Approximability Distance in the Space of H-Colourability Problems

Tommy Färnqvist*, Peter Jonsson**, and Johan Thapper

Department of Computer and Information Science
Linköpings universitet
SE-581 83 Linköping, Sweden
{tomfa, petej, johth}@ida.liu.se

Abstract. A graph homomorphism is a vertex map which carries edges from a source graph to edges in a target graph. We study the approximability properties of the *Weighted Maximum H-Colourable Subgraph* problem (MAX H-COL). The instances of this problem are edge-weighted graphs G and the objective is to find a subgraph of G that has maximal total edge weight, under the condition that the subgraph has a homomorphism to H; note that for $H = K_k$ this problem is equivalent to MAX k-CUT. To this end, we introduce a metric structure on the space of graphs which allows us to extend previously known approximability results to larger classes of graphs. Specifically, the approximation algorithms for MAX CUT by Goemans and Williamson and MAX k-CUT by Frieze and Jerrum can be used to yield non-trivial approximation results for MAX H-COL. For a variety of graphs, we show near-optimality results under the Unique Games Conjecture. We also use our method for comparing the performance of Frieze & Jerrum's algorithm with Håstad's approximation algorithm for general MAX 2-CSP. This comparison is, in most cases, favourable to Frieze & Jerrum.
Keywords: optimisation, approximability, graph homomorphism, graph H-colouring, computational complexity

1 Introduction

Let G be a simple, undirected and finite graph. Given a subset $S \subseteq V(G)$, a *cut* in G with respect to S is the set of edges from vertices in S to vertices in $V(G) \setminus S$. The MAX CUT-problem asks for the size of a largest cut in G. More generally, a k-cut in G is the the set of edges going from S_i to S_j, $i \neq j$, where $S_1, \ldots, S_k$ is a partition of $V(G)$, and the MAX k-CUT-problem asks for the size of a largest k-cut. The problem is readily seen to be identical to finding a largest k-colourable subgraph of G. Furthermore, MAX k-CUT is known to be **APX**-complete for every $k \geq 2$ and consequently does not admit a polynomial-time approximation scheme (PTAS).

In the absence of a PTAS, it is interesting to determine the best possible approximation ratio c within which a problem can be approximated or alternatively, the smallest c for which it can be proved that no polynomial-time approximation algorithm exists

* Supported by the *National Graduate School in Computer Science* (CUGS, Sweden).
** Partially supported by the *Center for Industrial Information Technology* (CENIIT) under grant 04.01 and by the *Swedish Research Council* (VR) under grant 2006-4532.

(typically under some complexity-theoretic assumption such as **P** $\neq$ **NP**). Since the 1970s, the trivial approximation ratio 1/2 was the best known for MAX CUT. It was not until 1995 that Goemans and Williamson [16], using semidefinite programming (SDP), achieved a ratio of $.878567$. Until very recently no other method than SDP was known to yield a non-trivial approximation ratio for MAX CUT. Trevisan [34] broke this barrier by using algebraic graph theory techniques to reach an approximation guarantee of $.531$. Frieze and Jerrum [15] determined lower bounds on the approximation ratios for MAX k-CUT using SDP techniques. Sharpened results for small values of k have later been obtained by de Klerk et al. [9]. Under the assumption that the *Unique Games Conjecture* (UGC) holds, Khot et al. [24] showed the approximation ratio for $k = 2$ to be essentially optimal and also provided upper bounds on the approximation ratio for $k > 2$. Håstad [19] has shown that SDP is a universal tool for solving the general MAX 2-CSP problem over any domain, in the sense that it establishes non-trivial approximation results for all of those problems. Assuming UGC, Raghavendra's SDP algorithms have optimal performance for every MAX CSP [30], but the exact approximation ratios are not yet known. In fact, even though an algorithm (doubly exponential in the domain size) for computing these ratios for specific MAX CSP problems has emerged [31], this should be contrasted to the infinite classes of graphs our method gives new bounds for.

Here, we study approximability properties of a generalised version of MAX k-CUT called MAX H-COL for undirected graphs H. This is a specialisation of the MAX CSP problem. Jonsson et al. [20] have shown that whenever H is loop-free, MAX H-COL does not admit a PTAS, and otherwise MAX H-COL is trivial. Langberg et al. [26] have studied the approximability of MAX H-COL when H is part of the input. We present approximability results for MAX H-COL where H is taken from different families of graphs. Many of these results turn out to be close to optimal under UGC. Our approach is based on analysing approximability algorithms applied to problems which they are not originally intended to solve. This vague idea will be clarified below.

Denote by $\mathcal{G}$ the set of all simple, undirected and finite graphs. A *graph homomorphism* from G to H is a vertex map which carries the edges in G to edges in H. The existence of such a map will be denoted by $G \to H$. If both $G \to H$ and $H \to G$, the graphs G and H are said to be *homomorphically equivalent* (denoted $G \equiv H$). For a graph $G \in \mathcal{G}$, let $\mathcal{W}(G)$ be the set of *weight functions* $w : E(G) \to \mathbb{Q}^+$ assigning weights to edges of G. For a $w \in \mathcal{W}(G)$, we let $\|w\| = \sum_{e \in E(G)} w(e)$ denote the total weight of G. Now, *Weighted Maximum H-Colourable Subgraph* (MAX H-COL) is the maximisation problem with

Instance: An edge-weighted graph (G, w), where $G \in \mathcal{G}$ and $w \in \mathcal{W}(G)$.
Solution: A subgraph G' of G such that $G' \to H$.
Measure: The weight of G' with respect to w.

