Dominating sets of random 2-in 2-out directed graphs

Stephen Howe*

Department of Mathematics and Statistics
University of New South Wales
Sydney, NSW 2052, Australia
stephenh@maths.unsw.edu.au

Submitted: Aug 8, 2007; Accepted: Jan 30, 2008; Published: Feb 11, 2008
Mathematics Subject Classifications: 05C80, 05C69, 05C20

Abstract

We analyse an algorithm for finding small dominating sets of 2-in 2-out directed graphs using a deprioritised algorithm and differential equations. This deprioritised approach determines an a.a.s. upper bound of $0.39856n$ on the size of the smallest dominating set of a random 2-in 2-out digraph on n vertices. Direct expectation arguments determine a corresponding lower bound of $0.3495n$.

1 Introduction

A *directed multigraph* G is a set $V = V(G)$ of vertices with a multiset $E = E(G) \subseteq V \times V$ of (directed) edges. When E contains no repeated edges and no loops (edges of the form (v, v) for some $v \in V$) we say that G is *simple* and call G a *directed graph* or *digraph*. The *in-degree* of a vertex $u \in V$ is the number of edges of the form (v, u) for some $v \in V$; the *out-degree* of u is the number of edges of the form (u, v) for some $v \in V$. We consider only directed multigraphs (simple or otherwise) for which every vertex has in-degree 2 and out-degree 2. Such graphs are called *2-in 2-out* or *2-regular*.

A random 2-in 2-out digraph (on n vertices) is a digraph chosen uniformly at random from the set of all 2-in 2-out digraphs on n vertices. Often the probability of a random graph having a certain property, such as being connected, tends to 1 as n tends to infinity. In this case we say that a random graph has such a property *asymptotically almost surely* (a.a.s.). For example, a.a.s. a random 2-in 2-out digraph is connected [3].

In [5] Duckworth and Wormald determined a.a.s. upper and lower bounds for dominating sets of random cubic graphs. We determine similar bounds for random 2-in 2-out digraphs.

*Research supported by the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (MASCOS).

A *dominating set* of a digraph G is a subset $D \subseteq V(G)$ of the vertices such that for every vertex $v \in V(G)$, either $v \in D$ or for some $u \in D$ the edge (u, v) is present in G. If we change (u, v) to (v, u) in the above definition, then we define an *absorbent* set of G. The results of this paper, stated in Theorem 1.3, also hold for absorbent sets.

Dominating sets of small cardinality are the most interesting. For a general digraph, finding a minimum dominating set is NP-hard (which follows from a simple reduction from the undirected case). Some approximation results can be found in [2]. For example, for digraphs with in-degree bounded by a constant B, it is NP-hard to approximate the size of the minimum dominating set to within a constant less than $B - 1$ for $B \geq 3$ and 1.36 for $B = 2$ ([2] Theorem 10).

Other results about domination in digraphs can be found in [8] and [10]. Of particular interest are the following bounds on the minimum size of a dominating set of an arbitrary digraph on n vertices.

Theorem 1.1 ([8, 10]). *Let G be a digraph on n vertices.*

(i) If G has minimum in-degree $\delta \geq 1$ then the minimum size of a dominating set in G is less than

$$\frac{\delta + 1}{2\delta + 1} n + 1.$$

(ii) If G has maximum out-degree Δ then the minimum size of a dominating set in G is greater than

$$\frac{1}{1 + \Delta} n.$$

Theorem 1.1 can be found in [8] as Theorem 15.49 and Theorem 15.57; part (i) is originally from [10].

By Theorem 1.1, for 2-in 2-out digraphs the minimum size of a dominating set is bounded below by $n/3$ and above by $3n/5$. We will significantly improve these bounds for random 2-in 2-out digraphs (see Theorem 1.3).

As far as we are aware, dominating sets of random regular digraphs have not been studied. However domination has been studied in other models of random digraphs. Consider the following model: start with n vertices and for each pair of vertices (u, v) independently include (u, v) as an edge with probability p (for some $p \in [0, 1]$). We denote this model by $\mathcal{DG}_{n,p}$. Lee obtained the following result.

Theorem 1.2 ([10]). *Fix p with $0 < p < 1$ and let $k = \log n - 2 \log \log n + \log \log e$ where $\log$ denotes the logarithm to base $1/(1 - p)$. Then the minimum size of a dominating set of a random digraph $G \in \mathcal{DG}_{n,p}$ is a.a.s. $\lfloor k + 1 \rfloor$ or $\lfloor k + 2 \rfloor$.*

We study dominating sets in random 2-in 2-out digraphs using two techniques: by considering an algorithm for finding dominating sets of small cardinality and using direct expectation arguments. The algorithm, called DominatingSet, is described in Section 2. In Sections 3, 4, and 5 we approximate DominatingSet by another algorithm, known as a deprioritised algorithm. The behaviour of the deprioritised algorithm is then described by solutions to a certain system of differential equations. This analysis, which we call the *deprioritised approach*, was initially introduced by Wormald in [15]. The deprioritised approach determines the upper bound of the next theorem; the lower bound comes from the direct expectation arguments which are described in Section 6.

Theorem 1.3. *Asymptotically almost surely the minimum size of a dominating set of a random 2-in 2-out digraph is less than* $0.39856n$ *and greater than* $0.3495n$.

Previously, similar work has found bounds for independent dominating sets [6] and vertex and edge packing [1] on random regular graphs. We are not aware of any previous work applying the deprioritised approach to directed graphs. In [6] and [1] Theorem 2 of [15] was used. However this theorem cannot be applied for all algorithms on random regular graphs, for example [12] and [4]. Nor is it applicable for DominatingSet (and many other algorithms on random regular digraphs). A justification of this is given just before Section 3.1.

Further useful definitions and results about random graphs in general can be found in [9]. When working with probabilities, we use $\mathbb{P}(A)$ to denote the probability of the event A occurring and $\mathbb{E}(X)$ to denote the expected value of a random variable X.

2 Finding Small Dominating Sets

We start with some useful notations and definitions. An edge $(u, v) \in E(G)$ is called an *edge from* u *to* v; we also say that u *dominates* v. Given a vertex u, vertices v such that $(v, u) \in E(G)$ are called *in-neighbours* of u. Thus the in-degree of a vertex u is the number of in-neighbours of u. Out-neighbours are defined similarly.

The pair (p, q) where p is the in-degree of u and q is the out-degree of u is called the *degree pair* of u. A vertex with degree pair $(0, 0)$ is called *isolated* while a vertex with degree pair $(2, 2)$ is called *saturated*. Finally let $V_{(i,j)} = V_{(i,j)}(G)$ be the set of vertices of G with degree pair (i, j).

Now a dominating set for a given 2-in 2-out digraph G can be found by the following algorithm. We set $H := G$ and let $\mathcal{D}$ be empty. While $\mathcal{D}$ is empty or there are vertices of degree pair $(0, 1)$, $(0, 2)$, or $(1, 2)$ in H: select a vertex v uniformly at random from $V_{(p,q)}$ where

$$(p, q) = \min\{(i, j) \,:\, (i, j) \in \{(0, 1), (0, 2), (1, 2), (2, 2)\} \text{ and } V_{(i,j)}(H) \neq \emptyset\}.$$

Here degree pairs are ordered lexicographically. After selecting v, remove the edges of H incident with vertices dominated by v (in H) and then remove the edges incident with v. Then add v to $\mathcal{D}$ as well as any newly isolated vertices of H that are not dominated by v in G. When $\mathcal{D} \neq \emptyset$ and there are no more edges of degree pairs $(0, 1)$, $(0, 2)$, and $(1, 2)$, add any remaining non-isolated vertices to $\mathcal{D}$. Then $\mathcal{D}$ is a dominating set for G.

In order to obtain results about 2-in 2-out digraphs we analyse the algorithm DominatingSet given below. DominatingSet is based on the algorithm described above but, instead of taking a random 2-in 2-out digraph as input, DominatingSet *constructs* a random 2-in 2-out digraph along with a dominating set. To do so we use the *pairing* or *configuration* model which we describe next.

2.1 Generating Random 2-in 2-out Digraphs Uniformly

We generate a random 2-in 2-out directed multigraph (on the n vertices $v_1, \ldots, v_n$) with the pairing model as follows. For each vertex v_i we associate two in-points and two out-points. A bijection P from the set of $2n$ in-points to the set of $2n$ out-points is called a *pairing*. If P is only a partial function (from the in-points to the out-points) but still one-to-one then we call P a *partial pairing*. In both cases, a *pair* of P is an in-point a and an out-point b such that $P(a) = b$.

