SPECTRAL GEOMETRY, LINK COMPLEMENTS AND SURGERY DIAGRAMS

MARC LACKENBY*

1. INTRODUCTION

The spectrum of the Laplacian on a Riemannian manifold M has been the focus of an enormous amount of study. Of particular importance is $\lambda_1(M)$, which is the infimum of $\mathrm{Spectrum}(M) - \{0\}$. This contains a good deal of geometric information about M; in particular, it is closely related to a geometric quantity, the Cheeger constant $h(M)$ (see [8]). A collection $\{M_i\}$ of n-dimensional Riemannian manifolds (for some fixed n) is known as an *expanding family* if $\inf \lambda_1(M_i) > 0$. When the Ricci curvature of the manifolds M_i is bounded from below, this definition is equivalent to the condition that $\inf h(M_i) > 0$, by work of Cheeger [8] and Buser [7]. The construction of explicit expanding families of manifolds has been a major topic of research, with applications to such diverse fields as group theory [19], lattices in Lie groups [20], number theory [5] and coding theory [21]. More recently, it has become important to understand the behaviour of $\lambda_1(M)$ when M is a hyperbolic 3-manifold, because this has connections with the virtually Haken conjecture, which is a major unsolved problem in low-dimensional topology (see [15]). The Lubotzky-Sarnak conjecture [20] proposes that any closed hyperbolic manifold has a tower of finite covers which does not form an expanding family. This conjecture has major ramifications: it would imply, for example, that any arithmetic lattice in $\mathrm{PSL}(2, \mathbb{C})$ has a finite index subgroup with a non-abelian free quotient, by work of the author, Long and Reid [16].

It is the purpose of this paper to examine $h(M)$ and $\lambda_1(M)$ for hyperbolic 3-manifolds M. Our main theorem imposes upper bounds on $h(M)$ and $\lambda_1(M)$ in terms of data from any surgery diagram of M. Applying this to the case of trivial surgery, where no solid tori are attached, this gives upper bounds on $h(M)$ and $\lambda_1(M)$ when M is the complement of a hyperbolic link in the 3-sphere. There will be three main applications. Firstly, we will show that a sequence of hyperbolic alternating link complements with volumes tending to infinity cannot form an expanding family. Secondly, we apply this result to establish a finiteness theorem for alternating link complements that are congruence arithmetic 3-manifolds. Thirdly, we utilise the existence of expanding families of hyperbolic 3-manifolds to prove that some 3-manifolds must have 'complicated' surgery diagrams.

In order to state our main theorem, we need some terminology. Let D be a link diagram, and let $G(D)$ be the underlying 4-valent planar graph. A *bigon region* is the closure of a complementary region of $G(D)$ with two edges in its boundary. A *twist region* is either a maximal collection of bigon regions whose union is connected or a single crossing adjacent to no bigon regions. The crossings of a diagram are canonically partitioned into twist regions. The *twist number* $t(D)$ is the number of twist regions. Thus, $t(D)$ is always at most the crossing number $c(D)$ of the diagram.
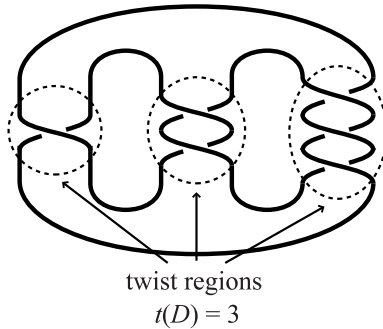

twist regions
$t(D) = 3$

Figure 1

Recall that a compact orientable 3-manifold M is obtained by *Dehn surgery* on a link L in the 3-sphere if there is a collection of properly embedded disjoint simple closed curves in M, such that removing an open regular neighbourhood of these curves results in the exterior of L. Equivalently, M can be constructed from the exterior of L by attaching solid tori, where the boundary of each solid torus is homeomorphically identified with a boundary component of the exterior of L. The essentially different ways of attaching each solid torus are parametrised by an element of $\mathbb{Q} \cup \infty$. A *rational surgery diagram* for M is a diagram for L, plus an assignment of an element of $\mathbb{Q} \cup \{\infty\}$ to some components of L. We permit some of the components of L to be unlabelled, in which case no solid torus is attached to the relevant component of $\partial N(L)$.

We say that a compact 3-manifold is *hyperbolic* if its interior admits a complete finite volume hyperbolic structure. We say that a link L in S^3 is *hyperbolic* if its exterior is hyperbolic.

Our main theorem is as follows.

Theorem 1.1. *Let D be a rational surgery diagram of a compact orientable hyperbolic 3-manifold M. Then the Cheeger constant $h(M)$ satisfies*

$$h(M) \leq 4\pi \frac{(24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)}}{\text{Volume}(M)} \leq 4\pi \frac{(24\sqrt{2} + 16\sqrt{3})\sqrt{c(D)}}{\text{Volume}(M)}.$$

By applying this to the case where no component of the link is filled in, we obtain the following corollary.

Corollary 1.2. *Let D be a diagram of a hyperbolic link L in the 3-sphere. Then*

$$h(S^3 - L) \leq 4\pi \frac{(24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)}}{\text{Volume}(S^3 - L)} \leq 4\pi \frac{(24\sqrt{2} + 16\sqrt{3})\sqrt{c(D)}}{\text{Volume}(S^3 - L)}.$$

We now give some applications of this corollary. If one wants to use this result to find upper bounds on $h(S^3 - L)$ from the diagram D, one needs to find lower bounds on the volume of $S^3 - L$. Fortunately, such bounds are known for two large classes of links: alternating links and highly twisted links.

We start with alternating links. If a link has an alternating diagram, then it has one where the underlying 4-valent graph has no edge loops. This is because any such loop may be removed to produce an alternating diagram of the same link with fewer crossings. A diagram is *twist-reduced* if, for every simple closed curve that meets the link projection transversely in four points away from the crossings, with two points of intersection adjacent to one crossing and the other two points of intersection adjacent to another crossing, the simple closed curve bounds a (possibly empty) collection of bigons arranged end to end between the crossings (see Figure 2). If an alternating diagram is not twist-reduced, then we may produce an alternating diagram of the same link with the same number of crossings but smaller twist number.
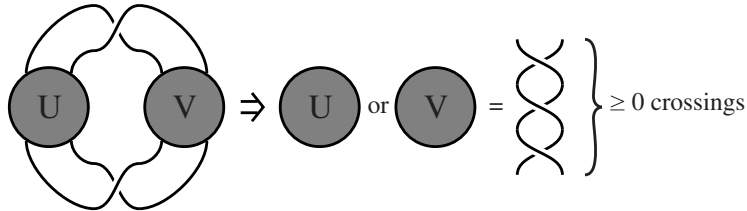

Figure 2

The main theorem of [13] is as follows.