We remark that we consider instances where the weight functions w are given explicitly. Given an edge-weighted graph (G, w), denote by $mc_H(G, w)$ the measure of the optimal solution to the problem MAX H-COL. Denote by $mc_k(G, w)$ the (weighted) size of a largest k-cut in (G, w). This notation is justified by the fact that $mc_k(G, w) = mc_{K_k}(G, w)$. In this sense, MAX H-COL generalises MAX k-CUT.

Let $\mathcal{G}_{\equiv}$ denote the set of equivalence classes of $\mathcal{G}$ under $\equiv$. The relation $\to$ is defined on $\mathcal{G}_{\equiv}$ in the obvious way and $(\mathcal{G}_{\equiv}, \to)$ is a lattice denoted by $\mathcal{C}_S$. For a more in-depth treatment of graph homomorphisms and the lattice $\mathcal{C}_S$, see [17]. In this paper, we endow $\mathcal{G}_\equiv$ with a metric d defined in the following way: for $M, N \in \mathcal{G}$, let

$$d(M,N) = 1 - \inf_{\substack{G \in \mathcal{G} \\ w \in \mathcal{W}(G)}} \frac{mc_M(G,w)}{mc_N(G,w)} \cdot \inf_{\substack{G \in \mathcal{G} \\ w \in \mathcal{W}(G)}} \frac{mc_N(G,w)}{mc_M(G,w)}. \tag{1}$$

We will show that d satisfies the following property: if $M, N \in \mathcal{G}$ and mc_M can be approximated within α, then mc_N can be approximated within $\alpha \cdot (1 - d(M,N))$ and conversely, if it is **NP**-hard to approximate mc_N within β, then mc_M is not approximable within $\beta / (1 - d(M,N))$ unless **P** $=$ **NP**. Hence, we can use d for extending known (in)approximability bounds on MAX H-COL to new and larger classes of graphs. For instance, we can apply the algorithm of Goemans and Williamson (which is intended for solving MAX K_2-COL) to MAX C_{11}-COL (i.e. the cycle on 11 vertices) and analyse how well the problem is approximated (it will turn out that Goemans and Williamson's algorithm approximates MAX C_{11}-COL within 0.79869). Furthermore, we present a linear program for $d(M,N)$ and show that the computation of $d(M,N)$ can be drastically simplified whenever M or N is edge-transitive.

The metric d is related to a well-studied graph parameter known as *bipartite density* $b(H)$ [1, 3, 6, 18, 27]: if H' is a bipartite subgraph of H with maximum number of edges, then $b(H) = \frac{e(H')}{e(H)}$, where $e(G)$ is the number of edges in a graph G. Lemma 5 shows that $b(H) = 1 - d(K_2, H)$ for edge-transitive graphs H. We note that while d is invariant under homomorphic equivalence, this is not true for bipartite density. There is also a close connection to work by Šámal on *cubical colourings* [32, 33]. In fact, it turns out that for a graph H, the cubical colouring number $\chi_q(H) = 1/(1 - d(K_2, H))$.

The paper comprises two main sections. Section 2 is used for proving the basic properties of d, showing that it is well-defined on $\mathcal{G}_\equiv$, and that it is a metric. After that, we describe how to construct the linear program for d. In section 3, we use d for studying the approximability of MAX H-COL and investigate optimality issues, for several classes of graphs. This is done by exploiting inapproximability bounds that are consequences of the Unique Games Conjecture. Comparisons are also made to the bounds achieved by the general MAX 2-CSP-algorithm by Håstad [19]. Our investigation covers a spectrum of graphs, ranging from graphs with few edges and/or containing long shortest cycles to dense graphs containing $\Theta(n^2)$ edges. The techniques used in this paper seem to generalise to larger sets of problems. This and other questions are discussed in Section 4. Due to space considerations, some proofs have been omitted.

2 Approximation via the Metric d

In this section we start out by proving basic properties of the metric d, that $(\mathcal{G}_\equiv, d)$ is a metric space, and that proximity of graphs M, N in this space lets us interrelate the approximability of MAX M-COL and MAX N-COL. Sections 2.2 and 2.3 are devoted to showing how to compute d.

2.1 The Space $(\mathcal{G}_\equiv, d)$

We begin by noting that $d(M,N) = 1 - s(N,M) \cdot s(M,N)$ if we define $s(M,N)$ (for $M, N \in \mathcal{G}$) as the infimum of $\frac{mc_M(G,w)}{mc_N(G,w)}$ over all $G \in \mathcal{G}$ and $w \in \mathcal{W}(G)$. We now see that the relation $mc_M(G, w) \geq s(M, N) \cdot mc_N(G, w)$ holds for all $G \in \mathcal{G}$ and $w \in \mathcal{W}(G)$. Using this observation, one can show that $s(M, N)$ and thereby $d(M, N)$ behaves well under graph homomorphisms and homomorphic equivalence.

Lemma 1. *Let $M, N \in \mathcal{G}$ and $M \rightarrow N$. Then, for every $G \in \mathcal{G}$ and every weight function $w \in \mathcal{W}(G)$, $mc_M(G, w) \leq mc_N(G, w)$.*

Corollary 2. *If M and N are homomorphically equivalent graphs, then $mc_M(G, w) = mc_N(G, w)$. Let $M_1 \equiv M_2$ and $N_1 \equiv N_2$ be two pairs of homomorphically equivalent graphs. Then, for $i, j, k, l \in \{1, 2\}$, $s(N_i, M_j) = s(N_k, M_l)$.*

Corollary 2 shows that s and d are well-defined as functions on the set $\mathcal{G}_{\equiv}$ and it is routine work to show that d is indeed a metric on this space.

