

**Figure 6.27** Two distinct rooted phylogenetic networks  $N_1$  and  $N_2$  that represent different sets of hardwired clusters, softwired clusters, and rooted phylogenetic trees.

### 6.13 Tree containment

The problem of determining whether a given rooted phylogenetic network  $N$  contains a given rooted phylogenetic tree  $T$ , on  $\mathcal{X}$ , is NP-complete [141].

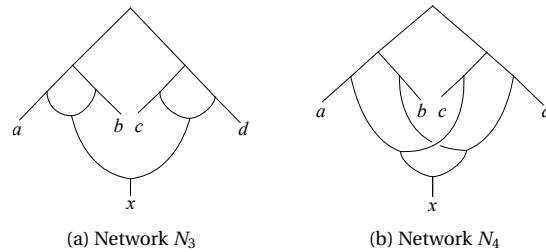
### 6.14 Comparing rooted networks

Suppose we are given two different phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$ , as shown in Figure 6.27, say. How to measure their similarity? For phylogenetic trees, this question can be answered in a number of ways. For example, we could use the Robinson-Foulds distance to compare two unrooted phylogenetic trees, or the cluster distance to compare two rooted phylogenetic trees (Section 3.16). Because splits and clusters correspond one-to-one to the edges of an associated unrooted or rooted phylogenetic tree, respectively, comparing the splits or clusters represented by two trees is equivalent to comparing the trees themselves.

For rooted phylogenetic networks, the situation is different. The association between a set of clusters or trees and the phylogenetic network used to represent the data is quite loose. For example, consider the two networks  $N_3$  and  $N_4$  in Figure 6.28. These two networks are topologically distinct, but, in a sense, *indistinguishable*, because they represent the same set of hardwired clusters, the same set of softwired clusters and the same set of rooted phylogenetic trees. So, comparing the clusters or trees represented by two rooted phylogenetic networks is not the same as comparing the networks themselves.

While biologists are not very interested in distinguishing between different networks that are indistinguishable in the above sense, this question has nevertheless obtained some attention in the computational literature and we report on some of the results in this section.

Below, we first discuss a number of distance functions that aim at comparing the represented data and then look at some measures that are based on topological invariants of the networks. With the exception of the subnetwork distance, none of them provide a proper metric for the set of rooted phylogenetic networks on  $\mathcal{X}$ , as they all return a distance of zero for the two distinct networks  $N_3$  and  $N_4$  displayed in Figure 6.28, and thus fail to fulfill the identity property of Definition 3.12.1.



**Figure 6.28** Two topologically distinct rooted phylogenetic networks  $N_3$  and  $N_4$  that represent the same hardwired clusters, softwired clusters, and rooted phylogenetic trees.

**Exercise 6.14.1** (Two indistinguishable networks) *Show that the two rooted phylogenetic networks  $N_3$  and  $N_4$  displayed in Figure 6.28 represent the same set of hardwired clusters, the same set of softwired clusters, and the same set of rooted phylogenetic trees.*

There is another issue to consider. Rooted phylogenetic networks are usually computed from an input set of incompatible clusters or an input set of incongruent rooted phylogenetic trees, on  $\mathcal{X}$ . The size of such an input dataset is usually polynomial in the number of taxa. However, a rooted phylogenetic network  $N$  that is computed to represent such a dataset will, in general, represent an exponential number of softwired clusters or rooted phylogenetic trees. This is due to the fact that each reticulation added to the network can potentially double the number of represented softwired clusters or trees. So, a direct comparison of the softwired clusters or rooted phylogenetic trees represented by two different networks will sometimes be infeasible, due to the large number of objects, and might be misleading, if most of the clusters or trees compared are not objects that the network is intended to represent.

