq k\leq n} \binom{n}{k} (\theta\gamma)^k (1-\gamma)^{n-k} \\
&= (1-\gamma)^n \sum_{0\leq k\leq n} \binom{n}{k} \left(\frac{\theta\gamma}{1-\gamma}\right)^k \\
&= (1-\gamma)^n \left(1 + \frac{\theta\gamma}{1-\gamma}\right)^n \\
&= (1 - \gamma + \theta\gamma)^n \\
&= (1 - \gamma(1-\theta))^n\,.
\end{aligned}$$

Theorem 10.1 then implies that, when $n\gamma \leq 1$, the extinction probability $\theta_{\text{ext}} = 1$, so that the survival probability is $1 - \theta_{\text{ext}} = 0$, whereas, when $n\gamma > 1$, the survival probability $1 - \theta_{\text{ext}}$ coincides with the unique solution x in $(0, 1)$ of the equation

$$(1 - \gamma x)^n = 1 - x\,.$$

See Figure 10.3 for a plot of the survival probability $1 - \theta_{\text{ext}}$ as a function of $\mu = n\gamma$.

Example 10.3 (Branching process and percolation)**.** Let us start with a branching process with some offspring distribution $\{p_k\}_{k\geq 0}$ of expected value $\mu = \sum_{k\geq 0} kp_k$. Then, for some given $\gamma \in [0, 1]$, we consider the sub-tree obtained by

keeping the root node and removing the other nodes independently at random with probability $(1-\gamma)$ and keeping them with probability γ: such random removal process is known as *percolation*. The result is another branching process (the percolated branching process) with offspring distribution

$$p_k^{(\gamma)} = \sum_{l \geq k} p_l \binom{l}{k} \gamma^k (1-\gamma)^{l-k}\,, \qquad k \geq 0\,.$$

Similarly, if we had removed links instead of nodes independently with probability $1-\gamma$ we would have obtained the same result for the connected component to which the root belongs.

The percolated branching process obtained by removing nodes independently with probability $(1-\gamma)$ from a branching process of degree distribution $\{p_k\}_{k\geq 0}$ has expected number of offsprings

$$\mu^{(\gamma)} = \sum_{k\geq 0} k \sum_{l\geq k} p_l \binom{l}{k} \gamma^k (1-\gamma)^{l-k} = \sum_{l\geq 0} p_l \sum_{k\leq l} k \binom{l}{k} \gamma^k (1-\gamma)^{l-k} = \sum_{l\geq 0} p_l l\gamma = \mu\gamma\,,$$

so that the necessary and sufficient condition for its survival is $\mu\gamma > 1$.

E.g., if the original branching process is simply a deterministic n-ary tree for some $n \geq 1$ (i.e., if $p_n = 1$ and $p_k = 0$ for $k \neq n$), then the percolated branching process is a branching process with binomial distribution as in Example 10.2, and hence has expected number of offsprings $n\gamma$. More in general, if the original branching process has binomial offspring distribution with parameters α and n, then, the offspring distribution of the percolated branching process is given by

$$\begin{aligned}
p_k^{(\gamma)} &= \sum_{k\leq l\leq n} \binom{n}{l} \alpha^l (1-\alpha)^{n-l} \binom{l}{k} \gamma^k (1-\gamma)^{l-k} \\
&= \binom{n}{k} (\alpha\gamma)^k (1-\alpha)^{n-k} \sum_{0\leq h\leq n-k} \binom{n-k}{h} \left(\frac{\alpha(1-\gamma)}{1-\alpha}\right)^h \\
&= \binom{n}{k} (\alpha\gamma)^k (1-\alpha)^{n-k} \left(1 + \frac{\alpha(1-\gamma)}{1-\alpha}\right)^{n-k} \\
&= \binom{n}{k} (\alpha\gamma)^k (1-\alpha\gamma)^{n-k}\,,
\end{aligned}$$

i.e., its binomial of parameters n and $\alpha\gamma$. Similarly, if the original branching process has Poisson offspring distribution with parameter λ, then the offspring distribution of the percolated process is given by

$$p_k^{(\gamma)} = \sum_{l\geq k} e^{-\lambda} \frac{\lambda^l}{l!} \binom{l}{k} \gamma^k (1-\gamma)^{l-k} = e^{-\lambda} \frac{(\lambda\gamma)^k}{k!} \sum_{h\geq 0} \frac{(\lambda(1-\gamma))^h}{h!} = e^{-\lambda\gamma} \frac{(\lambda\gamma)^k}{k!}\,,$$

i.e., it is a Poisson distribution with parameter $\lambda\gamma$.

Much more on the Galton-Watson branching process can be found in textbooks. See, e.g., [14, Sect.2.1].

10.2 The Erdös-Rényi random graph

The Erdös-Rényi (ER) random graph is a model of random graph depending on two parameters: the number of nodes n, and the probability p that a given pair of nodes is a link. Precisely, for a positive integer n and some $p \in [0,1]$, the ER random graph $\mathcal{G}(n,p)$ is a simple graph with node set $\mathcal{V} = \{1,\dots,n\}$ and where every unordered pair of nodes $\{i,j\}$ is a link independently with probability p. Due to its elementary definition, the ER random graph is amenable to a deep mathematical analysis and many results are available concerning, e.g., its degree distribution, connectedness, diameter, etc. We will quickly summarize a few of the main such results.