Now, from a given pairing P we construct a directed multigraph $G(P)$ (on $v_1, \ldots, v_n$); for each in-point a in a pair of P we add to (the multiset) $E(G(P))$ the edge (v_i, v_j) such that the out-point $P(a)$ is associated with v_i and the in-point a is associated with v_j. By construction $G(P)$ will be 2-in 2-out.

Selecting a pairing P uniformly at random we obtain a random directed multigraph $G(P)$. Although $G(P)$ is not distributed uniformly, by conditioning on $G(P)$ having no loops or repeated edges, we obtain a simple 2-in 2-out digraph uniformly at random. The probability that $G(P)$ is simple is bounded below by a constant, see Theorem 4.6 of [11]. Thus a property holding a.a.s. for random directed multigraphs generated by the pairing model, also holds a.a.s. for random 2-in 2-out digraphs.

The pairing model also allows us to use a random process to generate random 2-in 2-out directed multigraphs. Start with an empty partial pairing P where no in-point is mapped to any out-point. At each step of the process we extend the definition of P by one pair in the following way: select an in-point a, from the in-points not in the domain of P, and an out-point b, from the out-points not in the range of P, where a or b is selected uniformly at random; then extend the definition of P so that P maps a to b. The point not selected uniformly at random may be selected in any way we like. The process stops when P becomes a pairing. We call such a process a *random partial pairing process* and the resulting random pairing is distributed uniformly.

When we extend a partial pairing to map a to b we say we are *exposing* a pair (in particular, the pair corresponding to a and b), or exposing an in-point, or just exposing a point (when the pair corresponding to the in-point or point is clear from the context). Points that are not in the domain and not in the range of P are called *free*.

DominatingSet will expose pairs one at a time by determining one point of the next pair to be exposed. At the same time, vertices are added to a set $\mathcal{D}$ which will be a dominating set when the algorithm finishes. In this way DominatingSet generates a 2-in 2-out directed multigraph $G(P)$ (for some pairing P) and a dominating set for $G(P)$.

2.2 The Algorithm DominatingSet

Algorithms 1 and 2 define DominatingSet and its auxiliary algorithm Saturate. We will view DominatingSet as a sequence of operations where each operation involves selecting the vertex u, adding u to $\mathcal{D}$, and then calling Saturate with u. Let

$$P_0 \subset P_1 \subset \cdots \subset P_F$$

be the subsequence of the random partial pairing process defined by DominatingSet such that P_0 is the empty partial pairing, P_F is a pairing, and P_{t+1} is obtained from P_t by performing an operation. From this sequence we obtain a corresponding sequence $\{G_t\}_{t=0}^{F}$ of directed multigraphs where $G_t = G(P_t)$. We analyse DominatingSet using the random variables $Y_{(i,j)}(G_t) = \left|V_{(i,j)}(G_t)\right|$ and $D(G_t) = |\mathcal{D}(G_t)|$.

During each operation, some vertex v is added to $\mathcal{D}$ and the free points associated with v, and the free points associated with the out-neighbours of v are exposed by a call to Saturate. When all the in-points and out-points associated with a vertex v are exposed then v has been saturated. Any vertices other than v and its out-neighbours that become saturated are called *accidental saturates*. By adding accidental saturates to $\mathcal{D}$, after each operation, all saturated vertices are either in $\mathcal{D}$ or are dominated by a vertex in $\mathcal{D}$.

DominatingSet finishes when there are no vertices of degree pairs $(1,0)$, $(2,0)$, or $(2,1)$. By equating the sum of the in-degrees with the sum of the out-degrees, every vertex in the final graph G_F has degree pair $(0,0)$, $(1,1)$, or $(2,2)$. We complete the graph G_F to a 2-in 2-out digraph by calling Saturate on the remaining unsaturated vertices. This will add a subset of $V_{(1,1)}(G_F) \cup V_{(0,0)}(G_F)$ to $\mathcal{D}$. So $D(G_F) + Y_{(0,0)}(G_F) + Y_{(1,1)}(G_F)$ will be an upper bound on the smallest size of a dominating set for any 2-in 2-out digraph containing G_F as a subgraph. Note though, that we expect (but don't prove) that a.a.s. G_F has no vertices of degree pair $(0,0)$ or $(1,1)$.

3 The Differential Equations Method

As mentioned above, we view DominatingSet as a sequence of operations. Each operation involves selecting a vertex u uniformly at random from the vertices of a given degree pair, adding u to the dominating set, and then saturating u and its out-neighbours. We say that the operation *processes* the vertex u. There are four types of operations, given in Table 1, and the types depend solely on the degree pair of u. We also say that vertex v is of type k if the degree pair of v is associated with an operation of type k.

DominatingSet is a prioritised algorithm in the sense that the type of each operation is chosen deterministically. Such algorithms on undirected graphs have been analysed in [13] and [5]. Analysing prioritised algorithms on graphs is difficult and remains so for algorithms on digraphs. Wormald in [15] introduced the idea of deprioritised algorithms which are easier to analyse. These algorithms use the same operations as the prioritised algorithm but choose the type of operation to perform according to a probability distribution. We are free to choose this probability distribution however we like. With an appropriate choice the deprioritised algorithm will approximate the prioritised algorithm.

Algorithm 1 DominatingSet

Recall that $V_{(i,j)} = V_{(i,j)}(G(P))$ and $Y_{(i,j)} = \left|V_{(i,j)}\right|$
Set P to be the empty partial pairing;
Pick u uniformly at random from $V_{(0,0)}$;
$\mathcal{D} := \{u\}$;
Saturate(u);
while $Y_{(1,0)} + Y_{(2,0)} + Y_{(2,1)} \neq 0$ **do**
  if $Y_{(2,1)} \neq 0$ **then**
    Pick u uniformly at random from $V_{(2,1)}$;
  else if $Y_{(2,0)} \neq 0$ **then**
    Pick u uniformly at random from $V_{(2,0)}$;
  else
    Pick u uniformly at random from $V_{(1,0)}$;
  end if
  $\mathcal{D} := \mathcal{D} \cup \{u\}$;
  Saturate(u);
end while
return $\mathcal{D}$ and P;

Algorithm 2 Saturate(u)

Expose the free points associated with u;
Expose the free points associated with each out-neighbour of u in $G(P)$;
Add accidental saturates to $\mathcal{D}$;

Degree pair	Type
$(0,0)$	0
$(1,0)$	1
$(2,0)$	2
$(2,1)$	3

Table 1: Types of operations and vertices.

Algorithm 3 The deprioritised version of DominatingSet

Require: : $\epsilon > 0$ is given and sufficiently small.
  Set P to be the empty partial pairing;
  $\mathcal{D} := \emptyset$;
  for $i = 1, \ldots, \lfloor \epsilon n \rfloor$ **do**
    Pick u uniformly at random from $V_{(0,0)}$;
    $\mathcal{D} := \mathcal{D} \cup \{u\}$;
    Saturate(u);
  end for
  while $Y_{(1,0)} + Y_{(2,0)} + Y_{(2,1)} \neq 0$ **do**
    Set p_i for $i = 1, 2, 3$ as defined in Section 4.6;
    Choose a operation type k according to the distribution $\mathbb{P}(k = i) = p_i$;
    Choose u uniformly at random from the vertices of type k in $G(P)$;
    $\mathcal{D} := \mathcal{D} \cup \{u\}$;
    Saturate(u);
  end while
  return $\mathcal{D}$ and P;

The deprioritised version of DominatingSet is given in Algorithm 3. The for loop is called the preprocessing phase; it is required for reasons explained in Section 5. The probabilities p_1, p_2, and p_3 are derived in Section 4.6. Here we note the main difference between using the deprioritised approach on directed and undirected graphs and why the theorems of [15] are not applicable. For most of the algorithms that have been studied on undirected graphs, the type of operation to perform (except during the preprocessing phase) has been randomly selected from two possible types while we select from three possible types.

3.1 The Differential Equations Theorem

We analyse the deprioritised version of DominatingSet with Theorem 3.1 given below. A detailed introduction to this theorem can be found in [14]. Using Theorem 3.1 we show that, until near the very end of the algorithm, a.a.s. the scaled variables $Y_{(i,j)}(G_t)/n$ and $D(G_t)/n$ are approximated by the solutions $z_{(i,j)}(t/n)$ and $z(t/n)$ to some set of differential equations. The differential equations will be determined, in the next section, using the expected change in $Y_{(i,j)}$ and D due to an operation.

