Chapter 15
Distances in Graph Theory

A *graph* is a pair $G = (V, E)$, where V is a set, called the set of *vertices* of the graph G, and E is a set of unordered pairs of vertices, called the *edges* of the graph G. A *directed graph* (or *digraph*) is a pair $D = (V, E)$, where V is a set, called the set of *vertices* of the digraph D, and E is a set of ordered pairs of vertices, called *arcs* of the digraph D.

A graph in which at most one edge may connect any two vertices, is called a *simple graph*. If multiple edges are allowed between vertices, the graph is called a *multigraph*. A graph, together with a function which assigns a positive weight to each edge, is called a *weighted graph* or *network*.

The graph is called *finite* (*infinite*) if the set V of its vertices is finite (infinite, respectively). The *order* and *size* of a finite graph (V, E) are $|V|$ and $|E|$, respectively.

A *subgraph* of a graph $G = (V, E)$ is a graph $G' = (V', E')$ with $V' \subset V$ and $E' \subset E$. If G' is a subgraph of G, then G is called a *supergraph* of G'. A subgraph (V', E') of (V, E) is its *induced subgraph* if $E' = \{e = uv \in E : u, v \in V'\}$.

A graph $G = (V, E)$ is called *connected* if, for any $u, v \in V$, there exists a $(u-v)$ *walk*, i.e., a sequence of edges $uw_1 = w_0w_1, w_1w_2, \dots, w_{n-1}w_n = w_{n-1}v$ from E. A $(u - v)$ *path* is a $(u - v)$ walk with distinct edges. A graph is called *m-connected* if there is no set of $m - 1$ edges whose removal disconnects the graph; a connected graph is 1-connected. A digraph $D = (V, E)$ is called *strongly connected* if, for any $u, v \in V$, the *directed* $(u - v)$ and $(v - u)$ paths both exist. A maximal connected subgraph of a graph G is called its *connected component*.

Vertices connected by an edge are called *adjacent*. The *degree* $deg(v)$ of a vertex $v \in V$ of a graph $G = (V, E)$ is the number of its vertices adjacent to v.

A *complete graph* is a graph in which each pair of vertices is connected by an edge. A *bipartite graph* is a graph in which the set V of vertices is decomposed into two disjoint subsets so that no two vertices within the same subset are adjacent. A *simple path* is a simple connected graph in which two vertices have degree 1, and

M.M. Deza, E. Deza, *Encyclopedia of Distances*,
DOI 10.1007/978-3-662-44342-2_15

other vertices (if they exist) have degree 2; the *length* of a path is the number of its edges.

A *cycle* is a *closed simple path*, i.e., a simple connected graph in which every vertex has degree 2. The *circumference* of a graph is the length of the longest cycle in it. A *tree* is a simple connected graph without cycles. A tree having a path from which every vertex has distance ≤ 1 or ≤ 2, is called a *caterpillar* or *lobster*, respectively.

Two graphs which contain the same number of vertices connected in the same way are called *isomorphic*. Formally, two graphs $G = (V(G), E(G))$ and $H = (V(H), E(H))$ are called *isomorphic* if there is a bijection $f : V(G) \to V(H)$ such that, for any $u, v \in V(G)$, $uv \in E(G)$ if and only if $f(u)f(v) \in E(H)$.

We will consider mainly simple finite graphs and digraphs; more exactly, the equivalence classes of such isomorphic graphs.

15.1 Distances on the Vertices of a Graph

- **Path metric**

  The **path metric** (or **graphic metric**, *shortest path metric*) d_{path} is a metric on the vertex-set V of a connected graph $G = (V, E)$ defined, for any $u, v \in V$, as the length of a shortest $(u - v)$ path in G, i.e., a *geodesic*. Examples follow.

  Given an integer $n \geq 1$, the **line metric on** $\{1, \dots, n\}$ in Chap. 1 is the path metric of the path $P_n = \{1, \dots, n\}$. The path metric of the *Cayley graph* Γ of a finitely generated group $(G, \cdot, e)$ is called a **word metric**.

  The **hypercube metric** is the path metric of a *hypercube graph* $H(m, 2)$ with the vertex-set $V = \{0, 1\}^m$, and whose edges are the pairs of vectors $x, y \in \{0, 1\}^m$ such that $|\{i \in \{1, \dots, n\} : x_i \neq y_i\}| = 1$; it is equal to $|\{i \in \{1, \dots, n\} : x_i = 1\} \triangle \{i \in \{1, \dots, n\} : y_i = 1\}|$. The graphic metric space associated with a hypercube graph coincides with a **Hamming cube**, i.e., the metric space $(\{0, 1\}^m, d_{l_1})$.

  The **belt distance** (Garber–Dolbilin, 2010) is the path metric of a *belt graph* $B(P)$ of a polytope P with centrally symmetric facets. The vertices of $B(P)$ are the facets of P and two vertices are connected by an edge if the corresponding facets lie in the same *belt* (the set of all facets of P parallel to a given face of codimension 2).

  The reciprocal path metric is called **geodesic similarity**.

- **Weighted path metric**

  The **weighted path metric** d_{wpath} is a metric on the vertex-set V of a connected weighted graph $G = (V, E)$ with positive edge-weights $(w(e))_{e \in E}$ defined by

  $$\min_P \sum_{e \in P} w(e),$$

  where the minimum is taken over all $(u - v)$ paths P in G.

Sometimes, $\frac{1}{w(e)}$ is called the *length* of the edge e. In the theory of electrical networks, the edge-length $\frac{1}{w(e)}$ is identified with the *resistance* of the edge e. The **inverse weighted path metric** is $\min_P \sum_{e \in P} \frac{1}{w(e)}$.

- **Metric graph**

  A **metric** (or *metrized*) **graph** is a connected graph $G = (V, E)$, where edges e are identified with line segments $[0, l(e)]$ of length $l(e)$. Let x_e be the coordinate on the segment $[0, l(e)]$ with vertices corresponding to $x_e = 0, l(e)$; the ends of distinct segments are identified if they correspond to the same vertex of G. A *function* f on G is the $|E|$-tuple of functions $f_e(x_e)$ on the segments.

  A metric graph can be seen as an infinite metric space (X, d), where X is the set of all points on above segments, and the distance between two points is the length of the shortest, along the line segments traversed, path connecting them. Also, it can be seen as one-dimensional Riemannian manifold with singularities. There is a bijection between the metric graphs, the equivalence classes of finite connected edge-weighted graphs and the resistive electrical networks: if an edge e of a metric graph has length $l(e)$, then $\frac{1}{l(e)}$ is the weight of e in the corresponding edge-weighted graph and $l(e)$ is the resistance along e in the corresponding resistive electric circuit. Cf. the **resistance metric**.

  A **quantum graph** is a metric graph equipped with a self-adjoint differential operator (such as a *Laplacian*) acting on functions on the graph. The *Hilbert space* of the graph is $\oplus_{e \in E} L^2([0, w(e)])$, where the inner product of functions is $\langle f, g \rangle = \sum_{e \in E} \int_0^{w(e)} f_e^*(x_e) g_e(x_e) dx_e$.

- **Spin network**

  A **spin network** is (Penrose, 1971) a connected graph (V, E) with edge-weights $(w(e))_{e \in E}$ (*spins*), $w(e) \in \mathbb{N}$, such that for any distinct edges e_1, e_2, e_3 with a common vertex, it holds **spin triangle inequality** $|w(e_1) - w(e_2)| \le w(e_3) \le w(e_1) + w(e_2)$ and *fermion conservation*: $w(e_1) + w(e_2) + w(e_3)$ is an even number.

  The **quantum space-time** (Chap. 24) in *Loop Quantum Gravity* is a network of loops at Planck scale. Loops are represented by adapted spin networks: directed graphs whose arcs are labeled by irreducible representations of a compact Lie group and vertices are labeled by *interwinning operators* from the tensor product of labels on incoming arcs to the tensor product of labels on outgoing arcs. Such networks represent "quantum states" of the gravitational field on a 3D hypersurface.

- **Detour distance**

  Given a connected graph $G = (V, E)$, the **detour distance** is (Chartrand and Zhang, 2004) a metric on the vertex-set V defined, for $u \neq v$, as the length of the longest $(u - v)$ path in G. So, this distance is 1 or $|V| - 1$ if and only if uv is a *bridge* of G or, respectively, G contains a Hamiltonian $(u - v)$ path.

  The **monophonic distance** is (Santhakumaran and Titus, 2011) a distance (in general, not a metric) on the V defined, for $u \neq v$, as the length of a longest *monophonic* (or *minimal*), i.e., containing no chords, $(u - v)$ path in G.

The *height* of a DAG (acyclic digraph) is the number of vertices in a longest directed path.

- **Cutpoint additive metric**
  Given a graph $G = (V, E)$, Klein–Zhu, 1998, call a metric d on V **graph-geodetic metric** if, for $u, w, v \in V$, the **triangle equality** $d(u, w) + d(w, v) = d(u, v)$ holds if w is a (u, v)-*gatekeeper*, i.e., w lies on any path connecting u and v. Cf. **metric interval** in Chap. 1. Any gatekeeper is a *cutpoint*, i.e., removing it disconnects G and a *pivotal point*, i.e., it lies on any shortest path between u and v.
  Chebotarev, 2010, call a metric d on the vertices of a multigraph without loops **cutpoint additive** if $d(u, w)+d(w, v) = d(u, v)$ holds if and only if w lies on any path connecting u and v. The **resistance metric** is cutpoint additive (Gvishiani and Gurvich, 1992), while the **path metric** is graph-geodetic only (in the weaker Klein–Zhu sense). See also **Chebotarev–Shamis metric**.
- **Graph boundary**
  Given a connected graph $G = (V, E)$, a vertex $v \in V$ is (Chartrand et al., 2003) a *boundary vertex* if there exists a *witness*, i.e., a vertex $u \in V$ such that $d(u, v) \geq d(u, w)$ for all neighbors w of v. So, the end-vertices of a longest path are boundary vertices. The **boundary** of G is the set of all boundary vertices.
  The *boundary of a subset* $M \subset V$ is the set $\partial M \subset E$ of edges having precisely one endpoint in M. The **isoperimetric number** of G is (Buser, 1978) $\inf \frac{\partial M}{|M|}$, where the infimum is taken over all $M \subset V$ with $2|M| \leq |V|$.
- **Graph diameter**
  Given a connected graph $G = (V, E)$, its **graph diameter** is the largest value of the **path metric** between vertices of G.
  A connected graph is *vertex-critical* (*edge-critical*) if deleting any vertex (edge) increases its diameter. A graph G of diameter k is *goal-minimal* if for every edge uv, the inequality $d_{G-uv}(x, y) > k$ holds if and only if $\{u, v\} = \{x, y\}$.
  If G is m-connected and a is an integer, $0 \leq a < m$, then the a**-fault diameter** of G is the maximal diameter of a subgraph of G induced by $|V| - a$ of its vertices. For $0 < a \leq m$, the a**-wide distance** $d_a(u, v)$ between vertices u and v is the minimum integer l, for which there are at least a internally disjoint $(u - v)$ paths of length at most l in G: cf. **Hsu–Lyuu–Flandrin–Li distance**. The a*-wide diameter* of G is $\max_{u,v \in V} d_a(u, v)$; it is at least the $(a - 1)$-fault diameter of G.
  Given a *strong orientation* O of a connected graph $G = (V, E)$, i.e., a strongly connected digraph $D = (V, E')$ with arcs $e' \in E'$ obtained from edges $e \in E$ by orientation O, the *diameter* of D is the maximal length of shortest directed $(u-v)$ path in it. The **oriented diameter** of a graph G is the smallest diameter among strong orientations of G. If it is equal to the diameter of G, then any orientation realizing this equality is called *tight*. For example, a *hypercube graph* $H(m, 2)$ admits a tight orientation if $m \geq 4$ (McCanna, 1988).
- **Path quasi-metric in digraphs**
  The **path quasi-metric in digraphs** d_{dpath} is a quasi-metric on the vertex-set V of a strongly connected digraph $D = (V, E)$ defined, for any $u, v \in V$, as the length of a shortest directed $(u - v)$ path in D.

The **circular metric in digraphs** is a metric on the vertex-set V of a strongly connected digraph $D = (V, E)$, defined by $d_{dpath}(u, v) + d_{dpath}(v, u)$.