We start by discussing the degree distribution. First note that the expected number of undirected links is a binomial random variable with parameters $\binom{n}{2}$ and p, so that the average degree $\overline{w}$ has expectation

$$\mathbb{E}[\overline{w}] = \frac{2}{n}\binom{n}{2}p = (n-1)p\,.$$

In fact, the degree w_i of any given node $i \in \mathcal{V}$ is itself a random variable with binomial distribution with parameters $n-1$ and p, i.e.,

$$\mathbb{P}(w_i = k) = \binom{n-1}{k}p^k(1-p)^{n-1-k}\,, \qquad 0 \le k \le n\,. \tag{10.2}$$

Moreover, the node degrees are weakly correlated: if $i \neq j \in \mathcal{V}$, then the only coupling between w_i and w_j is through the possible presence of the link $\{i,j\}$. In particular, the following result holds true.

Lemma 10.2. Consider the ER random graph $\mathcal{G}(n,p)$ for some $n \ge 2$ and $p \in [0,1]$. Then, for any two distinct nodes $i \neq j \in \mathcal{V}$,

$$\mathbb{P}(w_i = 0, w_j = 0) = (1-p)^{2n-3} = \frac{\mathbb{P}(w_i = 0)\mathbb{P}(w_j = 0)}{1-p}\,, \tag{10.3}$$

and

$$\mathbb{P}(w_i = k, w_j = k) = \binom{n-2}{k-1}^2 p^{2k-1}(1-p)^{2n-2k-2} + \binom{n-2}{k}^2 p^{2k}(1-p)^{2n-2k-3}\,, \tag{10.4}$$

for all $k \ge 1$.

Proof. The first identity in (10.3) follows from the fact that $w_i = w_j = 0$ if and only if $\{i,j\}$ is not a link and $\{i,h\}$ and $\{j,h\}$ are not links for every node $h \in \mathcal{V}\setminus\{i,j\}$: this is the intersection of $2n-3$ independent events each occurring with probability $(1-p)$. Equation (10.4) follows from conditioning on the events $E = \{\{i,j\} \text{ is a link}\}$ and $F = \{\{i,j\} \text{ is not a link}\}$. □

We now focus on connectedness of the ER graph $\mathcal{G}(n,p)$. First, observe that, from (10.2), the probability that a given node i is isolated is $\mathbb{P}(w_i = 0) =$

$(1-p)^{n-1}$, hence the number of isolated nodes N_0 has expected value

$$\mathbb{E}[N_0] = \mathbb{E}\left[\sum\nolimits_i \chi_i\right] = \sum_{i\in\mathcal{V}} \mathbb{P}(w_i = 0) = n(1-p)^{n-1}\,, \tag{10.5}$$

where χ_i is the indicator function of the event $\{w_i = 0\}$, i.e., $\chi_i = 1$ if $w_i = 0$ and $\chi_i = 0$ if $w_i > 0$. On the other hand, the fact that $\chi_i^2 = \chi_i$ and the weak correlation property (10.3) imply that the variance of the number of isolated nodes satisfies

$$\begin{aligned}
\mathrm{Var}(N_0) &= \mathbb{E}\left[\left(\textstyle\sum_i \chi_i\right)^2\right] - \mathbb{E}\left[\textstyle\sum_i \chi_i\right]^2 \\
&= \mathbb{E}\left[\textstyle\sum_i \chi_i\right] + \mathbb{E}\left[\textstyle\sum_i \sum_{j\neq i} \chi_i\chi_j\right] - \mathbb{E}\left[\textstyle\sum_i \chi_i\right]^2 \\
&= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} - n^2(1-p)^{2n-2} \\
&= n\left((1-p)^{n-1} - (1-p)^{2n-3}\right) + n^2p(1-p)^{2n-3} \\
&= \frac{\mathbb{E}[N_0]^2}{n(1-p)}\left(\frac{1}{(1-p)^n} - 1 + np\right).
\end{aligned} \tag{10.6}$$

Then, if we choose $p = \frac{a}{n}\log n$ for some constant $a > 0$, then $p \to 0$ and $(1-p)^{1/p} \to e^{-1}$ as n grows large, so that

$$n^a(1-p)^n = n^a\,(1-p)^{\frac{a\log n}{p}} \overset{n\to\infty}{\longrightarrow} 1\,.$$

From (10.5) and the above, we get that, for $p = \frac{a}{n}\log n$,

$$\mathbb{E}[N_0] = n(1-p)^{n-1} = \frac{n}{1-p}\,(1-p)^{\frac{a}{n}\log n} \asymp ne^{-a\log n} = n^{1-a}\,, \tag{10.7}$$

where the symbol $\asymp$ stands for asymptotic equivalence, so that (10.7) means that the ratio $\mathbb{E}[N_0]/n^{1-a}$ converges to 1 as n grows large. Hence, if $a > 1$, then the expected number of isolated nodes vanishes as n grows large, whereas, if $a < 1$, then the expected number of isolated nodes grows unbounded as n grows large. In fact, we can prove the following result for the existence with high probability of isolated nodes in the ER graph.