We say that a maximisation problem Π can be approximated within $c < 1$ if there exists a randomised polynomial-time algorithm A such that $c \cdot Opt(x) \leq \mathbf{E}(A(x)) \leq Opt(x)$ for all instances x of Π. Proximity of graphs G and H in d allows us to determine bounds on the approximability of MAX H-COL from known bounds on the approximability of MAX G-COL:

Lemma 3. *Let M, N, K be graphs. If mc_M can be approximated within α, then mc_N can be approximated within $\alpha \cdot (1 - d(M, N))$. If it is* **NP***-hard to approximate mc_K within β, then mc_N is not approximable within $\beta / (1 - d(N, K))$ unless* **P** $=$ **NP**.

Proof. Let $A(G, w)$ be the measure of the solution returned by an algorithm which approximates mc_M within α. We know that for all $G \in \mathcal{G}$ and $w \in \mathcal{W}(G)$ we have the inequalities $mc_N(G, w) \geq s(N, M) \cdot mc_M(G, w)$ and $mc_M(G, w) \geq s(M, N) \cdot mc_N(G, w)$. As a consequence, $mc_N(G, w) \geq mc_M(G, w) \cdot s(N, M) \geq A(G, w) \cdot s(N, M) \geq mc_M(G, w) \cdot \alpha \cdot s(N, M) \geq mc_N(G, w) \cdot \alpha \cdot s(N, M) \cdot s(M, N) = mc_N(G, w) \cdot \alpha \cdot (1 - d(M, N))$. For the second part, assume to the contrary that there exists a polynomial-time algorithm B that approximates mc_N within $\beta/(1 - d(N, K))$. According to the first part mc_K can then be approximated within $(1 - d(N, K)) \cdot \beta/(1 - d(N, K)) = \beta$. This is a contradiction unless **P** $=$ **NP**. $\square$

2.2 Exploiting Symmetries

We will now consider general methods for computing s and d. In Lemma 4, we show that certain weight functions provide a lower bound on $mc_M(G, w)/mc_N(G, w)$, and in Lemma 5, we provide a simpler expression for $s(M, N)$ which depends directly on the automorphism group and thereby the symmetries of N. This expression becomes particularly simple when N is edge-transitive. An immediate consequence of this is that $s(K_2, H) = b(H)$ for edge-transitive graphs H.

We describe the solutions to MAX H-COL alternatively as follows: let G and $H \in \mathcal{G}$, and for any vertex map $f : V(G) \rightarrow V(H)$, let $f^{\#} : E(G) \rightarrow E(H)$ be the (partial) edge map induced by f. In this notation $h : V(G) \rightarrow V(H)$ is a graph homomorphism precisely when $(h^{\#})^{-1}(E(H)) = E(G)$ or, alternatively, when $h^{\#}$ is a total function. The set of solutions to an instance (G, w) of MAX H-COL can then be taken to be the set of vertex maps $f : V(G) \rightarrow V(H)$ with the measure $w(f) = \sum_{e \in (f^{\#})^{-1}(E(H))} w(e)$.

In the remaining part of this section, we will use this description of a solution. Let $f : V(G) \to V(H)$ be an optimal solution to the instance (G, w) of MAX H-COL. Define the weight $w_f \in \mathcal{W}(H)$ in the following way: for each $e \in E(H)$, let $w_f(e) = \sum_{e' \in (f^{\#})^{-1}(e)} \frac{w(e')}{mc_H(G,w)}$. The next result is now fairly obvious:

Lemma 4. *Let $M, N \in \mathcal{G}$ be two graphs. Then, for every $G \in \mathcal{G}$, every $w \in \mathcal{W}(G)$, and any optimal solution f to (G, w) of* MAX N-COL, $\frac{mc_M(G,w)}{mc_N(G,w)} \geq mc_M(N, w_f)$.

Let M and $N \in \mathcal{G}$ be graphs and let $A = \text{Aut}^*(N)$ be the (edge) automorphism group of N. We will let $\pi \in A$ act on $\{u, v\} \in E(N)$ by $\pi \cdot \{u, v\} = \{\pi(u), \pi(v)\}$. The graph N is edge-transitive if and only if A acts transitively on the edges of N. Let $\hat{\mathcal{W}}(N)$ be the set of weight functions $w \in \mathcal{W}(N)$ which satisfy $\|w\| = 1$ and for which $w(e) = w(\pi \cdot e)$ for all $e \in E(N)$ and $\pi \in \text{Aut}^*(N)$.

Lemma 5. *Let $M, N \in \mathcal{G}$. Then, $s(M, N) = \inf_{w \in \hat{\mathcal{W}}(N)} mc_M(N, w)$. In particular, when N is edge-transitive, $s(M, N) = mc_M(N, 1/e(N))$.*

Proof. Clearly, $s(M, N) \leq \inf_{w \in \hat{\mathcal{W}}(N)} \frac{mc_M(N,w)}{mc_N(N,w)} = \inf_{w \in \hat{\mathcal{W}}(N)} mc_M(N, w)$. For the first part of the lemma, it will be sufficient to prove that the following inequality holds for some $w' \in \hat{\mathcal{W}}$: $\alpha = \frac{mc_M(G,w)}{mc_N(G,w)} \geq mc_M(N, w')$. By taking the infimum over graphs G and weight functions $w \in \mathcal{W}(G)$ in the left-hand side of this inequality, we see that $s(M, N) \geq mc_M(N, w') \geq \inf_{w \in \hat{\mathcal{W}}(N)} mc_M(N, w)$.