- **Strong distance in digraphs**
The **strong distance in digraphs** is a metric between vertices v and v of a strongly connected digraph $D = (V, E)$ defined (Chartrand–Erwin–Raines–Zhang, 1999) as the minimum *size* (the number of edges) of a strongly connected subdigraph of D containing v and v. Cf. **Steiner distance of a set**.
- **Υ-metric**
Given a class Υ of connected graphs, the metric d of a metric space (X, d) is called a **Υ-metric** if (X, d) is isometric to a subspace of a metric space (V, d_{wpath}), where $G = (V, E) \in \Upsilon$, and d_{wpath} is the **weighted path metric** on V with positive edge-weight function w; cf. **tree-like metric**.
- **Tree-like metric**
A **tree-like metric** (or **weighted tree metric**) d on a set X is a **Υ-metric** for the class Υ of all trees, i.e., the metric space (X, d) is isometric to a subspace of a metric space (V, d_{wpath}), where $T = (V, E)$ is a tree, and d_{wpath} is the **weighted path metric** on the vertex-set V of T with a positive weight function w. A metric is a tree-like metric if and only if it satisfies the **four-point inequality**.
A metric d on a set X is called a **relaxed tree-like metric** if the set X can be embedded in some (not necessary positively) edge-weighted tree such that, for any $x, y \in X$, $d(x, y)$ is equal to the sum of all edge weights along the (unique) path between corresponding vertices x and y in the tree. A metric is a relaxed tree-like metric if and only if it is a **relaxed four-point inequality metric**.
- **Katz similarity**
Given a connected graph $G = (V, E)$ with positive edge-weight function $w = (w(e))_{e \in E}$, let $V = \{v_1, \dots, v_n\}$. Denote by A the $(n \times n)$-matrix $((a_{ij}))$, where $a_{ij} = a_{ji} = w(ij)$ if ij is an edge, and $a_{ij} = 0$, otherwise. Let I be the identity $(n \times n)$-matrix, and let t, $0 < t < \frac{1}{\lambda}$, be a parameter, where $\lambda = \max_i |\lambda_i|$ is the *spectral radius* of A and λ_i are the eigenvalues of A. Define the $(n \times n)$-matrix

$$K = ((k_{ij})) = \sum_{i=1}^{\infty} t^i A^i = (I - tA)^{-1} - I.$$

The number k_{ij} is called the **Katz similarity** between v_i and v_j. Katz, 1953, proposed it for evaluating social status.
Chebotarev, 2011, defined, for a similar $(n \times n)$-matrix $((c_{ij})) = \sum_{i=0}^{\infty} t^i A^i = (I - tA)^{-1}$ and connected edge-weighted multigraphs allowing loops, the **walk distance** between vertices v_i and v_j as any positive multiple of $d_t(i, j) = -\ln \frac{c_{ij}}{\sqrt{c_{ii} c_{jj}}}$ (cf. the **Nei standard genetic distance** in Chap. 23). He proved that d_t is a **cutpoint additive metric** and the **path metric** in G coincides with the *short walk distance* $\lim_{t \to 0+} \frac{d_t}{-\ln t}$ in G, while the **resistance metric** in G coincides with the *long walk distance* $\lim_{t \to \frac{1}{\lambda}-} \frac{2d_t}{n(t^{-1} - \lambda)}$ in the graph G' obtained from G by attaching weighted loops that provide G' with uniform weighted degrees.

If G is a simple unweighted graph, then A is its adjacency matrix. Let J be the $(n \times n)$-matrix of all ones and let $\mu = \min_i \lambda_i$. Let $N = ((n_{ij})) = \mu(I-J)-A$. Neumaier, 1980, remarked that $((\sqrt{n_{ij}}))$ is a semimetric on the vertices of G.

- **Resistance metric**

  Given a connected graph $G = (V, E)$ with positive edge-weight function $w = (w(e))_{e \in E}$, let us interpret the edge-weights as electrical conductances and their inverses as resistances. For any two different vertices u and v, suppose that a battery is connected across them, so that one unit of a current flows in at u and out in v. The voltage (potential) difference, required for this, is, by Ohm's law, the effective resistance between u and v in an electrical network; it is called the **resistance** (or *electric*) **metric** $\Omega(u, v)$ between them (Sharpe, 1967, Gvishiani–Gurvich, 1987, and Klein–Randic, 1993 [KlRa93]). So, if a potential of one volt is applied across vertices u and v, a current of $\frac{1}{\Omega(u,v)}$ will flow. The number $\frac{1}{\Omega(u,v)}$ is a measure of the *connectivity* between u and v.

  Let $r(u, v) = \frac{1}{w(e)}$ if uv is an edge, and $r(u, v) = 0$, otherwise. Formally,

  $$\Omega(u, v) = (\sum_{w \in V} f(w) r(w, v))^{-1},$$

  where $f : V \to [0, 1]$ is the unique function with $f(u) = 1$, $f(v) = 0$ and $\sum_{z \in V}(f(w) - f(z)) r(w, z) = 0$ for any $w \neq u, v$.

  The resistance metric is a weighted average of the lengths of all $(u - v)$ paths. It is applied when the number of $(u - v)$ paths, for any $u, v \in V$, matters.

  A probabilistic interpretation (Gobel–Jagers, 1974) is: $\Omega(u, v) = (deg(u)Pr(u \to v))^{-1}$, where $deg(u)$ is the degree of the vertex u, and $Pr(u \to v)$ is the probability for a random walk leaving u to arrive at v before returning to u. The expected commuting time between u and v is $2\sum_{e \in E} w(e)\Omega(u, v)$.

  Then $\Omega(u, v) \leq \min_P \sum_{e \in P} \frac{1}{w(e)}$, where P is any $(u - v)$ path (cf. **inverse weighted path metric**), with equality if and only if such a path P is unique. So, if $w(e) = 1$ for all edges, the equality means that G is a **geodetic graph**, and hence the path and resistance metrics coincide. Also, it holds that $\Omega(u, v) = \frac{|\{t : uv \in t \in T\}|}{|T|}$ if uv is an edge, and $\Omega(u, v) = \frac{|T' - T|}{|T|}$, otherwise, where T, T' are the sets of spanning trees for $G = (V, E)$ and $G' = (V, E \cup \{uv\})$.

  If $w(e) = 1$ for all edges, then $\Omega(u, v) = (g_{uu} + g_{vv}) - (g_{uv} + g_{vu})$, where $((g_{ij}))$ is the Moore–Penrose *generalized inverse* of the *Laplacian matrix* $((l_{ij}))$ of the graph G: here l_{ii} is the degree of vertex i, while, for $i \neq j$, $l_{ij} = 1$ if the vertices i and j are adjacent, and $l_{ij} = 0$, otherwise. A symmetric (for an undirected graph) and positive-semidefinite matrix $((g_{ij}))$ admits a representation KK^T. So, $\Omega(u, v)$ is the squared Euclidean distance between the u-th and v-th rows of K. The distance $\sqrt{\Omega(u, v)}$ is a **Mahalanobis distance** (cf. Chap. 17) with a weighting matrix $((g_{ij}))$. So, $\Omega_{u,v} = a_{uv}|((g_{ij}))|a_{uv}$, where a_{uv} are the vectors of zeros except for $+1$ and -1 in the u-th and v-th positions. This distance is called a *diffusion metric* in [CLMNWZ05] because it depends on a random walk.

The number $\frac{1}{2}\sum_{u,v\in V}\Omega(u,v)$ is called the *total resistance* (or *Kirchhoff index*) of G.

- **Hitting time quasi-metric**
  Let $G = (V, E)$ be a connected graph. Consider random walks on G, where at each step the walk moves to a vertex randomly with uniform probability from the neighbors of the current vertex. The **hitting** (or *first-passage*) **time quasi-metric** $H(u, v)$ from $u \in V$ to $v \in V$ is the expected number of steps (edges) for a random walk on G beginning at u to reach v for the first time; it is 0 for $u = v$. This quasi-metric is a **weightable quasi-semimetric** (cf. Chap. 1).
  The **commuting time metric** is $C(u, v) = H(u, v) + H(v, u)$.
  Then $C(u, v) = 2|E|\Omega(u, v)$, where $\Omega(u, v)$ is the **resistance metric** (or *effective resistance*), i.e., 0 if $u = v$ and, otherwise, $\frac{1}{\Omega(u,v)}$ is the current flowing into v, when grounding v and applying a 1 volt potential to u (each edge is seen as a resistor of 1 ohm). Also, $\Omega(u, v) = \sup_{f:V\to\mathbb{R},\, D(f)>0}\frac{(f(u)-f(v))^2}{DE(f)}$, where $DE(f)$ is the *Dirichlet energy* of f, i.e., $\sum_{st\in E}(f(s) - f(t))^2$.
  The above setting can be generalized to weighted digraphs $D = (V, E)$ with arc-weights c_{ij} for $ij \in E$ and the *cost* of a directed $(u - v)$ path being the sum of the weights of its arcs. Consider the random walk on D, where at each step the walk moves by arc ij with *reference probability* p_{ij} proportional to $\frac{1}{c_{ij}}$; set $p_{ij} = 0$ if $ij \notin E$. Saerens et al., 2008, defined the *randomized et al.* shortest path quasi-distance $d(u, v)$ on vertices of D as the minimum expected cost of a directed $(u - v)$ path in the probability distribution minimizing the expected cost among all distributions having a fixed **Kullback–Leibler distance** (cf. Chap. 14) with reference probability distribution. In fact, their biased random walk model depends on a parameter $\theta \geq 0$. For $\theta = 0$ and large θ, the distance $d(u, v) + d(v, u)$ become a metric; it is proportional to the commuting time and the usual path metric, respectively.
- **Chebotarev–Shamis metric**
  Given $\alpha > 0$ and a connected weighted *multigraph* $G = (V, E; w)$ with positive edge-weight function $w = (w(e))_{e\in E}$, denote by $L = ((l_{ij}))$ the *Laplacian* (or *Kirchhoff*) matrix of G, i.e., $l_{ij} = -w(ij)$ for $i \neq j$ and $l_{ii} = \sum_{j\neq i} w(ij)$. The **Chebotarev–Shamis metric** $d_\alpha(u, v)$ (Chebotarev and Shamis, 2000, called $\frac{1}{2}d_\alpha(u, v)$ α**-forest metric**) between vertices u and v is defined by

$$2q_{uv} - q_{uu} - q_{vv}$$

  for the **protometric** $((g_{ij})) = -(I + \alpha L)^{-1}$, where I is the identity matrix.
  Chebotarev and Shamis showed that their metric of $G = (V, E; w)$ is the **resistance metric** of another weighted multigraph, $G' = (V', E'; w')$, where $V' = V \cup \{0\}$, $E' = E \cup \{u0 : u \in V\}$, while $w'(e) = \alpha w(e)$ for all $e \in E$ and $w'(u0) = 1$ for all $u \in V$. In fact, there is a bijection between the forests of G and trees of G'. This metric becomes the resistance metric of $G = (V, E; w)$ as $\alpha \to \infty$.
  Their **forest metric** (1997) is the case $\alpha = 1$ of the α-forest metric.

Chebotarev, 2010, remarked that $2\ln q_{uv} - \ln q_{uu} - \ln q_{vv}$ is a **cutpoint additive metric** $d''_\alpha(u,v)$, i.e., $d''_\alpha(u,w) + d''_\alpha(w,v) = d''_\alpha(u,v)$ holds if and only if w lies on any path connecting u and v. The metric d''_α is the **path metric** if $\alpha \to 0^+$ and the **resistance metric** if $\alpha \to \infty$.

- **Truncated metric**
  The **truncated metric** is a metric on the vertex-set of a graph, which is equal to 1 for any two adjacent vertices, and is equal to 2 for any nonadjacent different vertices. It is the 2**-truncated metric** for the path metric of the graph. It is the $(1,2) - B$**-metric** if the degree of any vertex is at most B.
- **Hsu–Lyuu–Flandrin–Li distance**
  Given an *m-connected* graph $G = (V, E)$ and two vertices $u, v \in V$, a *container* $C(u,v)$ *of width* m is a set of m $(u - v)$ paths with any two of them intersecting only in u and v. The *length of a container* is the length of the longest path in it.
  The **Hsu–Lyuu–Flandrin–Li distance** between vertices u and v (Hsu–Lyuu, 1991, and Flandrin–Li, 1994) is the minimum of container lengths taken over all containers $C(u,v)$ of width m. This generalization of the path metric is used in parallel architectures for interconnection networks.
- **Multiply-sure distance**
  The **multiply-sure distance** is a distance on the vertex-set V of an *m-connected* weighted graph $G = (V, E)$, defined, for any $u, v \in V$, as the minimum weighted sum of lengths of m disjoint $(u-v)$ paths. This generalization of the path metric helps when several disjoint paths between two points are needed, for example, in communication networks, where $m - 1$ of $(u - v)$ paths are used to code the message sent by the remaining $(u - v)$ path (see [McCa97]).
- **Cut semimetric**
  A *cut* is a *partition* of a set into two parts. Given a subset S of $V_n = \{1, \dots, n\}$, we obtain the partition $\{S, V_n \backslash S\}$ of V_n. The **cut semimetric** (or **split semimetric**) δ_S defined by this partition, is a semimetric on V_n defined by

$$\delta_S(i,j) = \begin{cases} 1, & \text{if } i \neq j, |S \cap \{i,j\}| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

  Usually, it is considered as a vector in $\mathbb{R}^{|E_n|}$, $E(n) = \{\{i,j\} : 1 \le i < j \le n\}$.
  A *circular cut* of V_n is defined by a subset $S_{[k+1,l]} = \{k+1, \dots, l\} (\text{mod}\, n) \subset V_n$: if we consider the points $\{1, \dots, n\}$ as being ordered along a circle in that circular order, then $S_{[k+1,l]}$ is the set of its consecutive vertices from $k + 1$ to l. For a circular cut, the corresponding cut semimetric is called a **circular cut semimetric**.
  An **even cut semimetric** (**odd cut semimetric**) is δ_S on V_n with even (odd, respectively) $|S|$. A k**-uniform cut semimetric** is δ_S on V_n with $|S| \in \{k, n - k\}$. An **equicut semimetric** (**inequicut semimetric**) is δ_S on V_n with $|S| \in \{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil\}$ ($|S| \notin \{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil\}$, respectively); see, for example, [DeLa97].

