CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

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

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

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

By

Clinton Bowen

Master of Science in Applied Mathematics

Chapter 1

Realizability Problems for Weighted Trees

In this chapter our goal is to prove Theorem **??**: It is NP-Hard to decide whether a given tree with positive vertex weights is the contact graph of a disk arrangements with specified radii. This chapter's approach to proving Theorem **??** introduces an ordered weighted tree T, perturbed ordered weight tree T_ε, the Hausdorff distance, and then prove a lemma which shows that hexagons can be approximated by an ordered disk contact graph corresponding to the weighted tree T_ε.

1.1 Hausdorff Distance

Let A and B be sets in the plane. The *directed Hausdorff distance* is:

$$d(A,B) = \sup_{a\in A} \inf_{b\in B} ||a-b|| \tag{1.1}$$

$d(A,B)$ finds the furthest point $a \in A$ from any point in B. *Hausdorff distance* is

$$D(A,B) = \max\{d(A,B), d(B,A)\} \tag{1.2}$$

In Figure 1.1, we have two sets X and Y and illustrate $d(X,Y)$ and $d(Y,X)$. From this, it is possible to calculate the Hausdorff distance between X and Y.

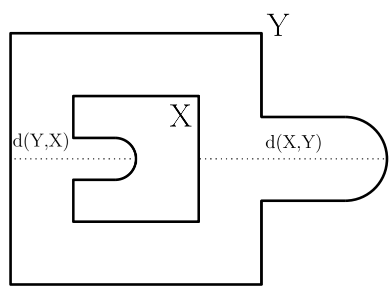

Figure 1.1: An illustrative example of $d(X,Y)$ and $d(Y,X)$ where X is the inner curve, and Y is the outer curve.

ε-approximation The weighted graph, G, is an ε*-approximation* of a polygon P if the Hausdorff distance between every realization of G as a contact graph of disks and a congruent copy of P is at most epsilon. A weighted graph G is said to be a $O(f(x))$*-approximation* of a polygon P if there is a positive constant M such that for all sufficiently large values of x the Hausdorff distance between every realization such realization of G as a contact graph of disks and a congruent copy of P is at $M \cdot |f(x)|$. A weighted graph G is said to be a *stable* if it has the property that for every two such realizations of G, the distance between the centers of the corresponding disks is at most ε after a suitable rigid transformation.

Suppose we have a unit disk U and we have a grid overlayed on the disk with side length δ. Let $S_1(\delta)$ be the union of grid squares completely in the interior of U. Let $S_2(\delta)$ be the union of squares with some point of the boundary of the square contained in the interior of disk U. The Hausdorff distance of U and $S_1(\delta)$ is at most $H(S_1(\delta),U) = \sqrt{2}\delta$. Similarly, the Hausdorff distance of U and $S_2(\delta)$ is at most $H(S_2(\delta),U) = \sqrt{2}\delta$. Thus if for any $\varepsilon > 0$, choose a δ such that $\sqrt{2}\delta \leq \varepsilon$, the Hausdorff distance between U and $S_1(\delta)$, U and $S_2(\delta)$ is

$$H(S_1(\delta),U) = \sqrt{2}\delta = H(S_2(\delta),U).$$

Similiaryly, the Hausdorff distance of U and $S_2(\delta)$ is at most $H(S_2(\delta),U) = \sqrt{2}\delta$.

Lemma 1. *For every $\varepsilon > 0$ and $x > 0$, there exists an ordered weighted tree T_ε and regular hexagon h of side length x as an ordered disk contact graph such that:*

1. *Every realization r of T_ε as an ordered disk contact graph where the radii of the disks equal the vertex weights, approximates the hexagon in the sense that:*

$$H\left(r\left(T_\varepsilon\right), h\right) = \varepsilon$$

2. *The number of nodes in T_ε and the weights are polynomial in ε and x.*

Lemma 1 can be generalized in three different ways: (1) if all weights are equal, (2) order does not matter, and (3) if we relax the regular hexagon to any arbitrary polygon. If all weights are equal, then every realization of the ordered weighted tree T_ε in Lemma 1.... If order is removed from the weighted tree T_ε, then ... If the regular hexagon is relaxed to an arbitrary polygon, then ..

