

## Chapter 2

# Graph Theoretical Approach

### Notations

$A-B$	Difference between sets A and B
$A \cup B$	Union of sets A and B
$A \cap B$	Intersection of sets A and B
$A \subset B$	A is a subset of B
$X \in A$	x is an element of A
$X \notin A$	x is not an element of A
$ A $	Cardinality of A (=number of elements of A)
$\emptyset$	Empty set
$G = (X, E)$	Non-oriented graph (coexistences)
$G_s = (X, U)$	Oriented graph (superpositions)
$G^* = (X, E + U)$	Biostratigraphic graph
$U$	Set of arcs of $G_s$ and $G^*$
$E$	Set of edges of $G$ and $G^*$
$X$	Set of vertices of $G$ , $G_s$ and $G^*$
$\{x, y\}$	Edges of $G$ and $G^*$
$(x, y)$	Arcs of $G_s$ and $G^*$
$G^c$	Complement of $G$
$\Gamma(x)$	Set of neighbors of x in $G$ (neighborhood)
$\Gamma^+(x)$	Set of successors of x in $G_s$ and $G^*$
$\Gamma^-(x)$	Set of predecessors of x in $G_s$ and $G^*$
$Z_n$	Circuit of length n without chord in $G$ and $G^*$
$C_n$	Cycle of length n in $G_s$ and $G^*$
$S_n$	Semi-oriented circuit of length n in $G^*$
$\Rightarrow$	Implies that
$\Leftrightarrow$	If and only if
$ $	Such that
$\exists$	There exists
$\neq$	Different from

## 2.1 Representing Stratigraphic Relationships

The ideas summarized in Figs. 1.1–1.3 can be expressed with the help of certain aspects of graph theory. This field of applied mathematics provides several useful theorems for overcoming the concrete difficulties so often met in biochronology. In addition, its notation is convenient to describe the algorithms which are used in the UAgaph program both to explain the structure of difficult biostratigraphic data, and to construct unitary associations and identify them in fossil-bearing beds.

The basic concepts of graph theory have been made comprehensible to non-specialists in several classical works: Golumbic (2004) and especially Roberts (1976, 1978). Stratigraphers interested in the mathematical aspects of the problems discussed in the following chapters can refer to these books.

## 2.2 Definitions

### 2.2.1 Graph

In introducing the technical definitions that follow, we will first appeal to the paleontologist's intuition. Imagine that the fossil species whose relationships are being analyzed are represented by a set of points (or vertices) and by oriented or non oriented lines (arcs and edges): the resulting diagram is called a graph.

1. the stratigraphic superpositions observed among these species can be represented by arcs (arrows) joining the points. These arcs are, by convention, oriented in the sense  $x \rightarrow y$  when species  $x$  has been observed below species  $y$ ;
2. coexistences (observed or virtual) among the species are represented by edges (lines) joining the points  $y$  and  $z$  and indicating that  $y$  and  $z$  occur together.

The absence of known stratigraphic relation between two species (vertices of the graph) means that their relation is not defined.

Three kind of graphs should be distinguished in our discussion.

### 2.2.2 The Non Oriented Graph $G$

A non-oriented graph  $G = (X, E)$  is composed of a finite set of points  $X = \{x_1, x_2, \dots, x_n\}$  and a family  $E = \{e_1, e_2, \dots, e_m\}$  of unordered pairs of distinct points of  $X$ .

Each pair  $e = \{x, y\}$  of  $E$  is called an edge of the graph, and the points of  $X$  are called the vertices of the graph.

The order of the graph is the number of points of  $X$ . It is denoted  $|X|$ .

The edges of  $G$  (i.e. the elements of  $E$ ) represent neighbour (or compatible) species. Edges that represent truly associated pairs of species (i.e. observed

coexistences) are called real edges; those that represent virtually associated pairs of species are called virtual edges.

