The domination number of on-line social networks and random geometric graphs⋆

Anthony Bonato[1], Marc Lozier[1], Dieter Mitsche[2],
Xavier Pérez-Giménez[1], and Paweł Prałat[1]

[1] Ryerson University, Toronto, Canada,
abonato@ryerson.ca, marc.lozier@ryerson.ca,
xperez@ryerson.ca, pralat@ryerson.ca
[2] Université de Nice Sophia-Antipolis, Nice, France
dmitsche@unice.fr

Abstract. We consider the domination number for on-line social networks, both in a stochastic network model, and for real-world, networked data. Asymptotic sublinear bounds are rigorously derived for the domination number of graphs generated by the memoryless geometric protean random graph model. We establish sublinear bounds for the domination number of graphs in the Facebook 100 data set, and these bounds are well-correlated with those predicted by the stochastic model. In addition, we derive the asymptotic value of the domination number in classical random geometric graphs.

1. Introduction

On-line social networks (or OSNs) such as Facebook have emerged as a hot topic within the network science community. Several studies suggest OSNs satisfy many properties in common with other complex networks, such as: power-law degree distributions [2, 13], high local clustering [36], constant [36] or even shrinking diameter with network size [23], densification [23], and localized information flow bottlenecks [12, 24]. Several models were designed to simulate these properties [19, 20], and one model that rigorously asymptotically captures all these properties is the geometric protean model (GEO-P) [5–7] (see [16, 25, 30, 31] for models where various ranking schemes were first used, and which inspired the GEO-P model). For a survey of OSN models see [8], and for more general complex networks [3]. A fundamental difference with GEO-P versus other models [2, 21, 23, 22] is that it posits an underlying feature or metric space. This metric space mirrors a construction in the social sciences called *Blau space* [26]. In Blau space, agents in the social network correspond to points in a metric space, and the relative position of nodes follows the principle of *homophily* [27]: nodes with similar socio-demographics are closer together in the space. We give the precise definition of the GEO-P model (actually, one of its variants, the so-called MGEO-P model) below. We focus on the MGEO-P model, since it simpler than GEO-P and generates graphs with similar properties.

The study of domination and dominating sets plays a prominent role in graph theory with a number of application to real-world networks. A *dominating set* in a graph G is a set of nodes S in G such that every node not in S is adjacent to at least one node in S. The *domination number* of G, written $\gamma(G)$, is the minimum cardinality of a dominating set in G. Computing $\gamma(G)$ is a well-known **NP**-complete problem, so typically heuristic algorithms are used to compute it for large-scale networks. Dominating sets appear in numerous applications such as: network controllability [11], as a centrality measure for efficient data routing [34], and detecting

---

biologically significant proteins in protein-protein interaction network [28]. For more additional background on domination in graph theory, see [15].

In social networks, we consider the hypothesis that minimum order dominating sets contain agents with strong influence over the rest of the network. Our goal in the present paper is to consider the problem of finding bounds on dominating sets in stochastic models of OSNs, and also in real-world data derived from OSNs. We consider bounds on the domination number of a stochastic model (see next paragraph), and upper bounds for that model are well-correlated with real-world OSN data. We note that the domination number has been studied previously in complex network models, including preferential attachment [9], and recently in [29].

The OSN model we consider is called the *memoryless geometric protean model (MGEO-P)*, first introduced in [4]. The MGEO-P model depends on five parameters which consists of: the number of nodes n, the dimension of the metric space m, the attachment parameter $0 < \alpha < 1$, the density parameter $0 < \beta < 1 - \alpha$, and the connection probability $0 < p \leq 1$.

