

The Expected Size of the Rule k Dominating Set

Jennie C. Hansen
Actuarial Mathematics
and Statistics Department
Herriot-Watt University
J.Hansen@ma.hw.ac.uk

Eric Schmutz
Department of Mathematics
Drexel University
Philadelphia, Pa. 19104
Eric.Jonathan.Schmutz@drexel.edu

Li Sheng
Department of Mathematics
Drexel University
Philadelphia, Pa. 19104
lsheng@math.@drexel.edu

November 30, 2005

Abstract

Dai, Li, and Wu proposed Rule k , a localized approximation algorithm that attempts to find a small connected dominating set in a graph. In this paper we consider the “average case” performance of two closely related versions of Rule k for the model of random unit disk graphs constructed from n random points in an $\ell_n \times \ell_n$ square. We show that, if $k \geq 3$ and $\ell_n = o(\sqrt{n})$, then for both versions of Rule k , the expected size of the Rule k dominating set is $\Theta(\ell_n^2)$ as $n \rightarrow \infty$. It follows that, for ℓ_n in a suitable range, the expected size of the Rule k dominating sets are within a constant factor of the optimum.

Keywords and phrases: *dominating set, localized algorithm, approximation algorithm, performance analysis, probabilistic analysis, Rule k , unit disk graph,*

1 Introduction

In this paper we consider the problem of finding a small connected dominating set for a *unit disk graph* $G = (V, E)$, where the vertex set, V , is a set of points in \mathbb{R}^2 . Given the vertex set V , the edge set E is determined as follows: an *undirected* edge $e \in E$ connects vertices $u, v \in V$ (and in this case we say that u and v are adjacent) iff the Euclidean distance between them is less than or equal to one. Unit disk graphs have been used by many authors as simplified mathematical models for the interconnections between hosts in a wireless network, and random unit disk graphs have been used as stochastic models for these networks [9],[13], [16],[17],[23],[24], [19],[20].

A *dominating set* in any graph $G = (V, E)$ is a subset $\mathcal{C} \subseteq V$ such that every vertex $v \in V$ either is in the set \mathcal{C} , or is adjacent to a vertex in \mathcal{C} . We say \mathcal{C} is a *connected dominating set* if \mathcal{C} is a dominating set and the subgraph induced by \mathcal{C} is connected. Obviously G cannot have a connected dominating set if G itself is not connected. We use the acronym ‘‘CDS’’ for a dominating set \mathcal{C} such that the subgraph induced by \mathcal{C} has the same number of components that G has. In this paper we consider a *random* unit disk graph model, \mathcal{G}_n , which is connected with asymptotic probability one. So, in this case, any CDS for \mathcal{G}_n will also be connected with high probability.

There has been considerable interest in designing good approximation algorithms for finding small connected dominating sets [2],[6],[8],[15],[25],[28]. One motivation for studying connected dominating sets is that there are CDS-based broadcasting methods that appear to be better than simple flooding [10],[27],[28], [30]. There have been various efforts to determine the average-case performance of such algorithms using simulations. However, with the exception of the mathematical parts of [4],[19],and [20], we are not aware of any probabilistic analysis that is mathematically rigorous. In this paper we consider a family of localized approximation algorithms, called ‘‘Rule k ’’, which were proposed by Dai, Li, and Wu [11],[30] and which find a CDS in a graph. To analyze the average-case performance of the Rule k algorithms, we first choose an appropriate probability model. Then, in the context of the probability model, we prove explicit asymptotic bounds on the expected size of the dominating set that is selected by (strict and relaxed versions of) Rule k .

2 The Algorithms

Some notation is needed first to describe the algorithms. We assume that each vertex has a unique identifier taken from a totally ordered set. For convenience, when $|V| = n$, we will use the numbers $1, 2, \dots, n$ as IDs, and will number the vertices accordingly. If x_i is a vertex whose ID is i , let let $N(x_i)$ be the set consisting of x_i and any vertices that are adjacent to x_i . The CDS constructed by the (Strict) Rule k algorithm is denoted $\mathcal{C}_k(V)$, and its cardinality is $C_k = |\mathcal{C}_k(V)|$. The set $\mathcal{C}_k(V)$ consists of all vertices $x_i \in V$ that are not excluded under the following ‘‘strict’’ version of Rule k :

Rule k: Vertex x_i is excluded from $C_k(V)$ iff $N(x_i)$ contains at least one set of k vertices $x_{i_1}, x_{i_2}, \dots, x_{i_k}$ such that

- $i_1 > i_2 > \dots > i_k > i$, and
- The subgraph induced by $\{x_{i_1}, x_{i_2}, \dots, x_{i_k}\}$ is connected, and
- $N(x_i) \subseteq \bigcup_{t=1}^k N(x_{i_t})$.

