Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Encoding undirected and semi-directed binary phylogenetic networks by quarnets

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

LEONIE NIPIUS

Delft, the Netherlands
July 2022

MSc thesis APPLIED MATHEMATICS

"Encoding undirected and semi-directed binary phylogenetic networks by quarnets"

LEONIE NIPIUS

Delft University of Technology

Supervisor

Dr.ir. L.J.J. van Iersel

Other thesis comittee members

Dr. Y. Murakami

Dr.ir. G.F. Nane

July, 2022

Delft

Abstract

Phylogenetic networks generalize evolutionary trees and are commonly used to represent evolutionary relationships between species that undergo reticulate evolutionary processes such as hybridization, recombination and lateral gene transfer. In this thesis all quarnets, networks on four species, of a network are assumed to be known. We prove that each recoverable undirected or semi-directed binary level-2 phylogenetic network without redundant biconnected components is encoded by its set of quarnets, meaning that the network is uniquely determined by its quarnets. Furthermore, two decomposition theorems for undirected and semi-directed binary phylogenetic networks are presented. These decomposition theorems are proved for undirected binary phylogenetic networks for all levels and for semi-directed binary phylogenetic networks that are at most level-2.

Contents

Chapter 1

Introduction

Many biological studies use the evolutionary histories of species. Therefore it is important that these relationships can be represented in a suitable way. Nowadays phylogenetic trees and networks are commonly used to represent this. A rooted phylogenetic tree is a rooted (graph theoretical) tree that has no indegree-1 outdegree-1 vertices and in which the leaves are bijectively labelled by the elements in X, where X is a set of species.

Although phylogenetic trees are often used, they can not represent all evolutionary relationships between the different species. We need another way of representing the relationships for species that undergo reticulate evolutionary processes such as hybridization, recombination and lateral gene transfer. For this reason there has been growing interest in using phylogenetic networks instead of phylogenetic trees.

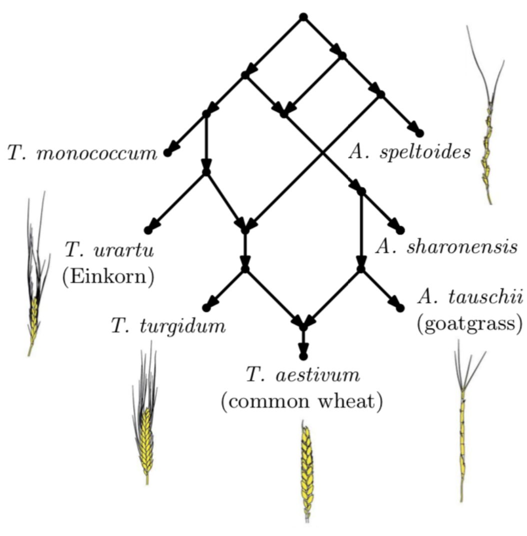


Figure 1.1: A directed phylogenetic network on wheat species. [9], [10], [12]

A directed phylogenetic network is a directed acyclic graph that has a single root, has no indegree-1 outdegree-1 vertices and has its leaves bijectively labelled by the elements in X. A reticulate evolutionary process is represented by a vertex with two incoming arcs. Such a vertex is called a reticulation. In Figure 1.1 an example of a directed phylogenetic network with three reticulations (representing hybridization events) can be found.

The intended purpose is to determine the relationships between the different species from the known biological data. Hereby the focus is often on the different reticulate evolutionary processes. If this is the case, then the relationships between the different species can also be represented by semi-directed phylogenetic networks. A semi-directed phylogenetic network can be obtained from a directed phylogenetic network by replacing the (directed) arcs by (undirected) edges, except for the reticulate events, and then suppressing the root.

A phylogenetic network is level-k if each biconnected component (non tree-like part of the network) has at most k reticulate events. The network in Figure 1.1 is an example of a network that is level-3. In [3], [4] and [1] there have been results regarding when two phylogenetic networks can be distinguished on the basis of data generated by Markov processes. Semi-directed phylogenetic networks come up in these papers because the root location is often not identifiable. These papers first show distinguishability results for 4-leaf phylogenetic networks, which are called quarnets. These results are then used to prove more general theorems for phylogenetic networks with more leaves. This approach has been successful for level-1 but the results for level-2 are still severely limited. The reason for this is that it is still mostly unclear which level-2 quarnets can be distinguished.

Here, we consider the question whether semi-directed level-2 phylogenetic networks can be distinguished if we assume that all quarnets can be distinguished. In other words, this thesis is about phylogenetic networks that can be reconstructed uniquely from their quarnets. In [11] and [13] there are proofs of this encoding by quarnets for directed phylogenetic networks that are level-3. In [14] it is proved that directed level-2 phylogenetic networks are even encoded by trinets, networks on three species.

Further we also consider undirected phylogenetic networks, networks without any given direction. These networks can be used if it is unclear which vertices represent reticulate events and which vertices represent speciation events. Note that we refer to Chapter 2 for all full definitions with their explanations.

Both semi-directed and undirected phylogenetic networks are networks for which the root is not known. There has been more research on these types of phylogenetic networks. We give some examples. In [5] is studied how an undirected phylogenetic network can be oriented. Note that this then results in a certain directed phylogenetic network. In [6] is discussed how semi-directed level-1 phylogenetic networks can be reconstructed from their quarnets.

In this thesis we will prove some results for undirected and semi-directed binary (phylogenetic) networks. We show two decomposition theorems. A CE-split is a partition of the leaves of the network through a cut-edge. The first decomposition theorem shows that the CE-splits of a network are determined by the CE-splits of the quarnets of the network. This first decomposition theorem holds for recoverable undirected binary networks for all levels and recoverable semi-directed binary level-2 networks. With a restriction of a network to a certain biconnected component we mean this biconnected component with its incident cut-edges ending in leaves. Roughly speaking, the second decomposition shows that a network is encoded by its quarnets if and only if the restrictions to the biconnected components are encoded by their quarnets. This second decomposition theorem holds for recoverable undirected binary networks without redundant biconnected components for all levels and recoverable semi-directed binary level-2 networks without redundant biconnected components.

Using these two decomposition theorems we will be able to prove some more results. We prove that every undirected and semi-directed binary simple level-2 network is encoded by its set of quarnets. Moreover, we prove that every recoverable undirected and semi-directed binary level-2 network without redundant biconnected components is encoded by its set of quarnets.

To conclude the introduction, we give an overview of this thesis. In Chapter 2 some preliminaries will be presented. In Chapter 3 we will prove that the class of simple level-2 networks with at least four leaves is weakly encoded by quarnets. In Chapter 4 we will prove the two decomposition theorems. In Chapter 5 we will combine the different results to prove that each recoverable level-2 network without redundant biconnected components is encoded by its set of quarnets. Note that in each of these already mentioned chapters we discuss undirected and semi-directed binary networks in separate sections. In Chapter 6 there is a discussion. Finally, the undirected and semi-directed binary level-3 generators are given in Appendix A.

Chapter 2

Preliminaries

In this chapter we will present some preliminaries. We will discuss in the first section undirected binary networks and in the second section semi-directed binary networks. These two sections can be read in parallel as many definitions are similar. Note that some definitions are missing or defined in a different way for undirected networks because of the differences between undirected and semi-directed networks. Further note that some definitions for directed binary networks are discussed in the second section.

The definitions in this chapter are based on the definitions of directed binary networks as defined in [14]. These definitions of directed networks are also extensively explained in [11]. Note that we refer to this type of networks as directed networks while they call these networks rooted networks. In this way it is easier to distinguish undirected, semi-directed and directed networks in this thesis. Further it is important to keep in mind that however the definitions of undirected and semi-directed networks mainly agree with definitions in other papers, some definitions are defined in a slightly different way.

2.1 Undirected

In this section we will discuss undirected binary networks. Note that sometimes this type of networks will be shortened to undirected networks in this thesis. The definition of an undirected binary (phylogenetic) network can be found below.

Definition 2.1. An *undirected binary (phylogenetic) network* on a set X is a connected undirected graph without loops or parallel edges such that each vertex has degree one or three and the vertices with degree one are bijectively labelled by the elements of X.

A degree-1 vertex in this definition is called a *leaf*. The set of all leaves is then called X. Here we assume that X denotes a non-empty finite set. As the leaves are bijectively labelled by the elements of X, we can identify each leaf with its label. Since X often consists of species, this means each leaf represents one of the species. Let N be an undirected binary network. A directed binary network N_D is an *orientation* of N if the undirected binary network obtained from N_D by replacing all arcs with edges and subsequently suppressing its root equals N (see Definition 2.16). By suppressing the root r, which is a degree-2 vertex, we mean that if there are edges ur and rv, then this becomes a single edge uv. Note that an orientation is not always unique. Furthermore, N is *orientable at a cut-edge* uv of N if there exists an orientation N_D of N such that N_D has arcs (r, u) and (r, v) where r is the root of N_D. Moreover, N is *orientable at a leaf* $\rho \in X$ if there exists an orientation N_D of N such that N_D has the arc (r, ρ) where r is the root of N_D. In Figure 2.1 an example of an undirected binary network can be found.

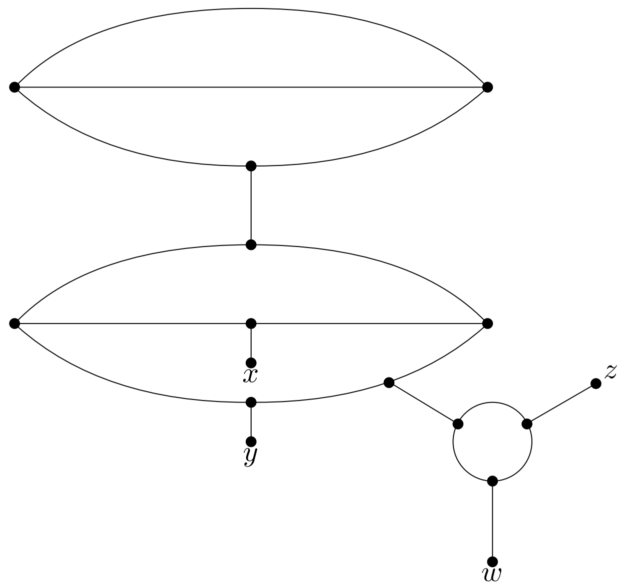


Figure 2.1: An example of an undirected binary network.

Later in this thesis we want to prove some results for an undirected binary network. As the network can be very large, it is a good idea to cut the network in different components. With this it is possible to look at each component of the network separately. In order to do this we need some definitions for an undirected binary network. First we give the definitions of cut-vertices and cut-edges of an undirected binary network.

Definition 2.2. Let N be an undirected binary network. A vertex v of N is a *cut-vertex* if its removal disconnects the graph of N. Similarly, an edge e of N is a *cut-edge* if its removal disconnects the graph of N.

Further we need the definition of a CE-split of an undirected binary network, which is given below.

Definition 2.3. Let N be an undirected binary network on X and $\{P, Q\}$ a partition of X. $\{P, Q\}$ is a *CE-split (Cut-Edge split)* of N if there exists a cut-edge uv of N such that its removal gives two connected graphs with leaves P and Q, respectively.

Now we use the definition of a cut-vertex in the definition below.

Definition 2.4. An undirected binary network is *biconnected* if it has no cut-vertices.

With above definition we can define a component of an undirected binary network.

Definition 2.5. A *biconnected component* of an undirected binary network is a maximal biconnected subgraph (i.e. a biconnected subgraph that is not contained in any other biconnected subgraph).

An undirected binary network often consists of many small components. A biconnected component of an undirected binary network is *trivial* if it is a cut-edge. The undirected binary network that is given in Figure 2.1 has three nontrivial biconnected components.

Now we will consider some different types of undirected binary networks by defining several conditions for undirected binary networks. In the following definition we first define when a nontrivial component is redundant or strongly redundant.

Definition 2.6. Let B be a nontrivial biconnected component of an undirected binary network. An edge uv is an *incident cut-edge* of B if u is in B and v is not in B. B is *redundant* if B has exactly two incident cut-edges. B is *strongly redundant* if B has exactly one incident cut-edge.

With the following definition we can easily define networks without components that are strongly redundant. Later in this thesis we will mainly look at networks that fulfil this requirement.

Definition 2.7. An undirected binary network is *recoverable* if it has no strongly redundant biconnected components.

The undirected binary network that is given in Figure 2.1 is not recoverable, because the upper nontrivial biconnected component is strongly redundant. Another requirement that is used a lot for a network in this thesis is simple, which is defined below.

Definition 2.8. An undirected binary network is *simple* if for each cut-edge uv holds that u or v is a leaf.

Note that a simple network has only one nontrivial component. Each undirected binary network can be transformed in such a way that it has only one nontrivial component and is therefore simple. In the definition below we can see that this can be done for each nontrivial biconnected component of the network.

Definition 2.9. Let N be an undirected binary network and B a nontrivial biconnected component with b cut-edges $e_1 = u_1v_1, \ldots, e_b = u_bv_b$ such that u_i is in B and v_i is not in B. Consider the undirected binary network N_B obtained from N by deleting all biconnected components except B, $e_1, \ldots, e_b$ and labelling $v_1, \ldots, v_b$ by new labels $y_1, \ldots, y_b$ that are not in X. Then, N_B is a *restriction* of N to B.

Note that N_B in the definition above is unique up to the choice of the new labels $y_1, \ldots, y_b$. We continue with the definition of another condition for an undirected binary network, namely a network that is level-k.

Definition 2.10. An undirected binary network is *level-k* if each biconnected component has at most $|V| + k - 1$ edges.

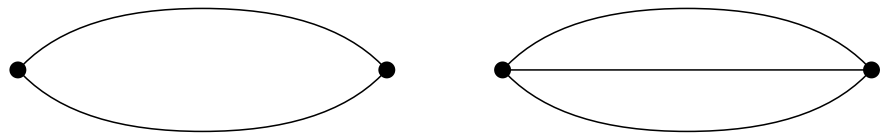
To define it more precisely we introduce the definition of strict. An undirected binary network is *strict level-k* if it is level-k but not level-$(k-1)$. An undirected binary simple (strict) level-k network is then a network with one nontrivial biconnected component that is (strict) level-k. The undirected binary network that is given in Figure 2.1 is strict level-2 since it has two nontrivial biconnected components that are strict level-2 and one nontrivial biconnected component that is strict level-1.

Another way to look at the biconnected components of an undirected binary network is considering the underlying structure of each component. Therefore we first give the definition of a generator.

Definition 2.11 (Definition 2.6 in [1])**.** An undirected binary level-k *generator* G is a multigraph as follows:

- G is a single vertex (if $k = 0$);
- G is the 2-regular multigraph with 2 vertices (if $k = 1$);
- G is a 3-regular biconnected multigraph with $2k-2$ vertices (if $k \geq 2$).

In Figure 2.2 the undirected binary generators for level-1 and level-2 are given. It can be seen that there are multiple generators for $k \geq 3$. For the undirected binary level-3 generators and their discussion we refer to Appendix A.1.



(a) Level-1 generator. (b) Level-2 generator.

Figure 2.2: The level-1 and level-2 generator of undirected binary networks.

We continue with the definition of a side of a generator. Note that where directed generators can have arc sides and reticulation sides undirected generators can only have edge sides.

Definition 2.12. The *sides* of an undirected binary level-k generator are the edges of the generator.

Note that deleting all leaves of an undirected binary simple strict level-k network N gives an undirected binary level-k generator G_N. G_N is then the *underlying generator* of N. Conversely, N can be reconstructed from G_N by hanging leaves on its sides as follows:

- for each edge e of G_N, replace e by an undirected path P_e with $l \geq 0$ internal vertices $v_1, \ldots, v_l$ and, for each such internal vertex v_i, add a leaf $x_i \in X$ and an edge $v_i x_i$.

A leaf x of an undirected binary simple strict level-k network *is on side s* if it is hung on side s in the construction of N from G_N. More precisely, for a leaf $x \in X$ of an undirected binary simple strict level-k network N with underlying generator G_N and a side s of G_N, x is on side s if the following holds:

- s is an edge uv of G_N and there is a vertex y such that xy is an edge of N and y lies on the undirected path P_s from u to v in N.

In the next definition is explained when a generator has symmetry. Intuitively, symmetry of a generator means that there exists a relabelling of the sides of the generator giving an isomorphic generator.

Definition 2.13. An undirected binary generator G has *symmetry* if it has parallel edges or if there exists a bijective function $f : V(G) \rightarrow V(G)$ such that for all $u, v \in V$ the number of edges between u and v is equal to the number of edges between $f(u)$ and $f(v)$ but $f(w) \neq w$ for at least one $w \in V$.

Finally we look at definitions that are needed to consider the relationship between networks and their subnetworks. Below the definitions of a trinet and a quarnet are given.

Definition 2.14. An undirected binary network is a *trinet (quarnet)* if it has three (four) leaves.

Trinets and quarnets are examples of smaller networks. A subnetwork can be exhibited by a larger network, as explained in the following definition. By suppressing parallel edges we mean that each set of parallel edges is replaced by a single edge. By suppressing strongly redundant biconnected components we mean that each such component is replaced by a single vertex.

Definition 2.15. Given an undirected binary network N on X and $A \subseteq X$, the undirected binary network on A *exhibited* by N is the undirected binary network obtained from N by deleting all leaves except the leaves in A and repeatedly applying the following operations until none is applicable:

- deleting all unlabelled vertices with degree one;
- suppressing all vertices with degree two;
- suppressing all parallel edges; and
- suppressing all strongly redundant biconnected components.

If $|A| = 3$ or $|A| = 4$, then the undirected binary network on A exhibited by N is a trinet or quarnet, respectively. With different sets A all trinets or quarnets of an undirected binary network can be exhibited. Let N be an undirected binary network. Then we denote the set of all trinets exhibited by N with $Tn(N)$ and the set of all quarnets exhibited by N with $Qn(N)$. In Figure 2.3 an example of an exhibited trinet is given.

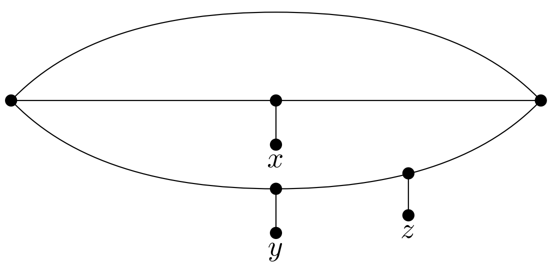


Figure 2.3: A trinet exhibited by the quarnet that is given in Figure 2.1.

In the following definition is explained when two undirected binary networks are equal. Shortly, this is the case when there is a graph isomorphism between these two networks that preserves leaf labels.

Definition 2.16. Given two undirected binary networks N and N' on X, we write $N = N'$ if there exists a bijective function $f : V(N) \to V(N')$ such that $f(x) = x$ for each leaf x of N and such that for every $u, v \in V(N)$ holds that uv is an edge of N if and only if $f(u)f(v)$ is an edge of N'.