- **Decomposable semimetric**
  A **decomposable semimetric** is a semimetric on $V_n = \{1, \dots, n\}$ which can be represented as a nonnegative linear combination of **cut semimetrics**. The set of all decomposable semimetrics on V_n is a *convex cone*, called the *cut cone* CUT_n. A semimetric on V_n is decomposable if and only if it is a **finite l_1-semimetric**.
  A **circular decomposable semimetric** is a semimetric on $V_n = \{1, \dots, n\}$ which can be represented as a nonnegative linear combination of **circular cut semimetrics**. A semimetric on V_n is circular decomposable if and only if it is a **Kalmanson semimetric** with respect to the same ordering (see [ChFi98]).
- **Finite l_p-semimetric**
  A **finite l_p-semimetric** d is a semimetric on $V_n = \{1, \dots, n\}$ such that (V_n, d) is a semimetric subspace of the l_p^m*-space* $(\mathbb{R}^m, d_{l_p})$ for some $m \in \mathbb{N}$.
  If, instead of V_n, is taken $X = \{0, 1\}^n$, the metric space (X, d) is called the l_p^n*-cube*. The l_1^n-cube is called a **Hamming cube**; cf. Chap. 4. It is the graphic metric space associated with a hypercube graph $H(n, 2)$, and any subspace of it is called a **partial cube**.
- **Kalmanson semimetric**
  A **Kalmanson semimetric** d *with respect to the ordering* $1, \dots, n$ is a semimetric on $V_n = \{1, \dots, n\}$ which satisfies the condition

$$\max\{d(i, j) + d(r, s), d(i, s) + d(j, r)\} \le d(i, r) + d(j, s)$$

  for all $1 \le i \le j \le r \le s \le n$.
  Equivalently, if the points $\{1, \dots, n\}$ are ordered along a circle C_n in that circular order, then the distance d on V_n is a Kalmanson semimetric if the inequality

$$d(i, r) + d(j, s) \le d(i, j) + d(r, s)$$

  holds for $i, j, r, s \in V_n$ whenever the segments $[i, j]$, $[r, s]$ are crossing chords of C_n.
  A **tree-like metric** is a Kalmanson metric for some ordering of the vertices of the tree. The Euclidean metric, restricted to the points that form a convex polygon in the plane, is a Kalmanson metric.
- **Multicut semimetric**
  Let $\{S_1, \dots, S_q\}$, $q \ge 2$, be a *partition* of the set $V_n = \{1, \dots, n\}$, i.e., a collection $S_1, \dots, S_q$ of pairwise disjoint subsets of V_n such that $S_1 \cup \dots \cup S_q = V_n$.
  The **multicut semimetric** $\delta_{S_1,\dots,S_q}$ is a semimetric on V_n defined by

$$\delta_{S_1,\dots,S_q}(i, j) = \begin{cases} 0, & \text{if } i, j \in S_h \text{ for some } h, 1 \le h \le q, \\ 1, & \text{otherwise.} \end{cases}$$

- **Oriented cut quasi-semimetric**
  Given a subset S of $V_n = \{1, \dots, n\}$, the **oriented cut quasi-semimetric** δ'_S is a quasi-semimetric on V_n defined by

$$\delta'_S(i,j) = \begin{cases} 1, & \text{if } i \in S, j \notin S, \\ 0, & \text{otherwise.} \end{cases}$$

  Usually, it is considered as the vector of $\mathbb{R}^{|I_n|}$, $I(n) = \{(i,j) : 1 \le i \ne j \le n\}$. The **cut semimetric** δ_S is $\delta'_S + \delta'_{V_n \setminus S}$.
- **Oriented multicut quasi-semimetric**
  Given a *partition* $\{S_1, \dots, S_q\}$, $q \ge 2$, of V_n, the **oriented multicut quasi-semimetric** $\delta'_{S_1,\dots,S_q}$ is a quasi-semimetric on V_n defined by

$$\delta'_{S_1,\dots,S_n}(i,j) = \begin{cases} 1, & \text{if} \quad i \in S_h, j \in S_m, h < m, \\ 0, & \text{otherwise} \quad . \end{cases}$$

15.2 Distance-Defined Graphs

Below we first give some graphs defined in terms of distances between their vertices. Then some graphs associated with metric spaces are presented.

A graph (V, E) is, say, *distance-invariant* or *distance monotone* if its metric space (V, d_{path}) is **distance invariant** or **distance monotone**, respectively (cf. Chap. 1). The definitions of such graphs, being straightforward subcases of corresponding metric spaces, will be not given below.

- **k-Power of a graph**
  The **k-power** of a graph $G = (V, E)$ is the supergraph $G^k = (V, E')$ of G with edges between all pairs of vertices having path distance at most k.
- **Distance-residual subgraph**
  For a connected finite graph $G = (V, E)$ and a set $M \subset V$ of its vertices, a **distance-residual subgraph** is (Luksic and Pisanski, 2010) a subgraph induced on the set of vertices u of G at the maximal **point-set distance** $\min_{v \in M} d_{\text{path}}(u, v)$ from M. Such a subgraph is called *vertex-residual* if M consists of a vertex, and *edge-residual* if M consists of two adjacent vertices.
- **Isometric subgraph**
  A subgraph H of a graph $G = (V, E)$ is called an **isometric subgraph** if the path metric between any two points of H is the same as their path metric in G.
  A subgraph H is called a *convex subgraph* if it is isometric, and for any $u, v \in H$ every vertex on a shortest $(u - v)$ path belonging to H also belongs to H.
  A subset $M \subset V$ is called *gated* if for every $u \in V \setminus M$ there exists a unique vertex $g \in M$ (called a *gate*) lying on a shortest $(u - v)$ path for every $v \in M$. The subgraph induced by a gated set is a convex subgraph.

- **Retract subgraph**
  A subgraph H of G is called a **retract subgraph** if it is induced by an idempotent **metric mapping** of G into itself, i.e., $f^2 = f : V \to V$ with $d_{\text{path}}(f(u), f(v)) \leq d_{\text{path}}(u, v)$ for $u, v \in V$. Any retract subgraph is isometric.
- **Partial cube**
  A **partial cube** is an **isometric subgraph** of a **Hamming cube**, i.e., of a hypercube $H(m, 2)$. Similar topological notion was introduced by Acharya, 1983: any graph (V, E) admits a *set-indexing* $f : V \cup E \to 2^X$ with injective $f|_V, f|_R$ and $f(uv) = f(u)\Delta f(v)$ for any $(uv) \in E$. The *set-indexing number* is $\min |X|$.
- **Median graph**
  A connected graph $G = (V, E)$ is called a **median graph** if, for any three vertices $u, v, w \in V$, there exists a unique vertex that lies simultaneously on a shortest $(u - v)$, $(u - w)$ and $(w - v)$ paths, i.e., (V, d_{path}) is a **median metric space**.
  The median graphs are exactly **retract subgraphs** of hypercubes. Also, they are exactly **partial cubes** such that the vertex-set of any *convex subgraph* is *gated* (cf. **isometric subgraph**).
- **Geodetic graph**
  A graph is called **geodetic** if there exists at most one shortest path between any two of its vertices. A graph is called *strongly geodetic* if there exists at most one path of length less than or equal to the diameter between any two of its vertices.
  A *uniformly geodetic graph* is a connected graph such that the number of shortest paths between any two vertices u and v depends only on $d(u, v)$.
  A graph is a *forest* (disjoint union of trees) if and only if there exists at most one path between any two of its vertices.
  The *geodetic number* of a finite connected graph (V, E) [BuHa90] is $\min |M|$ over sets $M \subset V$ such that any $x \in V$ lies on a shortest $(u - v)$ path with $u, v \in M$.
- **k-geodetically connected graph**
  A k-connected graph is called (Entringer–Jackson–Slater, 1977) **k-geodetically connected** ($k - GC$) if the removal of less than k vertices (or, equivalently, edges) does not affect the **path metric** between any pair of the remaining vertices.
  $2 - GC$ graphs are called *self-repairing*. Cf. **Hsu–Lyuu–Flandrin–Li distance**.
- **Interval distance monotone graph**
  A connected graph $G = (V, E)$ is called **interval distance monotone** if any of its intervals $I_G(u, v)$ induces a *distance monotone graph*, i.e., its path metric is **distance monotone**, cf. Chap. 1.
  A graph is interval distance monotone if and only if (Zhang–Wang, 2007) each of its intervals is isomorphic to either a path, a cycle or a hypercube.
- **Distance-regular graph**
  A connected *regular* (i.e., every vertex has the same degree) graph $G = (V, E)$ of diameter T is called **distance-regular** (or *drg*) if, for every two its vertices u, v and any integers $0 \leq i, j \leq T$, the number $|\{w \in V : d_{\text{path}}(u, w) = i,$

$d_{\text{path}}(v, w) = j\}|$ depends only on i, j and $k = d_{\text{path}}(u, v)$, but not on the choice of u and v.

A special case of it is a **distance-transitive graph**, i.e., such that its group of automorphisms is transitive, for any $0 \leq i \leq T$, on the pairs of vertices (u, v) with $d_{\text{path}}(u, v) = i$. An analog of drg is an *edge-regular graph* (Fiol–Carriga, 2001).

Any drg is a **distance-balanced graph** (or *dbg*), i.e., $|W_{u,v}| = |W_{v,u}|$, where $W_{u,v} = \{x \in V : d(x, u) < d(x, v)\}$. Such graph is also called *self-median* since it is exactly one, **metric median** (cf. **eccentricity** in Chap. 1) of which is V. A gbg is called *nicely distance-balanced* if $|W_{u,v}|$ is the same for all edges uv.

Any drg is a **distance degree-regular graph** (i.e., $|\{x \in V : d(x, u) = i\}|$ depends only on i; such graph is also called *strongly distance-balanced*), and a **walk-regular graph** (i.e., the number of closed walks of length i starting at u depends only on i). van Dam–Omidi, 2013, call a graph *strongly walk-regular* if there is an $l \geq 2$ such that the number of walks of length l from u to v depends only on whether the $d(u, v)$ is 0, 1, or $\geq$2; for $l = 2$, it is a *strongly regular graph*, i.e., a drg of diameter 2. A *d-Deza graph* (Gu, 2013) is a regular graph (V, E) in which there are exactly d different values of $|\{w \in V : d(u, w) = d(v, w) = 1\}|$ for distinct $u, v \in V$.

A graph G is a **distance-regularized graph** if for each $u \in V$, if admits an *intersection array at vertex u*, i.e., the numbers $a_i(u) = |G_i(u) \cap G_1(v)|$, $b_i(u) = |G_{i+1}(u) \cap G_1(v)|$ and $c_i(u) = |G_{i-1}(v) \cap G_1(v)|$ depend only on the distance $d(u, v) = i$ and are independent of the choice of the vertex $v \in G_i(u)$. Here, for any i, $G_i(w)$ is the set of all vertices at the distance i from w. Godsil–Shawe-Taylor, 1987, defined such graph and proved that it is either drg or *distance-biregular* (a bipartite one with vertices in the same class having the same intersection array).

A drg is also called a **metric association scheme** or *P-polynomial association scheme*. A finite **polynomial metric space** (cf. Chap. 1) is a special case of it, also called a *(P and Q)-polynomial association scheme*.

- **Distance-regular digraph**

  A strongly connected digraph $D = (V, E)$ is called **distance-regular** (Damerell, 1981) if, for any its vertices u, v with $d_{\text{path}}(u, v) = k$ and for any integer $0 \leq i \leq k + 1$, the number of vertices w, such that $d_{\text{path}}(u, w) = i$ and $d_{\text{path}}(v, w) = 1$, depends only on k and i, but not on the choice of u and v. In order to find interesting classes of distance-regular digraphs with unbounded diameter, the above definition was weakened by two teams in different directions. Call $\overline{d(x, y)} = (d(x, y), d(y, x))$ the **two-way distance in digraph** D. A strongly connected digraph $D = (V, E)$ is called **weakly distance-regular** (Wang and Suzuku, 2003) if, for any its vertices u, v with $\overline{d(u, v)} = (k_1, k_2)$, the number of vertices w, such that $\overline{d(w, u)} = (i_1, i_2)$ and $\overline{d(w, v)} = (j_1, j_2)$, depends only on the values $k_1, k_2, i_1, i_2, j_1, j_2$. Comellas et al., 2004, defined a **weakly distance-regular digraph** as one in which, for any vertices u and v, the number of $u \to v$ walks of every given length only depends on the distance $d(u, v)$.