Proposition 10.1. [Phase transition for existence of isolated nodes in Erdös-Rényi] For $a > 0$, consider the Erdös-Rényi random graph $\mathcal{G}(n,p)$ with $p = \frac{a}{n}\log n$. Then, with high probability as n grows large

(i) if $a > 1$, $\mathcal{G}(n,p)$ does not contain any isolated node;

(ii) if $a < 1$ the number of isolated nodes in $\mathcal{G}(n,p)$ satisfies $N_0 \asymp n^{1-a} \to +\infty$.

Proof. (i) Using Chebyshev's inequality, (10.6) and (10.5), one gets that, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{N_0}{\mathbb{E}[N_0]} - 1\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(N_0)}{\varepsilon^2\mathbb{E}[N_0]^2} = \frac{1}{n(1-p)\varepsilon^2}\left(\frac{1}{(1-p)^n} - 1 + np\right) \asymp \frac{n^{a-1}}{\varepsilon^2} \overset{n\to\infty}{\longrightarrow} 0\,.$$

The above implies that, for $0 < a < 1$, the number of isolated nodes grows as $N_0 \asymp \mathbb{E}[N_0] \asymp n^{1-a}$ with high probability as n grows large.

(ii) It follows from Markov's inequality and (10.7) that

$$\mathbb{P}(N_0 \geq 1) \leq \mathbb{E}[N_0] \asymp n^{1-a} \overset{n\to\infty}{\longrightarrow} 0\,,$$

so that with high probability as n grows large there are no isolated nodes. □

Proposition 10.1 is a first example of *phase transition* in the ER random graph: in $\mathcal{G}(n, \frac{a}{n}\log n)$, with high probability as n grows large, existence of isolated nodes or the lack thereof is determined by whether $a > 1$ or $a < 1$, respectively. Now, it is obvious that a graph containing isolated nodes is not connected, while the converse does not necessarily hold true. However, the following result shows that the phase transition for connectivity is the same as the one for the existence of isolated nodes.

Theorem 10.2 (Phase transition for connectivity of Erdös-Rényi)**.** For $a > 0$, consider the Erdös-Rényi random graph $\mathcal{G}(n,p)$ with $p = \dfrac{a}{n}\log n$. Then, with high probability as n grows large

(i) if $a < 1$, then $\mathcal{G}(n,p)$ is discconnected;

(ii) if $a > 1$, then $\mathcal{G}(n,p)$ is connected and its diameter satisfies

$$\operatorname{diam}(\mathcal{G}(n,p)) \leq A \log n\,,$$

for some positive constant A independent of n.

The result above shows that, in order to be connected with high probability, the ER random graph must have an average degree of the order of $\log n$ at least.

A different, and in some respect more relevant, regime is when $p = \lambda/n$ for some constant $\lambda > 0$. In this regime, the average degree $\mathbb{E}[w_i] = (n-1)p$ of any node i converges to the constant λ as n grows large and in fact it is not hard to show that the total number of undirected links $M = \sum_i w_i$ in $\mathcal{G}(n,p)$, which has binomial distribution of parameters $\binom{n}{2}$ and p, has asymptotic behavior $M \asymp n\lambda/2$ with high probability as n grows large. In fact, Lemma 10.2 on the weak correlation among the node degrees has implications on the asymptotic degree distribution of $\mathcal{G}(n,p)$. In fact, in such $p = \lambda/n$ regime, the degree distribution of the ER random graph $\mathcal{G}(n,\lambda/n)$ can be proven to converge to a Poisson distribution with parameter λ with high probability as stated in the following result.[2]

Proposition 10.2. Consider the ER random graph $\mathcal{G}(n,p)$ with $p = \lambda/n$ for some constant $\lambda > 0$ independent from n. Then, degree distribution satisfies

$$p_k := \frac{1}{n}\,|\{i \in \mathcal{V} :\, w_i = k\}| \xrightarrow{\mathbb{P}} e^{-\lambda}\frac{\lambda^k}{k!}\,, \qquad k \geq 0\,, \tag{10.8}$$

as n grows large.

[2]We use the notation $\xrightarrow{\mathbb{P}}$ for convergence in probability, i.e., $a_n \xrightarrow{\mathbb{P}} b$ means that $\lim \mathbb{P}(|a_n - b| > \varepsilon) = 0$ for every $\varepsilon > 0$.

Proof. We start with the case $k = 0$. In this case, (10.5) yields

$$\mathbb{E}[p_0] = \frac{\mathbb{E}[N_0]}{n} = (1-p)^{n-1} = \left(1 - \frac{\lambda}{n}\right)^{n-1} \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\,. \tag{10.9}$$

On the other hand, Chebyshev's inequality and (10.6) give

$$\mathbb{P}\left(\left|\frac{p_0}{\mathbb{E}[p_0]} - 1\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(p_0)}{\varepsilon^2 \mathbb{E}[p_0]^2} = \frac{\mathrm{Var}(N_0)}{\varepsilon^2 \mathbb{E}[N_0]^2} = \frac{\frac{1}{(1-p)^n} - 1 + np}{\varepsilon^2 n(1-p)} \asymp \frac{e^{\lambda} - 1 + \lambda}{\varepsilon^2 \lambda n} \overset{n\to\infty}{\longrightarrow} 0\,.$$

The above and (10.9) imply that $p_0 \xrightarrow{\mathbb{P}} e^{-\lambda}$ as n grows large.