Now we have defined when two undirected binary networks are equal, we can define when an undirected binary network is encoded by its sets of trinets or quarnets. Intuitively, it means that a recoverable undirected binary network can be uniquely reconstructed with all of its trinets or quarnets, respectively.

Definition 2.17. An undirected binary network N is *encoded* by its sets of trinets (quarnets) if there is no recoverable undirected binary network $N' \neq N$ with $Tn(N) = Tn(N')$ ($Qn(N) = Qn(N')$).

As it is not always possible to show that a network is encoded, there is also a weaker definition for a network being encoded. In this definition the network can be reconstructed uniquely from a certain class of undirected binary networks instead from all undirected binary networks.

Definition 2.18. A class of undirected binary networks $\mathcal{C}$ is *weakly encoded* by trinets (quarnets) if there are no two recoverable undirected binary networks N and N', with $N \neq N'$, in class $\mathcal{C}$ such that $Tn(N) = Tn(N')$ ($Qn(N) = Qn(N')$).

2.2 Semi-directed

In this section we will discuss semi-directed binary networks. Note that sometimes this type of networks will be shortened to semi-directed networks in this thesis. Some of the definitions for semi-directed networks use definitions for directed networks. Note that most of the definitions for directed networks are not given in this section since these definitions are already clearly explained in [11]. One of the definitions for directed networks we will give explicitly is the definition of a directed binary network, which is stated below.

Definition 2.19. A *directed binary (phylogenetic) network* on a set X is a directed acyclic graph for which holds the following:

- there is a single indegree-0 vertex;
- there are no indegree-1 outdegree-1 vertices;
- the vertices have at most indegree two;
- the vertices have at most outdegree two;
- the vertices with indegree two have outdegree one; and
- the outdegree-0 vertices are bijectively labelled by the elements of X.

In this definition of a directed binary network the indegree-0 vertex is called the root and the outdegree-0 vertices are called leaves. Using above definition of a directed network we can define a semi-directed network. Note that the suppressed degree-2 vertex in the definition was the root of the directed binary network. By suppressing a degree-2 vertex r we mean that if there are edges ur and rv, then this becomes a single edge uv, and if there is an edge ur and an arc (r, v), then this becomes a single arc (u, v).

Definition 2.20. A *semi-directed binary (phylogenetic) network* on a set X can be obtained from a directed binary network by replacing the arcs by edges except for the incoming arcs of an indegree-2 vertex and then suppressing the degree-2 vertex.

A degree-1 vertex of a semi-directed binary network is called a *leaf.* The set of all leaves is then called X. Here we assume that X denotes a non-empty finite set. As the leaves are bijectively labelled by the elements of X, we can identify each leaf with its label. Since X often consists of species, this means each leaf represents one of the species. Let N be a semi-directed binary network. If N_D is a directed binary network from which N can be *obtained* (as described in above definition), then N_D is called an *orientation* of N. Note that by definition there exists at least one orientation of N. Furthermore, N is *orientable at a cut-edge uv* of N if there exists an orientation N_D of N such that N_D has arcs (r, u) and (r, v) where r is the root of N_D. Moreover, N is *orientable at a leaf $\rho \in X$* if there exists an orientation N_D of N such that N_D has the arc (r, ρ) where r is the root of N_D. In Figure 2.4 an example of a semi-directed binary network can be found.

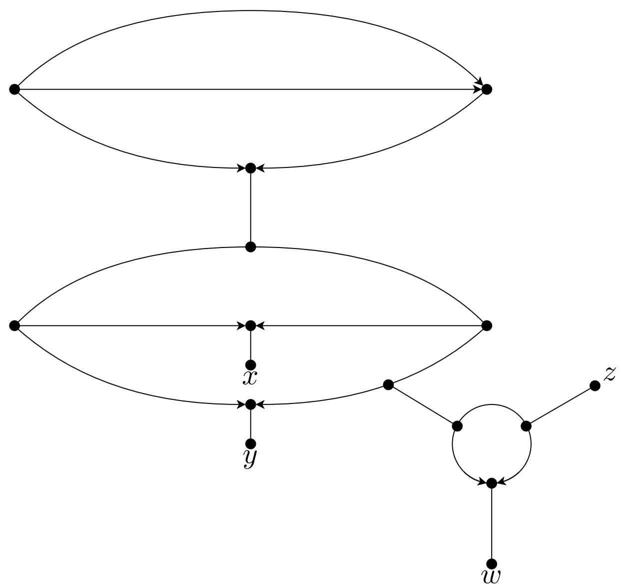


Figure 2.4: An example of a semi-directed binary network.

Later in this thesis we want to prove some results for a semi-directed network. As the network can be very large, it is a good idea to cut the network in different components. With this it is possible to look at each component of the network separately. In order to do this we need some definitions for a semi-directed binary network. First we give the definitions of cut-vertices, cut-arcs and cut-edges of a semi-directed binary network.

Definition 2.21. Let N be a semi-directed binary network. A vertex v of N is a *cut-vertex* if its removal disconnects the underlying undirected graph of N. Similarly, an arc a of N is a *cut-arc* if its removal disconnects the underlying undirected graph of N. An edge e of N is a *cut-edge* if its removal disconnects the underlying undirected graph of N.

The following lemma shows that semi-directed binary networks have no cut-arcs. Keep this in mind when comparing the definitions of directed networks (see [11]) with the definitions of semi-directed networks.

Lemma 2.22. *Let N be a semi-directed binary network. Then N has no cut-arcs.*

Proof. Let N be a semi-directed binary network. Assume that N has a cut-arc (x, z). Then by definition z is a reticulation vertex with two incoming arcs (x, z) and (y, z). Let N_D be any directed binary network from which N can be obtained. Let s be the lowest stable ancestor $LSA(\{x, y\})$ in N_D (see definition LSA in [11] or [14]). Let P_x be a directed path from s to x and P_y a directed path from s to y in N_D. Note that these paths only have vertex s in common. Then the underlying undirected graph of P_x, P_y, (x, z) and (y, z) is an undirected cycle. By definition this implies that the underlying undirected graph of N has also an undirected cycle containing x, y and z. Then (x, z) is not a cut-arc since its removal does not disconnect the underlying undirected graph of N because of the undirected cycle. This gives a contradiction. So N has no cut-arcs. □

We need an extra definition for directed binary networks next to the definitions given in [11], namely the definition of a CA-split. Therefore we first recall the definition of a CA-set of a directed binary network.

Definition 2.23. Let N be a directed binary network on X and $Q \subseteq X$. Q is a *CA-set (Cut-Arc set)* of N if there exists a cut-arc (u, v) of N such that $Q = \{x \in X | x \text{ is below } v\}$.

Now we can define a CA-split of a directed binary network.

Definition 2.24. Let N be a directed binary network on X and $\{P, Q\}$ a partition of X. $\{P, Q\}$ is a *CA-split (Cut-Arc split)* of N if P or Q is a CA-set of N.

For a semi-directed binary network we define a CE-split. Note that we do not define CA-splits for semi-directed binary networks as these networks do not have any cut-arcs.

Definition 2.25. Let N be a semi-directed binary network on X and $\{P, Q\}$ a partition of X. $\{P, Q\}$ is a *CE-split (Cut-Edge split)* of N if there exists a cut-edge uv of N such that its removal gives two connected graphs with leaves P and Q, respectively.

Now we use the definition of a cut-vertex in the definition below.

Definition 2.26. A semi-directed binary network is *biconnected* if it has no cut-vertices.

With above definition we can define a component of a semi-directed binary network.

Definition 2.27. A *biconnected component* of a semi-directed binary network is a maximal biconnected subgraph (i.e. a biconnected subgraph that is not contained in any other biconnected subgraph).

A semi-directed binary network often consists of many small components. A biconnected component of a semi-directed binary network is *trivial* if it is a cut-edge. The semi-directed binary network that is given in Figure 2.4 has three nontrivial biconnected components.

Now we will consider some different types of semi-directed binary networks by defining several conditions for semi-directed binary networks. In the following definition we first define when a nontrivial component is redundant or strongly redundant.

Definition 2.28. Let B be a nontrivial biconnected component of a semi-directed binary network. An edge uv is an *incident cut-edge* of B if u is in B and v is not in B. B is *redundant* if B has exactly two incident cut-edges. B is *strongly redundant* if B has exactly one incident cut-edge.

With the following definition we can easily define networks without components that are strongly redundant. Later in this thesis we will mainly look at networks that fulfil this requirement.

Definition 2.29. A semi-directed binary network is *recoverable* if it has no strongly redundant biconnected components.

The semi-directed binary network that is given in Figure 2.4 is not recoverable, because the upper nontrivial biconnected component is strongly redundant. Another requirement that is used a lot for a network in this thesis is simple, which is defined below.

Definition 2.30. A semi-directed binary network is *simple* if for each cut-edge uv holds that u or v is a leaf.

Note that a simple network has only one nontrivial component. Each semi-directed binary network can be transformed in such a way that it has only one nontrivial component and is therefore simple. In the definition below we can see that this can be done for each nontrivial biconnected component of the network.

Definition 2.31. Let N be a semi-directed binary network and B a nontrivial biconnected component with b cut-edges $e_1 = u_1v_1, \ldots, e_b = u_bv_b$ such that u_i is in B and v_i is not in B. Consider the semi-directed binary network N_B obtained from N by deleting all biconnected components except B, $e_1, \ldots, e_b$ and labelling $v_1, \ldots, v_b$ by new labels $y_1, \ldots, y_b$ that are not in X. Then, N_B is a *restriction* of N to B.

Note that N_B in the definition above is unique up to the choice of the new labels $y_1, \ldots, y_b$. In the definition below we define a reticulation vertex of a semi-directed binary network.

Definition 2.32. A *reticulation (vertex)* of a semi-directed binary network is a vertex with two incoming arcs.

The following definition explains another condition for a semi-directed binary network, namely a network that is level-k.

Definition 2.33. A semi-directed binary network is *level-k* if each biconnected component has at most k reticulations.

To define it more precisely we introduce the definition of strict. A semi-directed binary network is *strict level-k* if it is level-k but not level-$(k-1)$. A semi-directed binary simple (strict) level-k network is then a network with one nontrivial biconnected component that is (strict) level-k. This implies that the nontrivial biconnected component has exactly k reticulations if the simple network is strict level-k. The semi-directed binary network that is given in Figure 2.4 is strict level-2 since it has two nontrivial biconnected components with two reticulations and one nontrivial biconnected component with one reticulation.

Another way to look at the biconnected components of a semi-directed binary network is considering the underlying structure of each component. Before we give the definition of a semi-directed generator, we first recall the definition of a directed generator.

Definition 2.34. A directed binary level-k *generator* is a directed acyclic biconnected multigraph with exactly k reticulations with indegree-2 and outdegree at most one, a single vertex with indegree-0 and outdegree-2, and apart from that only vertices with indegree-1 and outdegree-2.

In Figures 2.5 and 2.6 the directed binary generators for level-1 and level-2 are given. Note that the directed binary level-2 generators are also discussed in [14] and [11].

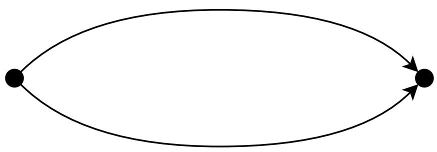
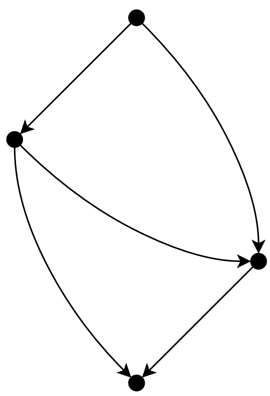
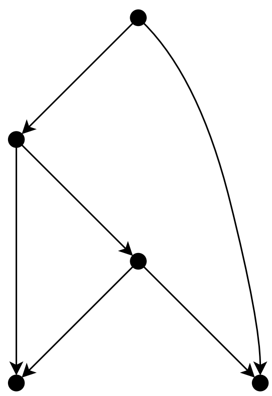


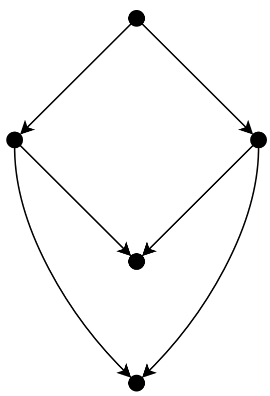
Figure 2.5: The level-1 generator of directed binary networks.



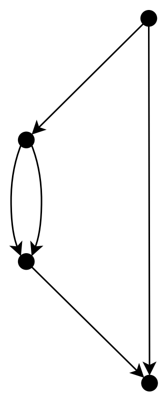
(a) Generator $2a$.



(b) Generator $2b$.



(c) Generator $2c$.



(d) Generator $2d$.

Figure 2.6: The four level-2 generators of directed binary networks. [14]

Below the definition of a semi-directed binary generator can be found, which is based on the definition of a directed binary generator. Note that the suppressed degree-2 vertex in the definition was the root of the directed binary generator.

Definition 2.35. A semi-directed binary level-k *generator* can be obtained from a directed binary level-k generator by replacing the arcs by edges except for the incoming arcs of an indegree-2 vertex and then suppressing any degree-2 vertex that does not have two incoming arcs or two outgoing arcs.

In Figures 2.7 and 2.8 the semi-directed binary generators for level-1 and level-2 are given. Note that these generators are the only generators that can be obtained from the generators in Figures 2.5 and 2.6 in the way as described in the definition above and are therefore the only semi-directed generators for level-1 and level-2. Hereby note that we do not suppress the degree-2 vertex for the semi-directed binary level-1 generator. Otherwise we would get an double directed arc in the generator which results in a loop. The semi-directed binary generator 2.1 can be obtained from the directed binary generator $2a$ or $2d$. The semi-directed binary generator 2.2 can be obtained from the directed binary generator $2b$ or $2c$. For the semi-directed binary level-3 generators and their discussion we refer to Appendix A.2.

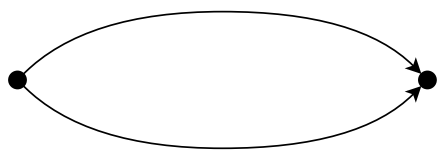
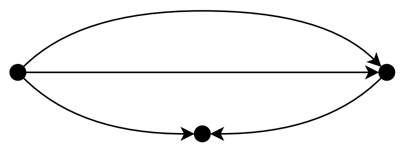
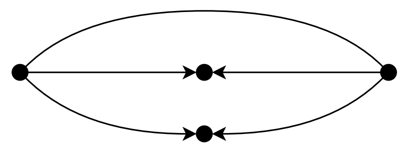


Figure 2.7: The level-1 generator of semi-directed binary networks.



(a) Generator 2.1.



(b) Generator 2.2.

Figure 2.8: The two level-2 generators of semi-directed binary networks.

We continue with the definition of a side of a generator. Note that where directed generators can have arc sides and reticulation sides semi-directed generators can have next to arc sides and reticulation sides also edge sides. Further note that the sides that are degree-2 vertices by definition have always two incoming arcs.

Definition 2.36. The *sides* of a semi-directed binary level-k generator are the arcs, edges and the degree-2 vertices of the generator.

Note that deleting all leaves of a semi-directed binary simple strict level-k network N gives a semi-directed binary level-k generator G_N. G_N is then the *underlying generator* of N. Conversely, N can be reconstructed from G_N by hanging leaves on its sides as follows:

- for each arc a of G_N, replace a by a semi-directed path P_a (only the last part is directed) with $l \geq 0$ internal vertices $v_1, \ldots, v_l$ and, for each each such internal vertex v_i, add a leaf $x_i \in X$ and an edge $v_i x_i$;
- for each edge e of G_N, replace e by an undirected path P_e with $l \geq 0$ internal vertices $v_1, \ldots, v_l$ and, for each such interval vertex v_i, add a leaf $x_i \in X$ and an edge $v_i x_i$; and
- for each degree-2 vertex v, add a leaf $x \in X$ and an edge vx.

A leaf x of a semi-directed binary simple strict level-k network *is on side* s if it is hung on side s in the construction of N from G_N. More precisely, for a leaf $x \in X$ of a semi-directed binary simple strict level-k network N with underlying generator G_N and a side s of G_N, x is on side s if one of the following holds:

- s is an arc (u, v) of G_N and there is a vertex y such that xy is an edge and y lies on the semi-directed path P_s from u to v in N;
- s is an edge uv of G_N and there is a vertex y such that xy is an edge and y lies on the undirected path P_s from u to v in N; or
- s is a degree-2 vertex of G_N and sx is an edge of N.

With the following definition the order of two leaves on an arc side of the generator can be distinguished.

Definition 2.37. Let x and y be two leaves on an arc side $s = (u, v)$ of the generator of a semi-directed binary network. If the parent of x is after the parent of y on the semi-directed path P_s from u to v in N, then x is *below* y on arc side s.

In the next definition is explained when a generator has symmetry. Intuitively, symmetry of a generator means that there exists a relabelling of the sides of the generator giving an isomorphic generator.

Definition 2.38. A semi-directed binary generator G has *symmetry* if it has parallel arcs or if there exists a bijective function $f : V(G) \to V(G)$ such that for all $u, v \in V$ the number of edges between u and v is equal to the number of edges between $f(u)$ and $f(v)$ and the number of arcs between u and v is equal to the number of arcs between $f(u)$ and $f(v)$ but $f(w) \neq w$ for at least one $w \in V$.

Finally we look at definitions that are needed to consider the relationship between networks and their subnetworks. Below the definitions of a trinet and a quarnet are given.

Definition 2.39. A semi-directed binary network is a *trinet (quarnet)* if it has three (four) leaves.

Trinets and quarnets are examples of smaller networks. A subnetwork can be exhibited by a larger network, as explained in the following definition. By suppressing parallel arcs we mean that each set of parallel arcs is replaced by a single arc. By suppressing strongly redundant biconnected components we mean that each such component is replaced by a single vertex.

Definition 2.40. Given a semi-directed binary network N on X and $A \subseteq X$, the semi-directed binary network on A *exhibited* by N is the semi-directed binary network obtained from N by deleting all leaves except the leaves in A and repeatedly applying the following operations until none is applicable:

- deleting all unlabelled vertices with degree one;
- suppressing all vertices with degree two;
- suppressing all parallel arcs (except parallel arcs that were already in N); and
- suppressing all strongly redundant biconnected components.

If $|A| = 3$ or $|A| = 4$, then the semi-directed binary network on A exhibited by N is a trinet or quarnet, respectively. With different sets A all trinets or quarnets of a semi-directed binary network can be exhibited. Let N be a semi-directed binary network. Then we denote the set of all trinets exhibited by N with $Tn(N)$ and the set of all quarnets exhibited by N with $Qn(N)$. In Figure 2.9 an example of an exhibited trinet is given.