Before stating the theorem we need a few definitions. Let $S^{(n)}$ be the set of all possible partial and complete pairings for a 2-in 2-out digraph on n vertices. A *history* $h_t^{(n)}$ of the process after t time units is a sequence $h_t^{(n)} = (q_0^{(n)}, \ldots, q_t^{(n)})$ where $q_i^{(n)} \in S^{(n)}$ for all $i = 0, 1, \ldots, t$. Let $S^{(n)+}$ denote all the possible histories of the process after t time units for $t = 0, 1, \ldots$ and let $H_t^{(n)}$ be the history of a given run of the process over t time units. Since we are interested in the asymptotic behaviour of the process as n tends to infinity, we often drop n from the notation.

Let $Y_1, \ldots, Y_a$ be random variables defined on a random process $G_0, \ldots, G_T$. Given a domain $W \subseteq \mathbb{R}^{a+1}$, we define the *stopping time* T_W to be the minimum t such that

$$(t/n, Y_1(t)/n, \ldots, Y_a(t)/n) \notin W.$$

A function $f : \mathbb{R}^m \to \mathbb{R}$ is *Lipschitz on* W (for $W \subseteq \mathbb{R}^m$) with Lipschitz constant L if, for L a positive constant, for all $\mathbf{x}$ and $\mathbf{y}$ in W,

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L \max_{1 \leq i \leq m} |x_i - y_i|.$$

The function $\|\cdot\|$ defined by $\|\mathbf{x}\| = \max_{1 \leq i \leq n} |x_i|$ is the ℓ^∞ norm.

Finally, a sequence of functions f_n *uniformly converges* to a function f for $x \in X$ if, for every $\epsilon > 0$, there exists an N such that

$$|f(x) - f_n(x)| < \epsilon$$

for all $x \in X$ and all $n > N$. Now we are ready to state Theorem 3.1 (which appears as Theorem 5.1 in [14]).

Theorem 3.1 ([14]). *For $1 \leq \ell \leq a$ with a fixed, let $y_\ell : S^{(n)+} \to \mathbb{R}$ and $f_\ell : \mathbb{R}^{a+1} \to \mathbb{R}$, such that for some constant C_0 and all ℓ, we have $|y_\ell(h_\ell)| < C_0 n$ for all $h_\ell \in S^{(n)+}$ and for all n. Let $Y_\ell(t)$ denote the random counterpart of $y_\ell(h_\ell)$. Assume the following three conditions hold where W is a bounded connected open set containing the closure of*

$$\{(0, z_1, \ldots, z_a) \mid \mathbb{P}(Y_\ell(0) = z_l n,\ 1 \leq \ell \leq a) \neq 0 \text{ for some } n\}.$$

(i) (Boundedness Hypothesis) For some functions $\beta = \beta(n) \geq 1$ and $\gamma = \gamma(n)$, the probability that

$$\max_{1 \leq \ell \leq a} |Y_\ell(t+1) - Y_\ell(t)| \leq \beta$$

conditional upon H_l, is at least $1 - \gamma$ for $t < T_W$.

(ii) (Trend Hypothesis) For some function $\lambda_1 = \lambda_1(n) = o(1)$, for all $1 \leq \ell \leq a$,

$$|\mathbb{E}(Y_\ell(t+1) - Y_\ell(t) \mid H_\ell) - f_\ell(t/n, Y_1(t)/n, \ldots, Y_a(t)/n)| \leq \lambda_1$$

for $t < T_W$.

(iii) (Lipschitz Hypothesis) Each function f_ℓ is continuous and satisfies a Lipschitz condition on

$$W \cap \{(t, z_1, \ldots, z_a) \mid t \geq 0\}$$

with the same Lipschitz constant for each ℓ.

Then the following are true:

(a) For $(0, \hat{z}_1, \ldots, \hat{z}_a) \in W$ the system of differential equations

$$\frac{dz_\ell}{dx} = f_\ell(x, z_1, \ldots, z_a) \text{ for } \ell = 1, \ldots, a$$

has a unique solution in W for $z_\ell : \mathbb{R} \to \mathbb{R}$ such that $z_\ell(0) = \hat{z}_\ell$ for $1 \leq \ell \leq a$ and which extends to points arbitrarily close to the boundary of W.

(b) Let $\lambda > \lambda_1 + C_0 n\gamma$ with $\lambda = o(1)$. For a sufficiently large constant C, with probability $1 - O(n\gamma + \frac{\beta}{\lambda}\exp(-\frac{n\lambda^3}{\beta^3}))$ we have

$$Y_\ell(t) = nz_\ell(t/n) + O(\lambda n)$$

uniformly for $0 \leq t \leq \sigma n$, for each ℓ, where $z_\ell(x)$ is the solution in (a) with $\hat{z}_\ell = Y_\ell(0)/n$ and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within ℓ^∞-distance $C\lambda$ of the boundary of W.

4 Determining the Differential Equations

First we determine functions $f^{(r)}_{(i,j)}$ and $f^{(r)}$ such that, for $0 \leq i, j \leq 2$ and $r \in \{0, 1, 2, 3\}$,

$$f^{(r)}_{(i,j)}(t/n, Y_{(0,0)}(t)/n, \ldots, Y_{(2,2)}(t)/n, D(t)/n) + o(1)$$

is the expected change in $Y_{(i,j)}$ due to an operation of type r at time t and

$$f^{(r)}(t/n, Y_{(0,0)}(t)/n, \ldots, Y_{(2,2)}(t)/n, D(t)/n) + o(1)$$

is the expected change in D due to an operation of type r at time t.

During an operation there are five sorts of vertices:

- vertices that have none of their associated free points exposed,
- the vertex u chosen at the start of the operation and added to the dominating set,
- the out-neighbours of u from exposing the free out-points associated with u, called *rems*,
- vertices, other than u and its out-neighbours, that have an associated in-point exposed, called *in-incs*, and
- vertices, other than u and its out-neighbours, that have an associated out-point exposed, called *out-incs*.

We determine the expected change in the random variables (and thus the differential equations) by considering the contribution from the different sorts of vertices. Since more than one edge may be exposed during an operation, the random variables $Y_{(i,j)}$ change during an operation. However they will only change by a constant amount (since only a constant number of edges are exposed); so if the number of free in-points ρ is at least a constant times n, the value of $Y_{(i,j)}/\rho$ for each (i,j) during an operation will be within $o(1)$ of its value at the start of the operation. Thus we will assume that ρ is $\Omega(n)$ and treat each $Y_{(i,j)}$ as a constant throughout each operation.

First let $\mathbb{P}_{\rm in}(w \in V_{(i,j)})$ be the probability that a vertex w, selected via a free in-point chosen uniformly at random, has degree pair (i,j). And similarly for $\mathbb{P}_{\rm out}(w \in V_{(i,j)})$. Then

$$\mathbb{P}_{\rm in}(w \in V_{(i,j)}) = (2-i)Y_{(i,j)}/\rho \quad \text{and} \quad \mathbb{P}_{\rm out}(w \in V_{(i,j)}) = (2-j)Y_{(i,j)}/\rho$$

where $\rho = \sum_{p=0}^{2}\sum_{q=0}^{2}(2-p)Y_{(p,q)} = \sum_{p=0}^{2}\sum_{q=0}^{2}(2-q)Y_{(p,q)}$.