For $k \geq 1$, (10.2) implies that

$$\begin{aligned}
\mathbb{E}[p_k] &= \frac{\mathbb{E}[N_k]}{n} \\
&= \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&= \frac{(n-1)(n-2)\dots(n-k)\lambda^k}{k! \cdot n^k (1-p)^{k+1}} \left(1 - \frac{\lambda}{n}\right)^n \quad \overset{n\to\infty}{\longrightarrow} \quad \frac{\lambda^k}{k!} e^{-\lambda}\,.
\end{aligned}$$

On the other hand, using (10.4) and arguing as in (10.6), one can compute the variance $\mathrm{Var}[N_k]$ of the number of degree-k nodes, show that the ratio $\mathrm{Var}[N_k]/\mathbb{E}[N_k]^2$ vanishes as n grows large, and finally use Chebyshev's inequality as before to prove that $p_k/\mathbb{E}[p_k] \xrightarrow{\mathbb{P}} 1$ and conclude that (10.8) holds true for all $k \geq 1$. □

Observe that Theorem 10.2 implies that, with high probability as n grows large, the ER random graph $\mathcal{G}(n,p)$ with $p = \lambda/n$ is disconnected irrespective of the value of $\lambda > 0$. In fact, another phase transition takes place in this regime, affecting the size of the largest connected component. In order to get an intuition, we consider the following exploration process of the neighborhood of a random node in $\mathcal{G}(n, \lambda/n)$. We start in a given node i, then explore the set of nodes at distance 1 from i, then the set of nodes at distance 2, and so on, until we have explored the whole connected component of $\mathcal{G}(n, \lambda/n)$ that node i belongs to. We will denote the size of such connected component by C_i. This clearly a random variable whose stochasticity depends on the realization of the graph. We will approximate this exploration process on $\mathcal{G}(n, \lambda/n)$, and consequently C_i, by a branching process with Poisson offspring distribution $p_k = e^{-\lambda}\lambda^k/k!$, for $k \geq 0$. The rationale for this approximation is that on the one hand Proposition 10.2 states that the degree distribution of the ER random graph $\mathcal{G}(n, \lambda/n)$ converges to a Poisson distribution with expected value λ as the network size n grows large. On the other hand, the neighborhood of a random node in the ER random graph can be shown to be tree-like up to some distance proportional to $\log n$ with high probability as n grows large. We can then easily use the branching process theory of Section 10.1 (see in particular Example 10.1) to determine that a necessary and sufficient conditions for the branching process with Poisson offspring distribution to have positive survival probability is that

its expected value $\lambda > 1$. Moreover, when $\lambda > 1$, the extinction probability θ_{ext} coincides with the unique solution $x \in (0,1)$ of the equation $x = e^{\lambda(x-1)}$. While this is just a statement on the Poisson branching process approximation of the exploration process of a random node in $\mathcal{G}(n, \lambda/n)$, it turns put to provide the correct prediction for the phase transition of the giant connected component of the actual ER random graph, which is summarized in the following result.

Theorem 10.3 (Phase transition for the giant connected component in Erdös-Rényi)**.** For $\lambda > 0$, and $n \geq 1$, let $p = \lambda/n$. Then, with high probability as n grows large

(i) if $\lambda < 1$, then there exists a constant $A > 0$ independent of n such that the size of every connected component in the Erdös-Rényi random graph $\mathcal{G}(n,p)$ satisfies
$$C_i \leq A \log n\,, \qquad 1 \leq i \leq n\,;$$

(ii) if $\lambda > 1$, the size $C_{\max}$ of the largest connected component of the Erdös-Rényi random graph $\mathcal{G}(n,p)$ satisfies
$$\frac{1}{n} C_{\max} \stackrel{n\to\infty}{\longrightarrow} 1 - x\,,$$
where x is the unique solution in $(0,1)$ of the equation $x = e^{\lambda(x-1)}$.

Part (i) of Theorem 10.3 states that, for $\lambda < 1$, the largest connected component of the ER random graph $\mathcal{G}(n, \lambda/n)$ has size at most logarithmic in n: in other words, $\mathcal{G}(n, \lambda/n)$ is divided in many small connected components. On the other hand, part (ii) of Theorem 10.3 states that, for $\lambda > 1$, a non-trivial fraction x of the nodes in the ER random graph $\mathcal{G}(n, \lambda/n)$ belongs to the same connected component. Observe that, as we know from Theorem 10.2, the ER random graph is not connected in this regime, and in fact it can be shown that the second largest connected component is of size at most logarithmic in n. Thus, when $\lambda > 1$, we should picture $\mathcal{G}(n, \lambda/n)$ as containing one giant component (of size linear in n) and many small components (of size logarithmic in n).

