Ordered Sets for Data Analysis: An Algorithmic Approach

Ordered sets

Sergei O. Kuznetsov

National Research University Higher School of Economics, Moscow, Russia

September 27, 2022

Ordered sets

- A hierarchical relations of the type “belong to a class”, “is a”, “be a more general description than”, which reflect taxonomies of subject domains.
- There are also important hierarchical relations of the type “to be a part” describing *meronomy* of the subject domain. Both taxonomies and meronomies are naturally described by order relations.

Quasi-order and partial order

Quasi-order **is reflexive** and **transitive binary relation**.

Example. Entailment relation $\vdash$ on formulas of various logical languages is reflexive and transitive, but not antisymmetric. For example,

$$\overline{x} \vee (x \to y) \vdash \overline{x} \vee y \overline{x} \vee y \vdash \overline{x} \vee (x \to y).$$

Note: formulas $\overline{x} \vee (x \to y)$ and $\overline{x} \vee y$ are (syntactically) different, which can be important, e.g., from the viewpoint of hardware realization.

Analysis of consumer baskets

Let M be the set of goods (items) in a supermarket and C be the set of customers during a particular day.

Customers can be described by tuples of length $|M|$, where the ith component of the tuple stays for ith item and the value of the component stays for the amount bought by the customer.

Analysis of consumer baskets

Customer x bought more or equal than customer y if all components of $x-$tuple are not less than all components of $y-$tuple. The relation *"bought more or equal than"* on the set of customers C is **reflexive**, **transitive**, but **not antisymmetric**, since equality of a purchase does not mean identity of customers.

Subgraph isomorphism relation

A **labeled (weighted) graph** is a tuple of the form $\Gamma := ((V, lv), (E, le))$, where $lv : v \to L_v$ and $le : e \to L_e$ are labeling functions, which take a vertex and an edge to their labels in sets L_v and L_e, respectively.

Subgraph isomorphism relation

Labeled graph $\Gamma_1 := ((V_1, lv_1), (E_1, le_1))$ is **isomorphic to a subgraph** of labeled graph $\Gamma_2 := ((V_2, lv_2), (E_2; le_2))$ or $\Gamma_1 \lhd \Gamma_2$, if there exists an injection $\varphi : V_1 \to V_2$ such that:

- it defines isomorphism of unlabeled graph (V_1, E_1) and a subgraph of labeled graph (V_2, E_2)
- respects labels of vertices and edges: $lv_1(v) = lv_2(\varphi(v)), le_1(v, w) = le_2(\varphi(v), \varphi(w))$.

Subgraph isomorphism relation

If graph $\Gamma_1 = (V_1, E_1)$ is isomorphic to a subgraph of graph $\Gamma_2 = (V_2, E_2)$ and graph Γ_2 is isomorphic to a subgraph of graph Γ_1, i.e., $\Gamma_1 \trianglelefteq \Gamma_2$ and $\Gamma_2 \trianglelefteq \Gamma_1$, then graphs Γ_1 and Γ_2 are said to be **isomorphic**. However, Γ_1 and Γ_2 do not coincide, since they have different sets of vertices and edges.

Note: The union of quasi-order $\trianglelefteq$ and the inverse relation $\trianglerighteq$, which is also a quasiorder, i.e., $\trianglelefteq \cap \trianglerighteq$ is an equivalence relation.

Equivalence classes relation given by quasi-order

Let for two equivalence classes π, σ relation $\pi \leq \sigma$ hold if $p \trianglelefteq s$ $\forall p \in \pi, s \in \sigma$.

Statement. Relation $\leq$ on classes of equivalence given by a quasi-order $\trianglelefteq$ is reflexive, transitive, and antisymmetric, i.e., it is a partial order.

Partial order

Definition

Partially ordered set (poset) is a pair $(P, \leq)$ where $\leq$ is a *partial order relation*.

Definition

Strict order is *antireflexive, asymmetric, and transitive relation*.
Strict order $<$ related to partial order $\leq$ is obtained from $\leq$ by deleting pairs of the form (a, a):

$$x < y := x \leq y \text{ and } x \neq y$$

Partial order on functions

Consider the set of real-valued functions on real numbers: $f : R \to R$. For any two functions g and h put $g \leq h$ if $g(x) \leq h(x)\ \forall x \in R$. Thus defined "pointwise" relation on functions $\leq$ is **reflexive**, **transitive**, and **antisymmetric**.

