A note on Random Homomorphism from Arbitrary Graphs to $\mathbb{Z}$

Martin Loebl
Jaroslav Nešetřil*
Department of Applied Mathematics
and
Institute for Theoretical Computer Science (ITI)
Charles University
Malostranské nám.25, 11800 Praha 1
Czech Republic
Email: loebl/nesetril@kam.ms.mff.cuni.cz

Bruce Reed
C.N.R.S., Paris
and
Department of Computer Science
McGill University
Montréal, Canada
Email: breed@cs.mcgill.ca

January 29, 2003

*Partially supported by the Project LN00A056 of the Czech Ministry of Education and a NATO Collaborative Research Grant

Abstract

We discuss the space of mappings f from the vertices of a fixed graph G to $\mathbb{Z}$ which satisfy: $|f(u) - f(v)| \leq 1$ whenever $uv \in E(G)$. In particular we focus on the (random) range of such mappings.

1 Introduction

By a *Lipschitz mapping* of a (finite) graph G, we mean a function f from its vertices to $\mathbb{Z}$ such that if uv is an edge of G then $|f(u) - f(v)| \leq 1$. We shall be interested in the number of values that a typical Lipschitz mapping takes. We cannot pick a Lipschitz mapping uniformly at random, as there are infinitely many such objects. So, we restrict our attention to connected G, fix some v_0 in V and consider the space Ω_G consisting of all Lipschitz mappings of G taking v_0 to 0 under the uniform distribution. We let f_G be a random element chosen from this space. We note that the space of Lipschitz mappings where $v_0 = i$ is a translate of Ω_G so we are indeed considering a typical Lipschitz mapping (up to translation).

We are interested in the random variable $R_G = |\{x | \exists u \in V(G) \ s.t. \ x = f_G(u)\}|$. For a disconnected graph H we define R_H to be the sum over all components G of H of R_G. This definition captures the fact that if we take a typical Lipschitz mapping from H to $\mathbb{Z}$ then the images of its components are disjoint.

We propose:

Conjecture 1 *The maximum over all connected n vertex graphs G of $\mathbf{Ex}(R_G)$ is achieved by an n vertex path.*

We shall prove that this conjecture is true modulo a constant factor. We also present other partial results on the conjecture.

Our work was motivated by two papers, one written by Benjamini, Haggstromm, and Mossel [1], the other by Benjamini and Schechtman[2]. They considered mappings from connected bipartite graphs to $\mathbb{Z}$ which satisfy $uv \in E(G)$ implies $|f(u) - f(v)| = 1$, proposing conjectures and obtaining results similar to ours. Our definition of a Lipschitz mapping corresponds more closely to what is generally meant by the term Lipschitz (our maps are 1-Lipschitz), and has the advantage that it extends to arbitrary graphs.

Their definition is more closely related to various classical random walks. In the language of graph homomorphisms [5, 7], they study homomorphisms to a two-way infinite path and we study homomorphisms to the graph obtained from a two-way infinite path by adding a loop at each vertex. We call the mappings they study, *strongly Lipschitz*.

Let us add one more remark which puts this note in a broader combinatorial context. Homomorphisms can be viewed as a generalization of colorings (see [7] for a general algebraic framework). Particularly, this led to the study of complexity problems (see e.g. [5, 4]) and to a structural theory (see e.g. [8]). While this research deals mainly with optimization and extremal cases (such as chromatic number), the study of random homomorphisms has been initiated only recently. This has been motivated by statistical mechanics (where one studies random homomorphisms into small targets, see e.g. [3]) and by probability theory (as a generalization of random walks, see e.g.[1, 2, 6]). In fact, as argued in [1], the homomorphisms may serve as a typical example of a Lipschitz map. This view is also taken here.

2 The Results

Let G be a connected graph on n vertices. We show:

Theorem 2 *For some absolute constant C, $\mathbf{Ex}(R_G) \le C\sqrt{n}$.*

Theorem 3 *For some absolute constant C we have: for any u and v at distance d in G, $\mathbf{Ex}(|f_G(u) - f_G(v)|) \le C\sqrt{d}$.*

Theorem 4 *For any pair of vertices u and v in any n-vertex graph, $\mathbf{Ex}(|f_G(u) - f_G(v)|)$ is no more than the expected value obtained when u and v are the endpoints of an n vertex path.*

The proof of the first two theorems follows almost immediately from results in [1]. The last result is more delicate but also relies on a result from that paper, which we state here.