In-incs and Out-incs

The expected change in $Y_{(i,j)}$ due to an in-inc w is $\mathrm{In}_{(i,j)} + o(1)$ where

$$\mathrm{In}_{(i,j)} = \mathbb{P}_{\rm in}(w \in V_{(i-1,j)}) - \mathbb{P}_{\rm in}(w \in V_{(i,j)}) = ((3-i)Y_{(i-1,j)} - (2-i)Y_{(i,j)})/\rho$$

and taking $Y_{(i,j)} = 0$ if $i < 0$ or $j < 0$. Similarly, the expected change in $Y_{(i,j)}$ due to an out-inc w is $\mathrm{Out}_{(i,j)} + o(1)$ where

$$\mathrm{Out}_{(i,j)} = \mathbb{P}_{\rm out}(w \in V_{(i,j-1)}) - \mathbb{P}_{\rm out}(w \in V_{(i,j)}) = ((3-j)Y_{(i,j-1)} - (2-j)Y_{(i,j)})/\rho.$$

Rems

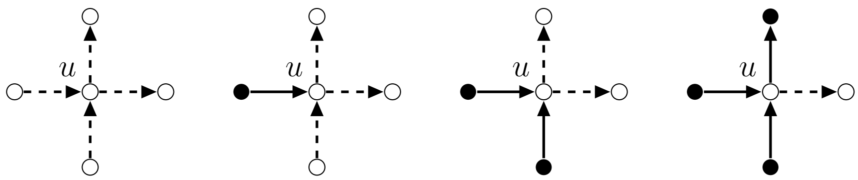
A rem is a vertex that is a new out-neighbour of u. Let w be a rem. Contributions to the expected change in $Y_{(i,j)}$ from saturating w come from three sources: w moving to $V_{(2,2)}$, in-incs from exposing the free out-points associated with w, and out-incs from exposing the free in-points associated with w. Let $F_{\rm in}$ and $F_{\rm out}$ be the number of free in-points and out-points associated with w (respectively) before the edge (u,w) is added. Then the expected change in $Y_{(i,j)}$ due to a rem is $\mathrm{Rem}_{(i,j)} + o(1)$ where

$$\begin{aligned}\mathrm{Rem}_{(i,j)} &= \delta_{(i,j)=(2,2)} - \mathbb{P}_{\rm in}(w \in V_{(i,j)}) + \mathbb{E}(F_{\rm in}-1)\,\mathrm{Out}_{(i,j)} + \mathbb{E}(F_{\rm out})\,\mathrm{In}_{(i,j)} \\ &= \delta_{(i,j)=(2,2)} - (2-i)Y_{(i,j)}/\rho + (2/\rho)(Y_{(0,0)} + Y_{(0,1)} + Y_{(0,2)})\mathrm{Out}_{(i,j)} \\ &\quad + (1/\rho)(4Y_{(0,0)} + 2Y_{(1,0)} + 2Y_{(0,1)} + Y_{(1,1)})\mathrm{In}_{(i,j)}\end{aligned}$$

and $\delta_b = 1$ if b is true and 0 otherwise.

Operations

There are 4 operation types as displayed in Figure 1. The black circles represent vertices that are saturated prior to the operation. The empty circles represent vertices which are not saturated prior to the operation. Similarly, the black edges are edges present at the start of the operation and dashed edges are edges added during the operation.



Type 0 $u \in V_{(0,0)}$ Type 1 $u \in V_{(1,0)}$ Type 2 $u \in V_{(2,0)}$ Type 3 $u \in V_{(2,1)}$

Figure 1: The four types of operations.

From Figure 1 we can see that for an operation processing a vertex u of degree pair (p,q) there are $2-q$ rems and $2-p$ out-incs from exposing the free points associated with u. Also u moves from $V_{(p,q)}$ to $V_{(2,2)}$. Therefore the expected change in $Y_{(i,j)}$ due to an operation of type r is $\mathrm{Op}r_{(i,j)} + o(1)$ where

$$\mathrm{Op}r_{(i,j)} = \delta_{(i,j)=(2,2)} - \delta_{(i,j)=(p,q)} + (2-q)\mathrm{Rem}_{(i,j)} + (2-p)\mathrm{Out}_{(i,j)}$$

and (p,q) is the degree pair of a vertex of type r (see Table 1).

Changes in the Size of the Dominating Set

The expected change in the size of the dominating set due to a type r operation is 1 plus the expected number of accidental saturates. Accidental saturates are either in-incs or out-incs; they are never rems. Thus the expected change in D due to a type r operation is $\mathrm{dom}_r + o(1)$ where

$$\mathrm{dom}_r = 1 + (2-q)(\mathrm{Rem}_{(2,2)} - 1) + (2-p)\mathrm{Out}_{(2,2)}$$

and (p,q) is the degree pair of a vertex of type r.

The Functions $f^{(r)}_{(i,j)}$ and $f^{(r)}$

To obtain the functions $f^{(r)}_{(i,j)}$ and $f^{(r)}$ we set $Y_{(i,j)}(t) = nz_{(i,j)}(t/n)$ and $D(t) = nz(t/n)$ and write $\mathrm{Op}r_{(i,j)}$ and dom_r in terms of $z_{(i,j)}$ (for $0 \le i,j \le 2$) and z. For example

$$\begin{aligned}
\mathrm{In}_{(i,j)} &= \frac{(3-i)z_{(i-1,j)} - (2-i)z_{(i,j)}}{s},\\
\mathrm{Out}_{(i,j)} &= \frac{(3-j)z_{(i,j-1)} - (2-j)z_{(i,j)}}{s}, \text{ and}\\
\mathrm{Rem}_{(i,j)} &= \delta_{i,2}\delta_{2,j} - (2-j)z_{(i,j)}/s + (2/s)(z_{(0,0)} + z_{(1,0)} + z_{(2,0)})\mathrm{In}_{(i,j)}\\
&\quad + (1/s)(4z_{(0,0)} + 2z_{(1,0)} + 2z_{(0,1)} + z_{(1,1)})\mathrm{Out}_{(i,j)}
\end{aligned}$$

where $s = \sum_{p=0}^{2}\sum_{q=0}^{2}(2-p)z_{(p,q)}$. The other equations follow from those above.

Next we determine the probability distribution that will approximate the prioritised algorithm.

Probabilities for Operation Types

Let p_r be the probability of choosing an operation of type r. For the deprioritised algorithm to approximate DominatingSet we want p_r to be approximately the proportion of operations of type r performed by DominatingSet.

In the prioritised algorithm the only Type 0 operation occurs as the first operation, so we set $p_0 = 0$. Asymptotically almost surely the second operation of DominatingSet is of type 1. Between any two type 1 operations there is a sequence of operations of type 2 and of type 3. We define a $(1,3)$-clutch to be an operation of type 1 and all operations of types 2 or 3 that follow until the next type 1 operation (or the algorithm ends). Then DominatingSet can be viewed as a sequence of $(1,3)$-clutches. Let M_r be the expected number of type r operations in a $(1,3)$-clutch. Then we set $p_r = M_r/(M_1 + M_2 + M_3)$.

Let

$$G_0 \subset G_1 \subset \cdots \subset G_L \tag{1}$$

be a $(1,3)$-clutch and let

$$H_0 \subset H_1 \subset \cdots \subset H_{L'}$$

be the subsequence of (1) of digraphs with no vertices of type 3. Thus H_1 is obtained from H_0 via a type 1 operation followed by a sequence of type 3 operations and, for $t > 1$, H_t is obtained from H_{t-1} via a type 2 operation followed by a sequence of type 3 operations.

Now consider

$$G_0 \subset G_1 \subset \cdots \subset G_K$$

where $G_K = H_1$. Let $N_3(t)$ be the number of type 3 vertices in G_t for $t = 0, \ldots, K$ (note that K is a random variable). Then we have $\mathbb{E}(N_3(0)) = 0$, $\mathbb{E}(N_3(K+1)) = 0$ and $\mathbb{E}(N_3(t)) = \mathrm{Op1}_{(2,1)} + (t-1)\mathrm{Op3}_{(2,1)}$. Thus

$$0 = \mathbb{E}(N_3(K+1)) = \mathbb{E}(\mathbb{E}(N_3(K+1) \mid K = k)) = \mathrm{Op1}_{(2,1)} + \mathbb{E}(K)\mathrm{Op3}_{(2,1)}$$

and so $\mathbb{E}(K) = \mathrm{Op1}_{(2,1)}/(-\mathrm{Op3}_{(2,1)})$. In a similar fashion we can show that the expected number of type 3 operations following a type 2 operation until the next type 1 or type 2 operation is $\mathrm{Op2}_{(2,1)}/(-\mathrm{Op3}_{(2,1)})$.