We do not provide a proof of Theorem 10.3, that can be found in many textbooks including [7, 14, 13]. However, we conclude this section by discussing how the exploration process of the neighborhood of a given node in the ER random graph $\mathcal{G}(n,p)$ (the actual exploration process, not its branching process approximation) can be seen as the outbreak of an epidemics. Consider a homogeneous population of n nodes, each of which can be in 3 possible states: S = susceptible; I = infected; and R = recovered. Then consider a discrete-time Markov chain $X(t)$ on $\{S, I, R\}^n$, whose i-th entry $X_i(t)$ describes the state of the i-th individual of the population at time $t = 0, 1, \ldots$. Assume that an infected individual always becomes recovered at the next step and that recovered individuals remain so (i.e., $X_i(t+1) = R$ if $X_i(t) = I$ or $X_i(t) = R$). On the other hand, susceptible individuals can get infected by independent probability-γ contagion from infected individuals: precisely, if $X_i(t) = S$, then $X_i(t+1) = S$ unless at least one j such $X_j(t) = I$ succeeds in passing him the infection. I.e.,

if $X_i(t) = S$, then $X_i(t+1) = S$ with conditional probability $(1-\gamma)^{I_t}$ and $X_i(t+1) = I$ with conditional probability $1-(1-\gamma)^{I_t}$, where I_t stands for the number of infected individuals at time t. This epidemics model is known as the Reed-Frost SIR epidemics. It is not hard to verify that (do it as an exercise) the absorbing states of this Markov chain are all and only those configurations $x \in \{S, I, R\}^n$ such that $x_i \in \{S, R\}$ for all i, i.e., all the configurations where there are no infected individuals.

What does the Reed-Frost SIR epidemic have to do with the exploration of neighborhood of a random node i in the ER random graph $\mathcal{G}(n,p)$? If $\gamma = p$ and one starts the epidemics with $X_i(0) = I$, and $X_j(0) = S$ all other individuals j, then for all $t \geq 0$ one can identify the infected individuals (i.e., those j such that $X_j(t) = I$) with the nodes of $\mathcal{G}(n,p)$ at distance t from node i. The question of the size C_i of the connected component of node i then gets translated into the question of what the number of ultimately recovered individuals in in the Reed-Frost SIR epidemic.

10.3 The configuration model

Let $\{p_k\}_{k\geq 0}$ be a probability distribution over the nonnegative integers and let

$$\mu := \sum_{k\geq 0} k p_k$$

be its first moment. Throughout, we will assume that μ is finite. For every positive integer n such that np_k is integer and $n\mu$ is even, we can define the configuration model with degree distribution $\{p_k\}$ as follows. For all $k \geq 1$, draw $p_k n$ nodes each with k half-links stemming from it, so that there are in total n nodes and μn half-links. Then, match the μn half-links uniformly at random. The resulting hyper-graph may contain self-loops and parallel links. However, the number of such self-loops and parallel links is a constant that does not depend on n (see the exercise). In fact, one can prove that in the limit as n grows large, the probability distributions of the number of self-loops and of the number of parallel links converge to Poisson distributions with expected values $\nu/2$ and ,respectively, $\nu^2/4$, respectively, where

$$\nu = \mu^{-1} \sum_{k\geq 0} p_k k^2 - 1$$

(see, e.g., [14, Chapter 3]).

More in general, one can prove that, if one selects a node at random with uniform probability, its neighborhood looks like a tree (i.e., no cycles are encountered) with high probability up to a distance proportinal to $\log n$. More precisely, one can prove that (at least under the additional assumption that the second moment of the degree distribution $\sum_k k^2 p_k$ remains finite as n grows large) there exists a constant $\alpha > 0$ such that the probability that a random node belongs to a cycle of length less than $\alpha \log n$ converges to 0 as n grows

large. Observe that that does not mean that there are no short cycles with high probability! In fact, we already observed that the number of self-loops and parallel links converges in distribution to a Poisson random variable which is non-zero with positive probability, and the same turns out to be true for cycles of any fixed length $k = 3, 4, \ldots$. What happens however is that encountering any of this short cycles when selecting a node at random is an event whose probability vanishes as n grows large.

In fact, one can prove that the neighborhood of a random node in the configuration model can be approximated, with high probability, by a *two-phase branching process* where the first generation has offspring distribution coinciding with the (*node-perspective*) degree distribution $\{p_k\}$, and all the other generations have offspring distribution coinciding with the *link-perspective* degree distribution $\{q_k\}_{k\geq 0}$, defined by

$$q_k = \frac{(k+1)p_{k+1}}{\mu}\,, \qquad k \geq 0\,. \tag{10.10}$$

The intuition behind the formula above is that, when selecting a link uniformly at random, and choosing one of its two end nodes with equal probability $1/2$, the probability that the node found has degree k is proportional to kp_k because nkp_k of the μn half-links are stemming out from nodes of degree k. The first moment of the link-perspective distribution $\{q_k\}$ is given by

$$\sum_{k\geq 0} kq_k = \frac{\sum_{k\geq 0} k(k+1)p_{k+1}}{\mu} = \frac{\sum_{k\geq 0} p_k k^2}{\mu} - 1 = \nu\,, \tag{10.11}$$

i.e., it coincides with the ratio between the second and first moment of $\{p_k\}$ minus 1.