Let $A = \text{Aut}^*(N)$ be the automorphism group of N. Let $\pi \in A$ be an arbitrary automorphism of N. If f is an optimal solution to (G, w) as an instance of MAX N-COL, then so is $f_\pi = \pi \circ f$. Let $w_\pi = w_{\pi \circ f}$. By Lemma 4, $\alpha \geq mc_M(N, w_\pi)$. Summing π in this inequality over A gives $|A| \cdot \alpha \geq \sum_{\pi \in A} mc_M(N, w_\pi) \geq mc_M(N, \sum_{\pi \in A} w_\pi)$ (the straightforward proof for the last inequality is omitted). The weight function $\sum_{\pi \in A} w_\pi$ can be determined as follows.

$$\sum_{\pi \in A} w_\pi(e) = \sum_{\pi \in A} \frac{\sum_{e' \in (f^{\#})^{-1}(\pi \cdot e)} w(e')}{mc_N(G, w)} = \frac{|A|}{|Ae|} \cdot \frac{\sum_{e' \in (f^{\#})^{-1}(Ae)} w(e')}{mc_N(G, w)},$$

where Ae denotes the orbit of e under A. Thus, $w' \sum_{\pi \in A} w_\pi / |A| \in \hat{\mathcal{W}}(N)$ and w' satisfies $\alpha \geq mc_M(N, w')$ so the first part follows.

For the second part, note that when the automorphism group A acts transitively on $E(N)$, there is only one orbit $Ae = E(N)$. Then, the weight function w' is given by

$$w'(e) = \frac{1}{e(N)} \cdot \frac{\sum_{e' \in (f^{\#})^{-1}(E(N))} w(e')}{mc_N(G, w)} = \frac{1}{e(N)} \cdot \frac{mc_N(G, w)}{mc_N(G, w)}.$$

□

2.3 Computing Distances

From Lemma 5 it follows that in order to determine $s(M, N)$, it is sufficient to minimise $mc_M(N, w)$ over $\hat{\mathcal{W}}(N)$. We will now use this observation to describe a linear program for computing $s(M, N)$. For $i \in \{1, \ldots, r\}$, let A_i be the orbits of $\text{Aut}^*(N)$ acting on $E(N)$. The measure of a solution f when $w \in \hat{\mathcal{W}}(N)$ is equal to $\sum_{i=1}^{r} w_i \cdot f_i$, where w_i is the weight of an edge in A_i and f_i is the number of edges in A_i which are mapped to an edge in M by f. Note that given a w, the measure of a solution f depends only on the vector $(f_1, \ldots, f_r) \in \mathbb{N}^r$. Therefore, take the solution space to be the set of such vectors: $F = \{ (f_1, \ldots, f_r) \mid f$ is a solution to (N, w) of MAX M-COL $\}$. Let the variables of the linear program be $w_1, \ldots, w_r$ and s, where w_i represents the weight of each element in the orbit A_i and s is an upper bound on the solutions.

$\min s$
$\sum_i f_i \cdot w_i \leq s$ for each $(f_1, \ldots, f_r) \in F$
$\sum_i |A_i| \cdot w_i = 1$ and $w_i, s \geq 0$

Given a solution w_i, s to this program, a weight function which minimises $mc_M(G, w)$ is given by $w(e) = w_i$ when $e \in A_i$. The measure of this solution is $s = s(M, N)$.

Example 6. The *wheel graph* on k vertices, W_k, is a graph that contains a cycle of length $k-1$ plus a vertex v not in the cycle such that v is connected to every other vertex. We call the edges of the $k-1$-cycle *outer edges* and the remaining $k-1$ edges *spokes*. It is easy to see that for odd k, the wheel graphs are homomorphically equivalent to K_3. We will now determine $s(K_3, W_n)$ for even $n \geq 6$ using the previously described construction of a linear program. Note that the group action of Aut$^*(W_n)$ on $E(W_n)$ has two orbits, one which consists of all outer edges and one which consists of all the spokes. If we remove one outer edge or one spoke from W_k, then the resulting graph can be mapped homomorphically onto K_3. Therefore, it suffices to choose $F = \{f, g\}$ with $f = (k-1, k-2)$ and $g = (k-2, k-1)$ since all other solutions will have a smaller measure than at least one of these. The program for W_k looks like this:

$\min s$
$(k-1) \cdot w_1 + (k-2) \cdot w_2 \leq s$
$(k-2) \cdot w_1 + (k-1) \cdot w_2 \leq s$
$(k-1) \cdot w_1 + (k-1) \cdot w_2 = 1$
$w_i, s \geq 0$

The solution is $w_1 = w_2 = 1/(2k-2)$ with $s(K_3, W_k) = s = (2k-3)/(2k-2)$.

In some cases, it may be hard to determine the distance between H and M or N. If we know that H is homomorphically sandwiched between M and N so that $M \rightarrow H \rightarrow N$, then we can provide an upper bound on the distance of H to M or N by using the distance between M and N. The following result can readily be proved from the definition of s:

Lemma 7. *Let $M \rightarrow H \rightarrow N$. Then, $s(M, H) \geq s(M, N)$ and $s(H, N) \geq s(M, N)$.*

3 Approximability of MAX H-COL

Let A be an approximation algorithm for MAX H-COL. Our method basically allows us to measure how well A performs on other problems MAX H'-COL. In this section, we will apply the method to various algorithms and various graphs. We do two things for each kind of graph under consideration: compare the performance of our method with that of some existing, leading, approximation algorithm and investigate how close to optimality we can get. Let $v(G), e(G)$ denote the number of vertices and edges in G, respectively. Our main algorithmic tools will be the following:

Theorem 8. *mc_2 can be approximated within $\alpha_{GW} \approx 0.878567$ [16] and mc_k can be approximated within $\alpha_k \sim 1 - \frac{1}{k} + \frac{2 \ln k}{k^2}$ [15]. Let H be a graph. There is an absolute constant $c > 0$ such that mc_H can be approximated within $1 - \frac{t(H)}{d^2} \cdot (1 - \frac{c}{d^2 \log d})$ where $d = v(H)$ and $t(H) = d^2 - 2 \cdot e(H)$ [19].*