Applying the same argument to $H_0 \subset \cdots \subset H_{L'}$ we can show that

$$M_2 = \frac{-\mathrm{Op1}_{(2,0)}\mathrm{Op3}_{(2,1)} + \mathrm{Op1}_{(2,1)}\mathrm{Op3}_{(2,0)}}{\mathrm{Op2}_{(2,0)}\mathrm{Op3}_{(2,1)} - \mathrm{Op2}_{(2,1)}\mathrm{Op3}_{(2,0)}}.$$

Therefore $p_r = q_r/(q_1 + q_2 + q_3)$ where

$$\begin{aligned}
q_1 &= \mathrm{Op2}_{(2,0)}\mathrm{Op3}_{(2,1)} - \mathrm{Op2}_{(2,1)}\mathrm{Op3}_{(2,0)},\\
q_2 &= -\mathrm{Op1}_{(2,0)}\mathrm{Op3}_{(2,1)} + \mathrm{Op1}_{(2,1)}\mathrm{Op3}_{(2,0)}, \text{ and}\\
q_3 &= -\mathrm{Op1}_{(2,1)}\mathrm{Op2}_{(2,0)} + \mathrm{Op1}_{(2,0)}\mathrm{Op2}_{(2,1)}.
\end{aligned}$$

For p_i $(i = 1, 2, 3)$ to approximate the proportion of operations in a $(1, 3)$-clutch we must have $-\mathrm{Op3}_{(2,1)}$ and q_1 bounded above zero. Thus the differential equations (which involve each p_i) will be solved on a domain where $-\mathrm{Op3}_{(2,1)}$ and q_1 (written in terms of the variables $z_{(i,j)}$ for $0 \leq i, j \leq 2$) are bounded above zero. It turns out that p_i $(i = 1, 2, 3)$ as given above approximate the proportion of operations during the entire algorithm.

The Differential Equations

The differential equations we use when applying Theorem 3.1 are a combination of the work in this section. We use the differential equation

$$\frac{dz_{(i,j)}}{dx} = \sum_{r=1}^{3} p_r \cdot f_{(i,j)}^{(r)} \tag{2}$$

to approximate $\mathbb{E}(Y_{(i,j)}(t+1) - Y_{(i,j)}(t) \,|\, G_0, \ldots, G_t)$ for $0 \leq i, j \leq 2$ and the differential equation

$$\frac{dz}{dx} = \sum_{r=1}^{3} p_r f^{(r)} \tag{3}$$

to approximate $\mathbb{E}(D(t+1) - D(t) \,|\, G_0, \ldots, G_t)$.

5 Analysing the Deprioritised Algorithm

Before applying Theorem 3.1 we consider how to satisfy hypotheses (i), (ii), and (iii). Since each step of the algorithm is one operation, and each operation exposes only a bounded (independently of n) number of edges, the Boundedness Hypothesis is satisfied.

The algorithm chooses the type of operation to perform using p_i (for $i = 1, 2, 3$) as a probability distribution. So the Trend Hypothesis will be satisfied when each $p_i \geq 0$, $p_1 + p_2 + p_3 = 1$, when an operation of the type chosen is able to be performed (that is, if $p_r > 0$ then there is at least one vertex of type r), and when $\rho > cn$ for some constant $c > 0$ (as this was assumed when deriving $f_{(i,j)}^{(r)}$).

Notice that at the start of the algorithm there are no vertices of types 1, 2, or 3; so the Trend hypothesis is not satisfied. Thus we start the deprioritised algorithm with a preprocessing phase where only type 0 operations are performed. During the preprocessing phase vertices of types 1, 2, and 3 build up.

We run the preprocessing phase for $\lfloor \epsilon n \rfloor$ steps for some $\epsilon = o(1)$. This ensures that the number of vertices of types 1, 2, and 3 are sufficient to satisfy the Trend Hypothesis and that the contribution to $Y_{(i,j)}$ $(0 \leq i, j \leq 2)$ during the preprocessing phase is $o(n)$.

The functions $dz_{(i,j)}/dx$, dz/dx, and their first order partial derivatives (with respect to $z_{(0,0)}$ etc) all have a similar form: a polynomial in the variables $z_{(i,j)}$ $(0 \leq i, j \leq 2)$ divided by a polynomial in s and $q_1+q_2+q_3$. So these functions are Lipschitz on a domain where each $z_{(i,j)}$ (for $0 \leq i, j \leq 2$) is bounded and s and $q_1 + q_2 + q_3$ are bounded above zero.

For any $\delta > 0$, let D_δ be the domain

$$\begin{aligned} D_\delta = \{(x, z_{(0,0)}, \ldots, z_{(2,2)}, z) : & -\delta < x < 4,\ -\delta < z < 1+\delta, \\ & -\delta < z_{(i,j)} < 1+\delta \text{ for } 0 \le i,j \le 2, \\ & z_{(0,0)} > \delta,\ q_1 > \delta,\ -f^{(3)}_{(2,1)} > \delta, \\ & q_2 > 0,\ q_3 > 0,\ z_{(1,0)} > 0\}. \end{aligned}$$

Define $z^{(\delta)}_{(i,j)}$ and $z^{(\delta)}$ to be the solutions on the closure of D_δ to the differential equations (2) and (3) with initial conditions $z^{(\delta)}(0) = 0$, $z^{(\delta)}_{(i,j)}(0) = 0$ for $(i,j) \neq (0,0)$, and $z^{(\delta)}_{(0,0)}(0) = 1$.

Let $\mathbf{z}^{(\delta)} = (x, z^{(\delta)}_{(0,0)}(x), \ldots, z^{(\delta)}_{(2,2)}(x), z(x))$. Then we have the following theorem.

Theorem 5.1. *For any fixed $\delta > 0$, let $x_f(\delta)$ be the infimum of all $x > 0$ for which $\mathbf{z}^{(\delta)} \notin D_\delta$. Then the minimum size of a dominating set of a random 2-in 2-out digraph is a.a.s. less than*

$$z^{(\delta)}(x_f(\delta))n + (1 - z^{(\delta)}_{(2,2)}(x_f(\delta)))n + o(n).$$

To prove this theorem we apply Theorem 3.1 twice: first to the preprocessing phase and then to rest of the deprioritised algorithm.

Now let

$$\widehat{D}_\delta = D_\delta \cap \{(x, z_{(0,0)}, \ldots, z_{(2,2)}, z) : z_{(2,0)} > 0,\ z_{(2,1)} > 0\}$$

and $\mathbf{Y}(t) = (t, Y_{(0,0)}(t), \ldots, Y_{(2,2)}(t), D(t))$. The next lemma allows us to apply Theorem 3.1 on the domain $\widehat{D}_\delta$ after the preprocessing phase.

Lemma 5.2. *For all sufficiently small $\epsilon > 0$ and $\delta > 0$, asymptotically almost surely $\mathbf{Y}(\lfloor \epsilon n \rfloor)/n \in \widehat{D}_\delta$. Moreover, $\mathbf{Y}(\lfloor \epsilon n \rfloor)/n$ is at least some distance $\kappa = \kappa(\epsilon) > 0$ from the boundary of $\widehat{D}_\delta$.*

Proof. Before the algorithm starts we have $Y_{(0,0)}(0) = n$, $Y_{(i,j)}(0) = 0$ for $(i,j) \neq (0,0)$, and $D(0) = 0$. Each operation during the preprocessing phase is of type 0. So we apply Theorem 3.1 with the differential equations

$$\frac{dz_{(i,j)}}{dx} = \mathrm{Op0}_{(i,j)} \quad \text{and} \quad \frac{dz}{dx} = \mathrm{dom}_0 \tag{4}$$

(written in terms of the $z_{(i,j)}$'s) on the domain

$$\begin{aligned} W_{\delta,\epsilon} = \{(x, z_{(0,0)}, \ldots, z_{(2,2)}, z) : & -\delta < x < \epsilon, \\ & -\delta < z_{(i,j)} < 1+\delta \text{ for } 0 \le i,j \le 2, \\ & -\delta < z < 1+\delta,\ z_{(0,0)} > \delta\}. \end{aligned}$$

Let $z^{(p)}_{(i,j)}$ and $z^{(p)}$ be the solutions to the system of differential equations (4) in $W_{\delta,\epsilon}$ with initial conditions $z^{(p)}_{(0,0)}(0) = 1$, $z^{(p)}_{(i,j)}(0) = 0$ for $(i,j) \neq (0,0)$, and $z^{(p)}(0) = 0$. Also let $\mathbf{z}^{(p)}(x) = (x, z^{(p)}_{(0,0)}(x), \ldots, z^{(p)}_{(2,2)}(x), z^{(p)}(x))$. The conditions of Theorem 3.1 are easily seen to be satisfied, so we conclude that a.a.s.

$$Y_{(i,j)}(t) = nz^{(p)}_{(i,j)}(t/n) + o(n)\ (0 \le i, j \le 2) \text{ and } D(t) = nz^{(p)}(t/n) + o(n)$$

for $0 \le t \le \sigma n$ where σ is the supremum of those x to which the solutions $z^{(p)}_{(i,j)}$ and $z^{(p)}$ can be extended before $\mathbf{z}^{(p)}(x)$ is within some distance $o(1)$ of the boundary of $W_{\delta,\epsilon}$.