Wu Li and Dai proved that $C_k(V)$ is a CDS, and they conjectured that its cardinality C_k is, in some sense, small on average. Theorem 4 of this paper confirms their conjecture, for a suitable model, for any $k \geq 3$. If $k < 3$, then the performance of the strict Rule k is not as good, and the analysis is considerably more complicated[18]. We therefore omit the cases $k = 1$ and $k = 2$ from consideration in this paper.

For each vertex x_i let $N^+(x_i)$ consist of those neighbors of x_i with ID's larger than i . The following "Relaxed Rule k" was suggested to us by a referee and attributed to Wu and Dai. The Relaxed Rule k has also been referred to as the " $k = \infty$ " version, although " $k < \infty$ " would be a more consistent terminology. For the remainder of the paper, we refer to it as the Relaxed Rule k, but let $C_\infty(V)$ denote the dominating set that it selects.

Relaxed Rule k: Vertex x_i is excluded from $C_\infty(V)$ iff

- $N^+(x_i)$ is connected, and
- all the neighbors of x_i are covered by higher numbered neighbors of x_i , i.e.

$$N(x_i) \subseteq \bigcup_{w \in N^+(x_i)} N(w)$$

The rest of this paper is organized as follows. In the next section we specify the model and define the random unit disk graph, \mathcal{G}_n . In Sections 3 we prove a local coverage theorem that is needed in section 4 to prove an upper bound for $E(C_k)$. The results in section 4 are used in section 5 to analyze Relaxed Rule k. Finally, in the remainder of the paper, we discuss lower bounds and optimality issues.

3 Choice of Models

Before estimating the expected size of the Rule k dominating set, we must specify the underlying probability model. For any real number $\ell > 1$, let $\mathcal{Q}(\ell)$ be an $\ell \times \ell$ square in \mathbb{R}^2 . The particular choice of a square will be immaterial, but its size will be very important. Let $\Omega_{n,\ell} = \mathcal{Q}(\ell) \times \mathcal{Q}(\ell) \times \dots \times \mathcal{Q}(\ell)$ be the n -fold product space with the usual product topology. For each $n \geq 1$, let $X_{n,\ell,1}, X_{n,\ell,2}, \dots, X_{n,\ell,n}$ be a sequence of random points selected independently from a uniform distribution on $\mathcal{Q}(\ell)$ and let $\mathbf{P}_{n,\ell}$ denote the uniform probability measure on $\Omega_{n,\ell}$ induced by the random variables $X_{n,\ell,1}, X_{n,\ell,2}, \dots, X_{n,\ell,n}$.

Finally, let $\mathcal{G}(n, \ell)$ be the random unit disk graph with vertex set $V_{n, \ell} = \{X_{n, \ell, 1}, X_{n, \ell, 2}, \dots, X_{n, \ell, n}\}$ that is formed from these vertices by putting an edge between two vertices iff the Euclidean distance between the two vertices is less than or equal to one.

We want to estimate the “average” size of $\mathcal{C}_k(V_{n, \ell})$ for large networks. As it stands, the expected value $E_{n, \ell}(C_k) [= E(C_k(V_{n, \ell}))]$ is defined with respect to the probability measure $\mathbf{P}_{n, \ell}$ on $\Omega_{n, \ell}$ and depends on both n and ℓ . We shall not however attempt any multivariate asymptotic estimates. Instead, we choose a suitable sequence, $(\ell_n)_{n=1}^\infty$, and consider the expected value $E_{n, \ell_n}(C_k)$ with respect to \mathbf{P}_{n, ℓ_n} as $n \rightarrow \infty$. To simplify notation throughout, we will (usually) suppress the dependence on the choice of a sequence $(\ell_n)_{n=1}^\infty$. Thus we write \mathcal{G}_n instead of $\mathcal{G}(n, \ell_n)$, and write $E_n(C_k)$ instead of $E_{n, \ell_n}(C_k)$. Suppressing even n , we write \mathcal{Q} instead of $\mathcal{Q}(\ell_n)$, and \mathbf{P} instead of \mathbf{P}_{n, ℓ_n} .

Conditions on the growth rate of ℓ_n will be clear from the statements of theorems. However, to provide some perspective on our choice of growth rates for ℓ_n , we mention that it is known that the threshold for connectivity is $\ell_n = \Theta(\sqrt{n/\log n})$; if ℓ_n grows faster than this, then the random unit disk graph \mathcal{G}_n will be disconnected with probability $1 - o(1)$ as $n \rightarrow \infty$. In this case, with high probability, $\mathcal{C}_k(V_{n, \ell})$ will not be a *connected* dominating set for \mathcal{G}_n . More precise versions of these remarks are provided in the new book by Penrose[26] which gives an up to date survey of random geometric graphs.