Theorem 1.3. *Let L be a hyperbolic link with a twist-reduced alternating diagram D having no edge loops. Then*

$$v_3(t(D) - 2) \leq \text{Volume}(S^3 - L) < 16v_3(t(D) - 1),$$

where $v_3 \simeq 1.0149$ is the volume of a regular hyperbolic ideal tetrahedron.

In fact, the constants in this theorem have been improved upon. The upper bound on the volume was reduced to $10v_3(t(D) - 1)$ by Agol and D. Thurston [13]. The lower

bound on volume was increased to $v_8(t(D)/2 - 1)$ by Agol, Storm and W. Thurston [2], where $v_8 \simeq 3.6639$ is the volume of a regular hyperbolic ideal octahedron. It is these improved bounds that we use.

This result has been extended to another class of links by Futer, Kalfagianni and Purcell [11]. A diagram is known as *highly twisted* if each twist region contains at least 7 crossings, and the diagram is alternating within each twist region. We say that a link is *highly twisted* if it has a twist-reduced highly twisted diagram having no edge loops. Their result is as follows.

Theorem 1.4. *Let L be a hyperbolic link with a twist-reduced highly twisted diagram D having no edge loops. Then*

$$0.70735\,(t(D) - 1) < \text{Volume}(S^3 - L) < 10v_3(t(D) - 1).$$

By combining Corollary 1.2, Theorem 1.3 and Theorem 1.4, we obtain the following consequence for the geometry of these link complements.

Corollary 1.5. *Let L be a hyperbolic link with an alternating or highly twisted diagram D that is twist-reduced, has no edge loops and where $t(D) > 2$. Then*

$$h(S^3 - L) \leq c_1/\sqrt{t(D)}$$
$$h(S^3 - L) \leq c_2/\sqrt{\text{Volume}(S^3 - L)}$$

where $c_1 \leq 1643$ and $c_2 \leq 1129$.

Buser's inequality [7] gives that $\lambda_1(S^3 - L) \leq 4h(S^3 - L) + 10(h(S^3 - L))^2$, and hence we have the following spectral consequence.

Corollary 1.6. *Let L be a hyperbolic link with an alternating or highly twisted diagram D that is twist-reduced, has no edge loops and where $t(D) > 2$. Then*

$$\lambda_1(S^3 - L) \leq c_3/\sqrt{t(D)} + c_4/t(D)$$
$$\lambda_1(S^3 - L) \leq c_5/\sqrt{\text{Volume}(S^3 - L)} + c_6/\text{Volume}(S^3 - L),$$

where $c_3 \leq 6572$, $c_4 \leq 2.7 \times 10^7$, $c_5 \leq 4516$ and $c_6 \leq 1.3 \times 10^7$.

This implies one direction of the following.

Theorem 1.7. *A collection of alternating or highly twisted hyperbolic link complements forms an expanding family if and only if their volumes are bounded.*

For the other direction, suppose that an infinite collection of alternating or highly twisted hyperbolic link complements have bounded volume, but do not form an expanding family. Then, we may pass to a subsequence where the smallest positive eigenvalue of

the Laplacian tends to zero. We may pass to a further subsequence M_i which converges in the Gromov-Hausdorff topology to a fixed finite-volume hyperbolic 3-manifold M_∞ (see Chapter E in [3] for example). By [9], the eigenvalues $\lambda_1(M_i)$ tend to $\lambda_1(M_\infty)$, and hence are bounded away from zero, which is a contradiction.

This result can be interpreted in several different ways. On the one hand, it may mean that expanding families of hyperbolic 3-manifolds are 'rare'. If so, this would provide support for the Lubotzky-Sarnak conjecture. On the other hand, it may mean that alternating and highly twisted link complements are not representative of 'generic' hyperbolic 3-manifolds.

Given the above non-expansion results for two large classes of knot and link complements, it is natural to speculate about all hyperbolic knot and link complements. This leads to the following interesting question:

Question. Does there exist a collection of hyperbolic link complements with volumes that tend to infinity and which forms an expanding family?

A consequence of Theorem 1.7 is the following finiteness result. We refer the reader to [1] for the definition of a *congruence* arithmetic 3-manifold. According to Theorem 5.9 of [1] (see also Corollary 1.3(a) of [6]), any such 3-manifold M has $\lambda_1(M) \geq 3/4$. Also, by [4], there are only finitely many arithmetic 3-manifolds with volume less than any given real number. Thus, we obtain the following corollary.

Corollary 1.8. *There are only finitely many alternating or highly twisted link complements that are congruence arithmetic hyperbolic 3-manifolds.*

This leads to the following interesting question.

Question. Are there only finitely many alternating or highly twisted link complements that are arithmetic?

The methods of this paper cannot be immediately applied here, since there exist arithmetic 3-manifolds with arbitrarily small Cheeger constant. However, any arithmetic 3-manifold finitely covers a congruence arithmetic 3-manifold. So, this may perhaps give a route to tackling this question.

We now return to surgery diagrams for 3-manifolds. It is natural to ask how complex must a surgery diagram of a 3-manifold be. In this direction, Constantino and D. Thurston have the following theorem [10].

Theorem 1.9. *There is a positive constant c such that any closed hyperbolic 3-manifold M has a rational surgery diagram with at most $c(\text{Volume}(M))^2$ crossings.*

It is not hard to show that the number of crossings required in such a surgery diagram D is at least $\text{Volume}(M)/(10v_3)$. This is because the exterior of the link specified by D has Gromov norm at most $10c(D)$ by the appendix in [13]. Gromov norm does not increase under Dehn filling [23], and so the Gromov norm of M is at most $10c(D)$. But the Gromov norm of a hyperbolic 3-manifold M is equal to $\text{Volume}(M)/v_3$, by [23]. This establishes the required bound.

Given this linear lower bound on crossing number, it is natural to ask whether the quadratic upper bound of Constantino and Thurston is sharp. Here, we show that it is, by applying Theorem 1.1 to an expanding family of hyperbolic 3-manifolds.