First we show that $\mathbf{z}^{(p)}(x)$ approaches the boundary $x = \epsilon$ of $W_{\delta,\epsilon}$ for all sufficiently small positive ϵ. From the definitions of $Y_{(i,j)}$ and D we have

$$-\delta < 0 \le Y_{(i,j)}(t)/n \le 1 < 1 + \delta$$

and similarly for $D(t)/n$ for $t = 0, \ldots, T_W$. So $\mathbf{Y}(t)/n$ approaches the boundary $x = \epsilon$ or $z_{(0,0)} = \delta$. In $\lfloor \epsilon n \rfloor$ operations $Y_{(0,0)}$ changes by $O(\lfloor \epsilon n \rfloor)$ and so, for ϵ and δ sufficiently small, $Y_{(0,0)}(t)/n = 1 + O(\epsilon)$ is bounded away from δ for $t = 0, \ldots, \lfloor \epsilon n \rfloor$. Therefore, for ϵ and δ sufficiently small, $\mathbf{Y}(t)/n$ approaches the boundary $x = \epsilon$ of $W_{\delta,\epsilon}$, while remaining at least a constant distance from the other boundaries.

Now assume that $\mathbf{z}^{(p)}(x)$ approaches a boundary of $W_{\delta,\epsilon}$ other than $x = \epsilon$, for example, the $z_{(0,0)} = \delta$ boundary. Then a.a.s.

$$Y_{(0,0)}(\sigma n)/n = z^{(p)}_{(0,0)}(\sigma) + o(1) = \delta + o(1).$$

Therefore $\mathbf{Y}(t)/n$ approaches arbitrarily close to the boundary $z_{(0,0)} = \delta$ of $W_{\delta,\epsilon}$. This contradicts the conclusion of the previous paragraph.

Hence $\mathbf{z}^{(p)}(x)$ approaches the boundary $x = \epsilon$ of $W_{\delta,\epsilon}$ while being bounded away from the others. Since $\lfloor \epsilon n \rfloor = \sigma n + o(n)$ (as $\sigma(n) \to \epsilon$) we have a.a.s.

$$Y_{(i,j)}(\lfloor \epsilon n \rfloor)/n = z^{(p)}_{(i,j)}(\sigma) + o(1) \text{ and } D(\lfloor \epsilon n \rfloor)/n = z^{(p)}(\sigma) + o(1) \tag{5}$$

for $0 \le i, j \le 2$.

By definition, $Y_{(i,j)}(\lfloor \epsilon n \rfloor)/n$ and $D(\lfloor \epsilon n \rfloor)/n$ are bounded away from the boundaries $z_{(i,j)} = 1 + \delta$, $z_{(i,j)} = -\delta$, $z = 1 + \delta$, and $z = -\delta$ (for $0 \le i, j \le 2$) of $\widehat{D}_\delta$. Earlier we showed that $Y_{(0,0)}(\lfloor \epsilon n \rfloor)$ is bounded above δn (for sufficiently small ϵ and δ), so $Y_{(0,0)}(\lfloor \epsilon n \rfloor)/n$ is bounded above δ. In a similar way we can show that $q_1(\mathbf{Y}(\lfloor \epsilon n \rfloor)/n)$ and $-f^{(3)}_{(2,1)}(\mathbf{Y}(\lfloor \epsilon n \rfloor)/n)$ are also bounded above δ.

Now consider the boundary $q_3 > 0$. We want to show a.a.s.

$$q_3(\mathbf{Y}(\lfloor \epsilon n \rfloor)/n) > \kappa$$

for some $\kappa > 0$. By (5) we have a.a.s.

$$\begin{aligned} q_3(\mathbf{Y}(\lfloor \epsilon n \rfloor)/n) &= q_3(\sigma, z^{(p)}_{(0,0)}(\sigma) + o(1), \ldots, z^{(p)}_{(0,0)}(\sigma) + o(1), z^{(p)}(\sigma) + o(1)) \\ &= q_3(\mathbf{z}^{(p)}(\sigma)) + o(1) \end{aligned}$$

where the last step holds since q_3 is Lipschitz. Now at $x = 0$ we have

$$q_3(\mathbf{z}^{(p)}(0)) = 0,$$
$$\frac{d}{dx}\left(q_3(\mathbf{z}^{(p)}(x))\right)|_{x=0} = 0, \text{ and}$$
$$\frac{d^2}{dx^2}\left(q_3(\mathbf{z}^{(p)}(x))\right)|_{x=0} = 88.$$

Note that during the preprocessing phase, the derivative of a function ϕ with respect to x is calculated via

$$\frac{d\phi}{dx} = \sum_{0\le i,j\le 2} \frac{\partial \phi}{\partial z_{(i,j)}} \mathrm{Op0}_{(i,j)}.$$

The derivative of ϕ can then be easily computed using the techniques of automatic differentiation, see [7].

Since the second derivative of q_3 is continuous it must remain positive on $[0, \epsilon]$ for some sufficiently small $\epsilon > 0$. Thus $q_3(\mathbf{z}^{(p)}(\sigma)) > \kappa > 0$ for some κ and so a.a.s. we have $q_3(\mathbf{Y}(\lfloor \epsilon n \rfloor)/n) > \kappa > 0$. The remaining boundaries, $q_2 > 0$ and $z_{(i,j)} > 0$ for $(i, j) = (1, 0), (2, 0), (2, 1)$ are dealt with similarly. □

We now define functions $\hat{z}_{(i,j)}$ and $\hat{z}$ by applying Theorem 3.1 to the deprioritised algorithm after the preprocessing phase. Hypotheses (i) and (iii) of Theorem 3.1 are satisfied as explained at the beginning of this section. For Hypothesis (ii) we note that, for $t < T_{\widehat{D}_\delta}$, we have $Y_{(1,0)}$, $Y_{(2,0)}$, and $Y_{(2,1)}$ all greater than zero; thus an operation of each type may be performed. Also $q_1 + q_2 + q_3 > \delta$ and $q_i > 0$ (for $i = 1, 2, 3$) ensure that p_1, p_2, and p_3 define a probability distribution. Together with Section 4, this shows that the Trend Hypothesis is satisfied. Finally, Lemma 5.2 allows us to apply Theorem 3.1 with initial conditions given below.

So let $\hat{\mathbf{z}}(x) = (x, \hat{z}_{(0,0)}(x), \ldots, \hat{z}_{(2,2)}(x), \hat{z}(x))$ where $\hat{z}_{(i,j)}$ and $\hat{z}$ are the solutions to the differential equations (2) and (3) with initial conditions $\hat{z}_{(i,j)}(0) = Y_{(i,j)}(\lfloor \epsilon n \rfloor)/n$ and $\hat{z}(0) = D(\lfloor \epsilon n \rfloor)/n$. Then applying Theorem 3.1 we obtain a.a.s.

$$Y_{(i,j)}(t) = n\hat{z}_{(i,j)}(t/n) + o(n) \ (0 \le i, j \le 2) \text{ and } D(t) = n\hat{z}(t/n) + o(n) \qquad (6)$$

for $0 \le t \le \sigma n$ where σ is now the supremum of all x to which the solutions $\hat{z}_{(i,j)}$ and $\hat{z}$ can be extended before $\hat{\mathbf{z}}(x)$ is within some distance $d(n) = o(1)$ of the boundary of $\widehat{D}_\delta$. Here σ depends on ϵ.

Note that

$$\frac{d\hat{z}_{(2,0)}}{dx} = 0 \quad \text{and} \quad \frac{d\hat{z}_{(2,1)}}{dx} = 0$$

for all x. Since $Y_{(i,j)}(\lfloor \epsilon n \rfloor)/n$ is bounded away from zero for $(i, j) = (2, 0)$ and $(2, 1)$, the solution $\hat{\mathbf{z}}(x)$ cannot approach the boundaries $z_{(2,0)} > 0$ and $z_{(2,1)} > 0$. Hence $\hat{\mathbf{z}}$ approaches a boundary of D_δ. This boundary will depend on $z^{(\delta)}_{(i,j)}$ and $z^{(\delta)}$.

To finish the proof of Theorem 5.1 we need the following standard lemma (which appears in [15] as Lemma 1).