Finally, throughout the remainder of this paper we adopt the following notation. For any points p and q in \mathbb{R}^2 , let $d(p, q)$ denote the ordinary Euclidean distance between p and q in \mathbb{R}^2 . Also, for any $v \in \mathbb{R}^2$ and $r > 0$, let $D_r(v) = \{w \in \mathbb{R}^2 : d(v, w) \leq r\}$.

4 Local Coverage by exactly k vertices

The next lemma is a purely geometric result which we require for the proof of Theorem 2. To state the lemma, we need some notation. Let $\delta = \frac{1}{2} - \frac{\sqrt{3}}{4} = .0669\dots$, and let $\rho = \frac{\sqrt{3}}{2} = .866\dots$ (so $\rho + 2\delta = 1$). Let p be any point in \mathcal{Q} , and let $D'_1(p) = D_1(p) \cap \mathcal{Q}$ be the set of points in the square \mathcal{Q} whose distance from p is one or less.

Lemma 1 *There exist points $z_0, z_1, z_2 \in D'_1(p)$ such that the following two conditions are satisfied:*

- for $s = 0$ and $s = 1$, $d(z_s, z_{s+1}) \leq 1 - 2\delta$
- $D'_1(p) \subseteq \bigcup_{s=0}^2 D_\rho(z_s)$.

Proof: Consider first the case where $D'_1(p) = D_1(p) \subseteq \mathcal{Q}$, i.e. p is a point that is not near the boundary of the square. We may, without loss of generality, choose the coordinate system such that $p = (0, 0)$ and such that the axes are parallel to the sides of the square \mathcal{Q} . For $s = 0, 1, 2$, let S_s be the sector of $D_1(p)$ consisting of those points whose polar coordinates (r, θ) satisfy $r \leq 1$ and

$\frac{(2s-1)\pi}{3} \leq \theta \leq \frac{(2s+1)\pi}{3}$. Let z_s be the point in S_s whose polar coordinates are $(\frac{1}{2}, \frac{2\pi s}{3})$. Then the first condition is satisfied: $d(z_s, z_{s+1}) = \sin \frac{\pi}{3} = 1 - 2\delta$. It is also straightforward to check that for $s = 0, 1, 2$, $S_s \subseteq D_\rho(z_s)$ and so the second condition is satisfied.

Now consider the remaining case $D'_1(p) \neq D_1(p)$, i.e. where $D_1(p)$ meets the boundary of \mathcal{Q} . Choose points z_0, z_1, z_2 as before so that $D_1(p) \subseteq \bigcup_{s=0}^2 D_\rho(z_s)$ and $d(z_s, z_{s+1}) \leq 1 - 2\delta$. We are not done because one or more of the points z_s may not lie in \mathcal{Q} . In particular, if $z_s \notin \mathcal{Q}$, then there is a (unique) $z'_s \in \mathcal{Q}$ such that $d(z_s, z'_s) = \inf\{d(z_s, z) : z \in \mathcal{Q}\}$. We replace z_s by z'_s and observe that every point of $D'_1(p)$ is closer to z'_s than it is to the original point z_s . Hence $S_s \cap \mathcal{Q} \subseteq D_\rho(z'_s)$. After replacing all z_s such that $z_s \notin \mathcal{Q}$ by the corresponding z'_s we obtain three points that satisfy the conditions of the lemma. \square

Fix $k \geq 3$, the k in ‘‘Rule k ’’. Suppose m points P_1, P_2, \dots, P_m are selected independently and uniform randomly in $D'(p)$. Let $\mathcal{K}_m(k)$ be the event that, for some $1 \leq i_0 < i_1 < i_2 < \dots < i_{k-1} \leq m$, we have:

- $D'_1(p) \subseteq \bigcup_{s=0}^{k-1} D_1(P_{i_s})$, and
- the unit disk graph with vertices $P_{i_0}, P_{i_1}, \dots, P_{i_{k-1}}$ is connected.

