3
Mimp-graphs: Graphs for Minimal Implicational Logic

In this chapter we define graph representations using "mixed" formulas and inferences in Natural Deduction in the purely implicational minimal logic; then we obtain a (weak) normalization theorem that, in fact, is a strong normalization theorem. The choice of purely implicational minimal logic ($M^{\rightarrow}$) is motivated by the fact that the computational complexity of the validity of $M^{\rightarrow}$ is PSPACE-complete and can polynomially simulate classical, intuitionistic and full minimal logic (Statman 1979) as well as any propositional logic with a Natural Deduction system satisfying the subformula property (Haeusler 2013).

3.1
Definition of Mimp-graphs

Mimp-graphs are special directed graphs whose nodes and edges are assigned with labels. Moreover we distinguish two parts, one representing the inferences of a proof, and the other the formulas. For the formula-part of a mimp-graph, we use formula graphs as a basis and consist only of formula nodes (see Definition 8). A formula can also be presented as a formula tree, but sharing within a formula is possible with formula graphs. An example is shown in the Figure 3.1: the propositions p and q each only need to occur once in the graph.

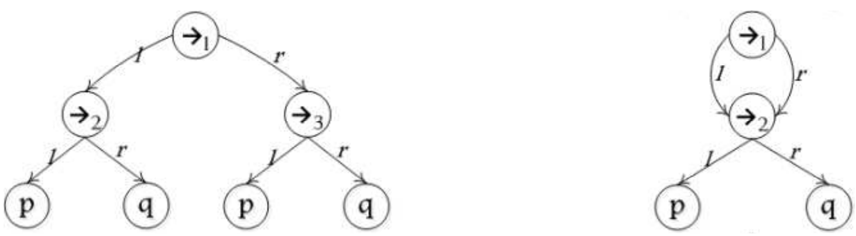


Figure 3.1: Formula $(p \rightarrow q) \rightarrow (p \rightarrow q)$ depicted as a formula tree (left side) and as a formula graph (right side).

The formula nodes (F-nodes) are labeled with formulas as being encoded/represented by their principal connectives (in particular, atoms). Additionally we will use delimiter nodes (D-nodes) H_i and C to indicate which are the hypothesis formulas and the final conclusion formula of the mimp-graph.

As for the inference-part of a mimp-graph we have the rule nodes (R-nodes) that are labeled with the names of the inference rules (→I and →E). Both logic connectives and inference names may be indexed, in order to achieve an 1–1 correspondence between formulas (inferences) and their representations (names).

Since formulas are uniquely determined by the representations in question, i.e. F-node labels, in the sequel we will sometimes identify both; to emphasize the difference we will refer to the former as *formula graphs*, i.e. those whose F-node labels are formulas, instead of principal connectives.

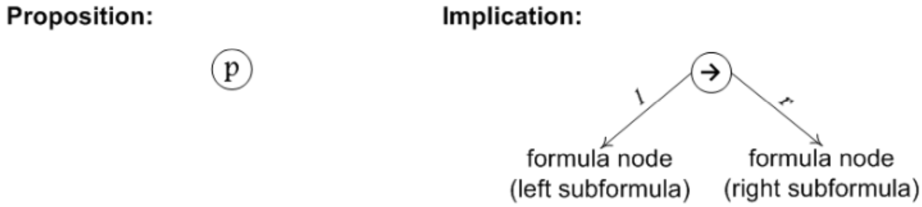


Figure 3.2: Types of formula nodes.

Edges are labeled with tokens that identify the connections between the respective R-nodes and F-nodes. Note that formulas may occur only once in the mimp-graph. Subformulas are indicated by outgoing edges with labels l (left) and r (right), see Figure 3.2.

R-nodes, like in Natural Deduction, require the correct number of premises. The premises are indicated by incoming edges and there are edges from the R-nodes to the conclusion formulas. In the terminology about R-nodes, the R-node →E has two incoming edges, these are distinguished by calling them major (with label M) or minor (with label m) and so also the F-node 'premise' associated with these edges. Thus, an R-node →E has a major premise and a minor premise, the major premise contains the connective that is eliminated; the other premise is called 'minor', and an R-node →I has only one premise. Figure 3.3 shows the R-nodes →I (implication introduction), →Iv[1] (implication introduction vacuously) and →E (implication elimination). In the case in which the discharge of hypotheses is vacuous, a mimp-graph is represented by a disconnected graph, where the discharged F-node is not linked to the conclusion of the rule by any directed path.

In the R-nodes, formulas are re-used, which is indicated by putting several arrows towards them, hence the number of edges with label p, M, m and c coming/going to an F-node could be arbitrarily large. To make all this a bit

[1]the "v" stands for "vacuous", this case of the rule →I discharges a hypothesis vacuously.

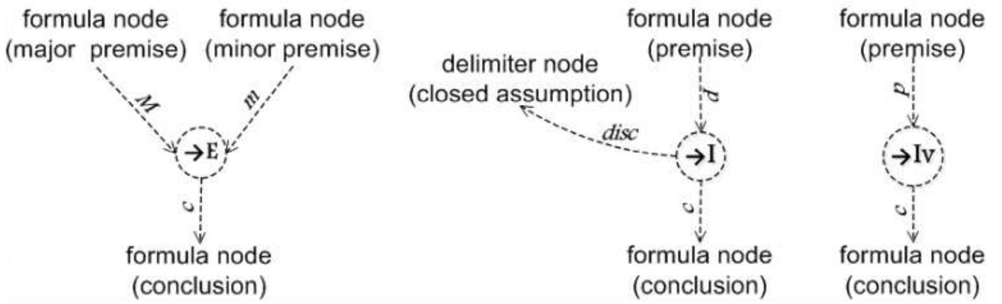


Figure 3.3: Types of rule nodes of the mimp-graph.

$$\dfrac{\dfrac{\dfrac{\dfrac{[p]^1 \quad p \to q}{q}\to E_1 \qquad [q \to r]^2}{r}\to E_2}{p \to r}\,(\to I_3{,}1)}{(q \to r) \to (p \to r)}\,(\to I_4{,}2)$$

$$\Downarrow trans$$

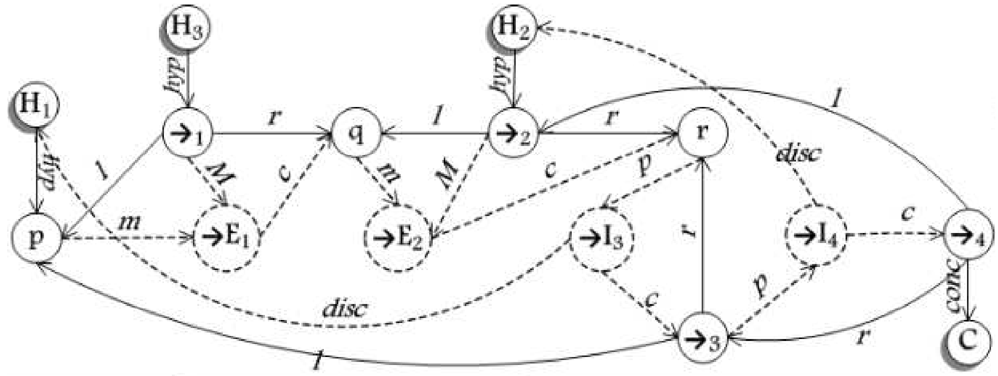


Figure 3.4: The transition from a derivation in ND to a mimp-graph.