1.2 Weighted Trees T_k

In this section we describe a particular family of unit weight trees and corresponding contact graphs disk arrangements called *snowflakes*. Note that we regard snowflakes with unit weight as a weight of r. For $i \in \mathbb{N}$, the construction of the snowflake tree, T_i, is as follows:

- Let v_0 be a vertex that has six paths attached to it: $p_1, p_2, \ldots, p_6$. Each path has i vertices.
- In botany, the stalk that attaches to a stem of a plant is called a *petiole*; petioles usually have leaves attached to their ends. We will now attach paths (petioles) onto every other path p_1, p_3, and p_5:
  - Every third vertex on that path has two petioles attached, one petiole on each side of p_k.
  - The number of vertices that lie on a petiole attached to the j^{th} vertex of p_k is $i - j$.
  - The first vertex of the $i - j$ vertices has one petiole attached; the remaining $i - j - 1$ vertices contain two paths. Each of these paths contain only one vertex. These paths are called *leaves*.

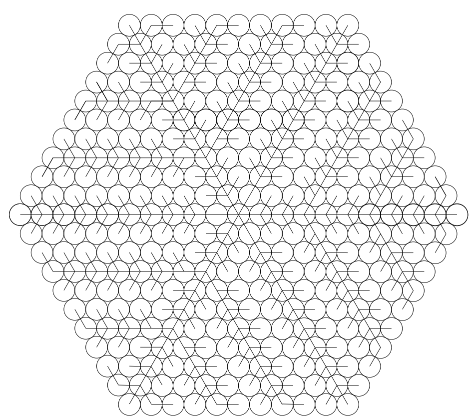

Figure 1.2: A contact graph that resembles the shape of concentric hexagons.

A *perfectly weighted snowflake tree* is a snowflake tree with all vertices having weight r. A *perturbed snowflake tree* is a snowflake tree with all vertices having weight of 1 with the exception of v_0; in a perturbed snowflake tree, v_0 will have a weight of $r + \zeta(\varepsilon)$. The value of $\zeta(\varepsilon)$ will be determined later on in the proof of the Lemma 1. For our analysis, all realizations of any snowflake, perfect or perturbed, shall have the disk corresponding to v_0 is centered at the origin. We can assume one of the paths is on the x-axis.

Perfectly Weighted Snowflake Tree. Consider the graph of the triangular lattice with unit distant edges:

$$\begin{aligned} V &= \left\{ a\cdot(1,0)+b\cdot\left(\frac{1}{2},\frac{\sqrt{3}}{2}\right) : a,b\in\mathbb{Z}\right\} \\ E &= \{\{u,v\} : ||u-v|| = 1 \text{ and } u,v\in V\} \end{aligned}$$

The following graph, $G=(V,E)$ is said to be the *unit distance graph* of the triangular lattice. We can show that no two distinct edges of this graph are non-crossing. First suppose that there were two distinct edges that crossed, $\{u_1,v_1\}$ and $\{u_2,v_2\}$. With respect to u_1, there are 6 possible edges corresponding to it, with each edge $\frac{\pi}{3}$ radians away from the next. No two edges cross.

The perfectly weighted snowflake tree that is a subgraph over the *unit distance graph*, $G=(V,E)$, of the triangular lattice. For the remainder of the thesis, a *snowflake*, S_i is a realization of a weighted tree T_i. To show this, for any S_i, fix $v_0 = 0\cdot(1,0)+0\cdot\left(\frac{1}{2},\frac{\sqrt{3}}{2}\right) = (0,0)\in V$ at origin. Next consider the six paths attached from origin. Fix each consecutive path $\frac{\pi}{3}$ radians away from the next such that the following points like on the corresponding paths: $(1,0)\in p_1, \left(\frac{1}{2},\frac{\sqrt{2}}{3}\right)\in p_2, \left(-\frac{1}{2}\mathsf{p}4,\frac{\sqrt{3}}{2}\right)\in p_3, (-1,0)\in p4, \left(-\frac{1}{2},-\frac{\sqrt{3}}{2}\right)\in p_5, \left(\frac{1}{2},-\frac{\sqrt{3}}{2}\right)\in p_6$. For S_i, there are i vertices on each path.