Partitions

Recall that *partition* of set S is a set of subsets of S (called *blocks*) $\{S_1, \ldots, S_n\}$ such that

$$\bigcup_{i \in [1,n]} S_i = S, \qquad \forall i, j \in [1, n] \qquad S_i \cap S_j = \emptyset.$$

One writes $S_1 \mid S_2 \mid \ldots \mid S_n$.

Partial order on partitions

Partition P_1 is **finer than** (or **equal to**) partition P_2 (equivalently, partition P_2 is **rougher than** (or **equal to**) partition P_1), denoted $P_1 \leq P_2$, if for every block A of partition P_1 there is block B of partition P_2 so that $A \subseteq B$.

Example.

For $S = \{a, b, c, d\}$ one has $\{a, b\} \mid \{c\} \mid \{d\} \leq \{a, b, c\} \mid \{d\}$.

Statement. Relation $\leq$ on partitions is a partial order.

Multisets

Multiset on set S is set S together with function $r : S \to N \cup \{0\}$ giving multiplicity of elements of S.

For example, consumer baskets, molecular formulas, portfolio shares, etc.

Partial order on multisets

Multiset M on S is usually denoted by $\{a_{m_a} | a \in M\}$, where m_a is multiplicity of element a. Multiset $L = \{a_{l_a} | a \in L\}$ is a submultiset of multiset $M = \{a_{m_a} | a \in M\}$ ($L \subseteq M$) if for all a one has $l_a \leq m_a$. For example, for

$$S = \{a, b, c, d\} \text{ one has } \{a_1, b_5; c_2\} \subseteq \{a_3, b_6, c_2, d_2\}.$$

Proposition Relation $\subseteq$ on multisets is a partial order.

Examples

- The relation "greater or equal" on real numbers.
- The relation "to be a divisor" on natural numbers.
- Containment ("to be a subset") relation on subsets of a set.

Linear order

Definition

A partially ordered set where every two elements are comparable (i.e., $x < y$ or $y < x$) is called *linearly ordered* or *chain*. A partially ordered set where every two elements are incomparable is called an *antichain*.

Example. The sets of natural, rational, real numbers with the standard "greater than or equal" relation is a linear order.

Examples

Let A be a finite set of symbols (alphabet) which is linearly ordered by relation $\prec \subseteq A \times A$.

A **word** in alphabet A is a finite (may be empty) sequence of symbols from A.

The set of all words is denoted by A^*.

Lexicographic order

Lexicographic order $<$ on words from A^* is defined as follows:

$w_1 < w_2$ for $w_1, w_2 \in A^*$ if either w_1 is a prefix w_2 (i.e., $w_2 = w_1 v$, where $v \in A^*$) if the first symbol from the left which is different for w_1 and w_2 for w_1 is less than that of w_2 w.r.t. $\prec$. (i.e. $w_2 = wav_1$, $w_1 = wbv_2$, where $w, v_1, v_2 \in A^*$, $a, b \in A, a \prec b$).

Statement. Lexicographic order on sets of words A is a strict linear order.

Order morphisms

Mapping $\varphi : M \to N$ between *two partially ordered sets* $(M, \leq_1)$ and $(N, \leq_2)$ **respects order** if for every $x, y \in M$ one has

$$x \leq_1 y \Rightarrow \varphi x \leq_2 \varphi y$$

.

For inverse implication, when $x \leq_1 y \Leftarrow \varphi x \leq_2 \varphi y$, φ is **order embedding**.

A bijective order embedding is called **order isomorphism**. Not every bijection respecting order is an order isomorphism.

Topological sorting

Theorem of Szpilraijn (on the possibility of *topological sorting*).

Theorem

Let $(S, \leq)$ be a finite poset. Then elements of S can be enumerated in a way such that

$$S = \{s_1, \ldots, s_n\}, \qquad s_i \leq s_j \Rightarrow i \leq j.$$

Topological sorting. Proof

Proof.

Let us arbitrarily enumerate the elements of S.

- Take an arbitrary minimal element of S with the least number, let it be q.
- Put $s_1 := q$.
- Consider the poset $(S_1, \leq_1)$, $S_1 := S \setminus \{q\}$, $\leq_1 := \leq \cap S \setminus \{q\}$ and repeat the previous actions we did for $(S_1, \leq_1)$:
  - find the minimal element r with the least number;
  - put $s_2 := r$, $S_2 := S_1 \setminus \{r\}$, $\leq_2 := \leq_1 \cap S_1 \setminus \{r\}$ and arrive at the poset $(S_2, \leq_2)$.
  - Iterate the procedure until the step k such that $S_k = \emptyset$.