Definition A *Simple Random Walk (or SRW)* of length l consists of a sequence $t_1, ..., t_l$ of independent random variables each of which is 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$. The *displacement* of this walk is $|\sum_{i=1}^{l} t_i|$.

Theorem 2.1 of [1] *Let G be a bipartite connected graph and u and v two vertices of G at distance d. If we choose a uniform Strongly Lipschitz mapping conditioned on $f(u) = 0$ then the distribution of $|f(v)|$ is stochastically dominated by the displacement of a simple random walk of length d.*

3 A Reduction

Consider a Lipschitz mapping g defined on G and let $C(g)$ be the subgraph of G induced by those edges on which g is a constant. We note that the fact that g is a Lipschitz mapping implies that $G - E(C(g))$ is a bipartite graph, and furthermore the graph obtained by contracting each component of $C(g)$ to a vertex is bipartite.

We let $D_H(G)$ be the graph obtained by contracting each component of H to a vertex. We can restrict our attention to H for which $D_H(G)$ is bipartite as for any H not in this set there is no g in Ω_G for which $C(g) = H$. We want to think of $D_H(G)$ as a labelled graph, so we label the vertex of $D_H(G)$ corresponding to a component U of $C(G)$ with the set of vertices of G in U.

We note that there is a one to one correspondence between Lipschitz mappings g for which $C(g) = H$ and strongly Lipschitz mappings h from $D_H(G)$ to the integers. These are the functions that the papers [1, 2] alluded to earlier study. (The papers normalize by setting $g(v_0) = 0$ for some fixed v_0 just as we do) Applying Theorem 2.1 from [1], we obtain: for some absolute constant C, if we condition on $C(f_G) = H$ then for any u and v at distance d in G, the conditional expected value of $|f_G(u) - f_G(v)|$ is at most $C\sqrt{d}$. Summing over all choices for H yields Theorem 3. Similarly, applying Theorem 2 of [2], to the h above, we obtain: for some absolute constant C, if we condition on $C(G) = H$ then the conditional expected value of R_G is at most $C\sqrt{n}$. Summing over all choices for H yields Theorem 2.

4 The Key to the Proof of Theorem 4

In proving our first two theorems, we could afford to lose a constant factor and so could easily reduce our problem to one on strongly Lipschitz functions. In proving Theorem 4, we want once again to think of conditioning on $C(G) = H$ for various subgraphs H of G and then applying Theorem 2.1 of [1]. However, as we want to obtain an exact result, we will need to perform a much more careful analysis.

Simple calculations show that, assymptotically, the expected distance between the endpoints of a path P_n with n vertices under a random Lipschitz mapping h approaches the expected distance between the endpoints of a path of length $\frac{2n}{3}$ under a random strongly Lipschitz mapping (because for any edge xy of the path $h(x)$ is equally likely to be $h(y)-1, h(y)+1$, or $h(y)$, so we expect about a third of the edges to be in $C(f_{P_n})$).

Thus, if u and v are joined in G by a path P with $|V(P)| << \frac{2n}{3}$ then we can apply Theorem 2.1 of [1] directly to obtain Theorem 4. For example, if G is (vertex) 2-connected then u and v are joined in G by a path of length at most $\frac{n}{2}$ so if n is large we are done.

More generally, consider a shortest u, v path P and let n_P be the number of edges of P which are blocks of G (recall that an edge is a block of G if it does not belong to a circuit in G).

Now, an edge xy of P which is a block is like an edge of P_n in that $f_G(x) - f_G(y)$ is equally likely to be $1, 0$, or -1. Thus, each such edge can be coupled with an edge of P_n and we can then apply the argument of the last paragraph to the remaining $n - n_P$ uncoupled edges of P and P_n to obtain the desired result.