more intuitive we give an example of a mimp-graph in Figure 3.4, which can be seen as a derivation of $(q \to r) \to (p \to r)$ from $(p \to q)$. Hypotheses are replaced by D-nodes H and indexes of discarded hypotheses are replaced by additional edges assigned with the label: *disc*. Note that the D-node H can only be discharged once. The re-using of formulas is necessary. We remind the reader that some valid implicational formulas, such as $((((r \to s) \to r) \to r) \to s) \to s$ (see Figure 3.5), need to use any number of times a subformula, in this case the subformula $(((r \to s) \to r) \to r) \to s$ is used twice. Because of this, edges p, m, M and c in Figure 3.5 are indexed with the same index of the R-node to which they belong.

F-nodes in the graph (Figure 3.4) are labeled with propositional letters p, q and r, the connective $\to$; R-nodes are labeled with $\to$E and $\to$I. The underlying idea is that there is an *inferential order* between R-nodes that provides the corresponding derivability order; F-node labeled $\to_4$ linked to the delimiter node

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[r]^1}{((r \to s) \to r) \to r} \quad [(((r \to s) \to r) \to r) \to s]^3}{s}}{r \to s} \quad [(r \to s) \to r]^2}{r}}{(((r \to s) \to r) \to r)} \quad [(((r \to s) \to r) \to r) \to s]^3}{\dfrac{s}{((((r \to s) \to r) \to r) \to s) \to s}}
$$

$\Downarrow$ *trans*

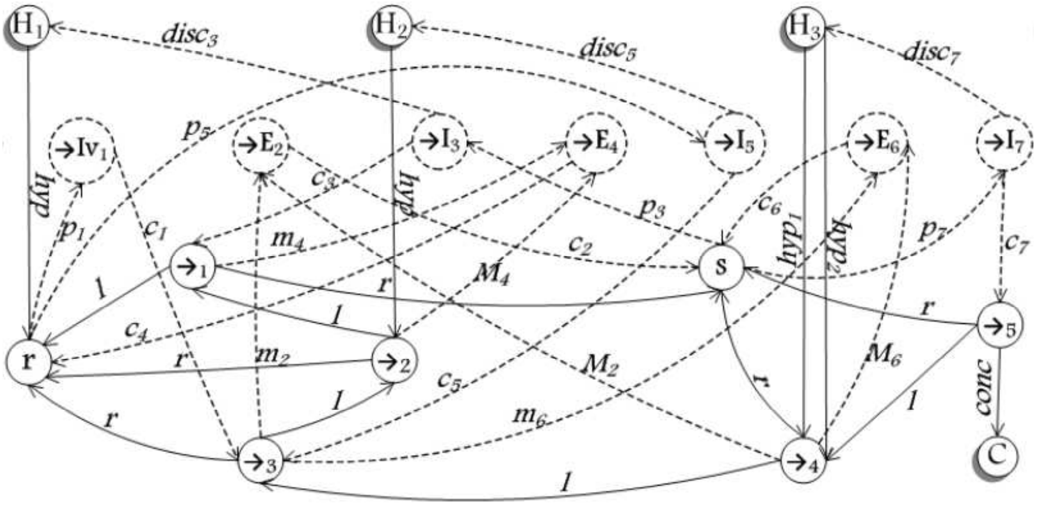


Figure 3.5: The transition from a natural deduction proof to a mimp-graph.

with labeled C by an edge labeled *conc* is the root node and the conclusion of the proof represented by the graph. Besides, the node $\to_1$ linked to the delimiter node labeled H_3 by the edge labeled *hyp* (hypothesis) in the graph is representing the premise $(p \to q)$.

We want to emphasize that the mimp-graphs put together information on formula-graphs and R-nodes. To make it more transparent we can use different types of lines. In this way F-nodes and edges between them are used solid lines, whereas inference nodes and edges between them and adjacent premises and/or conclusions are used dashed lines and additionally delimiter nodes have been shaded. So nodes of types $\to$ and p (propositions) together with adjacent edges (l, r) have solid lines, whereas nodes labeled $\to$I and $\to$E together with adjacent edges $(m, M, p, c, disc)$ have dashed lines.

We now give a formal definition of mimp-graphs.

Definition 5 (Label types) *There are four types of labels:*

- R-Labels *is the set of inference labels:* $\{\to\mathrm{I}_n / n \in \mathbb{N}\} \cup \{\to\mathrm{E}_m / m \in \mathbb{N}\}$,
- F-Labels *is the set of formula labels:* $\{\to_i / i \in \mathbb{N}\}$ *and propositional letters* $\{p, q, r, ...\}$,

- E_F-Labels *is the set of edge labels:* $\{l$ *(left)*, r *(right)*$\}$,
- E_M-Labels *is the set of edge labels:* $\{$ p_i *(premise)*, m_j *(minor premise)*, M_k *(major premise)*, c_r *(conclusion)*, $disc_s$ *(discharge)*, hyp_t *(hypothesis)*, *conc (final conclusion)* $/ \, i, j, k, r, s, t \in \mathbb{N}\}$,
- D-Labels *is the set of delimiter labels:* $\{H_k / k \in \mathbb{N}\} \cup \{C\}$.

The union of these four sets of label types will be called LBL.