Here, the relation $\sim$ indicates two expressions whose ratio tends to 1 as $k \to \infty$. We note that de Klerk et al. [9] have presented the sharpest known bounds on α_k for small values of k; for instance, $\alpha_3 \geq 0.836008$. We will compare the performance of Håstad's algorithm on MAX H-COL with the performance of the algorithms for mc_2 and mc_k in Theorem 8 analysed using Lemma 3 and estimates of the distance d. For this purpose, we introduce two functions, FJ_k and *Hå*, such that, if H is a graph, $FJ_k(H)$ denotes the best bound on the approximation guarantee when Frieze and Jerrum's algorithm for MAX k-CUT is applied to the problem mc_H, while *Hå*(H) is the guarantee when Håstad's algorithm is used to approximate mc_H. We note that the comparison is not entirely fair since Håstad's algorithm was probably not designed with the goal of providing optimal results—the goal was to beat random assignments. However, it is the currently best algorithm, with known bounds, that can approximate MAX H-COL for arbitrary $H \in \mathcal{G}$. This is in contrast with the algorithms of Raghavendra [30].

To be able to investigate the eventual near-optimality of our approximation method we will rely on the Unique Games Conjecture by Khot [23]. Thus, we assume henceforth that UGC is true, which gives us the following inapproximability results:

Theorem 9 (Khot et al. [24]). *For every $\varepsilon > 0$, it is NP-hard to approximate mc_2 within $\alpha_{GW} + \varepsilon$. It is NP-hard to approximate mc_k within $1 - \frac{1}{k} + \frac{2 \ln k}{k^2} + O(\frac{\ln \ln k}{k^2})$.*

3.1 Sparse Graphs

In this section, we investigate the performance of our method on graphs which have relatively few edges, and we see that the *girth* of the graphs plays a central role. The girth of a graph is the length of a shortest cycle contained in the graph. Similarly, the odd girth of a graph gives the length of a shortest odd cycle in the graph.

Before we proceed we need some facts about cycle graphs. Note that the odd cycles form a chain in the lattice $\mathcal{C}_S$ between K_2 and $C_3 = K_3$ in the following way: $K_2 \to \cdots \to C_{2i+1} \to C_{2i-1} \to \cdots \to C_3 = K_3$. Note that $C_{2k+1} \not\to K_2$ and $C_{2k+1} \not\to C_{2m+1}$. However, after removing one edge from C_{2k+1}, the remaining subgraph is isomorphic to the path P_{2k+1} which in turn is embeddable in both K_2 and C_{2m+1}. Since C_{2k+1} is edge-transitive, Lemma 5 gives us the following result:

Lemma 10. *Let $0 < k < m$ be odd integers. Then, $s(K_2, C_k) = s(C_m, C_k) = \frac{k-1}{k}$.*

Proposition 11. *Let $k \geq 3$ be odd. Then, $FJ_2(C_k) \geq \frac{k-1}{k} \cdot \alpha_{GW}$ and Hå$(C_k) = \frac{2}{k} + \frac{c}{k^2 \log k} - \frac{2c}{k^3 \log k}$. For any $\varepsilon > 0$, mc_{C_k} cannot be approximated within $\frac{k}{k-1} \cdot \alpha_{GW} + \varepsilon$.*

Proof. From Lemma 10 we see that $s(K_2, C_k) = \frac{k-1}{k}$ which implies (using Lemma 3) that $FJ_2(C_k) \geq \frac{k-1}{k} \cdot \alpha_{GW}$. Furthermore, mc_2 cannot be approximated within $\alpha_{GW} + \varepsilon'$ for any $\varepsilon' > 0$. From the second part of Lemma 3, we get that mc_{C_k} cannot be approximated within $\frac{k}{k-1} \cdot (\alpha_{GW} + \varepsilon')$ for any ε'. With $\varepsilon' = \varepsilon \cdot \frac{k-1}{k}$ the result follows. Finally, the bound on *Hå*(C_k) can be obtained by noting that $e(C_k) = k$. $\square$

Håstad's algorithm does not perform particularly well on sparse graphs; this is reflected by its performance on cycle graphs C_k where the approximation guarantee tends to zero when $k \to \infty$. We will see that this trend is apparent for all graph types studied in this section. Using results of Lai & Liu [25] and Dutton & Brigham [10], we continue with a result on a class of graphs with large girth:

Proposition 12. *Let $n > k \geq 4$. If H is a graph with odd girth $g \geq 2k+1$ and minimum degree $\geq \frac{2n-1}{2(k+1)}$, where $n = v(H)$, then $FJ_2(H) \geq \frac{2k}{2k+1} \cdot \alpha_{GW}$ and mc_H cannot be approximated within $\frac{2k+1}{2k} \cdot \alpha_{GW} + \varepsilon$ for any $\varepsilon > 0$. Asymptotically, Hå(H) is bounded by $\frac{c}{n^2 \log n} + \frac{2(n^{g/(g-1)})^3}{n^4 n^{1/(g-1)}} - \frac{2n^{g/(g-1)} n^{1/(g-1)} c}{n^4 \log n}$.*

Stronger results are possible if we restrict ourselves to planar graphs: Borodin et al. [7] have proved that if H is a planar graph with girth at least $\frac{20k-2}{3}$, then H is $(2+\frac{1}{k})$-colourable, i.e. there exists a homomorphism from H to C_{2k+1}. By applying our method, the following can be proved:

Proposition 13. *Let H be a planar graph with girth at least $g = \frac{20k-2}{3}$. If $v(H) = n$, then $FJ_2(H) \geq \frac{2k}{2k+1} \cdot \alpha_{GW}$ and Hå$(H) \leq \frac{6}{n} - \frac{12}{n^2} + \frac{c}{n^2 \log n} - \frac{6c}{n^3 \log n} + \frac{12c}{n^4 \log n}$. mc_H cannot be approximated within $\frac{2k+1}{2k} \cdot \alpha_{GW} + \varepsilon$ for any $\varepsilon > 0$.*