This is the main idea in the argument, In order to obtain an exact result which also holds for small values of n, we need to both complicate the proof slightly and carry out some tedious calculations. Specifically, we prove the following two results

Definition A 3-*variable* is one which is 0 with probability $\frac{1}{3}$ and 1 with probability $\frac{2}{3}$. We use σ_i to denote the (random) sum of i independent 3-variables.

Lemma 5 *For any two vertices u and v in an n vertex graph there exist integers s, t with $2s + t = n - 1$ and $s \neq 1$ such that the $\mathbf{Ex}(|f_G(u) - f_G(v)|)$ is at most the expected displacement of a random walk whose (random) length is $s + \sigma_t$.*

Lemma 6 *For any two integers s, t with $2s+t = n-1$ and $s \neq 1$ the expected displacement of a random walk whose (random) length is $s + \sigma_t$ is at most the expected value of $|f_{P_n}(x_1) - f_{P_n}(x_n)|$ for the endpoints x_1 and x_n of P_n.*

Combining these two lemmas, yields Theorem 4. We prove them in the next section.

5 The Details

We will need a simple result on the displacement of SRWs.

Definition We use e_i to denote the expected displacement of an SRW of length i. We use d_i to denote $e_{i+1} - e_i$.

Fact 7 *For $i \geq 0$, $d_{2i+1} = 0$ and $d_{2i} = \frac{\binom{2i}{i}}{2^{2i}}$.*

Proof It is easy to see that d_n is simply the probability that the displacement of an SRW of length n is 0. The result follows. $\square$

Corollary 8 *If $i < j$ then $e_i \leq e_j$.*

Recall that the length of a path is the number of edges it contains and let l_{uv} be the (random) length of a shortest path between u and v in $D_{C(f_G)}(G)$. We also need the following immediate corollary of Theorem 2.1 of [2]:

Theorem 9 *The expected value of $|f_G(u) - f_G(v)|$ is at most the expected displacement of a random walk whose length is l_{uv}.*

Proof of Lemma 5 Combining Theorem 9 with Corollary 8, we see that to prove Lemma 5, we need only verify the following claim:

Recall that one probability distribution p_1 on the non-negative integers *stochastically dominates* another p_2 if for every i,

$$\sum_{j \leq i} p_1(j) \geq \sum_{j \leq i} p_2(j).$$

Claim 10 *For any two vertices u and v in an n vertex graph there exist integers s, t with $2s + t = n - 1$ and $s \neq 1$ such that l_{uv} is stochastically dominated by the random variable which is s plus the sum of t independent 3-variables.*

Proof Let P be a shortest path of G between u and v. Note that each block which intersects P in an edge intersects P in a subpath. (Recall that a block of G is any inclusion maximal 2-connected subgraph of G.) Let

B_1,...,B_l be the set of such blocks, enumerated in the order that we encounter these subpaths when traversing P from u to v.

Set $x_0 = u, x_l = v$ and for i strictly between 0 and l, let $x_i = B_i \cap B_{i+1}$. Let l_i be the (random) length of a shortest path between x_{i-1} and x_i in $D_{C(f_B)}(B)$.

Now, suppose that x is a cutvertex of a graph H, and H_1 and H_2 are two subgraphs of H with union H and intersection x. Then, given any Lipschitz mapping f_1 on H_1 and any Lipschitz mapping f_2 on H_2 such that $f_2(x) = 0$, we obtain a Lipschitz mapping f on H by setting (i) $f(y) = f_1(y)$ for $y \in V(H_1)$ and (ii) $f(y) = f_1(x) + f_2(y)$= for $y \in V(G_2)$. Conversely, given any Lipschitz mapping f for H, restricting f to $V(H_1)$ clearly yields another Lipschitz mapping and setting $f_2(y) = f(y) - f(x)$ yields a Lipschitz mapping on H_2 with $f_2(x) = 0$. Using this bijection, one can easily obtain:

Fact 11 *For any block B of G and subset H of $E(B)$,* $\mathbf{Pr}(H = C(f_G) \cap E(B)) = \mathbf{Pr}(H = C(f_B))$.