Definition 6 *Let G be a graph. l_V is a labeling function from the* nodes *of G to* R∪F∪D-Labels, *i.e. it assigns a label to each node; l_E is a labeling function from the* edges *of G to* $\mathrm{E}_F \cup \mathrm{E}_M$-Labels, *i.e. it assigns a label to each edge.*

Definition 7 *Let* $G_1 = \langle \mathrm{V}^1, \mathrm{E}^1, \mathrm{L}^1 \rangle$ *and* $G_2 = \langle \mathrm{V}^2, \mathrm{E}^2, \mathrm{L}^2 \rangle$ *be two graphs, where:* V^1 *and* V^2 *are sets of vertices,* E^1 *and* E^2 *are sets of labeled* edges, L^1 *and* L^2 *are subsets of* LBL. *The operation* $G_1 \oplus G_2 = \langle \mathrm{V}^1 \sqcup \mathrm{V}^2, \mathrm{E}^1 \not\ddagger \mathrm{E}^2, \mathrm{L}^1 \cup \mathrm{L}^2 \rangle$ *equalizes nodes of G_1 with nodes of G_2 that have the same label, and equalizes edges with the same source, target and label into one. To be precise, the sets* $\mathrm{V}^1 \sqcup \mathrm{V}^2$ *and* $\mathrm{E}^1 \not\ddagger \mathrm{E}^2$ *are of the form*

- $\mathrm{V}^1 \sqcup \mathrm{V}^2 = \{x_1 \in V^1\} \cup \{x_2 \in V^2 / \forall x_1 \in V^1 \; l_{V^1}(x_1) \neq l_{V^2}(x_2)\}$.
- $\mathrm{E}^1 \not\ddagger \mathrm{E}^2 = \{x_1 \xrightarrow{t_1} y_1 \in E^1\} \cup \{x_2 \xrightarrow{t_2} y_2 \in E^2 / \; \forall (x_1 \xrightarrow{t_1} y_1) \in E^1 \; (l_{V^1}(x_1) \neq l_{V^2}(x_2) \vee l_{E^1}(x_1 \xrightarrow{t_1} y_1) \neq l_{E^2}(x_2 \xrightarrow{t_2} y_2) \vee l_{V^1}(y_1) \neq l_{V^2}(y_2))\}$.

Note: This operation does not include nodes created by duplication or marked, according what is explained in Definition 28.

We will use the terms α_m, β_n and γ_r to represent the principal connective of the formula α, β and γ respectively. In next definitions, the graph has been simplified to improve readability, and to explain the details.

Definition 8 (Formula graph) *A* formula graph *G is a directed graph* $\langle \mathrm{N}, \mathrm{A}, \mathrm{B} \rangle$ *where:* N *is a set of vertices (or nodes),* A *is a set of labeled* edges $\langle v \in \mathrm{N}, \, t \in \mathrm{E}_F\text{-Labels}, \, v' \in N \rangle$ *of source v, target v' and label* t *and is identified with the arrow* $v \xrightarrow{t} v'$, B *is a set of labels* $b \in$ F∪E_F-Labels.

Formula graphs are recursively defined as follows:

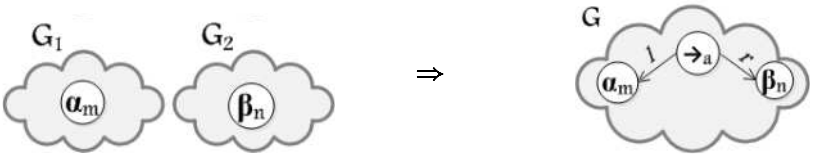
Basis *One propositional letter* p *is a formula graph.*

➔ *If G_1 is a formula graph with root node α_m and G_2 is a formula graph with root node β_n, then the graph G that is defined as $G_1 \oplus G_2$ with*

1. *an F-node* $\rightarrow_q$,

2. *edges:* $\rightarrow_q \xrightarrow{l} \alpha_m$ *and* $\rightarrow_q \xrightarrow{r} \beta_n$*,*

is a formula graph (see the following figure).



Definition 9 (Mimp-graph) *A* mimp-graph G *is a directed graph* $\langle \mathrm{V}, \mathrm{E}, \mathrm{L}\rangle$ *where:* V *is a set of nodes,* L *is a subset of* LBL, E *is a set of labeled* edges $\langle v \in \mathrm{V},\ t \in \mathrm{E}_{F \cup M}\text{-Labels},\ v' \in \mathrm{V}\ \rangle$ *of source* v*, target* v' *and label* t *and identified with the arrow* $v \xrightarrow{t} v'$*.*

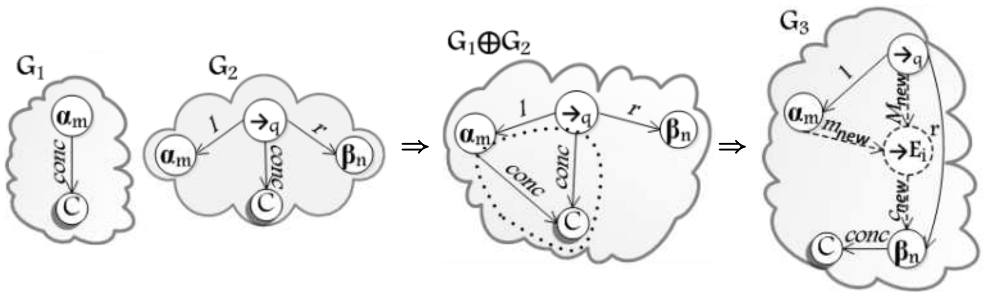
Mimp-graphs are recursively defined as follows:

Basis *If* G_1 *is a formula graph with root node* α_m *then the graph* G_2 *defined as* G_1 *with delimiter nodes* H_n *and* C *and edges* $\alpha_m \xrightarrow{conc} C$ *and* $H_n \xrightarrow{hyp} \alpha_m$ *is a mimp-graph.*

→E *If* G_1 *and* G_2 *are mimp-graphs, and the graph (intermediate step) obtained by* $G_1 \oplus G_2$ *contains the edge* $\rightarrow_q \xrightarrow{l} \alpha_m$ *and two nodes* $\rightarrow_q$ *and* α_m *linked to the delimiter node* C*, then the graph* G_3 *that is defined as* $G_1 \oplus G_2$ *with*

1. *the removal of ingoing edges in the node* C *which were generated in the intermediate step (see the figure below, dotted area in* $G_1 \oplus G_2$*);*
2. *an R-node* $\rightarrow\mathrm{E}_i$ *at the top position;*
3. *edges:* $\alpha_m \xrightarrow{m_{new}} \rightarrow\mathrm{E}_i$*,* $\rightarrow_q \xrightarrow{M_{new}} \rightarrow\mathrm{E}_i$*,* $\rightarrow\mathrm{E}_i \xrightarrow{c_{new}} \beta_n$ *and* $\beta_n \xrightarrow{conc} C$*, where new is a fresh (new) index ranging over all edges of kind c, m and M ingoing and/or outgoing of F-nodes* α_m*,* β_n *and* $\rightarrow_q$*;*

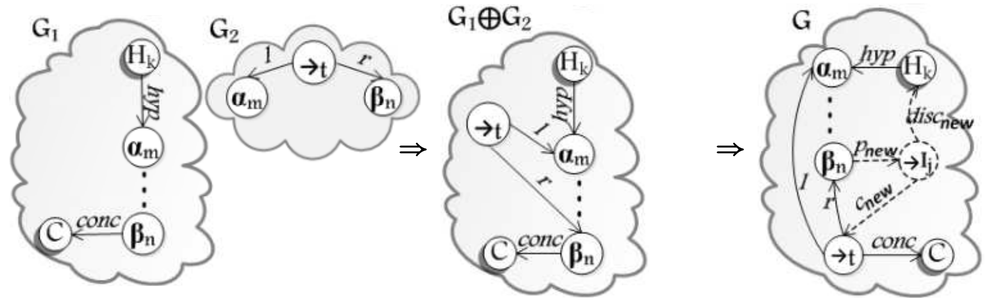
is a mimp-graph (see the following figure).