Proposition 13 can be further strengthened and extended in different ways: one is to consider a result by Dvořák et al. [11]. They have proved that every planar graph H of odd-girth at least 9 is homomorphic to the Petersen graph P. The Petersen graph is edge-transitive and it is known (cf. [3]) that the bipartite density of P is $4/5$ or, in other words, $s(K_2, P) = 4/5$. Consequently, mc_H can be approximated within $\frac{4}{5} \cdot \alpha_{GW}$ but not within $\frac{4}{5} \cdot \alpha_{GW} + \varepsilon$ for any $\varepsilon > 0$. This is better than Proposition 13 for planar graphs with girth strictly less than 13. Another way of extending Proposition 13 is to consider graphs embeddable on higher-genus surfaces. For instance, the lemma is true for graphs embeddable on the projective plane, and it is also true for graphs of girth *strictly* greater than $\frac{20k-2}{3}$ whenever the graphs are embeddable on the torus or Klein bottle. These bounds are direct consequences of results in Borodin et al. [7].

We conclude the section by looking at a class of graphs that have small girth. Let $0 < \beta < 1$ be the approximation threshold for mc_3, i.e. mc_3 is approximable within β but not within $\beta + \varepsilon$ for any $\varepsilon > 0$. Currently, we know that $\alpha_3 \leq 0.836008 \leq \beta \leq \frac{102}{103}$ [9, 21]. The wheel graphs from Section 2.3 are homomorphically equivalent to K_3 for odd k and we conclude (by Lemma 3) that mc_{W_k} has the same approximability properties as mc_3 in this case. For even $k \geq 6$, the following result says that $FJ_3(W_k) \to \alpha_3$ when $k \to \infty$, and Hå(W_k) tends to 0.

Proposition 14. *For $k \geq 6$ and even, $FJ_3(W_k) \geq \alpha_3 \cdot \frac{2k-3}{2k-2}$ but mc_{W_k} is not approximable within $\beta \cdot \frac{2k-2}{2k-3}$. Hå$(W_k) = \frac{4}{k} - \frac{4}{k^2} + \frac{c}{k^2 \log k} - \frac{4c}{k^3 \log k} + \frac{4c}{k^4 \log k}$.*

3.2 Dense and Random Graphs

We will now study *dense* graphs, i.e. graphs H containing $\Theta(v(H)^2)$ edges. For a graph H on n vertices, we obviously have $H \to K_n$. Let $\omega(G)$ denote the size of the largest clique in G and $\chi(G)$ denote the chromatic number of G. If we assume that $\omega(H) \geq r$, then we also have $K_r \to H$. Thus, if we determine $s(K_r, K_n)$, then we can use Lemma 7 to bound $FJ_n(H)$. According to Turán [35], there exists a family of graphs $T(n,r)$ such that $v(T(n,r)) = n$, $e(T(n,r)) = \lfloor \left(1 - \frac{1}{r}\right) \cdot \frac{n^2}{2} \rfloor$, $\omega(T(n,r)) = \chi(T(n,r)) = r$, and if G is a graph such that $e(G) > e(T(v(G), r))$, then $\omega(G) > r$.

Lemma 15. *Let r and n be positive integers. Then, $s(K_r, K_n) = e(T(n,r))/e(K_n)$.*

Proposition 16. *Let $v(H) = n$ and pick $r \in \mathbb{N}$, $\sigma \in \mathbb{R}$ such that $\left\lfloor \left(1 - \frac{1}{r}\right) \cdot \frac{n^2}{2} \right\rfloor \leq \sigma \cdot n^2 = e(H) \leq \frac{n(n-1)}{2}$. Then, $FJ_n(H) \geq \alpha_n \cdot s(K_r, K_n) \sim 1 - \frac{1}{r} - \frac{1}{n} + \frac{2 \ln n}{n(n-1)}$ and Hå$(H) = 2\sigma + \frac{(1-2\sigma)c}{n^2 \log n}$.*

Note that when r and n grow, $FJ_n(H)$ tends to 1. This means that, asymptotically, we cannot do much better. If we compare the expression for $FJ_n(H)$ with the inapproximability bound for mc_n (Theorem 9), we see that all we could hope for is a faster convergence towards 1. As σ satisfies $\left(1 - \frac{1}{r}\right) \cdot \frac{1}{2} \leq \sigma \leq \left(1 - \frac{1}{n}\right) \cdot \frac{1}{2}$, we conclude that *Hå*$(H)$ also tends to 1 as r and n grow. To get a better grip on how *Hå*(H) behaves we look at two extreme cases.

For a maximal $\sigma = \left(1 - \frac{1}{r}\right) \cdot \frac{1}{2}$, *Hå*$(H)$ becomes $1 - \frac{1}{n} + \frac{c}{n^3 \log n}$. On the other hand, this guarantee, for a minimal $\sigma = \left(1 - \frac{1}{r}\right) \cdot \frac{1}{2}$ is $1 - \frac{1}{r} + \frac{c}{rn^2 \log n}$. At the same time, it is easy to see that Frieze and Jerrum's algorithm makes these points approximable within α_n (since, in this case, $H \equiv K_n$) and α_r (since Turán's theorem tells us that $H \to K_r$ holds in this case), respectively. Our conclusion is that Frieze and Jerrum's and Håstad's algorithms perform almost equally well on these graphs asymptotically.