Using this fact, we see that l_{uv} is the sum of the l_i.=20 Furthermore, by the tree structure of the block graph, $n - 1 \leq \sum_{i=1}^{l}(|B_i| - 1)$. Thus, we need only prove our claim for 2-connected graphs, as for arbitrary graphs we can sum the result over the blocks $B_1, ..., B_l$. So, from now on, we assume G is 2-connected. Now, we may assume that G is not an edge, as otherwise the desired result is clearly true with $t = 1$ and $s = 0$. This implies that there are two paths from u to v in G and so P has length at most $\lfloor \frac{n}{2} \rfloor$. (As before n is the number of vertices of G.)

Case 1: $n = 2i + 1$ for $i \geq 2$.

Now, setting $s = \lfloor \frac{n}{2} \rfloor = i$, we see that $s \geq 2$ and $l_{uv} \leq |P| \leq s$. Thus setting $t = 0$ yields the desired result (Claim 10).

Case 2: $n = 2i + 2$ for $i \geq 2$.

Let $s = \lfloor \frac{n}{2} \rfloor - 1 = i$, and $t = 1$. Clearly $l_{uv} \leq |P| \leq s + 1$. So, to prove our Claim 10, we need only show that $\mathbf{Pr}(l_{uv} = s + 1) \leq \frac{2}{3}$.

If $|P|$ has length $\leq s$ then $\mathbf{Pr}(l_{uv} = s + 1)$ is zero, so we are done. Thus, there are two internally disjoint paths between u and v, each of length at least $s + 1$. Since $n = 2s + 2$, each of these must have length exactly $s + 1$, and their= union contains every vertex of G. We enumerate one of these paths as $y_0 = u, y_1, .., y_{s+1} = v$ and the other as $x_0 = u, ..., x_{s+1} = v$.

We let Ω' be the set of those Lipschitz mappings f on G satisfying $f(u) = 0$. We let Ω^* be the subset of Ω', for which the shortest path between u and v in $D_{C(f)}(G)$ has length $s + 1$. We need to show that $|\Omega^*| \leq \frac{2}{3}|\Omega'|$.

We say that a mapping f in Ω' is *level on the pair (j,j+1)*, if one of the following two situations occurs:

(a) $f(x_j) = f(y_j)$ and $f(x_{j+1}) = f(y_{j+1})$, or:

(b) $f(x_j) - f(x_{j+1}) = f(y_j) - f(y_{j+1})$, and neither $x_j y_{j+1}$ or $x_{j+1} y_j$ is an edge of G.

Now, if $f \in \Omega^*$ is level on $(j, j+1)$ then we can define a Lipschitz function f' which agrees with f on $x_0, ..., x_j, y_0, ..., y_j$ and such that $f'(z) = f(z) + (f(x_j) - f(x_{j+1}))$ elsewhere. Note that both $x_j x_{j+1}$ and $y_j y_{j+1}$ are edges of $C(f')$, so f' is in $\Omega' - \Omega^*$. We call such an (f, f') a *corresponding pair.* Note, that $C(f')$ and $C(f)$ agree except on the edges between sets $\{x_j, y_j\}$ and $\{x_{j+1}, y_{j+1}\}$, where they disagree. Thus, f' is in a corresponding pair with exactly one other element f^* of Ω^* which satisfies $f^*(x_j) - f^*(x_{j+1}) = f(x_{j+1}) - f(x_j)$. We note further that f and f' are level on the same pairs of consecutive vertices and that the number of corresponding pairs containing f is precisely the number of consecutive integers on which it is level.

We let Ω_k^* (resp. Ω_k') denote the subset of Ω^* (resp. Ω') consisting of those f which are level on exactly k pairs of consecutive integers. The above remarks imply that $2(\Omega_k' - \Omega_k^*) \geq k\Omega_k^*$. Thus, for k between 1 and s, $\Omega_k^* \leq \frac{2}{k+2}\Omega_k'(G) \leq \frac{2}{3}\Omega_k'$. So to complete the proof of this case, we need only show that $\Omega_0^* + \Omega_{s+1}^* \leq \frac{2}{3}(\Omega_0' + \Omega_{s+1}')$.