→I *If* G_1 *is a mimp-graph and contains a node* β_n *linked to the delimiter node* C *and the node* α_m *linked to the delimiter node* H_k*, then the graph* G *that is defined as*

1. *$G := G_1 \oplus G_2$, such that G_2 is a formula graph with root node $\rightarrow_t$ linked to F-nodes α_m and β_n by edges: $\rightarrow_t \xrightarrow{l} \alpha_m$, $\rightarrow_t \xrightarrow{r} \beta_n$;*
2. *with the removal of edges: $\beta_n \xrightarrow{conc} C$;*
3. *an R-node $\rightarrow\mathrm{I}_j$ at the top position;*
4. *edges: $\beta_n \xrightarrow{p_{new}} \rightarrow\mathrm{I}_j$, $\rightarrow\mathrm{I}_j \xrightarrow{c_{new}} \rightarrow_t$, $\rightarrow_t \xrightarrow{conc} C$ and $\rightarrow\mathrm{I}_j \xrightarrow{disc_{new}} H_k$, where new is a fresh (new) index considering all edges of kind p, disc and c ingoing and/or outgoing of F-nodes α_m, β_n and $\rightarrow_q$;*

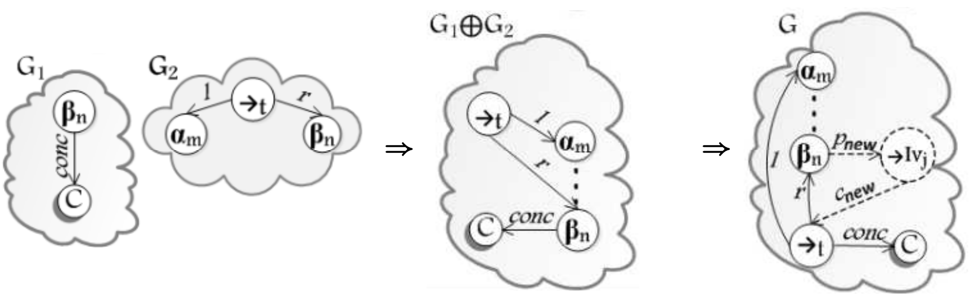
is a mimp-graph (see the following figure; the α_m-node is discharged*).*



$\rightarrow$Iv *If G_1 is a mimp-graph and contains a node β_n linked to the delimiter node C, then the graph G that is defined as*

1. *$G := G_1 \oplus G_2$, such that G_2 is a formula graph with root node $\rightarrow_t$ linked to F-nodes α_m and β_n by edges: $\rightarrow_t \xrightarrow{l} \alpha_m$, $\rightarrow_t \xrightarrow{r} \beta_n$;*
2. *with the removal of the edge $\beta_n \xrightarrow{conc} C$;*
3. *an R-node $\rightarrow\mathrm{Iv}_j$ at the top position;*
4. *edges: $\beta_n \xrightarrow{p_{new}} \rightarrow\mathrm{Iv}_j$, $\rightarrow_t \xrightarrow{conc} C$ and $\rightarrow\mathrm{Iv}_j \xrightarrow{c_{new}} \rightarrow_t$, where new is an index under the same conditions of the previous case;*

is a mimp-graph (see the following figure).



Definition 10 *Let G be a mimp-graph. An* inferential order $<$ *on nodes of G is a partial ordering of the R-nodes of G such that $n < n'$ iff n and n' are*

R-nodes and there is an F-node f such that $n \xrightarrow{lbl_1} f \xrightarrow{lbl_2} n'$ and lbl_1 is c and lbl_2 is m, or lbl_1 is c and lbl_2 is M, or lbl_1 is c and lbl_2 is p. Node n is a top position *node if n is maximal w.r.t. $<$.*

In order to avoid overloading of indexes, we will omit, whenever possible, the indexing of edges of kind c, m, M, p and *disc*, keeping in mind that the coherence of indexing is established by the kind of rule-node to which they are linked.

Lemma 1 enables us to prove that a given graph G is a mimp-graph. We just have to check that G has an inferential ordering on all R-nodes and that each node of G is of one of the possible types that generate the Basis and the construction cases $\rightarrow$E, $\rightarrow$I and $\rightarrow$Iv of Definition 9.

Lemma 1 *G is a mimp-graph if and only if the following properties hold:*

1. *There exists a well-founded (hence acyclic) inferential order $<$ on all R-nodes of the mimp-graph*[2]*.*

2. *Every node N of G is of one of the following six types:*

   P *N is labeled with one of the propositional letters: $\{p,\ q,\ r,\ \ldots\ \}$. N has no outgoing edges l and r.*

   K *N has label $\rightarrow_n$ and has exactly two outgoing edges with label l and r, respectively. N may has outgoing edges with labels p_i, m_j or M_k; and ingoing edges with label c_l and hyp_m.*

   E *N has label $\rightarrow\mathrm{E}_i$ and has exactly one outgoing edge $\rightarrow\mathrm{E}_i \xrightarrow{c} \beta_n$, where β_n is a node type* **P** *or* **K**. *N has exactly two ingoing edges $\alpha_m \xrightarrow{m} \rightarrow\mathrm{E}_i$ and $\rightarrow_q \xrightarrow{M} \rightarrow\mathrm{E}_i$, where α_m is a node type* **P** *or* **K**. *There are two outgoing edges from the node $\rightarrow_q$: $\rightarrow_q \xrightarrow{l} \alpha_m$ and $\rightarrow_q \xrightarrow{r} \beta_n$.*

   I *N has label $\rightarrow\mathrm{I}_j$ (or $\rightarrow\mathrm{Iv}_j$, if discharges an hypothesis vacuously), has one outgoing edge $\rightarrow\mathrm{I}_j \xrightarrow{c} \rightarrow_t$, and one (or zero for the case $\rightarrow$Iv) outgoing edge $\rightarrow\mathrm{I}_j \xrightarrow{disc} H_k$. N has exactly one ingoing edge: $\beta_n \xrightarrow{p} \mathrm{I}_j$, where β_n is a node type* **P** *or* **K**. *There are two outgoing edges from the node $\rightarrow_t$: $\rightarrow_t \xrightarrow{l} \alpha_m$ and $\rightarrow_t \xrightarrow{r} \beta_n$ such that there is one (or zero for the case $\rightarrow$Iv) ingoing edge to the node α_m: $\mathrm{H}_k \xrightarrow{hyp} \alpha_m$.*

   H *N has label H_k and has outgoing edges with label hyp.*

[2]We can extend this "dashed" inferential order $<$ to the full "mixed" order $<^*$ by adding new "solid" relations $<$ corresponding to arrows $\xrightarrow{l}$ and $\xrightarrow{r}$ between F-nodes. Note that $<^*$ may contain cycles (see Figure 3.4). However all recursive definitions and inductive proofs to follow are based on the well-founded "dashed" order $<$, hence being legitimate.