We define the six paths from origin as follows:

$$\begin{aligned} p_1 &= \{a\cdot(1,0)=\vec{v}\,|a=1,2,\ldots,i\,\} \\ p_2 &= \left\{ a\cdot\left(\frac{1}{2},\frac{\sqrt{3}}{2}\right)=\vec{v}\,|a=1,2,\ldots,i \right\} \\ p_3 &= \left\{ -a\cdot(1,0)+a\cdot\left(\frac{1}{2},\frac{\sqrt{3}}{2}\right) = a\left(-\frac{1}{2},\frac{\sqrt{3}}{2}\right)=\vec{v}\,|a=1,2,\ldots,i \right\} \\ p_4 &= \{a\cdot(-1,0)=\vec{v}\,|a=1,2,\ldots,i\,\} \\ p_5 &= \left\{ a\cdot\left(-\frac{1}{2},-\frac{\sqrt{3}}{2}\right)=\vec{v}\,|a=1,2,\ldots,i \right\} \\ p_6 &= \left\{ a\cdot(1,0)-a\cdot\left(\frac{1}{2},\frac{\sqrt{3}}{2}\right) = a\cdot\left(\frac{1}{2},-\frac{\sqrt{3}}{2}\right)|a=1,2,\ldots,i \right\} \end{aligned}$$

For S_i there exists i vertices on each path. We shall denote the i^{th} vertex on the j^{th} path as $v_{j,i}$. For each path defined above, the paths are defined as a set of vectors, $\vec{v}=a\cdot\vec{p}$ for some $a\in\mathbb{N}$ and $\vec{p}\in\mathbb{R}^2$. By setting $a=1,2,\ldots,i$, we obtain points that are contained in V. For $j=1$, 3, 5 and $\ell=3b\le i$ where $b\in\mathbb{N}$, there exists two paths attached to each vertex $v_{j,\ell}$. For S_i, each petiole attached to the ℓ^{th} vertex of p_j, there are $i-\ell$ vertices. For each vertex v on a petiole, which is not in the paths p_1, p_3, or p_5, there are two *leaves* on either side of the vertex; each leaf is a vertex that has an edge with v. The exceptions to the two leaves rule is on the first and last vertices of the petiole off of p_1, p_3, or p_5. In these exception, attach one leaf to the side of the vertex that is closest to center vertex v_0.

The triangular lattice is symmetric under rotation about v_0 by $\frac{\pi}{3}$ radians. For each vertex $v_{1,l}$ and $l=3b\le i$ where $b\in\mathbb{N}$, we place two petioles from it; the first petiole $\frac{\pi}{3}$ above p_1 at $v_{1,l}$ and $\frac{-\pi}{3}$ below p_1 at $v_{1,l}$ and call these petioles $p^+_{1,l}$ and $p^-_{1,l}$ respectively. With respect to $v_{1,l}$, one unit along $p^+_{1,l}$ is a point on the triangular lattice and similarly so on $p^-_{1,l}$. Continuing the walk along these paths, unit distance-by-unit distance, we obtain the next point on the triangular lattice up to $i-k$ distance away from $v_{1,l}$. Without loss of generality, for each vertex v of the petiole which are not in p_1 has two associated leaf nodes v^+ and v^-; v^+ is placed $\frac{\pi}{3}$ and one unit above v and v^- is placed $\frac{-\pi}{3}$ and one unit below v. Thus all leaf nodes are in the triangular lattice. This shows that each of the $i-k$ vertices on $p^-_{1,l}$, $p^+_{1,l}$, and leaves are in V. By rotating all of the paths along p_1 by $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$, we obtain the paths p_3 and p_5 respectively, completing the construction.

In Figure **??**, we have a set of unit radii disks arranged in a manner that outlines the perfectly weighted snowflake description above.

1.2.1 Perturbed Weighted Trees T_ε

Given $\varepsilon > 0$, we define T_ε as follows: the tree T_i with weight ε for every vertex except v_0; v_0 has a weight $\varepsilon + \zeta(\varepsilon)$ for some $\zeta(\varepsilon) > 0$ that is specified later. A perturbed weighted tree T_ε can be realized as a disk touching graph (a disk arrangement). A perturbed snowflake realization has some distinct qualities from perfect snowflake realizations. The angular relationships between adjacent vertices may vary, the distance between adjacent and neighboring vertices may vary as well.