Lemma 5.3 ([15]). *Let W be a bounded and open set. Suppose that $(x, \mathbf{y}_n(x))$ and $(x, \mathbf{z}_n(x))$ satisfy the same differential equations on W with differing initial conditions $(0, \mathbf{y}_n(0))$ and $(0, \mathbf{z}_n(0))$, respectively. If the differential equations are Lipschitz on W and $|\mathbf{y}_n(0) - \mathbf{z}_n(0)| \to 0$ as $n \to \infty$, then*

$$|\mathbf{y}_n(x) - \mathbf{z}_n(x)| \to 0$$

uniformly for $x \in [0, x_n^\star)$ where $x_n^\star$ is the infimum of those $x > 0$ for which $(x, \mathbf{y}_n(x)) \notin W$ or $(x, \mathbf{z}_n(x)) \notin W$.

In particular we use Lemma 5.3 to show that $\hat{z}_{(i,j)}$ and $\hat{z}$ approach $z^{(\delta)}_{(i,j)}$ and $z^{(\delta)}$ as n tends to infinity, as described by Lemma 5.4. Recall that $z^{(\delta)}_{(i,j)}$ and $z^{(\delta)}$ are defined immediately before Theorem 5.1, $x_f^{(\delta)}$ is defined in Theorem 5.1, and that $\hat{z}_{(i,j)}$, $\hat{z}$, and $\sigma(n)$ are defined before and after (6).

Lemma 5.4. *There exists sequences $\epsilon(n) = o(1)$ and $\kappa(n) = o(1)$ such that a.a.s.*

(i) $\sigma(n) \geq x_f(\delta) - \kappa(n) > 0$ *and*

(ii)
$$\left| \hat{z}_{(i,j)}(x) - z^{(\delta)}_{(i,j)}(x) \right| \to 0$$
uniformly for $x \in [0, x_f(\delta) - \kappa(n)]$.

Proof. We start by showing that $x_f(\delta)$ is at least a positive constant (for sufficiently small δ). At $x = 0$ we have $\mathbf{z}^{(\delta)}$ bounded away from all the boundaries of D_δ except $q_2 = 0$, $q_3 = 0$, and $z_{(1,0)} = 0$ (for sufficiently small δ). Thus by continuity $\mathbf{z}^{(\delta)}(x)$ remains bounded away from the same boundaries for $x \in [0, c]$ for some $c > 0$. For the boundaries $q_2 = 0$, $q_3 = 0$, and $z_{(1,0)} = 0$ we need to consider the derivatives of q_2, q_3, and $z_{(1,0)}$. Since we are no longer in the preprocessing phase, the derivative of a function ϕ (w.r.t. x) is calculated via

$$\frac{d\phi}{dx} = \sum_{0 \leq i,j \leq 2} \frac{\partial \phi}{\partial z_{(i,j)}} \frac{dz_{(i,j)}}{dx}$$

where $dz_{(i,j)}/dx$ is given by (2).

The derivatives of q_2, q_3, and $z_{(1,0)}$ at $x = 0$ are as follows:

$$\frac{dq_2}{dx}\left(\mathbf{z}^{(\delta)}(0)\right) = 6, \ \frac{dq_3}{dx}\left(\mathbf{z}^{(\delta)}(0)\right) = 0,$$
$$\frac{d^2 q_3}{dx^2}\left(\mathbf{z}^{(\delta)}(0)\right) = 42, \text{ and } \frac{dz_{(1,0)}}{dx}\left(\mathbf{z}^{(\delta)}(0)\right) = 3.$$

Hence for sufficiently small $c > 0$, we have that $q_2(x)$, $q_3(x)$, and $z_{(1,0)}(x)$ are greater then zero for $x \in (0, c)$. Therefore $x_f(\delta) > c$ for some sufficiently small constant $c > 0$. We also note that the derivatives considered above are all Lipschitz on D_δ.

Next we show that a.a.s. $\sigma(n) > C$ for some constant $C > 0$ for all ϵ sufficiently small. Since $|\mathbf{Y}(\lfloor \epsilon n \rfloor) - \mathbf{Y}(0)| = O(\epsilon n)$ we have

$$\left|\hat{z}_{(i,j)}(0) - z^{(\delta)}_{(i,j)}(0)\right| = O(\epsilon).$$

Arguments similar to the above show that a.a.s. $\hat{\mathbf{z}}(x)$ is bounded from the boundaries of D_δ other than $q_3 = 0$ for $x \in [0, C]$ (for C sufficiently small). Note that for the boundaries $q_2 = 0$ and $z_{(1,0)} = 0$ we need to use the Lipschitz property of the derivatives of q_2 and $z_{(1,0)}$, and to assume that ϵ is sufficiently small.

For the boundary $q_3 = 0$ it is sufficient to show that

$$\frac{dq_3}{dx}(\hat{\mathbf{z}}(0)) > 0. \tag{7}$$

So we consider

$$\phi(x) = \frac{dq_3}{dx}(x) = \sum_{0 \le i,j \le 2} \frac{\partial q_3}{\partial z_{(i,j)}}(x) \frac{dz_{(i,j)}}{dx}(x)$$

during the preprocessing phase. At the start of the preprocessing phase we have $\phi(0) = 0$ but the derivative of ϕ (calculated as in the proof of Lemma 5.2) is 56. Thus a.a.s. (7) holds and so a.a.s. $q_3(\hat{\mathbf{z}}(x)) > 0$ for $x \in [0, C]$ (redefining C if necessary). Therefore a.a.s. $\sigma(n) > C$ for some sufficiently small C and for all sufficiently small ϵ.

Now taking $\epsilon = \epsilon(n) = o(1)$, from Lemma 5.3, for some function $g(x)$ with $g(x) \to 0$ as $x \to 0$, we obtain

$$\left|\hat{z}_{(i,j)}(x) - z^{(\delta)}_{(i,j)}(x)\right| = g(\epsilon) = o(1) \tag{8}$$

uniformly for $x \in [0, \min\{x_f(\delta), x^\star(n)\})$ where $x^\star(n)$ is the infimum of those x for which $\hat{\mathbf{z}}(x) \notin D_\delta$. Note that $\sigma(n) < x^\star(n)$ by definition.

We complete the proof by showing that there exists a sequence $\kappa(n)$ tending to 0 such that a.a.s. $\sigma(n) \ge x_f(\delta) - \kappa(n)$. Fix $\kappa > 0$ and assume that $\sigma(n) < x_f(\delta) - \kappa$ for infinitely many n. By the definition of $x_f(\delta)$, for all $x \in [C, x_f(\delta) - \kappa]$, the distance from $\mathbf{z}^{(\delta)}(x)$ to the boundary of D_δ is bounded below by a constant. But we can also write this distance (at $x = \sigma$) as $\mathbf{z}^{(\delta)}(\sigma) - \hat{\mathbf{z}}(\sigma)$ plus the distance from $\hat{\mathbf{z}}(\sigma)$ to the boundary of D_δ. However, both these values tend to zero (from (8) and the definition of σ). Thus we have a contradiction (using $\sigma(n) \in [C, x_f(\delta) - \kappa]$) and so

$$\sigma(n) \ge x_f(\delta) - \kappa \tag{9}$$

for sufficiently large n.

Now, given a decreasing sequence $\{\kappa_i\}_{i \ge 0}$ tending to zero, we use the argument of the previous paragraph to construct a sequence $\kappa(n)$ such that (i) holds. □

Now we finish the proof of Theorem 5.1. At every stage of the algorithm, the set $\mathcal{D} \cup \big(V(G) \backslash V_{(2,2)}(G)\big)$ is a dominating set for the final graph G. Thus $D(t) + (n - Y_{(2,2)}(t))$

is an upper bound on the minimal size of a dominating set of a random 2-in 2-out digraph for any t. In particular, taking

$$t = t^\star = \lfloor n\big(x_f(\delta) - \kappa(n)\big)\rfloor,$$

by (6) and Lemma 5.4 (i), we have a.a.s.

$$D(t^\star) + (n - Y_{(2,2)}(t^\star)) = n\hat{z}(t^\star/n) + n(1 - \hat{z}_{(2,2)}(t^\star/n)) + o(n).$$

Then using Lemma 5.4 (ii) we have a.a.s.

$$\begin{aligned} D(t^\star) + (n - Y_{(2,2)}(t^\star)) &= nz^{(\delta)}\big(x_f(\delta) - \kappa(n) + o(1)\big) \\ &\qquad + n\left[1 - z^{(\delta)}_{(2,2)}\big(x_f(\delta) - \kappa(n) + o(1)\big)\right] + o(n) \\ &= nz^{(\delta)}(x_f(\delta)) + n(1 - z^{(\delta)}_{(2,2)}(x_f(\delta))) + o(n) \end{aligned}$$

since $z^{(\delta)}_{(i,j)}$ and $z^{(\delta)}$ are Lipschitz (on the closure of D_δ) and $\kappa(n) = o(1)$.