#### 6.14.1 Comparing hardwired networks

Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ , interpreted in the hardwired sense. A natural way to compare two such networks is to compute their *hardwired cluster distance*, a simple extension of the cluster distance for rooted phylogenetic trees (see Section 3.16), defined as follows:

**Definition 6.14.2** (Hardwired cluster distance) *Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ . We define the hardwired cluster distance  $d_{\text{hard}}$  between  $N_1$  and  $N_2$  as the cardinality of the symmetric difference of the two sets of hardwired clusters represented by the two networks, divided by two:*

$$d_{\text{hard}}(N_1, N_2) = \frac{|\mathcal{C}_{\text{hard}}(N_1) \Delta \mathcal{C}_{\text{hard}}(N_2)|}{2}.$$

The hardwired distance between any two rooted phylogenetic networks on  $\mathcal{X}$  is easy

to compute, because the number of hardwired clusters represented by a rooted phylogenetic network equals the number of tree edges in the network. For example, the hardwired distance between the two rooted phylogenetic networks  $N_1$  and  $N_2$  displayed in Figure 6.27 equals one, because there is exactly one cluster,  $\{c, e\}$ , that is only represented by  $N_1$ , and there is exactly one cluster,  $\{d, e\}$ , that is only represented by  $N_2$ .

The two rooted phylogenetic networks  $N_3$  and  $N_4$  shown in Figure 6.28 are topologically distinct, however they both represent exactly the same set of hardwired clusters. So, although the hardwired cluster distance is a proper metric for sets of clusters (represented by hardwired networks), it is not a proper metric on the networks themselves.

### 6.14.2 Comparing softwired networks

Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ , interpreted in the softwired sense. A natural way to compare two such networks is to compute their *softwired cluster distance*, which is another simple extension of the cluster distance for rooted phylogenetic trees (see Section 3.16), defined as follows:

**Definition 6.14.3** (Softwired cluster distance) *Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ . We define the softwired cluster distance  $d_{soft}$  between  $N_1$  and  $N_2$  as the cardinality of the symmetric difference of the two sets of softwired clusters represented by the two networks, divided by two:*

$$d_{soft}(N_1, N_2) = \frac{|\mathcal{C}_{soft}(N_1) \Delta \mathcal{C}_{soft}(N_2)|}{2}.$$

For example, the softwired cluster distance between the two rooted phylogenetic networks  $N_1$  and  $N_2$  displayed in Figure 6.27 is  $\frac{5}{2}$ , because there are exactly four softwired clusters that are only contained in  $N_1$ , namely  $\{a, b\}$ ,  $\{b, d\}$ ,  $\{c, e\}$ , and  $\{a, b, d\}$ , and there is exactly one that is only contained in  $N_2$ , namely  $\{d, e\}$ .

A rooted phylogenetic network  $N$  on  $\mathcal{X}$  can represent an exponential number of different softwired clusters, in the worst case, making this distance impractical to use in general. However, in Section 6.8 we saw that the number of softwired clusters depends exponentially on the maximal number of reticulations that span any given tree edge in a rooted phylogenetic network. While this number might be quite large in the context of the population genetics of sexually reproducing organisms, say, in phylogenetic applications this number should often be fairly small.

Moreover, for some topologically restricted classes of rooted phylogenetic networks, such as galled trees and level- $k$  networks for a fixed level  $k$ , the number of softwired clusters is polynomial and so for such networks, the softwired cluster distance can always be computed efficiently.

The two rooted phylogenetic networks  $N_3$  and  $N_4$  shown in Figure 6.28 both represent the same set of softwired clusters, but are topologically distinct. Thus, while the softwired cluster distance is a proper metric on sets of clusters (represented by softwired networks), it is not a proper metric for the networks themselves.

One way to avoid the computational difficulties of the softwired cluster distance is to use the hardwired cluster distance, even when comparing rooted phylogenetic networks that are being interpreted in the softwired sense. This seems to be a reasonable heuristic,

because every cluster that is represented by a rooted phylogenetic network in the hard-wired sense, is also represented by that network in the softwired sense. However, it is unclear how well the two different distance measures are correlated.

#### The tripartition distance

The *tripartition distance* was introduced in an attempt to provide a metric for softwired networks that is easy to compute [175].