Modification of S_1. We will show for any $\varepsilon > 0$ and arbitrary position of vertices, the placement of vertices is close to canonical position. In order to show this, we show the components of a perturbed snowflake in arbitrary position are close to canonical position. The argument comprises of three parts: (1) Showing that the pertubation of S_1 is small, (2) show that the displacement along the arms for all S_i for $i \geq 1$ is small, and (3) show that the displacement along the petioles is small.

Given a instance of a perturbed snowflake with v_0 having weight $\varepsilon + \zeta(\varepsilon)$ where $\varepsilon > 0$, vertices neighboring v_0 each have a range of placement on the plane when realizaed as a disk arrangement. Figure **??** shows a realization of S_1 and illustrates one such example of possible gaps, $\zeta(\varepsilon)$, that could be created between adjacent disks of S_1 in a perfect snowflake.

Displacement on S_1 is small. Note that (1) the adjacent disks in a perfect snowflake may or may not be adjacent in a given perturbed snowflake of S_1 and (2) $S_1 \subseteq S_i$ for any $i \in \mathbb{N}$. Given a snowflake in arbirary position with n unit segments per arm, the arms of the snowflake has a maximal length of n, end to end, if in canonical position; otherwise, the arm will have an end to end length less than n. Figure 1.3 shows an arm of a snowflake in arbitrary position corresponds to a compression and shift of vertices. The arm realized in arbitrary position in Figure 1.3 is analgous to a tree realized in arbitrary position where vertices are in a different position than canonical.

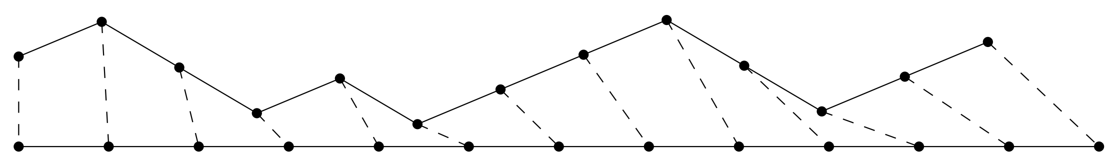

Figure 1.3: The polyline at the bottom represents a snowflake arm in canonical position. The polyline above represents a snowflake arm in non-canonical position.

Suppose we are given an ordered weighted tree T_ε such that the corresponding disk arrangement is a perturbed S_1. In a perfect snowflake of S_1 the six disks around the central disk kiss each other. The angle formed from the center of the central disk to the centers of any two adjacent disks is $\frac{\pi}{3}$. The side lengths of the equalateral triangle formed by the centers of three adjacent disks, one of which is the central disk, is $2r$. For a perturbed S_1 the the central disk is weighted $r + \zeta(\varepsilon)$. This can yield a change of angular displacement $\frac{\pi}{3}$ to $\frac{\pi}{3} - 2\chi$. To find the bounds of how large or small χ can be, we show the trigonometric relation of the half angle of the triangle corresponding to three adjacent disks (See Figure 1.4):

$$
\begin{aligned}
\sin\left(\frac{\pi}{6}-\chi\right) &= \frac{1}{2r+\zeta(\varepsilon)} \\
\frac{1}{2}\cos\chi = \sin\frac{\pi}{6}\cos\chi &= \frac{1}{2r+\zeta(\varepsilon)} + \cos\frac{\pi}{6}\sin\chi = \frac{1}{2r+\zeta(\varepsilon)} + \frac{\sqrt{3}}{2}\sin\chi \\
&\Longleftrightarrow \\
\frac{1}{2r+\zeta(\varepsilon)} + \frac{\sqrt{3}}{2}\left(\chi - \frac{\chi^3}{6}\right) &\leq \frac{1}{2r+\zeta(\varepsilon)} + \frac{\sqrt{3}}{2}\sin\chi = \frac{1}{2}\cos\chi \\
&\Longleftrightarrow \\
\frac{\sqrt{3}}{2}\left(\chi - \frac{\chi^3}{6}\right) &\leq \frac{1}{2} - \frac{1}{2r+\zeta(\varepsilon)} \qquad \text{if } \chi < 1 \\
\frac{5\sqrt{3}}{12}\chi &\leq \frac{2r+\zeta(\varepsilon)-2}{2(2+\zeta(\varepsilon)} \\
& \text{for } r = 1 \\
\chi &\leq \frac{3\zeta(\varepsilon)}{5\sqrt{3}} = \frac{12}{5\sqrt{3}}\frac{\zeta(\varepsilon)}{4}
\end{aligned}
$$