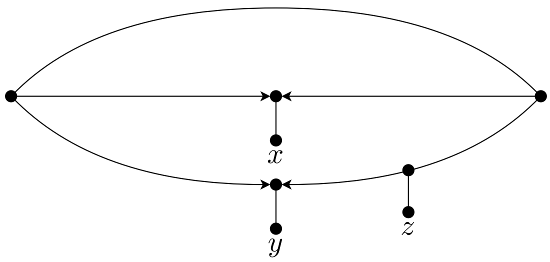


Figure 2.9: A trinet exhibited by the quarnet that is given in Figure 2.4.

In the following definition is explained when two semi-directed binary networks are equal. Shortly, this is the case when there is a graph isomorphism between these two networks that preserves leaf labels.

Definition 2.41. Given two semi-directed binary networks N and N' on X, we write $N = N'$ if there is a graph isomorphism between N and N' that preserves leaf labels, i.e. if there exists a bijective function $f : V(N) \to V(N')$ such that $f(x) = x$ for each leaf x of N and such that for every $u, v \in V(N)$ holds that (u, v) is an arc of N if and only if $(f(u), f(v))$ is an arc of N' and uv is an edge of N if and only if $f(u)f(v)$ is an edge of N'.

Now we have defined when two semi-directed binary networks are equal, we can define when a semi-directed binary network is encoded by its sets of trinets or quarnets. Intuitively, it means that a recoverable semi-directed binary network can be uniquely reconstructed with all of its trinets or quarnets, respectively.

Definition 2.42. A semi-directed binary network N is *encoded* by its sets of trinets (quarnets) if there is no recoverable semi-directed binary network $N' \neq N$ with $Tn(N) = Tn(N')$ $(Qn(N) = Qn(N'))$.

As it is not always possible to show that a network is encoded, there is also a weaker definition for a network being encoded. In this definition the network can be reconstructed uniquely from a certain class of semi-directed binary networks instead from all semi-directed binary networks.

Definition 2.43. A class of semi-directed binary networks $\mathcal{C}$ is *weakly encoded* by trinets (quarnets) if there are no two recoverable semi-directed binary networks N and N', with $N \neq N'$, in class $\mathcal{C}$ such that $Tn(N) = Tn(N')$ $(Qn(N) = Qn(N'))$.

Chapter 3

Simple networks

In this chapter we consider undirected and semi-directed binary simple networks. We will show that the class of simple level-2 networks is weakly encoded by quarnets. First we prove that this holds for undirected binary networks, then for semi-directed binary networks. Note that the different level-1 and level-2 generators are given in Chapter 2.

3.1 Undirected

In this section we prove that the class of undirected binary simple level-2 networks is weakly encoded by quarnets. We will show this by first proving it for strict level-1 and strict level-2 networks separately. In the theorem below we prove the result for undirected strict level-1 networks.

Theorem 3.1. *The class of undirected binary simple strict level-1 networks with at least four leaves is weakly encoded by quarnets.*

Proof. Assume there are two recoverable phylogenetic networks N and N' in the class of undirected binary simple strict level-1 networks with at least four leaves such that $Qn(N) = Qn(N')$. To prove that the class is weakly encoded by quarnets we have to show that $N = N'$.

By definition of the class the networks are all of the same form, which is given in Figure 3.1. Therefore we only have to show that the circular order of the leaves is the same for N and N'.

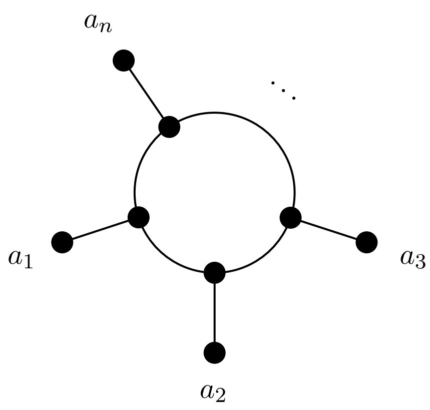


Figure 3.1: The form of an undirected binary simple strict level-1 network.

Let x and y be two leaves that are neighbours in N. Consider any quarnet containing x and y. This quarnet is level-1 and is therefore of the same form as N and N'. This holds also for the quarnets that will be considered later in the proof. Since x and y are neighbours in N, x and y are also neighbours in the quarnet. Suppose x and y are not neighbours in N'. Then there exists a quarnet with leaves p and q such that the circular order of these four leaves is x, p, y and q. This gives a contradiction since $Qn(N) = Qn(N')$. So x and y are also neighbours in N'.

Now we can see the circular order of the leaves as starting with x and ending with y. Then the remaining leaves are between x and y in the circular order, as can be seen in Figure 3.2.

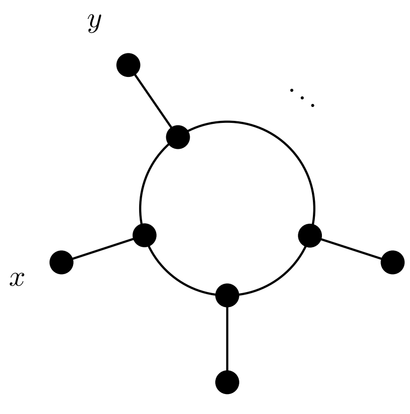


Figure 3.2: The form of an undirected binary simple strict level-1 network with two leaves fixed for the start and end of the circular order.

Let v and w be any two leaves, not x or y. Consider the quarnet with x, y, v and w. If v is earlier in the circular order than w in N, then the circular order of the quarnet is x, v, w and y. Suppose w is earlier in the circular order than v in N'. Then the circular order of the considered quarnet is x, w, v and y, which is a contradiction since $Qn(N) = Qn(N')$. Similarly, we get that if w is earlier in the circular order than v in N, w is earlier in the circular order than v in N'. As this can be done for all the different pairs of leaves, the circular order of the leaves is the same for N and N'. With this we can conclude that $N = N'$. □

In the theorem below we prove the wanted result for undirected strict level-2 networks. In this proof we use the underlying structure of such a network, the undirected level-2 generator as given in Figure 2.2b. Note that in the theorem above we did not use the undirected level-1 generator (see Figure 2.2a) as underlying structure explicitly since the two different sides can not be distinguished. Therefore we did not look at the order of the leaves on both sides separately, but at the order of the leaves on the two sides together. In this way we were still able to obtain the wanted result for level-1.

Theorem 3.2. *The class of undirected binary simple strict level-2 networks with at least four leaves is weakly encoded by quarnets.*

Proof. Assume there are two recoverable phylogenetic networks N and N' in the class of undirected binary simple strict level-2 networks with at least four leaves such that $Qn(N) = Qn(N')$. To prove that the class is weakly encoded by quarnets we have to show that $N = N'$.

By definition of the class the networks in this class have only one level-2 generator. Therefore N and N' have the same underlying generator (see Figure 3.3).

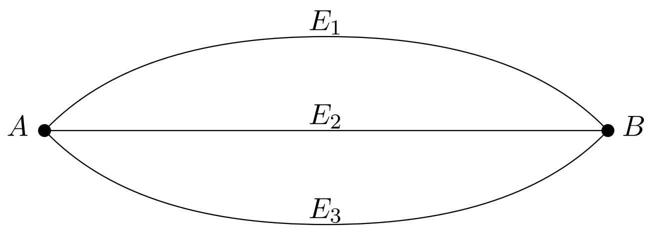


Figure 3.3: The level-2 generator of undirected binary networks.

Observe that the generator has symmetry since sides E_1, E_2 and E_3 can be interchanged to obtain an isomorphic generator. Let x and y be two leaves on the sides E_1 and E_2 in N, respectively. If there is no leaf on side E_1 or E_2 in N, interchange the label of this side with E_3 in N in order to have leaves x and y on sides E_1 and E_2, respectively. Note that there are at least two sides with a leaf on it, otherwise the underlying generator was not level-2 as an undirected network has no parallel edges by definition. Consider a quarnet that contains x and y. This quarnet has the same underlying generator as N and N'. This holds also for the quarnets that will be considered later in the proof. Since x and y are, respectively, on sides E_1 and E_2, x and y are on different sides in the quarnet. This shows x and y are also on different sides in N' as $Qn(N) = Qn(N')$. The sides E_1, E_2 and E_3 in N' can be interchanged such that x and y are, respectively, on sides E_1 and E_2 in N' because of the symmetry.

Let z be a leaf, not x or y. Consider a quarnet that contains x, y and z. This quarnet has again the same underlying generator as N and N'. If z is on side E_1 in N, then x and z are on the same side of the generator in the quarnet which implies z is also on side E_1 in N'. Note that the symmetry of the sides of the generator is already fixed. If z is on side E_2 in N, then y and z are on the same side of the generator in the quarnet which implies z is also on side E_2 in N'. If z is on side E_3 in N, then z is not on the same side of the generator in the quarnet as x or y which implies z is also on side E_3 in N'. This shows that N and N' have the same leaves on each side.

We continue with the order of the leaves on each side to end the proof. Without loss of generality assume that side E_1 is a side with the largest number of leaves and side E_3 is a side with the least number of leaves. Side E_1 has then at least two leaves as each network in the class has at least four leaves by definition while the generator has three sides. Let p be the first leaf on side E_1 and s the first leaf on side E_2, both viewing from label A in N (see Figure 3.4). Consider any quarnet containing p and s. Then p is always the first leaf on side E_1 viewing from A or B in the quarnet. Suppose this is not the case for N'. Then p is between two other leaves on side E_1. This is only possible if side E_1 has more than two leaves. Then p has also to be between two other leaves on side E_1 in a quarnet, which is a contradiction since $Q(N) = Q(N')$. So p is always the first leaf on side E_1 viewing from A or B in N'. Note that this especially becomes clear from the considered quarnets that have two or, if possible, three leaves on side E_1.

If p is the first leaf on side E_1 in N' viewing from B, the labels A and B have to be interchanged for N'. Note that this does not change the network itself as the labels A and B are not part of the network as only the vertices that are leaves are labelled. The labels A and B are useful for distinguishing the order of the leaves from left to right versus right to left on the different sides in the proof.

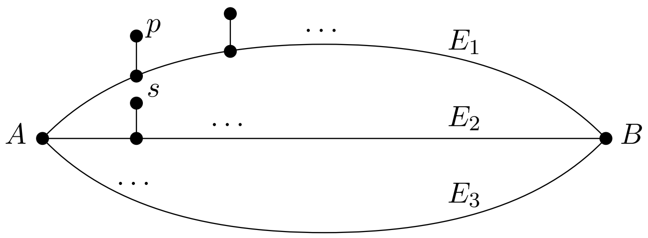


Figure 3.4: The level-2 generator of undirected binary networks with some leaves labelled. Side E_1 has at least two leaves, side E_2 at least one leaf and side E_3 has no minimum amount of leaves.

If side E_1 has two leaves, then the order of the leaves on side E_1 is already fixed as p is fixed. Now assume that side E_1 has more than two leaves. Let v and w be any two leaves on side E_1, not p. Consider the quarnet with p, s, v and w. Since p is fixed, the order of v and w on side E_1 can be determined. If v is nearer to p than w on side E_1 in the quarnet, then this is also the case in N'. If w is nearer to p than v on side E_1 in the quarnet, then this is also the case in N'. As this can be done for all the different pairs of leaves, the order of the leaves on side E_1 is the same for N and N'.

If side E_2 has one leaf, the order of the leaves on side E_2 is already determined and therefore the order is the same for N and N'. We continue with the case that side E_2 has more than one leaf. Again consider any quarnet containing p and s. Then s is always the first leaf on side E_2 viewing from A or B in the quarnet. Suppose this is not the case for N'. Then s is between two other leaves on side E_2. This is only possible if side E_2 has more than two leaves. Then s has also to be between two other leaves on side E_2 in a quarnet, which is a contradiction since $Q(N) = Q(N')$. So s is always the first leaf on side E_2 viewing from A or B in N'. Note that this especially becomes clear from the considered quarnets that have two or, if possible, three leaves on side E_2.

Let q be the second leaf on side E_1 and t the second leaf on side E_2, both viewing from A in N. Consider the quarnet with p, q, s and t. Then p and s are both first leaves viewing from A or both first leaves viewing from B on, respectively, side E_1 and E_2. As p is the first leaf on side E_1 viewing from A in N', s is also the first leaf on side E_2 viewing from A in N'.

If side E_2 has one leaf or two leaves, then the order of the leaves on side E_2 is already fixed as s is fixed. Now assume that side E_2 has more than two leaves. Let v and w be any two leaves on side E_2, not s. Consider the quarnet with p, s, v and w. Since s is fixed, the order of v and w on side E_2 can be determined. If v is nearer to s than w on side E_2 in the quarnet, then this is also the case in N'. If w is nearer to s than v on side E_2 in the quarnet, then this is also the case in N'. As this can be done for all the different pairs of leaves, the order of the leaves on side E_2 is the same for N and N'.

For side E_3 we can do the same as for side E_2 to show that the leaves are in the same order viewing from A for N and N'. Note that the only difference is that side E_3 can also have no leaves, but then the order of the leaves is always the same for N and N'.

Now the order of the leaves on each of the sides of the generator in the same for N and N'. Also the order of the leaves on each side with respect to the different sides is the same for N and N'. Hence we have showed that $N = N'$. □

We conclude this section by combining the results for strict level-1 and strict level-2 networks to obtain the result for level-2 networks as stated in the theorem below.

Theorem 3.3. *The class of undirected binary simple level-2 networks with at least four leaves is weakly encoded by quarnets.*

Proof. Assume there are two recoverable phylogenetic networks N and N' in the class of undirected binary simple level-2 networks with at least four leaves such that $Qn(N) = Qn(N')$. To prove that the class is weakly encoded by quarnets we have to show that $N = N'$.

By definition of the class the networks in this class are level-1 or level-2. Note that there are no level-0 networks in the class. Level-0 networks are networks without reticulations, also known as trees. Each edge in a tree is a cut-edge. If a tree has more than two leaves, then it has an cut-edge that not ends in a leaf. This implies a level-0 network with at least four leaves can not be simple.

Suppose N and N' are not both level-1 or level-2. Without loss of generality, say N is a level-1 network and N' is a level-2 network. The underlying generator of N' is the level-2 generator (see Figure 3.3). Let x and y be two leaves on two different sides of the generator in N'. Note that there exist at least two sides with a leaf on it since N' is level-2 and there are no parallel edges in N' by the definition of a network. Consider a quarnet of N that contains x and y. The quarnet has the same underlying generator as N' and is therefore level-2. This is a contradiction as N is level-1 and $Qn(N) = Qn(N')$. So N and N' are both level-1 or level-2.

If N and N' are both level-1, then by Theorem 3.1 we have that $N = N'$. If N and N' are both level-2 networks, then by Theorem 3.2 we have that $N = N'$. So in both cases we can conclude that $N = N'$. □

3.2 Semi-directed

In this section we prove that the class of semi-directed binary simple level-2 networks is weakly encoded by quarnets. We will show this by first prove it for strict level-1 and strict level-2 networks separately. In the theorem below we prove the result for semi-directed strict level-1 networks.

Theorem 3.4. *The class of semi-directed binary simple strict level-1 networks with at least four leaves is weakly encoded by quarnets.*

Proof. Assume there are two recoverable phylogenetic networks N and N' in the class of semi-directed binary simple strict level-1 networks with at least four leaves such that $Qn(N) = Qn(N')$. To prove that the class is weakly encoded by quarnets we have to show that $N = N'$.

By definition of the class the networks are all of the same form, which is given in Figure 3.5. Therefore we only have to show that the leaf below the reticulation and the circular order of the remaining leaves are the same for N and N'.

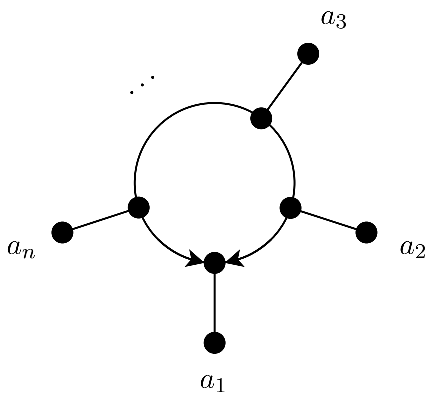


Figure 3.5: The form of a semi-directed binary simple strict level-1 network.

Let x be the leaf below the reticulation vertex in N. Let y be one of the two neighbours of x in N. Consider any quarnet containing x and y. This quarnet is level-1 and is therefore of the same form as N and N'. This holds also for the quarnets that will be considered later in the proof. Now x is also the leaf below the reticulation vertex in the quarnet. Since the direction of the order of the remaining leaves is the only symmetry, it follows that x is the leaf below the reticulation vertex in N'.

Since y is a neighbour of x in N, y is also a neighbour of x in the quarnet. Suppose y is not a neighbour of x in N'. Then there exists a quarnet with leaves p and q such that the circular order of these four leaves is x, p, y and q. This gives a contradiction since $Qn(N) = Qn(N')$. So y is also a neighbour of x in N'.

Now we can see the circular order of the leaves as starting with x and ending with y. Then the remaining leaves are between x and y in the circular order, as can be seen in Figure 3.6.

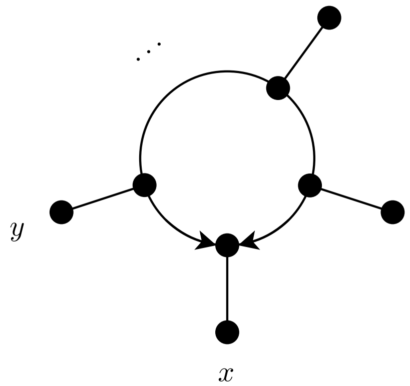


Figure 3.6: The form of a semi-directed binary simple strict level-1 network with two leaves fixed for the start and end of the circular order.

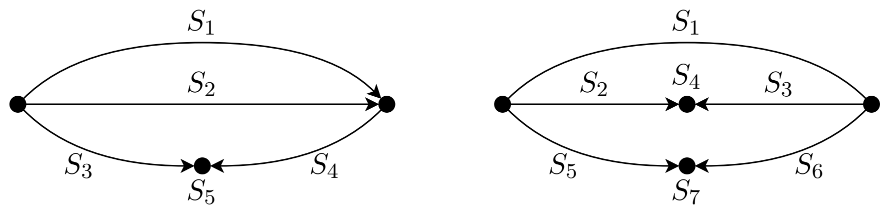
Let v and w be any two leaves, not x or y. Consider the quarnet with x, y, v and w. If v is earlier in the circular order than w in N, then the circular order of the quarnet is x, v, w and y. Suppose w is earlier in the circular order than v in N'. Then the circular order of the considered quarnet is x, w, v and y, which is a contradiction since $Qn(N) = Qn(N')$. Similarly, we get that if w is earlier in the circular order than v in N, w is earlier in the circular order than v in N'. As this can be done for all the different pairs of leaves, the circular order of the leaves is the same for N and N'. Now we can conclude that $N = N'$. □

In the theorem below we prove the wanted result for semi-directed strict level-2 networks. In this proof we use the possible underlying structures of such a network, the two semi-directed level-2 generators as given in Figure 2.8. In contrast to undirected strict level-2 networks we have to distinguish the possible underlying generators for semi-directed strict level-2 networks as there are now multiple generators for the class of networks. Note that in the theorem above we did not use the semi-directed level-1 generator (see Figure 2.7) as underlying structure explicitly since the two different arc sides can not be distinguished. Therefore we did not look at the order of the leaves on these sides separately, but at the order of the leaves on these two sides together. In this way we were still able to obtain the wanted result for level-1.