Example 10.4 (Link-perspective degree distribution of random regular graph)**.** Consider the degree distribution of a j-regular graph, i.e.,

$$p_k = \begin{cases} 1 & \text{if} \quad k = j \\ 0 & \text{if} \quad k \neq j\,. \end{cases}$$

Then, the link-oriented degree distribution is given by

$$q_k = \begin{cases} 1 & \text{if} \quad k = j-1 \\ 0 & \text{if} \quad k \neq j-1\,. \end{cases}$$

Clearly, the first moment μ of $\{p_k\}$ and the first moment ν of $\{q_k\}$ satisfy

$$\mu = j = \nu + 1\,.$$

Example 10.5 (Link-perspective degree distribution of Poisson distribution)**.** For $\lambda > 0$, let

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}\,, \qquad k \geq 0$$

be a Poisson distribution with expected value $\mu = \lambda$. Then, the associated link-perspective degree distribution satisfies

$$q_k = \frac{(k+1)p_{k+1}}{\mu} = e^{-\lambda}\frac{(k+1)\lambda^{k+1}}{\lambda(k+1)!} = e^{-\lambda}\frac{\lambda^k}{k!} = p_k\,, \qquad k \geq 0\,.$$

I.e., the node-and link-perspective degree distributions coincide for Poisson distributions, and, obviously, so do their expectations $\mu = \nu = \lambda$.

We can now adapt the arguments of Section 10.1 to determine the phase transition for the extinction probability θ_{ext} of the two-phase branching process were the first node has offspring distribution $\{p_k\}_{k\geq 0}$ coinciding with the node-perspective degree distribution of the configuration model and with all subsequent generations having offspring distribution $\{q_k\}_{k\geq 0}$ coinciding with the link-perspective degree distribution defined by (10.10). Observe that extinction of the two-phase branching process is equivalent to extinction of all the standard branching processes originated from the ξ_1^0 nodes in generation 1. Since the subpopulations originated by these nodes are independent single-phase branching processes with offspring distribution $\{q_k\}$, we obtain that the necessary and sufficient condition for survival is

$$\nu > 1\,, \qquad \text{or} \qquad q_1 = 1\,. \tag{10.12}$$

More specifically, the extinction probability θ_{ext} of the two-phase branching process satisfies

$$\theta_{\text{ext}} = \Phi(\overline{\theta}_{\text{ext}})\,, \qquad \overline{\theta}_{\text{ext}} = \min\{\theta \in [0,1] : \overline{\Phi}(\theta) = \theta\}$$

where

$$\Phi(\theta) = \sum_{k\geq 0} p_k\theta^k\,, \qquad \overline{\Phi}(\theta) = \sum_{k\geq 0} q_k\theta^k$$

are the moment generating functions of $\{p_k\}$ and $\{q_k\}$, respectively. Observe that, by virtue of (10.11), condition (10.12) is equivalent to $\sum_{k\geq 0} p_k k^2 > 2\mu$ or $p_2 = 1$, i.e.,

$$\sum_{k\geq 0} p_k k(k-2) > 0\,, \qquad \text{or} \qquad p_2 = 1\,. \tag{10.13}$$

It is easy to verify that (10.13) is always verified when $p_0 = p_1 = 0$, since in this case the summation in the leftmost condition of (10.13) involves only nonnegative terms and at least a positive one if $p_2 < 1$.

While the argument above concerns the two-stage branching process, it turns out to give the correct prediction for the existence of a giant connected component in the configuration model itself, as for the following result.

Theorem 10.4 (Existence of a giant component in the configuration model)**.** Let $\{p_k\}_{k\geq 0}$ be a probability distribution with finite first moment μ and such that $p_2 < 1$. Consider the configuration model random graph with size n and prescribed degree distribution $\{p_k\}$. Then, with high probability as n grows large

(i) if $\sum_{k\geq 0} p_k k(k-2) \leq 0$, then the size of largest connect component is sub-linear in n;

(ii) if $\sum_{k\geq 0} p_k k(k-2) > 0$, then the largest connect component has size linear in n.

We also have the following result, stating that the configuration model is connected with high probability as the size n grows large provided that the minimum degree is not smaller than 3.

Theorem 10.5 (Connectivity of the configuration model)**.** Let $\{p_k\}_{k\geq 0}$ be a probability distribution with finite first moment μ and such that $p_0 = p_1 = p_2 = 0$. Then, with high probability as n grows large, the configuration model random graph with size n and prescribed degree distribution $\{p_k\}$ is connected.

10.4 The preferential attachment model

The preferential attachment is a model of growing random network. The idea is the following: we start at time $t = 1$ with an initial graph $\mathcal{G}_1$, then at every time $t \geq 2$, create a new graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$ by adding a new node to $\mathcal{G}_{t-1}$ and connecting it to some of the existing nodes $\mathcal{V}_{t-1}$ of $\mathcal{G}_{t-1}$ chosen according to some rule. For convenience, we will typically start with an initial graph $\mathcal{G}_1$ comprising just two nodes with $c \geq 1$ parallel links, and label nodes in such a way that $\mathcal{V}_t = \{0, 1, \ldots, t\}$ for $t \geq 1$. However, other choices are possible as well.

It remains to be specified the rule prescribing which nodes in $\mathcal{V}_{t-1}$ the new node t chooses to connect to. One possibility would be to select a certain fixed number $c \geq 1$ of nodes uniformly at random among the set $\mathcal{V}_{t_1}$ and add a link from node t to each of these nodes. We may call this a *uniform attachment* model. In contrast, a *preferential attachment* model tends to favor the choice of those nodes in $\mathcal{V}_{t-1}$ which have already a higher degree in $\mathcal{G}_{t-1}$. In particular, the mechanism first proposed by Price [34, 29] prescribes that the new node t chooses the nodes to attach to from $\mathcal{V}_{t-1}$ with probability proportional to a constant $a > -1$ plus their degree in $\mathcal{G}_{t-1}$, which we denote by $w_i(t-1)$. In the special case considered by Barabasi and Albert [3] (that coined the term preferential attachment) the constant $a = 0$, so that existing nodes in $\mathcal{V}_{t-1}$ are chosen with probability proportional to their degree in $\mathcal{G}_{t-1}$.