The coordinates of the centers of the disks of the construction are close to cannonical position following the angular displacement argument above. That is if u is a center of a disk in a disk arrangement corresponding to T_ε, u lies in a ball $b_{\chi(\varepsilon)}(u_c)$ where u_c is the cannonical position of the u.

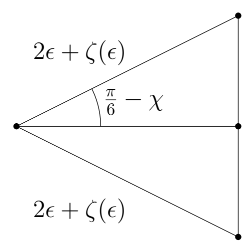

Figure 1.4: This figure depicts a triangle corresponding to the center of the central disk and two adjacent disks.

For any $\varepsilon > 0$, the bounds for angular displacement formed at the center of the central disk and two adjacent disks is:

$$\frac{\pi}{3} - \frac{6\zeta(\varepsilon)}{5\sqrt{3}} \leq \frac{\pi}{3} - 2\chi = 2\chi_{\min} \leq 2\chi \leq 2\chi_{\max} = \frac{\pi}{3} + 2\chi \leq \frac{\pi}{3} + \frac{6\zeta(\varepsilon)}{5\sqrt{3}}$$

Displacement on the arms is small. To show that the angluar displacement along the arm is small, we extend the angular argument on the perturbed S_1 and by induction, show that it is small for all i.

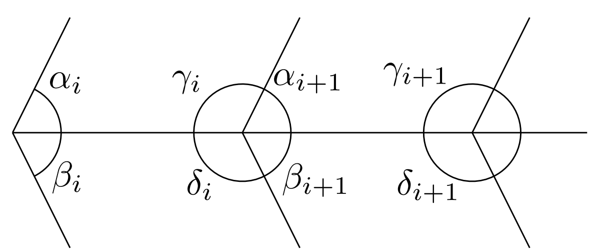

Figure 1.5: An arm depicted at the i^{th} and $(i+1)^{\text{st}}$ vertex.

Denote the angles on the concave side of the i^{th} vertex as α_i and β_i and the convex side of the $(i+1)^{\text{st}}$ vertex as γ_i and δ_i respectively (see Figure 1.5 for reference).

For any vertex, the sum of angles about the vertex is 2π, e.g.:

$$\gamma_i + \delta_i + \alpha_{i+1} + \beta_{i+1} = 2\pi$$

Suppose we numbered the the disks about the central disk 1 through 6. Without loss of generality, the angles α_0 and β_0 correspond to the angles formed between the central angle, disks i and $i+1$ and disks $i+1$ and $i+2$ respectively, for $i = 1,2,3$. The bounds for α_0 and β_0 are the same as 2χ in the earlier argument, i.e.:

$$\begin{array}{rcccl} \frac{\pi}{3} - \frac{6\zeta(\varepsilon)}{5\sqrt{3}} & \leq & \alpha_0 & \leq & \frac{\pi}{3} + \frac{6\zeta(\varepsilon)}{5\sqrt{3}} \\ \frac{\pi}{3} - \frac{6\zeta(\varepsilon)}{5\sqrt{3}} & \leq & \beta_0 & \leq & \frac{\pi}{3} + \frac{6\zeta(\varepsilon)}{5\sqrt{3}} \end{array}$$

We know that $\alpha_0 + \beta_0 \leq \frac{2\pi}{3} + \frac{12\zeta(\varepsilon)}{5\sqrt{3}}$. We also know that in canonical position:

$$\begin{array}{rcl} \pi & = & \alpha_0 + \gamma_0 \\ \pi & = & \beta_0 + \delta_0 \end{array}$$