To do so we consider generating all possible Lipschitz functions starting from left to right. It is easy to verify that given the choices for $f(x_j)$ and $f(y_j)$, if f in Ω^* is level on $(j, j+1)$ then there are at most two choices for the pair $f(x_{j+1}), f(y_{j+1})$. Thus, $|\Omega_{s+1}^*| \leq 2^{s+1}$.

In the same vein, the reader may easily verify that given the choices for $f(x_j)$ and $f(y_j)$, if f in Ω^* is not level on $(j, j+1)$ then there are at most two chocies for the pair $f(x_{j+1}), f(y =_{j+1})$. Thus, $|\Omega_0^* \leq 2^{s+1}|$ and $\Omega_0^* + \Omega_{s+1}^* \leq 2^{s+2}$.

On the other hand, given any sequence of values $a_1, ..., a_{s+1}$ in $\{0, 1, -1\}$, we can generate a Lipschitz function by setting $f(x_0) = 0$ and then for j between 1 and $s + 1$, = setting $f(x_j) = f(y_j) = f(x_{j-1}) + a_j$. Thus, $|\Omega_{s+1}'| \geq 3^{s+1}$. So, since $s \geq 2$, $\Omega_0^* + \Omega_{s+1}^* \leq \frac{2}{3}(\Omega_0' + \Omega_{s+1}')$, as desired.

Case 3: n is less than 5. Set $s = 0$, and $t = n - 1$. We need to show that l_{uv} is stochastically dominated by the sum of t 3-variables. Since there are only four graphs to verify, each with at most four vertices, we leave it to the reader to check this by hand.

□

Proof Lemma 6

Given a labelling of the edges of P_n using $0, 1$, and -1, we can obtain a Lipschitz mapping f of P_n with $f(x_0) = 0$ by setting $f(x_i)$ to be the sum of the labels on the subpath between x_0 and x_i. Clearly, this yields a bijection between such edge labellings and Lipschitz mappings of P_n with $f(x_0) = 0$.

Since we can choose our edge labelling by first specifying which edges receive 0, we see that the expected value of $|f_{P_n}(x_0) - f_{P_n}(x_n)|$ is the expected displacement of a random walk whose length is the sum of $n - 1 = 2s + t$ independent 3-variables.

Now, we can couple the choice of t of these variables with the choice of the 3-variables for the walk of length $s + \sigma_t$ whose length we are trying to bound. So, to prove Lemma 6, we need only show:

Claim 12 *For any i, s with $s \neq 1$, the expected displacement of a SRW of length $i + \sigma_{2s}$ is at most the expected displacement of a SRW of length $i + s$.*

Proof We use B to denote the expected displacement of a random walk of length $i + \sigma_{2s}$. We note that there are 3^{2s} possible outcomes for our $2s$ 3-variables and that there are $\binom{2s}{j}2^j$ of these for which j of the 3-variables are non-zero. Thus,

$$B = \sum_{j=0}^{2s} e_{i+j}(\binom{2s}{j}2^j)/3^{2s}.$$

So, we need to show

$$\sum_{j=0}^{2s}(e_{j+i} - e_{s+i})(\binom{2s}{j}2^j) > 0.$$

In doing so, it helps to rearrange the sum slightly, pairing the terms for $s + a$ and $s - a$ for each a. I.e we show the following sum is positive:

$$\sum_{a=1}^{s}(e_{s+i+a} - e_{s+i})(\binom{2s}{s+a}2^{s+a}) + (e_{s+i-a} - e_{s+i})(\binom{2s}{s-a}2^{s-a}).$$

We shall prove that every term of this sum except the first is positive. We claim further that if $s+i$ is even then the first term is positive whilst if $s+i$ is odd then the sum of the first two terms is positive. This will yield the desired result.