The result is the enumeration we are looking for.

□

Topological sorting

Both the process and the result of linear ordering of a poset are called *topological sorting*.

A linear extension of the original partial order is the result of topological sorting.

Question: How many linear extensions does an order have?

Different extensions

Theorem

Let $(S, \leq)$ be a finite poset where elements x and y are not comparable. Then there are two different topological sortings of $(S, \leq)$ such that x has a greater number than y in one of the sortings and smaller number in the other sorting.

Proof.

(Idea) The algorithm of topological sorting presented above allows for various orders of enumerating incomparable elements, so both x coming before y and y coming before x are possible. $\square$

Example

- There are usually several prerequisites for obtaining an entrance visa to a foreign country: a valid passport, an invitation, and some other documents.

  Consider the relation **"obtaining document** x **comes before obtaining document** y**"**.

  This relation defines a partial order on the starting times of applications for documents. A person who needs to get a visa makes *topological sorting* of the respective actions w.r.t. *linear order* on time instances.

Example

- A computation process running on a single processor is based on the precedence relation on intermediate computation results. This relation defines *a partial order* on time events, which is transformed by the operating system into the linear order of computations executed by the processor.

Topological sorting of a poset is a mapping that respects order. Topological sorting of a poset (where order is not linear) is an example of *bijection*, which is *not an order isomorphism*.

The covering relation and diagram of partial order

Interval $[a, b]$ is set $\{x \mid a \leq x \leq b\}$ of all elements of partial order which lie between a and b.
Let $(P, \leq)$ be a partially ordered set. Then the respective **covering relation** $\prec$ is given as follows:

$$x \prec y := x \leq y,\ x \neq y,\ \nexists z \neq x, y\ x \leq z \leq y,$$

or equivalently,

$$x \prec y := x < y,\ \nexists z\ x < z < y.$$

Representing finite order with its covering relation

Theorem

Let $a < b$ in a finite poset $(P, \leq)$. Then P contains a subset of elements $\{x_1, \ldots, x_l\}$ such that $a = x_1 \prec \cdots \prec x_l = b$.

Proof.

By induction on the number of elements y such that $a < y < b$ (i.e., y lies in the interval $[a, b]$). $\square$

Covering relation and order isomorphisms

Theorem

Let $(P, \leq_p)$ and $(Q, \leq_q)$ be finite posets with covering relations $\prec_p$ and $\prec_q$, and let $\varphi : P \to Q$ be a bijection. Then the following two statements are equivalent:

1. *Bijection φ is an order isomorphism, i.e., $x \leq_p y$ iff $\varphi(x) \leq_q \varphi(y)$.*
2. *$x \prec_p y$ iff $\varphi(x) \prec_q \varphi(y)$.*

Proof.

$1 \to 2$.

- Let φ be an order isomorphism and let $x \prec_p y$, then $x \leq_p y$ and $\varphi(x) \leq_q \varphi(y)$.
- Suppose that there exists w such that $\varphi(x) <_q w <_q \varphi(y)$.
- Since φ is a bijection, there should exist $u \in P$, a unique preimage of w: $\phi(u) = w$. Since φ is an order isomorphism, one has $x <_p u <_p y$, which contradicts the fact that $x \prec_p y$.
- Hence, $\varphi(x) \prec_q \varphi(y)$.

- Assume that $\varphi(x) \prec_q \varphi(y)$.
- Since φ is an order isomorphism, the inverse mapping φ^{-1} will also be an order isomorphism and by reasoning similar to that above we obtain $x \prec_p y$.

$\square$

Proof (continuation)

$2 \to 1$

- Assume that 2 holds and $x <_p y$.
- By the previous theorem one can find a sequence of covering elements $x = x_0 \prec_p x_1 \prec_p \cdots \prec_p x_n$.
- By condition 2, $\varphi(x) = \varphi(x_0) \prec_q \varphi(x_1) \prec_q \cdots \prec_q \varphi(x_n)$, hence $\varphi(x) <_q \varphi(y)$.
- Analogously, due to bijectivity of φ and inverse mapping φ^{-1}, having 2 and $\varphi(x) <_q \varphi(y)$ we obtain $x <_p y$.

