On the relative complexity of approximate counting problems*

Martin Dyer[†]
School of Computer Studies
University of Leeds

Leslie Ann Goldberg[‡]
Dep't of Computer Science
University of Warwick

Catherine Greenhill[§]
Dep't of Mathematics and Statistics
University of Melbourne

Mark Jerrum[¶]
Division of Informatics
University of Edinburgh

August 24, 2000

Abstract

Two natural classes of counting problems that are interreducible under approximation-preserving reductions are: (i) those that admit a particular kind of efficient approximation algorithm known as an "FPRAS," and (ii) those that are complete for #P with respect to approximation-preserving reducibility. We describe and investigate not only these two classes but also a third class, of intermediate complexity, that is not known to be identical to (i) or (ii). The third class can be characterised as the hardest problems in a logically defined subclass of #P.

*This work was supported in part by the EPSRC Research Grant "Sharper Analysis of Randomised Algorithms: a Computational Approach" and by the ESPRIT Projects RAND-APX and ALCOM-FT.

[†]dyer@scs.leeds.ac.uk, School of Computer Studies, University of Leeds, Leeds LS2 9JT, United Kingdom.

[‡]leslie@dcs.warwick.ac.uk, http://www.dcs.warwick.ac.uk/~leslie/, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

[§]csg@ms.unimelb.edu.au, Department of Mathematics and Statistics, University of Melbourne, Parkville VIC, Australia 3052. Supported by an Australian Research Council Postdoctoral Fellowship.

[¶]mrj@dcs.ed.ac.uk, http://www.dcs.ed.ac.uk/~mrj/, Division of Informatics, University of Edinburgh, JCMB, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom.

1 The setting

Not a great deal is known about the complexity of obtaining approximate solutions to counting problems. A few problems are known to admit an efficient approximation algorithm or "FPRAS" (definition below). Some others are known not to admit an FPRAS under some reasonable complexity-theoretic assumptions. In light of the scarcity of absolute results, we propose to examine the relative complexity of approximate counting problems through the medium of approximation-preserving reducibility. Through this process, a provisional landscape of approximate counting problems begins to emerge. Aside from the expected classes of interreducible problems that are "easiest" and "hardest" within the counting complexity class #P, we identify an interesting class of natural interreducible problems of apparently intermediate complexity.

A *randomised approximation scheme* (RAS) for a function $f : \Sigma^* \to \mathbb{N}$ is a probabilistic Turing machine[1] (TM) that takes as input a pair $(x, \varepsilon) \in \Sigma^* \times (0, 1)$ and produces as output an integer random variable Y satisfying the condition $\Pr(e^{-\varepsilon} \leq Y/f(x) \leq e^{\varepsilon}) \geq 3/4$. A randomised approximation scheme is said to be *fully polynomial* if it runs in time $\mathrm{poly}(|x|, \varepsilon^{-1})$. The unwieldy phrase "fully polynomial randomised approximation scheme" is usually abbreviated to *FPRAS*.

Suppose $f, g : \Sigma^* \to \mathbb{N}$ are functions whose complexity (of approximation) we want to compare. An *approximation-preserving reduction* from f to g is a probabilistic oracle TM M that takes as input a pair $(x, \varepsilon) \in \Sigma^* \times (0, 1)$, and satisfies the following three conditions: (i) every oracle call made by M is of the form (w, δ), where $w \in \Sigma^*$ is an instance of g, and $0 < \delta < 1$ is an error bound satisfying $\delta^{-1} \leq \mathrm{poly}(|x|, \varepsilon^{-1})$; (ii) the TM M meets the specification for being a randomised approximation scheme for f whenever the oracle meets the specification for being a randomised approximation scheme for g; and (iii) the run-time of M is polynomial in $|x|$ and ε^{-1}. If an approximation-preserving reduction from f to g exists we write $f \leq_{\mathrm{AP}} g$, and say that *f is AP-reducible to g*. If $f \leq_{\mathrm{AP}} g$ and $g \leq_{\mathrm{AP}} f$ then we say that *f and g are AP-interreducible*, and write $f \equiv_{\mathrm{AP}} g$.

In arriving at a precise definition of AP-reducibility a number of issues had to be resolved. Should the reduction be deterministic or randomised? Should it be Turing or many-one/Karp? Should ε enter explicitly into the time bound for the reduction? As a general principle, we have always chosen the most liberal option, i.e., the one leading to the largest class of reductions.[2] However, we shall only rarely make use of the full generality our definition, preferring in the main to work within a more restricted class of reductions.

Two counting problems play a special role in this article.

Name. #SAT.

Instance. A Boolean formula φ in conjunctive normal form (CNF).

[1] All our Turing machines will be *transducers*, i.e., equipped with a write-only output tape. In what follows, we shall not mention this fact explicitly.

[2] At the other extreme, Saluja, Subrahmanyam and Thakur [16] propose a very demanding notion of approximation-preserving reduction, which is probably not suitable for our purposes.

Output. The number of satisfying assignments to φ.

Name. #BIS.

Instance. A bipartite graph B.

Output. The number of independent sets in B.

The problem #SAT is the counting version of the familiar decision problem SAT, so its special role is not surprising. The (apparent) significance of #BIS will only emerge from an extended empirical study using the tool of approximation-preserving reducibility. This is not the first time the problem #BIS has appeared in the literature. Provan and Ball show it to be #P-complete [14], while (in the guise of "2BPMONDNF") Roth raises, at least implicitly, the question of its approximability [15].

Three classes of AP-interreducible problems are studied in this paper. The first is the class of counting problems (functions $\Sigma^* \to \mathbb{N}$) that admit an FPRAS. These are trivially AP-interreducible, since all the work can be embedded into the reduction (which declines to use the oracle). The second is the class of counting problems AP-interreducible with #SAT. As we shall see, these include the "hardest to approximate" counting problems within the class #P. The third is the class of counting problems AP-interreducible with #BIS. These problems are naturally AP-reducible to functions in #SAT, but we have been unable to demonstrate the converse relation. Moreover, no function AP-interreducible with #BIS is known to admit an FPRAS. Since a number of natural and reasonably diverse counting problems *are* AP-interreducible with #BIS, it remains a distinct possibility that the complexity of this class of problems in some sense lies strictly between the class of problems admitting an FPRAS and #SAT. Perhaps significantly, #BIS and its relatives can be characterised as the hardest to approximate problems within a logically defined subclass of #P that we name $\#\mathrm{RH}\Pi_1$.

2 Problems that admit an FPRAS

A very few non-trivial combinatorial structures may be counted *exactly* using a polynomial-time deterministic algorithm; a fortiori, they may be counted using an FPRAS. The two key examples are spanning trees in a graph (Kirchhoff), and perfect matchings in a planar graph (Kasteleyn). Intriguingly, both of these algorithms rely on a reduction to a determinant, which may be computed in polynomial time by Gaussian elimination. Details of both algorithms may be found in Kasteleyn's survey article [13].

There are some additional specimens that are more interesting in the context of this article: problems that admit an FPRAS despite being complete (with respect to usual Turing reducibility) in #P. These are more common than exactly solvable counting problems, but still not numerous. Two representative examples are:

Name. #MATCH.

Instance. A graph G.[3]

Output. The number of matchings (of all sizes) in G.

Name. #DNF-SAT.

Instance. A Boolean formula φ in disjunctive normal form (DNF).

Output. The number of satisfying assignments to φ.

#MATCH may be approximated in the FPRAS sense by "Markov chain Monte Carlo" (Jerrum and Sinclair [9]), and #DNF-SAT by a more direct sampling technique (Karp, Luby and Madras [12]).

3 Problems AP-interreducible with #Sat

Suppose $f, g : \Sigma^* \to \mathbb{N}$. A *parsimonious reduction* (Simon [17]) from f to g is a function $\varrho : \Sigma^* \to \Sigma^*$ satisfying (i) $f(w) = g(\varrho(w))$ for all $w \in \Sigma^*$, and (ii) ϱ is computable by a polynomial-time deterministic Turing transducer. In the context of counting problems, parsimonious reductions "preserve the number of solutions." The generic reductions used in the usual proofs of Cook's theorem are parsimonious, i.e., the number of satisfying assignments of the constructed formula is equal to the number of accepting computations of the given Turing machine/input pair. Since a parsimonious reduction is a very special instance of an approximation-preserving reduction, we see that all problems in #P are AP-reducible to #SAT. Thus #SAT is complete for #P w.r.t. (with respect to) AP-reducibility. This means that #SAT cannot have an FPRAS unless NP = RP. The same is obviously true of any problem in #P to which #SAT is AP-reducible.

Let $\mathrm{A} : \Sigma^* \to \{0, 1\}$ be some decision problem in NP. One way of expressing membership of A in NP is to assert the existence of a polynomial p and a polynomial-time computable predicate R (witness-checking predicate) satisfying the following condition: $\mathrm{A}(x)$ iff there is a word $y \in \Sigma^*$ such that $|y| = p(|x|)$ and $R(x, y)$. The counting problem, $\#\mathrm{A} : \Sigma^* \to \mathbb{N}$, corresponding to A is defined by

$$\#\mathrm{A}(x) = \left| \left\{ y \mid |y| = p(|x|) \text{ and } R(x, y) \right\} \right|.$$

Formally, the counting version #A of A depends on the witness-checking predicate R and not just on A itself; however, there is usually a "natural" choice for R, so our notation should not confuse. Note that our notation for #SAT and SAT is consistent with the convention just established, where we take "y is a satisfying assignment to formula x" as the witness-checking predicate.

Many "natural" NP-complete problems A have been considered, and in every case the corresponding counting problem #A is complete for #P with

[3]Note that the graph G is no longer restricted to be planar.

respect to (conventional) polynomial-time Turing reducibility. No counterexamples to this phenomenon are known, so it remains a possibility that this empirically observed relationship is actually a theorem. If so, we seem to be far from proving it or providing a counterexample. Strangely enough, the corresponding statement for AP-reducibility *is* a theorem.

Theorem 1 *Let* A *be an NP-complete decision problem. Then the corresponding counting problem,* #A, *is complete for* #P *w.r.t. AP-reducibility.*

Proof. That $\#\mathrm{A} \in \#\mathrm{P}$ is immediate. The fact that #SAT is AP-reducible to #A is more subtle. Using the bisection technique of Valiant and Vazirani, we know [21, Cor. 3.6] that #SAT can be approximated (in the FPRAS sense) by a polynomial-time probabilistic TM M equipped with an oracle for the *decision* problem SAT.[4] Furthermore, the decision oracle for SAT may be replaced by an approximate counting oracle (in the RAS sense) for #A, since A is NP-complete, and a RAS must, in particular, reliably distinguish none from some. (Note that the failure probability may be made negligible through repeated trials [11, Lemma 6.1].) Thus the TM M, with only slight modification, meets the specification for an approximation-preserving reduction from #SAT to #A. We conclude that the counting version of every NP-complete problem is complete for #P w.r.t. AP-reducibility. □

The following problem is a useful starting point for reductions.

Name. #LARGEIS.

Instance. A positive integer m and a graph G in which every independent set has size at most m.

Output. The number of size-m independent sets in G.

Garey et al. [7] have shown that the decision problem corresponding to #LARGEIS is NP-complete. Therefore, Theorem 1 implies the following:

Observation 2 #LARGEIS $\equiv_{\mathrm{AP}}$ #SAT.