To describe this approach, we need to introduce some additional simple definitions. Let  $N$  be a rooted phylogenetic network on  $\mathcal{X}$ . Recall that an ancestor  $u$  of a node  $v$  in  $N$  is called a *stable ancestor*, if every path from the root to  $v$  passes through  $u$ . If this is not the case, then we refer to  $u$  as an *unstable ancestor*. By symmetry, we call  $v$  a *stable descendant* of  $u$ , if  $u$  is a stable ancestor of  $v$ , and we call  $v$  a *unstable descendant* of  $u$ , if  $u$  is an unstable ancestor of  $v$ .

**Definition 6.14.4** (Tripartitions of a network) *Let  $N$  be a rooted phylogenetic network on  $\mathcal{X}$ . The tripartition associated with a tree edge  $e = (v, w)$  in  $N$  is defined as*

$$\theta(e) = (A(e), B(e), C(e)),$$

where

- (i)  $A(e) = \{x \in \mathcal{X} \mid x \text{ labels a stable descendant of } w\}$ ,
- (ii)  $B(e) = \{x \in \mathcal{X} \mid x \text{ labels an unstable descendant of } w\}$ , and
- (iii)  $C(e) = \{x \in \mathcal{X} \mid x \text{ labels a node that is not a descendant of } w\}$ .

We use  $\Theta(N)$  to denote the (multi-)set of tripartitions obtained from all internal tree edges in  $N$ .

The set of all tripartitions of a rooted network is easy to compute, and so the following distance based on comparing the tripartitions can be easily calculated:

**Definition 6.14.5** (Tripartition distance) *The tripartition distance between two rooted phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$  is defined as:*

$$d_{\text{tripart}}(N_1, N_2) = \frac{|\Theta(N_1) \Delta \Theta(N_2)|}{2}.$$

As an example for this distance function, consider the two rooted phylogenetic networks  $N_1$  and  $N_2$  shown in Figure 6.27. The five internal tree edges of  $N_1$  give rise to the following five tripartitions:

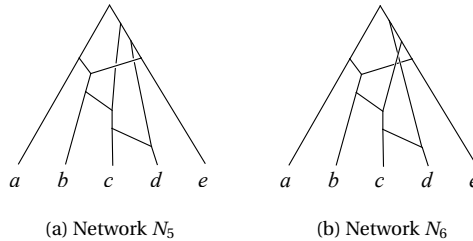
$$\begin{aligned} &(\{a, b, d\}, \{c\}, \{e\}), (\{a\}, \{b, c\}, \{d, e\}), (\{d\}, \{b, c\}, \{a, e\}), \\ &(\{e\}, \{c\}, \{a, b, d\}), (\{b\}, \{c\}, \{a, d, e\}). \end{aligned}$$

The five internal tree edges of  $N_2$  give rise to the following five tripartitions:

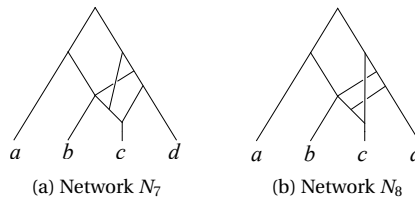
$$\begin{aligned} &(\{a, b, c\}, \{d\}, \{e\}), (\{a\}, \{b, c\}, \{d, e\}), (\emptyset, \{b, c, d\}, \{a, e\}), \\ &(\{e\}, \{d\}, \{a, b, c\}), (\{b, c\}, \emptyset, \{a, d\}). \end{aligned}$$

As the two networks have only one tripartition in common, the tripartition distance is  $d_{\text{tripart}}(N_1, N_2) = \frac{8}{2} = 4$ .

On rooted phylogenetic trees, the tripartition distance equals the cluster distance. In



**Figure 6.29** Two distinct rooted phylogenetic networks  $N_5$  and  $N_6$  that represent different softwired clusters, but have tripartition distance zero.