Diagram

(Hasse) Diagram of a partially ordered set $(P; \leq)$ is a plain geometrical object consisting of circles which centers correspond to elements of the poset and edges that stay for the pairs of the covering relation $(P; \prec)$, connect centers of circles and respect the following properties:

1. $a \prec b \implies$ center of the circle corresponding to element a has strictly smaller vertical coordinate than the center of the circle corresponding to element b.
2. An edge intersects only two circles that correspond to two vertices of the edge.

Remarks on diagrams

Remark 1. The possibility of matching condition 1 is guaranteed by the possibility of topological sorting of the poset.

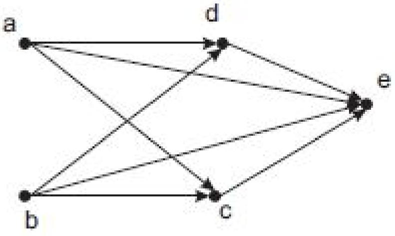
Remark 2. By the previous theorem, partial orders are isomorphic iff they can be represented by same diagram.

Remark 3. By the definition, diagrams cannot contain triangles and horizontal edges.

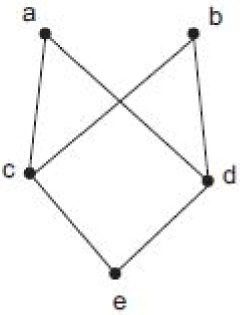
Example

Binary matrix of a partial order.

	a	b	c	d	e
a	1	0	1	1	1
b	0	1	1	1	1
c	0	0	1	0	1
d	0	0	0	1	1
e	0	0	0	0	1



acyclic graph



order diagram

Duality principle for partial orders

Relation $\geq$ inverse to partial order $\leq$ on set M, is called **dual** partial order $(M; \leq)^d$. Let A be a statement about partial order $(M; \leq)$. Statement A^d **dual** to statement A is obtained by replacing $\leq$ by $\geq$.

Duality principle. Statement A is valid for poset $(M, \leq)$ if dual statement A^d is valid for dual poset $(M, \leq)^d$.

Duality principle is used for simplification of definitions and proofs.

Important elements and subsets of posets

An element $p \in P$ of partial order $(P, \leq)$ is called **maximal** if there is no element of P strictly larger than p: $\forall x \in P \; x \not> p$. **Minimal** element is defined dually.

In general, maximal (minimal) elements can be not unique.

Important elements and subsets of posets

The **largest** element of $(P, \leq)$ is an element $1 \in P$ that is larger than all other elements: $\forall x \in P \; x \leq 1$. The **least** element is defined dually, as the one that is smaller than all other elements.

By definition, the largest (smallest) element, if it exists, is the unique maximal (minimal) element.

Important elements and subsets of posets

Let $(P, \leq)$ be a poset. Subset $J \subseteq P$ is called **order ideal** if $x \in J$, $y \leq x \Rightarrow y \in J$. Dually, subset $F \subseteq P$ is called **order filter** if $x \in F$, $y \geq x \Rightarrow y \in F$. In finite case an order filter can be represented by the set of its minimal elements, which are pairwise incomparable (make an **antichain**).

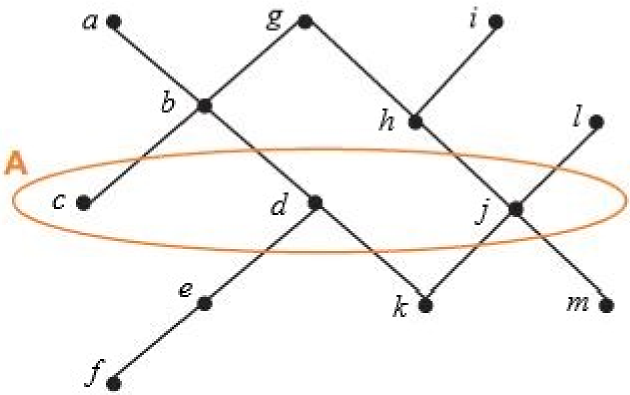
Dually, order ideals can be represented by sets of maximal elements (make an **antichain**).