Theorem 1.10. *Let $\{M_i\}$ be an expanding family of finite-volume orientable hyperbolic 3-manifolds. Then, there is a positive constant c such that any rational surgery diagram for M_i requires at least $c(\text{Volume}(M_i))^2$ crossings.*

Such expanding families of 3-manifolds are known to exist. Indeed, the following result of Long, Lubotzky and Reid [18] shows that they arise as covering spaces of any given finite-volume hyperbolic 3-manifold.

Theorem 1.11. *Any finite-volume hyperbolic 3-manifold has an infinite sequence of distinct finite-sheeted regular covers that forms an expanding family.*

The proof of our main result, Theorem 1.1, utilises the relationship between the following various notions of 'width':

1. The max-width of a link in the 3-sphere. This is closely related to a concept introduced by Gabai in his proof of the Property R conjecture [12]. We give its definition below.

2. The width of an abstract graph. (By 'abstract' here, we just mean that the graph is not necessarily embedded anywhere, and if it is, then this embedding is immaterial.) This notion was first defined by the author in [14], where it was used to prove a group-theoretic result. We recall it in Section 2.

3. The Morse width of a planar graph. This is a new concept, which measures the interaction between a graph embedded in the 2-sphere and Morse functions for that sphere. We will define it in Section 3.

4. The Heegaard width of a compact orientable 3-manifold M. This measures the complexity of generalised Heegaard splittings for M. A very closely related concept was defined by the author in [15] (and was denoted $c_+(M)$ there). We will give its definition in Section 5.

Let us start with the definition of the max-width of a link. One considers diagrams D for the link in $\mathbb{R}^2$. The *max-width* of D is the maximum, over all $t \in \mathbb{R}$, of the number of intersections of $\mathbb{R} \times \{t\}$ with the link projection. The *max-width* of the link is defined to be the minimal max-width of any of its diagrams. This is a slight variant of Gabai's definition in [12]. He considered the sum of the number of intersections, over a finite collection of representative level sets $\mathbb{R} \times \{t\}$. We call this the *sum-width* of the diagram. (See Section 6 where the definition of sum-width is recalled in detail.) However, here, it is more appropriate to consider the maximum rather than the sum.

Let D be a rational surgery diagram for a compact orientable hyperbolic 3-manifold M, and let $G(D)$ be the underlying 4-valent planar graph. For convenience, we assume that D is a diagram in the 2-sphere rather than $\mathbb{R}^2$. The first step in the proof of Theorem 1.1 is to bound the width of $G(D)$ as an abstract graph. To do this, we use a famous result of Lipton and Tarjan [17] that provides an efficient 'separator' for a planar graph. This is a collection of vertices in the graph with relatively small size, which divides the graph into two subgraphs, each with relatively large size. Using an inductive argument involving these separators, we can find an upper bound on the width of $G(D)$ as an abstract graph. This is then used to construct a Morse function on the diagram 2-sphere, so that each level set of the function has relatively few intersections with the graph. This extends to a Morse function on the 3-sphere, in which level sets have relatively few intersections with the link. We arrange that this Morse function is equivalent to the standard Morse function on the 3-sphere, and so this provides an upper bound on the max-width of the link, as defined above. This leads to the following result, which is of independent interest.

Theorem 1.12. *Let D be a diagram of a link L in the 3-sphere. Then the max-width of L is at most*

$$2 + (24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)}.$$

Note that we are not claiming here that the max-width of D, or any diagram ambient isotopic to it, is bounded by the above function.

A theorem of Thompson [22] relates the width of a knot to its bridge number, provided that the knot is not tangle-composite. However, some care is required here, because Thompson used the sum-width of a knot (as defined by Gabai) rather than max-width. In Section 6, we show how these notions are related, and apply Theorem 1.12 to prove the following.

Theorem 1.13. *Let D be a diagram of a knot K in the 3-sphere that is not tangle-composite. Then the bridge number of K is at most*

$$1 + (12\sqrt{2} + 8\sqrt{3})\sqrt{t(D)}.$$

We now return to the proof of Theorem 1.1. By Theorem 1.12, the link specified by D has a diagram with max-width at most $2 + (24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)}$. Using this diagram, we construct a generalised Heegaard splitting of the link exterior, with control over the Euler characteristic of the splitting surfaces. This extends to a generalised Heegaard splitting of the filled-in 3-manifold M. Finally, we use a result of the author from [15] which places an upper bound on the Cheeger constant of a hyperbolic 3-manifold in terms of data from a generalised Heegaard splitting.

2. The width of planar graphs

In this and the following two sections, we give the proof of Theorem 1.12. The first step is as follows. Given a link diagram D, let $G(D)$ be its underlying planar graph. Let $T(D)$ be the corresponding *twist graph*, which has a vertex for each twist region of $G(D)$ and an edge for each edge of $G(D)$ not lying in a twist region. Thus, $T(D)$ is a planar graph which is obtained by collapsing $G(D)$. (See Figure 3.)
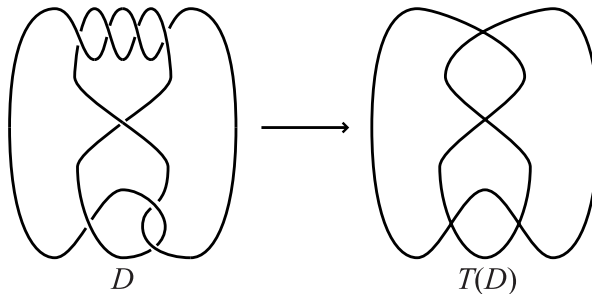

Figure 3

The width of a finite graph G was defined in [14]. We recall the definition here. Let $V(G)$ be the vertex set of G with cardinality $v(G)$. For any subset A of $V(G)$, let ∂A denote the set of edges with one endpoint in A and one endpoint not in A. We consider all possible bijections $\phi \colon [1, v(G)] \cap \mathbb{N} \to V(G)$. (Such a bijection is effectively just a total ordering on the vertices.) The *width* of ϕ is

$$\text{width}(\phi) = \max_{i \in [1, v(G)] \cap \mathbb{N}} |\partial(\phi([1, i] \cap \mathbb{N}))|.$$

The *width* of G, denoted width(G), is the minimum, over all such ϕ, of width(ϕ).