C *N has label* C *and has exactly one ingoing edge with label conc.*

Proof:

$\Rightarrow$: By induction on the construction of mimp-graph (Definition 9). For every construction case for mimp-graphs we have to check the three properties stated in Lemma. Property (2) is immediate. For property (1), we know from the induction hypothesis that there is an inferential order $<$ on R-nodes of the mimp-graph. In the construction cases $\rightarrow$I, $\rightarrow$Iv or $\rightarrow$E, we make the new R-node that is introduced highest in the <-ordering, which yields an inferential ordering on R-nodes. In the construction case $\rightarrow$E, when we have two inferential orderings, $<_1$ on G_1 and $<_2$ on G_2. Then $G_1 \oplus G_2$ can be given an inferential ordering by taking the union of $<_1$ and $<_2$ and in addition putting $n < m$ for every R-node n, m such that $n \in G_1$ and $m \in G_2$.

$\Leftarrow$: By induction on the number of R-nodes of G. Let $<$ be the topological order that is assumed to exist. Let n be the R-node that is maximal w.r.t. $<$. Then n must be on the top position. When we remove node n, including its edges linked (if n is of type **I**) and the node type **C** is linked to the premise of the R-node, we obtain a graph G' that satisfies the properties listed in Lemma. By induction hypothesis we see that G' is a mimp-graph. Now we can add the node n again, using one of the construction cases for mimp-graphs: *Basis* if n is a **P** node or **K** node, $\rightarrow$E if n is an **E** node, $\rightarrow$I if n is an **I** node. ∎

It is natural to consider minimal mimp-graph-like representations of given natural deductions. Actually one can try to minimize the number of F-nodes and/or R-nodes, but in this version we consider only the F-option, as it helps to reduce the size under standard normalization (see the next section). To grasp the point, note that mimp-graph in Figure 3.4 (see above) is F-minimal, i.e., its F-labeled nodes refer to pairwise distinct formulas. This observation is summarized by Theorem 3.

Theorem 3 (F-minimal representation) *Every standard tree-like natural deduction* Π *has a uniquely determined (up to graph-isomorphism)* F-minimal mimp-like representation G_Π *that satisfies the following four conditions.*

1. G_Π *is a mimp-graph whose size does not exceed the size of* Π.
2. Π *and* G_Π *both have the same (set of) hypotheses and the same conclusion.*
3. *There is graph homomorphism* $h : \Pi \to G_\Pi$ *that is injective on R-Labels.*
4. *All F-Labels occurring in* G_Π *denote pairwise distinct formulas.*

Proof: Let N and Form be the set of nodes and formulas, respectively, occurring in Π. Note that Π determines a fixed surjection f : N$\to$Form that may not be injective (for in Π, one and the same formula may be assigned to different nodes). In order to obtain G_Π take as R-nodes the inferences occurring in Π assigned with the corresponding R-Labels representing inferences' names (possibly indexed, in order to achieve a 1–1 correspondence between inferences and R-Labels, cf. Figure 3.4). Define basic F-nodes of G_Π as formulas from Form assigned with the corresponding F-Labels representing formulas' principal connectives (possibly indexed, in order to achieve an 1–1 correspondence between formulas and F-Labels, cf. Figure 3.4). So the total number of all basic F-nodes of G_Π is the cardinality of the set Form, while f being a mapping from nodes of Π onto the basic F-nodes of G_Π. To complete the construction of G_Π we add, if necessary, the remaining F-nodes labeled by failing representations of subformulas of $f(x)$, $x \in$ N, and define the E-Labels of G_Π, accordingly. Note that by the definition all nodes of G_Π have pairwise distinct labels. In particular, every F-Label occurs only once in G_Π, which yields the crucial condition 4. ■

3.2 Normalization for mimp-graphs

In this section we define the normalization procedure for mimp-graphs. It is based on the standard normalization method given by Prawitz. Thus a *maximal formula* in mimp-graphs is a $\to$I followed by a $\to$E of the same formula graph (see Definition 11). It is the same notion of maximal formulas that is being used in natural deduction derivations. So a maximal formula occurrence is the consequence of an application of an introduction rule and major premise of an application of an elimination rule. But here we assume that derivations represented by mimp-graphs. We wish to eliminate such maximal formula by dropping nodes and edges that are involved in the maximal formula.

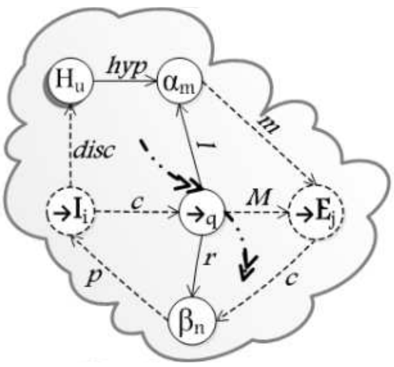
Definition 11 *A* maximal formula m *in a mimp-graph* G *(see the figure below, where the double-headed arrows represent several edges) is a sub-graph of* G *consisting of:*

1. *F-nodes* α_m, β_n, $\to_q$, *the R-node* $\to\mathrm{I}_i$ *and the delimiter node* H_u*;*

2. *the R-node* $\to\mathrm{E}_j$ *at the top position;*

3. *edges:* $\to_q \xrightarrow{l} \alpha_m$, $\to_q \xrightarrow{r} \beta_n$, $\beta_n \xrightarrow{p} \to\mathrm{I}_i$, $\to\mathrm{I}_i \xrightarrow{c} \to_q$, $H_u \xrightarrow{hyp} \alpha_m$, $\to\mathrm{I}_i \xrightarrow{disc} H_u$, $\alpha_m \xrightarrow{m} \to\mathrm{E}_j$, $\to_q \xrightarrow{M} \to\mathrm{E}_j$ *and* $\to\mathrm{E}_j \xrightarrow{c} \beta_n$*;*

$$\dfrac{\begin{array}{c}\Pi_1\\ \alpha\end{array} \qquad \dfrac{\begin{array}{c}[\alpha]^u\\ \Pi_2\\ \beta\end{array}}{\alpha \to \beta}\,(\to\text{I, u})}{\begin{array}{c}\beta\\ \Pi_3\end{array}}\,\to\text{E} \qquad \Rightarrow$$