Important elements and subsets of posets

For a partial order $(P, \leq)$ and $Q \subseteq P$

- $\downarrow Q = \{x \in P \mid (\exists q \in Q)\ x \leq q\}$.
- $\uparrow Q = \{x \in P \mid (\exists q \in Q)\ x \geq q\}$.
    - If Q is **an antichain**, then $\downarrow Q$ is an order ideal, for which Q is **the set of maximal elements** and $\uparrow Q$ is **an order filter**, for which Q is **the set of minimal elements**.
    - If $Q = \{x\}$ for $x \in P$, then the order filter is called **principle filter**.

Example: Order filter and order ideal



$A\uparrow$ is order filter

$A\downarrow$ is order ideal

Order filter and order ideal

The set of order ideals of poset $(P, \leq)$ is usually denoted by $O(P)$. This set is partially ordered w.r.t. containment.

Theorem

Let $(P, \leq)$ be a poset, then the following three statements are equivalent:

- $x \leq y$
- $\downarrow \{x\} \subseteq \downarrow \{y\}$
- $(\forall Q \in O(P))\ y \in Q \Leftrightarrow x \in Q$

Partial order and linear order

Partial order has "more complicated" structure than linear order, since one needs several linear orders to represent partial order adequately.

Theorem (Dushnik-Miller theorem)

Any intersection of linear orders on a set is a partial order, and any partial order is an intersection of linear orders that are its linear extensions.

Dushnik-Miller theorem

Proof. 1.

- Let $(A, \leq_i)$, $i \in I$ be linear orders on set A and set of indices I.
- Consider their intersection, relation R. All linear orders are reflexive: $(a, a) \in (A, \leq_i)$, hence $(a, a) \in R$.
- If xRy and yRz, then $x \leq_i y$, $y \leq_i z$ for any partial order $\leq_i$, therefore, $x \leq_i z$ in all linear orders and xRz in their intersection, i.e., R is transitive.
- Similar for antisymmetricity: assume that xRy and yRx, then for all linear orders one has $x \leq_i y$, $y \leq_i x$, which, by antisymmetricity of linear orders contradicts the fact that x and y are different. Hence, R is antisymmetric.

Dushnik-Miller theorem

Proof. 2.

- Let $(A, \leq)$ be a partial order, then it is contained (as a set of pairs making relation $\leq$) in all its linear extensions (topological sortings) $(A, \leq_i)$.
- By 1, the intersection of all $(A, \leq_i)$ is a partial order. This intersection contains partial order $(A, \leq)$, since it is contained in all linear orders $(A, \leq_i)$.
- It remains to show that the intersection of $(A, \leq_i)$ does not contain elements not belonging to $(A, \leq)$.
- Assume the converse: $\exists\ x \neq y$: the pair (x, y) is not contained in $(A, \leq)$, but it is contained in the intersection of $(A, \leq_i)$, hence, in all linear orders $(A, \leq_i)$.
- The pair (y, x) cannot be contained in $(A, \leq)$, since by antisymmetry this would imply $x = y$.

Dushnik-Miller theorem

Proof. 3.

- It remains to consider the possibility that x and y are incomparable in $(A, \leq)$.
- It implies that there exists a linear extension of $(A, \leq)$ containing (y, x) and not containing (x, y).
- Hence, the intersection of all linear extensions of $(A, \leq)$ would not contain (x, y), which contradicts our assumption.

$\square$

Order dimension

Dushnik-Miller theorem allows one to introduce the following definition of order dimension. **Order dimension** of partial order $(P, \leq)$ is the least number of linear orders $(P, \leq_i)$, $i \in \{1, \dots, k\}$ on P so that their intersection gives $(P, \leq) : (P, \leq) = \bigcap_{i \in \{1, \dots, k\}} (P, \leq_i)$.

Order dimension

Another definition of order dimension is based on the product of linear orders. **Multiplicative dimension** of partial order $(P, \leq)$ is the least number k of linear orders $(P, \leq_i)$, $i \in \{1, \ldots, k\}$ on P such that $(P, \leq)$ is order embedded in the Cartesian product $\times_{i\in\{1,\ldots,k\}}(P, \leq_i)$.