Another insight that comes out of the proof of Theorem 1 is that the set of functions AP-reducible to #SAT has a "structural" characterisation as the class of functions that may be approximated (in the FPRAS sense) by a polynomial-time probabilistic Turing transducer equipped with an NP oracle. Informally, in a complexity-theoretic sense, approximate counting is much easier that exact counting: the former lies "just above" NP [19], while the latter lies above the entire polynomial hierarchy [20].

Theorem 1 shows that counting versions of NP-complete problems are all AP-interreducible. Simon, who introduced the notion of parsimonious reduction [17], noted that many of these counting problems are in fact parsimoniously interreducible with #SAT. In other words, many of the problems covered by

[4] Only a sketch of the proof of this fact is presented in [21]; for a detailed proof, consult Goldreich's lecture notes [8].

Theorem 1 are in fact related by direct reductions, often parsimonious, as opposed to the rather arcane reductions implicit in that theorem. Since we are interested in investigating exactly when the full power of AP-reducibility is necessary, we also offer a proof of Observation 2 by direct reduction, in Appendix A.[5]

An interesting fact about exact counting, discovered by Valiant, is that a problem may be complete for #P w.r.t. usual Turing reducibility even though its associated decision problem is polynomial-time solvable. So it is with approximate counting. A counting problems may be complete for #P w.r.t. AP-reducibility when its associated decision problem is not NP-complete, and even when it is trivial, as in the next example.

Name. #IS.

Instance. A graph G.

Output. The number of independent sets (of all sizes) in G.

Theorem 3 $\#\mathrm{IS} \equiv_{\mathrm{AP}} \#\textsc{Sat}$.

Proof. We need only demonstrate that $\#\textsc{Sat} \leq_{\mathrm{AP}} \#\mathrm{IS}$, since the opposite direction comes from the generic reduction of Cook's theorem. We'll actually show $\#\textsc{Large}\mathrm{IS} \leq_{\mathrm{AP}} \#\mathrm{IS}$, which is sufficient by Observation 2. The "boosting" technique we use was presented by Sinclair [18], but is repeated here with a view to providing a simple, concrete example of an approximation-preserving reduction.

Let m and $G = (V, E)$ be an instance of $\#\textsc{Large}\mathrm{IS}$, and set $n = |V|$. Construct an instance $G' = (V', E')$ of #IS as follows:

$$V' = V \times [r],$$

and

$$E' = \Big\{ \big\{(u,i),(v,j)\big\} : \{u,v\} \in E \text{ and } i,j \in [r] \Big\},$$

where r is a sufficiently large number, to be chosen later, and $[r] = \{0, \dots, r-1\}$ denotes the set containing the first r natural numbers. Informally, vertices in G are transformed to r-independent sets in G', and edges to complete bipartite graphs on $r + r$ vertices.

An independent set I' in G' projects to an independent set $I = \pi(I')$ in G in the following natural way

$$I = \pi(I') = \big\{ v \in V : \text{ there exists } i \in [r] \text{ such that } (v,i) \in I' \big\}.$$

[5]In Appendix A, we give a parsimonious reduction from $\#\textsc{Sat}$ to $\#\textsc{Large}\mathrm{IS}$. This provides a (direct) proof of Observation 2. It turns out that Observation 2 remains true even when the definition of $\#\textsc{Large}\mathrm{IS}$ is modified so that a "witness" is provided along with every problem instance. In particular, along with m and G, a proper m-vertex-colouring of the complement of G is provided. The colouring serves as a witness that every independent set of G has size at most m. The reduction in Appendix A shows how such witnesses can be incorporated into the constructed problem instance.

Furthermore, every independent set of size k in G arises in exactly $(2^r - 1)^k$ ways as a projection of this kind. Thus, denoting by $\mathcal{I}_m(G)$ the set of all size-m independent sets in G and by $\mathcal{I}(G')$ the set of all independent sets in G',

$$|\mathcal{I}(G')| \geq (2^r - 1)^m \cdot |\mathcal{I}_m(G)|.$$

On the other hand, at most $(2^r - 1)^{m-1}$ independent sets I' in G' project to each independent set $I = \pi(I')$ in G of size strictly less than m. Thus

$$|\mathcal{I}(G')| \leq (2^r - 1)^m \cdot |\mathcal{I}_m(G)| + (2^r - 1)^{m-1} 2^n.$$

It follows from the two inequalities that

$$\mathcal{I}_m(G) = \left\lfloor \frac{|\mathcal{I}(G')|}{(2^r - 1)^m} \right\rfloor, \tag{1}$$

provided we choose $r \geq n + 1$. Thus we have constructed an AP-reduction from #LargeIS to #IS: use an oracle for #IS to approximate $|\mathcal{I}(G')|$, divide by $(2^r - 1)^m$, and round[6] to the nearest integer. (The reduction is of a rather degenerate form, with one oracle call and no use of randomisation.)

As this is the first concrete example of an approximation-preserving reduction, we add some technical details concerning the choice of the accuracy parameter δ in the definition of reduction. If it were not for the floor function in (1), we could simply set $\delta = \varepsilon$, since division by a constant preserves relative error. The discontinuous floor function could spoil the approximation when its argument is small. However, we shall only apply the floor function in situations where its argument is in the range (say) $[N, N + 1/4]$ for some integer N. This avoids technical problems, as we now see.

Suppose more generally that the true result N is obtained by rounding a fraction Q with $|Q - N| \leq 1/4$. Suppose further that the oracle provides an approximation $\widehat{Q}$ to Q satisfying $Qe^{-\delta} \leq \widehat{Q} \leq Qe^{\delta}$ (as it is required to do with probability at least 3/4). Set $\delta = \varepsilon/21$, where ε is the accuracy parameter governing the final result. There are two cases. If $N \leq 2/\varepsilon$, then a short calculation yields $|\widehat{Q} - Q| < 1/4$ implying that the result returned is exact. If $N > 2/\varepsilon$, then the result returned is in the range $[(N - 1/4)e^{-\delta} - 1/2, (N + 1/4)e^{\delta} + 1/2]$ which, for the chosen δ, is contained in $[Ne^{-\varepsilon}, Ne^{\varepsilon}]$. □

Other counting problems can be shown to be complete for #P w.r.t. AP-reducibility using similar "boosting reductions." There is a paucity of examples that are complete for some more "interesting" reason. One result that might qualify is the following:

Theorem 4 #IS *remains complete for* #P *w.r.t. AP-reducibility even when restricted to graphs of maximum degree 25.*

Proof. This follows from a result of Dyer, Frieze and Jerrum [4], though rather indirectly. In the proof of Theorem 2 of [4] it is demonstrated that an FPRAS for bounded-degree #IS could be used (as an oracle) to provide a polynomial-time

[6] We will explain below why one should round rather than taking the floor.

randomised algorithm for an NP-complete problem, such as the decision version of satisfiability. Then #SAT $\leq_{\mathrm{AP}}$ #IS follows, as before, via the bisection technique of Valiant and Vazirani. □

Let H be any fixed, q-vertex graph, possibly with loops. An *H-colouring* of a graph G is simply a homomorphism from G to H. If we regard the vertices of H as representing colours, then a homomorphism from G to H induces a q-colouring of G that respects the structure of H: two colours may be adjacent in G only if the corresponding vertices are adjacent in H. Some examples: K_q-colourings, where K_q is the complete q-vertex graph, are simply the usual (proper) q-colourings; K_2^1-colourings, where K_2^1 is K_2 with one loop added, are independent sets; and S_q^*-colourings, where S_q^* is the q-leaf star with loops on all $q+1$ vertices, are configurations in the "q-particle Widom-Rowlinson model" from statistical physics.

Name. #q-PARTICLE-WR-CONFIGS.

Instance. A graph G.

Output. The number of q-particle Widom-Rowlinson configurations in G, i.e., S_q^*-colourings of G, where S_q^* denotes the q-leaf star with loops on all $q+1$ vertices.

We will return to the problem of counting Widom-Rowlinson configurations later in the paper. In particular, we will show (in §4) that #2-PARTICLE-WR-CONFIGS is AP-interreducible with #BIS and (in §6) that #3-PARTICLE-WR-CONFIGS is at least as hard as #BIS in the sense that #BIS $\leq_{\mathrm{AP}}$ #3-PARTICLE-WR-CONFIGS. We will also show (in §7) that for $q \geq 4$, #q-PARTICLE-WR-CONFIGS is AP-interreducible with #SAT.

Aside from containing many problems of interest, H-colourings provide an excellent setting for testing our understanding of the complexity landscape of (exact and approximate) counting. To initiate this programme we considered all 10 possible 3-vertex connected Hs (up to symmetry, and allowing loops). The complexity of *exactly* counting H-colourings was completely resolved by Dyer and Greenhill [5]. Aside from $H = K_3^*$ (the complete graph with loops on all three vertices) and $H = K_{1,2} = P_3$ (P_n will be used to denote the path of length $n-1$ on n vertices), which are trivially solvable, the problem of counting H-colourings for connected three-vertex Hs is #P-complete. Of the eight Hs for which exact counting is #P-complete, seven can be shown to be complete for #P w.r.t. AP-reducibility using reductions very similar to those appearing elsewhere in this article. The remaining possibility for H is S_2^* (i.e, 2-particle Widom-Rowlinson configurations) which we return to in the next section. Other complete problems could be mentioned here but we prefer to press on to a potentially more interesting class of counting problems.

4 Problems AP-interreducible with #BIS

The reduction described in the proof of Theorem 3 does not provide useful information about #BIS, since we do not have any evidence that the restriction of #LARGEIS to bipartite graphs is complete for #P w.r.t. AP-reducibility.[7] The fact that #BIS is interreducible with a number of other problems not known to be complete (or to admit an FPRAS) prompts us to study #BIS and its relatives in some detail. The following list provides examples of problems AP-interreducible with #BIS; more will be added later.

Name. #P_4-COL.

Instance. A graph G.

Output. The number of P_4-colourings of G, where P_4 is the path of length 3.

Name. #DOWNSETS.

Instance. A partially ordered set $(X, \preceq)$.

Output. The number of downsets in $(X, \preceq)$.

Name. #1P1NSAT.

Instance. A Boolean formula φ in conjunctive normal form (CNF), with at most one unnegated literal per clause, and at most one negated literal.

Output. The number of satisfying assignments to φ.

Name. #BEACHCONFIGS.

Instance. A graph G.

Output. The number of "Beach configurations" in G, i.e., P_4^*-colourings of G, where P_4^* denotes the path of length 3 with loops on all four vertices.

Note that an instance of #1P1NSAT is a conjunction of Horn clauses, each having one of the restricted forms $x \Rightarrow y$, $\neg x$, or y, where x and y are variables.

Theorem 5 *The problems #BIS, #P_4-COL, #2-PARTICLE-WR-CONFIGS, #BEACHCONFIGS, #DOWNSETS and #1P1NSAT are all AP-interreducible.*

Proof. The problems #BIS and #P_4-COL are essentially the same. A graph G is P_4-colourable iff it is bipartite, in which case two of the colours (the end ones) point out an independent set. Conversely, each independent set in a connected bipartite graph G arises from one of two distinct P_4 colourings in this manner.[8] The correspondence between independent sets and P_4-colourings (trivially) constitutes a matching pair of approximation-preserving reductions between the two problems.

[7] Note that this statement does not contradict the general principle, enunciated in §3, that counting-analogues of NP-complete decision problems are complete w.r.t. AP-reducibility, since a maximum cardinality independent set can be located in a bipartite graph using network flow.

[8] The symmetry of P_4 allows a renaming of colours; in general, the correspondence between colourings and independent sets is $2^\kappa : 1$, where κ is the number of connected components of G.