The nodes and edges of the network arise from the following process. Initially, the network is empty. At each of n steps, a new node v arrives and is assigned both a random position q_v in $\mathbb{R}^m$ within the unit-hypercube $[0, 1)^m$ and a random rank r_v from those unused ranks remaining in the set 1 to n. The influence radius of any node v is computed based on the formula:

$$I(r_v) = \tfrac{1}{2}\big(r_v^{-\alpha} n^{-\beta}\big)^{1/m}.$$

With probability p, the node v is adjacent to each existing node u satisfying $\mathcal{D}(v, u) \leq I(r_u)$, where the distances are computed with respect to the following metric:

$$\mathcal{D}(v, u) = \min\left\{\|q_v - q_u - z\|_\infty : z \in \{-1, 0, 1\}^m\right\},$$

and where $\|\cdot\|_\infty$ is the infinity-norm. We note that this implies that the geometric space is symmetric in any point as the metric "wraps" around like on a torus. The volume of the space influenced by the node is $r_v^{-\alpha} n^{-\beta}$. Then the next node arrives and repeats the process until all n nodes have been placed. We refer to this model by MGEO-P(n, m, α, β, p).

We give rigorous bounds on the domination number of a typical graph generated by the MGEO-P. An event A_n holds *asymptotically almost surely* (*a.a.s.*) if it holds with probability tending to 1 as n tends to infinity. Our main result on MGEO-P is the following.

Theorem 1. *If $m = o(\log n)$, then a.a.s. the domination number of a graph G sampled from the MGEO-P(n, m, α, β, p) model satisfies*

$$\gamma(G) = \Omega(C^{-m/(1-\alpha)} n^{\alpha+\beta}) \quad and \quad \gamma(G) = O(n^{\alpha+\beta} \log n),$$

where C is any constant greater than 6. In particular, a.a.s. $\gamma(G) = n^{\alpha+\beta+o(1)}$.

We defer the proof of Theorem 1 to Section 2. It is noteworthy that the domination number of the preferential attachment model is linear in the order of the graphs sampled; see [9]; this fact further demonstrates the differences between MGEO-P and other complex graph models.

Theorem 1 suggests a sublinear bound on the domination number for OSNs, and we evidence for this in real-world data. In Section 3, we find bounds for graphs in the Facebook 100 data set, and compare these results to those for the stochastic models. We chose to work with the so-called *Facebook 100* (or FB100) data set, as it provides representative samples from the network of increasing orders. Hence, we may consider trends for the domination number in the data.

While the data presented is our first and initial study, the bounds we find for the domination number of FB100 are of sublinear order, and these bounds are well-correlated with those from MGEO-P. Sublinear domination results for other complex networks were also reported in [29]; our approach is distinct as we consider social networks of increasing orders.

In addition to the results above, we find rigorous bounds on the domination number for classical random geometric graphs. Given a positive integer n, and a non-negative real r, we consider a *random geometric graph* $G = (V, E) \in \mathcal{G}(n, r)$ defined as follows. The node set V of G is obtained by choosing n points independently and uniformly at random in the square $\mathcal{S} = [0, 1]^2$. (Note that, with probability 1, no point in $\mathcal{S}$ is chosen more than once, and hence, we may assume that $|V| = n$.) For notational purposes, we identify each node $v \in V$ with its corresponding geometric position $v = (v_x, v_y) \in \mathcal{S}$, where v_x and v_y denote the usual x- and y-coordinates in $\mathcal{S}$, respectively. Finally, the edge set E is constructed by connecting each pair of nodes u and v by an edge if and only if $d_E(u, v) \leq r$, where d_E denotes the Euclidean distance in $\mathcal{S}$.

Random geometric graphs were first introduced in a slightly different setting by Gilbert [14] to model the communications between radio stations. Since then several closely related variants on these graphs have been widely used as a model for wireless communication, and have also been extensively studied from a mathematical point of view. The basic reference on random geometric graphs is the monograph by Penrose [32].

We note that our study is the first to explicitly provide provable bounds on the domination number of random geometric graphs. In particular, we derive the following result.