Theorem 3.5. *The class of semi-directed binary simple strict level-2 networks with at least four leaves is weakly encoded by quarnets.*

Proof. Assume there are two recoverable phylogenetic networks N and N' in the class of semi-directed binary simple strict level-2 networks with at least four leaves such that $Qn(N) = Qn(N')$. To prove that the class is weakly encoded by quarnets we have to show that $N = N'$.

By definition of the class the networks in this class can have two different level-2 generators. In Figure 3.7 the two generators can be find with labelled sides.



(a) Generator 2.1.

(b) Generator 2.2.

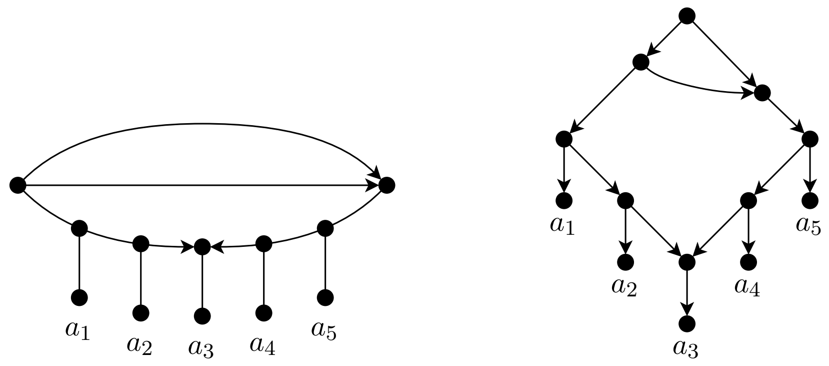
Figure 3.7: The two level-2 generators of semi-directed binary networks with labelled sides.

If N has underlying generator 2.1, consider any quarnet of N with a leaf on the reticulation side S_5. The quarnet is a simple strict level-2 network with underlying generator 2.1. This quarnet is also a quarnet of N' since $Qn(N) = Qn(N')$. This implies N' has underlying generator 2.1 as N' and the quarnet are both simple level-2 networks. If N has underlying generator 2.2, consider any quarnet of N with two leaves on the reticulation sides S_4 and S_7. In the same way we get that N' has underlying generator 2.2. So N and N' have the same underlying generator. In the remainder of the proof we show for each of the generators that $N = N'$.

Generator 2.1

Observe that generator 2.1 has symmetry. Sides S_1 and S_2 can be interchanged with each other to obtain an isomorphic generator. First consider the case that there are no leaves on sides S_1 and S_2 in N (see Figure 3.8). Note that in this case N has parallel arcs. Since any directed binary network that is an orientation of the semi-directed binary network N can not have any parallel arcs by definition, the root has to be on side S_1 or S_2 in N. Note that it does not matter if the root is on side S_1 or S_2 in N because of the symmetry. Since the root of N is known, the directions of all edges are known. This implies there is only one possible directed binary network N_D that is an orientation of N. Since N has a pair of parallel arcs, each $q \in Q(N)$ has also parallel arcs by definition. Since $Qn(N) = Qn(N')$ and each $q \in Qn(N)$ has parallel arcs, N' has also parallel arcs by definition of an exhibited quarnet. In the same way as for N, this implies that there is only one possible directed binary network N'_D that is an orientation of N'.

Recall that each $q \in Q(N)$ has parallel arcs. Similar as for N, we now get that q_D is known for each $q \in Q(N)$. In other words, for each four leaves we have a quarnet in $Qn(N)$ that corresponds to a quarnet in $Qn(N_D)$. Since $Qn(N) = Qn(N')$, we have $Qn(N_D) = Qn(N'_D)$. We get by Corollary 1 of [14] that $N_D = N'_D$ since N'_D is recoverable as N' is simple. This implies $N = N'$ as wanted.



(a) N. (b) N_D.

Figure 3.8: An example of a semi-directed binary network N with underlying generator 2.1 that has a unique orientation N_D.

Now we assume that there is at least one leaf on side S_1 or S_2 in N. Let x be the leaf on the reticulation side S_5 in N. Consider any quarnet containing x. The quarnet is level-2 and has therefore the same underlying generator as N and N'. Now x is also the leaf on the reticulation side S_5 in the quarnet. Since generator 2.1 has no symmetry which involves side S_5, it follows that x is also the leaf on the reticulation side S_5 in N'.

We assumed that there is at least one leaf on side S_1 or S_2 in N. Assume without loss of generality that this leaf is on side S_1 in N. Let p be a leaf on side S_1 in N. Let q be any other leaf on any side S_i, not side S_5, in N. Consider a quarnet containing x, p and q. As before we have that the quarnet is level-2 with underlying generator 2.1. The quarnet implies that p is on side S_1 or S_2 in N'. If p is on side S_2 in N', sides S_1 and S_2 have to be interchanged for N'. Note that this is allowed because of the symmetry of the generator. Since the symmetry of the generator is now fixed and $Qn(N) = Qn(N')$, using the same quarnet we have that q is a leaf on side S_i in N'. We can do this for each leaf q on sides S_1, S_2, S_3 and S_4. Now N and N' have the same leaves on each side of the generator.

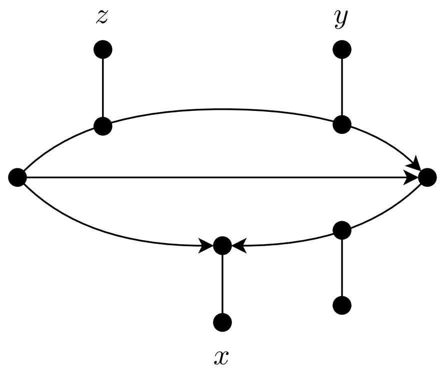


Figure 3.9: A possible quarnet that can be used for a network to determine the order of the leaves y and z on side S_1. In this case y is below z on arc side S_1.

Let S_i be any side with at least two leaves. Note that S_i can be S_1, S_2, S_3 or S_4. Let y and z be two leaves on side S_i. Consider a quarnet containing x, y and z. Again we have that the quarnet is level-2 with underlying generator 2.1. The order of y and z on side S_i can be determined as the side is an arc side (see Figure 3.9). If y is below z on arc side S_i in N, then this holds also for the quarnet. Since $Qn(N) = Qn(N')$, y is also below z on arc side S_i in N'. Similarly, we get that if z is below y on arc side S_i in N, y is also below z on arc side S_i in N'. So the order of the leaves on the different sides is the same for N and N'. Since N and N' have on each side the same leaves in the same order, we can conclude $N = N'$.

Generator 2.2

Observe that generator 2.2 has some symmetry. Sides S_2, S_3 and S_4 can be interchanged with sides S_5, S_6 and S_7, respectively, to obtain an isomorphic generator. Sides S_2 and S_5 can be interchanged with sides S_3 and S_6, respectively, again yielding an isomorphic generator. Note that for the last mentioned symmetry the order of the leaves on the edge side S_1 turns around.

Let x be a leaf on side S_4 and y a leaf on side S_7 in N. Consider any quarnet containing x and y. The quarnet is level-2 and has therefore the same underlying generator as N and N'. The quarnet implies that x and y are on sides S_4 and S_7 in N'. Assume without loss of generality that x is on side S_4 and y on side S_7 in N'.

Let z be a leaf on any side S_i, not side S_4 or S_7, in N. Consider a quarnet containing x, y and z. As before we have that the quarnet is level-2 with underlying generator 2.2. Now we get some results from the quarnet using the symmetry and $Qn(N) = Qn(N')$. If z is a leaf on side S_1 in N, then z is a leaf on side S_1 in N'. If z is a leaf on side S_2 or S_3 in N, then z is a leaf on side S_2 or S_3 in N'. If z is a leaf on side S_5 or S_6 in N, then z is a leaf on side S_5 or S_6 in N'.

First we consider the case that there is at least one leaf on side S_2, S_3, S_5 or S_6 in N. Let a be such a leaf. Without loss of generality, assume that a is on the same side in N' as in N. Let b a leaf on side $s \in \{S_2, S_3, S_5, S_6\}$, not a, in N. Consider the quarnet with x, y, a and b. Again we have that the quarnet is level-2 with underlying generator 2.2. Since a is fixed, b is now on side s in N'. So N and N' have on each side the same leaves.

Let p and q be two leaves that are both on side S_2, S_3, S_5 of S_6 in N. Consider the quarnet with x, y, p and q. Again the quarnet is level-2 with underlying generator 2.2. Using this quarnet we get that p and q are in the same order in N' as in N.

If side S_1 has no leaves or just one leaf, the order on side S_1 is already determined. For the case that side S_1 has more than one leaf, the order of the leaves have to determined. Let k be such a leaf on side S_2, S_3, S_5 or S_6. Let l be the first leaf on side S_1 viewing from side S_2 in N. Consider the quarnet with x, y, k and l. Again the quarnet is level-2 with underlying generator 2.2. Using the quarnet l is also the first leaf on side S_1 viewing from side S_2 in N' since the symmetry is already fixed.

If side S_1 has two leaves, the order is now determined. For the case that side S_1 has more than two leaves, we continue with the order of the leaves. Let m and n be two leaves on side S_1, not l. Consider the quarnet with x, l, m and n. Note that this quarnet is level-1. Since l is fixed, the order of m and n on side S_1 can be determined. If m is a neighbour of l in the quarnet, then m is nearer to l than n on side S_1 in N'. If n is a neighbour of l in the quarnet, then n is nearer to l than m on side S_1 in N'. As this can be done for all leaves, the order of the leaves on side S_1 is determined.

Now we consider the case that there are no leaves on sides S_2, S_3, S_5 and S_6 in N. Then there are at least two leaves on side S_1. If side S_1 has two leaves, the order of the two leaves on side S_1 does not change the network and we are done. For the case that S_1 has more than two leaves, we continue with the order of the leaves. Let r be the first leaf on either end of side S_1 in N. Consider a quarnet containing x, y and r. Again the quarnet is level-2 with underlying generator 2.2. Using the quarnet r is also the first leaf on either end of side S_1 in N'. Note that it does not matter which first leaf it is as the order of the leaves on side S_1 can be turned around without changing the network.

Let s and t be two leaves on side S_1, not r. Consider the quarnet with x, r, s and t. Note that this quarnet is level-1. Since r is fixed, the order of s and t on side S_1 can be determined. If s is a neighbour of r in the quarnet, then s is nearer to r than t on side S_1 in N'. If t is a neighbour of r in the quarnet, then t is nearer to r than s on side S_1 in N'. As this can be done for all leaves, the order of the leaves on side S_1 is determined.

For all different cases N and N' have on each side the same leaves in the same order. Therefore we can conclude $N = N'$. □

In the proof above we had to deal with networks that have parallel arcs. We did not have these parallel arcs in the previous proofs of undirected and semi-directed networks. Now we had a semi-directed network with parallel arcs if the semi-directed network has underlying generator 2.1 and has no leaves on sides S_1 and S_2 of the generator. Since we then know the root and therefore also the directed network from which the semi-directed network is obtained from, we use this unique orientation to prove the case of semi-directed networks with parallel arcs. But note that the proof of this case could also be shown in the same way as the other cases by proving that the leaves are on the same sides of the generator using the quarnets. If we prove the case of parallel arcs in the same way as the other cases, the orientation of the network is no longer needed and only semi-directed networks are considered. We preferred the way of proving using directed networks as then the proof uses that the directions in the whole network are known.

If the network N has underlying generator 2.1, we only need quarnets with at most three leaves fixed in the proof. Therefore the result for the semi-directed binary generator 2.1 can also be proved by using trinets instead of quarnets. This does not hold for the other considered generators in this chapter.

We conclude this section by combining the results for strict level-1 and strict level-2 network to obtain the result for level-2 networks as stated in the theorem below.

Theorem 3.6. *The class of semi-directed binary simple level-2 networks with at least four leaves is weakly encoded by quarnets.*

Proof. Assume there are two recoverable phylogenetic networks N and N' in the class of semi-directed binary simple level-2 networks with at least four leaves such that $Qn(N) = Qn(N')$. To prove that the class is weakly encoded by quarnets we have to show that $N = N'$.

By definition of the class the networks in this class are level-1 or level-2. Note that there are no level-0 networks in the class. Level-0 networks are networks without reticulations, also known as trees. Since level-0 networks have no reticulations, semi-directed level-0 networks have no arcs, only edges. Each edge in a tree is a cut-edge. If a tree has more than two leaves, then it has an cut-edge that not ends in a leaf. This implies a level-0 network with at least four leaves can not be simple.

Suppose N and N' are not both level-1 or level-2. Without loss of generality, say N is a level-1 network and N' is a level-2 network. The underlying generator of N' is level-2 generator 2.1 or 2.2 (see Figure 3.7). Generator 2.1 has one reticulation side and generator 2.2 has two reticulation sides. If the underlying generator of N' is 2.1, let x be the leaf on the reticulation side S_5 in N'. If the underlying of N' is 2.2, let x and y be the leaves on the reticulation sides S_4 and S_7 in N'. Consider a quarnet of N that contains x (and y). The quarnet has the same underlying generator as N' and is therefore level-2. This is a contradiction as N is level-1 and $Qn(N) = Qn(N')$. So N and N' are both level-1 or level-2.

If N and N' are both level-1, then by Theorem 3.4 we have that $N = N'$. If N and N' are both level-2 networks, then by Theorem 3.5 we have that $N = N'$. So in both cases we can conclude that $N = N'$. □

Chapter 4

Decomposition theorems

In this chapter we will prove two decomposition theorems for undirected and semi-directed binary networks. The first theorem considers CE-splits and the second theorem considers restrictions.

4.1 Undirected

In this section we prove the two decomposition theorems for undirected binary networks. For the first theorem, which is about CE-splits, we first give some results. In the observation below the relation of CE-splits between a network and its exhibited networks, for example its quarnets, is discussed.

Observation 4.1. *Let N be an undirected binary network on X. Let $\{A, B\}$ a partition of X and $C \subseteq X$ such that $A \cap C \neq \emptyset$ and $B \cap C \neq \emptyset$. Denote N' as the undirected binary network on C exhibited by N. If $\{A, B\}$ is a CE-split of N, then $\{A \cap C, B \cap C\}$ is a CE-split of N'.*

In contrast to semi-directed networks undirected networks do not always have an orientation (see Figure 4.1).

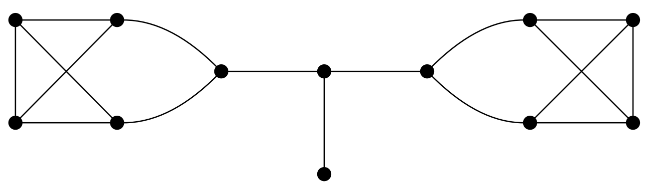


Figure 4.1: An example of an undirected binary network that has no orientation. Both nontrivial biconnected components are strongly redundant. Since these components are not connected to a leaf, they must both contain the root in order to have an orientation of the network, which is not possible.

Following lemma implies that if we add the condition recoverable to the undirected binary network, it is guaranteed that there is an orientation of the network. This lemma is therefore used in Observations 4.3 and 4.4. Note that it is even guaranteed that there is an orientation at a leaf.

Lemma 4.2 (Lemma 2.69 in [8])**.** *Let N be an undirected binary network on X. Then N is orientable at any leaf $\rho \in X$ if and only if N is recoverable.*

Above theorem considers orientations at a leaf. If an undirected network is not recoverable, it can still have an orientation (see Figure 4.2).

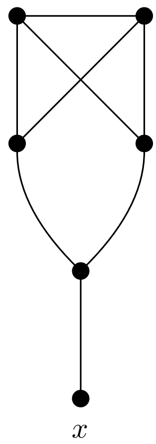
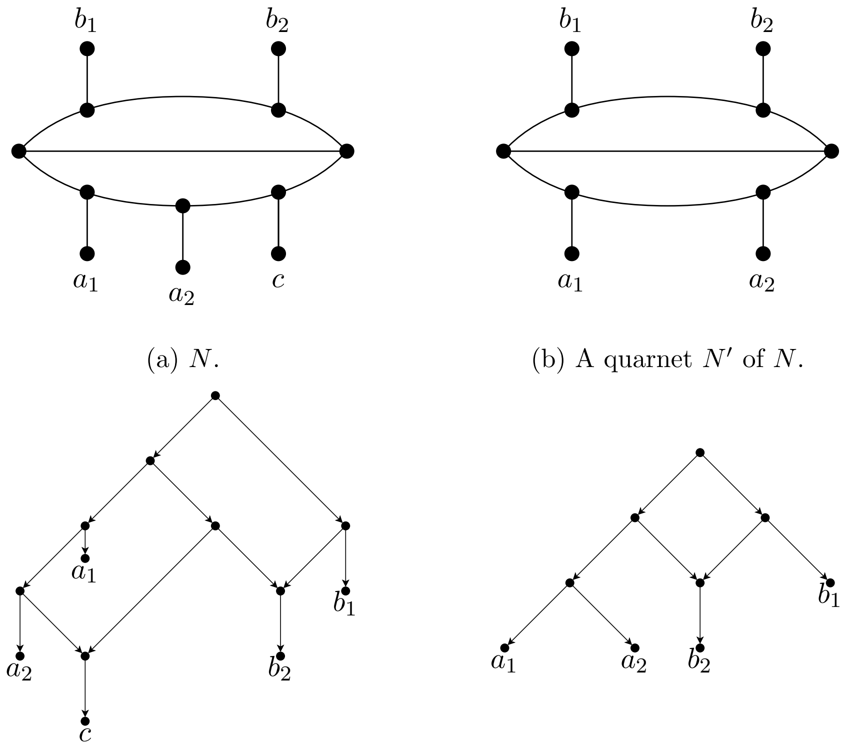


Figure 4.2: An example of an undirected binary network that is not redundant, but has an orientation. In an orientation of the network the root is in the nontrivial biconnected component. For example, the root can be placed on the upper edge of the network.

The following observation describes the relationship between the CE-splits of an undirected network and the CA-splits of any of its orientations.

Observation 4.3. *Let N be a recoverable undirected binary network on X, N_D any directed binary network that is an orientation of N (N_D exists by Lemma 4.2) and $\{A, B\}$ a partition of X. Then $\{A, B\}$ is a CE-split of N if and only if $\{A, B\}$ is a CA-split of N_D.*

For semi-directed networks the choice for an orientation does not have any influence on its exhibited networks (see Observation 4.8). This is different for undirected networks as it can make a difference.



(a) N.

(b) A quarnet N' of N.

(c) An orientation N_D of N.

(d) A quarnet N'_D of N_D.

