CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

PROTEIN FOLDING: PLANAR CONFIGURATION SPACES OF DISC ARRANGEMENTS AND HINGED POLYGONS: *PROTEIN FOLDING IN FLATLAND*

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Computer Science

by

Clinton Bowen

August 2014

The thesis of Clinton Bowen is approved:

Dr. Silvia Fernandez	Date
Dr. John Dye	Date
Dr. Csaba Tóth, Chair	Date

California State University, Northridge

Table of Contents

ABSTRACT

PROTEIN FOLDING: PLANAR CONFIGURATION SPACES OF DISC ARRANGEMENTS AND HINGED POLYGONS: *PROTEIN FOLDING IN FLATLAND*

By

Clinton Bowen

Master of Science in Computer Science

abstract goes here

Chapter 1

Background

In this section, we cover the background subjects needed to formally pose the problem and present solutions in this thesis. We start with two types of combinatorial structures, linkages and polygonal linkages. We then discuss the configuration spaces of linkages and polygonal linkages. We then look into an alternate representation of linkages, disk arrangements and state the disk arrangement theorem. We then look at satisfiability problems and then review a framework, the logic engine, which can encode a type of satisfiability problem. Finally, we cover the basic definitions of algorithm complexity for **P** and **NP**.

1.1 Graphs

A *graph* is an ordered pair $G = (V, E)$ comprising of a set V of vertices and a set E of edges or lines. Every edge $e \in E$, is an unordered pair of distinct vertices $u, v \in V$ (the edge represents their adjacency, $e = \{u, v\}$). With this definition of a graph, there are no loops (self adjacent vertices, $\{v, v\}$) or multi-edges (several edges between the same pair of vertices).

A motivation for using graphs is modelling physical objects like molecules. This requires an embedding into the plane or $\mathbb{R}^3$. An *embedding* of the graph $G = (V, E)$ is an injective mapping $\Pi : V \mapsto \mathbb{R}^2$ (see Figure 1.1).

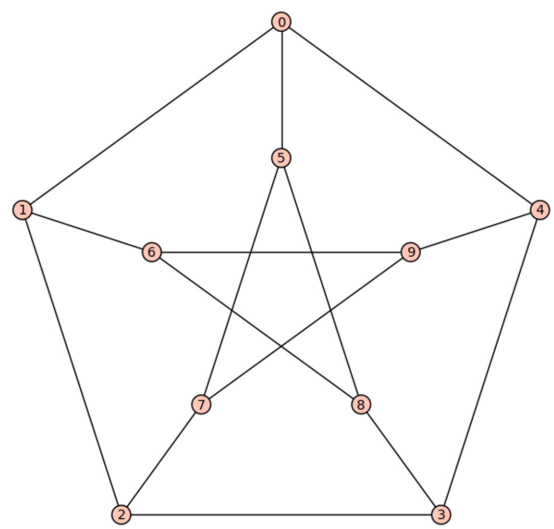


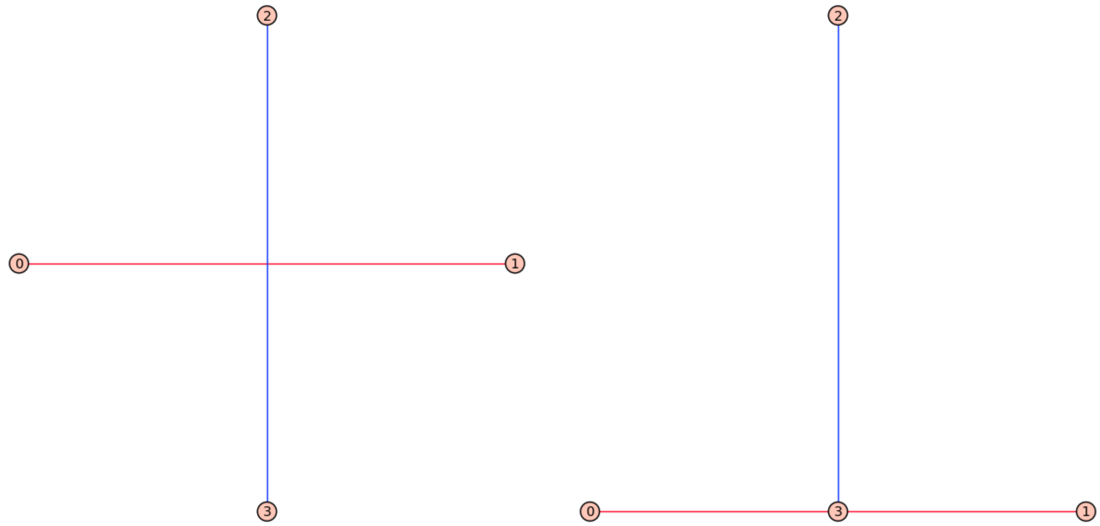
Figure 1.1: An embedding of the Peterson graph.