Preferential attachment random graphs turn out to exhibit power law degree distributions. In particular, for the Barabasi and Albert model, where $a = 0$ and every new node is added at time t with degree $w_t(t) = c$, it can be shown that the fraction of degree-k nodes in $\mathcal{G}_t$, which we denote by $p_k(t)$, converges

with probability one to

$$\lim_{t\to+\infty} p_k(t) = p_k = \begin{cases} 0 & \text{if} \quad k < c \\ \dfrac{2}{2+c} & \text{if} \quad k = c \\ \dfrac{2c(c+1)}{k(k+1)(k+2)} & \text{if} \quad k \geq c\,. \end{cases} \tag{10.14}$$

Hence, for the Barabasi and Albert preferential attachment model the limit degree distribution satisfies

$$p_k \asymp k^{-3}\,, \qquad \text{as } k \to \infty\,,$$

i.e., it behaves like a *power law* with decay exponent $\gamma = 3$. If one considers the more general preferential attachment mechanism of Price, where the new node joining at time t is attached to c nodes of the existing graph $\mathcal{G}_{t-1}$ chosen with probability proportional to $a + w_i(t-1)$, where $a > -1$, then one can show that the asymptotic degree distribution $p_k = \lim_{t\to\infty} p_k(t)$ satisfies

$$p_k \asymp k^{-(3+a)}\,, \qquad \text{as } k \to \infty\,,$$

i.e., it behaves like a power law with decay exponent $3 + a$.

We conclude this section by observing that, while we have described the preferential attachment model as generating an undirected graph, the Price mechanism with $a > 0$ can be easily used to generate directed graphs (as in fact, it was meant to in its original formulation), by simply letting the new node at time t join the existing graph $\mathcal{G}_{t-1}$ by linking to c of the existing nodes chosen with probability proportional to $a + w_i^-(t-1)$, where $w_i^-(t-1)$ is the in-degree of node i in $\mathcal{G}_{t-1}$.[3] In this way, one obtains a random directed graph whose asymptotic in-degree distribution is a power law. In fact, Price proposed this as a model for a citation network, where nodes represent articles and links citations. A further generalization consists in letting the out-degree of the new node (i.e., the number of articles cited by a newly published one) be an independent random variable with given probability distribution, instead of constant c. This generalization allows one to account for arbitrary out-degree distributions.

10.5 Small-world models

Small-world random graph models were introduced to account for the empirical observation that many real-world networks have small diameter and high clustering. Typical small-world models combine links due to geographic proximity with a few long-distance connections.

[3] Observe that we need $a > 0$ instead of simply $a > -1$ in this case, because new nodes join with 0 in-degree and we need to assign them a positive —though small— probability of being linked form the nodes joining later on.

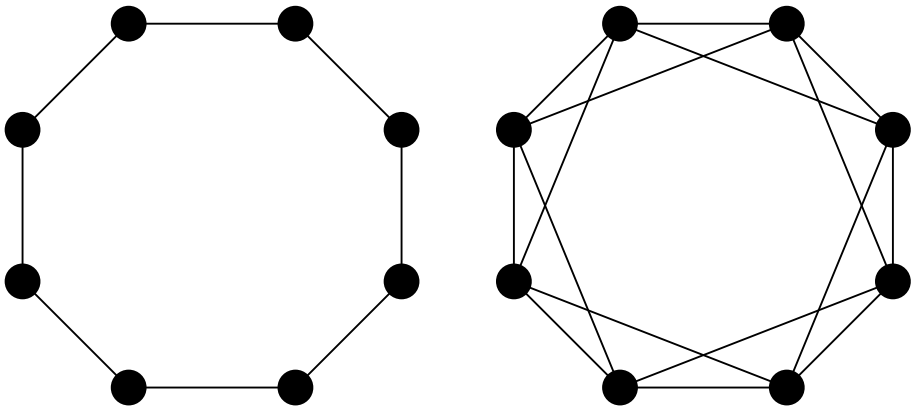


Figure 10.4: Left: a ring graph with $n = 8$ nodes. A symmetric k-regular undirected graph with node set $\mathcal{V} = \{1, \dots, n\}$ where every node is directly connected to the $k = 4$ nodes whose index is the closest to their own modulo $n = 8$.

The basic example we consider is obtained starting with a simple graph with node set $\mathcal{V} = \{1, \dots, n\}$, where every node is directly connected to the k nodes whose index is the closest to their own modulo n, where k is a positive even number such that $k \leq n - 1$. So, e.g., node 1 is connected to nodes $2, 3, \dots, k/2 + 1$ and $n, n - 1, \dots, n - k/2 + 1$, and so on. For $k = 2$, this is simply a ring graph as in the left plot in Figure 10.4, for larger even k it looks like the one reported in the right plot of Figure 10.4. Observe that, for $k \geq 4$, this graph contains a large number of triangles, hence has high clustering coefficient.