Figure 4.3: An undirected binary network N with an orientation N_D for which the exhibited quarnets N' and N'_D on $\{a_1, a_2, b_1, b_2\}$ holds that $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CA-split of N'_D but not a CE-split of N'.

Figure 4.3 shows that not both implications of Observation 4.3 hold for a quarnet of an undirected network and a quarnet of any orientation of the undirected network. Therefore we get the following observation for exhibited networks.

Observation 4.4. *Let N be a recoverable undirected binary network on X and N_D any directed binary network that is an orientation of N (N_D exists by Lemma 4.2). Let $C \subseteq X$ and $\{A, B\}$ a partition of C. Denote N' as the undirected binary network on C exhibited by N and N'_D as the directed binary network on C exhibited by N_D. If $\{A, B\}$ is a CE-split of N', then $\{A, B\}$ is CA-split of N'_D.*

Using the results above we are able to prove the first decomposition theorem for undirected networks, which is given below. We already saw that recoverable undirected binary networks are always orientable at a leaf. Therefore we use this relation with directed networks in the proof. Note that in order to do this we use some results of directed networks.

Theorem 4.5. *Let N be a recoverable undirected binary network on X and $\{A, B\}$ a partition of X.*

(i) If $|A| = 1$ or $|B| = 1$, then $\{A, B\}$ is a CE-split of N.

(ii) Otherwise, if $|A| \geq 2$ and $|B| \geq 2$, then $\{A, B\}$ is a CE-split of N if and only if for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N.

Proof. Let N be a recoverable undirected binary network on X and $\{A, B\}$ a partition of X.

(i) Assume $|A| = 1$ or $|B| = 1$. Since each leaf of an undirected binary network is connected to a cut-egde, $\{A, B\}$ is a CE-split of N.

(ii) Assume $|A| \geq 2$ and $|B| \geq 2$.

" $\implies$ " Assume $\{A, B\}$ is a CE-split of N. Let $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$. There exists a unique quarnet Q on $\{a_1, a_2, b_1, b_2\}$ in $Qn(N)$. Now by Observation 4.1 $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of Q.

" $\impliedby$ " Assume that for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N. Let $\rho \in X$. By Lemma 4.2 N is orientable at ρ. Let N_D be the directed binary network that is the orientation of N that corresponds to orienting N at ρ. Note that $Qn(N_D)$ is the set of quarnets that can be exhibited by N_D as defined in [11] or [14].

Since $\{A, B\}$ is a partition of X, we have $\rho \in A$ or $\rho \in B$. Without loss of generality, we can say that $\rho \in A$. The leaves in $X \setminus \{\rho\}$ are below a different cut-arc leaving the root of N_D than leaf ρ. Therefore ρ is not in a CA-set of N_D that contains at least one other $x \in X$. This implies that A is not a CA-set of N_D. Further observe that with the same reasoning we get that for each quarnet $q_D \in Qn(N_D)$ there exists no CA-set that contains ρ and at least one other $x \in X$.

By Observation 4.4 our assumption implies that for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CA-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N_D. Choosing $a_2 = \rho$ implies that for all $a_1 \in A$ with $a_1 \neq \rho$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, \rho\}, \{b_1, b_2\}\}$ is a CA-split of the quarnet on $\{a_1, \rho, b_1, b_2\}$ exhibited by N_D. Then $\{a_1, \rho\}$ or $\{b_1, b_2\}$ is a CA-set of the quarnet on $\{a_1, \rho, b_1, b_2\}$ exhibited by N_D. Earlier we saw that $\{a_1, \rho\}$ is not a CA-set. So $\{b_1, b_2\}$ is a CA-set of the quarnet on $\{a_1, \rho, b_1, b_2\}$ exhibited by N_D.

Now, by Observation 4 of [14], $\{b_1, b_2\}$ is a CA-set of the trinet on $\{a_1, b_1, b_2\}$ exhibited by N_D, where $a_1 \in A \setminus \{\rho\}$. Again by Observation 4 of [14], $\{b_1, b_2\}$ is a CA-set of the trinet on $\{\rho, b_1, b_2\}$ exhibited by N_D. Combining these results gives that for all $a \in A$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{b_1, b_2\}$ is a CA-set of the trinet on $\{a, b_1, b_2\}$ exhibited by N_D.

Then, by Theorem 1 of [14], B is a CA-set of N_D. So $\{A, B\}$ is a CA-split of N_D. By Observation 4.3 $\{A, B\}$ is a CE-split of N as wanted. □

Now we have proved the first decomposition theorem for undirected binary networks, we continue with the second decomposition theorem. The proof of this theorem is based on the proof of Theorem 2 in [14] for directed binary networks. The structure of the proof is therefore very similar. The main difference is that we now make an extra assumption, namely that the network has no redundant biconnected component.

Theorem 4.6. *A recoverable undirected binary network N on X, with $|X| \geq 4$ and no redundant biconnected components, is encoded by its quarnets $Qn(N)$ if and only if, for each nontrivial biconnected component B of N with at least five incident cut-edges, N_B is encoded by $Qn(N_B)$.*

Proof. Let N be a recoverable undirected binary network on X with $|X| \geq 4$ and no redundant biconnected components.

" $\implies$ " Assume N is encoded by its quarnets $Qn(N)$. Consider any nontrivial biconnected component B of N with at least five incident cut-edges. Suppose that N_B is not encoded by $Qn(N_B)$. Then there exists a recoverable undirected binary network $N'_B \neq N_B$ such that $Qn(N_B) = Qn(N'_B)$. By Theorem 4.5 N'_B has the same CE-splits as N_B. By definition of N_B this implies that each CE-split of N'_B has one set that is a singleton.

Suppose N'_B has a redundant biconnected component R. Note that R has exactly two incident cut-edges. Since each CE-split of N'_B has one set that is a singleton, one incident cut-edge of R leads to a single leaf x and the other incident cut-edge of R leads to the other leaves. Now each quarnet containing x also contains R. This is not possible because $Qn(N_B) = Qn(N'_B)$ and an undirected binary simple network without redundant biconnected components has no quarnets with redundant biconnected components. This shows that N'_B has no redundant biconnected components.

If we combine that N'_B has no redundant biconnected components and that all CE-splits of N'_B have one set that is a singleton, we get that N'_B consists of one nontrivial biconnected component with incident cut-edges to the leaves. So N'_B is a simple network.

Let B' be the nontrivial biconnected component of N'_B. Let N' be the result of replacing B by B' in N. Note that N' is recoverable. We will show that $Qn(N) = Qn(N')$.

Let $Q \in Qn(N)$ and let w, x, y and z be the leaves of Q. If the incident cut-edges of B that are leading to w, x, y and z are all different or all the same, then we have that $Q \in Qn(N')$ since the only difference between N and N' is that B is replaced by B' and that $Qn(N_B) = Qn(N'_B)$.

Now suppose the incident cut-edges of B that are leading to w, x, y and z are not all different and also not all the same. Let A be a maximal subset of $\{w, x, y, z\}$ such that each incident cut-edge of B is leading to at most one leaf in A. Note that $|A| = 2$ or $|A| = 3$. Since B has at least five incident cut-edges, we can find one leaf (if $|A| = 3$) or two leaves (if $|A| = 2$) such that together with A we have four leaves which corresponds to four different incident cut-edges of B. Since $Qn(N_B) = Qn(N'_B)$ we now have that the subnetworks on A exhibited by N and N' are the same. Now the quarnet on w, x, y and z is the same for N and N' since the only differences are in the parts corresponding to incident cut-edges of B and these parts outside B are the same for N and N'. Now we have showed that $Qn(N) = Qn(N')$. This contradicts the assumption that N is encoded by $Qn(N)$.

" $\Longleftarrow$ " Assume that for each nontrivial biconnected component B of N with at least five incident cut-edges, N_B is encoded by $Qn(N_B)$. Suppose that N is not encoded by $Qn(N)$. Then there exists a recoverable undirected binary network $N' \neq N$ with $Qn(N) = Qn(N')$. We will show that $N = N'$.

Suppose N' has a redundant biconnected component R. If R is connected to other redundant biconnected components, we redefine R as the maximum set of connected redundant biconnected components that contains the redundant biconnected component R. The two incident cut-edges that connect R with the rest of N' we then call the incident cut-edges of R. By Theorem 4.5 N' has the same CE-splits as N. Since R has exactly two incident cut-edges, R is always connected to two other biconnected components by its incident cut-edges. If one of the incident cut-edges of R is connected to a leaf, then with a similar reasoning as we used for N_B and N'_B in the proof for the other implication of the theorem we get that N and N' have no redundant biconnected components. Now if both incident cut-edges of R are not connected to a leaf, then both incident cut-edges of R are connected to a biconnected component with at least two incident cut-edges besides the incident cut-edge connected to R. So both incident cut-edges of R are leading to at least two leaves. Consider the quarnet on $\{w, x, y, z\}$ such that one incident cut-edge of R leads to w and x and the other incident cut-edges of R leads to y and z. Then this quarnet contains R. This is a contradiction since $Qn(N) = Qn(N')$ and N has no redundant biconnected components. So in both cases we get that N' has no redundant biconnected components.

At this point, we observe that, for a biconnected component B with exactly four incident cut-edges, N_B is trivially encoded by $Qn(N_B)$, since in that case N_B is isomorphic to the single quarnet in $Qn(N_B)$.

The rest of the proof is by induction on $|X|$. If $|X| = 4$, then, since N and N' are recoverable, N and N' are both equal to the single quarnet in $Qn(N)$ and we are done.

Now assume $|X| \geq 5$. Let B_0 an arbitrary biconnected component of N. Let $e_1 = u_1v_1, \ldots, e_b = u_bv_b$ be the incident cut-edges of B_0, where $u_1, \ldots, u_b$ are in B_0. Note that $b \geq 3$ since N is recoverable and has no redundant biconnected components.

Let $N_1, \ldots, N_b$ be the connected components after deleting B_0 and new leaves $p_1, \ldots, p_b$ with edges $v_i p_i$ for $1 \leq i \leq b$, such that N_i contains v_i. We call X_i the set of leaves of N_i for $1 \leq i \leq b$. Then, since $b \geq 3$, we have $|X_i| < |X|$. Note that N_i is recoverable and has no redundant biconnected components for $1 \leq i \leq b$ since N has no redundant biconnected components.

We saw that N' has the same CE-splits as N. Thus, $\{X_i, X \setminus \{X_i\}\}$ is a CE-split of N' for $1 \leq i \leq b$. Now, since a tree is uniquely defined by its splits ([2]), viewing the biconnected components of N and N' as vertices of two trees shows that N' has a biconnected component B'_0 with b incident cut-edges for which the CE-splits agree with the CE-splits of the incident cut-edges of B_0. Let $e'_1 = u'_1 v'_1, \ldots, e'_b = u'_b, v'_b$ be the incident cut-edges of B'_0, where $u'_1, \ldots, u'_b$ are in B'_0. Let $N'_1, \ldots, N'_b$ be the connected components after deleting B'_0 and new leaves $p'_1, \ldots, p'_b$ with edges $v'_i p'_i$ for $1 \leq i \leq b$, such that N'_i contains v'_i. Assume without loss of generality that N'_i is a network on X_i for $1 \leq i \leq b$. Note that we can choose the new leaves $p'_1, \ldots, p'_b$ such that $p'_i = p_i$ holds for $1 \leq i \leq b$. Further note that N'_i is recoverable and has no redundant biconnected components for $1 \leq i \leq b$ since N' has no redundant biconnected components.

Now we look at the quarnets of N_i and N'_i. Let $i \in \{1, \ldots, b\}$. Let l_1, l_2, l_3 and l_4 be four different leaves of N_i. If p_i is not of the four leaves, then the quarnet exhibited on $\{l_1, l_2, l_3, l_4\}$ by N_i is also a quarnet of N and therefore known. If p_i is one of the four leaves, say l_4, then we can obtain the quarnet exhibited on $\{l_1, l_2, l_3, p_i\}$ by N_i in the following way. Let $p \in X$ a leaf that is not in N_i. Let Q be the quarnet exhibited on $\{l_1, l_2, l_3, p\}$ by N and subsequently suppressing all the redundant biconnected components that are connected to p. Then Q is the wanted quarnet of N_i. For N'_i we can obtain the quarnets in the same way. So $Qn(N_i)$ and $Qn(N'_i)$ are known. To show that $N = N'$, it remains to show that $N_{B_0} = N_{B'_0}$ and that $N_i = N'_i$ for $1 \leq i \leq b$.

First, we show that $N_{B_0} = N_{B'_0}$. Observe that $Qn(N_{B_0}) = Qn(N_{B'_0})$ (if for any four leaves y_j, y_k, y_l, y_m the quarnet in $Qn(N_{B_0})$ and the quarnet in $Qn(N_{B'_0})$ would be different, then for any four leaves x_j, x_k, x_l, x_m in the parts of the network corresponding to incident cut-edges e_j, e_k, e_l, e_m, respectively, the quarnet in $Qn(N)$ and the quarnet in $Qn(N')$ would be different). If $b \geq 5$, then $N_{B_0} = N_{B'_0}$ holds because $Qn(N_{B_0}) = Qn(N_{B'_0})$ and by assumption N_{B_0} is encoded by $Qn(N_{B_0})$. Moreover, $b \geq 3$ since N is recoverable and has no redundant biconnected components. For $b = 4$ the statement $N_{B_0} = N_{B'_0}$ is trivially true.

The only case left is $b = 3$. Consider a leaf in each of the three parts corresponding to the three incident cut-edges of B_0, namely leaves l_1, l_2 and l_3. Since $|X| \geq 4$, we can find another leaf l_4 of N. Consider the quarnet Q in $Qn(N)$ on $\{l_1, l_2, l_3, l_4\}$. Since $Qn(N) = Qn(N')$ we now get that the trinets on $\{l_1, l_2, l_3\}$ exhibited by N and N', respectively, are the same. Let $B_0(T)$ be the biconnected component of this trinet T such that the component has three incident cut-edges that lead to l_1, l_2 and l_3, respectively. Then we have that $N_{B_0(T)} = N_{B_0}$. Moreover, since N' also exhibits T, we have that $N_{B_0(T)} = N_{B'_0}$. It follows that $N_{B_0} = N_{B'_0}$.

Now let $i \in \{1, \ldots, b\}$. We will show that $N_i = N'_i$. Observe that $Qn(N_i) = Qn(N'_i)$ for similar reasons as we used for showing that $Qn(N_{B_0}) = Qn(N_{B'_0})$. Since $|X_i| < |X|$, the statement $N_i = N'_i$ follows by induction if (a) N_i and N'_i are recoverable and have no redundant biconnected components and (b) $|X_i| \geq 4$. Note that (a) holds since N and N' have no redundant biconnected components. If $|X_i| = 1$, then clearly $N_i = N'_i$ because both consist of a single leaf. The only cases left are $|X_i| = 2$ and $|X_i| = 3$.

First we consider the case that $|X_i| = 2$. Denote the two leaves of X_i with l_1 and l_2. Since $|X| \geq 4$, we can find two other leaves l_3 and l_4 of N. Consider the quarnet Q in $Qn(N)$ on $\{l_1, l_2, l_3, l_4\}$. Since $Qn(N) = Qn(N')$ we now get that the binets on $\{l_1, l_2\}$ exhibited by N and N', respectively, are the same. So $N_i = N'_i$ holds.

Now we consider the case that $|X_i| = 3$. Denote the three leaves of X_i with l_1, l_2 and l_3. Since $|X| \geq 4$, we can find another leaf l_4 of N. Consider the quarnet Q in $Qn(N)$ on $\{l_1, l_2, l_3, l_4\}$. Since $Qn(N) = Qn(N')$ we now get that the trinets on $\{l_1, l_2, l_3\}$ exhibited by N and N', respectively, are the same. So $N_i = N'_i$ holds.

Since for all cases $N_{B_0} = N_{B'_0}$ and $N_i = N'_i$ for $1 \leq i \leq b$ holds, we can conclude that $N = N'$, which gives a contradiction. □

4.2 Semi-directed

In this section we prove the two decomposition theorems for semi-directed binary networks. For the first theorem, which is about CE-splits, we first give some results. Recall from Lemma 2.22 that semi-directed binary networks have no cut-arcs and therefore we do not have to consider cut-arcs or CA-splits for semi-directed binary networks. In the observation below the relation of CE-splits between a network and its exhibited networks, for example its quarnets, is discussed.

Observation 4.7. *Let N be a semi-directed binary network on X. Let $\{A, B\}$ a partition of X and $C \subseteq X$ such that $A \cap C \neq \emptyset$ and $B \cap C \neq \emptyset$. Denote N' as the semi-directed binary network on C exhibited by N. If $\{A, B\}$ is a CE-split of N, then $\{A \cap C, B \cap C\}$ is a CE-split of N'.*

In contrast to undirected networks semi-directed networks always have an orientation by definition. Now we do not need the condition recoverable to guarantee this.

The following observation shows that the order of exhibiting a network and obtaining a semi-directed binary network from a directed binary network gives the same semi-directed binary network as result. In other words, the choice for an orientation does not have any influence on its exhibited networks. Note that this was not the case for undirected binary networks.

Observation 4.8. *Let N be a semi-directed binary network on X and N_D any directed binary network that is an orientation of N. Then for all $C \subseteq X$ the semi-directed binary network on C exhibited by N can be obtained from the directed binary network on C exhibited by N_D.*

By how a semi-directed network is constructed from a directed network it is easy to obtain the following observation about the relationship between the CE-splits of a semi-directed network and the CA-splits of any of its orientations. Note that this also holds for the exhibited networks by Observation 4.8.

Observation 4.9. *Let N be a semi-directed binary network on X, N_D any directed binary network that is an orientation of N and $\{A, B\}$ a partition of X. Then $\{A, B\}$ is a CE-split of N if and only if $\{A, B\}$ is a CA-split of N_D.*

Using the results above we are able to prove the first decomposition theorem for semi-directed networks, which is given below. Note that recoverable semi-directed binary networks are not always orientable at a leaf. This differs with recoverable undirected binary networks as they are always orientable at a leaf. Therefore we distinguish for semi-directed networks the case that the network is orientable at a leaf and the case that this does not hold.

Theorem 4.10. *Let N be a recoverable semi-directed binary level-2 network on X and $\{A, B\}$ a partition of X.*

(i) If $|A| = 1$ or $|B| = 1$, then $\{A, B\}$ is a CE-split of N.

(ii) Otherwise, if $|A| \geq 2$ and $|B| \geq 2$, then $\{A, B\}$ is a CE-split of N if and only if for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N.

Proof. Let N be a recoverable semi-directed binary level-2 network on X and $\{A, B\}$ a partition of X.

(i) Assume $|A| = 1$ or $|B| = 1$. Since each leaf of a semi-directed binary network is connected to a cut-edge, $\{A, B\}$ is a CE-split of N.

(ii) Assume $|A| \geq 2$ and $|B| \geq 2$.