A non-rigorous numerical analysis of the differential equations gives us the upper bound of Theorem 1.3. We also note that the numerical analysis suggests that at the end of the algorithm we have $z_{(0,0)} = z_{(1,1)} = 0$. So it seems reasonable to conjecture that DominatingSet and the deprioritised version of DominatingSet a.a.s. return 2-in 2-out digraphs.

6 Proving the Lower Bound

Let P be a pairing selected uniformly at random from the pairings on some set V of vertices. We will determine an a.a.s. lower bound on the minimum size of a dominating set of the directed multigraph $G(P)$ obtained from P. The lower bound will hold for random simple 2-in 2-out digraphs just as the upper bound did.

Let $N(k)$ be the number of dominating sets of size k in $G(P)$. Note that for any two integers k and $k^\star$ with $k < k^\star$, a dominating set of size k can be extended to a dominating set of size $k^\star$. Thus, using Markov's inequality, the probability that $G(P)$ has a dominating set of size less than or equal to $k^\star$ is

$$\mathbb{P}\left(\sum_{k \le k^\star} N(k) \ge 1\right) = \mathbb{P}(N(k^\star) \ge 1) \le \mathbb{E}(N(k^\star)).$$

So if $\mathbb{E}(N(k^\star)) = o(1)$ then $k^\star$ is an a.a.s. lower bound on the minimum size of a dominating set of $G(P)$.

Now consider a 2-in 2-out digraph D and a dominating set X of D. Each vertex of D not in X has either one or two out-edges to a vertex in X. With this in mind, for any two subsets Σ and A of V, the pair (Σ, A) is called a *domination pair* if Σ is a dominating set for $G(P)$, $\Sigma \cap A = \emptyset$, and each vertex of A has exactly one out-edge to a vertex in Σ.

Now let $N(k,a)$ be the number of domination pairs (Σ, A) of $G(P)$ such that $|\Sigma| = k$ and $|A| = a$. Then

$$\mathbb{E}(N(k)) = \sum_{a=0}^{n-k} \mathbb{E}(N(k,a)). \tag{10}$$

To calculate $\mathbb{E}(N(k,a))$ we write $N(k,a)$ as a sum of indicator variables $I_{(\Sigma,A)}$, where $I_{(\Sigma,A)} = 1$ if (Σ, A) is a domination pair and 0 otherwise, for all subsets Σ and A of V such that

$$\Sigma \cap A = \emptyset,\ |\Sigma| = k, \text{ and } |A| = a. \tag{11}$$

So choose a pair of sets Σ and A such that (11) holds; there are $\binom{n}{k}\binom{n-k}{a}$ ways to do so. Each of the $2k$ in-points associated with the vertices of Σ must be mapped to an out-point. Of these out-points, $a + 2(n-k-a)$ are associated with vertices in $V(G(P))) \backslash \Sigma$; so the remaining $4k + a - 2n$ out-points are associated with vertices in Σ. Thus there are $2^a\binom{2k}{4k+a-2n}$ ways to choose which out-points will be the images of the in-points associated with vertices in Σ, and there are $(2k)!$ ways to map the out-points to the in-points. We still have $2n-2k$ in-points left, so we can complete the pairing in $(2n-2k)!$ ways. Thus

$$\mathbb{E}(N(k,a)) = 2^a \binom{n}{k}\binom{n-k}{a}\binom{2k}{4k+a-2n}\binom{2n}{2k}^{-1}.$$

We also obtain the following bounds on a and k:

$$\max\{0, 2n-4k\} \le a \le n-k \text{ and } n/3 \le k \le n.$$

By (10) if $\max\{\mathbb{E}(N(k,a)) : a = 0, \ldots, n-k\}$ tends to zero exponentially quickly (as $n \to \infty$) then $\mathbb{E}(N(k)) = o(1)$. Let $k = \kappa n$, $a = \alpha n$, and $\phi(x) = x^x$ (with $\phi(0) = 1$). Then using Stirling's approximation we have

$$\mathbb{E}(N(P,k,a))^{\frac{1}{n}} \backsim \frac{2^{\alpha-2}\phi(2\kappa)^2\phi(2-2\kappa)}{\phi(\kappa)\phi(\alpha)\phi(1-\kappa-\alpha)\phi(4\kappa+\alpha-2)\phi(2-2\kappa-\alpha)} \tag{12}$$

for $1/3 \le \kappa \le 1$ and $\max\{0, 2-4\kappa\} \le \alpha \le 1-\kappa$.

Denote the right hand side of (12) by $f(\kappa, \alpha)$. Then

$$\frac{\partial}{\partial\alpha}(\ln f(\kappa,\alpha)) = \ln\left(\frac{2(2-2\kappa-\alpha)(1-\kappa-\alpha)}{\alpha(4\kappa+\alpha-2)}\right).$$

Thus $\frac{\partial}{\partial\alpha}(\ln f(\kappa,\alpha)) = 0$ when $2(2-2\kappa-\alpha)(1-\kappa-\alpha) = \alpha(4\kappa+\alpha-2)$. Solving for α gives $\alpha = 2-\kappa \pm \sqrt{\kappa(4-3\kappa)}$. The only solution lying in the domain arising from the constraints on k and a given above is $\alpha = 2-\kappa-\sqrt{\kappa(4-3\kappa)}$.

Therefore, for a given κ, the maximum value of $f(\kappa,\alpha)$ will occur at $\alpha = 1-\kappa$, $\alpha = \max\{0, 2-4\kappa\}$, or $\alpha = 2-\kappa-\sqrt{\kappa(4-3\kappa)}$. For $\kappa = 0.3495$ we have $f(\kappa, 1-\kappa)$, $f(\kappa, 2-4\kappa)$, and $f(\kappa, 2-\kappa-\sqrt{\kappa(4-3\kappa)})$ all less than 1. This proves the lower bound of Theorem 1.3.

References

[1] M. Beis, W. Duckworth, and M. Zito. Packing vertices and edges in random regular graphs. *Random Structures and Algorithms*, 32(1):20–37, 2008.

[2] M. Chlebík and J. Chlebíková. Approximation hardness of dominating set problems. In *Proc. of 12th European Symposium on Algorithms (ESA 2004)*, volume 3221 of *Lecture Notes in Computer Science*, pages 192–203. Springer Verlag, 2004.

[3] C. Cooper. A note on the connectivity of 2-regular digraphs. *Random Structures Algorithms*, 4(4):469–472, 1993.

[4] J. Diaz, M.J. Serna, and N.C. Wormald. Computation of the bisection width for random d-regular graphs. *Theoretical Computer Science*, 382(2):120–130, 2007.

[5] W. Duckworth and N.C. Wormald. Minimum independent dominating sets of random cubic graphs. *Random Structures Algorithms*, 21(2):147–161, 2002.

[6] W. Duckworth and N.C. Wormald. On the independent domination number of random regular graphs. *Combinatorics, Probability and Computing*, 15(4):513–522, 2006.

[7] A. Griewank. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Frontiers in Appl. Math. SIAM, Philadelphia, PA, 2000.

[8] T.W. Haynes, S.T. Hedetniemi, and P.J. Slater. *Domination In Graphs: Advanced Topics*. Monographs and textbooks in pure and applied mathematics, 209. Marcel Dekker, Inc., New York, 1998.

[9] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. John Wiley and Sons, New York, 2000.

[10] C. Lee. *On the domination number of a digraph*. PhD thesis, Department of Mathematics, Michigan State University, 1994.

[11] B.D. McKay. Asymptotics for 0-1 matrices with prescribed line sums. In *Enumeration and design (Waterloo, Ont., 1982)*, pages 225–238. Academic Press, Toronto, ON, 1984.

[12] L. Shi and N.C. Wormald. Colouring random regular graphs. *Combinatorics, Probability and Computing*, 16(3):459–494, 2007.

[13] N.C. Wormald. Differential equations for random processes and random graphs. *The Annals of Applied Probability*, 5(4):1217–1235, Nov. 1995.

[14] N.C. Wormald. The differential equation method for random graph process and greedy algorithms. In M. Karoński and H. Prömel, editors, *Lectures in Approximation and Randomized Algorithms*, pages 73–155. PWN, Warsaw, 1999.

[15] N.C. Wormald. Analysis of greedy algorithms on graphs with bounded degree. *Discrete Mathematics*, 273:235–260, 2003.