1.1.0.1 Edge Crossings

We define *plane embeddings* of a graph to be an embedding where the following degenerate configurations do not occur:

(i) the interiors of two or more edges intersect, or

(ii) an edge passes through a vertex



(a) The interior of the edges intersect.

(b) An edge passes through a vertex.

Figure 1.2: These figures exhibit the 4 types of edge crossings.

A graph is called *planar* if it admits a plane embedding. A *plane graph* is a graph together with a plane embedding.

1.1.1 Trees

A *path* is a sequence of vertices in which every two consecutive vertices are connected by an edge. A *simple cycle* of a graph is a sequence, $(v_1, v_2, \dots, v_{t-1}, v_t)$, of distinct vertices such that every two consecutive vertices are connected by an edge, and the last vertex, v_t, connects to v_1. A graph is *connected* if for any two vertices, there exists a path between the two points.

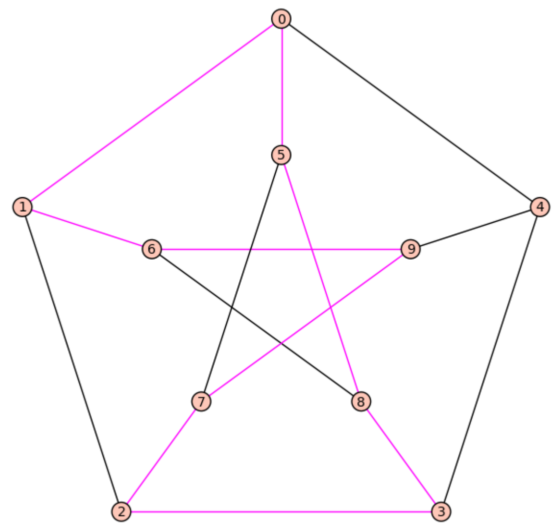
A *tree* is a graph that has no simple cycles and is connected.

1.1.2 Ordered Trees

An *ordered tree* is a tree together with a cyclic order of the neighbors for each vertex. Embeddings of ordered trees are equivalent if for each node the counter-clockwise ordering of adjacent nodes are the same.

1.1.3 Graph Isomorphism

To determine when two graphs are equivalent, we need to define an isomorphism for graphs. Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, a *graph isomorphism* a bijective function $f : V_1 \mapsto V_2$ such that for any two vertices $u, v \in V_1$, we have $\{u, v\} \in E_1$, if and only if $(f(u), f(v)) \in E_2$.



(a) An embedding of the Peterson graph with a simple cycle of (2,7,9,6,1,0,5,8,3).