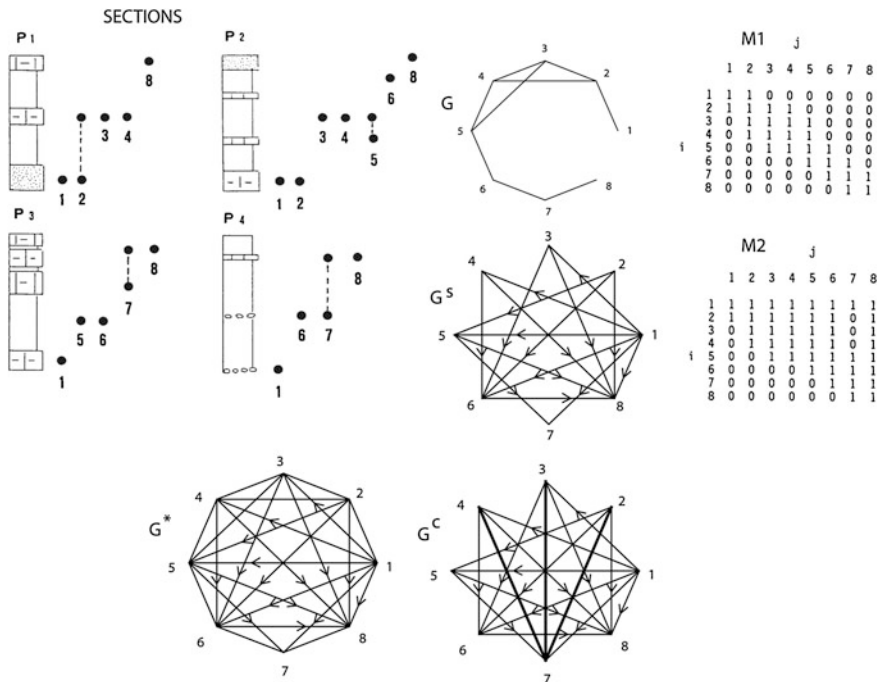
Two vertices connected by an edge are neighbors in  $G$ . The set of neighbors of a vertex  $x$  is denoted  $\Gamma(x)$ : this set corresponds to the neighbourhood of the species  $x$ . It contains  $x$ .

The graph  $G$  in Fig. 2.1 represents the coexistences of species 1–8. Such a graph is called a coexistence graph.

The neighbourhood of species 4 (i.e. the set of neighbours of vertex 4 in  $G$ ) is denoted  $\Gamma(4) = \{2, 3, 5\}$ .

### 2.2.3 The Adjacency Matrix of $G$

Let  $G = (X, E)$  be a non-oriented graph. The adjacency matrix  $A = (a_{ij})$  of  $G$  is defined by



**Fig. 2.1** Stratigraphic distribution of 8 species (1–8) in four imaginary sections ( $P_1$ – $P_4$ ).  $G$  Coexistence graph of the 8 species.  $M1$  Species-species matrix associated with  $G$ .  $G^S$  Superposition graph of species 1–8.  $G^*$  Biostratigraphic graph of species 1–8.  $M2$  Species-species matrix associated with  $G^*$ .  $G^C$  Semi-oriented complementary graph of  $G$

$$a_{ij} = \begin{cases} 1 & \text{if } x_i \text{ and } x_j \text{ are neighbors} \\ 0 & \text{otherwise.} \end{cases}$$

The matrix  $M_1$  given in Fig. 2.1 is the adjacency matrix of the graph  $G$  in Fig. 2.1.

### 2.2.4 The Oriented Graph $G_s$

An oriented graph  $G_s = (X, U)$  is a finite set  $X = \{x_1, x_2, \dots, x_n\}$  of vertices and a family  $U = \{u_1, u_2, \dots, u_m\}$  of ordered pairs of distinct vertices of  $X$ . Each such pair  $u = (x, y)$  is called an arc of the graph.

Two incompatible species whose superposition is known form an arc of  $G_s$ ; by convention the arc  $(x, y)$  means that species  $x$  lies beneath species  $y$ .

The species that have been observed above species  $x$  in stratigraphic sections form the set of successors of  $x$ : this set is denoted  $I^+(x)$ . Similarly, the species that have been observed beneath  $x$  form the set of predecessors of  $x$ ; this set is denoted  $I^-(x)$ .

Such a graph is called a superposition graph  $G_s$ .

We will use the words pair and couple interchangeably, unlike some specialists in graph theory for whom pairs are unordered and couples are ordered.

### 2.2.5 The Biostratigraphic Graph $G^*$

The semi-oriented graph  $G^* = (X, E + U)$ , which combines the edges (or arcs) of  $G$  and  $G_s$ , is called a biostratigraphic graph. It contains all available information about the stratigraphic relationships of species: neighborhoods (i.e. coexistences) and exclusions with known or unknown stratigraphic relationships (Fig. 2.1).