Although it is not strictly relevant here, there is an attractive pictorial interpretation of the width of a finite graph G (without loops and without isolated vertices). Given a bijection $\phi\colon [1, v(G)] \cap \mathbb{N} \to V(G)$, draw the graph in $\mathbb{R}^2$ so that the vertices have height given by ϕ^{-1}, and so that the height function on each edge has no critical points and so that the edges intersect transversely. Then, the width of ϕ is equal to the maximal number of intersections, over all $t \in \mathbb{R}$, between the graph and $\mathbb{R} \times \{t\}$. Hence, the width of G is the minimum of this maximum, over all such realisations of G in $\mathbb{R}^2$. There are obvious analogies with the max-width of a link defined in Section 1.
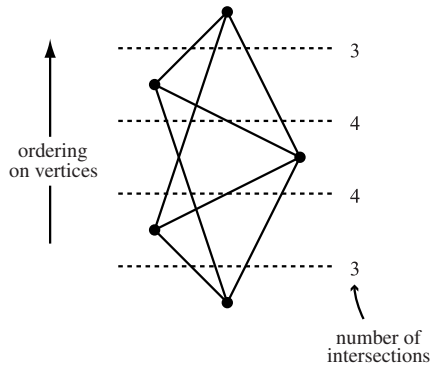

Figure 4

One reason for studying the width of a graph G is its relation to the graph's Cheeger constant. Recall [19] that this is defined to be

$$h(G) = \min\left\{\frac{|\partial A|}{|A|} : A \subset V(G), 0 < |A| \leq v(G)/2\right\}.$$

Setting i to be $\lfloor v(G)/2 \rfloor$ in the definition of graph width, we obtain the inequality

$$h(G) \leq \frac{\text{width}(G)}{\lfloor v(G)/2 \rfloor}.$$

The following result is central to this paper. It gives an upper bound on the width of a planar graph G, which grows as a function of $\sqrt{v(G)}$.

Theorem 2.1. *Let G be a finite planar graph, with $v(G)$ vertices and maximal vertex degree $\Delta(G)$. Then*

$$\text{width}(G) \leq (6\sqrt{2} + 4\sqrt{3})\Delta(G)\sqrt{v(G)}.$$

The proof relies heavily on the important work of Lipton and Tarjan on separators for planar graphs [17]. Recall that a set of vertices S in a graph G *separates* G into

disjoint graphs G_1 and G_2 if one obtains $G_1 \cup G_2$ by removing S and all adjacent edges from G, and if there is no edge joining G_1 to G_2. (We do not require either G_i to be connected.)

Theorem 2.2. (Lipton-Tarjan [17]) *Any finite planar graph has a set of vertices S which separates G into G_1 and G_2, such that*

$$v(G_1) \leq (2/3)v(G), \qquad v(G_2) \leq (2/3)v(G), \qquad |S| \leq \sqrt{8v(G)}.$$

Proof of Theorem 2.1. We prove this by induction on $v(G)$. For brevity, set $n = v(G)$. The induction starts trivially with $n = 1$. Let us therefore prove the inductive step. Let S be the set of vertices provided by Lipton and Tarjan's theorem, which separates G into G_1 and G_2. By induction, for $j = 1$ and 2, there is a bijection $\phi_j \colon [1, v(G_j)] \cap \mathbb{N} \to V(G_j)$ with width at most

$$(6\sqrt{2} + 4\sqrt{3})\Delta(G_j)\sqrt{v(G_j)} \leq 4(\sqrt{3} + \sqrt{2})\Delta(G)\sqrt{n}.$$

We use these to define a bijection $\phi \colon [1, n] \cap \mathbb{N} \to V(G)$, as follows. Send the first $|S|$ integers to S via some arbitrary bijection. Then send the next $v(G_1)$ vertices to $V(G_1)$ via ϕ_1. (In other words, define $\phi(i) = \phi_1(i - |S|)$ in this part of the domain.) Then send the final $v(G_2)$ vertices to $V(G_2)$ via ϕ_2. It is clear that

$$\text{width}(\phi) \leq \Delta(G)|S| + \max\{\text{width}(\phi_1), \text{width}(\phi_2)\}.$$

To see this, consider some integer i between 1 and $v(G)$, and let $A = \phi([1, i] \cap \mathbb{N})$. If $i \leq |S|$, then any edge in ∂A is adjacent to a vertex in S. There are at most $\Delta(G)|S|$ such edges. If i lies between $|S| + 1$ and $|S| + v(G_1)$, any edge in ∂A is either adjacent to S or runs between two vertices in G_1. In the latter case, this edge contributes to $\text{width}(\phi_1)$. A similar argument holds when i lies between $|S| + v(G_1) + 1$ and $v(G)$. Thus,

$$\text{width}(G) \leq \text{width}(\phi) \leq \Delta(G)\sqrt{8n} + 4(\sqrt{3} + \sqrt{2})\Delta(G)\sqrt{n} = (6\sqrt{2} + 4\sqrt{3})\Delta(G)\sqrt{n},$$

as required. □

We will apply this result to the twist graph $T(D)$ of a diagram D. The following can then be used to provide an upper bound on the width of $G(D)$.

Lemma 2.3. *Let D be a link diagram. Then* $\text{width}(G(D)) \leq \text{width}(T(D)) + 2$.

Proof. Start with a total ordering ϕ on the vertices of $T(D)$, with minimal width. Use this to construct a total ordering on the vertices of $G(D)$, as follows. The first vertex

10

of $T(D)$ corresponds to a twist region of $G(D)$. We set the vertices of this twist region to be the first vertices in the ordering, taken in order along the twist region. Then consider the second vertex of $T(D)$, and so on. At each stage, we need to choose which end of the relevant twist region that we start at, and we need to be a little careful how we make this choice. We ensure that we start at a vertex in the twist region that is adjacent to some earlier vertex in another twist region, if possible. Let ψ be the resulting bijection $[1, c(D)] \cap \mathbb{N} \to V(G(D))$. Then, for each positive integer i, $\psi([1, i] \cap \mathbb{N})$ consists of the vertices in the twist regions of $\phi([1, j] \cap \mathbb{N})$, for some positive integer j, plus a proper (possibly empty) subset of the vertices in a single twist region. Thus, the edges in $\partial(\psi([1, i] \cap \mathbb{N}))$ correspond to the edges in $\partial(\phi([1, j] \cap \mathbb{N}))$, except possibly for edges that lie in or are adjacent to the twist region. It is easy to see that the total number of edges in $\partial(\psi([1, i] \cap \mathbb{N}))$ is therefore at most

$$\max\{|\partial(\phi([1, j] \cap \mathbb{N}))| + 2, |\partial(\phi([1, j+1] \cap \mathbb{N}))| + 2\}.$$

Hence, the width of ψ is at most width$(\phi) + 2$, which equals width$(T(D)) + 2$. $\square$

3. The Morse width of planar graphs

Let G be a finite graph embedded in the 2-sphere, in which each vertex has valence at most 4. In this section, our goal is to construct a Morse function $f \colon S^2 \to \mathbb{R}$, starting from a total ordering of the vertices of G.