However, as a special case of maximal formula could also happen that between the R-nodes →I and →E there are several other maximal formulas such as the case in the example of Figure 3.6, where there is a maximal formula with R-nodes $\to I_4$ and $\to E_5$ (dotted area with white background) inside of the maximal formula with R-nodes $\to I_3$ and $\to E_6$ (area with shaded background). That is, the inferential orders of R-nodes are intermediate to those of the R-nodes $\to I_3$ and $\to E_6$. In these cases eliminate in one step the maximal formula, with the exception of the R-nodes $\to_1$ and $\to_2$ because they are still related with other nodes, becoming as shown in the same figure. We can visualize another maximal formula (with R-nodes $\to I_9$ and $\to E_{10}$) that is pending removal.

Definition 12 *(1) For $n_i \in V$, a* path *in a proof-graph is a sequence of vertices and edges of the form $n_1 \xrightarrow{l_1} n_2 \xrightarrow{l_2} \ldots \xrightarrow{l_{k-2}} n_{k-1} \xrightarrow{l_{k-1}} n_k$ such that n_1 is a hypothesis F-node, n_k is the conclusion F-node, n_i alternating between an R-node and an F-node. edges l_i alternate between two types of edges: $l_j \in \{m, M, p\}$ and $l_j = c$. (2) A* branch *is an initial part of a* path *which stops at the conclusion F-node or at the first minor premise whose major premise is the conclusion of an R-node.*

The term *R-node sequence* is representing a deduction, and if it is a smaller part of another R-node sequence (subdeduction), then it is called *a subsequence* of the latter. A subsequence that derives a premise of the last R-node application in an R-node sequence is called a direct R-node subsequence. Instead of writing "the direct R-node subsequence that derives the minor premise of the last inference of an R-node sequence D", we simply write *"the minor subsequence of D".*

Definition 13 *A reordering of a given mimp-graph G is obtaining by supplying G with the following (new) inferencial order on the R-nodes of G.*

- $o(t_m) = 0$ *for an R-node t_m starting with hypothesis.*
- $o(t) = o(t') + 1$ *if the conclusion formula of R-node t' is premise, right premise or major premise of t.*

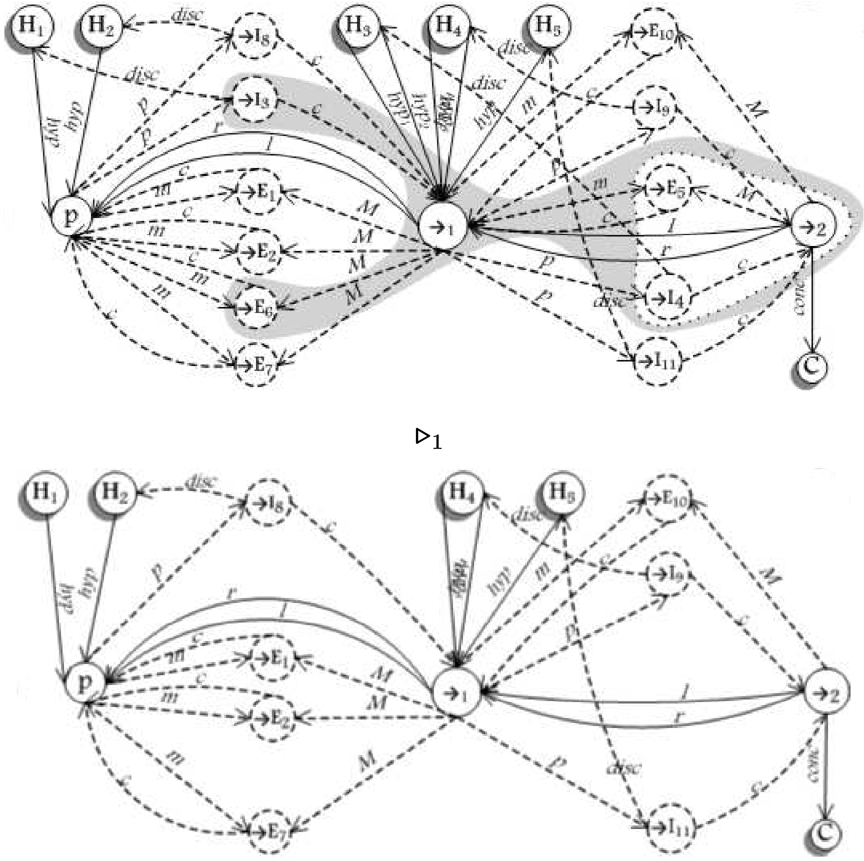


Figure 3.6: Example of a maximal formula with an intermediate maximal formula.

Proposition 1 *A graph obtained by a reordering according to Definition 13 is a mimp-graph.*

Definition 14 *Consider a mimp-graph G with a maximal formula m, that satisfies the following requirements:*

1. *Between the R-nodes* $\rightarrow\mathrm{I}_i$ *and* $\rightarrow\mathrm{E}_l$ *there are zero or more maximal formulas with inferential orders within the range of these rule nodes.*

2. *There is an edge* $\rightarrow\mathrm{I}_i \xrightarrow{c} \rightarrow_q$ *and the F-node* $\rightarrow_q$ *has zero or more ingoing edges.*

3. *There is an edge* $\rightarrow_q \xrightarrow{M} \rightarrow\mathrm{E}_l$ *and the F-node* $\rightarrow_q$ *is the premise of zero or more of another R-nodes.*

4. *If a branch will be separated from the inferential order this branch must be insertable in the following branch, according to the order, i.e., the conclusion of this separated branch is the premise in the following branch.*

The elimination of a maximal formula *m from G is the following operation on a mimp-graph (see Figure 3.7, the double-headed arrows are representing several edges):*

1. *If there is no maximal formula between the R-nodes* $\rightarrow\mathrm{I}_i$ *and* $\rightarrow\mathrm{E}_l$ *then follow these steps:*

   (a) *If the edge* $\rightarrow\mathrm{I}_i \xrightarrow{c} \rightarrow_q$ *is the only ingoing edge to* $\rightarrow_q$ *and the edge* $\rightarrow_q \xrightarrow{M} \rightarrow\mathrm{E}_l$ *is the only outgoing edge from* $\rightarrow_q$ *then remove edges to and from the F-node* $\rightarrow_q$*, and remove the F-node* $\rightarrow_q$*.*