In the following pages,  $X = \{x_1, \dots, x_n\}$  denotes a set of fossil species. The stratigraphic relationships observed among the species of  $X$  can be expressed using a matrix  $A = (a_{ij})$  as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } x_i \text{ is in the same level or in a lower level than } x_j \\ 0 & \text{otherwise.} \end{cases}$$

This matrix  $A$  is the *adjacency matrix* of the biostratigraphic graph  $G^*$ . The edges of  $G^*$  represent pairs of chronologically coexisting species. Edges that represent truly associated pairs of species (i.e. observed coexistences) are called *real* edges; those that represent virtually associated pairs of species are called *virtual* edges (Fig. 2.1).

### 2.2.6 *The Complementary Graph $G^c$*

If  $G$  is a non-oriented graph, then the complementary graph  $G^c$  of  $G$  is the graph whose vertices are the same as the vertices of  $G$ , but where an edge (or an arc) connects two vertices precisely if no edge connects the vertices in  $G$  (Fig. 2.1).

In our biochronologic problem, the edges of  $G^c$  represent any exclusions of species (i.e. superpositions and undetermined stratigraphic relationships). In practice then, we must admit that some of those edges are already oriented: these orientations represent pairs of stratigraphically superposed species. This is why, in our problem,  $G^c$  is semi-oriented (Fig. 2.1).

### 2.2.7 *Maximal Horizons and Residual Maximal Horizons*

A fossil-bearing bed in an isolated stratigraphic section is a maximal horizon if the set of species that coexist in the bed is maximal (NB: in the framework of a study limited to the section).

The maximal horizons of a given section are separated from each other by separation intervals corresponding to the minimal intersections of existence intervals. In other words, all the beds located between two minimal intersections (= separation intervals) are equivalent to the maximal horizon framed by them.

Maximal horizons that are strictly distinct from each other in all sections (with regard to the inclusion relationship) are called residual maximal horizons.

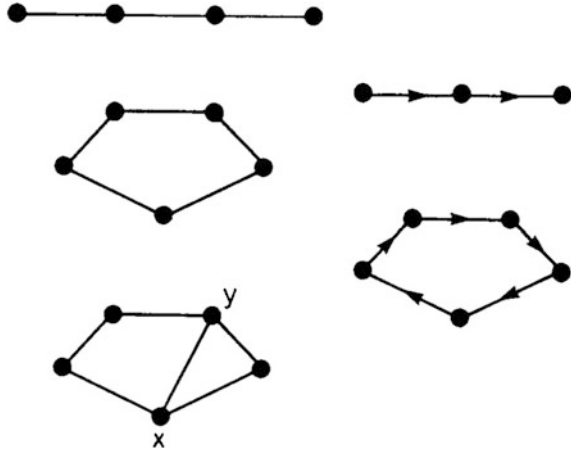
## 2.3 Basic Technical Terms

### 2.3.1 *Chain, Circuit and Chord*

In a non-oriented graph  $G$ , a chain is an alternating sequence of distinct vertices and edges beginning and ending in a vertex. The length of the chain is the number of its vertices.

A circuit is a closed chain. A chord is an edge connecting two non-consecutive vertices of a circuit. A circuit of length  $n$  admitting no chords is denoted  $Z_n$ . See Fig. 2.2.

**Fig. 2.2** Chain of length 4.  
 $Z_5$  circuit. Chord: the edge  $\{x, y\}$  is a chord of the  $Z_5$  circuit.  
 Path of length 3.  $C_5$  cycle



### 2.3.2 Path, Maximal Path, Cycle and Strongly Connected Graph

A path in an oriented graph  $G_s = (X, U)$  is a sequence of arcs  $(u_1, u_2, \dots, u_n)$ , such that the end vertex of each arc coincides with the beginning vertex of the next arc in the sequence (Fig. 2.2). A path is maximal if it is not a subpath of any longer path.

A  $C_n$  cycle is a closed path (Fig. 2.2), i.e. the end-point of its last arc coincides with the origin of its first arc.

An oriented graph  $G_s = (X, U)$  is strongly connected if, for every pair of vertices  $x$  and  $y$ , there is a path from  $x$  to  $y$  and a path from  $y$  to  $x$ .