The problems #DOWNSETS and #1P1NSAT are also very close; indeed, #DOWNSETS is a restricted version of #1P1NSAT in which (a) all clauses have two literals, i.e., are of the form $x \Rightarrow y$, and (b) there are no cyclic chains of implications $x_0 \Rightarrow x_1 \Rightarrow \cdots \Rightarrow x_{\ell-1} \Rightarrow x_0$. But, given an arbitrary instance of #1P1NSAT, any forced variables as in (a) may be removed by substituting FALSE or TRUE and then simplifying; and any set of ℓ variables forming a cyclic chain as in (b) may be replaced by a single variable. So #DOWNSETS and #1P1NSAT are certainly AP-interreducible.

AP-interreducibility of all the problems other than #P_4-COL and #1P1NSAT follows from the cycle of reductions

$$\begin{aligned}\#\text{BIS} &\leq_{\text{AP}} \#\text{2-PARTICLE-WR-CONFIGS}\\ &\leq_{\text{AP}} \#\text{BEACHCONFIGS}\\ &\leq_{\text{AP}} \#\text{DOWNSETS}\\ &\leq_{\text{AP}} \#\text{BIS}\end{aligned}$$

which are presented in Lemmas 6, 7, 8 and 9. Note that a reduction from #2-PARTICLE-WR-CONFIGS to #BIS was already known [1]. □

Lemma 6 #BIS $\leq_{\text{AP}}$ #2-PARTICLE-WR-CONFIGS.

Proof. Suppose $B = (X, Y, A)$ is an instance of #BIS, where $A \subseteq X \times Y$. For convenience, $X = \{x_0, \ldots, x_{n-1}\}$ and $Y = \{y_0, \ldots, y_{n-1}\}$. Construct an instance $G = (V, E)$ of #2-PARTICLE-WR-CONFIGS as follows. Let $U_i : 0 \leq i \leq n-1$ and K all be disjoint sets of size $3n$. Then define

$$V = \bigcup_{i\in[n]} U_i \cup \{v_0, \ldots, v_{n-1}\} \cup K$$

and

$$E = \bigcup_{i\in[n]} U_i^{(2)} \cup \big(\{v_0, \ldots, v_{n-1}\} \times K\big) \cup K^{(2)} \cup \bigcup \big\{U_i \times \{v_j\} : (x_i, y_j) \in A\big\},$$

where $U_i^{(2)}$, etc., denotes the set of all unordered pairs of elements from U_i. So U_i and K all induce cliques in G, and all v_j are connected to all of K. Let the Widom-Rowlinson (W-R) colours be red, white and green, where white is the centre colour. Say that a W-R configuration (colouring) is *full* if all the sets $U_0, \ldots, U_{n-1}$ and K are bichromatic. (Note that each set is either monochromatic, or bichromatic red/white or green/white.) We shall see presently that full W-R configurations account for all but a vanishing fraction of the set of all W-R configurations.

Consider a full W-R configuration $C : V \to \{\text{red}, \text{white}, \text{green}\}$ of G. Assume $C(K) = \{\text{red}, \text{white}\}$; the other possibility, with green replacing red is symmetric. Every full colouring in G may be interpreted as an independent set in B as follows:

$$I = \big\{x_i : \text{green} \in C(U_i)\big\} \cup \big\{y_j : C(v_j) = \text{red}\big\}.$$

Moreover, every independent set in B can be obtained in this way from exactly $(2^{3n}-2)^{n+1}$ full W-R configurations of G satisfying the condition $C(K) = \{\text{red}, \text{white}\}$. So

$$|\mathcal{W}'(G)| = 2(2^{3n}-2)^{n+1} \cdot |\mathcal{I}(B)|,$$

where $\mathcal{W}'(G)$ denotes the set of full W-R configurations of G, and the factor of two comes from symmetry between red and green.

Crude counting estimates provide

$$|\mathcal{W}(G) \setminus \mathcal{W}'(G)| \leq 3(n+1)(2 \cdot 2^{3n})^n 3^n,$$

where $\mathcal{W}(G)$ denotes the set of all W-R configurations of G. (One of the $n+1$ sets in $\{U_0, \ldots, U_{n-1}, K\}$ is not bichromatic, and has at most 3 colourings. Each of the other sets has at most 2×2^{3n} colourings. There are at most 3^n ways to colour $v_0, \ldots, v_{n-1}$.) Since

$$\frac{3(n+1)(2 \cdot 2^{3n})^n 3^n}{2(2^{3n}-2)^{n+1}} < \frac{1}{4}$$

for n sufficiently large (actually $n \geq 17$) we have

$$|\mathcal{I}(B)| = \left\lfloor \frac{|\mathcal{W}(G)|}{2(2^{3n}-2)^{n+1}} \right\rfloor$$

and the result follows as in the proof of Theorem 3. □

Lemma 7 #2-Particle-WR-Configs $\leq_{\mathrm{AP}}$ #BeachConfigs.

Proof. Let $G = (V, E)$ be an instance of #2-Particle-WR-Configs, with $|V| = n$. Construct an instance $G' = (V', E')$ of #BeachConfigs as follows:

$$V' = V \cup \{s\} \cup [r],$$

and

$$E' = E \cup (V \times \{s\}) \cup (\{s\} \times [r]),$$

where r is a sufficiently large number, to be chosen later. There are four possible colours that can be applied to the vertex s, but only two distinct ones, up to symmetry. If one of the "end" colours is used to colour s, then all the other vertices must receive one of two colours, and any assignment of the two colours is permissible; thus there are 2^{n+r} ways to complete the colouring of G'. If one of the "middle" colours is used to colour s, then the induced colouring on V is a W-R configuration, and the remaining r vertices may be tricoloured. Combining these counts,

$$|\mathcal{B}(G')| = 2 \cdot 3^r \cdot |\mathcal{W}(G)| + 2 \cdot 2^{n+r},$$

where $\mathcal{B}(G')$ denotes the set of all beach configurations of G'. Hence

$$|\mathcal{W}(G)| = \left\lfloor \frac{|\mathcal{B}(G')|}{2 \cdot 3^r} \right\rfloor,$$

provided r is large enough. In fact $r = 2n$ will do, as then $2^{n+r}/3^r = (8/9)^n$, which is less than $1/4$ when $n \geq 12$. □

Lemma 8 #BeachConfigs $\leq_{\mathrm{AP}}$ #Downsets.

Proof. Let $G = (V, E)$ be an instance of #BeachConfigs, with $|V| = n$. We construct, as an instance of #Downsets, a partial order on the $3n$-element set $V\times[3]$. For each vertex v, we impose the relationships $(v,0) \prec (v,1) \prec (v,2)$; for each edge (u,v), the relationships $(v,0) \prec (u,1)$, $(v,1) \prec (u,2)$, $(u,0) \prec (v,1)$ and $(u,1) \prec (v,2)$. Given a downset D and a vertex v, there are four possibilities for the set $D \cap \{(v,0),(v,1),(v,2)\}$: these are the four colours of a Beach configuration. So there is a 1-1 correspondence between Beach configurations in G and downsets in $(V\times[3],\prec)$. □

Lemma 9 #Downsets $\leq_{\mathrm{AP}}$ #BIS.

Proof. Let $(X,\preceq)$ be an instance of #Downsets. For convenience, identify X with $[n]$. Define a bipartite graph $B = (U, V, E)$ as follows. Let $\{U_i, V_i : i \in X\}$ be a collection of disjoint sets with $|U_i| = |V_i| = 2n$. Then define $U = \bigcup_{i\in X} U_i$, $V = \bigcup_{i\in X} V_i$, and

$$E = \big\{(u,v) : u \in U_i \wedge v \in V_j \wedge i \preceq j\big\}.$$

(Note that equality is allowed between i and j, so that $U_i \cup V_i$ induces a complete bipartite graph on $2n + 2n$ vertices.) Call an independent set $I \in \mathcal{I}(B)$ *full* iff $I \cap (U_i \cup V_i) \neq \emptyset$ for all $i \in X$. Denote by $\mathcal{I}'(B)$ the set of all full independent sets in B, and by $\mathcal{D}(X,\preceq)$ the set of all downsets in the partial order $(X,\preceq)$. Every full independent set $I \in \mathcal{I}'(B)$ corresponds to a downset $D = \{i \in X : I \cap V_i \neq \emptyset\}$, and every downset $D \in \mathcal{D}(X,\preceq)$ arises from exactly $(2^{2n}-1)^n$ full independent sets I in this way; thus

$$|\mathcal{I}'(B)| = (2^{2n}-1)^n \cdot |\mathcal{D}(X,\preceq)|.$$

By a crude estimation of non-full independent sets,

$$|\mathcal{I}(B) \setminus \mathcal{I}'(B)| \leq 3^n(2^{2n}-1)^{n-1}.$$

Since

$$\frac{3^n(2^{2n}-1)^{n-1}}{(2^{2n}-1)^n} < \frac{1}{4}$$

(at least for $n \geq 5$),

$$|\mathcal{D}(X,\preceq)| = \left\lfloor \frac{|\mathcal{I}(B)|}{(2^{2n}-1)^n} \right\rfloor$$

and the result follows as in the proof of Theorem 3 □

#2-Particle-WR-Configs and #BeachConfigs are in fact the first two examples in an infinite sequence of #BIS-equivalent problems. Consider the following sequence of counting problems, parameterised by a positive integer parameter q:

Name. $\#P_q^*$-COL.

Instance. A graph G.

Output. The number of P_q^*-colourings of G, where P_q^* is the path of length $q-1$ with loops on all q vertices.

Observe that #2-PARTICLE-WR-CONFIGS and #BEACHCONFIGS are the special cases $q = 3$ and $q = 4$, respectively. The reductions presented in the proofs of Lemmas 7 and 8 easily generalise to higher q so we immediately obtain.

Theorem 10 $\#P_q^*$-COL $\equiv_{\mathrm{AP}}$ #BIS, *for all* $q \geq 3$.

Clearly, the case $q = 2$ is trivially solvable.

5 A logical characterisation of #BIS and its relatives

Saluja, Subrahmanyam and Thakur [16] have presented a logical characterisation of the class #P (and of some of its subclasses), much in the spirit of Fagin's logical characterisation of NP [6]. In their framework, a counting problem is identified with a sentence φ in first-order logic, and the objects being counted with models of φ. By placing a syntactic restriction on φ, it is possible to identify a subclass $\#\mathrm{RH}\Pi_1$ of #P whose complete problems include all the ones mentioned in Theorem 5.

We follow as closely as possible the notation and terminology of [16], and direct the reader to that article for further information and clarification. A *vocabulary* is a finite set $\sigma = \{\widetilde{R}_0, \ldots, \widetilde{R}_{k-1}\}$ of relation symbols of arities $r_0, \ldots, r_{k-1}$. A *structure* $\mathbf{A} = (A, R_0, \ldots, R_{k-1})$ over σ consists of a *universe* (set of objects) A, and relations $R_0, \ldots, R_{k-1}$ of arities $r_0, \ldots, r_{k-1}$ on A; naturally, each relation $R_i \subseteq A^{r_i}$ is an interpretation of the corresponding relation symbol $\widetilde{R}_i$.[9] We deal exclusively with ordered finite structures; i.e., the size $|A|$ of the universe is finite, and there is an extra binary relation that is interpreted as a total order on the universe. Instead of representing an instance of a counting problem as a word over some alphabet Σ, we represent it as a structure $\mathbf{A}$ over a suitable vocabulary σ. For example, an instance of #IS is a graph, which can be regarded as a structure $\mathbf{A} = (A, \sim)$, where A is the vertex set and $\sim$ is the (symmetric) binary relation of adjacency.