- **Metrically almost transitive graph**
  An *automorphism* of a graph $G = (V, E)$ is a map $g : V \to V$ such that u is adjacent to v if and only if $g(u)$ is adjacent to $g(v)$, for any $u, v \in V$. The set $Aut(G)$ of automorphisms of G is a group with respect to the composition of functions.
  A graph G is **metrically almost transitive** (Krön–Möller, 2008) if there is an integer r such that, for any vertex $u \in V$ it holds

$$\cup_{g \in Aut(G)}\{g(\overline{B}(u, r) = \{v \in V : d_{\text{path}}(u, v) \leq r\})\} = V.$$

- **Metric end**
  Given an infinite graph $G = (V, E)$, a *ray* is a sequence $(x_0, x_1, \dots)$ of distinct vertices such that x_i and x_{i+1} are adjacent for $i \geq 0$.
  Two rays R_1 and R_2 are equivalent whenever it is impossible to find a bounded set of vertices F such that any path from R_1 to R_2 contains an element of F.
  Metric ends are defined as equivalence classes of *metric rays* which are rays without infinite, bounded subsets.
- **Graph of polynomial growth**
  Let $G = (V, E)$ be a transitive locally finite graph. For a vertex $v \in V$, the *growth function* is defined by

$$f(n) = |\{u \in V : d(u, v) \leq n\}|,$$

  and it does not depend on v. Cf. **growth rate of metric space** in Chap. 1.
  The graph G is a **graph of polynomial growth** if there are some positive constants k, C such that $f(n) \leq Cn^k$ for all $n \geq 0$. It is a **graph of exponential growth** if there is a constant $C > 1$ such that $f(n) > C^n$ for all $n \geq 0$.
  A group with a finite symmetric set of generators has *polynomial growth rate* if the corresponding *Cayley graph* has polynomial growth. Here the metric ball consists of all elements of the group which can be expressed as products of at most n generators, i.e., it is a closed ball centered in the identity in the **word metric**, cf. Chap. 10.
- **Distance-polynomial graph**
  Given a connected graph $G = (V, E)$ of diameter T, for any $2 \leq i \leq T$ denote by G_i the graph (V, E') with $E' = \{e = uv \in E : d_{\text{path}}(u, v) = i\}$. The graph G is called a **distance-polynomial** if the adjacency matrix of any G_i, $2 \leq i \leq T$, is a polynomial in terms of the adjacency matrix of G.
  Any **distance-regular** graph is a distance-polynomial.
- **Distance-hereditary graph**
  A connected graph is called **distance-hereditary** (Howorka, 1977) if each of its connected induced subgraphs is isometric.
  A graph is distance-hereditary if each of its induced paths is isometric. A graph is distance-hereditary, bipartite distance-hereditary, **block graph**, tree if and only if its path metric is a **relaxed tree-like metric** for edge-weights being, respectively, nonzero half-integers, nonzero integers, positive half-integers, positive integers.

A graph is called a **parity graph** if, for any $u, v \in V$, the lengths of all induced $(u-v)$ paths have the same parity. A graph is a parity graph (moreover, distance-hereditary) if and only if every induced subgraph of odd (moreover, any) order of at least five has an even number of Hamiltonian cycles (McKee, 2008).

- **Distance magic graph**
  A graph $G = (V, E)$ is called a **distance magic graph** if it admits a *distance magic labeling*, i.e., a *magic constant* $k > 0$ and a bijection $f : V \to \{1, 2, \dots, |V|\}$ with $\sum_{uv \in E} f(v) = k$ for every $u \in V$. Introduced by Vilfred, 1994, these graphs generalize *magic squares* (such complete n-partite graphs with parts of size n).
  Among trees, cycles and K_n, only P_1, P_3, C_4 are distance magic. The *hypercube graph* $H(m, 2)$ is distance magic if $m = 2, 6$ but not if $m \equiv 0, 1, 3 \pmod 4$.
- **Block graph**
  A graph is called a **block graph** if each of its *blocks* (i.e., a maximal 2-connected induced subgraph) is a complete graph. Any tree is a block graph.
  A graph is a block graph if and only if its path metric is a **tree-like metric** or, equivalently, satisfies the **four-point inequality**.
- **Ptolemaic graph**
  A graph is called **Ptolemaic** if its path metric satisfies the **Ptolemaic inequality**

$$d(x, y)d(u, z) \le d(x, u)d(y, z) + d(x, z)d(y, u).$$

  A graph is Ptolemaic if and only if it is distance-hereditary and *chordal*, i.e., every cycle of length greater than 3 has a chord. So, any **block graph** is Ptolemaic.
- **k-cocomparability graph**
  A graph $G = (V, E)$ is called (Chang–Ho–Ko, 2003) **k-cocomparability graph** if its vertex-set admits a linear ordering $<$ such that for any three vertices $u < v < w$, $d(u, w) \le k$ implies $d(u, v) \le k$ or $d(v, w) \le k$.
- **Distance-perfect graph**
  Cvetković et al., 2007, observed that any graph of diameter T has at most $k + T^k$ vertices, where k is its **location number** (cf. Chap. 1), i.e., the minimal cardinality of a set of vertices, the path distances from which uniquely determines any vertex. They called a graph **distance-perfect** if it meets this upper bound and proved that such a graph has $T \neq 2$.
- **t-irredundant set**
  A set $S \subset V$ of vertices in a connected graph $G = (V, E)$ is called **t-irredundant** (Hattingh–Henning, 1994) if for any $u \in S$ there exists a vertex $v \in V$ such that, for the path metric d_{path} of G, it holds

$$d_{\text{path}}(v, x) \le t < d_{\text{path}}(v, V \backslash S) = \min_{u \notin S} d_{\text{path}}(v, u).$$

  The **t-irredundance number** ir_t of G is the smallest cardinality $|S|$ such that S is t-irredundant but $S \cup \{v\}$ is not, for every $v \in V \backslash S$.

The *t-domination number* γ_t and *t-independent number* α_t of G are, respectively, the cardinality of the smallest $(t+1)$-*covering* (by the open balls of the radius $r+1$) and largest $\lceil \frac{t}{2} \rceil$-*packing* of the metric space $(V, d_{\text{path}}(u, v))$; cf. the **radii of metric space** in Chap. 1. Then it holds that $\frac{\gamma_t+1}{2} \leq ir_t \leq \gamma_t \leq \alpha_t$.

Let B_S denote $\{v \in V : d(v, S) = 1\}$. Then $\max_{S \subset V} |B_S| = |V| - \gamma_1$ and $\max_{S \subset V}(|B_S| - |S|)$ are called the *enclaveless number* and the *differential* of G.

- **r-Locating-dominating set**

  Let $D = (V, E)$ be a digraph and $C \subset V$, and let $B_r^-(v)$ denote the set of all vertices x such that there exists a directed $(x - v)$ path with at most r arcs.

  If $B_r^-(v) \cap C$, $v \in V \setminus C$ (respectively, $v \in V$), are nonempty distinct sets, C is called (Slater, 1984) an **r-locating-dominating set** (respectively, an **r-identifying code**; cf. Chap. 16) of D. Such sets of smallest cardinality are called *optimal*.
- **Locating chromatic number**

  The **locating chromatic number** of a graph $G = (V, E)$ is the minimum number of color classes $C_1, \dots, C_t$ needed to color vertices of G so that any two adjacent vertices have distinct colors and each vertex $u \in V$ has distinct *color code* $(\min_{v \in C_1} d(u, v), \dots, \min_{v \in C_k} d(u, v))$.
- **k-Distant chromatic number**

  The **k-distant chromatic number** of a graph $G = (V, E)$ is the minimum number of colors needed to color vertices of G so that any two vertices at distance at most k have distinct colors, i.e., it is the chromatic number of the **k-power of** G.
- **Distance between edges**

  The **distance between edges** in a connected graph $G = (X, E)$ is the number of vertices in a shortest path between them. So, adjacent edges have distance 1.

  A **distance-k matching** of G is a set of edges no two of which are within distance k. For $k = 1$, it is the usual matching. For $k = 2$, it is also *induced* (or *strong*) matching. A distance-k matching of G is equivalent to an independent set in the **k-power** of the line graph of G. A **distance-k edge-coloring** of G is an edge-coloring such that each color class induces a distance-k matching.

  The **distance-k chromatic index** $\mu_k(G)$ is the least integer t such that there exists a distance-t edge-coloring of G. The **distance-k matching number** $\nu_k(G)$ is the largest integer t such that there exists a distance-t matching in G with t edges. It holds that $\mu_k(G)\nu_k(G) \geq |E|$.

  The **distance between faces** of a plane graph is the number of vertices in a shortest path between them. A **distance-k face-coloring** is a face-coloring such that any two faces at distance at most k have different colors. The **distance-k face chromatic index** is the least integer t such that such coloring exists.
- **Rainbow distance**

  In an edge-colored graph, the **rainbow distance** is (Chartrand and Zhang, 2005) the length of a shortest *rainbow* (i.e., containing no color twice) path.

  In a vertex-colored graph, the **colored distance** is (Dankelmann et al., 2001) the sum of distances between all unordered pairs of vertices having different colors.

- **D-distance graph**
  Given a set D of positive numbers containing 1 and a metric space (X, d), the **D-distance graph** is a graph $G = (V = X, E)$ with the edge-set $E = \{uv : d(u, v) \in D\}$ (cf. **D-chromatic number** in Chap. 1). If (X, d) is path metric of a graph H, then G is called the **distance power** H^D of H.
  Alon–Kupavsky, 2014, call G (in the case $(X, d) = \mathbb{E}^n$, $d = \{1\}$) the *faithful unit-distance graph*, using term *unit-distance graph* for $E \subseteq \{(u, v) : ||u-v||_2 = 1\}$.
  For a positive number t, the *signed distance graph* is (Fiedler, 1969) a signed graph with the vertex-set X in which vertices x, y are joined by a positive edge if $t > d(x, y)$, by a negative edge if $d(x, y) > t$, and not joined if $d(x, y) = t$.
  A D-distance graph is called a **distance graph** (or *unit-distance graph*) if $D = \{1\}$, an *ϵ-unit graph* if $D = [1 - \epsilon, 1 + \epsilon]$, a *unit-neighborhood graph* if $D = (0, 1]$, an *integral-distance graph* if $D = \mathbb{Z}_+$, a *rational-distance graph* if $D = \mathbb{Q}_+$, and a *prime-distance graph* if D is the set of prime numbers (with 1).
  Every finite graph can be represented by a D-distance graph in some $\mathbb{E}^n$. The minimum dimension of such a Euclidean space is called the *D-dimension* of G.
  A *matchstick graph* is a crossingless unit-distance graph in $\mathbb{E}^2$.
- **Distance-number of a graph**
  Given a graph $G = (V, E)$, its *degenerate drawing* is a mapping $f : V \to \mathbb{R}^2$ such that $|f(V)| = |V|$ and $f(uv)$ is an open straight-line segment joining the vertices $f(u)$ and $f(v)$ for any edge $uv \in E$; it is a *drawing* if, moreover, $f(w) \notin f(uv)$ for any $uv \in E$ and $w \in V$.
  The **distance-number** $dn(G)$ **of a graph** G is (Carmi et al., 2008) the minimum number of distinct edge-lengths in a drawing of G.
  The *degenerate distance-number* of G, denoted by $ddn(G)$, is the minimum number of distinct edge-lengths in a degenerated drawing of G. The first of the **Erdös-type distance problems** in Chap. 19 is equivalent to determining $ddn(K_n)$.
- **Dimension of a graph**
  The **dimension** $dim(G)$ **of a graph** G is (Erdös–Harary–Tutte, 1965) the minimum k such that G has a *unit-distance representation* in $\mathbb{R}^k$, i.e., every edge is of length 1. The vertices are mapped to distinct points of $\mathbb{R}^k$, but edges may cross.
  For example, $dim(G) = n - 1, 4, 2$ for $G = K_n, K_{m,n}, C_n$ $(m \geq n \geq 3)$.
- **Bar-and-joint framework**
  A n-dimensional **bar-and-joint framework** is a pair (G, f), where $G = (V, E)$ is a finite graph (no loops and multiple edges) and $f : V \to \mathbb{R}^n$ is a map with $f(u) \neq f(v)$ whenever $uv \in E$. The **framework** is a straight line realization of G in $\mathbb{R}^n$ in which the length of an edge $uv \in E$ is given by $||f(u) - f(v)||_2$.
  The vertices and edges are called *joints* and *bars*, respectively, in terms of Structural Engineering. A **tensegrity structure** (Fuller, 1948) is a mechanically stable bar framework in which bars are either *cables* (tension elements which cannot get further apart), or *struts* (compression elements which cannot get closer together).

A framework (G, f) is *globally rigid* if every framework (G, f'), satisfying $||f(u) - f(v)||_2 = ||f'(u) - f'(v)||_2$ for all $uv \in E$, also satisfy it for all $u, v \in V$. A framework (G, f) is *rigid* if every continuous motion of its vertices which preserves the lengths of all edges, also preserves the distances between all pairs of vertices. The framework (G, f) is *generic* if the set containing the coordinates of all the points $f(v)$ is algebraically independent over the rationals. The graph G is *n-rigid* if every its n-dimensional generic realization is rigid. For generic frameworks, rigidity is equivalent to the stronger property of infinitesimal rigidity.

An *infinitesimal motion* of (G, f) is a map $m : V \to \mathbb{R}^n$ with $(m(u) - m(v))(f(u) - f(v)) = 0$ whenever $uv \in E$. A motion is *trivial* if it can be extended to an isometry of $\mathbb{R}^n$. A framework is an *infinitesimally rigid* if every motion of it is trivial, and it is *isostatic* if, moreover, the deletion of any its edge will cause loss of rigidity. (G, f) is an *elastic framework* if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for every edge-weighting $w : E \to \mathbb{R}_{>0}$ with $\max_{uv \in E} |w(uv) - ||f(u) - f(v)||_2| \leq \delta$, there exist a framework (G, f') with $\max_{v \in V} ||f(u) - f'(v)||_2 < \epsilon$.