$G = (V, E)$
$V = \{1, 2, 3, 4, 5, 6, 7, 8\}$
$E = \{12, 23, 24, 35, 36, 37, 48, \}$
$Order_1 = \{(2)_1, (134)_2, (2567)_3, (8)_4, (3)_5, (3)_6, (3)_7, (4)_8\}$
$Order_2 = \{(2)_1, (143)_2, (2657)_3, (8)_4, (3)_5, (3)_6, (3)_7, (4)_8\}$

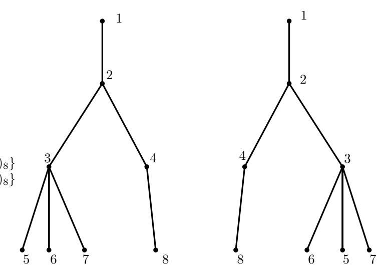


Figure 1.4: A tree with two embeddings with different cyclic orderings around vertices.

Graph	Vertices	Edges
G_1	$\{a,b,c,d,e\}$	$\{ab,(b,c),(c,d),(d,e),(e,a)\}$
G_2	$\{1,2,3,4,5\}$	$\{(1,2),(2,3),(3,4),(4,5),(5,1)\}$

Table 1.1: Two graphs that are isomorphic with the alphabetical isomorphism $f(a) = 1$, $f(b) = 2$, $f(c) = 3$, $f(d) = 4$, $f(e) = 5$.

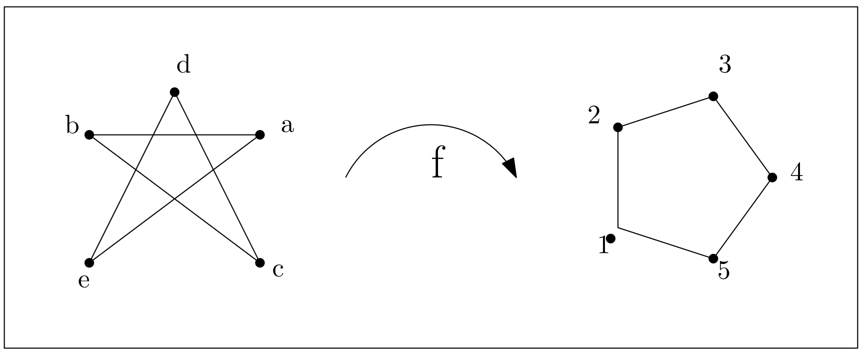


Figure 1.5: This figure depicts the graph isomorphism shown in Table (**??**) between V_1 and V_2 in the plane.