Let $f \colon S^2 \to \mathbb{R}$ be a Morse function. We say that f is *generic* with respect to G if its critical points have distinct values, and the vertices of G have distinct values. However, we allow the possibility that a critical point and a vertex take the same value; indeed, we allow vertices to be critical points of f. The *width* of G with respect to f is defined to be $\max\{|f^{-1}(t) \cap G| : t \in \mathbb{R}\}$. The *Morse width* of the embedded graph G is the minimal width of G with respect to f, over all generic Morse functions f.

Proposition 3.1. *Let G be a finite graph embedded in S^2, in which each vertex has valence at most 4. Then the Morse width of G is equal to its width as an abstract graph.*

Proof. We first show that the width of G is at most its Morse width. Consider a generic Morse function f on the 2-sphere. The vertices are ordered according to their values under f, and so this defines a bijection $\phi \colon [1, v(G)] \cap \mathbb{N} \to V(G)$. Let t_i be the value under f of the ith vertex. If $t \in (t_i, t_{i+1})$, then clearly, any edge in $\partial(\phi([1, i] \cap \mathbb{N}))$ must go through $f^{-1}(t)$. Hence, $|f^{-1}(t) \cap G|$ is at least $|\partial(\phi([1, i] \cap \mathbb{N}))|$. The width of ϕ is therefore at most the width of G with respect to f, which establishes the required inequality.

To prove the inequality in the other direction, suppose that we are given a bijection $\phi \colon [1, v(G)] \cap \mathbb{N} \to V(G)$. We will extend ϕ^{-1} to a Morse function f on S^2 with the same width. First define f to be monotonic on the edges. Since each vertex has valence at most 4, we may extend f to a regular neighbourhood of G. (See Figure 5 for some examples.) Then extend f over each complementary region of G to create a generic Morse function. Clearly, the width of G with respect to f is equal to the width of ϕ, as required. $\square$

$\underline{\hspace{2cm}}$ = G

$\underline{\hspace{2cm}}$ = level set of f

$\longrightarrow\!\!\!\!\!\underline{\hspace{1cm}}$ = direction of increasing f
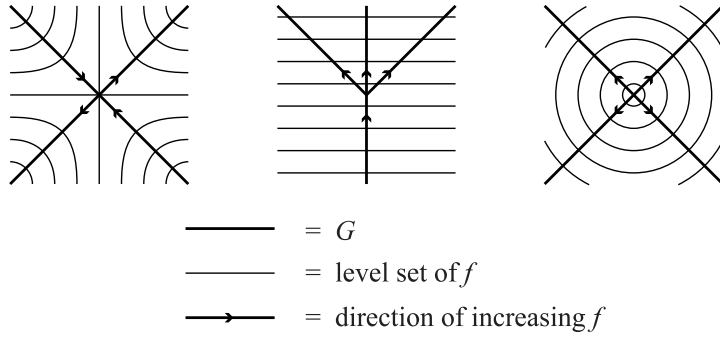
Figure 5

An example of an embedded planar graph G is given in Figure 6. There, a Morse function on S^2 is constructed from a bijection $[1, v(G)] \cap \mathbb{N} \to V(G)$, using the recipe in the above proof.

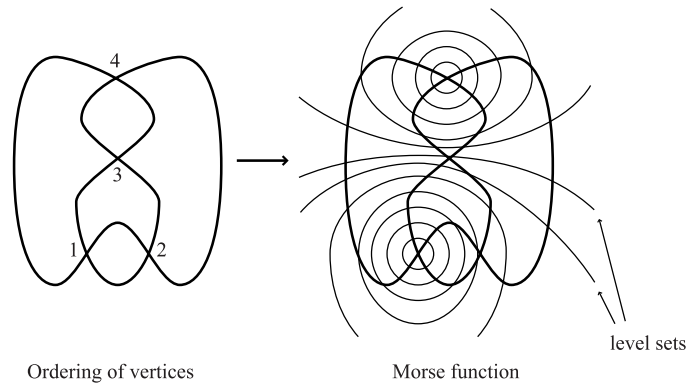Ordering of vertices   Morse function   level sets

Figure 6

4. The max-width of links

In Section 1, we gave the definition of the max-width of a link L in the 3-sphere. We now give an interpretation of max-width which is slightly more topological. Let $f \colon S^3 \to \mathbb{R}$ be the standard Morse function with a single maximum, a single minimum

12

and no other critical points. Let L be a link embedded in S^3 so that $f|L$ is a Morse function. The *max-width* of this embedding is $\max\{|f^{-1}(t) \cap L| : t \in \mathbb{R}\}$. It is clear that the minimum max-width over all links ambient isotopic to L is equal to the max-width of L, as defined in Section 1.

We now give a bound on the max-width of L in terms of Morse width, as defined in Section 3.

Proposition 4.1. *Let D be a diagram of a link L. Then the max-width of L is at most the Morse width of the embedded graph $G(D)$.*

Proof. Let $f\colon S^2 \to \mathbb{R}$ be a Morse function that is generic with respect to $G(D)$ and which realises the Morse width of $G(D)$. From this, we will construct a Morse function $F\colon S^3 \to \mathbb{R}$, equivalent to the standard Morse function, such that the max-width of L with respect to F is equal to the width of $G(D)$ with respect to f. Some care is required here, because f may have many critical points, whereas we require F to have just two.

Let us first rescale f so that its image lies in $(-1, 1)$.

Let $S^2 \times [-3, 3]$ be a regular neighbourhood of the equatorial sphere in S^3. We may pick an embedding of L in $S^2 \times [-1, 1]$ as specified by the diagram D. More specifically, we may arrange that the projection of L onto the first factor S^2 of the product is equal to the link projection in D, and that the behaviour of L near the inverse image of the double points coincides with the crossing information of D.

We will now specify a foliation of $S^2 \times [-3, 3]$ by 2-spheres. The leaves will be indexed by $t \in [-1, 1]$, and will be denoted by S_t^2. This foliation will be equivalent to the product foliation, and so the leaves S_t^2 can be viewed as the level sets of a smooth function $F\colon S^2 \times [-3, 3] \to [-1, 1]$ without critical points. We can then extend this to a Morse function on S^3 that is equivalent to the standard Morse function.