Theorem 2. *Let $G \in \mathcal{G}(n, r)$ and let $\omega = \omega(n)$ be any function tending to infinity as $n \to \infty$. Then a.a.s. the following holds:*

(a) *Denote by $N(x)$ the minimal number of balls of radius x needed to cover $\mathcal{S}$. If $r = \Theta(1)$, then*

$$\Omega(1) = N(r + \sqrt{\omega \log n/n}) \leq \gamma(G) \leq N(r - \omega/\sqrt{n}).$$

(b) *Define $C = 2\pi\sqrt{3}/9 \approx 1.209$. If $\omega\sqrt{\log n/n} \leq r = o(1)$, then*

$$\gamma(G) = (C/\pi + o(1))r^{-2}.$$

(c) *If $1/\sqrt{n} \leq r < \omega\sqrt{\log n/n}$, then*

$$\gamma(G) = \Theta(r^{-2}).$$

(d) *If $r < 1/\sqrt{n}$, then*

$$\gamma(G) = \Theta(n).$$

It is straightforward to verify that the bounds on $\gamma(G)$ in part (a) differ by at most 1 if ω is sufficiently small, but in general we do not give accurate estimations of $N(r)$ for $r = \Theta(1)$. The proof of Theorem 2 is deferred to Section 4. The final section summarizes our results and presents open problems.

2. Proof of Theorem 1

For each node $v \in [n]$, we consider the ball $B_v = \{x \in [0, 1)^m : \mathcal{D}(x, v) \leq I(r_v)\}$, which has volume $b_v = r_v^{-\alpha} n^{-\beta}$. The next lemma will be useful to estimate the sum of volumes of the balls corresponding to a set of nodes.

Lemma 1. *Let T be a set of t nodes (fixed before ranks are chosen) with $\omega n^\alpha \log n \le t \le n$, for a function ω going to infinity with n arbitrarily slowly.*

(a) Then a.a.s.

$$\sum_{i \in T} r_i^{-\alpha} = (1 + o(1)) \frac{tn^{-\alpha}}{1 - \alpha}. \tag{1}$$

(b) Furthermore, given any integer s such that $1 \le s \le t$ and $s^{1-\alpha} \ge \omega(n/t)^\alpha \log n$, a.a.s. all subsets $S \subseteq T$ of s nodes satisfy

$$\sum_{i \in S} r_i^{-\alpha} \le (1 + o(1)) \frac{s^{1-\alpha}(t/n)^\alpha}{1 - \alpha}. \tag{2}$$

Observe that the sum in (1) is asymptotic to what one would expect. Indeed, if the ranks of the nodes in T are distributed evenly, then one would obtain $\sum_{i \in T}^t r_i^{-\alpha} = \sum_{i=1}^t (in/t)^{-\alpha} = tn^{-\alpha}/(1 - \alpha) + O(1)$.

Proof. Let s and t be integers satisfying all the conditions of the statement in part (b). Set $\hat\omega = \omega^{1/4} \to \infty$, so we have $t \ge \hat\omega^4 (t/s)^{1-\alpha} n^\alpha \log n$. This also implies $\hat\omega = o((sn/t)^{1-\alpha})$. Let Y_j be the number of elements in T with rank at most j. Observe that Y_j has expectation jt/n, and follows a hypergeometric distribution. For $(sn/t)^{1-\alpha}/\hat\omega \le j \le n$, a Chernoff bound (see e.g. [17]) gives that

$$\mathbf{Pr}\left(\left| Y_j - \frac{jt}{n} \right| \ge (1/\hat\omega)\frac{jt}{n} \right) \le 2\exp\left(-\frac{jt}{3\hat\omega^2 n} \right) \le 2e^{-\hat\omega \log n/3} = o(1/n^2).$$

We apply a union bound over all j, and conclude that a.a.s., for every $(sn/t)^{1-\alpha}/\hat\omega \le j \le n$,

$$(1 - 1/\hat\omega)jt/n < Y_j < (1 + 1/\hat\omega)jt/n.$$

In order to estimate the sums in the statement, we assume w.l.o.g. that $T = [t]$ and $r_1 < r_2 < \cdots < r_t$ (otherwise we permute the indices of the vertices in T). It follows that a.a.s., for every $(sn/t)^{1-\alpha}/\hat\omega \le j \le n$,

$$r_{\lfloor (1-1/\hat\omega)jt/n \rfloor} \le j \le r_{\lceil (1+1/\hat\omega)jt/n \rceil}.$$