"$\Longrightarrow$" Assume $\{A, B\}$ is a CE-split of N. Let $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$. There exists a unique quarnet Q on $\{a_1, a_2, b_1, b_2\}$ in $Qn(N)$. Now by Observation 4.7 $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of Q.

“ $\Longleftarrow$ ” Suppose N is orientable at a leaf $\rho \in X$. Assume that for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N. Let N_D be the directed binary network that is the orientation of N that corresponds to orienting N at ρ. Note that $Qn(N_D)$ is the set of quarnets that can be exhibited by N_D as defined in [11] or [14]. Further note that the quarnets $Qn(N)$ can be obtained from $Qn(N_D)$ by Observation 4.8.

Since $\{A, B\}$ is a partition of X, we have $\rho \in A$ or $\rho \in B$. Without loss of generality, we can say that $\rho \in A$. The leaves in $X \setminus \{\rho\}$ are below a different cut-arc leaving the root of N_D than leaf ρ. Therefore ρ is not in a CA-set of N_D that contains at least one other $x \in X$. This implies that A is not a CA-set of N_D. Further observe that with the same reasoning we get that for each quarnet $q_D \in Qn(N_D)$ there exists no CA-set that contains ρ and at least one other $x \in X$.

By Observation 4.9 our assumption implies that for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CA-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N_D. Choosing $a_2 = \rho$ implies that for all $a_1 \in A$ with $a_1 \neq \rho$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, \rho\}, \{b_1, b_2\}\}$ is a CA-split of the quarnet on $\{a_1, \rho, b_1, b_2\}$ exhibited by N_D. Then $\{a_1, \rho\}$ or $\{b_1, b_2\}$ is a CA-set of the quarnet on $\{a_1, \rho, b_1, b_2\}$ exhibited by N_D. Earlier we saw that $\{a_1, \rho\}$ is not a CA-set. So $\{b_1, b_2\}$ is a CA-set of the quarnet on $\{a_1, \rho, b_1, b_2\}$ exhibited by N_D.

Now, by Observation 4 of [14], $\{b_1, b_2\}$ is a CA-set of the trinet on $\{a_1, b_1, b_2\}$ exhibited by N_D, where $a_1 \in A \setminus \{\rho\}$. Again by Observation 4 of [14], $\{b_1, b_2\}$ is a CA-set of the trinet on $\{\rho, b_1, b_2\}$ exhibited by N_D. Combining these results gives that for all $a \in A$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{b_1, b_2\}$ is a CA-set of the trinet on $\{a, b_1, b_2\}$ exhibited by N_D.

Then, by Theorem 1 of [14], B is a CA-set of N_D. So $\{A, B\}$ is a CA-split of N_D. By Observation 4.9 $\{A, B\}$ is a CE-split of N as wanted.

Now suppose N is not orientable at a leaf $\rho \in X$. Assume that for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N. Let N_D be any directed binary network that is an orientation of N. Note that $Qn(N_D)$ is the set of quarnets that can be exhibited by N_D as defined in [11] or [14]. Further note that the quarnets $Qn(N)$ can be obtained from $Qn(N_D)$ by Observation 4.8.

By Observation 4.9 our assumption implies that for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CA-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N_D.

Denote the biconnected component of N_D that contains the root as B_r. Further denote the number of outgoing cut-arcs of B_r with k. Since N has no strongly redundant components, B_r has at least two outgoing cut-arcs. So we have $k \geq 2$. Since the outgoing cut-arcs of B_r are not ordered, we fix the cut-arcs in an arbitrary order for the prove. We define the following sets for $i = 1, \ldots, k$:

$$A_i = \{a \in A | a \text{ is below the } i\text{-th outgoing cut-arc of } B_r\}$$

$$B_i = \{b \in B | b \text{ is below the } i\text{-th outgoing cut-arc of } B_r\}$$

$$C_i = A_i \cup B_i$$

Note that $\{\cup_{i=1}^k A_i, \cup_{i=1}^k B_i\} = \{A, B\}$ and $\cup_{i=1}^k C_i = X$.

Claim: At least one of the following holds:

(a) there exists $i \in \{1, \ldots, k\}$ such that $A_i = A$,

(b) there exists $j \in \{1, \ldots, k\}$ such that $B_j = B$,

(c) for all $l \in \{1, \ldots, k\}$ holds that $C_l = A_l$ or $C_l = B_l$ (which implies $B_l = \emptyset$ or $A_l = \emptyset$).

Proof of claim: Assume (a) and (b) are both not true. Then there exist $i_1, i_2 \in \{1, \ldots, k\}$ with $i_1 \neq i_2$ such that $A_{i_1} \neq \emptyset$ and $A_{i_2} \neq \emptyset$. There exist also $j_1, j_2 \in \{1, \ldots, k\}$ with $j_1 \neq j_2$ such that $B_{j_1} \neq \emptyset$ and $B_{j_2} \neq \emptyset$. Let $a_1 \in A_{i_1}$, $a_2 \in A_{i_2}$, $b_1 \in B_{j_1}$ and $b_2 \in B_{j_2}$.

Suppose a_1, a_2, b_1 and b_2 are not below four different outgoing cut-arcs of B_r. Without loss of generality, a_1 and b_1 are below the same outgoing cut-arc of B_r. Then a_2 and b_2 are not below this outgoing cut-arc of B_r. Note that then $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is not a CA-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N_D. This gives a contradiction.

So a_1, a_2, b_1 and b_2 are below four different outgoing cut-arcs of B_r. Since $a_1 \in A$ and $b_1 \in B$ were chosen arbitrarily, we now see that for any $a_1 \in A$ and $b_1 \in B$ holds that they are below different outgoing cut-arcs of B_r. This implies that (c) holds, which concludes the proof of the claim.

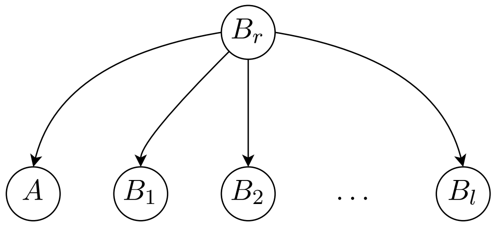


Figure 4.4: B_r with outgoing arcs to different components when (a) (or (b)) of the claim holds.

Assume (a) or (b) of the claim holds (see Figure 4.4). Without loss of generality, we can say there exists $i \in \{1, \ldots, k\}$ such that $A_i = A$. This implies all $a \in A$ are below the same outgoing cut-arc of B_r.

Assume there is no $b \in B$ that is below the same outgoing cut-arc of B_r as the leaves of A. Consider the outgoing cut-arc of B_r that has the leaves of A below it. In other words, we consider the i-th outgoing cut-arc of B_r. This cut-arc implies that A is a CA-set. So $\{A, B\}$ is a CA-split of N_D. By Observation 4.7 $\{A, B\}$ is a CE-split of N.

Now assume there is a $b \in B$ that is below the same outgoing cut-arc of B_r as the leaves of A. Note that there is at least one $b^* \in B$ that is not below the same outgoing cut-arc of B_r as the leaves of A since there are at least two outgoing cut-arcs leaving B_r. Choose such a b^* arbitrary. Further note that B is not a CA-set of N_D since the biconnected component B_r contains no cut-arcs by definition and at least two outgoing cut-arcs of B_r have a leaf $b \in B$ below it. Therefore we need to prove that A is a CA-set of N_D.

To show this we consider a part of N_D. Therefore we define the directed binary network N'_D. We get this network by first replacing the root component B_r of N_D by a single vertex r. Subsequently r has the outgoing arcs (r, v) and (r, b^*), where v comes from the i-th outgoing cut-arc (u, v) of B_r. This implies the part below the i-th outgoing cut-arc of B_r of N_D is also in N'_D.

Note that we can get the quarnets $Qn(N'_D)$ from the corresponding quarnets of N_D. If such a quarnet does not contain b^*, then the quarnet of N_D and N'_D are equal. On the other hand, if such a quarnet contains b^*, we can get the quarnet exhibited by N'_D in the same way from the quarnet exhibited by N_D as we constructed N'_D from N_D. Note that the root component of a quarnet containing b^* can already be a single vertex before it is replaced. Further note that the parts below outgoing cut-arcs of B_r without leaves of A are removed from the network except for b^*.

Now we know the directed binary network N'_D and its quarnets $Qn(N'_D)$. Observe that N'_D has the arc (r, b^*) where r is the root of N'_D. Now using the proof for the case that N is orientable at a leaf $\rho \in X$ gives that A is now a CA-set of N'_D. By how we constructed N'_D from N_D we know that A is also a CA-set of N_D. Then $\{A, B\}$ is a CA-split of N_D. By Observation 4.9 $\{A, B\}$ is now a CE-split of N as wanted.

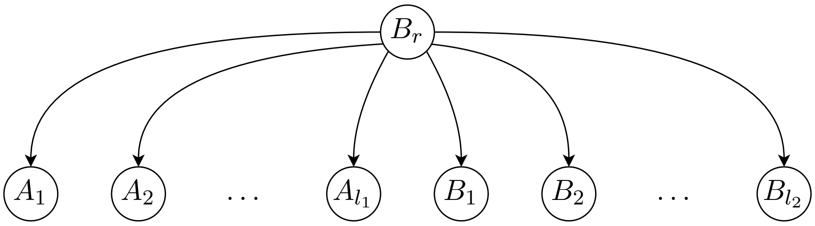


Figure 4.5: B_r with outgoing arcs to different components when (c) of the claim holds.

Now assume (a) and (b) of the claim do not hold. By the claim we know that (c) holds (see Figure 4.5). Now we know that B_r then has at least four outgoing cut-arcs since it has at least two outgoing cut-arcs with leaves of A below it and at least two outgoing cut-arcs with leaves of B below it.

Note that since N is level-2, N_D is also level-2. Therefore B_r has at most 2 reticulations. If the underlying generator of B_r is level-0, B_r is a single vertex. Then B_r has at most two outgoing cut-arcs, which is a contradiction. Now we know that the underlying generator of B_r is the level-1 generator or one of the two level-2 generators 2.1 and 2.2.

First we consider the case that B_r has the level-1 generator as underlying generator. Suppose without loss of generality that the leaves below the reticulation are in B. Let b_1 one of the leaves below the reticulation. Let $b_2 \in B$ a leaf below another outgoing cut-arc of B_r. Let $a_1, a_2 \in A$ be two leaves below two different outgoing cut-arcs of B_r. We already saw that a_1, a_2, b_1 and b_2 are now below four different outgoing cut-arcs of B_r. Consider the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N_D. The quarnet is simple, because there is one reticulation side and we have a leaf below it. This implies that the only cut-arcs are the outgoing cut-arcs to the leaves a_1, a_2, b_1 and b_2. So only the four singletons $\{a_1\}$, $\{a_2\}$, $\{b_1\}$ and $\{b_2\}$ are CA-sets of the quarnet. Therefore $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is not a CA-split of the quarnet. By Observation 4.9 $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is not a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N. This gives a contradiction.

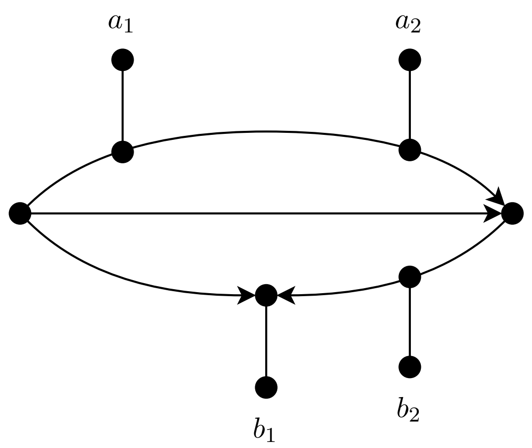


Figure 4.6: A possible quarnet without CE-split $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ that has underlying generator 2.1.

We can deal with the case that B_r has level-2 generator 2.1 as underlying generator in the same way (see Figure 4.6). This is because the level-1 generator and level-2 generator 2.1 have both one reticulation side. Now we consider the case that B_r has level-2 generator 2.2 as underlying generator. Let x_1 be a leaf below the reticulation that corresponds to side S_4 of generator 2.2. Let x_2 be a leaf below the reticulation that corresponds to side S_7 of generator 2.2. Let y and z be two leaves such that x_1, x_2, y and z are below four different outgoing cut-arcs of B_r and that two leaves are in A and two leaves are in B. Consider the quarnet on $\{x_1, x_2, y, z\}$ exhibited by N_D. The quarnet is simple, because there are two reticulation sides and we have a leaf below both of them. This implies that the only cut-arcs are the outgoing cut-arcs to the leaves x_1, x_2, y and z. So only the four singletons $\{x_1\}$, $\{x_2\}$, $\{y\}$ and $\{z\}$ are CA-sets of the quarnet. Therefore $\{\{x_1, x_2, y, z\} \cap A, \{x_1, x_2, y, z\} \cap B\}$ is not a CA-split of the quarnet. Note that we have chosen the leaves in such a way that each part of the partition consists of two leaves. By Observations 4.8 and 4.9 $\{\{x_1, x_2, y, z\} \cap A, \{x_1, x_2, y, z\} \cap B\}$ is not a CE-split of the quarnet on $\{x_1, x_2, y, z\}$ exhibited by N. Since the quarnet has two leaves in A and two leaves in B, this gives a contradiction. □

The decomposition theorem above has another difference with the decomposition theorem for undirected networks. The decomposition theorem for semi-directed networks is only stated for level-2 networks. The level-2 condition is only used for semi-directed binary networks that are not orientable at a leaf. We have used the level-2 condition there only in the proof for the part that (a) and (b) of the claim both do not hold. In other words, we have used there the level-2 condition if only (c) of the claim holds. For the rest of the proof the level-2 condition is not used and is therefore proved for general recoverable semi-directed binary networks. This means, for example, that for semi-directed binary networks that are orientable at a leaf the theorem always holds.

Since the proof for the part that only (c) of the claim holds is the only part of the proof that depends on level-k of the root component B_r, we only have to investigate this part of the second decomposition theorem for $k \geq 3$. For semi-directed binary level-k generators ($k \geq 3$) with one or two reticulation sides the proof follows in the same way as we did in the proof for the level-1 generator or level-2 generator 2.2, respectively. If such a generator has more than two reticulation sides more proof is needed in order to show the wanted result for higher level networks.

If we consider the part that only (c) of the claim holds for a recoverable semi-directed binary level-3 networks, then we know that B_r is level-3. The cases that B_r is level-3 but not strict level-3 we already considered in the proof of the theorem. If B_r is strict level-3, then B_r has one of the 17 semi-directed binary level-3 generators (see Appendix A.2) as underlying generator. These generators have at most three reticulation sides. If the level-3 generator of B_r has one or two reticulation sides, we can deal with this generator in the same way as we did in the proof for the level-1 generator or level-2 generator 2.2, respectively.

The proof for a semi-directed binary level-3 generator with three reticulation sides does not follow easily and is therefore different from the proofs for the generators we already have discussed. If we could assume that N is not orientable at any cut-edge of N, then the rest of the proof for level-3 would be as follows: In the beginning of (c) of the claim we showed that B_r has at least four outgoing cut-arcs. Three of these outgoing cut-arcs are below the three reticulation sides. Since there are at least four outgoing cut-arcs, there exists a fourth outgoing cut-arc that is above these reticulations. Then N is orientable at the cut-edge in N corresponding to this fourth outgoing cut-arc of B_r in N_D. This gives a contradiction.

Note that in order to prove the second decomposition theorem for level-3 we still need to show that the theorem holds in case that N is orientable at a cut-edge of N. In the proof of the theorem we now have showed that the theorem holds in case that N is orientable at a leaf $\rho \in X$, which is only a special case of N being orientable at a cut-edge of N.

Now we have proved the first decomposition theorem for semi-directed binary networks, we continue with the second decomposition theorem. The proof of this theorem is based on the proof of Theorem 2 in [14] for directed binary networks. The structure of the proof is therefore very similar. The main difference is that we now make an extra assumption, namely that the network has no redundant biconnected component. A difference with the second decomposition for undirected binary networks is that we now consider only semi-directed binary networks that are level-2 since Theorem 4.10 is only proved for semi-directed binary networks that are level-2. For the rest of the proof we do not use that the network is level-2 and holds therefore for semi-directed binary networks in general.

Theorem 4.11. *A recoverable semi-directed binary level-2 network N on X, with $|X| \geq 4$ and no redundant biconnected components, is encoded by its quarnets $Qn(N)$ if and only if, for each nontrivial biconnected component B of N with at least four incident cut-edges, N_B is encoded by $Qn(N_B)$.*

Proof. Let N be a recoverable semi-directed binary level-2 network on X with $|X| \geq 4$ and no redundant biconnected components.

" $\implies$ " Assume N is encoded by its quarnets $Qn(N)$. Consider any nontrivial biconnected component B of N with at least four incident cut-edges. Suppose that N_B is not encoded by $Qn(N_B)$. Then there exists a recoverable semi-directed binary network $N'_B \neq N_B$ such that $Qn(N_B) = Qn(N'_B)$. By Theorem 4.10 N'_B has the same CE-splits as N_B. By definition of N_B this implies that each CE-split of N'_B has one set that is a singleton.

Suppose N'_B has a redundant biconnected component R. Note that R has exactly two incident cut-edges. Since each CE-split of N'_B has one set that is a singleton, one incident cut-edge of R leads to a single leaf x and the other incident cut-edge of R leads to the other leaves. Now each quarnet containing x also contains R. This is not possible because $Qn(N_B) = Qn(N'_B)$ and it is easily checked that for each such a leaf x there exists a quarnet in $Qn(N_B)$ with no such redundant biconnected component since N_B is simple. This shows that N'_B has no redundant biconnected components.

If we combine that N'_B has no redundant biconnected components and that all CE-splits of N'_B have one set that is a singleton, we get that N'_B consists of one nontrivial biconnected component with incident cut-edges to the leaves. So N'_B is a simple network.

Let B' be the nontrivial biconnected component of N'_B. Let N' be the result of replacing B by B' in N. Note that N' is recoverable. We will show that $Qn(N) = Qn(N')$.

Let $Q \in Qn(N)$ and let w, x, y and z be the leaves of Q. If the incident cut-edges of B that are leading to w, x, y and z are all different or all the same, then we have that $Q \in Qn(N')$ since the only difference between N and N' is that B is replaced by B' and that $Qn(N_B) = Qn(N'_B)$.

Now suppose the incident cut-edges of B that are leading to w, x, y and z are not all different and also not all the same. Let A be a maximal subset of $\{w, x, y, z\}$ such that each incident cut-edge of B is leading to at most one leaf in A. Note that $|A| = 2$ or $|A| = 3$. Since B has at least four incident cut-edges, we can find one leaf (if $|A| = 3$) or two leaves (if $|A| = 2$) such that together with A we have four leaves which corresponds to four different incident cut-edges of B. Since $Qn(N_B) = Qn(N'_B)$ we now have that the subnetworks on A exhibited by N and N' are the same. Now the quarnet on w, x, y and z is the same for N and N' since the only differences are in the parts corresponding to incident cut-edges of B and these parts outside B are the same for N and N'. Now we have showed that $Qn(N) = Qn(N')$. This contradicts the assumption that N is encoded by $Qn(N)$.