   (b) *If the D-node* H_u*, discharged by* $\rightarrow\mathrm{I}$*, has* n *outgoing edges with label hyp then repeat* n*-times edges in the minor subsequence of* $\rightarrow\mathrm{E}_l$*.*

   (c) *Remove edges to and from nodes* $\rightarrow\mathrm{I}_i$*,* $\rightarrow\mathrm{E}_l$ *and* H_u*.*

   (d) *Remove nodes* $\rightarrow\mathrm{I}_i$*,* $\rightarrow\mathrm{E}_l$ *and* H_u*.*

   (e) *Apply the operation defined in Definition 13 to the resulting graph. Note that Proposition 1 ensures that the result is a mimp-graph.*

2. *Otherwise eliminate the maximal formulas between the R-nodes* $\rightarrow\mathrm{I}_i$ *and* $\rightarrow\mathrm{E}_l$ *as in the previous step.*

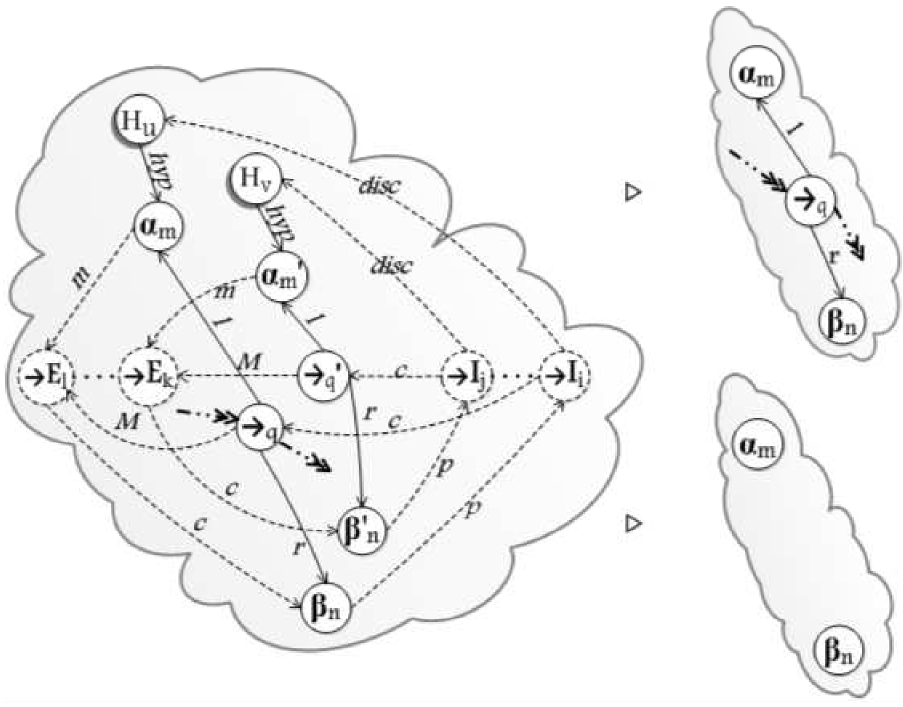


Figure 3.7: Elimination of a maximal formula in mimp-graphs.

Note that the removal of a node type I generated by case $\rightarrow$Iv in Definition 23, disconnects the graph, meaning that the sub-graph hypotheses linked, by the edge with label m, to the node labeled $\rightarrow$E removed, is no longer connected to the delimiter node type C.

Figure 3.8 shows an instance of the elimination of a maximal formula in tree form. Note that this example shows the reason why essentially our (weak)

$$\dfrac{\begin{array}{c}\Pi_2\\ \beta\end{array} \qquad \dfrac{\begin{array}{c}\Pi_1\\ \alpha\end{array} \qquad \dfrac{\dfrac{\begin{array}{c}[\beta]^v[\alpha]^u\\ \Pi_0\\ \gamma\end{array}}{\beta \to \gamma}\,v}{\alpha \to (\beta \to \gamma)}\,u}{\beta \to \gamma}}{\gamma} \quad \triangleright_1 \quad \dfrac{\begin{array}{c}\Pi_2\\ \beta\end{array} \qquad \dfrac{\begin{array}{c}[\beta]^v \quad \begin{array}{c}\Pi_1\\ \alpha\end{array}\\ \Pi_0\\ \gamma\end{array}}{\beta \to \gamma}}{\gamma} \quad \triangleright_2 \quad \begin{array}{c}\Pi_2 \quad \Pi_1\\ \beta \qquad \alpha\\ \Pi_0\\ \gamma\end{array}$$