For values of t away from small neighbourhoods of the critical values, we define S_t^2 initially to be

$$(f^{-1}[t, 1] \times \{t - 2\}) \cup (f^{-1}(t) \times [t - 2, t + 2]) \cup (f^{-1}[-1, t] \times \{t + 2\}).$$

See Figure 7 for an example of such an S_t^2. It is clear that for regular values of t, S_t^2 is a 2-sphere. For it is obtained by cutting S^2 along $f^{-1}(t)$ to give $f^{-1}[-1, t]$ and $f^{-1}[t, 1]$ and then inserting annuli between the corresponding boundary components of these surfaces. It is also straightforward to check that these spheres are pairwise disjoint.
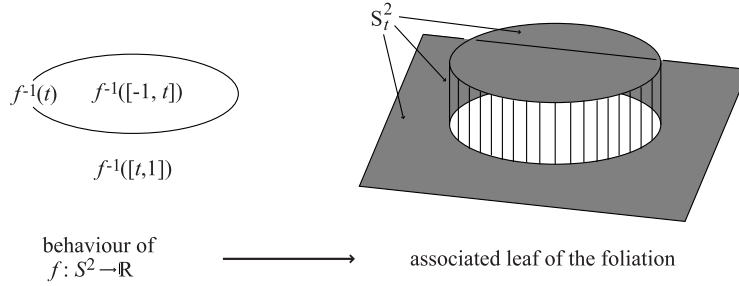
13

Figure 7

Near the critical values of f, we need to modify this definition slightly, in order that S_t^2 is genuinely a sphere. (For example, if a local maximum of f has value t, then S_t^2 as defined above would be the union of a sphere and an interval.) However, this modification can clearly be done. We also need to modify the entire foliation so that it is smooth. Again, this is clearly possible.

An example is given in Figure 8. In the first diagram, the level sets of the Morse function are shown. In the second diagram, a cross-section of the preliminary 'foliation' is depicted. In the third diagram, the modification that makes the foliation smooth is shown.
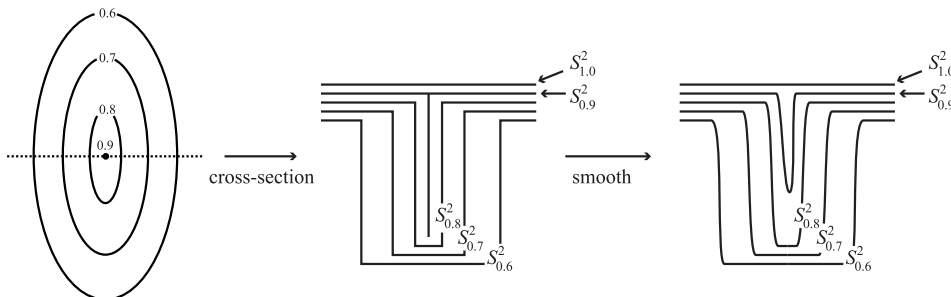

Figure 8

For any value of t in $[-1, 1]$, the number of intersections between S_t^2 and L is equal to the number of intersections between $f^{-1}(t)$ and $G(D)$. Hence, the max-width of L with respect to F is indeed equal to the width of $G(D)$ with respect to f, as required. $\square$

We can now complete the proof of Theorem 1.12.

Proof. Let D be a diagram of a link L in the 3-sphere. By Proposition 4.1, the max-width of L is at most the Morse width of $G(D)$. By Proposition 3.1, this is equal to the width of $G(D)$ as an abstract graph. By Lemma 2.3, this is at most width$(T(D)) + 2$. This is at most $(24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)} + 2$, by Theorem 2.1. $\square$

14

5. The Heegaard width of 3-manifolds

Let M be a compact orientable 3-manifold. In this section, we will study an invariant of M known as its *Heegaard width*. This is defined in terms of the complexity of surfaces in generalised Heegaard splittings of M.

Recall that, for a compact connected surface S, $\chi_-(S)$ is $\max\{0, -\chi(S)\}$. When S is a compact, possibly disconnected surface, $\chi_-(S)$ is defined to be the sum of $\chi_-(S_i)$, as S_i ranges over all the components of S.

Let $\{C_1, \ldots, C_n\}$ be a generalised Heegaard splitting for M. Thus, each C_i is a disjoint union of compression bodies; M is union of these compression bodies; their interiors are disjoint; their boundary components are either properly embedded or components of ∂M; and, for each relevant odd integer i, $\partial_+ C_i = \partial_+ C_{i+1}$ and $\partial_- C_i \cap \text{int}(M) = \partial_- C_{i-1} \cap \text{int}(M)$. We define the *width* of the splitting to be $\max_i \chi_-(\partial_+ C_i)$. The *Heegaard width* of M is the minimal width of any generalised Heegaard splitting. We denote it by Heeg-width(M).

In this section, we will prove the following.

Proposition 5.1. *Let L be a link in the 3-sphere. Then the Heegaard width of the exterior of L is at most* max-width$(L) - 2$.

Proof. Let $f: S^3 \to \mathbb{R}$ be the standard Morse function, and place L so that it has minimal max-width with respect to f. We may assume that $f|L$ is a Morse function and that the critical points of $f|L$ occur at distinct heights. Let $S_1, \ldots, S_n$ be a minimal collection of level sets of f, with increasing values under f, so that each S_i avoids the critical points of $f|L$ and so that between S_i and S_{i+1} only local minima or only local maxima of $f|L$ appear. In other words, if we view the critical points of $f|L$ as a sequence of local minima, followed by a sequence of local maxima, followed by a sequence of local minima, and so on, then the surfaces S_i occur between the local minima and the local maxima, and between the local maxima and the local minima. From the surfaces S_i, we construct closed surfaces F_i in the exterior of L, as follows.

Around L, place n parallel tori $T_1, \ldots, T_n$, where T_1 is closest to $\partial N(L)$, T_2 is adjacent to T_1, and so on. For each i, the intersection $T_i \cap S_i$ is a collection of simple closed curves, which bound discs in S_i. Remove the interiors of these discs, and attach the parts of T_i that lie above S_i. Let F_i be the resulting surface. (See Figure 9.)