" $\Longleftarrow$ " Assume that for each nontrivial biconnected component B of N with at least five incident cut-edges, N_B is encoded by $Qn(N_B)$. Suppose that N is not encoded by $Qn(N)$. Then there exists a recoverable semi-directed binary network $N' \neq N$ with $Qn(N) = Qn(N')$. We will show that $N = N'$.

Suppose N' has a redundant biconnected component R. If R is connected to other redundant biconnected components, we redefine R as the maximum set of connected redundant biconnected components that contains the redundant biconnected component R. The two incident cut-edges that connect R with the rest of N' we then call the incident cut-edges of R. By Theorem 4.10 N' has the same CE-splits as N. Since R has exactly two incident cut-edges, R is always connected to two other biconnected components by its incident cut-edges. If one of the incident cut-edges of R is connected to a leaf, then with a similar reasoning as we used for N_B and N'_B in the proof for the other implication of the theorem we get that N and N' have no redundant biconnected components. Now if both incident cut-edges of R are not connected to a leaf, then both incident cut-edges of R are connected to a biconnected component with at least two incident cut-edges besides the incident cut-edge connected to R. So both incident cut-edges of R are leading to at least two leaves. Consider the quarnet on $\{w, x, y, z\}$ such that one incident cut-edge of R leads to w and x and the other incident cut-edges of R leads to y and z. Then this quarnet contains R. This is a contradiction since $Qn(N) = Qn(N')$ and N has no redundant biconnected components. So in both cases we get that N' has no redundant biconnected components.

At this point, we observe that, for a biconnected component B with exactly four incident cut-edges, N_B is trivially encoded by $Qn(N_B)$, since in that case N_B is isomorphic to the single quarnet in $Qn(N_B)$.

The rest of the proof is by induction on $|X|$. If $|X| = 4$, then, since N and N' are recoverable, N and N' are both equal to the single quarnet in $Qn(N)$ and we are done.

Now assume $|X| \geq 5$. Let B_0 any biconnected component of N that contains the root in a directed binary network N_D from which N can be obtained. Let $e_1 = u_1v_1, \ldots, e_b = u_bv_b$ be the incident cut-edges of B_0, where $u_1, \ldots, u_b$ are in B_0. Note that $b \geq 3$ since N is recoverable and has no redundant biconnected components.

Let $N_1, \ldots, N_b$ be the connected components after deleting B_0 and new leaves $p_1, \ldots, p_b$ with edges v_ip_i for $1 \leq i \leq b$, such that N_i contains v_i. We call X_i the set of leaves of N_i for $1 \leq i \leq b$. Then, since $b \geq 3$, we have $|X_i| < |X|$. Note that N_i is recoverable and has no redundant biconnected components for $1 \leq i \leq b$ since N has no redundant biconnected components.

We saw that N' has the same CE-splits as N. Thus, $\{X_i, X \setminus \{X_i\}\}$ is a CE-split of N' for $1 \leq i \leq b$. Now, since a tree is uniquely defined by its splits ([2]), viewing the biconnected components of N and N' as vertices of two trees shows that N' has a biconnected component B'_0 with b incident cut-edges for which the CE-splits agree with the CE-splits of the incident cut-edges of B_0. Let $e'_1 = u'_1 v'_1, \ldots, e'_b = u'_b, v'_b$ be the incident cut-edges of B'_0, where $u'_1, \ldots, u'_b$ are in B'_0. Let $N'_1, \ldots, N'_b$ be the connected components after deleting B'_0 and new leaves $p'_1, \ldots, p'_b$ with edges $v'_i p'_i$ for $1 \leq i \leq b$, such that N'_i contains v'_i. Assume without loss of generality that N'_i is a network on X_i for $1 \leq i \leq b$. Note that we can choose the new leaves $p'_1, \ldots, p'_b$ such that $p'_i = p_i$ holds for $1 \leq i \leq b$. Further note that N'_i is recoverable and has no redundant biconnected components for $1 \leq i \leq b$ since N' has no redundant biconnected components.

Now we look at the quarnets of N_i and N'_i. Let $i \in \{1, \ldots, b\}$. Let l_1, l_2, l_3 and l_4 be four different leaves of N_i. If p_i is not of the four leaves, then the quarnet exhibited on $\{l_1, l_2, l_3, l_4\}$ by N_i is also a quarnet of N and therefore known. If p_i is one of the four leaves, say l_4, then we can obtain the quarnet exhibited on $\{l_1, l_2, l_3, p_i\}$ by N_i in the following way. Let $p \in X$ a leaf that is not in N_i. Let Q be the quarnet exhibited on $\{l_1, l_2, l_3, p\}$ by N and subsequently suppressing all the redundant biconnected components that are connected to p. Then Q is the wanted quarnet of N_i. For N'_i we can obtain the quarnets in the same way. So $Qn(N_i)$ and $Qn(N'_i)$ are known. To show that $N = N'$, it remains to show that $N_{B_0} = N_{B'_0}$ and that $N_i = N'_i$ for $1 \leq i \leq b$.

First, we show that $N_{B_0} = N_{B'_0}$. Observe that $Qn(N_{B_0}) = Qn(N_{B'_0})$ (if for any four leaves y_j, y_k, y_l, y_m the quarnet in $Qn(N_{B_0})$ and the quarnet in $Qn(N_{B'_0})$ would be different, then for any four leaves x_j, x_k, x_l, x_m in the parts of the network corresponding to incident cut-edges e_j, e_k, e_l, e_m, respectively, the quarnet in $Qn(N)$ and the quarnet in $Qn(N')$ would be different). If $b \geq 5$, then $N_{B_0} = N_{B'_0}$ holds because $Qn(N_{B_0}) = Qn(N_{B'_0})$ and by assumption N_{B_0} is encoded by $Qn(N_{B_0})$. Moreover, $b \geq 3$ since N is recoverable and has no redundant biconnected components. For $b = 4$ the statement $N_{B_0} = N_{B'_0}$ is trivially true.

The only case left is $b = 3$. Consider a leaf in each of the three parts corresponding to the three incident cut-edges of B_0, namely leaves l_1, l_2 and l_3. Since $|X| \geq 4$, we can find another leaf l_4 of N. Consider the quarnet Q in $Qn(N)$ on $\{l_1, l_2, l_3, l_4\}$. Since $Qn(N) = Qn(N')$ we now get that the trinets on $\{l_1, l_2, l_3\}$ exhibited by N and N', respectively, are the same. Let $B_0(T)$ be the biconnected component of this trinet T such that the component has three incident cut-edges that lead to l_1, l_2 and l_3, respectively. Then we have that $N_{B_0(T)} = N_{B_0}$. Moreover, since N' also exhibits T, we have that $N_{B_0(T)} = N_{B'_0}$. It follows that $N_{B_0} = N_{B'_0}$.

Now let $i \in \{1, \ldots, b\}$. We will show that $N_i = N'_i$. Observe that $Qn(N_i) = Qn(N'_i)$ for similar reasons as we used for showing that $Qn(N_{B_0}) = Qn(N_{B'_0})$. Since $|X_i| < |X|$, the statement $N_i = N'_i$ follows by induction if (a) N_i and N'_i are recoverable and have no redundant biconnected components and (b) $|X_i| \geq 4$. Note that (a) holds since N and N' have no redundant biconnected components. If $|X_i| = 1$, then clearly $N_i = N'_i$ because both consist of a single leaf. The only cases left are $|X_i| = 2$ and $|X_i| = 3$.

First we consider the case that $|X_i| = 2$. Denote the two leaves of X_i with l_1 and l_2. Since $|X| \geq 4$, we can find two other leaves l_3 and l_4 of N. Consider the quarnet Q in $Qn(N)$ on $\{l_1, l_2, l_3, l_4\}$. Since $Qn(N) = Qn(N')$ we now get that the binets on $\{l_1, l_2\}$ exhibited by N and N', respectively, are the same. So $N_i = N'_i$ holds.

Now we consider the case that $|X_i| = 3$. Denote the three leaves of X_i with l_1, l_2 and l_3. Since $|X| \geq 4$, we can find another leaf l_4 of N. Consider the quarnet Q in $Qn(N)$ on $\{l_1, l_2, l_3, l_4\}$. Since $Qn(N) = Qn(N')$ we now get that the trinets on $\{l_1, l_2, l_3\}$ exhibited by N and N', respectively, are the same. So $N_i = N'_i$ holds.

Since for all cases $N_{B_0} = N_{B'_0}$ and $N_i = N'_i$ for $1 \leq i \leq b$ holds, we can conclude that $N = N'$, which gives a contradiction. □

Chapter 5

Recoverable networks

In this chapter we will combine some of the different results we obtained earlier in this thesis for undirected and semi-directed binary networks. Note that the proofs in this chapter are based on the proofs in [14] that are about directed binary networks. Therefore the structure of the proofs in this chapter are quite similar to these proofs.

5.1 Undirected

In this section we prove some results for recoverable undirected binary networks. First we prove that an undirected binary simple level-2 network is encoded by its set of quarnets. This is a more general result of Theorem 3.3 since this theorem showed that the class of undirected binary simple level-2 networks is weakly encoded by quarnets. Note that an undirected binary simple network is recoverable by definition.

Theorem 5.1. *Every undirected binary simple level-2 network N on X, with $|X| \geq 4$, is encoded by its set of quarnets $Qn(N)$.*

Proof. Let N be an undirected binary simple level-2 network on X with $|X| \geq 4$. Assume that this network is not encoded by its set of quarnets $Qn(N)$. Then there exists a recoverable undirected binary network $N' \neq N$ with $Qn(N) = Qn(N')$. We will show that $N = N'$.

We want to show that N' is an undirected binary simple level-2 network. In other words, we need to show that N' is simple and level-2.

First we show that N' is simple. By Theorem 4.5 the set of CE-splits of N' equals the set of CE-splits of N. Since N is a simple network this implies that each CE-split of N' has one set that is a singleton.

Suppose N' has a redundant biconnected component R. Note that R has exactly two incident cut-edges. Since each CE-split of N' has one set that is a singleton, one incident cut-edge of R leads to a single leaf x and the other incident cut-edge of R leads to the other leaves $X \setminus \{x\}$. Now each quarnet containing x also contains R. This is not possible because $Qn(N) = Qn(N')$ and an undirected binary simple network without redundant biconnected components has no quarnets with redundant biconnected components. This shows that N' has no redundant biconnected components.

If we combine that N' has no redundant biconnected components and that all CE-splits of N' has one set that is a singleton, we get that N' consists of one nontrivial biconnected component with incident cut-edges to the leaves. So N' is a simple network.

Now we show that N' is level-2. Suppose that N' is a simple strict level-k network with $k > 2$. If N' has no sets of parallel edges, choose four arbitrary leaves. If N' has exactly one set of parallel edges, choose one leaf that is on one of these edges and three other leaves arbitrary. If N' has at least two sets of parallel edges, choose two leaves on two different sets of parallel edges and two other leaves arbitrary. Consider the quarnet exhibited on these four leaves by N' for each of these cases.

Note that a quarnet exhibited by N' can only be a level-k' network with $k' < k$ if at least one pair of parallel edges of N' is suppressed. If there are at most two sets of parallel edges in N', then the quarnet is strict level-k since no parallel edges are suppressed. If there are more than two sets of parallel edges in N', then there are some parallel edges suppressed in the quarnet since we have only chosen two leaves on two different sets of parallel edges. Since the two sets of parallel edges that are kept in the quarnet are then in one biconnected component in N' they are also in the same biconnected component in the quarnet. This implies that the quarnet is strict level-k' with $k' > 2$.

In each of the cases N' has a quarnet that is level-k' with $k' > 2$. Since $Qn(N') = Qn(N)$ contains only level-2 quarnets, this gives a contradiction. It follows that N' is a level-2 network.

So N' is a recoverable undirected binary simple level-2 network. By Theorem 3.3 we now get that $N = N'$, which gives a contradiction. □

Note that in the above proof we could not use reticulations to show that N' is level-2 as is done in the proof for directed binary networks in Theorem 3 of [14] since we have not defined reticulations for undirected binary networks. This is also a difference between this proof and the proof in the next section for semi-directed binary networks since it is then possible to follow the proof for directed binary networks quite well.

Further note that Theorem 5.1 can be extended to a more general result. In this thesis we define undirected binary networks by a single definition. If undirected and binary networks are defined separately, we can define being encoded such that N' is not necessarily binary (see [14]). In that case we can show that N' is still binary using that $Qn(N) = Qn(N')$ and that N is binary. So with this different definition of encoding the results of this section still holds for undirected binary networks.

Finally, in the corollary below we combine the results in order to obtain a result for recoverable undirected binary level-2 networks.

Corollary 5.2. *Every recoverable undirected binary level-2 network N on X, with $|X| \geq 4$ and no redundant biconnected components, is encoded by its set of quarnets $Qn(N)$.*

Proof. Follows from Theorems 4.6 and 5.1. □

5.2 Semi-directed

In this section we prove some results for recoverable semi-directed binary networks. First we prove that a semi-directed binary simple level-2 network is encoded by its set of quarnets. This is a more general result of Theorem 3.6 since this theorem showed that the class of semi-directed binary simple level-2 networks is weakly encoded by quarnets. Note that a semi-directed binary simple network is recoverable by definition.

Theorem 5.3. *Every semi-directed binary simple level-2 network N on X, with $|X| \geq 4$, is encoded by its set of quarnets $Qn(N)$.*

Proof. Let N be a semi-directed binary simple level-2 network on X with $|X| \geq 4$. Assume that this network is not encoded by its set of quarnets $Qn(N)$. Then there exists a recoverable semi-directed binary network $N' \neq N$ with $Qn(N) = Qn(N')$. We will show that $N = N'$.

We want to show that N' is a semi-directed binary simple level-2 network. In other words, we need to show that N' is simple and level-2.

First we show that N' is simple. By Theorem 4.10 the set of CE-splits of N' equals the set of CE-splits of N. Since N is a simple network this implies that each CE-split of N' has one set that is a singleton.

Suppose N' has a redundant biconnected component R. Note that R has exactly two incident cut-edges. Since each CE-split of N' has one set that is a singleton, one incident cut-edge of R leads to a single leaf x and the other incident cut-edge of R leads to the other leaves $X \setminus \{x\}$. Now each quarnet containing x also contains R. This is not possible because $Qn(N) = Qn(N')$ and it is easily checked that for each such a leaf x there exists a quarnet in $Qn(N)$ with no such redundant biconnected component since N_B is simple. This shows that N' has no redundant biconnected components.

If we combine that N' has no redundant biconnected components and that all CE-splits of N' has one set that is a singleton, we get that N' consists of one nontrivial biconnected component with incident cut-edges to the leaves. So N' is a simple network.

Now we show that N' is level-2. First note that $Tn(N) = Tn(N')$ since the trinets can be exhibited by the quarnets and $Qn(N) = Qn(N')$. Now suppose we have any simple strict level-k network with $k > 2$. Then this network has exactly k reticulations. If there are at least three leaves on reticulation sides, take three such leaves. Otherwise take all leaves on reticulation sides and take the remaining leaves on sides that form parallel arcs in the underlying generator (if this is possible), choosing at most one leaf per pair of parallel arcs. If the network has parallel arcs, then the underlying network has two sides that form parallel arcs which have no leaves on it. In this case one leaf can be chosen arbitrarily since this set of parallel arcs will not be suppressed. Now the trinet on these three leaves has at least three reticulations. Note that if a leaf is chosen on one of the parallel arcs in the underlying generator, the pair of parallel arcs will not be suppressed, and so we get a reticulation. So a simple strict level-k network, with $k > 2$, has a level-k' trinet with $k' > 2$. Since $Tn(N') = Tn(N)$ contains only level-2 trinets, it follows that N' is a level-2 network.

So N' is a recoverable semi-directed binary simple level-2 network. By Theorem 3.6 we now get that $N = N'$, which gives a contradiction. □

Note that Theorem 5.3 can be extended to a more general result. In this thesis we define semi-directed binary networks by a single definition. If semi-directed and binary networks are defined separately, we can define being encoded such that N' is not necessarily binary (see [14]). In that case we can show that N' is still binary using that $Qn(N) = Qn(N')$ and that N is binary. So with this different definition of encoding the results of this section still holds for semi-directed binary networks.

Finally, in the corollary below we combine the results in order to obtain a result for recoverable semi-directed binary level-2 networks.

Corollary 5.4. *Every recoverable semi-directed binary level-2 network N on X, with $|X| \geq 4$ and no redundant biconnected components, is encoded by its set of quarnets $Qn(N)$.*

Proof. Follows from Theorems 4.11 and 5.3. □

Chapter 6

Discussion

A number of interesting open problems remain. First of all, this thesis focused mostly on level-2 networks. For $k \geq 3$, it would be of great interest to investigate which undirected and semi-directed binary level-k networks are encoded by quarnets. Furthermore, we can look for a largest class of level-k networks that is weakly encoded by quarnets.

In this thesis we have looked at networks that are encoded by quarnets, subnetworks on four leaves. It might also be interesting to investigate which undirected and semi-directed binary level-k networks are encoded by subnetworks with five or more leaves. For example, we can investigate for a certain k if all level-k networks are encoded by networks on l leaves for a certain l. Note that in order to get the strongest result, we need l to be as small as possible. Furthermore, note that if a network is encoded by subnetworks on p leaves, then the network is also encoded by subnetworks on q leaves if $q > p$. Some results in this direction are given in [7].

We have showed that simple level-2 networks are weakly encoded by using the undirected and semi-directed binary level-k generators for $k \leq 2$. To the best of my knowledge, the undirected and semi-directed binary level-k generators have not been studied a lot. In this thesis we have given the level-k generators for $k \leq 3$ explicitly. This can be extended to higher levels. This would be of interest since the proofs of encodings are often based on generators.

Similar as for directed binary level-k networks, the number of generators grows as k increases. Therefore we need a shorter and more efficient way to investigate the level-k networks if k is large. Possibly there is another way to decompose the undirected and semi-directed binary level-k networks in order to investigate these networks in a more efficient way.

Note that some of the directions for further research we mention here also hold for directed binary networks and were therefore already mentioned in the discussion of [11] for directed binary networks. The first decomposition theorem holds for undirected and directed binary networks for all levels. For semi-directed binary networks the first decomposition theorem can not be proved easily for all levels in a similar way. It could be of some interest to compare the current results for undirected, semi-directed and directed binary networks and to study these different types of networks in parallel.

Some definitions of semi-directed binary networks can be reconsidered. In this thesis they mainly depend on the definitions of directed binary networks. In [6], semi-directed binary networks are defined in a different way. The semi-directed binary networks have then not always a valid root location. It could be of some interest to generalize the results from this thesis to results that hold for semi-directed binary networks that are defined in this different way. This is especially relevant if there are possibly multiple roots. Note that if the definition of a semi-directed binary network changes in this way, the definition of a semi-directed binary level-k generator then also has to be changed.