$$\Downarrow trans$$

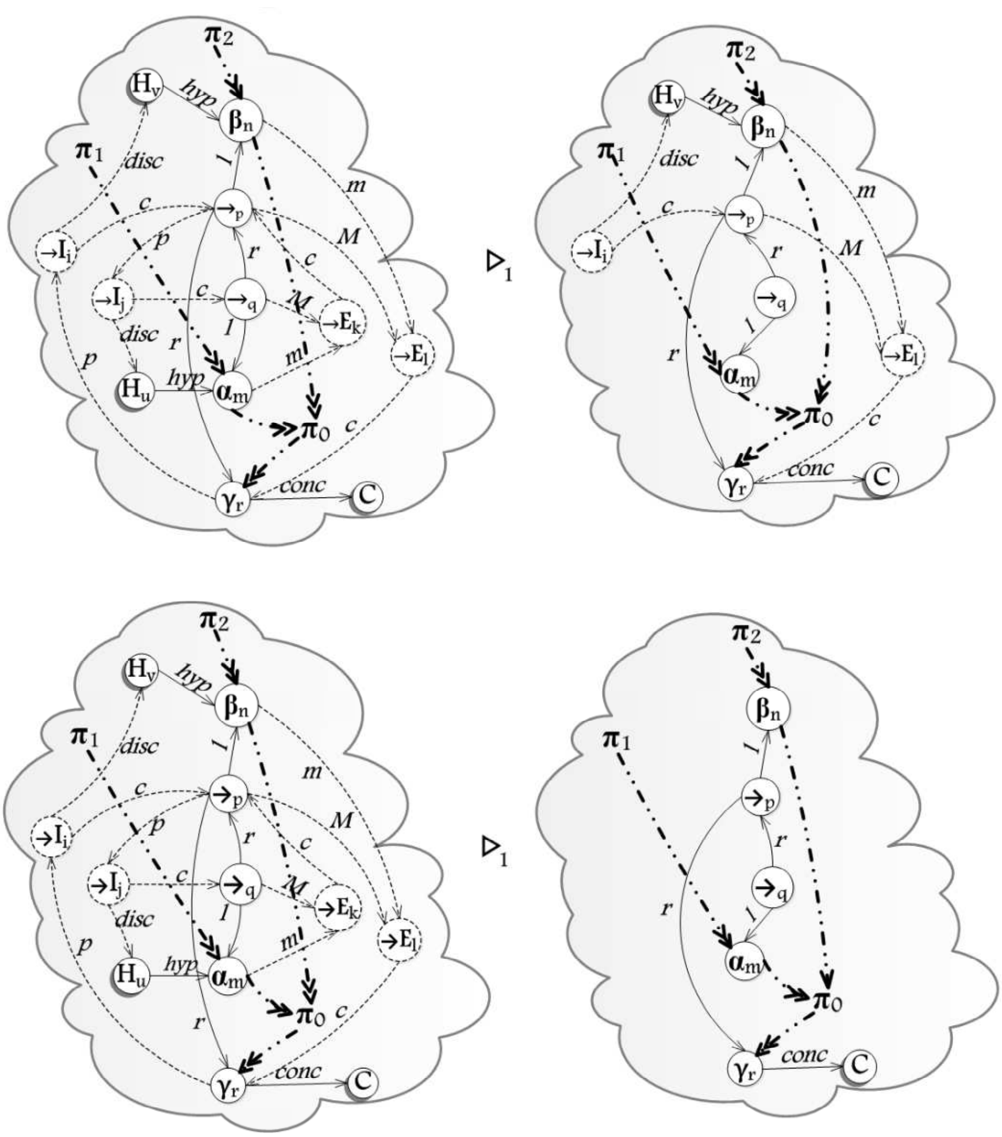


Figure 3.8: Eliminating a maximal formula in a natural deduction proof and its mimp-graph translation.

normalization theorem is directly a strong normalization theorem. The formula $\beta \rightarrow \gamma$ is not a maximal formula before a reduction is applied to eliminate the maximal formula $\alpha \rightarrow (\beta \rightarrow \gamma)$. This possibility of having hidden maximal formulas in Natural Deduction is the main reason to use more sophisticated methods whenever proving strong normalization. In mimp-graphs there is no possibility to hide a maximal formula because all formulas are represented only once in the graph. In this graph $\beta \rightarrow \gamma$ is already a maximal formula. We can choose to remove any of the two maximal formulas. If $\beta \rightarrow \gamma$ is chosen to be eliminated, by the mimp-graph normalization procedure, its reduction eliminates the $\alpha \rightarrow (\beta \rightarrow \gamma)$ too. On the other hand, the choice of $\alpha \rightarrow (\beta \rightarrow \gamma)$ to be reduced only eliminates itself. In any case, the number of maximal formula decreases.

Lemma 2 *If G is a mimp-graph with a maximal formula m and G' is obtained from G by eliminating m, then G' is also a mimp-graph.*

Proof: We use Lemma 1. All nodes in G' are of the right form: P, F, E, I, H or C. We verify that G' has one ingoing edge with label *conc* to the delimiter node type C and that is acyclic and connected. Finally, a referential order on G' (as defined in Definition 13) between R-nodes must preserve. ■

We shall construct the normalization proof for mimp-graphs. This proof is guided by the normalization measure. That is, the general mechanism from the proof determines that a given mimp-graph G should be transformed into a non-redundant mimp-graph by applying reduction steps and at each reduction step the measure must be decreased. The normalization measure will be the number of maximal formulas in the mimp-graph.

Also note an important observation concerning F-minimal mimp representations (see Theorem 3). Since F-minimal mimps can have at most one occurrence of hypotheses α and/or β, every proper reduction step will diminish the size of deduction. Hence the size of the graph (= the number of nodes) can serve as another inductive parameter, provided that the normalization is being applied to F-minimal mimp-graph representations.

Theorem 4 (Normalization) *Every mimp-graph G can be reduced to a normal mimp-graph G' having the same hypotheses and conclusion as G. Moreover, for any standard tree-like natural deduction Π, if $G := G_\Pi$ (the F-minimal mimp-like representation of Π, cf. Theorem 3), then the size of G' does not exceed the size of G, and hence also Π.*

Remark 1 *The second assertion sharply contrasts to the well-known exponential speed-up of standard normalization. Note that the latter is a consequence* *of the tree-like structure of standard deductions having different occurrences of equal hypotheses formulas, whereas all formulas occurring in F-minimal mimp-like representations are pairwise distinct.*

Proof: This characteristic of preservation of the premises and conclusions of the derivation is proved naturally. Through an inspection of each elimination of maximal formula is observed that the reduction step (see Definition 14) of the mimp-graph does not change the set of premises and conclusions (indicated by the delimiter nodes type H and C) of the derivation that is being reduced.

In addition, the demonstration of this theorem has two primary requirements. First, we guarantee that through the elimination of maximal formulas in the mimp-graph, cannot generate more maximal formulas. The second requirement is to guarantee that during the normalization process, the normalization measure adopted is always reduced.

The first requirement is easily verifiable through an inspection of each case in the elimination of maximal formulas. Thus, it is observed that no case produces more maximal formulas. The second requirement is established through the normalization procedure and demonstrated through an analysis of existing cases in the elimination of maximal formulas in mimp-graphs. To support this statement, it is used the notion of normalization measure, we adopt as measure of complexity (induction parameter) the number of maximal formulas $Nmax(G)$. Besides, as already mentioned, working with F-mimimal mimp-graph representations we can use as optional inductive parameter the ordinary size of mimp-graphs. ■

Normalization Process

We know that a specific mimp-graph G can have one or more maximal formulas represented by $M_1, ..., M_n$. Thus, the normalization procedure is:

1. Choose a maximal formula represented by M_k.
2. Identify the respective number of maximal formulas $Nmax(G)$.
3. Eliminate M_k as defined in Definition 14, creating a new graph G.
4. In this elimination, one of the following three cases may occur:

   a) The maximal formula is removed.

   b) The maximal formula is removed but the F-node is maintained, and, $Nmax(G)$ is decreased;

   c) All maximal formulas are removed.

5. Repeat steps 1 to 4 until the normalization measure $Nmax(G)$ is reduced to 0 and G becomes a normal mimp-graph.

Since the process of eliminating a maximal formula on mimp-graphs always ends in the elimination of at least one maximal formula, and with the decrease in the number of vertices of the graph, we can say that this normalization theorem is directly a strong normalization theorem.

3.3 Summary

In this chapter, we introduced the mimp-graph through the main definitions and examples, mainly devised for extracting proof-theoretic properties from proof system. That is, we have tackled one of our research tracks. Mimp-graphs preserve the ability to represent proofs in Natural Deduction and their minimal formula representation is a key feature of the mimp-graph structure, because as we saw earlier, it is easy to determine an upper bounds on the length of reduction sequences leading to normal proofs. It is the number of maximal formulas. This feature is of crucial importance because we intend to use this method in automated theorem provers.