We note that event $\mathcal{K}_m(k)$ implies that the random unit disk graph which is formed from the vertices P_1, P_2, \dots, P_m has a k -point connected dominating set. With this notation we can state

Theorem 2 *There is a positive constant $\alpha < 1$ and a positive constant m_k such that, for all $m > m_k$, $\Pr(\mathcal{K}_m(k)) > 1 - 3\alpha^m$.*

Proof: Choose points z_0, z_1, z_2 as in the proof of Lemma 1. If z is any point in $D_\delta(z_s)$, then for all $y \in S_s$, $d(z, y) \leq d(z, z_s) + d(z_s, y) \leq \delta + \rho < 1$. Let \mathcal{E}_s be the event that *none* of the m random points P_1, P_2, \dots, P_m lies in $D_\delta(z_s)$. Then

$$\Pr(\mathcal{E}_s) = \left(1 - \frac{\text{Area}(D_\delta(z_s) \cap \mathcal{Q})}{\text{Area}(D'(p))}\right)^m.$$

Note that $\text{Area}(D_\delta(z_s) \cap \mathcal{Q}) \geq \frac{1}{4} \text{Area}(D_\delta(z_s)) = \frac{\pi\delta^2}{4}$, and that $\text{Area}(D'(p)) \leq \text{Area}(D_1(p)) = \pi$. If we let $\alpha = 1 - \frac{\delta^2}{4} = .998\dots$, then $\alpha < 1$, and for $s = 0, 1, 2$,

$$\Pr(\mathcal{E}_s) \leq \alpha^m. \tag{1}$$

It follows from(1) that $\Pr(\mathcal{K}_m) \geq 1 - 3\alpha^m$ since $\mathcal{E}_0^c \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c \subseteq \mathcal{K}_m$, and the proof is complete if $k = 3$.

Now suppose that $k > 3$. In this case, $\mathcal{K}_m(3) \subseteq \mathcal{K}_m(k)$ whenever $m \geq k$. To see this, suppose for some $1 \leq i_0 < i_1 < i_2 \leq 1$:

- $D'_1(p) \subseteq \bigcup_{s=0}^2 D_1(P_{i_s})$, and
- the unit disk graph with vertices $P_{i_0}, P_{i_1}, P_{i_2}$ is connected.

Choose any $1 \leq i_3 < \dots < i_{k-1} \leq m$ such that $i_j \notin \{i_0, i_1, i_2\}$ for $3 \leq j \leq k-1$. Then we also have:

- $D'_1(p) \subseteq \bigcup_{s=0}^{k-1} D_1(P_{i_s})$, and
- the unit disk graph with vertices $P_{i_0}, P_{i_1}, \dots, P_{i_{k-1}}$ is connected (since, for $3 \leq j \leq k-1$, each P_{i_j} is connected to either P_{i_0}, P_{i_1} , or P_{i_2}).

It follows that $\Pr(\mathcal{K}_m(k)) \geq \Pr(\mathcal{K}_m(3)) > 1 - 3\alpha^m$ for all sufficiently large m . \square

5 Strict Rule k Analysis

In this section, we assume that $\ell_n = o(\sqrt{n})$ as $n \rightarrow \infty$. Also, in this section, let $U_k = \sum_{i=1}^n I_i$ be a sum of indicator variables where $I_i = 1$ iff node i is not included in $\mathcal{C}_k(V)$ under Rule k. Thus Rule k selects a dominating set $\mathcal{C}_k(V)$ having $C_k = n - U_k$ vertices, and it is desirable for U_k to be large. Our goal in this section is to prove that, for all $k > 2$, $E(U_k) \geq n - O(\ell_n^2)$.

Let $\lambda_n = n - \ell_n^2$, and let X_1, X_2, \dots, X_n be independent, uniformly distributed random points in \mathcal{Q} , namely the locations of vertices. (Here we are again simplifying notation by writing X_i instead of $X_{n, \ell_n, i}$.) Let ρ_i be the number of neighbors of vertex i having a larger ID, i.e. the number of $j > i$ such that $d(X_i, X_j) \leq 1$.

Lemma 3

$$\mathbf{P}\left(\rho_i < \frac{(n-i)\pi}{8\ell_n^2}\right) \leq \exp\left(\frac{-(n-i)\pi}{32\ell_n^2}\right).$$

Proof: Let $|D'_1(X_i)| = \text{Area}(D_1(X_i) \cap \mathcal{Q})$ be the area of the set of points in \mathcal{Q} whose distance from X_i is one or less. Thus $|D'_1(X_i)| = \pi$ unless X_i happens to fall near the border, and in all cases $|D'_1(X_i)| \geq \frac{\pi}{4}$. Given $|D'_1(X_i)|$, the variable ρ_i has a Binomial($n-i, \frac{|D'_1(X_i)|}{\ell_n^2}$) distribution. Therefore Chernoff's bound on the lower tail distribution gives

$$\begin{aligned} & \mathbf{P}\left(\rho_i < \frac{(n-i)\pi}{8\ell_n^2} \mid |D'_1(X_i)|\right) = \\ & \mathbf{P}\left(\rho_i < \frac{\pi}{8|D'_1(X_i)|} \cdot \frac{|D'_1(X_i)|(n-i)}{\ell_n^2} \mid |D'_1(X_i)|\right) \\ & \leq \exp\left(-\left(1 - \frac{\pi}{8|D'_1(X_i)|}\right)^2 \cdot \frac{|D'_1(X_i)|(n-i)}{2\ell_n^2}\right) \\ & \leq \exp\left(\frac{-(n-i)\pi}{32\ell_n^2}\right). \end{aligned}$$