The objects to be counted are represented as sequences $\mathbf{T} = (T_0, \ldots, T_{r-1})$ and $\mathbf{z} = (z_0, \ldots, z_{m-1})$ of (respectively) relations and first-order variables. We say that a counting problem f (a function from structures over σ to numbers) is in the class $\#\mathcal{FO}$ if it can be expressed as

$$f(\mathbf{A}) = \bigl|\bigl\{(\mathbf{T}, \mathbf{z}) : \mathbf{A} \models \varphi(\mathbf{z}, \mathbf{T})\bigr\}\bigr|,$$

where φ is a first-order formula with relation symbols from $\sigma \cup \mathbf{T}$ and (free) variables from $\mathbf{z}$. For example, by encoding an independent set as a unary

[9] We have emphasised here the distinction between a relation symbol $\widetilde{R}_i$ and its interpretation R_i. From now on, however, we simplify notation by referring to both as R_i. The meaning should be clear from the context.

relation I, we may express #IS quite simply as

$$f_{\mathrm{IS}}(\mathbf{A}) = \bigl|\bigl\{(I) : \mathbf{A} \models \forall x, y.\, x \sim y \Rightarrow \neg I(x) \vee \neg I(y)\bigr\}\bigr|.$$

Indeed, #IS is in the subclass $\#\Pi_1 \subset \#\mathcal{FO}$ (so named by Saluja et al.), since the formula defining f_{IS} contains only universal quantification. Saluja et al. [16] exhibit a strict hierarchy of subclasses

$$\#\Sigma_0 = \#\Pi_0 \subset \#\Sigma_1 \subset \#\Pi_1 \subset \#\Sigma_2 \subset \#\Pi_2 = \#\mathcal{FO} = \#\mathrm{P}$$

based on quantifier alternation depth. Among other things, they demonstrate that all functions in $\#\Sigma_1$ admit an FPRAS.[10]

All the problems introduced in §4, in particular those mentioned in Theorem 5, lie in a syntactically restricted subclass $\#\mathrm{RH}\Pi_1 \subseteq \#\Pi_1$ to be defined presently. Furthermore, they characterise $\#\mathrm{RH}\Pi_1$ in the sense of being complete for $\#\mathrm{RH}\Pi_1$ with respect to AP-reducibility (and even, as we shall see, with respect to a much more demanding notion of reducibility). We say that a counting problem f is in the class $\#\mathrm{RH}\Pi_1$ if it can be expressed in the form

$$f(\mathbf{A}) = \bigl|\bigl\{(\mathbf{T}, \mathbf{z}) : \mathbf{A} \models \forall \mathbf{y}.\, \psi(\mathbf{y}, \mathbf{z}, \mathbf{T})\bigr\}\bigr|, \tag{2}$$

where ψ is an unquantified CNF formula in which each clause has at most one occurrence of an unnegated relation symbol from $\mathbf{T}$, and at most one occurrence of a negated relation symbol from $\mathbf{T}$. The rationale behind the naming of the class $\#\mathrm{RH}\Pi_1$ is as follows: "Π_1" indicates that only universal quantification is allowed, and "RH" that the unquantified subformula ψ is in "restricted Horn" form. Note that the restriction on clauses of ψ applies only to terms involving symbols from $\mathbf{T}$; other terms may be arbitrary.

For example, suppose we represent an instance of #DOWNSETS as a structure $\mathbf{A} = (A, \preceq)$, where $\preceq$ is a binary relation (assumed to be a partial order). Then $\#\text{DOWNSETS} \in \#\mathrm{RH}\Pi_1$ since the number of downsets in the partially ordered set $(A, \preceq)$ may be expressed as

$$f_{\mathrm{DS}}(\mathbf{A}) = \bigl|\bigl\{(D) : \mathbf{A} \models \forall x \in A, y \in A.\, D(x) \wedge y \preceq x \Rightarrow D(y)\bigr\}\bigr|, \tag{3}$$

where we have represented a downset in an obvious way as a unary relation D on A. The problem #1P1NSAT is expressed by a formally identical expression, but with $\preceq$ interpreted as an arbitrary binary relation (representing clauses) rather than a partial order.[11]

The main result of this section is

Theorem 11 *#1P1NSAT is complete for $\#\mathrm{RH}\Pi_1$ under parsimonious reducibility.*

[10] The class $\#\Sigma_1$ is far from capturing all functions admitting an FPRAS. For example, #DNF-SAT admits an FPRAS even though it lies in $\#\Sigma_2 \setminus \#\Pi_1$ [16].

[11] To be absolutely precise, one also needs two unary relations, U and N say, to code the one-literal clauses.

Proof. Consider the generic counting problem in $\#\mathrm{RH}\Pi_1$, as presented in equation (2). Suppose $\mathbf{T} = (T_0, \ldots, T_{r-1})$, $\mathbf{y} = (y_0, \ldots, y_{\ell-1})$ and $\mathbf{z} = (z_0, \ldots, z_{m-1})$, where (T_i) are relations of arity (t_i), and (y_j) and (z_k) are first-order variables. Let $L = |A|^\ell$ and $M = |A|^m$, and let $(\eta_0, \ldots, \eta_{L-1})$ and $(\zeta_0, \ldots, \zeta_{M-1})$ be enumerations of A^ℓ and A^m. Then

$$\mathbf{A} \models \forall \mathbf{y}.\, \psi(\mathbf{y}, \mathbf{z}, \mathbf{T}) \quad \text{iff} \quad \mathbf{A} \models \bigwedge_{q=0}^{L-1} \psi(\eta_q, \mathbf{z}, \mathbf{T}),$$

and

$$f(\mathbf{A}) = \sum_{s=0}^{M-1} \Big| \Big\{ \mathbf{T} : \bigwedge_{q=0}^{L-1} \psi_{q,s}(\mathbf{T}) \Big\} \Big|, \tag{4}$$

where $\psi_{q,s}(\mathbf{T})$ is obtained from $\psi(\eta_q, \zeta_s, \mathbf{T})$ by replacing every subformula that is true (resp., false) in $\mathbf{A}$ by TRUE (resp., FALSE). Now $\bigwedge_{q=0}^{L-1} \psi_{q,s}(\mathbf{T})$ is a CNF formula with propositional variables $T_i(\alpha_i)$ where $\alpha_i \in A^{t_i}$. Moreover, there is at most one occurrence of an unnegated propositional variable in each clause, and at most one of a negated variable. Thus, expression (4) already provides an AP-reduction to #1P1NSAT, since $f(A)$ is the sum of the numbers of satisfying assignments to M (i.e. polynomially many) instances of #1P1NSAT. (To obtain a precise correspondence we must add, in each instance, trivial clauses $T_i(\alpha_i) \Rightarrow T_i(\alpha_i)$ for every propositional variable $T_i(\alpha_i)$ not already occurring in $\bigwedge_{q=0}^{L-1} \psi_{q,s}(\mathbf{T})$, otherwise the number of models T will be underestimated by a factor 2^u where u is the number of unrepresented variables $T_i(\alpha_i)$.)

The above reduction is not yet parsimonious. To accomplish this, let us distinguish the variables in the above set of instances of #1P1NSAT as $T_i{}^s(\alpha_i)$ $(s = 0, 1, \ldots, M-1)$. Also, write $\Psi^s = \bigwedge_{q=0}^{L-1} \psi_{q,s}(\mathbf{T}^s)$ $(s = 0, 1, \ldots, M-1)$. We may assume that Ψ^s contains no one-literal clauses, since the truth setting of any such literal is forced, and the corresponding variable may be set to TRUE or FALSE. Let $w_1, w_2, \ldots, w_{M-1}$ be new propositional variables, and suppose $w_0 =$ FALSE, $w_M =$ TRUE for the sake of exposition. Let

$$\Phi^s = \bigwedge_{i=0}^{r-1} \bigwedge_{\alpha_i \in A^{t_i}} (\, T_i{}^s(\alpha_i) \Rightarrow w_{s+1}\,) \qquad (s = 0, 1, \ldots, M-2)$$

$$\text{and} \quad \Xi^s = \bigwedge_{i=0}^{r-1} \bigwedge_{\alpha_i \in A^{t_i}} (\, w_s \Rightarrow T_i{}^s(\alpha_i)\,) \qquad (s = 1, 2, \ldots, M-1),$$

and consider the formula

$$\varphi = \bigwedge_{s=0}^{M-1} \Psi^s \wedge \bigwedge_{s=0}^{M-2} \Phi^s \wedge \bigwedge_{s=1}^{M-1} \Xi^s.$$

Observe that φ is an instance of #1P1NSAT. We claim that it has exactly $f(A)$ satisfying assignments. To see this note that if, for a given s, $T_i{}^s(\alpha_i)$ is assigned TRUE for *some* i, then for all $p > s$, $T_j{}^p(\alpha_j)$ must be assigned TRUE for *every* j. This is forced by the Φ^s, Ξ^s formulae. Thus there can only be one s such that the ${T_i}^s(\alpha_i)$ receive both truth assignments. This is the unique s such that w_s is assigned FALSE and w_{s+1} is assigned TRUE. Any $s = 0, 1, \ldots, M-1$ is possible but, once it is fixed, it is easy to see that φ is satisfied if and only if Ψ^s is satisfied. The satisfying assignments are clearly disjoint for different s, and the claim follows. □

Corollary 12 *The problems* #BIS, $\#P_4$-COL, $\#P_q^*$-COL *(for $q \geq 3$, including as special cases* #2-PARTICLE-WR-CONFIGS *and* #BEACHCONFIGS*) and* #DOWNSETS *are all complete for* $\#\mathrm{RH}\Pi_1$ *with respect to AP-reducibility.*

Proof (sketch). Hardness is immediate from Theorems 5, 10 and 11. That each of the problems is in the class $\#\mathrm{RH}\Pi_1$ can be established by constructing suitable logical formulas along the lines of (3). Suppose we represent an instance of $\#P_q^*$-COL as a structure $\mathbf{A} = (A, \sim)$ where A is the vertex set and $\sim$ is a binary relation (assumed to represent adjacency). We can express the number of P_q^*-colourings as follows, where, for $1 \leq j < q$, the unary relation C_j is "true" for a vertex iff its colour is in $\{c_1, \ldots, c_j\}$.

$$
\begin{aligned}
f_{\mathrm{P}_q^*}(\mathbf{A}) = \bigl|\bigl\{(C_1, \ldots, C_{q-1}) : \mathbf{A} \models {} & \forall x \in A, y \in A. \\
& (C_1(x) \Rightarrow C_2(x)) \wedge \cdots \wedge (C_{q-2}(x) \Rightarrow C_{q-1}(x)) \wedge \\
& (C_1(x) \wedge x \sim y \Rightarrow C_2(y)) \wedge \cdots \wedge (C_{q-2}(x) \wedge x \sim y \Rightarrow C_{q-1}(y))\bigr\}\bigr|.
\end{aligned}
$$

We can represent an instance of #BIS as a structure $\mathbf{A} = (A, L, \sim)$, where A is the vertex set, L is the set of "left" vertices and $\sim$ is a binary relation (assumed to represent adjacency). We can express the number of independent sets as follows, where the unary relation X is "true" for left-vertices which are in the independent set, and for right-vertices which are not in the independent set.

$$
f_{\mathrm{BIS}}(\mathbf{A}) = \bigl|\bigl\{(X) : \mathbf{A} \models \forall x \in A, y \in A.\, L(x) \wedge x \sim y \wedge X(x) \Rightarrow X(y)\bigr\}\bigr|.
$$

□

Corollary 12 tells us something about the complexity class $\#\mathrm{RH}\Pi_1$. In particular, it is likely to be a *strict* subset of $\#\Pi_1$. Indeed, since $\#\mathrm{IS} \in \#\Pi_1$, $\#\mathrm{RH}\Pi_1 = \#\Pi_1$ would imply $\#\mathrm{IS} \leq_{\mathrm{AP}} \#\mathrm{BIS}$ (i.e., it would imply that #BIS is complete for #P with respect to AP-reducibility). It is clear that $\#\mathrm{RH}\Pi_1$ is not a subset of the previous class in the hierarchy of Saluja et al. In particular, $\#1\mathrm{P}1\mathrm{NSAT} \in \#\mathrm{RH}\Pi_1 \setminus \#\Sigma_1$ (this can be proved using arguments similar to those used by Saluja et al. to show that counting satisfying assignments of a 3CNF formula is not in $\#\Sigma_1$). Every problem in $\#\Sigma_1$ is (trivially) AP-reducible to the complete problems in $\#\mathrm{RH}\Pi_1$, but we do not know whether $\#\Sigma_1 \subseteq \#\mathrm{RH}\Pi_1$.