A framework (G, f) with $||f(u) - f(v)||_2 > r$ if $u, v \in V, u \neq c$ and $||f(u), f(v)||_2 \leq R$ if $uv \in E$, for some $0 < r < R$, is called (Doyle–Snell, 1984) a *civilized drawing of a graph*. The random walks on such graphs are recurrent if $n = 1, 2$.

- **Distance constrained labeling**

  Given a sequence $\alpha = (\alpha_1, \ldots, \alpha_k)$ of **distance constraints** $\alpha_1 \geq \cdots \geq \alpha_k > 0$, a *$\lambda_\alpha$-labeling* of a graph $G = (V, E)$ is an assignment of labels $f(v)$ from the set $\{0, 1, \ldots, \lambda\}$ of integers to the vertices $v \in V$ such that, for any t with $0 \leq t \leq k$, $|f(v) - f(u)| \geq \alpha_t$ whenever the path distance between u and v is t.

  The *radio frequency assignment problem*, where vertices are transmitters (available channels) and labels represent frequencies of not-interfering channels, consists of minimizing λ. **Distance-two labeling** is the main interesting case $\alpha = (2, 1)$; its *span* is the difference between the largest and smallest labels used.

- **Distance-related graph embedding**

  An *embedding* of the guest graph $G = (V_1, E_1)$ into the host graph $H = (V_2, E_2)$ with $|V_1| \leq |V_2|$, is an injective map from V_1 into V_2.

  The **wire length**, *dilation* and *antidilation* of G in H are

$$\min_f \sum_{(uv) \in E_1} d_H(f(u), f(v)), \ \min_f \max_{(uv) \in E_1} d_H(f(u), f(v)), \ \max_f \min_{(uv) \in E_1} d_H(f(u), f(v)),$$

  respectively, where f is any embedding of G into H. The main **distance-related graph embedding** problems consist of finding or estimating these three parameters.

The *bandwidth* and *antibandwidth* of G is the dilation and antidilation, respectively, of G in a path H with V_1 vertices.

- **Bandwidth of a graph**
  Given a graph $G = (V, E)$ with $|V| = n$, its *ordering* is a bijective mapping $f : V \to \{1, \dots, n\}$. Given a number $b > 0$, the *bandwidth problem* for (G, b) is the existence of ordering f with the *stretch* $\max_{uv \in E} |f(u) - f(v)|$ at most b.
  The **bandwidth** of G, denoted by $bw(G)$, is the minimum stretch over all f.
  The *antibandwidth problem* for G is to find ordering f with maximal $\min_{uv \in E} |f(u) - f(v)|$ (*antibandwidth*).
- **Path distance width of a graph**
  Given a connected graph $G = (V, E)$, an ordered partition $V = \cup_{i=1}^{t} L_i$ of its vertices is called a *distance structure* on G if $L_i = \{v \in V : \min_{u \in L_1} d_{\text{path}}(u, v) = i - 1\}$ for $1 \leq i \leq t$. The structure is *rooted* if $|L_1| = 1$.
  The **path distance width** $pwd(G)$ of G is defined (Yamazaki et al., 1999) as $\min \max_{1 \leq i \leq t} |L_i|$ over all distance structures on G.
  An ordered partition $V = \cup_{i=1}^{t} L_i$ is called a *level structure* on G if for each edge uv with $u \in L_i$ and $v \in L_j$, it holds that $|i - j| \leq 1$. The *level width* (or *strong pathwidth*) $lw(G)$ is $\min \max_{1 \leq i \leq t} |L_i|$ over all level structures.
  Clearly, $lw(G) \leq pdw(G)$. Yamazaki et al., 1999, proved that $pdw(G)$ can be arbitrarily larger than the **bandwidth** $bw(G)$ and $lw(G) \leq bw(G) < 2lw(G)$.
- **Tree-length of a graph**
  A *tree decomposition* of a graph $G = (V, E)$ is a pair of a tree T with vertex-set W and a family of subsets $\{X_i : i \in W\}$ of V with $\cup_{i \in W} X_i = V$ such that

  1. for every edge $(uv) \in E$, there is a subset X_i containing u, v, and
  2. for every $v \in V$, the set $i \in W : v \in X_i$ induces a connected subtree of T.

  The *chordal graphs* (i.e., ones without induced cycles of length at least 4) are exactly those admitting a tree decomposition where every X_i is a clique.
  For tree decomposition, the *tree-length* is $\max_{i \in W} diam(X_i)$ ($diam(X_i)$ is the diameter of the subgraph of G induced by X_i) and *tree-width* is $\max_{i \in W} |X_i| - 1$.
  The **tree-length** of G (Dourisboure–Gavoille, 2004) and its **tree-width** (Robertson–Seymour, 1986) are the minima, over all tree decompositions, of above tree-length and tree-width. The *path-length* G is defined taking as trees only paths.
  Given a linear ordering $e_1, \dots, e_{|E|}$ of the edges of G, let, for $1 \leq i < |E|$, denote by $G_{\leq i}$ and $G_{i<}$ the graphs induced by the edges $\{e_1, \dots, e_i\}$ and $\{e_{i+1}, \dots, e_{|E|}\}$, respectively. The *linear-length* is $\max_{1 \leq i < |E|} diam(V(G_{\leq i}) \cap V(G_{i<}))$. The **linear-length** of G (Umezawa–Yamazaki, 2009) is the minimum of the above linear-length taken over all the linear orderings of its edges.
- **Spatial graph**
  A **spatial graph** (or *spatial network*) is a graph $G = (V, E)$, where each vertex v has a spatial position $(v_1, \dots, v_n) \in \mathbb{R}^n$. ($G$ is called a *geometric graph* if it is drawn on $\mathbb{R}^2$ and its edges are straight-line segments.)
  The *graph-theoretic dilation* and *geometric dilation* of G are, respectively:

$$\max_{v,u \in V} \frac{d(v,u)}{||v-u||_2} \text{ and } \max_{(vu)\in E} \frac{d(v,u)}{||v-u||_2}.$$

- **Distance Geometry problem**
Given a weighted finite graph $G = (V, E; w)$, the **Distance Geometry problem** (DGP) is the problem of realizing it as a **spatial graph** $G = (V', E')$, where $x : V \to V'$ is a bijection with $x(v) = (v_1, \dots, v_n) \in \mathbb{R}^n$ for every $v \in V$ and $E' = \{(x(u)x(v)) : (uv) \in E\}$, so that for every edge $(uv) \in E$ it holds that

$$||x(u) - x(u)||_2 = w(uv).$$

The main application of DGP is the *molecular DGP*: to find the coordinates of the atoms of a given molecular conformation are by exploiting only some of the distances between pairs of atoms found experimentally; cf. [MLLM13].
- **Arc routing problems**
Given a finite set X, a quasi-distance $d(x, y)$ on it and a set $A \subseteq \{(x, y) : x, y \in X\}$, consider the weighted digraph $D = (X, A)$ with the vertex-set X and arc-weights $d(x, y)$ for all arcs $(x, y) \in A$. For given sets V of vertices and E of arcs, the **arc routing problem** consists of finding a *shortest* (i.e., with minimal sum of weights of its arcs) (V, E)*-tour*, i.e., a circuit in $D = (X, A)$, visiting each vertex in V and each arc in E exactly once or, in a variation, at least once.
The *Asymmetric Traveling Salesman problem* corresponds to the case $V = X$, $E = \emptyset$; the *Traveling Salesman problem* is the symmetric version of it (usually, each vertex should be visited exactly once). The *Bottleneck Traveling Salesman problem* consists of finding a (V, E)-tour T with smallest $\max_{(x,y)\in T} d(x, y)$.
The *Windy Postman problem* corresponds to the case $V = \emptyset$, $E = A$, while the Chinese Postman problem is the symmetric version of it.
The above problems are also considered for general arc- or edge-weights; then, for example, the term *Metric TSP* is used when edge-weights in the Traveling Salesman problem satisfy the triangle inequality, i.e., d is a quasi-semimetric.
- **Steiner distance of a set**
The **Steiner distance of a set** $S \subset V$ of vertices in a connected graph $G = (V, E)$ is (Chartrand et al., 1989) the minimum *size* (number of edges) of a connected subgraph of G, containing S. Such a subgraph is a tree, and is called a *Steiner tree* for S. Cf. general **Steiner diversity** in **Steiner ratio** (Chap. 1).
The Steiner distance of the set $S = \{u, v\}$ is the path metric between u and v. The *Steiner k-diameter* of G is the maximum Steiner distance of any k-subset of V.
- **t-Spanner**
A *factor*, i.e., a spanning subgraph, $H = (V, E(H))$ of a connected graph $G = (V, E)$ is called a **t-spanner** (or *t-multiplicative spanner*) of G if, for every $u, v \in V$, the inequality $d^H_{\text{path}}(u, v)/d^G_{\text{path}}(u, v) \le t$ holds. The value t is called the *stretch factor* (or *dilation*) of H. Cf. **distance-related graph embedding** and **spatial graph**.
The graph $H = (V, E(H))$ is called a *k-additive spanner* of G if, for every $u, v \in V$, the inequality $d^H_{\text{path}}(u, v) \le d^G_{\text{path}}(u, v) + k$ holds.

Mulder and Nebeský, 2012, defined, for connected H, the *guide* of (H, G) as the ternary relation $R \subset V \times V \times V$ consisting of ordered triples (u, w, v) such that $uw \in E$ and $d^H_{\text{path}}(u, w) + d^H_{\text{path}}(w, v) = d^H_{\text{path}}(u, v)$. The guide of (G, G) is called the *step* ternary relation; cf. **metric betweenness** in Chap. 1.

- **Optimal realization of metric space**
  Given a finite metric space (X, d), a *realization* of it is a weighted graph $G = (V, E; w)$ with $X \subset V$ such that $d(x, y) = d_G(x, y)$ holds for all $x, y \in X$. The realization is **optimal** if it has minimal $\sum_{(uv) \in E} w(uv)$.
- **Proximity graph**
  Given a finite subset V of a metric space (X, d), its **proximity graph** is a graph representing neighbor relationships between points of V. Such graphs are used in Computational Geometry and many real-world problems. The main examples are presented below. Cf. **underlying graph of a metric space** in Chap. 1.
  A *spanning tree* of V is a set T of $|V|-1$ unordered pairs (x, y) of different points of V forming a tree on V; the *weight* of T is $\sum_{(x,y) \in T} d(x, y)$. A **minimum spanning tree** $MST(V)$ of V is a spanning tree with the minimal weight. Such a tree is unique if the edge-weights are distinct.
  A **nearest neighbor graph** is the digraph $NNG(V) = (V, E)$ with vertex-set $V = v_1, \dots, v_{|V|}$ and, for $x, y \in V$, $xy \in E$ if y is the *nearest neighbor* of x, i.e., $d(x, y) = \min_{v_i \in V \setminus \{x\}} d(x, v_i)$ and only v_i with maximal index i is picked. The *k-nearest neighbor graph* arises if k such v_i with maximal indices are picked. The undirect version of $NNG(V)$ is a subgraph of $MST(V)$.
  A **relative neighborhood graph** is (Toussaint, 1980) the graph $RNG(V) = (V, E)$ with vertex-set V and, for $x, y \in V$, $xy \in E$ if there is no point $z \in V$ with $\max\{d(x, z), d(y, z)\} < d(x, y)$. Also considered, for $(X, d) = (\mathbb{R}^2, ||x - y||_2)$, the related *Gabriel graph* $GG(V)$ (in general, *β-skeleton*) and *Delaunay triangulation* $DT(V)$; then $NNG(V) \subseteq MST(V) \subseteq RNG(V) \subseteq GG(V) \subseteq DT(V)$.
  For any $x \in V$, its *sphere of influence* is the open metric ball $B(x, r_x) = \{z \in X : d(x, z) < r\}$ in (X, d) centered at x with radius $r_x = \min_{z \in V \setminus \{x\}} d(x, z)$.
  Sphere of influence graph is the graph $SIG(V) = (V, E)$ with vertex-set V and, for $x, y \in V$, $xy \in E$ if $B(x, r_x) \cap B(y, r_y) \neq \emptyset$; so, it is a proximity graph and an *intersection graph*. The *closed sphere of influence graph* is the graph $CSIG(V) = (V, E)$ with $xy \in E$ if $\overline{B(x, r_x)} \cap \overline{B(y, r_y)} \neq \emptyset$.

15.3 Distances on Graphs

- **Chartrand–Kubicki–Schultz distance**
  The **Chartrand–Kubicki–Schultz distance** (or *ϕ-distance*, 1998) between two connected graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $|V_1| = |V_2| = n$ is

$$\min\{\sum |d_{G_1}(u, v) - d_{G_2}(\phi(u), \phi(v))|\},$$

where d_{G_1}, d_{G_2} are the path metrics of graphs G_1, G_2, the sum is taken over all unordered pairs u, v of vertices of G_1, and the minimum is taken over all bijections $\phi : V_1 \to V_2$.