We then add to the existing graph l additional undirected links, where l is a binomial random variable with parameters $nk/2$ and $p \in [0, 1]$, and the new links connect pairs of nodes chosen independently and uniformly at random. It can be shown that, with high probability as n grows large, the degree distribution is a shifted Poisson

$$p_j = 0\,, \qquad 0 \leq j < k\,, \qquad p_j = e^{-kp} \frac{(kp)^{j-k}}{(j-k)!}\,, \qquad j \geq k\,,$$

the clustering coefficient converges to

$$C = \frac{3(k-2)}{4(k-1) + 8kp + 4kp^2}\,.$$

and that the diameter satisfies

$$\mathrm{diam}(\mathcal{G}) \leq A \log n$$

for some constant A independent of n.

Appendix A

Linear Algebra

Appendix B

Fixed-Point Theorems

Appendix C

Dynamical Systems

Bibliography

[1] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, *Opinion fluctuations and disagreement in social networks*, Mathematics of Operation Research **38** (2013), no. 1, 1–27.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows: Theory, algorithms, and applications*, Prentice Hall, 1993.

[3] A.L. Barabasi and R. Albert, *Emergence of scaling in random networks*, Science **268** (1999), 509–512.

[4] D. Bertsekas, *Network optimization: Continuous and discrete models*, Athena Scientific, 1998.

[5] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont, MA, 1997.

[6] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis, *Convergence in multiagent coordination, consensus, and flocking*, Proceedings of IEEE CDC, 2005.

[7] B. Bollobás, *Modern graph theory*, Springer, 1998.

[8] P. Bonacich, *Power and centrality: a family of measures*, American Journal of Sociology **5** (1987), 1170–1182.

[9] D. Braess, *Über ein Paradoxen aus der Verkehrsplanung*, Unternehmensforschung **12** (1968), 258–268.

[10] S. Brin and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Proceedings of the 7th International World Wide Web Conference, 1998, pp. 107–117.

[11] G. Como, E. Lovisari, and K. Savla, *Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing*, IEEE Transactions on Control of Network Systems **2** (2015), no. 1, 57–67.

[12] M.H. DeGroot, *Reaching a consensus*, Journal of the American Statistical Association **69** (1974), no. 345, 118–121.

[13] M. Draief and Laurent Massoulié, *Epidemics and rumors in complex networks*, Cambridge University Press, 2010.

[14] R. Durrett, *Random graph dynamics*, Cambridge University Press, 2006.

[15] P. Elias, A. Feinsten, and C. Shannon, *A note on the maximum flow through a network*, IRE Transactions on Information Theory **2** (1956), no. 4, 117–119.

[16] L.R. Ford and D.R. Fulkerson, *Maximal flow through a network*, Canadian Journal of Mathematics **8** (1956), 399–404.

[17] J. R. P. French, *A formal theory of social power*, Psychological Review **63** (1956), 181=194.

[18] G. Frobenius, *Über matrizen aus positiven elementen, 1*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (1908), 471–476.

[19] ———, *Über matrizen aus positiven elementen, 2*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (1909), 514–518.

[20] ———, *Ueber matrizen aus nicht negativen elementen*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (1912), 456–477.

[21] C. Godsil and G. Royle, *Algebraic graph theory*, Springer, 2001.

[22] B. Golub and M.O. Jackson, *Naive learning in social networks and the wisdom of crowds*, American Economic Journal: Microeconomics **2** (2010), no. 1, 112–149.

[23] F. Harari, *A criterion for unanimity in french's theory of social power*, vol. Studies in Social Power, Institute for Social Research, 1959.

[24] H. Haveliwala, *Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search*, IEEE Transactions on Knowledge Data Engineering **15** (2003), no. 4, 784–796.

[25] L. Katz, *A new status index derived from sociometric analysis*, Psychometrika **18** (1953), no. 1, 39–43.

[26] D.A. Levin, Y. Peres, and E.L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2010.

[27] E. Lieberman, C. Hauert, and M.A. Nowak, *Evolutionary dynamics on graphs*, Nature **433** (2005), 312–316.

[28] D. Monderer and L.S. Shapley, *Potential games*, Games and Econoic Behavior **14** (1996), 124–143.

[29] M.E.J. Newman, *The structure and function of complex networks*, SIAM Review **45** (2003), no. 2, 167–256.

[30] J.F. Padgett and C.K. Ansell, *Robust action and the rise of the Medici, 1400-1434*, American Journal of Sociology **6** (1993), 1259–1319.

[31] O. Perron, *Zur theorie der matrices*, Mathematische Annalen **64** (1907), no. 2, 248–263.

[32] R.W. Rosenthal, *A class of games possessing pure-strategy Nash equilibria*, International Journal of Game Theory **2** (1973), no. 1, 65–67.

[33] T. Roughgarden, *Selfish routing and the price of anarchy*, MIT Press, 2005.

[34] H.A. Simon, *On a class of skew distribution functions*, Biometrika **42** (1955), 425–440.

[35] J. Surowiecki, *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*, Doubleday, 2004.

[36] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984.

[37] J. G. Wardrop, *Some theoretical aspects of road traffic research*, ICE Proceedings: Engineering Divisions **1** (1952), no. 3, 325–362.

[38] P. Whittle, *Networks: Optimisation and evolution*, Cambridge University Press, 2007.