**Figure 6.30** Two distinct rooted phylogenetic networks  $N_7$  and  $N_8$  that represent the same hardwired clusters, softwired clusters, and rooted phylogenetic trees, but have a non-zero tripartition distance.

[175] it was claimed that the tripartition distance is a proper metric on rooted phylogenetic networks. However, this assertion is now known to be false. For example, the networks  $N_5$  and  $N_6$  displayed in Figure 6.29 are distinct, and represent distinct sets of softwired clusters, and yet their tripartition distance is zero. Interestingly, the two networks  $N_7$  and  $N_8$  illustrated in Figure 6.30 are indistinguishable in the sense that they represent the same hardwired clusters, the same softwired clusters and the same rooted phylogenetic trees, and yet their tripartition distance is non-zero.

**Exercise 6.14.6** (Tripartition distance) *Show that the two networks  $N_7$  and  $N_8$  displayed in Figure 6.29 are indistinguishable, and yet they have a non-zero tripartition distance.*

The tripartition distance is a proper metric on the set of all rooted phylogenetic networks that have the tree-child property and are time consistent [38]:

**Lemma 6.14.7** (Metric on tree-child time-consistent networks) *The tripartition distance is a proper metric on the set of all rooted phylogenetic networks that have the tree-child property and are time-consistent.*

### 6.14.3 Comparing networks representing trees

Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ . If the networks are being used to represent a collection of rooted phylogenetic trees on  $\mathcal{X}$ , then it seems reasonable to

base their comparison on the (multi-)sets of trees that they represent, using the *displayed trees distance*:

**Definition 6.14.8** (Displayed trees distance) *Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ . We define the displayed tree distance  $d_{trees}$  between  $N_1$  and  $N_2$  as the cardinality of the symmetric difference of the two sets of rooted phylogenetic trees represented by the two networks, divided by two:*

$$d_{trees}(N_1, N_2) = \frac{|\mathcal{T}(N_1) \Delta \mathcal{T}(N_2)|}{2}.$$

The two rooted phylogenetic networks  $N_1$  and  $N_2$  displayed in Figure 6.27 both represent four rooted phylogenetic trees each (if we distinguish between multiple occurrences of the same tree), but they have no tree in common. Thus, the displayed tree distance between  $N_1$  and  $N_2$  is four.

As in the case of the softwired clusters distance, the use of this distance function is limited by the fact that the number of objects to be compared is exponential in the worst case.

The two rooted phylogenetic networks  $N_3$  and  $N_4$  displayed in Figure 6.28 represent the same set of trees, so their displayed trees distance is zero, although the two networks are distinct. Hence, the displayed trees distance is not a proper metric for rooted phylogenetic networks.

#### 6.14.4 Comparing the topology of networks

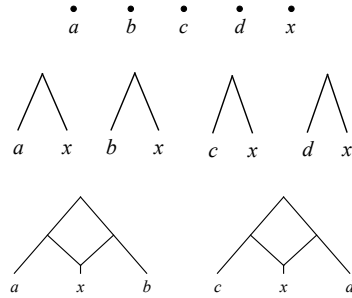
Two rooted phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$  are called *isomorphic*, if there exists a directed-graph isomorphism between them that preserves the labeling of the leaves. The *isomorphism distance function* assigns a distance of 0 between any two isomorphic networks and 1 between any two non-isomorphic ones. The complexity of determining whether two rooted phylogenetic networks are isomorphic, is unknown. However, if both the indegrees and outdegrees of all nodes are bounded, then isomorphism can be checked in polynomial time [164].