Together, we have the following result:

$$\begin{array}{rcl} 2\pi & = & \alpha_0 + \gamma_0 + \beta_0 + \delta_0 \\ 2\pi & = & \alpha_0 + \gamma_0 + (2\pi - \alpha_1 - \beta_1) \\ \alpha_1 + \beta_1 & = & \alpha_0 + \gamma_0 \\ & \leq & \dfrac{2\pi}{3} + \dfrac{12\zeta(\varepsilon)}{5\sqrt{3}} \end{array}$$

And so the error bounds on 2χ hold in general for α_i and β_i for all i.

$$\alpha_i + \beta_i \leq \frac{2\pi}{3} + \frac{12\zeta(\varepsilon)}{5\sqrt{3}}$$

Displacement on the petioles is small. Note that the petioles have the same geometric structure as the arms; the exception is the number of leaves on each side of the petioles. Since we've shown that the geometric shape in arbirary position is already close to canonical position for any $\varepsilon > 0$, the same argument applies here for the petioles.

We have shown the displacements of all components of the perturbed snowflake are small for any $\varepsilon > 0$. This shows that the structure has stability in preserving any information encoded with it.

Proof of Lemma 1

Proof. For any $\varepsilon > 0$, we construct an ordered weighted tree T_ε and regular hexagon of side length x.

In order to show that every realization r of T_ε as an ordered disk contact graph where the radii of the disks equal the vertex weights, approximates the hexagon such that the Hausdorff distance is ε, let the radii of all disks besides the perturbed center of corresponding to T_ε be $r = \varepsilon$. Let the corresponding disk arrangement for T_ε be S_j where

$$j = \frac{x}{2\varepsilon + 1}.$$

The Hausdorff distance between the regular hexagon that is the convex hull for the centers of the disks in the disk arrangement and the union of the disks themselves is

$$H(r(T_\varepsilon), h) = \left(\frac{2}{\sqrt{3}} - 1\right)\zeta(\varepsilon)$$

The number of disk in the contact graph corresponding is a polynomial $d(x, \varepsilon) \le \frac{4x}{\varepsilon}$.

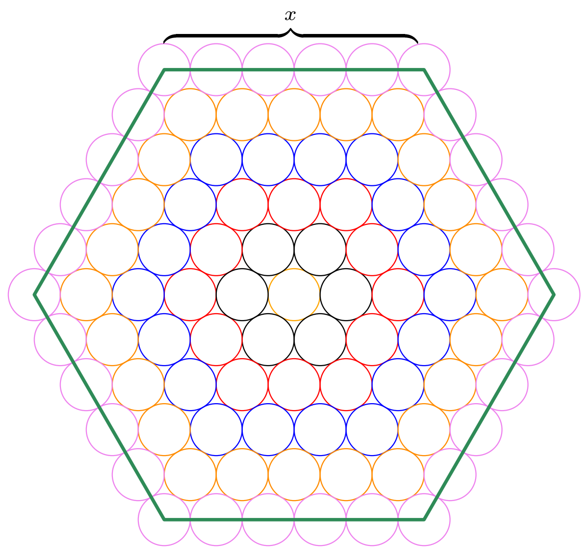

Figure 1.6: A regular hexagon of side length x as the convex hull of the centers of a disk arrangement in canonical position.

□

Proof of Theorem ??

Proof. Given an instance of a P3SAT boolean formula, we can use the snowflake reduction of the modified auxiliary construction. For any center of a disk in any realization the displacement of the center is in an open ball $b_{\zeta(\varepsilon)}(c)$ where c is the position of the center in canonical position. Using Lemma 1, we can approximate any hexagon with a tree T_ε. In the modified auxiliary construction in Chapter 3, we had four types of hexagons with different side lengths and the skinny rhombus. We can scale the weights (radii) of the corresponding ordered weighted disk contact graph corresponding to T_ε to the rigid frame, obstacle, flag, and half sized hexagons in a modified auxiliary contruction accordingly. The rhombus can be approximated by a chain of obstacle hexagons.

By approximating the polygons in the modified auxiliary construction with the snowflake, we show that Theorem **??** is a corollary by applying Lemma 1. □

Bibliography