Another way to study dense graphs is via random graphs. Let $\mathcal{G}(n,p)$ denote the random graph on n vertices in which every edge is chosen randomly and independently with probability $p = p(n)$. We say that $\mathcal{G}(n,p)$ has a property A *asymptotically almost surely* (a.a.s.) if the probability it satisfies A tends to 1 as n tends to infinity. Here, we let $p = c$ for some $0 < c < 1$. For $G \in \mathcal{G}(n,p)$ it is well known that a.a.s. $\omega(G)$ assumes one of at most two values around $\frac{2 \ln n}{\ln(1/p)}$ [5, 29]. It is also known that, almost surely $\chi(G) \sim \frac{n}{2 \ln(np)} \ln\left(\frac{1}{1-p}\right)$, as $np \to \infty$ [4, 28]. Let us say that $\chi(G)$ is concentrated in width s if there exists $u = u(n,p)$ such that a.a.s. $u \leq \chi(G) \leq u + s$. Alon and Krivelevich [2] have shown that for every constant $\delta > 0$, if $p = n^{-1/2-\delta}$ then $\chi(G)$ is concentrated in width $s = 1$. That is, almost surely, the chromatic number takes one of two values.

Proposition 17. *Let $H \in \mathcal{G}(n,p)$. When $np \to \infty$, $FJ_m(H) \sim 1 - \frac{2}{m} + \frac{2 \ln m}{m^2} + \frac{1}{m^2} - \frac{2 \ln m}{m^3}$, where $m = \omega(H)$. Hå$(H) = p - \frac{p}{n} + (1-p) \cdot \frac{c}{n^2 \log n} + \frac{pc}{n^3 \log n}$.*

We see that, in the limiting case, *Hå*(H) tends to p, while $FJ_m(H)$ tends to 1. Again, this means that, for large enough graphs, we cannot do much better. With a better analysis, one could possibly reach a faster convergence rate for $FJ_m(H)$.

It is interesting to look at what happens for graphs $H \in \mathcal{G}(n,p)$ where np does not tend to ∞ when $n \to \infty$. We have the following result by Erdős and Rényi [14]: let c be a positive constant and $p = \frac{c}{n}$. If $c < 1$, then a.a.s. no component in $\mathcal{G}(n,p)$

contains more than one cycle, and no component has more than $\frac{\ln n}{c-1-\ln c}$ vertices. Now we see that if $np \to \varepsilon$ when $n \to \infty$ and $0 < \varepsilon < 1$, then $\mathcal{G}(n,p)$ almost surely consists of components with at most one cycle. Thus, each component resembles a cycle where, possibly, trees are attached to certain cycle vertices, and each component is homomorphically equivalent to the cycle it contains. Since we know from Section 3.1 that Frieze and Jerrum's algorithm performs better than Håstad's algorithm on cycle graphs, it follows that the same relationship holds in this part of the $\mathcal{G}(n,p)$ spectrum.

4 Conclusions and Open Problems

We have defined a metric on graphs that measures how well one graph can be embedded in another. While not apparent from its definition, which involves taking infima over the set of all edge-weighted graphs, we have shown that the metric can be computed practically by using linear programming. Given a graph H and known approximability properties for MAX H-COL, this metric allows us to deduce bounds on the corresponding properties for graphs close to H. In other words, the metric measures how well an algorithm for MAX H-COL works on problems MAX H'-COL, for graphs H' close to H, it also translates inapproximability results between these problems. In principle, given a large enough set of graphs with known approximability results for MAX H-COL, our method could be used to derive good bounds on the approximability of MAX H-COL for all graphs. If the known results were in fact tight and the set of graphs dense in $\mathcal{G}_{\equiv}$ (in the topology induced by d), then we would have tight results for all graphs. In this paper we have considered the graphs with known properties to be the complete graphs. We have shown that this set of graphs is sufficient for achieving new bounds on several different classes of graphs, i.e. applying Frieze and Jerrum's algorithm to MAX H-COL gives comparable to or better results than when applying Håstad's MAX 2-CSP algorithm for the classes of graphs we have considered. One possible explanation for this is that the analysis of the MAX 2-CSP algorithm only aims to prove it better than a random solution on expectation, which may leave room for strengthening of the approximation guarantee. At the same time, we are probably overestimating the distance between the graphs. It is likely that both results can be improved. This immediately suggests two clear directions of research. On the one hand, we need approximability/inapproximability result pairs for MAX H-COL on a substantially larger class of graphs. This can be seen considering for example MAX C_5-COL. The closest complete graph to C_5 is K_2, which gives us the inconsequential inapproximability bound $\alpha_{GW} \cdot 5/4 > 1$. On the other hand, we do not measure the actual distance from each graph to the closest complete graph. Instead, we embed each graph between K_2 and a cycle or between its largest clique and K_k, where k is greater than or equal to the chromatic number. In the first case, Erdős [13] has proved that for any positive integers k and l there exists a graph of chromatic number k and girth at least l. It is obvious that such graphs cannot be sandwiched between K_2 and a cycle as was the case of the graphs of high girth in Section 3.1. Additionally, there are obviously graphs with an arbitrarily large gap between largest clique and chromatic number. A different idea is thus required to deal with these graphs. In general, to apply our method more precisely, we need a better understanding of the structure of $\mathcal{C}_S$ and how this interacts with our metric d. Clearly, progress in either one of the directions will influence what type of result to look for in the other direction. In light of this discussion, two interesting candidates for research are the circular complete graphs and the Kneser graphs, see for example [17]. Both of these classes generalise the complete graphs and have been subject to substantial previous research. Partial results for d on 3-colourable circular complete graphs have been obtained by Engström [12].

We conclude the paper by considering two other possible ways to extend our results. Firstly, Kaporis et al. [22] have shown that mc_2 is approximable within $.952$ for any given average degree d and asymptotically almost all random graphs G in $\mathcal{G}(n, m = \lfloor \frac{d}{2} n \rfloor)$, where $\mathcal{G}(n, m)$ is the probability space of random graphs on n vertices and m edges selected uniformly at random. In a similar vein, Coja-Oghlan et al. [8] give an algorithm that approximates mc_k within $1 - O(1/\sqrt{np})$ in expected polynomial time, for graphs from $\mathcal{G}(n, p)$. It would be interesting to know if these results could be carried further, to other graphs G, so that better approximability bounds on MAX H-COL, for H such that $G \rightarrow H$, could be achieved.