\square

Theorem 4 *If $k > 2$, then $E(C_k) = E(|\mathcal{C}_k(V)|) = O(\ell_n^2)$.*

Proof: Let \mathcal{B}_i be the event that $\rho_i \geq \frac{(n-i)\pi}{8\ell_n^2}$. By Lemma 3,

$$\mathbf{P}(I_i = 1) \geq \mathbf{P}(I_i = 1|\mathcal{B}_i)\mathbf{P}(\mathcal{B}_i) \geq \mathbf{P}(I_i = 1|\mathcal{B}_i)\left(1 - \exp\left(-\frac{(n-i)\pi}{32\ell_n^2}\right)\right) \quad (2)$$

Now suppose that $i \leq \lambda_n = n - \ell_n^2$, and observe that

$$\mathbf{P}(I_i = 1|\mathcal{B}_i) = \sum_{v \geq \frac{(n-i)\pi}{8\ell_n^2}} \mathbf{P}(I_i = 1|\rho_i = v)\mathbf{P}(\rho_i = v|\mathcal{B}_i). \quad (3)$$

To estimate this, observe that

$$\mathbf{P}(I_i = 1|\rho_i = v) = \int_{\mathcal{Q}} \mathbf{P}(I_i = 1|\rho_i = v, X_i = \vec{x})f_{X_i}(\vec{x}|\rho_i = v)d\vec{x} \quad (4)$$

where $f_{X_i}(\vec{x}|\rho_i = v)$ is the conditional density of X_i on the square \mathcal{Q} given that $\rho_i = v$. For $v > (n-i)\pi/8\ell_n^2$, Theorem 2 yields

$$\mathbf{P}(I_i = 1|\rho_i = v, X_i = x) \geq 1 - 3\alpha^v \geq 1 - 3\alpha^{(n-i)\pi/8\ell_n^2}. \quad (5)$$

Putting this back into (4) and then (3), we get

$$\mathbf{P}(I_i = 1|\mathcal{B}_i) \geq 1 - 3\alpha^{(n-i)\pi/8\ell_n^2}, \quad (6)$$

and therefore

$$\mathbf{P}(I_i = 1) \geq \mathbf{P}(I_i = 1|\mathcal{B}_i)\mathbf{P}(\mathcal{B}_i) \geq (1 - 3\alpha^{(n-i)\pi/8\ell_n^2})\left(1 - \exp\left(-\frac{(n-i)\pi}{32\ell_n^2}\right)\right) \quad (7)$$

$$\geq 1 - 3\alpha^{(n-i)\pi/8\ell_n^2} - \exp\left(-\frac{(n-i)\pi}{32\ell_n^2}\right). \quad (8)$$

Recall that $\lambda_n = n - \ell_n^2$, and that the foregoing estimates were valid for all $i \leq \lambda_n$. Putting $j = n - i$, we get

$$E(U_k) \geq \sum_{i=1}^{\lambda_n} \mathbf{P}(I_i = 1) = \sum_{i=1}^{\lambda_n} \left(1 - 3\alpha^{(n-i)\pi/8\ell_n^2} - \exp\left(-\frac{(n-i)\pi}{32\ell_n^2}\right)\right) \quad (9)$$

$$\geq \lambda_n - 3 \sum_{j \geq \ell_n^2} (\alpha^{\pi/8\ell_n^2})^j - \sum_{j \geq \ell_n^2} (\exp(-\pi/32\ell_n^2))^j \quad (10)$$

$$= n - O(\ell_n^2). \quad (11)$$

□

6 Relaxed Rule k Analysis

The following lemma is straight-forward, but is essential for the analysis of the Relaxed Rule k .

Lemma 5 $\mathcal{C}_\infty(V) \subseteq \mathcal{C}_k(V)$.