Equivalence of dimension definitions

Proof

- If partial order $(P, \leq)$ can be represented as intersection of linear orders $\times_i(P, \leq_i)$, $i \in \{1, \ldots, k\}$, then $(P, \leq)$ can be order embedded in the product of these orders, hence the multiplicative dimension is not larger than the order dimension.
- In the other direction, let the partial order $(P, \leq)$ be order embeddable in the Cartesian product $\times_i(P, \leq_i)$, $i \in \{1, \ldots, k\}$ of linear orders.
- Lexicographically (hence, linearly) ordered set $(P, \leq_i^*)$ for every i in the following way:

$$w \leq_i^* v \iff w_i >_i v_i \text{ or } w_i = v_i \text{ and } w_j > v_j \text{ for the first } j : w_j \neq v_j.$$

Equivalence of dimension definitions

Proof (Continuation)

- On the one hand, $\leq_i^*$ is a linear order. On the other hand, due to embedding of the initial poset $(P, \leq)$ in the product of linearly ordered sets $\times_i(P, \leq_i)$ the condition $w_j > v_j$ for the first $j : w_j \neq v_j$ denotes that $w_j \geq v_j$ for all other j.
- It can be easily tested that the initial poset $(P, \leq)$ can be embedded in the intersection of linear orders $(P, \leq_i^*)$, hence order dimension of $(P, \leq)$ is not larger than its multiplicative dimension.

$\square$

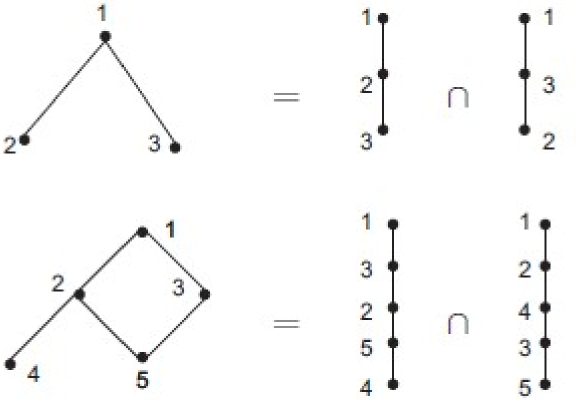
Application

Dushnik-Miller theorem and **the definition of the order dimension** can be applied to the analysis of preferences: for every finite poset of alternatives there exists a set of linearly ordered sets **("scales")** where all alternatives are comparable and the intersection of the scales gives the original order.

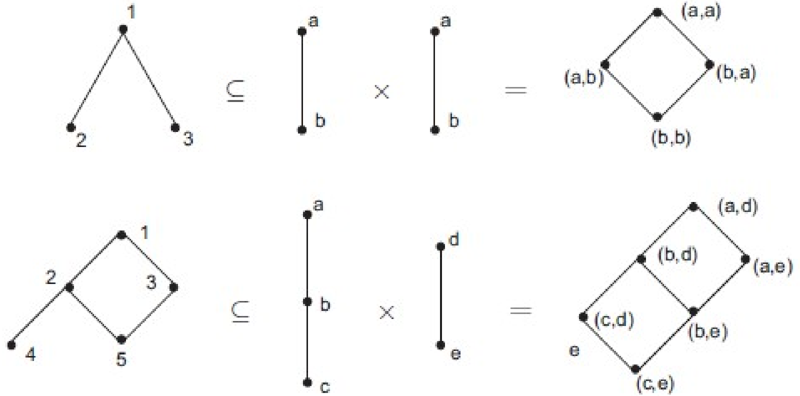
The scales can be considered as different aspects of alternatives, w.r.t. which all alternatives are pairwise comparable.

Example: in selecting a car, these scales can be price, year of production, producer, model, color, etc.

Determining order dimension of a poset



Determining multiplicative dimension of a poset



Theorem of Dilworth

Theorem of Dilworth gives a representation of a partial order through the union of orders.
The **width** of a partial order is the size of its maximal antichain.

Theorem

Let $(P, \leq)$ be a poset of width k. Then set P can be represented as a partition $P = P_1 \cup \cdots \cup P_k$, $i \neq j \Leftrightarrow P_i \cap P_j = \emptyset$, where each block P_i is linearly ordered w.r.t. $\leq$.

A proof of the theorem, which employs the theory of matchings, can be found e.g., in [Ore, 1962].

References

1. O.Ore, Theory of Graphs, AMS, 1962.
2. B.A. Davey and H.A. Priestley, Introduction to Lattices and Order, Cambridge University Press, 2nd Edition, 2002.
3. T.S. Blyth, M.F. Janowitz, Residuation Theory, Pergamon Press, 1972.