Clearly, Corollary 12 continues to hold even if "AP-reducibility" is replaced by a more stringent reducibility. In fact, most of our results remain true for more stringent reducibilities than AP-reducibility. One such tightening is to "restricted approximation-preserving reduction". The definition of *RAP-reduction* follows closely that of AP-reduction, but the operation of the Turing machine M

is greatly restricted. On input (x, ε), the machine M may make a single oracle call $(w, \delta) \in \Sigma^* \times \mathbb{R}^+$, and compute a positive rational $q \in \mathbb{Q}^+$ without recourse to the oracle. Suppose the result from the oracle call is $y \in \mathbb{N}$. Then the result returned by M is the integer closest to qy.

All the results based on *explicit* reductions in this article (not just Theorem 11 and Corollary 12) hold with "RAP-reducibility" replacing "AP-reducibility." The results that appeal to the bisection technique of Valiant and Vazirani [21] seem to require a more liberal notion of reducibility.

6 Problems to which #BIS is reducible

There are some problems that we have been unable to place in any of the three AP-interreducible classes considered in this article even though reductions from #BIS can be exhibited. The existence of such reductions may be considered as weak evidence for intractability, at least provisionally while the complexity status of the class $\#\mathrm{RH}\Pi_1$ is unclear. Two examples are #3-PARTICLE-WR-CONFIGS (the special case of #q-PARTICLE-WR-CONFIGS with $q = 3$) and #BIPARTITE q-COL:

Name. #BIPARTITE q-COL.

Instance. A bipartite graph B.

Output. The number of q-colourings of B.

Three observations concerning #BIPARTITE q-COL: (i) the special case $q = 2$ is trivially solvable; (ii) the special case $q = 3$ has an alternative characterisation as counting C_6-colourings of a general graph, where C_6 is the cycle on six vertices; and (iii) #BIPARTITE q-COL includes the q-state ferromagnetic Potts model as a special case. Observation (ii) follows from a similar argument to that used to relate #BIS and #P_4-COL in the proof of Theorem 5.

To interpret observation (iii), suppose G is a graph on n vertices, and set $q = 3$ (say). The configurations of the 3-state ferromagnetic Potts system based on G are the 3^n possible 3-colourings, not necessarily proper, of the graph G. Define the weight of a configuration σ to be $2^{m(\sigma)}$, where $m(\sigma)$ is the number of edges of G that are monochromatic under the 3-colouring σ. Consider the problem of computing the total weight of configurations: this is a simplified formulation of the problem of evaluating the partition function of the 3-state *ferromagnetic* Potts model at a certain non-zero temperature. The reduction of this (weighted) counting problem to #BIPARTITE3-COL is accomplished by mapping G to its "2-stretch," i.e., the graph G' obtained from G by subdividing each edge by a single additional vertex. An *antiferromagnetic* system is obtained by giving weight $\alpha^{m(\sigma)}$ to configuration σ, where $\alpha < 1$. Notice that (usual) graph colouring is obtained in the "zero temperature limit" as $\alpha \to 0$; notice also that an antiferromagnet (repulsive) Potts system on the bipartite graph G' effectively models a ferromagnetic (attractive) Potts system on the general graph G.

An intermediate problem that features in our reductions is:

Name. #BipartiteMaxIS.

Instance. A bipartite graph B.

Output. The number of *maximum* independent sets in B.

Theorem 13 #BIS *is AP-reducible to all three problems:* #BipartiteMaxIS, #3-Particle-WR-Configs *and* #Bipartite q-Col.

Proof. Follows from the reductions guaranteed by Lemmas 15, 16 and 17. □

The first of the three problems is actually AP-interreducible with #BIS, as the following lemma shows:

Lemma 14 #BipartiteMaxIS $\leq_{\mathrm{AP}}$ #BIS.

Proof. Since the maximum size, m, of an independent set in a bipartite graph can be determined in polynomial time, the reduction from the proof of Theorem 3 may be used. □

We now give the lemmas which we use to prove Theorem 13.

Lemma 15 #BIS $\leq_{\mathrm{AP}}$ #BipartiteMaxIS.

Proof. Let G be an instance of #BIS, with vertex set $\{v_0, \dots, v_{n-1}\}$. We construct an instance, G' of #BipartiteMaxIS as follows. The vertices of G' are $\{v_0, \dots, v_{n-1}\} \cup \{v'_0, \dots, v'_{n-1}\}$. The edges of G' are the edges of G together with $\{(v_i, v'_i)\}$. Now there is a bijection between the independent sets of G and the maximum independent sets of G'. □

Lemma 16 #BipartiteMaxIS $\leq_{\mathrm{AP}}$ #3-Particle-WR-Configs.

Proof. Let $B = (X, Y, A)$ be an instance of #BipartiteMaxIS, where $X = \{x_0, \dots, x_{n-1}\}$ and $Y = \{y_0, \dots, y_{n-1}\}$. Let M be the size of a maximum independent set in B. (Note that M can be determined from B in polynomial time.) Construct an instance $G = (V, E)$ of #3-Particle-WR-Configs as follows, where s and t are integers to be chosen below. Let $U_i : 0 \leq i \leq n-1$ be disjoint sets of size s, and $V_j : 0 \leq j \leq n-1$ be disjoint sets of size s. Further, let K be a set of size t. Then set

$$V = K \cup \bigcup_{i \in [n]} U_i \cup \bigcup_{j \in [n]} V_j$$

and

$$E = K^{(2)} \cup \bigcup_{j \in [n]} (V_j \times K) \cup \bigcup \{U_i \times V_j : (x_i, y_j) \in A\}.$$

Thus K is a clique, and there is a complete bipartite graph between $\bigcup_j V_j$ and K. An S_3^*-colouring corresponds to a colouring of G with colours b, r_1, r_2 and r_3 in which, for $\varrho \neq \pi$, there are no edges between vertices coloured r_ϱ and vertices coloured r_π. A colouring is *full* if, for some ϱ, K has vertices coloured b and r_ϱ (and no other colours). Every full colouring points out an independent set in B. The vertex y_j is in the independent set if V_j contains at least one vertex coloured r_ϱ. The vertex x_i is in the independent set if U_i contains at least one vertex whose colour is not b or r_ϱ. How many times does an independent set with k u_i's and ℓ v_i's come up (as a full colouring)?

$$\begin{aligned}
&3(2^t-2)(4^s-2^s)^k(2^s)^{n-k}(2^s-1)^\ell \\
&= 3(2^t-2)2^{sn}(2^s-1)^{k+\ell}.
\end{aligned}$$

Let $Z = 3(2^t-2)2^{sn}(2^s-1)^M$. Let N denote the number of maximum independent sets in B. We will say that a full colouring is *M-large* if the independent set that it points out has size M, and *M-small* otherwise. The number of M-small full colourings is at most

$$2^{2n}3(2^t-2)2^{sn}(2^s-1)^{M-1} \le \frac{2^{2n}Z}{2^s-1} \le Z/8,$$

if s is sufficiently large with respect to n. The number of non-full colourings is at most $4 \cdot 4^{2sn}$, which is at most $Z/8$ if t is sufficiently large with respect to s and n. Let Y denote the number of colourings. Then

$$N = \left\lfloor \frac{Y}{Z} \right\rfloor,$$

and the result follows. $\square$

Lemma 17 *For $q \ge 3$, #BIPARTITEMAXIS $\le_{\mathrm{AP}}$ #BIPARTITE q-COL.*

Proof. Let $B = (X, Y, A)$ be an instance of #BIPARTITEMAXIS, where $X = \{x_0, \dots, x_{n-1}\}$ and $Y = \{y_0, \dots, y_{n-1}\}$. Let M be the size of a maximum independent set in B. Construct an instance $G = (V, E)$ of #BIPARTITE q-COL as follows, where r, s and ℓ are integers to be chosen below. Let $U_i : 0 \le i \le n-1$ be disjoint sets of size r, and $V_i : 0 \le i \le n-1$ be disjoint sets of size s. Further, let I_1 be a set of size $(q-2)\ell$ and I_2 be a set of size 2ℓ. Let i_0 be a vertex that is not in any of these sets. Then set

$$V = \{i_0\} \cup I_1 \cup I_2 \cup \bigcup_{i \in [n]} U_i \cup \bigcup_{j \in [n]} V_j$$

and

$$\begin{aligned}
E = {} & (\{i_0\} \times I_1) \cup (I_1 \times I_2) \cup \bigcup_{i \in [n]} (\{i_0\} \times U_i) \cup \bigcup_{j \in [n]} (V_j \times I_1) \\
& \cup \bigcup \{U_i \times V_j : (x_i, y_j) \in A\}.
\end{aligned}$$

A q-colouring of G is *full* if exactly $q-2$ colours are used to colour the vertices in I_1. Every full colouring points out an independent set in B. Consider a full colouring in which blue is not used to colour any vertices in $I_1 \cup \{i_0\}$. Vertex x_i is in the independent set if U_i contains at least one blue vertex and vertex y_i is in the independent set if V_i contains at least one blue vertex. Let $f(a, b)$

denote the number of onto functions from a set of size a to a set of size b. Let $z = \lg((q-1)/(q-2))$. How many times does an independent set with k u_i's and j v_i's come up (as a full colouring)?

$$2\binom{q}{q-2} f((q-2)\ell, q-2)\, 2^{2\ell}\, (q-2)^{rn}\, (2^s-1)^{j+k} \left(\frac{2^{zr}-1}{2^s-1}\right)^k. \tag{5}$$

Let N denote the number of maximum independent sets in B. Let

$$Z = 2\binom{q}{q-2} f((q-2)\ell, q-2)\, 2^{2\ell}\, (q-2)^{rn}\, (2^s-1)^M.$$

As in the proof of Lemma 16, we wish to show that the total contribution of the non-full colourings is small. To this end, let

$$\varrho(y) = \binom{q}{y} f((q-2)\ell, y)\, (q-y)^{2\ell}.$$

$\varrho(y)$ is the number of colourings of $I_1 \cup I_2$ in which I_1 is coloured with exactly y colours. Thus, $\varrho(y) = 0$ unless $y \in \{1, \ldots, q-1\}$. We will choose ℓ to be sufficiently large that, for a positive constant c,

$$\varrho(q-2) \le \sum_{y=1}^{q-1} \varrho(y) \le \varrho(q-2)(1+\exp(-c\ell)). \tag{6}$$

(We will show later that equation (6) holds for an appropriate choice of ℓ.) Equation (6) implies that the total contribution of the non-full colourings is at most

$$\varrho(q-2)\exp(-c\ell) q^{1+rn+sn}.$$

If ℓ is at least a sufficiently large polynomial in q, n, r, and s then this is at most $\varrho(q-2)\exp(-c\ell/2)$ which is at most $Z/8$. As in the proof of Lemma 16, the number of M-small full colourings is also at most $Z/8$.