Proof: Suppose $x \notin \mathcal{C}_k$, i.e. x is removed from $\mathcal{C}_k(V)$ by the strict version of Rule k . This means there is a connected set \mathcal{S} of k vertices in $N(x)$ that dominate $N(x)$ and have higher ID's than x has. Because $\mathcal{S} \subseteq N^+(x)$, it is clear $N^+(x)$ also dominates $N(x)$. We need only check that $N^+(x)$ is connected. Let $u, v \in N^+(x)$. Since \mathcal{S} dominates $N(x)$, there are vertices $y, z \in \mathcal{S}$ such that $d(u, y) \leq 1$ and $d(v, z) \leq 1$. Because \mathcal{S} is connected, there must be a path γ from y to z in \mathcal{S} . Since $\mathcal{S} \subseteq N^+(X)$, this is also a path from y to z in $N^+(X)$. Let Γ be the path that consists of u , followed by the path γ , followed by v . Then Γ is a path from u to v in $N^+(x)$. Since u and v were arbitrary elements of $N^+(x)$, it follows that $N^+(x)$ is connected. So x is also removed from $\mathcal{C}_\infty(V)$. The result follows. \square

Now let $C_\infty = |\mathcal{C}_\infty(V)|$ be the size of the Relaxed Rule k dominating set. Since $\mathcal{C}_\infty(V) \subseteq \mathcal{C}_k(V)$, we certainly have $C_\infty \leq C_k$. The following theorem is therefore an immediate corollary to Theorem 4:

Theorem 6 $E(C_\infty) = O(\ell_n^2)$.

7 Lower Bound

If a vertex v has higher ID than any of its neighbors, then it cannot be eliminated under either version of Rule k . This simple observation is the basis for

Theorem 7 *If $\ell_n = o(\sqrt{n})$, then, for all sufficiently large n , the expected size of the Relaxed Rule k dominating set is more than $\ell_n^2/4$.*

Proof:

Let $L_n = \sum_{i=1}^n I_i$, where $I_i = 1$ iff node i has a higher ID than all the nodes in $D'_1(X_i) = D_1(X_i) \cap \mathcal{Q}$. Note that $I_i = 1$ iff the nodes $X_{i+1}, X_{i+2}, \dots, X_n$ all fall *outside* $D'_1(X_i)$. Therefore

$$\mathbf{P}(I_i = 1) = \left(1 - \frac{|D'_1(X_i)|}{\ell_n^2}\right)^{n-i} \geq \left(1 - \frac{\pi}{\ell_n^2}\right)^{n-i} \quad (12)$$

Therefore

$$E(L_n) \geq \sum_{i=1}^n \left(1 - \frac{\pi}{\ell_n^2}\right)^{n-i} = \frac{\ell_n^2}{\pi} \left(1 - \left(1 - \frac{\pi}{\ell_n^2}\right)^n\right) = \frac{\ell_n^2}{\pi} (1 - o(1)). \quad (13)$$

The result follows since $C_\infty \geq L_n$ always. \square

Corollary 8 *If $\ell_n = o(\sqrt{n})$, then, for all sufficiently large n , the expected size of the Strict Rule k dominating set is more than $\ell_n^2/4$.*

8 Optimality

For this section, $\ell_n \leq \sqrt{\frac{n}{a \log n}}$, where a is a constant greater than 9. It is easy to verify that, with asymptotic probability one, there exists a CDS, C_{rand} , having $O(\ell_n^2)$ vertices: simply partition the square \mathcal{Q} into $\lfloor 3\ell_n \rfloor^2$ equal-sized squares, each with sides of length $s_n = \frac{\ell_n}{\lfloor 3\ell_n \rfloor} = \frac{1}{3} + O(\frac{1}{\ell_n})$, and then pick one node from each of these small squares. More explicitly, for $0 \leq i, j < \lfloor 3\ell_n \rfloor$, let $Q_{i,j} = \{(x, y) : i s_n \leq x < (i+1)s_n \text{ and } j s_n \leq y < (j+1)s_n\}$. Let \mathcal{B} be the event that each of the $\lfloor 3\ell_n \rfloor^2$ small squares contains one or more nodes. By Boole's inequality,

$$\mathbf{P}(\mathcal{B}^c) \leq 9\ell_n^2 \mathbf{P}(Q_{1,1} \text{ is empty}) = 9\ell_n^2 \left(1 - \frac{1}{\lfloor 3\ell_n \rfloor^2}\right)^n \quad (14)$$

$$= 9\ell_n^2 \exp\left(-\frac{n}{9\ell_n^2} \left(1 + O\left(\frac{1}{\ell_n^2}\right)\right)\right) \quad (15)$$

$$< \frac{n}{\log n} e^{-\log n} = O\left(\frac{1}{\log n}\right). \quad (16)$$

Now given the vertices $V = \{X_1, X_2, \dots, X_n\}$, we construct $C_{rand} \subseteq V$ as follows: For each $1 \leq i, j \leq \lfloor 3\ell_n \rfloor$, if $Q_{i,j}$ contains at least one vertex, then select one vertex $V_{i,j}$ uniform randomly from among the vertices in $Q_{i,j}$, and include $V_{i,j}$ in C_{rand} . Thus C_{rand} is a (random) set of at most $\lfloor 3\ell_n \rfloor^2$ nodes. It can contain fewer nodes (possibly as few as one), but with asymptotic probability 1, C_{rand} contains exactly $\lfloor 3\ell_n \rfloor^2$ vertices and is a CDS.