### 2.3.3 Subgraphs, Generated Subgraphs and Strong Components

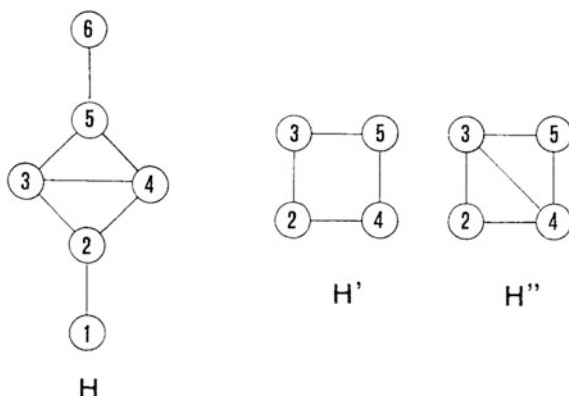
Let  $H = (X, F)$  be a graph. A graph  $H' = (X', F')$  is a subgraph of  $H = (X, F)$  if  $X'$  is a subset of  $X$  and  $F'$  is a subset of  $F$ .

If  $H'$  contains all the edges (resp. arcs) connecting vertices of  $X'$ , then  $H'$  is a generated subgraph of  $H$ .

The graph  $H'$  in Fig. 2.3 is a subgraph of  $H$ , and  $H''$  is a generated subgraph of  $H$ .

A strongly connected subgraph of an oriented graph  $G_s$  is called a strong component of  $G_s$ .

**Fig. 2.3**  $H'$  is a subgraph of  $H$  and  $H''$  is a generated subgraph of  $H$



### 2.3.4 Semi-oriented Circuit

Our biostratigraphic graph  $G^*$  is semi-oriented; similarly we will say that a sequence of edges and arcs is a semi-oriented circuit if all the edges in that sequence can be oriented to make a cycle. Such structures will be denoted  $S_n$ , where  $n$  is the number of vertices.

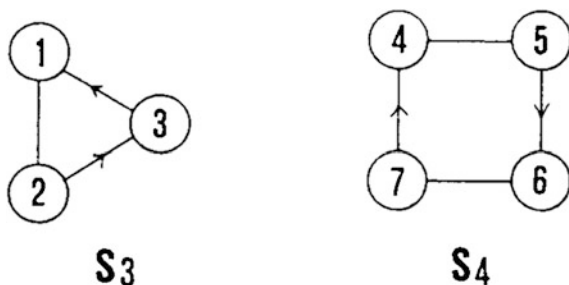
The graphs  $S_3$  and  $S_4$  in Fig. 2.4 are semi-oriented circuits: they represent conflicting stratigraphic relationships of species 1, 2 and 3 (for  $S_3$ ) and species 4, 5, 6 and 7 (for  $S_4$ ). These structures will be studied in detail in the following chapters.

### 2.3.5 Orientation of a Graph

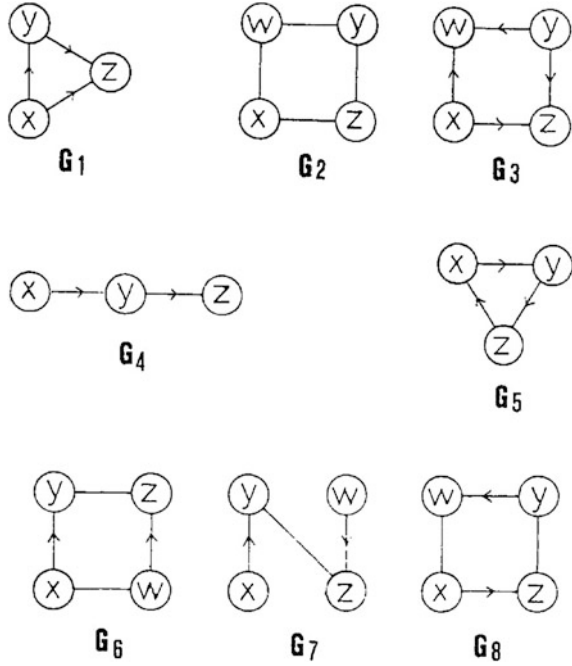
#### 2.3.5.1 Transitive Orientation

An oriented graph  $G_s = (X, U)$  is transitive if whenever  $(x, y) \in U$  and  $(y, z) \in U$ , then  $(x, z) \in U$ .