The isomorphism distance does not tell us how similar two non-isomorphic networks are. To obtain a more useful distance function, we proceed as follows. For a rooted phylogenetic network  $N = (V, E)$  on  $\mathcal{X}$  and any node  $v$  of  $N$ , we define the *subnetwork*  $N(v)$  associated with  $v$  to be the rooted phylogenetic network  $N|_{V'}$  that is embedded in  $N$  and is induced by the set  $V'$  of all nodes that are descendants of  $v$ , with root  $v$ . For an example, see Figure 6.31. We use  $\Omega(N)$  to denote the multi-set of all rooted phylogenetic networks that are subnetworks of  $N$  and we define [39]:

**Definition 6.14.9** (Subnetwork distance) *Let  $N_1$  and  $N_2$  be two rooted phylogenetic networks on  $\mathcal{X}$ . The subnetwork isomorphism distance between  $N_1$  and  $N_2$  is defined as*

$$d_{sub-net}(N_1, N_2) = \frac{|\Omega(N_1) \Delta \Omega(N_2)|}{2}.$$

As an example, consider the two rooted phylogenetic networks  $N_3$  and  $N_4$  displayed in Figure 6.28. They both have the same proper subnetworks, listed in Figure 6.31. But as



**Figure 6.31** The rooted phylogenetic networks  $N_3$  and  $N_4$  in Figure 6.28 share exactly the same set of proper subnetworks listed here. The two networks differ only as complete networks.

the full networks themselves are also considered subnetworks, the symmetric difference of the two sets has cardinality two and so the subnetwork distance is  $d_{sub-net}(N_1, N_2) = 1$ .

In addition to this approach, a number of simpler distance functions have been suggested that try and capture how topologically similar two rooted phylogenetic networks are. They are based on comparing basic topological invariants of the networks that can be computed efficiently. Also, they are easier to use than the isomorphism-based approach when comparing two networks “by hand”. In the following, we describe the two that are perhaps most useful.

#### The path-multiplicity distance

Let  $N$  be a rooted phylogenetic network on  $\mathcal{X} = \{x_1, \dots, x_n\}$ . The *path-multiplicity distance* is based on the idea of counting the number of different paths from any given node to each of the leaves in  $N$ . The *path-multiplicity vector* of a node  $v$  in  $N$  is defined as

$$\mu(v) = (m_1(v), \dots, m_n(v)),$$

where  $m_i(v)$  is the number of different directed paths from  $v$  to the leaf with label  $x_i$  (for  $i = 1, \dots, n$ ). Let  $\mu(N)$  denote the (multi-)set of all path-multiplicity vectors obtained for all internal nodes of  $N$ . The set of all path-multiplicity vectors is easy to compute and we define a distance by comparing the sets of path-multiplicity vectors [40]:

**Definition 6.14.10** (Path-multiplicity distance) *The path-multiplicity distance (also called the  $\mu$ -distance) between two rooted phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$  is defined as*

$$d_{path}(N_1, N_2) = \frac{|\mu(N_1) \Delta \mu(N_2)|}{2}.$$

As an example, consider the two rooted phylogenetic networks  $N_1$  and  $N_2$  displayed in Figure 6.27. Both have eight internal nodes. Processing the nodes from top-left to bottom-

right, the path vectors associated with network  $N_1$  are:

$$(1, 2, 3, 1, 1), (1, 2, 2, 1, 0), (1, 1, 1, 0, 0), (0, 1, 1, 1, 0), \\ (0, 0, 1, 0, 1), (0, 1, 1, 0, 0), (0, 1, 1, 0, 0), (0, 0, 1, 0, 0).$$

Similarly, for  $N_2$  we obtain:

$$(1, 2, 2, 2, 1), (1, 2, 2, 1, 0), (1, 1, 1, 0, 0), (0, 1, 1, 1, 0), \\ (0, 0, 0, 1, 1), (0, 1, 1, 0, 0), (0, 1, 1, 0, 0), (0, 0, 0, 1, 0).$$

The symmetric difference between the two listed sets contains six elements and so  $d_{\text{path}}(N_1, N_2) = 3$ .