It is worth pointing out that this existence argument cannot be used in a straight-forward way as the basis for a localized algorithm because the nodes do not know their own locations in the network. One of the main advantages of the Rule k algorithm is that a vertex makes its decision based on very limited information, namely its list of neighbors and their lists of neighbors. Nevertheless, the existence argument is useful for us because it leads to a lower bound on the size that a CDS can have. In particular, Theorem 9 below is based on the following observation: If v is any point in \mathcal{Q} , then at most 81 nodes of C_{rand} are in $D_1(v)$. In particular, if C_{opt} is a minimum sized CDS, and v is a node in C_{opt} , then $N(v)$ includes at most 81 nodes of C_{rand} . But C_{opt} is a dominating set; therefore every node in C_{rand} must be in $N(v)$ for at least one $v \in C_{opt}$. We therefore have a lower bound on the size of C_{opt} :

$$|C_{opt}| \geq \frac{1}{81} |C_{rand}|. \quad (17)$$

Combining (17) with (16), we get

Theorem 9 *Suppose $a > 9$, and $\ell_n \leq \sqrt{\frac{n}{a \log n}}$ for all n . Then there is a constant $B > 0$ such that, for all $n > 1$,*

$$\mathbf{P}_{n, \ell_n} \left(|C_{opt}| < \frac{1}{10} \ell_n^2 \right) < \frac{B}{\log n}.$$

The argument given above was influenced by [22]. The appendix of [11] is also pertinent, but we do not see how to turn the discussion there into a mathematically rigorous proof.

Corollary 10 $E(|C_{opt}|) = \Theta(\ell_n^2)$

Proof: For the upper bound, note that $|C_{opt}| \leq C_k$, and consequently $E(|C_{opt}|) \leq E(C_k) = O(\ell_n^2)$ by Theorem 4. For the lower bound, we use (17) to get

$$\begin{aligned} E(|C_{opt}|) &\geq \frac{1}{81} E(|C_{rand}|) \\ &\geq \frac{1}{81} \mathbf{P}(|C_{rand}| = \lfloor 3\ell_n^2 \rfloor) \cdot \lfloor 3\ell_n^2 \rfloor \\ &= \frac{1}{81} (1 - O(\frac{1}{\log n})) \lfloor 3\ell_n^2 \rfloor = \Theta(\ell_n^2). \end{aligned}$$

□

9 Discussion

We analyzed, for every $k > 2$, the strict version Rule k in which a node x is excluded from the dominating set $\mathcal{C}_k(V)$ iff it has a set of k higher numbered neighbors that dominate its entire neighborhood $N(x)$. We also analyzed the practical “ $k = \infty$ ” variant of Rule k in which a node x is excluded from the connected dominating set $\mathcal{C}_\infty(V)$ iff the subset $N^+(x)$ of its neighbors with greater IDs is connected and dominates the entire neighborhood of $N(x)$. The conclusion of this paper is that both the strict and relaxed versions of Rule k are optimal insofar as they yield a CDS whose expected size is bounded by a constant multiple of the size of the minimum CDS.

Acknowledgement We thank Harish Sethu for helpful comments. We are also grateful to an anonymous referee for suggesting the relaxed version of Rule k .