Therefore, setting $\ell = 2s^{1-\alpha}t^\alpha n^{-\alpha}/\hat\omega$, we have that a.a.s., for every $\ell \le i \le t$,

$$\left\lfloor \frac{1}{1 + 1/\hat\omega} in/t \right\rfloor \le r_i \le \left\lceil \frac{1}{1 - 1/\hat\omega} in/t \right\rceil.$$

For the lower bound on r_i below, we need to use the fact that $\left\lfloor \frac{1}{1+1/\hat\omega} in/t \right\rfloor \ge (sn/t)^{1-\alpha}/\hat\omega$, which is easily verified to be true since $\hat\omega = o\left((sn/t)^{1-\alpha}\right)$. Finally, we infer that a.a.s., for any choice of S,

$$\sum_{i \in S} r_i^{-\alpha} \le \sum_{i=1}^s r_i^{-\alpha} = (1 + o(1))\left(\frac{n}{t}\right)^{-\alpha}\sum_{i=\ell}^s i^{-\alpha} + O(\ell) = (1 + o(1))\frac{1}{1-\alpha}s^{1-\alpha}t^\alpha n^{-\alpha}.$$

This proves statement (b). For statement (a), take $s = t$ and note that for this choice of s, for any $\omega n^\alpha \log n \le t \le n$, the condition $s^{1-\alpha} \ge \omega(n/t)^\alpha \log n$ is satisfied. Observe that then $S = T = [t]$, so the first inequality in the above equation is an equality. $\qquad\square$

Upper bound: Fix a constant $K > \frac{1-\alpha}{p}$, and let D be the set containing the first $t = \lfloor Kn^{\alpha+\beta} \log n \rfloor$ nodes added in the process. We will show that a.a.s. D is a dominating set. By Lemma 1, we may condition on the event that (1) holds for $t = |D| = \lfloor Kn^{\alpha+\beta} \log n \rfloor$. Note that this assumption on the ranks does not affect the distribution of the location of the nodes in $[0,1)^m$. Therefore, given a node $u > t$ (appearing in the process later than nodes in D), the probability that u is not dominated by D is

$$\prod_{i=1}^{t}(1 - pr_i^{-\alpha}n^{-\beta}) \leq \exp\left(-pn^{-\beta}\sum_{i=1}^{t}r_i^{-\alpha}\right) = \exp\left(-(1+o(1))\frac{p}{1-\alpha}tn^{-\alpha-\beta}\right)$$

$$= \exp\left(-(1+o(1))\frac{Kp}{1-\alpha}\log n\right) = o(1/n).$$

Taking a union bound over all nodes not in D, we can guarantee that a.a.s. all nodes are dominated.

As an alternative and relatively simple approach, one may prove the same upper bound on the domination number as follows. First, show that a.a.s. the minimum degree δ is at least $(1+o(1))pn^{1-\alpha-\beta}$. Then we may use Theorem 1.2.2 in [1], which states that for every graph G with minimum degree δ,

$$\gamma(G) \leq n\frac{1 + \log(\delta+1)}{\delta+1}. \tag{3}$$

Lower bound: We consider for convenience a natural directed version of MGEO-P(n, m, α, β, p), by orienting each edge from its "younger" end node (that is, appearing later in the process) to its "older" end node. For a set of nodes $D \subseteq [n]$, $N_{\text{in}}(D)$ denotes the set of nodes $u \in [n] \setminus D$ such that there is a directed edge from u to some node in D or, equivalently, such that there is an edge from u to some node in D that is older than u. $N_{\text{out}}(D)$ is defined analogously, replacing older by younger.

Define $t = \lceil n^{\alpha+\beta} \rceil$, and let $T = [t]$ be the set of the oldest t nodes in the process. We want to show that a.a.s. there is no dominating set of order at most $\xi\mu^{-m}n^{\alpha+\beta}$, where ξ and μ are specified later. We give more power to our adversary by allowing her to pick (deterministically after the graph has been revealed) two sets of nodes D_1 and D_2 of order $\lfloor \xi\mu^{-m}n^{\alpha+\beta} \rfloor$ each (not necessary disjoint). Her goal is also easier than the original one; she needs to achieve that, for every node $v \in T$, either v is in-dominated by D_1 (that is, $v \in D_1 \cup N_{\text{in}}(D_1)$) or v is out-dominated by D_2 (that is, $v \in D_2 \cup N_{\text{out}}(D_2)$); nothing is required for young nodes in $[n] \setminus T$. We show that a.a.s. the adversary cannot succeed, that is, regardless of her choice of D_1, D_2 we have always some node in T not in $D_1 \cup D_2 \cup N_{\text{in}}(D_1) \cup N_{\text{out}}(D_2)$.