Let Y be the number of colourings. Now we are almost finished except that

1. we still need to show that equation (6) holds as long as ℓ is sufficiently large with respect to the constant q, and

2. unlike the situation in the proof of Lemma 16, the number of M-large full colourings is not precisely NZ. That is, we have ignored the extra factor of $\left(\frac{2^{zr}-1}{2^s-1}\right)^k$ in equation (5). To finish, we must show that the parameters r and s can be chosen such that for any $k \in [0, n]$

$$e^{-\varepsilon} \le \left(\frac{2^{zr}-1}{2^s-1}\right)^k \le e^{\varepsilon}, \tag{7}$$

where ε is a given accuracy parameter.

Now we show that equation (6) holds as long as ℓ is sufficiently large with respect to the constant q. In particular, we show that for sufficiently large ℓ there is a positive constant c such that for all $y \in \{1, \ldots, q-3, q-1\}$, we have $\varrho(y) \le \varrho(q-2)\exp(-c\ell)$.

First, consider $y \in \{1, \ldots, q-3\}$. In this case (as long as ℓ is at least $2(q-2)\ln(q-2)$), Lemma 18 and the definition of ϱ show that

$$\frac{\varrho(q-2)}{\varrho(y)} \ge \frac{\binom{q}{q-2}}{\binom{q}{y}} \left(\frac{q-2}{y}\right)^{(q-2)\ell} (1-\exp(-\ell/2)) \left(\frac{2}{q-y}\right)^{2\ell}.$$

If ℓ is sufficiently large then this is at least $\exp(c\ell)$, since

$$\begin{aligned}\left(\frac{q-2}{y}\right)^{(q-2)/2} &= \left(1+\frac{q-2-y}{y}\right)^{(q-2)/2} \ge 1 + \left(\frac{q-2}{2}\right)\left(\frac{q-2-y}{y}\right) \\ &= 1 + \left(\frac{q-2-y}{2}\right)\left(\frac{q-2}{y}\right) > 1 + \frac{q-2-y}{2} = \frac{q-y}{2}.\end{aligned}$$

Finally, consider $y = q-1$. As before,

$$\frac{\varrho(q-2)}{\varrho(q-1)} \ge \frac{\binom{q}{q-2}}{\binom{q}{q-1}} \left(\frac{q-2}{q-1}\right)^{(q-2)\ell} (1-\exp(-\ell/2))2^{2\ell}.$$

This is at least $\exp(c\ell)$, since

$$\left(\frac{q-1}{q-2}\right)^{q-2} = \left(1+\frac{1}{q-2}\right)^{q-2} < 2^2.$$

We now conclude the proof by showing that the parameters r and s can be chosen such that, for any $k \in [0,n]$ equation (7) holds. Note that we want r and s to be at most polynomial in n and ε^{-1}. Also, we must make s at least a sufficiently large multiple of n (say $1000n$) so that the number of M-small full colourings stays below $Z/8$. Let W be be a positive integer such that $\lfloor zW \rfloor$ is at least $1000n$. Let $R = \lceil (16(\ln 2)Wn)/(7\varepsilon) \rceil$. Finally, let $r = Wx$, where x is chosen from Corollary 20.

There are two cases. If $zr - \lfloor zr \rfloor \le W/R$ then we set $s = \lfloor zr \rfloor$. Otherwise, we set $s = \lceil zr \rceil$. To finish, we just need to show that equation (7) is satisfied either way. Let $\delta = \varepsilon/n$. For the first case,

$$(\ln 2)(zr - \lfloor zr \rfloor) \le (\ln 2)W/R \le 7\delta/16 \le \ln(1+\delta/2),$$

where the rightmost inequality relies on the fact that $\delta < 1/2$. Exponentiating both sides,

$$2^{zr} \le 2^{\lfloor zr \rfloor}(1+\delta/2) \le 2^{\lfloor zr \rfloor} + \delta(2^{\lfloor zr \rfloor} - 1).$$

Thus,

$$\frac{2^{zr} - 2^{\lfloor zr \rfloor}}{2^{\lfloor zr \rfloor} - 1} \le \delta.$$

Adding 1 to both sides,

$$\frac{2^{zr} - 1}{2^{\lfloor zr \rfloor} - 1} \le 1 + \delta \le e^{\delta}.$$

The second case is analogous.

□

Lemma 18 *If a and b are positive integers and $a \geq 2b \ln b$ then*

$$b^a\left(1 - \exp(-a/(2b))\right) \leq f(a,b) \leq b^a.$$

Proof. The right-hand inequality is straightforward, and the left-hand inequality can be derived as follows.

$$\begin{aligned} f(a,b) &\geq b^a - b(b-1)^a = b^a\left(1 - b\left(1 - \frac{1}{b}\right)^a\right) \\ &\geq b^a(1 - b\exp(-a/b)) = b^a\left(1 - \exp\left(-a\left(\frac{1}{b} - \frac{\ln b}{a}\right)\right)\right) \\ &\geq b^a\left(1 - \exp\left(\frac{-a}{2b}\right)\right). \end{aligned}$$

□

Lemma 19 *For any positive integer R there is an $x \in [1,\ldots,R]$ such that*

$$\min(zx - \lfloor zx \rfloor, \lceil zx \rceil - zx) \leq 1/R.$$

Proof. For $i \in [1,\ldots,R]$, let μ_i denote $zi - \lfloor zi \rfloor$. If there is an i such that $\mu_i \leq 1/R$ then take $x = i$. Otherwise, there are $i \neq j$ such that $0 \leq \mu_i - \mu_j \leq 1/R$, so take $x = |i-j|$. □

Corollary 20 *For any positive integer W and any positive integer R, there is an $x \in [1,\ldots,R]$ such that*

$$\min(zWx - \lfloor zWx \rfloor, \lceil zWx \rceil - zWx) \leq W/R.$$

7 An erratic sequence of problems

In this section, we consider a sequence of H-colouring problems. Let Wr_q be the graph with vertex set $V_q = \{a, b, c_1, \ldots, c_q\}$ and edge set

$$E_q = \{(a,b)\} \cup \{(b,b)\} \cup \bigcup_i \{(b,c_i)\} \cup \bigcup_i \{(c_i,c_i)\}.$$

Wr_0 is just K_2 with one loop added. Wr_1 is called "the wrench" in [2]. Consider the problem #q-Wrench-Col, which is defined as follows.

Name. #q-Wrench-Col.

Instance. A graph G.

Output. The number of Wr_q-colourings of G.

In this section, we prove the following theorem.

Theorem 21

- *For $q \leq 1$, #q-Wrench-Col is AP-interreducible with #Sat.*
- #2-Wrench-Col *is AP-interreducible with* #BIS.
- *For $q \geq 3$, #q-Wrench-Col is AP-interreducible with #Sat.*

Theorem 21 indicates that either (i) #BIS is AP-interreducible with #Sat (which would be surprising) or (ii) the complexity of approximately counting H-colourings is "non-monotonic": when H is chosen from a regularly constructed sequence, the complexity may jump down and then up again. Since Wr_0-colourings are independent sets, the theorem follows from Theorems 3 and 5 and Lemmas 15, 22, 23, 24 and 25. As starting points for our reductions, we will use the following problems.

Name. #LargeIS-Cubic.

Instance. A positive integer m and a connected cubic graph G in which every independent set has size at most m.

Output. The number of size-m independent sets in G.

Name. #LargeCut.

Instance. A positive integer k and a connected graph G in which every cut[12] has size at most k.

Output. The number of size-k cuts of G.

Garey et al. [7] have shown that the decision problems corresponding to these counting problems are NP-complete. Therefore, Theorem 1 shows that the counting problems are AP-interreducible with #Sat. A direct (nearly parsimonious) reduction from #Sat to #LargeIS-Cubic appears in Appendix A and a direct parsimonious reduction from #Sat to #LargeCut appears in [10].[13]

Lemma 22 #LargeCut $\leq_{\mathrm{AP}}$ #1-Wrench-Col.

[12] Recall that a "cut" of a graph is a partition of its vertex set into two subsets and that the size of the cut is the number of edges which span the two subsets.

[13] Recall that it was possible to modify the definition of #LargeIS so that a "witness" was provided along with the instance. Similarly, it is possible to modify the definitions of #LargeIS-Cubic and #LargeCut so that witnesses are provided along with the input. For example, a witness for #LargeCut could be used to check that the instance has no cuts of size exceeding k.

Proof. Let k and $G = (V, E)$ be an instance of #LARGECUT. Construct an instance $G' = (V', E')$ of #1-WRENCH-COL as follows, where the size of V is n and s and t are integers to be determined below. For every vertex u of G let A_u and A'_u be disjoint sets of size $2s$, let B_u and B'_u be disjoint sets of size $7s$, and let $V_u = A_u \cup B_u \cup B'_u \cup A'_u$. Let $B_u[i]$ denote the ith element of B_u. For every edge e of G let S_e and S'_e be disjoint sets of size t. Then set

$$V' = \left(\bigcup_{u \in V} V_u\right) \cup \left(\bigcup_{e \in E} S_e \cup S'_e\right)$$

and

$$E' = \left(\bigcup_{u \in V} A_u \times B_u \cup A'_u \times B'_u \cup \bigcup_{i \in \{1,\ldots,7s\}} \{(B_u[i], B'_u[i])\}\right)$$
$$\cup \left(\bigcup_{e=(u,v) \in E} B_u \times S_e \cup B'_v \times S_e \cup B'_u \times S'_e \cup B_v \times S'_e\right).$$

A wrench-colouring of G' is a colouring of the vertices of G' with colours g, b and r such that every neighbour of every colour-g vertex is coloured b. Thus, in a wrench-colouring of G', every edge is coloured with one of the six colourings (g, b), (b, g), (b, b), (b, r), (r, b) and (r, r). A wrench-colouring is *full* if, for every vertex u of G, the set of colourings assigned to edges between B_u and B'_u is either exactly $\mathcal{C}_1 = \{(g, b), (b, b), (b, r), (r, b), (r, r)\}$ or exactly $\mathcal{C}_2 = \{(b, g), (b, b), (r, b), (b, r), (r, r)\}$. Note that in the first case A_u is coloured b and A'_u has no g. In the second case, A'_u is coloured b and A_u has no g. Every full wrench-colouring points out a cut of G. The vertex u of G is in the left side of the partition in the first case and in the right side in the second case. Recall that $f(a, b)$ denotes the number of onto functions from a set of size a to a set of size b. How many times does a size-j cut come up (as a full colouring)?

$$2(f(7s, 5)2^{2s})^n 2^{jt}.$$

Let $Z = 2(f(7s, 5)2^{2s})^n 2^{kt}$. Let N denote the number of k-cuts. We say that a full colouring is k-large if the cut that it points out has size k and k-small otherwise. The number of k-small full colourings is at most $2^n Z/2^t$ which is at most $Z/8$ as long as $t \geq n + 3$. We conclude the proof by showing that the number of non-full colourings is at most $Z/8$. In particular, let $\mathcal{C}$ denote the set of colourings assigned to edges between B_u and B'_u. In each case (below) the number of colourings is exponentially smaller (as a function of s) than Z. In our calculations, we use Lemma 18 and we assume that s is sufficiently large compared to t, so we do not have to worry about any additional factor (up to $3^{2t\binom{n}{2}}$) which might arise due to having more possibilities for colouring vertices in S_e or S'_e (for any e).