References

- [1] C.Adjih,P.Jacquet, L.Viennot,Computing Connected Dominating Sets with Multipoint Relays, *INRIA RR-4597* (2002).
- [2] K.M.Alzoubi,P.J.Wan,and O.Frieder, Distributed Heuristics for Connected Dominating Sets in Wireless Ad Hoc Networks, *Journal of Communications and Networks*,**4** (1) (2002) 1–8.
- [3] M.J.B.Appel and R.P.Russo, The minimum vertex degree of a graph on uniform points in $[0, 1]^d$, *Advances in Appl.Probab* **29** (3) (1997) 582–594.
- [4] E.Baccelli and P.Jacquet, Flooding Techniques in Mobile Ad Hoc Networks, *INRIA RR-5002* (2003).
- [5] T.Camp and B.Williams, Comparison of Broadcasting Techniques for Mobile Ad Hoc Networks, Proceedings of the 3’rd ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc 2002 (2002) 194–205.
- [6] M.Cardei,X.Cheng,X.Cheng, D.Du,Connected Domination in Multihop Ad Hoc Wireless Networks, *JCIS* (2002) 251–255.
- [7] G.B.Chae, E.M. Palmer, and W.C. Siu., Geodetic Number of Random Graphs of Diameter Two , *Australasian Journal of Combinatorics* **26** (2002) 11-20.
- [8] B. Chen, K. Jamieson, H. Balakrishnan, R. Morris , Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks *Proc. of the 6th ACM MOBICOM Conf.*, Rome, Italy, July 2001.
- [9] B.N. Clark,C.J. Colburn, and D.J.Johnson, Unit Disk Graphs,*Discrete Mathematics* **86(1-3)** (1990) 165–177.
- [10] B.Das and V. Bharghavan Routing in Ad-Hoc Networks Using Minimum Connected Dominating Sets, International Conference on Communications **1** (1997) 376–380.
- [11] F.Dai and J.Wu, An Extended Localized Algorithm for Connected Dominating Set Formation in Ad Hoc Wireless Networks, *IEEE Transactions on Parallel and Distributed Systems*, **15** (10) 2004.
- [12] F.Dai and J.Wu, Performance analysis of broadcast protocols in ad hoc networks based on self pruning, *IEEE Trans.on Parallel and Distributed Systems* **15** (11) (2004).
- [13] E.N.Gilbert, Random Plane Networks, *J.Soc.Indust.Appl.Math.* **9** (1961) 533.
- [14] A.Godbole and B.Wielund, On the Domination Number of a Random Graph, *Electronic Journal of Combinatorics* **8** #R37 (2001)

- [15] S.Guha and S.Khuller, Approximation algorithms for connected dominating sets, *Algorithmica* **20** (4) (1998) 374–387.
- [16] P.Gupta and P.R.Kumar, “Critical power for asymptotic connectivity in wireless networks”, in *Stochastic Analysis, Control, Optimization and Applications*, Birkhauser (1999) 547–566.
- [17] W.K.Hale, Frequency Assignment: Theory and Applications, *Proc. IEEE* **68** (1980) 1497–1514.
- [18] J.C.Hansen, E.Schmutz, and L.Sheng, unpublished manuscript at arXiv:cs.DM/0408068.
- [19] P.Jacquet, Analytical Results on Connected Dominating Sets in Mobile Ad Hoc Networks, *INRIA RR-5173* (2004).
- [20] P.Jacquet, A.Laouiti, P.Minet, L.Viennot, Performance of Multipoint Relaying in Ad Hoc Mobile Routing Protocols, In “Networking 2002” *Lecture Notes in Computer Science* **2345** (2002) 387–398.
- [21] D.Lichtenstein, Planar formulae and their uses, *SIAM J.Comput.* **11**(2) (1982) 329–343.
- [22] M.V.Marathe, H.Breu, H.B.Hunt, S.S.Ravi, and D.J.Rosenkrantz, Simple Heuristics for Unit Disk Graphs, *Networks* **25** no.2 (1995) 59–68.
- [23] C.McDiarmid, Discrete mathematics and radio channel assignment, *Recent Advances in algorithms and combinatorics*, 27–63, CMS Books Math (2003).
- [24] C.McDiarmid, Random channel assignment in the plane, *Random Structures and Algorithms* **22** (2) 187–212.
- [25] W.Peng and X.Lu, On the reduction of broadcast redundancy in mobile ad hoc networks, *Proceedings of the 1st ACM International Symposium on Mobile Ad Hoc Networking* (2000) 129–130.
- [26] *Random Geometric Graphs*, Oxford Studies in Probability **5**, Oxford University Press, (2003) ISBN 0-19-850626-0.
- [27] R.Sivakumar, B.Das, and V. Bharghavan, Spine-based routing in ad hoc networks, *Cluster Computing* **1** (2) (1998) 237–248.
- [28] I. Stojmenovic, M. Seddigh, J. Zunic, Dominating sets and neighbor elimination based broadcasting algorithms in wireless networks, *IEEE Transactions on Parallel and Distributed Systems*, Vol. **13**, No. **1**, (2002), 14–25.
- [29] Y.C.Tseng, S.Y.Ni, Y.S.Chen, J.P.Sheu, The broadcast storm problem in a mobile ad hoc network, *Wireless Networks* **8**(2-3) (2002) 153–167.
- [30] J. Wu and H.Li, On calculating connected dominating set for efficient routing in ad hoc wireless networks, *Workshop on Discrete Algorithms and Methods for MOBILE Computing and Communications* (1999) 7–14.
- [31] A.C.C.Yao, On constructing spanning trees in k dimensional spaces and related problems, *SIAM J.Computing* **11** (4) (1982) 721–736.