- **Subgraph metric**
Let $\mathbb{F} = \{F_1 = (V_1, E_1), F_2 = (V_2, E_2), \dots, \}$ be the set of isomorphism classes of finite graphs. Given a finite graph $G = (V, E)$, denote by $s_i(G)$ the number of *injective homomorphisms* from F_i into G, i.e., the number of injections $\phi : V_i \to V$ with $\phi(x)\phi(y) \in E$ if $xy \in E_i$ divided by the number $\frac{|V|!}{(|V|-|V_i|)!}$ of such injections from F_i with $|V_i| \le |V|$ into $K_{|V|}$. Set $s(G) = (s_i(G))_{i=1}^{\infty} \in [0,1]^{\infty}$. Let d be the **Cantor metric** (cf. Chap. 18) $d(x,y) = \sum_{i=1}^{\infty} 2^{-i}|x_i - y_i|$ on $[0,1]^{\infty}$ or any metric on $[0,1]^{\infty}$ inducing the *product topology*. Then Bollobás–Riordan, 2007, defined the **subgraph metric** between the graphs G_1 and G_2 as
$$d(s(G_1), s(G_2))$$
and generalized it on *kernels* (or *graphons*), i.e., symmetric measurable functions $k : [0,1] \times [0,1] \to \mathbb{R}_{\ge 0}$, replacing G by k and the above $s_i(G)$ by
$$s_i(k) = \int_{[0,1]^{|V_i|}} \prod_{st \in E_i} k(x_s x_t) \prod_{s=1}^{|V_i|} dx_s.$$

- **Benjamini–Schramm metric**
The rooted graphs (G, o) and (G', o') (where $G = (V, E), G' = (V', E')$ and $o \in V, o' \in V'$) are *isomorphic* is there is a graph-isomorphism of G onto G' taking o to o'. Let X be the set of isomorphism classes of rooted connected locally finite graphs and let $(G, o), (G', o')$ be representatives of two classes.
Let k be the supremum of all radii r, for which rooted **metric balls** $(\overline{B}_G(o, r), o)$ and $(\overline{B}_{G'}(o', r), o')$ (in the usual **path metric**) are isomorphic as rooted graphs. Benjamini and Schramm, 2001, defined the metric 2^{-k} between classes represented by (G, o) and (G', o'). Here $2^{-\infty}$ means 0. Benjamini and Curien, 2011, defined the similar distance $\frac{1}{1+k}$.
- **Rectangle distance on weighted graphs**
Let $G = G(\alpha, \beta)$ be a complete weighted graph on $\{1, \dots, n\}$ with vertex-weights $\alpha_i > 0$, $1 \le i \le n$, and edge-weights $\beta_{ij} \in \mathbb{R}$, $1 \le i < j \le n$. Denote by $A(G)$ the $n \times n$ matrix $((a_{ij}))$, where $a_{ij} = \frac{\alpha_i \alpha_j \beta_{ij}}{(\sum_{1 \le i \le n} \alpha_i)^2}$.
The **rectangle distance** (or *cut distance*) between two weighted graphs $G = G(\alpha, \beta)$ and $G' = G(\alpha', \beta')$ (with vertex-weights (α'_i) and edge-weights (β'_{ij})) is defined (Borgs–Chayes–Lovász–Sós–Vesztergombi, 2007) by
$$\max_{I,J \subset \{1,\dots,n\}} \left| \sum_{i \in I, j \in J} (a_{ij} - a'_{ij}) \right| + \sum_{i=1}^{n} \left| \frac{\alpha_i}{\sum_{1 \le j \le n} \alpha_j} - \frac{\alpha'_i}{\sum_{1 \le j \le n} \alpha'_j} \right|,$$
where $A(G) = ((a_{ij}))$ and $A(G') = ((a'_{ij}))$.

In the case $(\alpha'_i) = (\alpha_i)$, the rectangle distance is $||A(G) - A(G')||_{cut}$, i.e., the **cut norm metric** (cf. Chap. 12) between matrices $A(G)$ and $A(G')$ and the *rectangle distance* from Frieze–Kannan, 1999. In this case, the l_1- and l_2-metrics between two weighted graphs G and G' are defined as $||A(G) - A(G')||_1$ and $||A(G) - A(G')||_2$, respectively. The subcase $\alpha_i = 1$ for all $1 \leq i \leq n$ corresponds to unweighted vertices. Cf. the **Robinson–Foulds weighted metric**.

Authors generalized the rectangle distance on *kernels* (or *graphons*), i.e., symmetric measurable functions $k : [0,1] \times [0,1] \to \mathbb{R}_{\geq 0}$, using the *cut norm* $||k||_{cut} = \sup_{S,T \subset [0,1]} | \int_{S \times T} k(x,y)dxdy|$.

A map $\phi : [0,1] \to [0,1]$ is *measure-preserving* if, for any measurable subset $A \subset [0,1]$, the measures of A and $\phi^{-1}(A)$ are equal. For a kernel k, define the kernel k^{ϕ} by $k^{\phi}(x,y) = k(\phi(x), \phi(y))$. The **Lovász–Szegedy semimetric** (2007) between kernels k_1 and k_1 is defined by

$$\inf_{\phi} ||k_1^{\phi} - k_2||_{cut},$$

where ϕ ranges over all measure-preserving bijections $[0,1] \to [0,1]$. Cf. **Chartrand–Kubicki–Schultz distance**.

- **Subgraph-supergraph distances**

A *common subgraph* of graphs G_1 and G_2 is a graph which is isomorphic to induced subgraphs of both G_1 and G_2. A *common supergraph* of graphs G_1 and G_2 is a graph which contains induced subgraphs isomorphic to G_1 and G_2.

The **Zelinka distance** d_Z [Zeli75] on the set **G** of all graphs (more exactly, on the set of all equivalence classes of isomorphic graphs) is defined by

$$d_Z = \max\{n(G_1), n(G_2)\} - n(G_1, G_2)$$

for any $G_1, G_2 \in \mathbf{G}$, where $n(G_i)$ is the number of vertices in G_i, $i = 1, 2$, and $n(G_1, G_2)$ is the maximum number of vertices of their common subgraph.

The **Bunke–Shearer metric** (1998) on the set of nonempty graphs is defined by

$$1 - \frac{n(G_1, G_2)}{\max\{n(G_1), n(G_2)\}}.$$

Given any set **M** of graphs, the **common subgraph distance** d_M on **M** is

$$\max\{n(G_1), n(G_2)\} - n(G_1, G_2),$$

and the **common supergraph distance** d_M^* is defined, for any $G_1, G_2 \in \mathbf{M}$, by

$$N(G_1, G_2) - \min\{n(G_1), n(G_2)\},$$

where $n(G_i)$ is the number of vertices in G_i, $i = 1, 2$, while $n(G_1, G_2)$ and $N(G_1, G_2)$ are the maximal order of a common subgraph $G \in \mathbf{M}$ and the minimal order of a common supergraph $H \in \mathbf{M}$, respectively, of G_1 and G_2.

d_M is a metric on **M** if the following condition (i) holds:

(i) if $H \in \mathbf{M}$ is a common supergraph of $G_1, G_2 \in \mathbf{M}$, then there exists a common subgraph $G \in \mathbf{M}$ of G_1 and G_2 with $n(G) \geq n(G_1)+n(G_2)-n(H)$.
 d_M^* is a metric on **M** if the following condition (ii) holds:

(ii) if $G \in \mathbf{M}$ is a common subgraph of $G_1, G_2 \in \mathbf{M}$, then there exists a common supergraph $H \in \mathbf{M}$ of G_1 and G_2 with $n(H) \leq n(G_1) + n(G_2) - n(G)$.

One has $d_M \leq d_M^*$ if the condition (i) holds, and $d_M \geq d_M^*$ if (ii) holds.
The distance d_M is a metric on the set **G** of all graphs, the set of all cycle-free graphs, the set of all bipartite graphs, and the set of all trees. The distance d_M^* is a metric on the set **G** of all graphs, the set of all connected graphs, the set of all connected bipartite graphs, and the set of all trees. The Zelinka distance d_Z coincides with d_M and d_M^* on the set **G** of all graphs. On the set **T** of all trees the distances d_M and d_M^* are identical, but different from the Zelinka distance.
The Zelinka distance d_Z is a metric on the set $\mathbf{G}(n)$ of all graphs with n vertices, and is equal to $n-k$ or to $K-n$ for all $G_1, G_2 \in \mathbf{G}(n)$, where k is the maximum number of vertices of a common subgraph of G_1 and G_2, and K is the minimum number of vertices of a common supergraph of G_1 and G_2.
On the set $\mathbf{T}(n)$ of all trees with n vertices the distance d_Z is called the **Zelinka tree distance** (see, for example, [Zeli75]).

- **Fernández–Valiente metric**
Given graphs G and H, let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be their *maximum common subgraph* and *minimum common supergraph*; cf. **subgraph-supergraph distances**. The **Fernández–Valiente metric** (2001) between G and H is

$$(|V_2| + |E_2|) - (|V_1| + |E_1|).$$

- **Graph edit distance**
The **graph edit distance** (Axenovich–Kézdy–Martin, 2008, and Alon–Stav, 2008) between graphs G and G' on the same labeled vertex-set is defined by

$$d_{ed}(G, G') = |E(G) \Delta E(G')|.$$

It is the minimum number of edge deletions or additions needed to transform G into G', and half of the Hamming distance between their adjacency matrices.
Given a *graph property* (i.e., a family $\mathcal{H}$ of graphs), let $d_{ed}(G, \mathcal{H})$ be $\min\{d_{ed}(G, G') : V(G') = V(G), G' \in \mathcal{H}\}$. Given a number $p \in (0, 1]$, the **edit distance function of a property** $\mathcal{H}$ is (if this limit exists) defined by

$$ed_{\mathcal{H}}(p) = \lim_{n \to \infty} \max \left\{ d_{ed}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \left\lfloor p \binom{n}{2} \right\rfloor \right\} \left(\binom{n}{2} \right)^{-1}.$$

If $\mathcal{H}$ is *hereditary* (closed under the taking induced subgraphs) and *nontrivial* (contains arbitrarily large graphs), then (Balogh–Martin, 2008) it holds

$$ed_{\mathcal{H}}(p) = \lim_{n \to \infty} \mathbb{E}[d_{ed}(G(n,p), \mathcal{H})] \left(\binom{n}{2} \right)^{-1};$$

$G(n, p)$ is the **random graph** (Chap. 1) on n vertices with edge probability p.
Bunke, 1997, defined the *graph edit distance* between vertex- and edge-labeled graphs G_1 and G_2 as the minimal total cost of matching G_1 and G_2, using deletions, additions and substitutions of vertices and edges. Cf. also **tree**, **top-down**, **unit cost** and **restricted edit distance** between rooted trees.
The **Bayesian graph edit distance** between two *relational graphs* (i.e., triples (V, E, A), where V, E, A are the sets of vertices, edges, *vertex-attributes*) is (Myers–Wilson–Hancock, 2000) their graph edit distance with costs defined by probabilities of operations along an editing path seen as a memoryless error process. Cf. **transduction edit distances** (Chap. 11) and **Bayesian distance** (Chap. 14).
The **structural Hamming distance** between two digraphs $G = (X, E)$ and $G' = (X, E')$ is defined (Acid–Campos, 2003) as $SHD(G, G') = |E \Delta E'|$.

- **Edge distance**
The **edge distance** on the set of all graphs is defined (Baláž et al., 1986) by

$$|E_1| + |E_2| - 2|E_{12}| + ||V_1| - |V_2||$$

for any graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, where $G_{12} = (V_{12}, E_{12})$ is a common subgraph of G_1 and G_2 with maximal number of edges. This distance has many applications in Organic and Medical Chemistry.

- **Contraction distance**
The **contraction distance** is a distance on the set $\mathbf{G}(n)$ of all graphs with n vertices defined by

$$n - k$$

for any $G_1, G_2 \in \mathbf{G}(n)$, where k is the maximum number of vertices of a graph which is isomorphic simultaneously to a graph, obtained from each of G_1 and G_2 by a finite number of *edge contractions*. To perform the *contraction* of the edge $uv \in E$ of a graph $G = (V, E)$ means to replace u and v by one vertex that is adjacent to all vertices of $V \backslash \{u, v\}$ which were adjacent to u or to v.

- **Edge move distance**
The **edge move distance** (Baláž et al., 1986) is a metric on the set $\mathbf{G}(n, m)$ of all graphs with n vertices and m edges, defined, for any $G_1, G_2 \in \mathbf{G}(m, n)$, as the minimum number of *edge moves* necessary for transforming the graph G_1 into the graph G_2. It is equal to $m - k$, where k is the maximum size of a common subgraph of G_1 and G_2.

An *edge move* is one of the *edge transformations*, defined as follows: H can be obtained from G by an edge move if there exist (not necessarily distinct) vertices u, v, w, and x in G such that $uv \in E(G)$, $wx \notin E(G)$, and $H = G - uv + wx$.

- **Edge jump distance**

  The **edge jump distance** is an extended metric (which in general can take the value ∞) on the set $\mathbf{G}(n, m)$ of all graphs with n vertices and m edges defined, for any $G_1, G_2 \in \mathbf{G}(m, n)$, as the minimum number of *edge jumps* necessary for transforming G_1 into G_2.

  An *edge jump* is one of the *edge transformations*, defined as follows: H can be obtained from G by an edge jump if there exist four distinct vertices u, v, w, and x in G, such that $uv \in E(G)$, $wx \notin E(G)$, and $H = G - av + wx$.

- **Edge flipping distance**

  Let $P = \{v_1, \dots, v_n\}$ be a collection of points on the plane. A *triangulation* T of P is a partition of the convex hull of P into a set of triangles such that each triangle has a disjoint interior and the vertices of each triangle are points of P.