1.2 Linkages

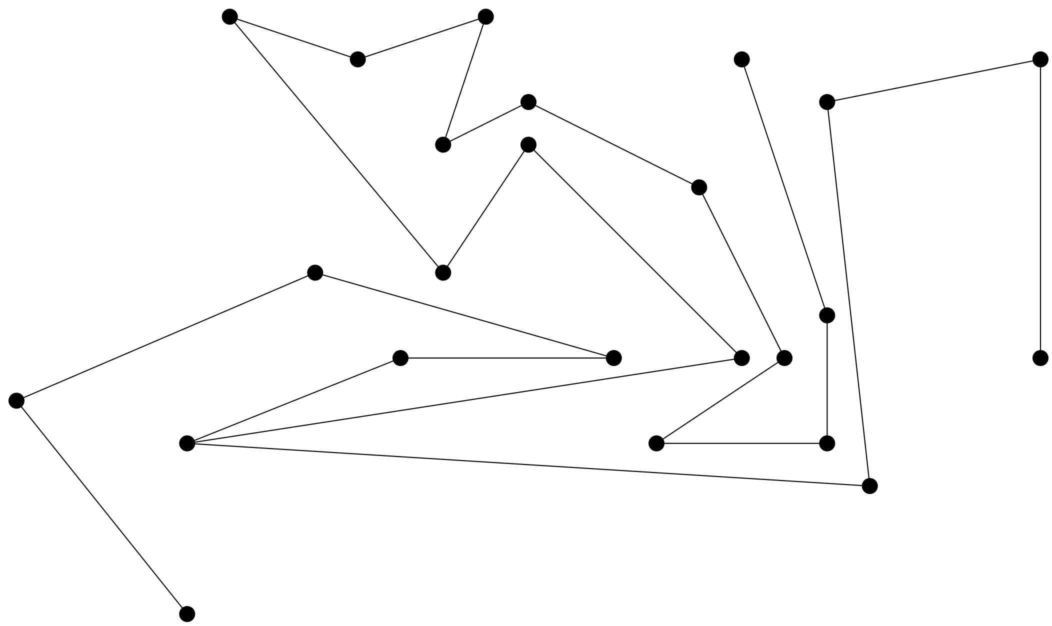


Figure 1.6: An embedded linkage

When graphs model physical objects, distances between adjacent vertices matter. The length assignment of a graph $G = (V,E)$ is $\ell : E \mapsto \mathbb{R}^+$. A *linkage* is a graph $G = (V,E)$ with a length assignment $l : E \mapsto \mathbb{R}^+$.

We consider embeddings of a graph that respects the length assignment. A *realization* of a linkage, G and l, is an embedding of a graph, Π, such that for every edge $\{u,v\} \in E$, $\ell(\{u,v\}) = |\Pi(u) - \Pi(v)|$. A *plane realization* is a plane embedding with the property, $l(\{u,v\}) = |\Pi(u) - \Pi(v)|$.

1.3 Polygonal Linkages

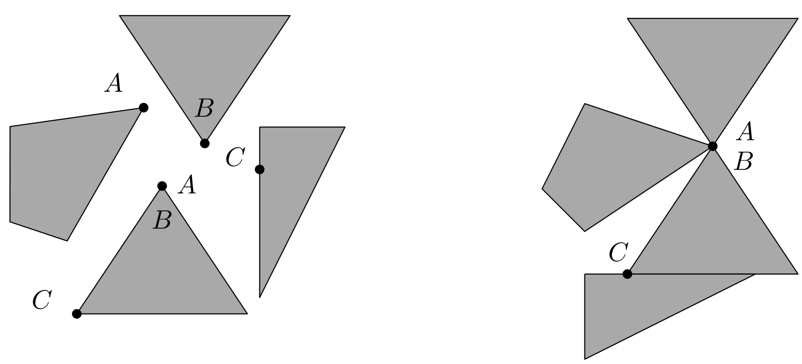


Figure 1.7: (a) A polygonal linkage with a non-convex polygon and two hinge points corresponding to three polygons. Note that hinge points correspond to two distinct polygons.(b) Illustrating that two hinge points can correspond to the same boundary point of a polygon.

A generalization of linkages are polygonal linkages where the edges of given lengths are replaced by rigid polygons. Formally, a *polygonal linkage* is an ordered pair $(\mathsf{PP}, \mathscr{H})$ where PP is a finite set of polygons and $\mathscr{H}$ is a finite set of hinges; a *hinge* $h \in \mathscr{H}$ corresponds to two points on the boundary of two distinct polygons in PP. A *realization* of a polygonal linkage is an interior-disjoint placement of congruent copies of the polygons in PP such that the points corresponding to each hinge are identified (Fig. **??**). This definition of realization rules well known geometric dissections (e.g. Fig. 1.8).

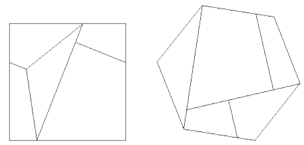


Figure 1.8: Two configurations of polygonal linkage where the polygons touch on boundary segments instead of hinges. These two realizations of the polygonal linkage are invalid to our definitions.

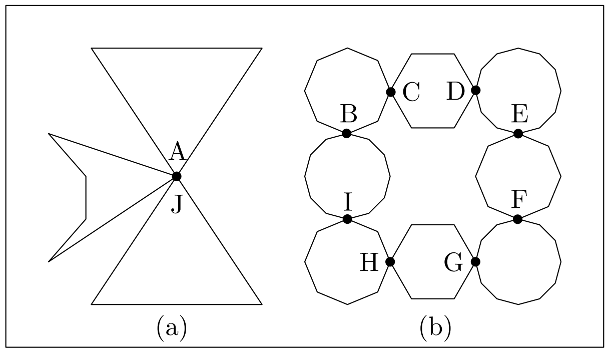


Figure 1.9: (a) A polygonal linkage with a non-convex polygon and a hinge point corresponding to three polygons. (b) A polygonal linkage with 8 regular polygons.