The semi-directed binary level-k generators can be obtained in different ways. A first way can be by defining the semi-directed binary level-k generators for each k directly as we did in this thesis for undirected binary level-k generators. A second way can be by using the directed binary level-k generators for a certain k. Note that these directed binary level-k generators must then be known. For the levels we have considered in this thesis this was already the case. A third way can be by using the undirected binary level-k generators to obtain the semi-directed binary level-k generators. This could then be possible by creating reticulations at vertices or on edges. A fourth way can be by using the semi-directed binary level-$(k-1)$ generators to obtain the semi-directed binary level-k generators. Then an extra reticulation has to be added to the generator. Although we already have a definition which can be used to obtain the different undirected binary level-k generators, we can possibly also look at the undirected binary level-k generators in similar ways to obtain the generators in a more intuitive way. Note that this can be done for all different levels. It is especially important to choose the most efficient way to obtain the generators if k becomes larger.

We now continue with discussing the decomposition theorems. We have proved that the first decomposition theorem holds for recoverable undirected binary networks in general. However, this is not the case for recoverable semi-directed binary networks. In the second implication of the proof for the case that (c) of the claim holds (and (a) and (b) of the claim do not hold) the proof depends on k. The first decomposition is now only proved for recoverable semi-directed binary level-2 networks. It would be of great interest to investigate this special case in the proof further for $k \geq 3$.

For level-3 this could possibly be showed by distinguishing in the cases that the network is or is not orientable at a cut-edge. We saw that the proof for the case that the network is not orientable at any cut-edge follows then quite easily. This gives a motivation for the following conjecture.

Conjecture 6.1. Let N be a recoverable semi-directed binary level-3 network on X and $\{A, B\}$ a partition of X.

(i) If $|A| = 1$ or $|B| = 1$, then $\{A, B\}$ is a CE-split of N.

(ii) Otherwise, if $|A| \geq 2$ and $|B| \geq 2$, then $\{A, B\}$ is a CE-split of N if and only if for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $b_1, b_2 \in B$ with $b_1 \neq b_2$, $\{\{a_1, a_2\}, \{b_1, b_2\}\}$ is a CE-split of the quarnet on $\{a_1, a_2, b_1, b_2\}$ exhibited by N.

Note that the case that the network is orientable at a cut-edge still has to be investigated before we can conclude if this conjecture holds.

The second decomposition theorem is now proved for recoverable undirected binary networks without redundant biconnected components and recoverable semi-directed binary level-2 networks without redundant biconnected components. In further research this decomposition theorem could possibly be extended to networks that may have redundant biconnected components. Note that if the first decomposition theorem can be proved for semi-directed binary level-k networks for a certain k ($k \geq 3$), then the second decomposition theorem can also be extended to this level of semi-directed binary networks.

Further note that the combined results as described in Chapter 5 can possibly be extended if undirected and semi-directed binary networks are investigated further. If it can be proved that simple level-k networks are weakly encoded for a certain k ($k \geq 3$) and the first decomposition theorem holds for this k, then we know by the proofs of this thesis that then each such a simple level-k network is encoded. Moreover, if also the second decomposition theorem holds for these networks, then each of these recoverable level-k networks is encoded.

Furthermore, this thesis can give some ideas for an algorithm to reconstruct recoverable undirected and semi-directed level-2 networks from their sets of quarnets. Note that it is not guaranteed that all level-2 networks can be reconstructed from their sets of quarnets since networks with redundant biconnected components have to be investigated further.

Finally, it would be of great interest to investigate if a recoverable undirected or recoverable semi-directed binary network can be uniquely reconstructed by its set of quarnets if not all the quarnets of the network are known. This is relevant for studies using Markov models, as in [3], [4] and [1], since then not all quarnets can be distinguished from each other. In further research it can be investigated if with less knowledge about the quarnets of the network, certain networks still can be reconstructed.

Bibliography

[1] Muhammad Ardiyansyah. Distinguishing level-2 phylogenetic networks using phylogenetic invariants. *arXiv preprint arXiv:2104.12479*, 2021.

[2] Peter Buneman. The recovery of trees from measures of dissimilarity. In *Mathematics the the Archeological and Historical Sciences*, pages 387–395. Edinburgh University Press, 1971.

[3] Elizabeth Gross and Colby Long. Distinguishing phylogenetic networks. *SIAM Journal on Applied Algebra and Geometry*, 2(1):72–93, 2018.

[4] Elizabeth Gross, Leo van Iersel, Remie Janssen, Mark Jones, Colby Long, and Yukihiro Murakami. Distinguishing level-1 phylogenetic networks on the basis of data generated by markov processes. *Journal of Mathematical Biology*, 83(3):1–24, 2021.

[5] Katharina T Huber, Leo van Iersel, Remie Janssen, Mark Jones, Vincent Moulton, Yukihiro Murakami, and Charles Semple. Orienting undirected phylogenetic networks. *arXiv preprint arXiv:1906.07430*, 2019.

[6] Sophia Huebler, Rachel Morris, Joseph Rusinko, and Yifei Tao. Constructing semi-directed level-1 phylogenetic networks from quarnets. *arXiv preprint arXiv:1910.00048*, 2019.

[7] Frank Janisse. Encoding level-k phylogenetic networks, 2021. MSc thesis, TU Delft, `http://resolver.tudelft.nl/uuid:11939b58-b834-4073-8de8-b61d9a5f9a81`.

[8] Remie Janssen. Rearranging phylogenetic networks, 2021. PhD thesis, TU Delft, `https://doi.org/10.4233/uuid:1b713961-4e6d-4bb5-a7d0-37279084ee57`.

[9] Thomas Marcussen, Simen R Sandve, Lise Heier, Manuel Spannagl, Matthias Pfeifer, International Wheat Genome Sequencing Consortium, Kjetill S Jakobsen, Brande BH Wulff, Burkhard Steuernagel, Klaus FX Mayer, and Odd-Arne Olsen. Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, 345(6194):1250092, 2014.

[10] David Morrison. Complex hybridizations in wheat. https://phylonetworks.blogspot.com/2015/01/complex-hybridizations-in-wheat.html.

[11] Leonie Nipius. Rooted binary level-3 phylogenetic networks are encoded by quarnets, 2020. BSc thesis, TU Delft, http://resolver.tudelft.nl/uuid:a9c5a8d4-bc8b-4d15-bdbb-3ed35a9fb75d.

[12] Leo van Iersel. Hoe zijn ze verwant? *Nieuw Archief voor Wiskunde*, 5/16(3):174–178, 2015. http://www.nieuwarchief.nl/serie5/pdf/naw5-2015-16-3-174.pdf.

[13] Leo van Iersel, Sjors Kole, Vincent Moulton, and Leonie Nipius. An algorithm for reconstructing level-2 phylogenetic networks from trinets. *arXiv preprint arXiv:2109.11244*, 2021.

[14] Leo van Iersel and Vincent Moulton. Trinets encode tree-child and level-2 phylogenetic networks. *Journal of Mathematical Biology*, 68(7):1707–1729, 2014.

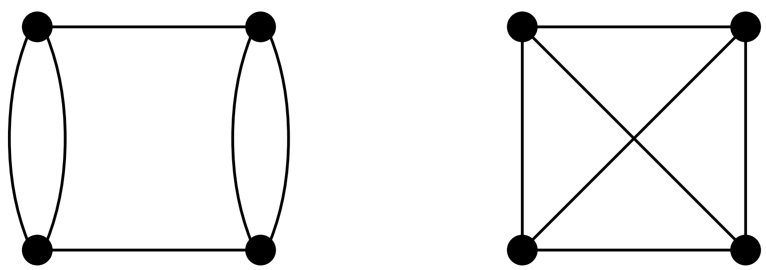
Appendix A

Level-3 generators

In the following sections the level-3 generators of undirected and semi-directed binary networks will be given. Note that, since we do not use the different sides of these level-3 generators in this thesis, the different sides of the level-3 generators will not be labelled.

A.1 Undirected

In this section we look at the undirected binary level-3 generators. Recall we have used Definition 2.6 of [1] for the definition of an undirected binary level-3 generator in this thesis. Using this definition we know that a multigraph G is an undirected binary level-3 generator if and only if G is a 3-regular biconnected multigraph with four vertices. This implies that the two undirected binary level-3 generators as given in Figure A.1 are level-3 generators and that these are the only possible level-3 generators for undirected binary networks.



(a) Generator 3.1. (b) Generator 3.2.

Figure A.1: The two undirected binary level-3 generators.

A.2 Semi-directed

In this section we look at the semi-directed binary level-3 generators. We recall the definition which states that a semi-directed binary level-3 generator can be obtained from a directed binary level-3 generator by replacing the arcs by edges except for the incoming arcs that belong to an indegree-2 vertex and then suppressing the degree-2 vertex that does not have two incoming arcs. Since each semi-directed binary level-3 generator can be obtained from a directed binary level-3 generator, we can use the level-3 generators of directed binary networks.

There are 65 different directed binary level-3 generators, which are given in Appendix A of [11]. These 65 level-3 generators are numbered by 3.1, . . . , 3.65. We keep this numbering for the directed binary level-3 generators. Therefore we will use another numbering for the semi-directed binary level-3 generators.

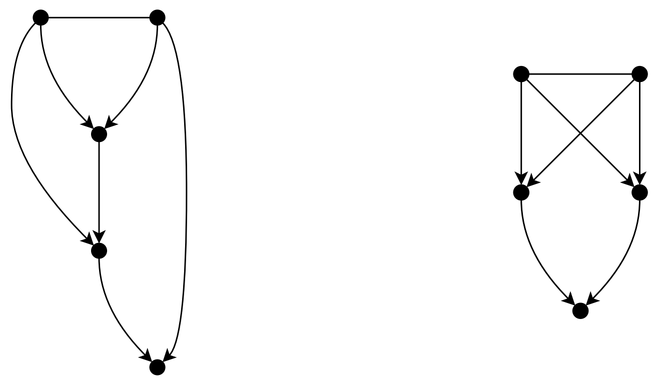
From each of the 65 directed binary level-3 generators we can obtain a semi-directed binary level-3 generator. This does not imply that there are 65 different semi-directed binary level-3 generators. In Table A.1 we can see which directed binary level-3 generators are corresponding to the same semi-directed binary level-3 generator. Note that each semi-directed binary level-3 generator can be obtained from at least two different directed binary level-3 generators.

Semi-directed	Directed
$3a$	3.15, 3.23, 3.25, 3.32, 3.62
$3b$	3.16, 3.24
$3c$	3.19, 3.22
$3d$	3.18, 3.20, 3.36, 3.65
$3e$	3.28, 3.64
$3f$	3.31, 3.63
$3g$	3.12, 3.17, 3.21, 3.35, 3.43, 3.55, 3.59
$3h$	3.4, 3.5, 3.29, 3.34, 3.42, 3.48, 3.49
$3i$	3.9, 3.26, 3.41
$3j$	3.13, 3.33, 3.44, 3.54
$3k$	3.27, 3.40
$3l$	3.11, 3.30, 3.38, 3.53, 3.56, 3.61
$3m$	3.2, 3.3, 3.47
$3n$	3.7, 3.8, 3.45, 3.51, 3.57
$3o$	3.1, 3.6, 3.46, 3.58
$3p$	3.10, 3.37, 3.39, 3.52, 3.60
$3q$	3.14, 3.50

Table A.1: The semi-directed binary level-3 generators with the directed binary level-3 generators from which they can be obtained.

Now we see that there are 17 different semi-directed binary level-3 generators. These 17 generators are given in Figures A.2 - A.10. Note that the generators with an $*$ have symmetry next to the symmetry that is caused by the possible parallel arcs.

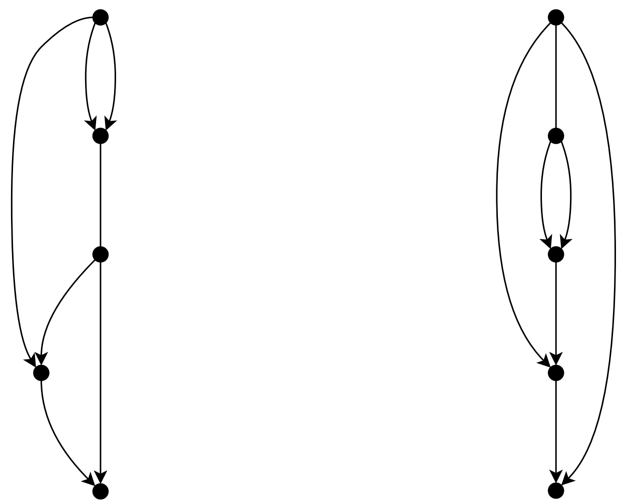
Further note that each semi-directed binary level-3 generator can be oriented in at least two different ways since we already saw that each semi-directed binary level-3 generator corresponds to at least two different semi-directed binary level-3 generators. Moreover, for a semi-directed binary level-3 generator the number of different possible roots is equal to the number of directed binary level-3 generators from which the semi-directed binary level-3 generator can be obtained.



(a) Generator $3a$.

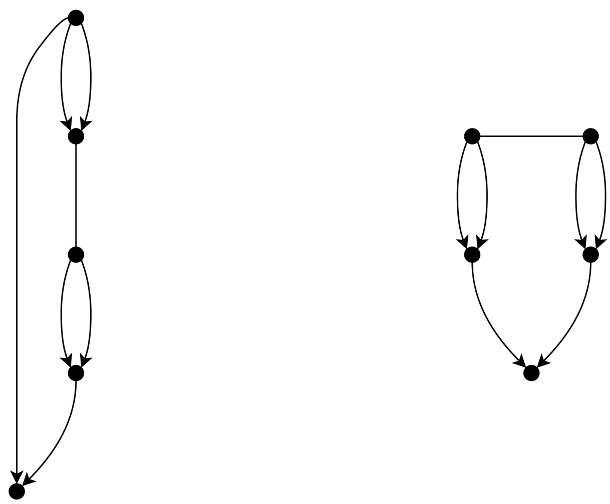
(b) Generator $3b$.*

Figure A.2: The two semi-directed binary level-3 generators that have one reticulation side and no parallel arcs.



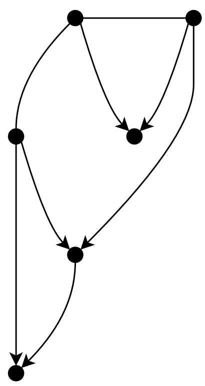
(a) Generator $3c$. (b) Generator $3d$.

Figure A.3: The two semi-directed binary level-3 generators that have one reticulation side and one set of parallel arcs.

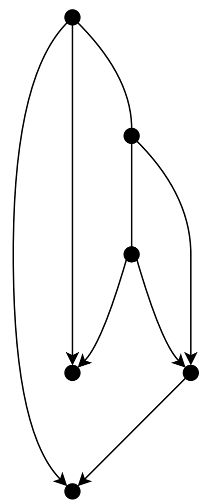


(a) Generator $3e$. (b) Generator $3f$.*

Figure A.4: The two semi-directed binary level-3 generators that have one reticulation side and two sets of parallel arcs.

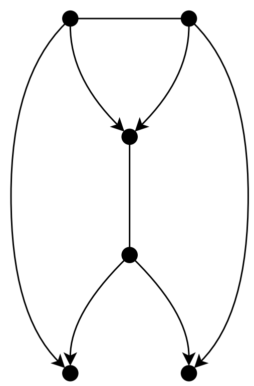


(a) Generator $3g$.

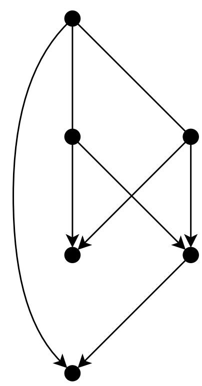


(b) Generator $3h$.

Figure A.5: Two of the four semi-directed binary level-3 generators that have two reticulation sides and no parallel arcs.

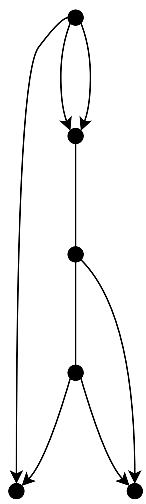


(a) Generator $3i$.*

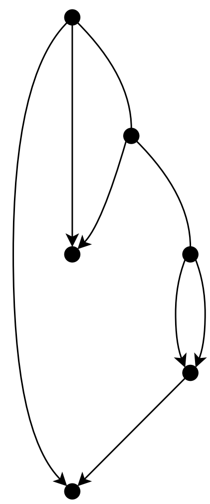


(b) Generator $3j$.*

Figure A.6: Two of the four semi-directed binary level-3 generators that have two reticulation sides and no parallel arcs.

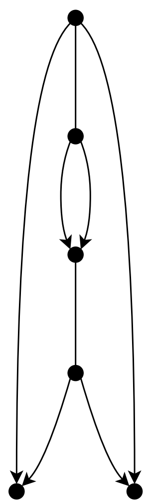


(a) Generator $3k$.



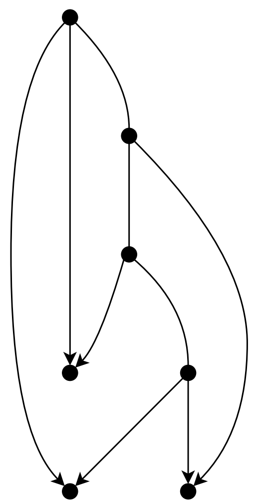
(b) Generator $3l$.

Figure A.7: Two of the three semi-directed binary level-3 generators that have two reticulation sides and one set of parallel arcs.

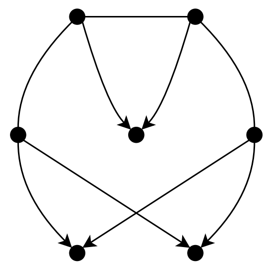


(a) Generator $3m$.*

Figure A.8: One of the three semi-directed binary level-3 generators that have two reticulation sides and one set of parallel arcs.

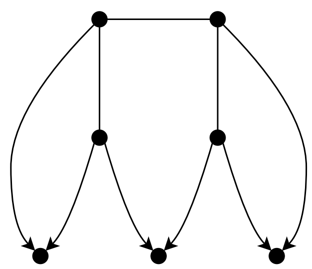


(a) Generator $3n$.

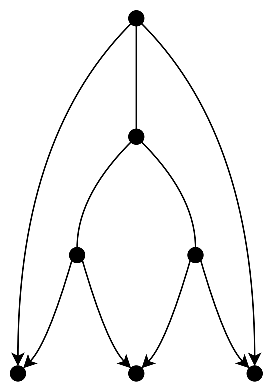


(b) Generator $3o$.*

Figure A.9: Two of the four semi-directed binary level-3 generators that have three reticulation sides and no parallel arcs.



(a) Generator $3p$.*



(b) Generator $3q$.*

Figure A.10: Two of the four semi-directed binary level-3 generators that have three reticulation sides and no parallel arcs.