  The **edge flipping distance** is a distance on the set of all triangulations of P defined, for any triangulations T and T_1, as the minimum number of edge flippings necessary for transforming T into T_1.

  An edge e of T is called *flippable* if it is the boundary of two triangles t and t' of T, and $C = t \cup t'$ is a convex quadrilateral. The *flipping* e is one of the *edge transformations*, which consists of removing e and replacing it by the other diagonal of C. Edge flipping is an special case of *edge jump*.

  The edge flipping distance can be extended on *pseudo-triangulations*, i.e., partitions of the convex hull of P into a set of disjoint interior *pseudo-triangles* (simply connected subsets of the plane that lie between any three mutually tangent convex sets) whose vertices are given points.

- **Edge rotation distance**

  The **edge rotation distance** (Chartand–Saba–Zou, 1985) is a metric on the set $\mathbf{G}(n, m)$ of graphs with n vertices and m edges, defined, for any G_1, G_2, as the minimum number of *edge rotations* needed for transforming G_1 into G_2.

  An *edge rotation* is one of the *edge transformations*, defined as follows: H can be obtained from G by an edge rotation if there exist distinct vertices u, v, and w in G, such that $uv \in E(G)$, $uw \notin E(G)$, and $H = G - uv + uw$.

- **Tree edge rotation distance**

  The **tree edge rotation distance** is a metric on the set $\mathbf{T}(n)$ of all trees with n vertices defined, for all T_1, $T_2 \in \mathbf{T}(n)$, as the minimum number of *tree edge rotations* necessary for transforming T_1 into T_2. A *tree edge rotation* is an *edge rotation* performed on a tree, and resulting in a tree.

  For $\mathbf{T}(n)$ the tree edge rotation and the edge rotation distances may differ.

- **Edge shift distance**

  The **edge shift distance** (or **edge slide distance**) is a metric (Johnson, 1985) on the set $\mathbf{G}_c(n, m)$ of all connected graphs with n vertices and m edges defined, for any $G_1, G_2 \in \mathbf{G}_c(m, n)$, as the minimum number of *edge shifts* necessary for transforming G_1 into G_2.

An *edge shift* is one of the *edge transformations*, defined as follows: H can be obtained from G by an edge shift if there exist distinct vertices u, v, and w in G such that $uv, vw \in E(G)$, $uw \notin E(G)$, and $H = G - uv + uw$. Edge shift is a special kind of *edge rotation* in the case when the vertices v, w are adjacent in G. The edge shift distance can be defined between any graphs G and H with components $G_i (1 \leq i \leq k)$ and $H_i (1 \leq i \leq k)$, respectively, such that G_i and H_i have the same order and the same size.

- **F-rotation distance**
  The **F-rotation distance** is a distance on the set $\mathbf{G}_F(n, m)$ of all graphs with n vertices and m edges, containing a subgraph isomorphic to a given graph F of order at least 2 defined, for all G_1, $G_2 \in \mathbf{G}_F(m, n)$, as the minimum number of *F-rotations* necessary for transforming G_1 into G_2.
  An *F-rotation* is one of the *edge transformations*, defined as follows: let F' be a subgraph of a graph G, isomorphic to F, let u, v, w be three distinct vertices of the graph G such that $u \notin V(F')$, $v, w \in V(F')$, $uv \in E(G)$, and $uw \notin E(G)$; H can be obtained from G by the F-rotation of the edge uv into the position uw if $H = G - uv + uw$.
- **Binary relation distance**
  Let R be a nonreflexive *binary relation* between graphs, i.e., $R \subset \mathbf{G} \times \mathbf{G}$, and there exists $G \in \mathbf{G}$ such that $(G, G) \notin R$.
  The **binary relation distance** is a metric (which can take the value ∞) on the set $\mathbf{G}$ of all graphs defined, for any graphs G_1 and G_2, as the minimum number of *R-transformations* necessary for transforming G_1 into G_2. We say that a graph H can be obtained from a graph G by an *R-transformation* if $(H, G) \in R$.
  An example is the distance between two *triangular embeddings of a complete graph* (i.e., its cellular embeddings in a surface with only 3-gonal faces) defined as the minimal number t such that, up to replacing t faces, the embeddings are isomorphic.
- **Crossing-free transformation metrics**
  Given a subset S of $\mathbb{R}^2$, a *noncrossing spanning tree* of S is a tree whose vertices are points of S, and edges are pairwise noncrossing straight line segments.
  The **crossing-free edge move metric** (see [AAH00]) on the set $\mathbf{T}_S$ of all noncrossing spanning trees of a set S, is defined, for any $T_1, T_2 \in \mathbf{T}_S$, as the minimum number of *crossing-free edge moves* needed to transform T_1 into T_2. Such move is an edge transformation which consists of adding some edge e in $T \in \mathbf{T}_S$ and removing some edge f from the induced cycle so that e and f do not cross.
  The **crossing-free edge slide metric** is a metric on the set $\mathbf{T}_S$ of all *noncrossing spanning trees* of a set S defined, for any $T_1, T_2 \in \mathbf{T}_S$, as the minimum number of *crossing-free edge slides* necessary for transforming T_1 into T_2. Such slide is one of the edge transformations which consists of taking some edge e in $T \in \mathbf{T}_S$ and moving one of its endpoints along some edge adjacent to e in T, without introducing edge crossings and without sweeping across points in S (that gives a new edge f instead of e). The edge slide is a special kind of crossing-free edge

move: the new tree is obtained by closing with f a cycle C of length 3 in T, and removing e from C, in such a way that f avoids the interior of the triangle C.

- **Traveling salesman tours distances**

  The *Traveling Salesman problem* is the problem of finding the shortest tour that visits a set of cities. We will consider only Traveling Salesman problem with undirected links. For an n-city traveling salesman problem, the space $\mathcal{T}_n$ of tours is the set of $\frac{(n-1)!}{2}$ cyclic permutations of the cities $1, 2, \dots, n$.

  The metric D on $\mathcal{T}_n$ is defined in terms of the difference in form: if tours $T, T' \in \mathcal{T}_n$ differ in m links, then $D(T, T') = m$.

  A *k-OPT transformation* of a tour T is obtained by deleting k links from T, and reconnecting. A tour T', obtained from T by a k-OPT transformation, is called a *k-OPT* of T. The distance d on the set $\mathcal{T}_N$ is defined in terms of the 2-OPT transformations: $d(T, T')$ is the minimal i, for which there exists a sequence of i 2-OPT transformations which transforms T to T'. In fact, $d(T, T') \leq D(T, T')$ for any $T, T' \in \mathcal{T}_N$ (see, for example, [MaMo95]). Cf. **arc routing problems**.

- **Orientation distance**

  The **orientation distance** (Chartrand–Erwin–Raines–Zhang, 2001) between two orientations D and D' of a finite graph is the minimum number of arcs of D whose directions must be reversed to produce an orientation isomorphic to D'.

- **Subgraphs distances**

  The standard distance on the set of all subgraphs of a connected graph $G = (V, E)$ is defined by

$$\min\{d_{\text{path}}(u, v) : u \in V(F), v \in V(H)\}$$

  for any subgraphs F, H of G. For any subgraphs F, H of a strongly connected digraph $D = (V, E)$, the standard quasi-distance is defined by

$$\min\{d_{dpath}(u, v) : u \in V(F), v \in V(H)\}.$$

  Using standard operations (rotation, shift, etc.) on the edge-set of a graph, one gets corresponding distances between its edge-induced subgraphs of given size which are subcases of similar distances on the set of all graphs of a given size and order.

  The **edge rotation distance** on the set $\mathbf{S}^k(G)$ of all edge-induced subgraphs with k edges in a connected graph G is defined as the minimum number of *edge rotations* required to transform $F \in \mathbf{S}^k(G)$ into $H \in \mathbf{S}^k(G)$. We say that H can be obtained from F by an *edge rotation* if there exist distinct vertices u, v, and w in G such that $uv \in E(F)$, $uw \in E(G) \backslash E(F)$, and $H = F - uv + uw$.

  The **edge shift distance** on the set $\mathbf{S}^k(G)$ of all edge-induced subgraphs with k edges in a connected graph G is defined as the minimum number of *edge shifts* required to transform $F \in \mathbf{S}^k(G)$ into $H \in \mathbf{S}^k(G)$. We say that H can be obtained from F by an *edge shift* if there exist distinct vertices u, v and w in G such that $uv, vw \in E(F)$, $uw \in E(G) \backslash E(F)$, and $H = F - uv + uw$.

The **edge move distance** on the set $\mathbf{S}^k(G)$ of all edge-induced subgraphs with k edges of a graph G (not necessary connected) is defined as the minimum number of *edge moves* required to transform $F \in \mathbf{S}^k(G)$ into $H \in \mathbf{S}^k(G)$. We say that H can be obtained from F by an *edge move* if there exist (not necessarily distinct) vertices u, v, w, and x in G such that $uv \in E(F)$, $wx \in E(G)\backslash E(F)$, and $H = F - uv + wx$. The edge move distance is a metric on $\mathbf{S}^k(G)$. If F and H have s edges in common, then it is equal to $k - s$.

The **edge jump distance** (which in general can take the value ∞) on the set $\mathbf{S}^k(G)$ of all edge-induced subgraphs with k edges of a graph G (not necessary connected) is defined as the minimum number of *edge jumps* required to transform $F \in \mathbf{S}^k(G)$ into $H \in \mathbf{S}^k(G)$. We say that H can be obtained from F by an *edge jump* if there exist four distinct vertices u, v, w, and x in G such that $uv \in E(F)$, $wx \in E(G)\backslash E(F)$, and $H = F - uv + wx$.

15.4 Distances on Trees

Let T be a *rooted tree*, i.e., a tree with one of its vertices being chosen as the *root*. The *depth* of a vertex v, $depth(v)$, is the number of edges on the path from v to the root. A vertex v is called a *parent* of a vertex u, $v = par(u)$, if they are adjacent, and $depth(u) = depth(v) + 1$; in this case u is called a *child* of v. A *leaf* is a vertex without child. Two vertices are *siblings* if they have the same parent.

The *in-degree* of a vertex is the number of its children. $T(v)$ is the subtree of T, rooted at a node $v \in V(T)$. If $w \in V(T(v))$, then v is an *ancestor* of w, and w is a *descendant* of v; $nca(u, v)$ is the *nearest common ancestor* of the vertices u and v.

T is called a *labeled tree* if a symbol from a fixed finite alphabet $\mathcal{A}$ is assigned to each node. T is called an *ordered tree* if a left-to-right order among siblings in T is given. On the set $\mathbb{T}_{rlo}$ of all rooted labeled ordered trees there are three *editing operations*:

- *Relabel*—change the label of a vertex v;
- *Deletion*—delete a nonrooted vertex v with parent v', making the children of v become the children of v'; the children are inserted in the place of v as a subsequence in the left-to-right order of the children of v';
- *Insertion*—the complement of deletion; insert a vertex v as a child of a v' making v the parent of a consecutive subsequence of the children of v'.

For unordered trees above operations (and so, distances) are defined similarly, but the insert and delete operations work on a subset instead of a subsequence.

We assume that there is a *cost function* defined on each editing operation, and the *cost* of a sequence of editing operations is the sum of the costs of these operations.

The *ordered edit distance mapping* is a representation of the editing operations. Formally, the triple (M, T_1, T_2) is an *ordered edit distance mapping* from T_1 to T_2, $T_1, T_2 \in \mathbb{T}_{rlo}$, if $M \subset V(T_1) \times V(T_2)$ and, for any (v_1, w_1), $(v_2, w_2) \in M$, the following conditions hold: $v_1 = v_2$ if and only if $w_1 = w_2$ (*one-to-one condition*),

v_1 is an ancestor of v_2 if and only if w_1 is an ancestor of w_2 (*ancestor condition*), v_1 is to the left of v_2 if and only if w_1 is to the left of w_2 (*sibling condition*).

We say that a vertex v in T_1 and T_2 is *touched by a line* in M if v occurs in some pair in M. Let N_1 and N_2 be the set of vertices in T_1 and T_2, respectively, not touched by any line in M. The *cost* of M is given by $\gamma(M) = \sum_{(v,w)\in M} \gamma(v \to w) + \sum_{v\in N_1} \gamma(v \to \lambda) + \sum_{w\in N_2} \gamma(\lambda \to w)$, where $\gamma(a \to b) = \gamma(a, b)$ is the *cost* of an editing operation $a \to b$ which is a relabel if $a, b \in \mathcal{A}$, a deletion if $b = \lambda$, and an insertion if $a = \lambda$. Here $\lambda \notin \mathcal{A}$ is a special *blank symbol*, and γ is a metric on the set $\mathcal{A} \cup \lambda$ (excepting the value $\gamma(\lambda, \lambda)$).