1.3.1 Disk Arrangements

It turns out the disk arrangements are an equivalent way to to represent plane graphs. By representing vertices as interioir disjoint disks and by representing edges as as points of intersections (contact), *kissing points* between two disks. The graph corresponding to a given disk arrangement, $\mathscr{D}$, is said to be the *contact graph*. A *disk arrangement* is a set, $\mathscr{D}$, of pairwise interior-disjoint disks in the plane, $\mathscr{D} = \{C_i\}_{i=1}^n$. $\{C_i\}_{i=1}^n$ such that for any circle $C \in \{C_i\}_{i=1}^n$, C

A classical result by Thurston and Koebe is that every disk arrangement embedded into the plane had a corresponding plane graph.

Theorem 1.3.1 (**??**Disk Packing Theorem)**.** *For every graph G, there is a disk arrangement in the plane whose contact graph is isomorphic to G.*

Proposition 1.3.1. *For every linkage L, there is a disk arrangement in the plane whose contact graph is isomorphic to L.*

1. Show the relation between polygonal linkages and disk arrangengements.

1.3.1.1 Ordered Disk Arrangement

Suppose we're given a tree. By the disk packing theorem we can ascertain a sense of order for the isomorphic disk packing. An *ordered disk arrangement* is a rooted tree in which the counter-clockwise ordering of adjacent vertices.

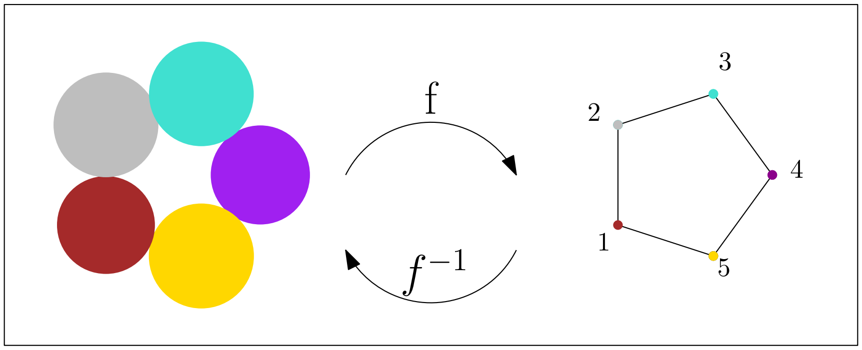


Figure 1.10: This example represents a disk arrangement transformed to and from its corresponding graph G_2

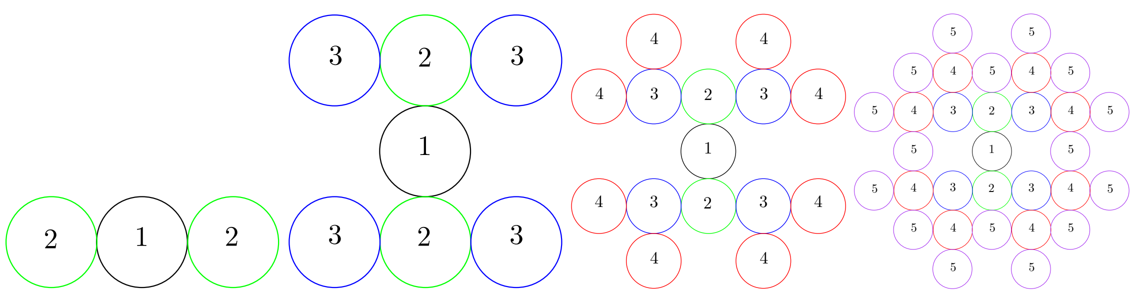
1.3.1.2 Disk Packing Confinement Problem

Given inputs of radii By adding constraints to the embeddings of disk arrangements, we can devise realizability problem by a volume argument.