**Exercise 6.14.11** (Indistinguishability) *Show that the path-multiplicity distance between the two distinct rooted phylogenetic networks  $N_3$  and  $N_4$  displayed in Figure 6.28 is zero.*

The two networks depicted in Figure 6.28 show that the path multiplicity distance is not a metric on the set of all rooted phylogenetic networks. However, the path-multiplicity distance is a proper metric on the set of all root phylogenetic networks that have the tree-child property [40]:

**Lemma 6.14.12** (Metric on tree-child networks) *The path-multiplicity distance is a proper metric on the set of all rooted phylogenetic networks that have the tree-child property.*

Moreover, the path-multiplicity distance is also proper metric on the following set of rooted phylogenetic networks [37]:

**Lemma 6.14.13** (Metric on a subset of tree-sibling networks) *The path-multiplicity distance is a proper metric on the set of all bicombinating rooted phylogenetic networks that have the tree-sibling property and are time consistent.*

#### *The nested-labels distance*

The final distance function that we consider is the nested-labels distance. In order to compute this distance function, we assign *nested labels* to all nodes of a rooted phylogenetic network  $N$  on  $\mathcal{X}$  recursively as follows:

- (i) The nested label of a leaf is the taxon that labels the leaf.
- (ii) The nested label of an internal node is defined as the multi-set of all nested labels of its children.

Let  $Y(N)$  denote the multi-set of all nested labels assigned to nodes of  $N$ . The nested-labels distance can be defined as follows [179, 39]:

**Definition 6.14.14** (Nested-labels distance) *The nested-labels distance between two rooted phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$  is defined as*

$$d_{\text{nested}}(N_1, N_2) = \frac{|Y(N_1) \Delta Y(N_2)|}{2}.$$



As an example, consider the two rooted phylogenetic networks  $N_1$  and  $N_2$  shown in Figure 6.27. The leaf nodes give rise to the nested labels  $a, b, \dots, e$ . Although the labels are computed bottom-up, we report them here from top-left to bottom-right. The nested labels of network  $N_1$  are:

$$\begin{aligned} & \{\{a, \{b, c\}\}, \{\{b, c\}, d\}, \{c, e\}, \\ & \{a, \{b, c\}\}, \{\{b, c\}, d\}, \\ & \quad \{a, \{b, c\}\}, \\ & \quad \{\{b, c\}, d\}, \\ & \quad \quad \{b, c\}, \\ & \quad \quad \{b, c\}, \\ & \quad \quad \{c, e\}, \\ & \quad \quad \quad \{c\}, \\ & \quad \quad \quad a, b, c, d, e. \end{aligned}$$

The nodes of the network  $N_2$  give rise to the following nested labels:

$$\begin{aligned} & \{\{\{a\}, \{b, c\}\}, \{\{b, c\}, d\}, \{d, e\}, \\ & \{\{\{a\}, \{b, c\}\}, \{\{b, c\}, d\}\}, \\ & \quad \{\{a\}, \{b, c\}\}, \\ & \quad \{\{b, c\}, d\}, \\ & \quad \quad \{d, e\}, \\ & \quad \quad \{b, c\}, \\ & \quad \quad \{b, c\}, \\ & \quad \quad \{d\}, \\ & \quad \quad a, b, c, d, e. \end{aligned}$$

The symmetric difference of these two sets contains sixteen labels and so we have  $d_{\text{nested}}(N_1, N_2) = \frac{16}{2} = 8$ .

**Exercise 6.14.15** (Indistinguishability) *Show that the nested label distance between the two distinct rooted phylogenetic networks  $N_3$  and  $N_4$  displayed in Figure 6.28 is zero.*

In [39] it is shown that this distance is a proper metric on a number of different classes of networks. Here we report one of the results:

**Lemma 6.14.16** (Metric on tree-child networks) *The nested label distance is a proper metric on the set of all rooted phylogenetic networks that have the tree-child property.*

### 6.14.5 Alignment of rooted phylogenetic networks

To aid a visual comparison of two rooted phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$ , it would be helpful to algorithmically match nodes that are similarly situated in the two networks. In general terms, an *alignment* of two rooted phylogenetic networks  $N_1$  and  $N_2$  on  $\mathcal{X}$  is given by a one-to-one mapping between a subset of the nodes of  $N_1$  and a subset of the nodes of  $N_2$ , which defines a set of *aligned pairs* of nodes.