**Fig. 2.4** Semioriented graphs of length 3 and 4 denoted  $S_3$  and  $S_4$



**Fig. 2.5** Examples of transitive graphs ( $G_1$  and  $G_3$ ), non-transitive graphs ( $G_4$  and  $G_5$ ), transitively orientable graphs ( $G_2$  and  $G_8$ ) and non-transitively orientable graphs ( $G_6$  and  $G_7$ )



Graphs  $G_1$  and  $G_3$  of Fig. 2.5 are transitive, but graphs  $G_4$  and  $G_5$  are not: the arc  $(x, z)$  is missing.

### 2.3.5.2 Transitively Orientable Graph

A non-oriented graph is transitively orientable if and only if an orientation can be chosen for all the edges so that the resulting oriented graph is transitively oriented.

Graph  $G_2$  of Fig. 2.5 is transitively orientable:  $G_3$  is a transitive orientation of  $G_2$ .

The properties of transitively orientable graphs are described in detail in Roberts (1976) and Golumbic (2004).

### 2.3.5.3 Orientability of a Semi-oriented Graph

A semi-oriented graph is transitively orientable if its partial orientation can be extended to an orientation making the resulting oriented graph transitively oriented.



### 2.3.5.4 Examples

Four examples of transitively orientable and non-transitively orientable graphs are given in Fig. 2.5.

1. The edge  $\{x, z\}$  of graph  $G_6$  is lacking; if the edge  $\{y, z\}$  of  $G_6$  is oriented in the direction  $y \rightarrow z$ , the graph is not transitive. Since the edge  $\{y, w\}$  is also lacking, the opposite orientation of  $\{y, z\}$  also fails to make  $G_6$  transitive.
2. As above, the edges  $\{x, z\}$  and  $\{y, w\}$  of graph  $G_7$  are both lacking; neither orientation of  $\{y, z\}$  makes  $G_7$  transitive.
3. Graph  $G_8$  of Fig. 2.5 is transitively orientable:  $G_3$  is a transitive orientation of  $G_8$ .
4. If the edges  $\{2, 7\}$ ,  $\{3, 7\}$  and  $\{4, 7\}$  are given the orientations  $(2, 7)$ ,  $(3, 7)$  and  $(4, 7)$ , the graph of Fig. 2.1 H becomes transitively oriented.

### 2.3.6 Clique, Maximal Clique and Unitary Association

A clique of  $G$  is a collection of vertices, each of which is a neighbor of every other. A clique is maximal if it is contained in no larger clique. Maximal cliques will be denoted  $k$ .

Strictly speaking, a maximal clique has only real edges. A *unitary association*, however, is a maximal clique of which some edges may be virtual. We will use the same notation for maximal cliques and for unitary associations.

### 2.3.7 Incidence Matrix and Maximal Clique Matrix

Let  $X = \{x_1, x_2, \dots, x_n\}$  be a set of  $n$  vertices and  $S = \{s_1, \dots, s_m\}$  be a family of  $m$  subsets of  $X$ . The subset-vertex incidence matrix is the  $m$  by  $n$  matrix ( $m$  rows and  $n$  columns) whose  $(i, j)$ th entry ( $i$ th row,  $j$ th column) is 1 if vertex  $x_i$  belongs to subset  $s_j$  and 0 otherwise.

The subset-vertex incidence matrix in which  $S$  represents the maximal cliques of a graph  $G$  and  $X$  represents the vertices of  $G$  is called the maximal clique matrix of  $G$ .

Figure 1.1d is a subset-vertex incidence matrix where the subsets  $UA_1, \dots, UA_4$  are precisely the maximal cliques of the graph  $G$  in Fig. 2.1. This matrix is the maximal clique matrix of  $G$  in Fig. 2.1.

### 2.3.8 *Consecutive 1's Property*

A matrix composed of 1's and 0's has the consecutive 1's property if its rows can be permuted so that the 1's in each column have no 0's between them (Fig. 2.1 M1).

### 2.3.9 *Triangular Matrix*

A matrix is upper (resp. lower) triangular if all the entries beneath (resp. above) its main diagonal are 0.

### 2.3.10 *Triangulated Graph*

A non-oriented graph is triangulated if all its circuits of length greater than 3 have a chord.