Secondly, the idea of defining a metric on a space of problems which relates their approximability can be extended to more general cases. It should not prove too difficult to generalise the framework introduced in this paper to MAX CSP over directed graphs or even languages consisting of a single, finitary relation. How far can this generalisation be carried out? Could it provide any insight into the approximability of MAX CSP on arbitrary constraint languages?

References

1. N. Alon. Bipartite subgraph. *Combinatorica*, 16:301–311, 1996.
2. N. Alon and M. Krivelevich. The concentration of the chromatic number of random graphs. *Combinatorica*, 17:303–313, 1997.
3. A. Berman and X.-D. Zhang. Bipartite density of cubic graphs. *Discrete Mathematics*, 260:27–35, 2003.
4. B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8(1):49–55, 1988.
5. B. Bollobás and P. Erdős. Cliques in random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 80(419):419–427, 1976.
6. J. Bondy and S. Locke. Largest bipartite subgraphs in triangle-free graphs with maximum degree three. *Journal of Graph Theory*, 10:477–504, 1986.
7. O. Borodin, S.-J. Kim, A. Kostochka, and D. West. Homomorphisms from sparse graphs with large girth. *Journal of Combinatorial Theory, ser. B*, 90:147–159, 2004.
8. A. Coja-Oghlan, C. Moore, and V. Sanwalani. MAX k-CUT and approximating the chromatic number of random graphs. *Random Structures and Algorithms*, 28:289–322, 2005.
9. E. de Klerk, D. Pasechnik, and J. Warners. Approximate graph colouring and MAX-k-CUT algorithms based on the θ function. *Journal of Combinatorial Optimization*, 8:267–294, 2004.
10. R. Dutton and R. Brigham. Edges in graphs with large girth. *Graphs and Combinatorics*, 7(4):315–321, 1991.
11. Z. Dvořák, R. Škrekovski, and T. Valla. Planar graphs of odd-girth at least 9 are homomorphic to the Petersen graph. To appear in SIAM Journal on Discrete Mathematics.
12. R. Engström. Approximability distances between circular complete graphs. Master's thesis, Linköping University, 2008. LITH-IDA-EX-A–08/062–SE.

13. P. Erdős. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.
14. P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.
15. A. Frieze and M. Jerrum. Improved approximation algorithms for MAX k-CUT and MAX BISECTION. *Algorithmica*, 18(1):67–81, 1997.
16. M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.
17. P. Hell and J. Nešetřil. *Graphs and Homomorphisms (Oxford Lecture Series in Mathematics and Its Applications)*. Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press, 2004.
18. G. Hophkins and W. Staton. Extremal bipartite subgraphs of cubic triangle-free graphs. *Journal of Graph Theory*, 6:115–121, 1982.
19. J. Håstad. Every 2-CSP allows nontrivial approximation. In *Proceedings of the 37th Annual ACM Symposium on the Theory of Computing (STOC-2005)*, pages 740–746, 2005.
20. P. Jonsson, A. Krokhin, and F. Kuivinen. Ruling out polynomial-time approximation schemes for hard constraint satisfaction problems. In *Proceedings of the 2nd International Computer Science Symposium in Russia (CSR-2007)*, pages 182–193, 2007. Full version available at http://www.dur.ac.uk/andrei.krokhin/papers/hardgap.pdf.
21. V. Kann, S. Khanna, J. Lagergren, and A. Panconesi. On the hardness of approximating MAX k-CUT and its dual. *Chicago Journal of Theoretical Computer Science*, 1997(2), 1997.
22. A. Kaporis, L. Kirousis, and E. Stavropoulos. Approximating almost all instances of MAX-CUT within a ratio above the Håstad threshold. In *Proceedings of the 14th Annual European Symposium on Algorithms (ESA-2006)*, pages 432–443, 2006.
23. S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on the Theory of Computing (STOC-2002)*, pages 767–775, 2002.
24. S. Khot, G. Kindler, E. Mossel, and R. O'Donnel. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? *SIAM Journal of Computing*, 37(1):319–357, 2007.
25. H.-J. Lai and B. Liu. Graph homomorphism into an odd cycle. *Bulletin of the Institute of Combinatorics and its Applications*, 28:19–24, 2000.
26. M. Langberg, Y. Rabani, and C. Swamy. Approximation algorithms for graph homomorphism problems. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 176–187. Springer, 2006.
27. S. Locke. A note on bipartite subgraphs of triangle-free regular graphs. *Journal of Graph Theory*, 14:181–185, 1990.
28. T. Łuczak. The chromatic number of random graphs. *Combinatorica*, 11(1):45–54, 1991.
29. D. Matula. The employee party problem. *Notices of the American Mathematical Society*, 19, 1972. A – 382.
30. P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th annual ACM symposium on the Theory of Computing (STOC-2008)*, pages 245–254, 2008.
31. P. Raghavendra and D. Steurer. How to round any CSP. Manuscript, 2009.
32. R. Šámal. Fractional covering by cuts. In *Proceedings of the 7th International Colloquium on Graph Theory (ICGT-2005)*, pages 455–459, 2005.
33. R. Šámal. *On XY mappings*. PhD thesis, Charles University in Prague, 2006.
34. L. Trevisan. Max Cut and the smallest eigenvalue. *CoRR*, abs/0806.1978, 2008.
35. P. Turán. On an extremal problem in graph theory. *Matematicko Fizicki Lapok*, 48:436–452, 1941.