1. Round 1: Start with a disk of unit radius.
2. Round 2: Add two kissing disks, each of diameter 2, that do not intersect with any other disk (they may kiss other disk).
3. Round 3 and Higher: For each new kissing disk added, add two more non-intersecting kissing disks of diameter 2 to it.

For each round i we are adding $2^{(i-1)}$ disks, each with an of π. The area that the disk arrangement is bounded by at round i is a box of length $2\cdot(2\cdot(i-1)+1)$ totalling to an area of $(4\cdot i^2-4\cdot i+1)$. Meanwhile the total area of the disk arrangement at round i is $\pi\cdot(2^i-1)$. The exponential growth rate of the disk packing will exceed its bounded area for sufficiently large i.

Figure (1.11) illustrates the iterative problem. The problem with this is that the area in which is necessary to contain this disk growing disk arrangement will exceed the area needed to contain it.



(a) A disk arrangement with two layers of disks (b) A disk arrangement with three layers of disks (c) A disk arrangement with four layers of disks (d) A disk arrangement with five layers of disks

Figure 1.11: The gradual growth of disk arrangements by adding two kissing disks to each of the previously generated disks. By continuing this arrangement growth, the space needed to contain the kissing disks will exceed the area containing the disk arrangements.

1.4 Configuration Spaces

Just as one can compose colors or forms, so one can compose motions.

Alexander Calder, 1933

We'd like to describe motions and range of motions of embedded graphs, linkages, polygonal linkages, and disk arrangements. Table 1.4 provides the definition of *reconfiguration* for each type of object covered so far:

Object Type	Definition of Reconfiguration
Graph	a continuous motion of the vertices that preserves the lengths of the edges and never causes the edges to intersect.
Linkage	same as graph
Polygonal Linkage	a continuous motion of polygons that preserves shapes of polygons, hinge point pairings, and never causes the polygonal sides to intersect.
Disk Arrangement	a continuous motion of disks that preserves disk radii, pairs of contact points, and never causes disks to intersect.

1.4.1 Configuration Spaces of Linkages

Let's focus on the space of embeddings of a linkage. If there are n vertices of a linkage, the *configuration space* of a linkage is said to be a vector space of dimension $2 \cdot n$ where edge length is preserved. A *configuration space* for a linkage G and corresponding proper embedding, L_1 is said to be for any other

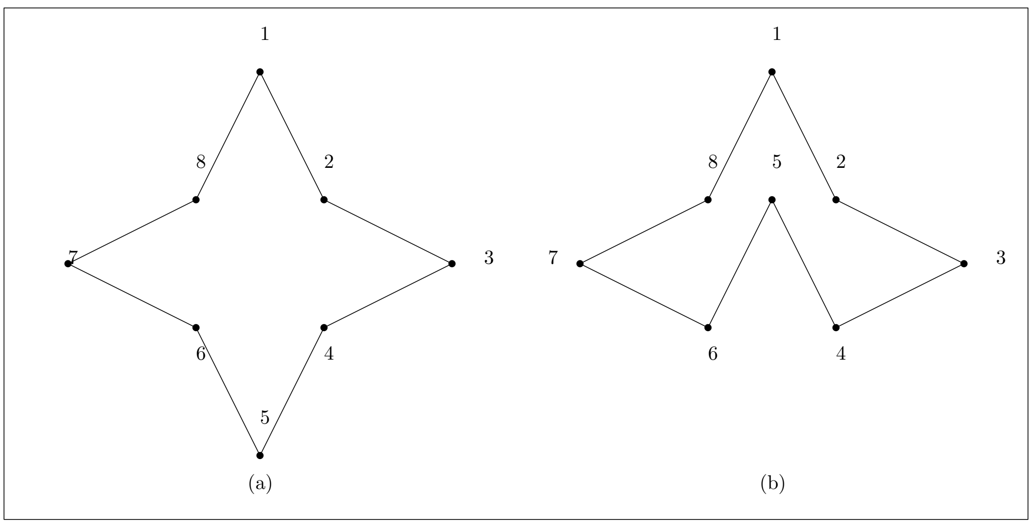


Figure 1.12: (a) and (b) show a linkage in two embeddings.