In the computation of the path-multiplicity distance in Section 6.14.4, each internal node in both rooted phylogenetic networks under comparison is assigned a path-

multiplicity vector. These vectors can be used to compute an alignment between the two networks [40].

In more detail, consider the *network alignment graph*  $A(N_1, N_2) = (V_1 \cup V_2, E)$ , whose node set is the union of the set  $V_1$  of internal nodes of network  $N_1$  and the set  $V_2$  of internal nodes of network  $N_2$ , and whose edge set  $E$  consists of all possible edges between a node in  $V_1$  and a node in  $V_2$ .

In this bipartite graph, we assign a weight  $\omega(e)$  to each edge  $e = (v_1, v_2)$ , with  $v_1 \in V_1$  and  $v_2 \in V_2$ , that is based on the negative sum of absolute differences of the components of the path-multiplicity vectors  $\mu(v_1)$  and  $\mu(v_2)$ , namely

$$\omega((v_1, v_2)) = W - \sum_{i=1}^n |m_i(v_1) - m_i(v_2)|,$$

where  $W$  is a large constant that ensures that all weights will be non-negative. Recall from graph theory that a *matching*  $M$  is a set of edges in a graph such that no two nodes are incident to more than one edge in  $M$ . It is called a *maximum weighted matching*, if, in addition, it maximizes the sum of edge weights  $\sum_{e \in M} \omega(e)$ . Determining a maximum weighted matching for a bipartite graph is computationally easy [152].

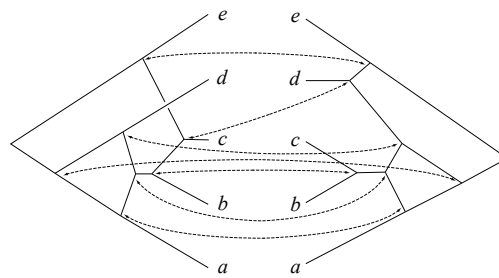
So, let  $M$  be a maximum weighted matching in the alignment graph  $A(N_1, N_2)$ . An alignment of  $N_1$  and  $N_2$  is then given as follows: Any two nodes that are connected by an edge that is contained in the matching  $M$  form an aligned pair. In addition, each pair of leaves that are labeled by the same taxon form an aligned pair.

**Exercise 6.14.17** (Special case) *In a rooted phylogenetic network  $N_1$ , assume that  $r$  is a reticulate node that has exactly one child,  $v$ . Compare the path vectors  $\mu(r)$  and  $\mu(v)$ . What problem can arise for  $r$  and  $v$  in a matching-based alignment to some other network  $N_2$ ? How can the problem be solved?*

To illustrate the alignment of networks, consider the two rooted phylogenetic networks  $N_1$  and  $N_2$  displayed in Figure 6.27. A maximum weighted matching  $M$  in  $A(N_1, N_2)$  is achieved by pairing the internal nodes in the two networks in the order that they occur when listed from top-left to bottom-right, as illustrated in Figure 6.32.

### Chapter notes

This chapter contains many new results and a fair amount of unpublished material. To the best of our knowledge, this is the first attempt to develop a theory of clusters and rooted phylogenetic networks. We became aware of the importance of distinguishing between “softwired” and “hardwired” rooted phylogenetic networks while writing this chapter. The hardness of the cluster containment problem (and also of the tree containment problem) for softwired networks suggests that simply reporting a rooted phylogenetic network without further annotations may be too uninformative and thus additional information indicating how clusters or trees are embedded in the network may be useful (for example, see Figures 4.5 and 4.6).



**Figure 6.32** An alignment of the two rooted phylogenetic networks  $N_1$  and  $N_2$ . Dashed lines are used to connect aligned pairs of nodes. For simplicity, we have omitted the lines between the two root nodes and between the paired leaves.