- **Tree edit distance**
  The **tree edit distance** (see [Tai79]) on the set $\mathbb{T}_{rlo}$ of all rooted labeled ordered trees is defined, for any $T_1, T_2 \in \mathbb{T}_{rlo}$, as the minimum cost of a sequence of editing operations (relabels, insertions, and deletions) turning T_1 into T_2.
  In terms of ordered edit distance mappings, it is equal to $\min_{(M,T_1,T_2)} \gamma(M)$, where the minimum is taken over all such mappings (M, T_1, T_2).
  The **unit cost edit distance** between T_1 and T_2 is the minimum number of three above editing operations turning T_1 into T_2, i.e., it is the tree edit distance with cost 1 of any operation.
- **Selkow distance**
  The **Selkow distance** (or *top-down edit distance*, *degree*-1 *edit distance*) is a distance on the set $\mathbb{T}_{rlo}$ of all rooted labeled ordered trees defined, for any $T_1, T_2 \in \mathbb{T}_{rlo}$, as the minimum cost of a sequence of editing operations (relabels, insertions, and deletions) turning T_1 into T_2 if insertions and deletions are restricted to leaves of the trees (see [Selk77]).
  The root of T_1 must be mapped to the root of T_2, and if a node v is to be deleted (inserted), then any subtree rooted at v is to be deleted (inserted).
  In terms of ordered edit distance mappings, it is equal to $\min_{(M,T_1,T_2)} \gamma(M)$, where the minimum is taken over all such mappings (M, T_1, T_2) such that $(par(v), par(w)) \in M$ if $(v, w) \in M$, where neither v nor w is the root.
- **Restricted edit distance**
  The **restricted edit distance** is a distance on the set $\mathbb{T}_{rlo}$ of all rooted labeled ordered trees defined, for any $T_1, T_2 \in \mathbb{T}_{rlo}$, as the minimum cost of a sequence of editing operations (relabels, insertions, and deletions) turning T_1 into T_2 with the restriction that disjoint subtrees should be mapped to disjoint subtrees.
  In terms of ordered edit distance mappings, it is equal to $\min_{(M,T_1,T_2)} \gamma(M)$, where the minimum is taken over all such mappings (M, T_1, T_2) satisfying the following condition: for all (v_1, w_1), (v_2, w_2), $(v_3, w_3) \in M$, $nca(v_1, v_2)$ is a proper ancestor of v_3 if and only if $nca(w_1, w_2)$ is a proper ancestor of w_3.
  This distance is equivalent to the *structure respecting edit distance* which is defined by $\min_{(M,T_1,T_2)} \gamma(M)$. Here the minimum is taken over all ordered edit distance mappings (M, T_1, T_2), satisfying the following condition: for all (v_1, w_1), (v_2, w_2), $(v_3, w_3) \in M$, such that none of v_1, v_2, and v_3 is an ancestor of the others, $nca(v_1, v_2) = nca(v_1, v_3)$ if and only if $nca(w_1, w_2) = nca(w_1, w_3)$.
  Cf. **constrained edit distance** in Chap. 11.

- **Alignment distance**
  The **alignment distance** (see [JWZ94]) is a distance on the set $\mathbb{T}_{rlo}$ of all rooted labeled ordered trees defined, for any $T_1, T_2 \in \mathbb{T}_{rlo}$, as the minimum *cost* of an *alignment* of T_1 and T_2. It corresponds to a restricted edit distance, where all insertions must be performed before any deletions.
  Thus, one inserts *spaces*, i.e., vertices labeled with a *blank symbol* λ, into T_1 and T_2 so that they become isomorphic when labels are ignored; the resulting trees are overlaid on top of each other giving the *alignment* $T_{\mathcal{A}}$ which is a tree, where each vertex is labeled by a pair of labels. The *cost* of $T_{\mathcal{A}}$ is the sum of the costs of all pairs of opposite labels in $T_{\mathcal{A}}$.
- **Splitting-merging distance**
  The **splitting-merging distance** (see [ChLu85]) is a distance on the set $\mathbb{T}_{rlo}$ of all rooted labeled ordered trees defined, for any $T_1, T_2 \in \mathbb{T}_{rlo}$, as the minimum number of vertex splittings and mergings needed to transform T_1 into T_2.
- **Degree-2 distance**
  The **degree-2 distance** is a metric on the set $\mathbb{T}_l$ of all labeled trees (*labeled free trees*), defined, for any $T_1, T_2 \in \mathbb{T}_l$, as the minimum number of editing operations (relabels, insertions, and deletions) turning T_1 into T_2 if any vertex to be inserted (deleted) has no more than two neighbors. This metric is a natural extension of the **tree edit distance** and the **Selkow distance**.

A *phylogenetic X-tree* is an unordered unrooted tree with the labeled leaf set X and no vertices of degree two. If every interior vertex has degree three, the tree is called *binary*. Let $\mathbb{T}(X)$ denote the set of all phylogenetic X-trees.

- **Robinson–Foulds metric**
  A *cut* $A|B$ of X is a *partition* of X into two subsets A and B (see **cut semimetric**). Removing an edge e from a phylogenetic X-tree induces a cut of the leaf set X which is called the *cut associated with* e.
  The **Robinson–Foulds metric** (or *Bourque metric*, *bipartition distance*) is a metric on the set $\mathbb{T}(X)$, defined, for any phylogenetic X-trees $T_1, T_2 \in \mathbb{T}(X)$, by
  $$\frac{1}{2}|\Sigma(T_1)\triangle\Sigma(T_2)| = \frac{1}{2}|\Sigma(T_1) \setminus \Sigma(T_2)| + \frac{1}{2}|\Sigma(T_2) \setminus \Sigma(T_1)|,$$
  where $\Sigma(T)$ is the collection of all cuts of X associated with edges of T.
  The **Robinson–Foulds weighted metric** is a metric on the set $\mathbb{T}(X)$ of all phylogenetic X-trees defined by
  $$\sum_{A|B\in\Sigma(T_1)\cup\Sigma(T_2)} |w_1(A|B) - w_2(A|B)|$$
  for all $T_1, T_2 \in \mathbb{T}(X)$, where $w_i = (w(e))_{e\in E(T_i)}$ is the collection of positive weights, associated with the edges of the X-tree T_i, $\Sigma(T_i)$ is the collection of all

cuts of X, associated with edges of T_i, and $w_i(A|B)$ is the weight of the edge, corresponding to the cut $A|B$ of X, $i = 1, 2$. Cf. more general **cut norm metric** in Chap. 12 and **rectangle distance on weighted graphs**.

- **μ-metric**
  Given a phylogenetic X-tree T with n leaves and a vertex v in it, let $\mu(v) = (\mu_1(v), \dots, \mu_n(v))$, where $\mu_i(v)$ is the number of different paths from the vertex v to the i-th leaf. Let $\mu(T)$ denote the multiset on the vertex-set of T with $\mu(v)$ being the multiplicity of the vertex v.
  The **μ-metric** (Cardona–Roselló–Valiente, 2008) is a metric on the set $\mathbb{T}(X)$ of all phylogenetic X-trees defined, for all $T_1, T_2 \in \mathbb{T}(X)$, by
  $$\frac{1}{2}|\mu(T_1)\Delta\mu(T_2)|,$$
  where Δ denotes the symmetric difference of multisets.
  Cf. the **metrics between multisets** in Chap. 1 and the **Dodge–Shiode WebX quasi-distance** in Chap. 22.
- **Nearest neighbor interchange metric**
  The **nearest neighbor interchange metric** (or **crossover metric**) on the set $\mathbb{T}(X)$ of all phylogenetic X-trees, is defined, for all $T_1, T_2 \in \mathbb{T}(X)$, as the minimum number of *nearest neighbor interchanges* required to transform T_1 into T_2.
  A *nearest neighbor interchange* consists of swapping two subtrees in a tree that are adjacent to the same internal edge; the remainder of the tree is unchanged.
- **Subtree prune and regraft distance**
  The **subtree prune and regraft distance** is a metric on the set $\mathbb{T}(X)$ of all phylogenetic X-trees defined, for all $T_1, T_2 \in \mathbb{T}(X)$, as the minimum number of *subtree prune and regraft transformations* required to transform T_1 into T_2.
  A *subtree prune and regraft transformation* proceeds in three steps: one selects and removes an edge uv of the tree, thereby dividing the tree into two subtrees T_u (containing u) and T_v (containing v); then one selects and subdivides an edge of T_v, giving a new vertex w; finally, one connects u and w by an edge, and removes all vertices of degree two.
- **Tree bisection-reconnection metric**
  The **tree bisection-reconnection metric** (or **TBR-metric**) on the set $\mathbb{T}(X)$ of all phylogenetic X-trees is defined, for all $T_1, T_2 \in \mathbb{T}(X)$, as the minimum number of *tree bisection and reconnection* transformations required to transform T_1 into T_2.
  A *tree bisection and reconnection transformation* proceeds in three steps: one selects and removes an edge uv of the tree, thereby dividing the tree into two subtrees T_u (containing u) and T_v (containing v); then one selects and subdivides an edge of T_v, giving a new vertex w, and an edge of T_u, giving a new vertex z; finally one connects w and z by an edge, and removes all vertices of degree two.

- **Quartet distance**
  The **quartet distance** (see [EMM85]) is a distance of the set $\mathbb{T}_b(X)$ of all binary phylogenetic X-trees defined, for all $T_1, T_2 \in \mathbb{T}_b(X)$, as the number of mismatched *quartets* (from the total number $\binom{n}{4}$ possible quartets) for T_1 and T_2.
  This distance is based on the fact that, given four leaves $\{1, 2, 3, 4\}$ of a tree, they can only be combined in a binary subtree in three ways: $(12|34)$, $(13|24)$, or $(14|23)$: the notation $(12|34)$ refers to the binary tree with the leaf set $\{1, 2, 3, 4\}$ in which removing the inner edge yields the trees with the leaf sets $\{1, 2\}$ and $\{3, 4\}$.
- **Triples distance**
  The **triples distance** (see [CPQ96]) is a distance of the set $\mathbb{T}_b(X)$ of all binary phylogenetic X-trees defined, for all $T_1, T_2 \in \mathbb{T}_b(X)$, as the number of triples (from the total number $\binom{n}{3}$ possible triples) that differ (for example, by which leaf is the outlier) for T_1 and T_2.
- **Perfect matching distance**
  The **perfect matching distance** is a distance on the set $\mathbb{T}_{br}(X)$ of all rooted binary phylogenetic X-trees with the set X of n labeled leaves defined, for any $T_1, T_2 \in \mathbb{T}_{br}(X)$, as the minimum number of interchanges necessary to bring the perfect matching of T_1 to the perfect matching of T_2.
  Given a set $A = \{1, \dots, 2k\}$ of $2k$ points, a *perfect matching* of A is a *partition* of A into k pairs. A rooted binary phylogenetic tree with n labeled leaves has a root and $n-2$ internal vertices distinct from the root. It can be identified with a perfect matching on $2n-2$, different from the root, vertices by following construction: label the internal vertices with numbers $n+1, \dots, 2n-2$ by putting the smallest available label as the parent of the pair of labeled children of which one has the smallest label among pairs of labeled children; now a matching is formed by peeling off the children, or sibling pairs, two by two.
- **Tree rotation distance**
  The **tree rotation distance** is a distance on the set $\mathbf{T}_n$ of all rooted ordered binary trees with n interior vertices defined, for all $T_1, T_2 \in \mathbf{T}_n$, as the minimum number of *rotations*, required to transform T_1 into T_2.
  Given interior edges uv, vv', vv'' and uw of a binary tree, the *rotation* is replacing them by edges uv, uv'', vv' and vw.
  There is a bijection between edge flipping operations in triangulations of convex polygons with $n+2$ vertices and rotations in binary trees with n interior vertices.
- **Attributed tree metrics**
  An *attributed tree* is a triple (V, E, α), where $T = (V, E)$ is the underlying tree, and α is a function which assigns an *attribute vector* $\alpha(v)$ to every vertex $v \in V$. Given two attributed trees (V_1, E_1, α) and (V_2, E_2, β), consider the set of all *subtree isomorphisms* between them, i.e., the set of all isomorphisms $f : H_1 \to H_2$, $H_1 \subset V_1$, $H_2 \subset V_2$, between their *induced subtrees*.
  Given a similarity s on the set of attributes, the similarity between isomorphic induced subtrees is defined as $W_s(f) = \sum_{v \in H_1} s(\alpha(v), \beta(f(v)))$. Let ϕ be the isomorphism with maximal similarity $W_s(\phi) = W(\phi)$.

The following four semimetrics on the set $\mathbf{T}_{att}$ of all attributed trees are used:

$$\max\{|V_1|, |V_2|\} - W(\phi), \quad |V_1| + |V_2| - 2W(\phi) \quad \text{and}$$

$$1 - \frac{W(\phi)}{\max\{|V_1|, |V_2|\}}, \quad 1 - \frac{W(\phi)}{|V_1| + |V_2| - W(\phi)}.$$

They become metrics on the set of equivalences classes of attributed trees: two such trees (V_1, E_1, α) and (V_2, E_2, β) are called *equivalent* if they are *attribute-isomorphic*, i.e., if there exists an isomorphism $g : V_1 \to V_2$ between the trees such that, for any $v \in V_1$, we have $\alpha(v) = \beta(g(v))$. Then $|V_1| = |V_2| = W(g)$.

- **Maximal agreement subtree distance**
  The **maximal agreement subtree distance** (MAST) is a distance of the set $\mathbf{T}$ of all trees defined, for all $T_1, T_2 \in \mathbf{T}$, as the minimum number of leaves removed to obtain a *(greatest) agreement subtree*.
  An *agreement subtree* (or *common pruned tree*) of two trees is an identical subtree that can be obtained from both trees by pruning leaves with the same label.