1. $|\mathcal{C}| \geq 5$ but $\mathcal{C} \neq \mathcal{C}_1$ and $\mathcal{C} \neq \mathcal{C}_2$: A_u and A'_u are coloured b, so there are at most 6^{7s} possibilities for colouring the vertices in V_u, which is exponentially fewer than $f(7s, 5)2^{2s}$ (since $6^7 < 5^7 2^2$).

2. $|\mathcal{C}| = 4$: A_u and A'_u have no vertices with colour g, so there are at most $4^{7s}2^{2s}2^{2s}$ possibilities for colouring the vertices in V_u, which is exponentially fewer than $f(7s,5)2^{2s}$ (since $4^7 2^2 2^2 < 5^7 2^2$).

3. $|\mathcal{C}| \leq 3$: There are at most $3^{7s}3^{2s}3^{2s}$ possibilities for colouring the vertices in V_u, which is exponentially fewer than $f(7s,5)2^{2s}$ (since $3^7 3^2 3^2 < 5^7 2^2$).

□

Lemma 23 #2-Wrench-Col $\leq_{\mathrm{AP}}$ #Downsets.

Proof. Let $G = (V, E)$ be an instance of #2-Wrench-Col. Following the proof of Lemma 8, we construct an instance of #Downsets, a partial order on the $2n$-element set $V \times [2]$. For each edge (u, v) of G, we impose the relationships $(u, 0) \prec (v, 1)$ and $(v, 0) \prec (u, 1)$. Given a downset D and a vertex u of G, there are four possibilities for the set $D_u = D \cap \{(u, 0), (u, 1)\}$. These possibilities correspond to the four colours of an Wr_2-colouring of G. If $D_u = \{(u, 1)\}$ then u is mapped to vertex a of Wr_1 and if $D_u = \{(u, 0)\}$ then u is mapped to vertex b of Wr_1. Now there is a 1-1 correspondence between Wr_1-colourings of G and downsets in $(V \times [2], \prec)$. □

Lemma 24 #BipartiteMaxIS $\leq_{\mathrm{AP}}$ #2-Wrench-Col.

Proof. Similar to the proof of Lemma 16. □

Lemma 25 *For $q \geq 3$,* #LargeIS-Cubic $\leq_{\mathrm{AP}}$ #q-Wrench-Col.

Proof. Let m and G be an instance of #LargeIS-Cubic. Let n be the number of vertices of G. First, construct a graph G' from G. For every vertex u of G, let $V[u]$ be the graph with vertex set $\{u_1, u_2, u_3, u_4, u_5\}$ and edge set $\{(u_1, u_4), (u_2, u_4), (u_3, u_4), (u_1, u_5), (u_2, u_5), (u_3, u_5)\}$. G' will be constructed from the graphs $V[u]$ and from some additional edges. In particular, if v is the i'th smallest neighbour of u in G and u is the j'th smallest neighbour of v in G, then we add (u_i, v_j) to G'. Next, construct a graph G'' from G'. Let r be sufficiently large with respect to n and let $s = 1.1\,r$. Every vertex u_1, u_2, or u_3 in G' corresponds to an independent set in G'' of size r. Every vertex u_4 or u_5 in G' corresponds to an independent set in G'' of size s. Every edge of G' corresponds to a complete bipartite graph in G''.

A G'-colouring is is a colouring which maps each of the $5n$ vertices of G' to a non-empty subset of V_q in such a way that

1. if vertices α and β of G' are adjacent and the colour of α includes a then the colour of β is $\{b\}$, and

2. if vertices α and β of G' are adjacent and the colour of α includes c_i (for any $i \in \{1, \ldots, q\}$) then the colour of β is a subset of $\{b, c_i\}$.

We will say that a G'-colouring is "independent" if, for every vertex u of G either

1. u_1, u_2 and u_3 are coloured V_q and u_4 and u_5 are coloured $\{b\}$, or

2. u_1, u_2 and u_3 are coloured $\{b\}$ and u_4 and u_5 are coloured V_q.

There is a 1-1 correspondence between independent sets of G and independent G'-colourings. (u is in the independent set iff u_1 is coloured V_q.) Furthermore, every Wr_q-colouring of G'' points out a G'-colouring and every size-M independent set of G corresponds to $f(r,q+2)^{3M} f(s,q+2)^{2(n-M)}$ Wr_q-colourings of G'', where $f(x,y)$ denotes the number of onto functions from a set of size x to a set of size y, as in the proof of Lemma 17. Let N denote the number of size-m independent sets in G. Let Y denote the number of Wr_q-colourings of G''. We will say that an independent G'-colouring is "full" if the independent set that it points out has size m. Claim 3 (below) shows that if C is a non-full G'-colouring then the fraction of Wr_q-colourings of G'' which correspond to C is exponentially small (as a function of r). This implies that

$$N = \left\lfloor \frac{Y}{f(r,q+2)^{3m} f(s,q+2)^{2(n-m)}} \right\rfloor.$$

We say that a G'-colouring C is "exponentially unlikely" when the fraction of Wr_q-colourings of G'' which correspond to C is exponentially small (as a function of r). We now complete the proof of the lemma by proving Claims 1–3. In each case, the fact that the specified fraction is exponentially large in r follows from Lemma 18.

Claim 1 *If, in G'-colouring C, some, but not all, of the vertices in $V[u]$ are coloured $\{b, c_i\}$ (for some vertex u of G and some $i \in \{1,\ldots,q\}$) then C is exponentially unlikely.*

Proof of Claim 1.

1. Suppose that u_1 is coloured $\{b, c_i\}$ and both u_4 and u_5 are coloured $\{b\}$. Then the G'-colouring C' obtained by recolouring u_1 with V_q and all neighbours of u_1 with $\{b\}$ corresponds to a factor of $f(r,q+2)/f(r,2)^2$ more Wr_q-colourings of G'' than C. This factor is exponentially large in r since $q > 2$. If r is sufficiently large with respect to n then it exceeds the number of G'-colourings, so C is exponentially unlikely.

2. Suppose that u_1 and u_4 are coloured $\{b, c_i\}$ and u_5 is coloured $\{b\}$. Then the G'-colouring C' obtained by recolouring u_5 with $\{b, c_i\}$ corresponds to a factor of $f(s,2)$ more Wr_q-colourings of G'' than C.

3. Suppose that u_1 and u_4 and u_5 are coloured $\{b, c_i\}$ and u_3 is coloured $\{b\}$. Then the G'-colouring C' obtained by recolouring u_4 and u_5 with V_q and u_1, u_2 and u_3 with $\{b\}$ corresponds to a factor of $f(s,q+2)^2/(f(s,2)^2 f(r,2)^2)$ more Wr_q-colourings of G'' than C.

4. Suppose that u_4 is coloured $\{b, c_i\}$ and all of its neighbours are coloured $\{b\}$. Then the G'-colouring C' obtained by recolouring u_4 with V_q corresponds to a factor of $f(s,q+2)/f(s,2)$ more Wr_q-colourings of G'' than C.

By symmetry, these are the only cases.

Claim 2 *If, in G'-colouring C, some vertex of G' has a colour other than V_q or $\{b\}$, then C is exponentially unlikely.*

Proof of Claim 2. Suppose (for contradiction) that C is not exponentially unlikely and that it has a vertex z whose colour is not $\{b\}$ or V_q. z must have a neighbour with a colour other than $\{b\}$ (otherwise C would be exponentially unlikely, since exponentially more Wr_q-colourings correspond to the G'-colouring obtained from C by recolouring z with V_q). Since the colour of z is not $\{c_i\}$ (otherwise C would be exponentially unlikely), it must be $\{b, c_i\}$ (for some $i \in \{1, \ldots, q\}$). Now consider the connected component U' of G' which contains z and has every vertex coloured $\{b, c_i\}$. By Claim 1, this corresponds to a connected component U of G, of size, say, ℓ. We will show that that C is exponentially unlikely. First, suppose that the maximum degree of a vertex in the subgraph of G induced by U is less than three. In this case, obtain a G'-colouring C' from C by recolouring $\lceil \ell/2 \rceil$ of the vertices in U with V_q and the rest of them with $\{b\}$. C' corresponds to a factor of

$$\frac{f(r, q+2)^{3\lceil \ell/2 \rceil} f(s, q+2)^{2\lfloor \ell/2 \rfloor}}{f(r, 2)^{3\ell} f(s, 2)^{2\ell}}$$

more Wr_q-colourings of G'' than C. If the subgraph of G induced by U has maximum degree three then, since it is not equal to K_4 (otherwise it would be all of G), it has an independent set of size I of size at least $\ell/3$. (This follows from Brooks' theorem [3], which says that if a connected graph Γ is not a complete graph and has maximum degree $\Delta \geq 3$, then it is Δ-colourable.) Now obtain C' from C by re-colouring the vertices in U' to encode the independent set I. (That is, if a vertex u is in the independent set, colour u_1, u_2 and u_3 with V_q as before.) Since $f(r, q+2)^3 \geq f(s, q+2)^2$, C' corresponds to a factor of at least

$$\frac{f(r, q+2)^{(\ell/3)3} f(s, q+2)^{(2\ell/3)2}}{f(r, 2)^{3\ell} f(s, 2)^{2\ell}}$$

more Wr_q-colourings of G'' than C. This factor is exponentially large in r since $q > 2$.

Claim 3 *If G'-colouring C is not full then it is exponentially unlikely.*

Proof of Claim 3. Suppose (for contradiction) that C is not exponentially unlikely and that for some vertex u of G, some but not all of the vertices in $\{u_1, u_2, u_3\}$ have colour V_q. (By Claim 2, the others and u_4 and u_5 have colour $\{b\}$.) Then, C corresponds to exponentially fewer Wr_q-colourings of G'' (by a factor of $f(s, q+2)^2/f(r, q+2)^2$) than the G'-colouring C' obtained from C by recolouring u_4 and u_5 with V_q and u_1, u_2 and u_3 with $\{b\}$. If all of u_1, u_2 and u_3 have colour $\{b\}$ and C is not exponentially unlikely then u_4 and u_5 have colour V_q. Thus, if C is not exponentially unlikely, it is independent. As we saw before, the number of Wr_q-colourings of G'' corresponding to a size-M independent set of G is $f(r, q+2)^{3M} f(s, q+2)^{2(n-M)}$. Since $f(r, q+2)^3/f(s, q+2)^2$ is exponentially large as a function of r, C is also full. $\square$

Essentially the same reduction yields:

Lemma 26 *For $q \geq 4$, #LargeIS-Cubic $\leq_{\mathrm{AP}}$ #q-Particle-WR-Configs.*

Appendix A: A direct reduction from #Sat to #LargeIS

Garey et al. [7] present a (conventional) many-one/Karp reduction from 3-Sat (the decision version of #Sat restricted to formulas with three literals per clause) to LargeIS-Cubic (the decision version of #LargeIS-Cubic). Let $\varphi = C_1 \wedge \cdots \wedge C_r$ be an instance of 3-Sat in the variables $x_1, \ldots, x_n$. For convenience, assume that each variable x_i in φ occurs t_i times unnegated and (the same number) t_i times negated; such a formula will be called *balanced.* A cubic graph $G = G(\varphi)$ is constructed that has an independent set of size $m = r + \sum_i t_i = 5r/2$ iff φ is satisfiable. For each variable x_i there is a cycle of length $2t_i$. For each clause C_j there is a triangle (complete graph on three vertices or K_3); each vertex in the triangle stands for a particular literal in C_j. Thus the total number of vertices in G is $3r + \sum_i 2t_i = 6r$. Note that G is the complement of a m-partite graph, with $m = 5r/2$, so there is certainly no independent set of size greater than m. (Each variable-cycle contains t_i disjoint copies of K_2, and each clause-triangle is a K_3.)