We claim that the surfaces $F_1, \ldots, F_n$ separate the exterior of L into compression bodies $C_1, \ldots, C_{n+1}$. To see this, consider the submanifold C_{i+1} between F_i and F_{i+1}. This is a modification of the region between S_i and S_{i+1}. Suppose first that this region

15

only contains local maxima of L and that $i \neq n$. For each local maximum of L between S_i and S_{i+1}, there is a disc D such that $L \cap D$ is an arc in ∂D running from S_i up to the local maximum of L and back down to S_i, and such that the remainder of ∂D is an arc in S_i. We may arrange that these discs are all disjoint and that the intersection between each disc and C_{i+1} is a compression disc for F_i. If we cut C_{i+1} along these discs, the resulting manifold is homeomorphic to $F_{i+1} \times I$ (see Figure 10).
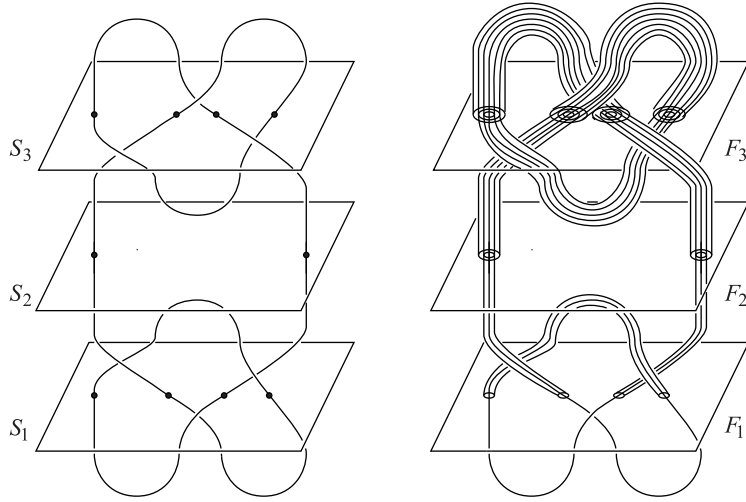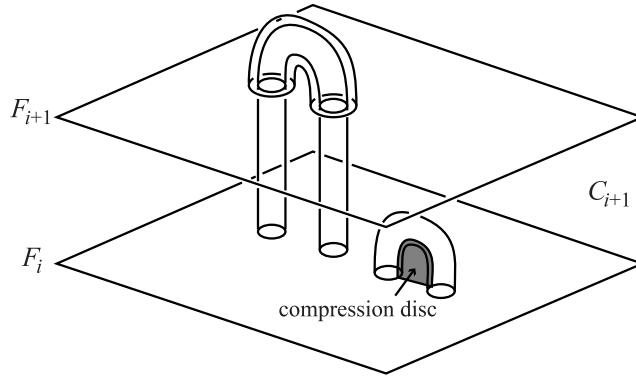

Figure 9

Figure 10

Suppose now that the region between S_i and S_{i+1} contains only local minima of L and that $i \neq 1$. As previously, for each local minimum, there is a disc D such that $L \cap D$ is an arc in ∂D running from S_{i+1} down to the local minimum of L and back up to S_{i+1}, and such that the remainder of ∂D is an arc in S_{i+1}. Let N be a big regular neighbourhood of D, and let D' be $\partial N \cap C_{i+1}$. Then, provided N was big enough, D' is a disc properly embedded in C_{i+1} with boundary in F_{i+1} (see Figure 11). We may

16

arrange that these discs D' are all disjoint. If we cut C_{i+1} along these discs, the result is a collection of copies of $T^2 \times I$ and a copy of $F_i \times I$. Thus, C_{i+1} is a compression body. Similarly, C_1 and C_{n+1} are compression bodies.

These compression bodies form a generalised Heegaard splitting for the exterior of L. Its width is $\max_i\{|S_i \cap L| - 2\}$, which equals max-width$(L) - 2$. $\square$
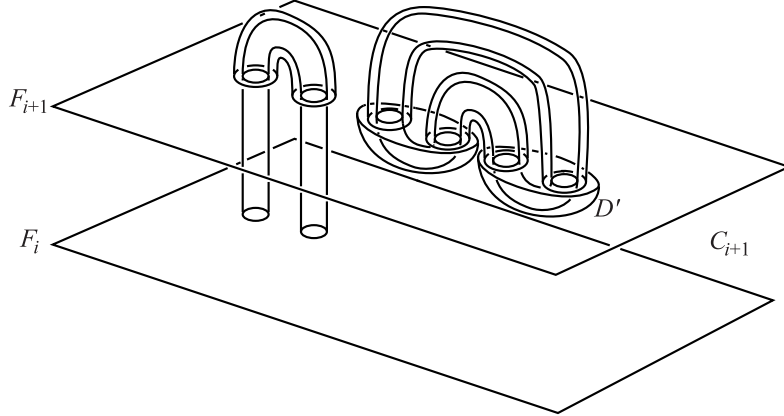

Figure 11

The key relationship between the Heegaard width of a finite-volume hyperbolic manifold and its Cheeger constant is contained in the following result. The following is a slight variant of Theorem 4.1 in [15], and it has essentially the same proof, which we will not repeat here.

Theorem 5.2. *Let M be a compact orientable finite-volume hyperbolic 3-manifold. Then*

$$h(M) \leq 4\pi \frac{\text{Heeg-width}(M)}{\text{Volume}(M)}.$$

Proof of Theorem 1.1. Let D be a rational surgery diagram of a compact orientable hyperbolic 3-manifold M. Let L be the link defined by D. By Proposition 5.1, the Heegaard width of the exterior of L is at most max-width$(L) - 2$. By Theorem 1.12, this at most $(24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)}$. Now, the Heegaard width of a compact orientable 3-manifold does not increase under Dehn filling, since a generalised Heegaard splitting for the unfilled manifold becomes a generalised Heegaard splitting for the filled-in one. Hence,

$$\text{Heeg-width}(M) \leq (24\sqrt{2} + 16\sqrt{3})\sqrt{t(D)}.$$

Applying Theorem 5.2 gives the required upper bound on $h(M)$. $\square$

17

6. Other notions of link width

In this section, our goal is to prove Theorem 1.13, which places an upper bound on the bridge number of any tangle-prime knot, in terms of the twist number of any of its diagrams. In order to do this, we need to investigate yet more notions of the width of a link. We have already defined the max-width of a link in Section 1. We now give some variants of this.

Sum-width

This is the original version of width which Gabai introduced in his proof of the Property R conjecture. Let D be a diagram for a link L in $\mathbb{R}^2$. Projection onto the second co-ordinate of $\mathbb{R}^2$ gives a height function. Let us suppose that the restriction to the link of this height function is Morse, and the critical points of the link have distinct values. These values divide $\mathbb{R}$ into open intervals, and for each value of t in one of these intervals, the number of intersection between the link projection and $\mathbb{R} \times \{t\}$ is constant. Let us define the *sum-width* of D to be the sum of the number of intersection points, over each of these intervals. The *sum-width* of L is the minimal sum-width of any diagram for L.