### 2.3.11 *Forbidden Generated Subgraph*

If a class of graphs is defined by requiring that no elements of some explicit list of subgraphs appear as generated subgraphs, this class will be said to be defined by forbidden subgraphs.

For instance, triangulated graphs are defined by forbidding the list of graphs  $Z_n$ ,  $n > 3$ .

### 2.3.12 *Asteroidal Triple*

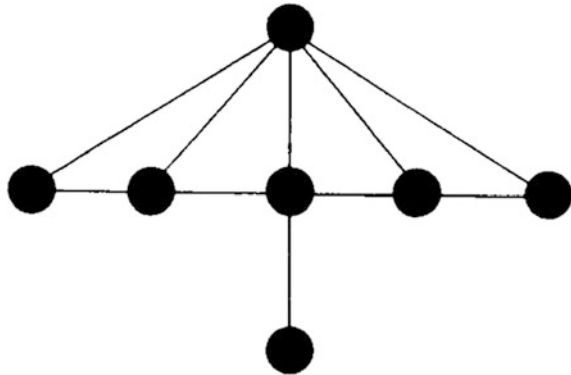
Vertices  $x, y, z$  of a non-oriented graph  $G$  form an asteroidal triple if there exists a chain  $C_1$  between  $x$  and  $y$ , a chain  $C_2$  between  $x$  and  $z$ , and a chain  $C_3$  between  $y$  and  $z$ , such that there is no edge between  $z$  and  $C_1$ ,  $y$  and  $C_2$  or  $x$  and  $C_3$ .

Figure 2.6 shows one example of asteroidal triple.

### 2.3.13 *Interval*

Technically, an interval is a line segment, which we will take to be closed, i.e. containing its own extremities.

**Fig. 2.6** Example of asteroidal triple



In our problem, we will be concerned with the time intervals corresponding to the life span of species, already called existence intervals. We should not expect an existence interval to be entirely recorded in a single section, or the record of an existence interval in different sections to be synchronous.

By convention we consider that if a fossil species has a discontinuous vertical distribution in a given stratigraphic section, it is considered as virtually present in all the beds that are flanked by its first local appearance and disappearance.

### 2.3.14 Intersection Graph

Let  $S$  be a set, and  $A = \{A_1, \dots, A_n\}$  be a family of subsets of  $S$ . The intersection graph of  $(A, S)$  has one vertex  $x_i$  for each subset  $A_i$ , and an edge joining  $x_i$  to  $x_j$  if and only if  $A_i \cap A_j \neq \emptyset$ .

For instance, to Fig. 2.1 we can associate a set  $S$  consisting of four distinct time axes  $T_1, T_2, T_3, T_4$ , one for each section, and subsets  $A_1, \dots, A_8$  of  $S$ ,  $A_i \cap T_k$  being the observed (and partially recorded) existence interval of the  $i$ th species in the  $k$ th section. Figure 2.1 (M1) is the intersection graph of this family.

More generally, the same procedure allows us to interpret any coexistence graph as an intersection graph.

### 2.3.15 Interval Graph

The intersection graph of a family of intervals on a line is called an interval graph.

This means that if  $J = \{J_1, \dots, J_n\}$  is a family of intervals, then the corresponding intersection graph has a set  $X = \{x_1, \dots, x_n\}$  of vertices, with an edge connecting  $x_i$  and  $x_j$  if and only if  $J_i$  and  $J_j$  intersect.

For instance, the graph in Fig. 2.1 is an interval graph. This is not obvious from the definition, and will not generally be true of coexistence graphs. However, it is easy to verify that this one is the interval graph associated to the intervals of Fig. 2.1.

## References

- Golumbic, M. (2004) Algorithmic graph theory and perfect graphs. *Annals of Discrete Mathematics* 57, 340.
- Roberts, F. (1976). *Discrete mathematical models*. New Jersey: Prentice Hall. 559 p.
- Roberts, F. (1978). *Graph theory and its application to problems of society* (Vol 29, 122 p.). Philadelphia: NFS-CBMS Monograph.

Discrete Biochronological Time Scales

Guex, J.; Galster, F.; Hammer, Ø.

2016, XVIII, 160 p. 41 illus., 9 illus. in color., Hardcover

ISBN: 978-3-319-21325-5