To achieve an independent set of size m it is necessary to choose one of two possible independent sets of size t_i in each variable-cycle. Interpret one of these as $x_i = 0$ and the other as $x_i = 1$. Additional edges are added to G joining variable-cycles to clause-triangles. These are placed so as to allow a vertex in a clause-triangle to be included in an independent set of size m iff the corresponding literal is true. Notice that this can be achieved by a collection of edges which are pairwise vertex disjoint. Thus G is cubic. Refer to [7] for a more formal description of G.

The reduction as it stands is not parsimonious: each satisfying assignment in φ corresponds to $\prod_j \mu_j$ independent sets in G, where μ_j is the number of literals in C_j made true by the assignment. Rather than change Garey et al.'s construction, we instead massage the formula φ to avoid the problem just identified. Starting with an arbitrary CNF formula φ we first construct a 3-CNF formula φ' (i.e., one with three literals per clause) that has the same number of satisfying assignments as φ. Next, we construct from φ' positive integers r_1, r_2 and r_3 and a 3-CNF formula φ'' that

1. has the same number of satisfying assignments as φ',
2. is balanced, and
3. has the property that in every satisfying assignment, exactly r_1 clauses have one true literal, exactly r_2 clauses have two true literals, and exactly r_3 clauses have three true literals.

Thus the composite reduction $\varphi \mapsto \varphi' \mapsto \varphi'' \mapsto G(\varphi'')$ expands the solution set by a constant factor $2^{r_2}3^{r_3}$: not a parsimonious reduction, but the next best thing.

The transformation $\varphi \mapsto \varphi'$ is based on the equivalence of the two formulas

$$(a \vee b \Leftrightarrow x) \quad \text{and} \quad (a \vee b \vee \neg x) \wedge (a \vee \neg b \vee x) \wedge (\neg a \vee b \vee x) \wedge (\neg a \vee \neg b \vee x). \tag{8}$$

This enables us to introduce a new variable x and force it to be the disjunction of two existing variables a and b. In particular, for $k > 3$, a k-term clause $\ell_0 \vee \cdots \vee \ell_{k-1}$ may be rewritten $(\ell_0 \vee \cdots \vee \ell_{k-3} \vee x) \wedge (\ell_{k-2} \vee \ell_{k-1} \Leftrightarrow x)$, where x is a new variable, and then rewritten further as a five-clause CNF formula using (8). By iterating this process we may efficiently transform an arbitrary CNF formula φ into a 3-CNF formula φ'. The transformation is clearly parsimonious.

Let r' denote the number of clauses in φ'. Let $r_1 = 31r'$, $r_2 = 16r'$, and $r_3 = 5r'$. We will construct φ'' using the equivalence of $(a \vee b \Leftrightarrow x)$ and the following expression, which is a balanced version of (8):

$$\begin{aligned}&(a \vee b \vee \neg x) \wedge (a \vee \neg b \vee x) \wedge (\neg a \vee b \vee x) \wedge (\neg a \vee \neg b \vee x)\\ &\quad \wedge (x \vee \neg x \vee \neg x) \wedge (x \vee \neg x \vee \neg x).\end{aligned} \tag{9}$$

Suppose that $a \vee b \vee c$ is a clause of φ'. Let $y, z_1, z_2, z_3, z_4, z_5, z_6$ be new variables. The clause $a \vee b \vee c$ of φ' will be transformed into the following equivalent expression. Note that y has the same truth value as $a \vee b \vee c$, and it is required by the expression to be TRUE. Also, the variables z_1–z_6 are only there to establish the third property required of φ''. Their values are "ignored".

$$\begin{aligned}&(a \vee b \Leftrightarrow x) \wedge (a \vee \neg b \Leftrightarrow z_1) \wedge (\neg a \vee b \Leftrightarrow z_2) \wedge (\neg a \vee \neg b \Leftrightarrow z_3)\\ &\quad \wedge (x \vee c \Leftrightarrow y) \wedge (x \vee \neg c \Leftrightarrow z_4) \wedge (\neg x \vee c \Leftrightarrow z_5) \wedge (\neg x \vee \neg c \Leftrightarrow z_6)\\ &\quad \wedge (y \vee y \vee y) \wedge (y \vee \neg y \vee \neg y) \wedge (y \vee \neg y \vee \neg y) \wedge (y \vee \neg y \vee \neg y).\end{aligned} \tag{10}$$

Each of the first eight clauses in (10) is further transformed using (9). The reader may verify that the resulting 52-clause expression

1. has exactly one satisfying assignment if $a \vee b \vee c$ is TRUE and none otherwise (regarding the truth assignments to a, b and c as fixed),

2. is balanced, and

3. has the property that for every satisfying assignment, exactly 31 clauses have one true literal, exactly 16 clauses have two true literals, and exactly 5 clauses have three true literals.

This completes the construction of φ''.

Appendix B: A glossary of problems

As an aid to navigation, Table 1 contains a complete list of problems considered in this article, with their complexity status and a note of where to find them.

Problem name	Def'd in	Status	Refer to
#BeachConfigs	§4	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 5
#Bipartite q-Col	§6	$\geq_{\mathrm{AP}}$ #BIS	Thm. 13
#P_4-Col	§4	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 5
#P_q^*-Col	§4	$\equiv_{\mathrm{AP}}$ #BIS ($q \geq 3$)	Thm. 10
#q-Wrench-Col	§7	$\equiv_{\mathrm{AP}}$ #Sat ($q \leq 1$)	Thm. 21
#2-Wrench-Col	§7	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 21
#q-Wrench-Col	§7	$\equiv_{\mathrm{AP}}$ #Sat ($q \geq 3$)	Thm. 21
#Downsets	§4	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 5
#IS	§3	$\equiv_{\mathrm{AP}}$ #Sat	Thm. 3
#BIS	§1	(primal)	Thm. 5
#LargeIS-Cubic	§7	$\equiv_{\mathrm{AP}}$ #Sat	App. A
#LargeIS	§3	$\equiv_{\mathrm{AP}}$ #Sat	Obs. 2
#BipartiteMaxIS	§6	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 13, Lem. 14
#Match	§2	FPRAS	[9]
#Sat	§1	(primal)	Section 3
#DNF-Sat	§2	FPRAS	[12]
#1p1nSat	§4	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 5
#2-Particle-WR-Configs	§3	$\equiv_{\mathrm{AP}}$ #BIS	Thm. 5
#3-Particle-WR-Configs	§3	$\geq_{\mathrm{AP}}$ #BIS	Thm. 13
#q-Particle-WR-Configs	§3	$\equiv_{\mathrm{AP}}$ #Sat ($q \geq 4$)	Lemma 26

Table 1: A list of counting problems

Acknowledgements

Lemma 22 is due to Mike Paterson. We thank Dominic Welsh for telling us about reference [16] and Marek Karpinski for stimulating discussions on the topic of approximation-preserving reducibility.

References

[1] G. Brightwell, O. Häggström and P. Winkler, Nonmonotonic behavior in hard-core and Widom-Rowlinson models, *Journal of Statistical Physics* **94** (1999), 415–435.

[2] G.R. Brightwell and P. Winkler, Graph homomorphisms and phase transitions, *CDAM Research Report, LSE-CDAM-97-01*, Centre for Discrete and Applicable Mathematics, London School of Economics, October 1997.

[3] R.L. Brooks, On colouring the nodes of a network, *Proceedings of the Cambridge Philosophical Society* **37** (1941) 194–197.

[4] M. Dyer, A. Frieze and M. Jerrum, On counting independent sets in sparse graphs, *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science* (FOCS'99), IEEE Computer Society Press, 1999, 210–217.

[5] M. Dyer and C. Greenhill, The complexity of counting graph homomorphisms, To appear in *Random Structures and Algorithms.* (A preliminary version appeared in *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms* (2000) 246–255.)

[6] R. Fagin, Generalized first-order spectra and polynomial time recognisable sets. In "Complexity of Computation" (R. Karp, ed.), *SIAM-AMS Proceedings* **7**, 1974, 43–73.

[7] M.R. Garey, D.S. Johnson and L. Stockmeyer, Some simplified NP-complete graph problems, *Theoretical Computer Science* **1** (1976) 237–267.

[8] O. Goldreich, Introduction to Complexity Theory, *Lecture Notes Series of the Electronic Colloquium on Computational Complexity*, 1999, Chapter 10: "#P and approximating it."
`http://www.eccc.uni-trier.de/eccc-local/ECCC-LectureNotes/`

[9] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-hard Problems* (Dorit Hochbaum, ed.), PWS, 1996, 482–520.

[10] M. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, *SIAM Journal on Computing* **22** (1993) 1087–1116.

[11] M.R. Jerrum, L.G. Valiant and V.V. Vazirani, *Random generation of combinatorial structures from a uniform distribution*, Theoretical Computer Science **43** (1986), 169–188.

[12] R.M. Karp, M. Luby and N. Madras, Monte-Carlo approximation algorithms for enumeration problems, *Journal of Algorithms* **10** (1989), 429–448.

[13] P.W. Kasteleyn, Graph theory and crystal physics. In *Graph Theory and Statistical Physics* (F. Harary, ed.), Academic Press, 1967, 43–110.

[14] J.S. Provan and M.O. Ball, The complexity of counting cuts and of computing the probability that a graph is connected, *SIAM Journal on Computing* **12** (1983), 777–788.

[15] D. Roth, On the Hardness of approximate reasoning, *Artificial Intelligence Journal* **82** (1996), 273–302.

[16] S. Saluja, K.V. Subrahmanyam and M.N. Thakur, Descriptive complexity of #P functions, *Journal of Computer and Systems Sciences* **50** (1995), 493–505.

[17] J. Simon, On the difference between one and many (Preliminary version), *Proceedings of the 4th International Colloquium on Automata, Languages and Programming* (ICALP), Lecture Notes in Computer Science **52**, Springer-Verlag, 1977, 480–491.

[18] A. Sinclair, *Algorithms for random generation and counting: a Markov chain approach*, Progress in Theoretical Computer Science, Birkhäuser, Boston, 1993.

[19] L. Stockmeyer, The complexity of approximate counting (preliminary version), *Proceedings of the 15th ACM Symposium on Theory of Computing* (STOC'83), ACM, 1983, 118–126.

[20] S. Toda, PP is as hard as the polynomial-time hierarchy, *SIAM Journal on Computing* **20** (1991), 865–877.

[21] L.G. Valiant and V.V. Vazirani, *NP is as easy as detecting unique solutions*, Theoretical Computer Science **47** (1986), 85–93.