Lex-width

Although sum-width is very useful and elegant, one loses a certain amount of information when performing the summation. Instead, one can combine the intersection numbers of the intervals into a multi-set (that is, a set where repetitions are retained), called the *lex-width* of D. One compares two multi-sets using the usual lexicographical ordering. Here, one re-orders each multi-set into a descending sequence. One compares the first two integers of each multi-set. If they are the same, one passes to the second two, and so on. The minimal lex-width (with respect to this ordering) over all diagrams D for L is known as the *lex-width* of L. Unlike sum-width and max-width, this is not a single integer, but is a multi-set of integers. The following is trivial.

Lemma 6.1. *The largest integer in the multi-set* lex-width(L) *is* max-width(L).

Proof. Let D be a diagram for L with minimal lex-width. The max-width of D is the maximum integer in the multi-set lex-width(D), which equals lex-width(L). Thus, the max-width of L is at most this maximum.

Conversely, let D' be a diagram for L with minimal max-width. The lex-width of D' is an upper bound for the lex-width of L. But when comparing multi-sets, one compares their largest integers first. So, the maximal integer in lex-width(L) is at most the max-width of D', which equals the max-width of L. $\square$

A diagram for a link L is a *bridge diagram* if projection onto the second factor of $\mathbb{R}^2$ restricts to a Morse function on L, in which all the local maxima occur above all the local minima. The *bridge number* $b(L)$ is the minimal number of local maxima in any bridge diagram. Hence, we trivially have that max-width$(L) \leq 2b(L)$. We will now see that this is, in fact, an equality for tangle-prime knots.

It is a theorem of Thompson [22] that, when K is a tangle-prime knot, then every diagram for K of minimal sum-width is a bridge diagram. However, exactly the same argument gives the following related result.

Theorem 6.2. *If a knot K is tangle-prime, then every diagram for K of minimal lex-width is a bridge diagram.*

Corollary 6.3. *If K is a tangle-prime knot, then* max-width$(K) = 2b(K)$.

Proof. The largest integer in the multi-set lex-width(K) is max-width(K), by Lemma 6.1. But, in a bridge diagram, the maximal number of intersections between $\mathbb{R} \times \{t\}$ and the link projection is at least $2b(K)$. Thus, by Theorem 6.2, max-width$(K) \geq 2b(K)$. Since the opposite inequality always holds, this must be an equality. $\square$

We can now prove Theorem 1.13.

Proof. Let D be a diagram of a knot K in the 3-sphere that is not tangle-composite. By Corollary 6.3, the bridge number of K is equal to half the max-width of K. By Theorem 1.12, this is at most $1 + (12\sqrt{2} + 8\sqrt{3})\sqrt{t(D)}$. $\square$

REFERENCES

1. I. AGOL, M. BELOLIPETSKY, P. STORM, K. WHYTE, *Finiteness of arithmetic hyperbolic reflection groups*, arXiv:math.GT/0612132

2. I. AGOL, P. STORM, W. THURSTON, *Lower bounds on volumes of hyperbolic Haken 3-manifolds*, J. Amer. Math. Soc. 20 (2007) 1053–1077.

3. R. BENEDETTI, C. PETRONIO, *Lectures on hyperbolic geometry.* Universitext. Springer-Verlag, Berlin, 1992

4. A. BOREL, *Commensurability classes and volumes of hyperbolic 3-manifolds*, Ann. Scuola Norm. Sup. Pisa 8 (1981) 1–33.

5. J. BOURGAIN, A. GAMBURD, P. SARNAK, *Sieving and expanders.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 3, 155–159.

6. M. BURGER, P. SARNAK, *Ramanujan duals II*, Invent. Math. 106 (1991) 1-11.

7. P. Buser, *A note on the isoperimetric constant*, Ann. Sci. Ecole Norm. Sup. 15 (1982) 213–230.

8. J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in analysis (Papers dedicated to Salomon Bochner, 1969), pp 195–199, Princeton University Press.

9. B. Colbois, G. Courtois, *Convergence de variétés et convergence du spectre du Laplacien*, Ann. Sci. Ecole Norm. Sup. 24 (1991) 507–518.

10. F. Constantino, D. Thurston, *3-manifolds efficiently bound 4-manifolds*, J. Topology 1 (2008) 703–745.

11. D. Futer, E. Kalfagianni, J. Purcell, *Dehn filling, volume and the Jones polynomial*, J. Differential Geom. 78 (2008) 429–464.

12. D. Gabai, *Foliations and the topology of 3-manifolds, III*, J. Differential Geom. 26 (1987) 479–536.

13. M. Lackenby, *The volume of hyperbolic alternating link complements*, Proc. London Math. Soc. 88 (2004) 204–224 (with an appendix by I. Agol and D. Thurston)

14. M. Lackenby, *A characterisation of large finitely presented groups*, J. Algebra 287 (2005) 458–473.

15. M. Lackenby, *Heegaard splittings, the virtually Haken conjecture and Property (τ)*, Invent. Math. 164 (2006) 317–359.

16. M. Lackenby, D. Long, A. Reid, *Covering spaces of arithmetic 3-orbifolds*, Int. Math. Res. Not. (2008)

17. R. Lipton, R. Tarjan, *A separator theorem for planar graphs*, SIAM J. Applied Mathematics 36 (1979) 177–189.

18. D. D. Long, A. Lubotzky, A. W. Reid, *Heegaard genus and Property (τ) for hyperbolic 3-manifolds*, J. Topology 1 (2008) 152–158.

19. A. Lubotzky, *Discrete Groups, Expanding Graphs and Invariant Measures*, Progress in Math. 125 (1994)

20. A. Lubotzky, *Eigenvalues of the Laplacian, the first Betti number and the congruence subgroup problem*, Ann. Math. 144 (1996) 441–452.

21. M. Sipser, D. Spielman, *Expander codes*, IEEE Trans. on Information Theory, 42 (1996) 1710–1722.

22. A. THOMPSON, *Thin position and bridge number for knots in the 3-sphere.* Topology 36 (1997) 505–507.

23. W. THURSTON, *The geometry and topology of 3-manifolds,* Lecture notes (1980).

Mathematical Institute, University of Oxford,
24-29 St Giles', Oxford OX1 3LB, United Kingdom.