To finish the proof of Claim 12 we will need the following immediate corollary of Fact 7:

Corollary 13 *For $k \geq 0$, $d_{2k+2} = (1 - \frac{1}{2k+2})d_{2k}$. Furthermore, $d_{2k+2j} \leq \frac{1}{2}(\frac{3}{4})^{j-1} d_{2k}$.*

Now,

$$(e_{s+i+a} - e_{s+i})(\binom{2s}{s+a}2^{s+a}) + (e_{s+i-a} - e_{s+i})(\binom{2s}{s-a}2^{s-a})$$
$$= \binom{2s}{s+a}2^{s-a}(2^{2a}(e_{s+i+a} - e_{s+i}) - (e_{s+i} - e_{s+i-a=})).$$

We can write $e_{s+i} - e_{s+i-a}$ as $\sum_{l=s+i-a}^{s+i-1} d_l$. and $e_{s+i+a} - e_{s+i}$ as $\sum_{l=s+i}^{s+i+a-1} d_l$. Note that all the terms in these two sums are non-negative. Now, the number of terms in each of these sums is the same, and since every other term is zero in both, the number of non zero terms in the two sums differ by at most 1.

If $a \geq 2$ each sum has at least one non-zero term , so the first sum has at most twice as many non-zero terms as the second. Applying, Corollary 13, we see that the smallest non-zero term in the second sum divided by the largest term in the first sum is at least $\frac{1}{2}(\frac{3}{4})^{a-1}$. Combining these two observations, we see that the second sum divided by the first sum is at least $\frac{1}{4}\frac{3}{4}^{a-1}$. Thus for $a \geq 2$, $2^{2a}(e_{s+i+a} - e_{s+i}) - (e_{s+i} - e_{s+i-a})$ is positive and so is

$$\binom{2s}{s+a}2^{s-a}(2^{2a}(e_{s+i+a} - e_{s+i}) - (e_{s+i} - e_{s+i-a}))$$

If $s+i$ is even then $e_{s+i} - e_{s+i-1}$ is 0, so the above equation is positive even when $a = 1$ and we are done.

If $s+i$ is odd then $e_{s+i+1} - e_{s+i}$ is 0 as is $e_{s+i-1} - e_{s+i-2}$, so

$$\sum_{a=1}^{2} \binom{2s}{s+a}2^{s-a}(2^{2a}(e_{s+i+a} - e_{s+i}) - (e_{s+i} - e_{s+i-a}))$$

$$= 2^{s-2}(16\binom{2s}{s+2}d_{s+i+1} - (\binom{2s}{s-2}d_{s+i-1} + 2\binom{2s}{s-1}d_{s+i-1}))$$

$$= 2^{s-2}\binom{2s}{s+2}(16d_{s+i+1} - (2\frac{s+2}{s-1} + 1)d_{s+i-1}).$$

Now, since $s \geq 2$, $\frac{s+2}{s-1} \leq 4$. Furthermore, $s+i$ is odd and at least three and hence by Corollary 13, $d_{s+i+1} \geq \frac{3}{4}d_{s+i-1}$. Thus, $(16d_{s+i+1} - (2\frac{s+2}{s-1}+1)d_{s+i-1})$ is positive and hence so is:

$$\sum_{a=1}^{2}\binom{2s}{s+a}2^{s-a}(2^{2a}(e_{s+a-i} - e_{s-i}) - (e_{s+=i} - e_{s+t-1})).$$

This completes the proof of our claim and the lemma.

□

References

[1] I. Benjamini, O. Haggstrom, and E. Mossel: *On random graph homomorphisms into* $\mathbb{Z}$, Journal of Combinatorial Theory(B) **78** (2000), 86-114.

[2] I. Benjamini and G. Schechtman:*Upper Bounds on The Height Differences of The Gaussian Random Field and the Range of Random Graph Homomorphisms into* $\mathbb{Z}$, Random Structures and Algorithms**17** (2000), 20-25.

[3] G. Brightwell, P. Winkler: Graph homomorphisms and phase transitions, J. Comb. Th. B

[4] M. Dyer, C. Greenhill: The complexity of counting graph homomorphisms Random Structures and Algorithms**17** (2000)

[5] P. Hell, J. Nešetřil: Complexity of H-coloring, J. Comb. Th B **48** (1990), 92-110.

[6] J. Kahn: Range of cube-indexed random walk (to appear)

[7] J. Nešetřil: Aspects of Structural Combinatorics (Graph homomorphisms and their use), Taiwanese J. Math. **3**, 4 (1999), 381-424.

[8] J. Nešetřil, C. Tardif: Duality Theorems for Finite Structures (Characterizing Gaps and Good Characterizations), J. Comb. Th.B **80** (2000), 80-97.