proper embedding of a linkage G, L_2, such that the lengths of every edge of G is preserved between the two embeddings, i.e.:

$$l((u,v)) = |L_1(u) - L_1(v)| = |L_2(u) - L_2(v)|$$

Equivalent embeddings include translations and rotations about the center of mass on $L(V)$. We further our embeddings by requiring that one vertice is pinned to the point of origin on the plane as well as a neighboring vertex.

Theorem 1.4.1 ([**?**, **?**]Carpenter's Rule Theorem)**.** *Every realization of a linkage can be continuously moved (without self-intersection) to any other realization. In other words, the realization space of such a linkage is always connected.*

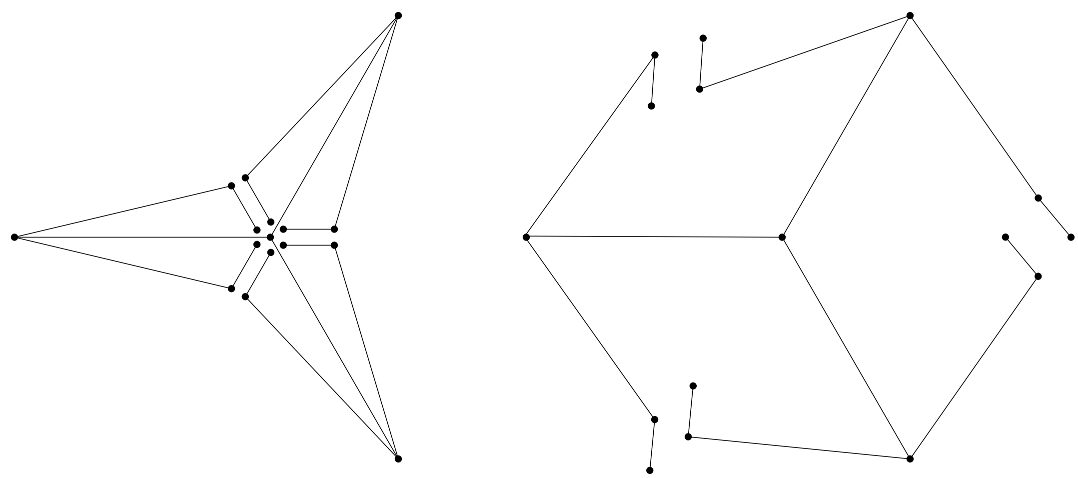


Figure 1.13: A linkage whose complete configuration space is discontinuous. These two examples above are two configurations of the same linkage that cannot continuously transform into the other without edge crossing.

A *reconfiguration* of a linkage whose graph is $G = (V, E)$ and length assignment is ℓ is a continuous function $f : [0,1] \mapsto \mathbb{R}^{2\cdot|V|}$ specifying a configuration of the linkage for every $t \in [0,1]$ where length assignment ℓ is preserved, edges do not cross and for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $|t_1 - t_2| < \delta$ implies

$$|f(t_1) - f(t_2)| < \varepsilon$$

1.4.2 Configuration Spaces of Polygonal Linkages

Some text goes here

1.4.3 Configuration Spaces of Disk Arrangements

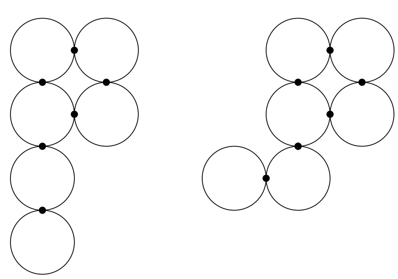


Figure 1.14: A linkage whose complete configuration space is discontinuous. These two examples above are two configurations of the same linkage that cannot continuously transform into the other without edge crossing.