Out-domination: Given a constant $0 < \varepsilon < 1$, we define T' to be the set of nodes in T with rank greater than $(1-\varepsilon)n$. Note that $|T'|$ has a hypergeometric distribution, so it follows easily from Chernoff's bound (see [17]) that a.a.s. $|T'| \geq (\varepsilon/2)n^{\alpha+\beta}$. For convenience, we choose $\varepsilon = \varepsilon(\alpha)$ to be the only real in $(0,1)$ satisfying $\varepsilon = 2(1-\varepsilon)^{\alpha}$. For every node $i \in T'$, the corresponding ball B_i has length at most

$$((1-\varepsilon)^{-\alpha}n^{-\alpha-\beta})^{1/m} = ((2/\varepsilon)n^{-\alpha-\beta})^{1/m}.$$

We consider a tessellation of $[0,1)^2$ into large cells. At the centre of each large cell we consider a smaller cell. Small cells have side length $((2/\varepsilon)n^{-\alpha-\beta})^{1/m}$ and large ones have side length

$2((2/\varepsilon)n^{-\alpha-\beta})^{1/m}$. There are

$$N = \left\lfloor \frac{1}{2}((\varepsilon/2)n^{\alpha+\beta})^{1/m} \right\rfloor^m = \frac{\varepsilon}{2}(2 + o(1))^{-m}n^{\alpha+\beta} \to \infty$$

large cells fully contained in $[0, 1)^m$ (we discard the rest), and thus N small cells inside of those. By construction, if a node in T' falls into a small cell, then its ball is contained into in the corresponding large cell. Let $\mathcal{X}$ be the set of small cells that contain at least one node in T', and let $T'' \subseteq T'$ be a set of $X = |\mathcal{X}|$ nodes such that each cell in $\mathcal{X}$ contains precisely one node in T'' (if a given small cell contains at least two nodes in T', then a node is selected arbitrarily to be placed in T''). Vertices in T'' are potentially dangerous for the adversary since, one node in D_2 can "out-dominate" at most one single node in T''. However, she may in theory get lucky and in-dominate many of these nodes (in the next section we will show that this will not happen a.a.s.).

We want to show that a.a.s. $X \geq N/4$. The probability that there are at least $3N/4$ small cells containing no nodes in T' is at most

$$\binom{N}{\lceil 3N/4 \rceil} \left(1 - (3N/4)(2/\varepsilon)n^{-\alpha-\beta}\right)^{(\varepsilon/2)n^{\alpha+\beta}} \leq 2^N \exp(-3N/4) = o(1).$$

Therefore,

$$X \geq N/4 = \frac{\varepsilon}{8}\lambda^{-m}n^{\alpha+\beta}, \qquad \text{for some } \lambda = 2 + o(1). \tag{4}$$

In-domination: Let $\xi = \xi(\alpha)$ be a sufficiently small positive constant, and define

$$\mu = \left(\lambda\left(1 + 2(2/\varepsilon)^{1/m}\right)\right)^{1/(1-\alpha)} > 3\lambda.$$

The adversary chooses a set $D_1 \subseteq T$ of $s = \lfloor \xi\mu^{-m}n^{\alpha+\beta} \rfloor$ nodes in her attempt to in-dominate T'. By Lemma 1(b), a.a.s. regardless of her choice,

$$\sum_{i \in D_1} r_i^{-\alpha} \leq \frac{(1+o(1))}{1-\alpha}s^{-\alpha+1}t^\alpha n^{-\alpha} = (1 + o(1))\frac{\xi^{1-\alpha}}{1-\alpha}\mu^{-(1-\alpha)m}n^\beta. \tag{5}$$

We tessellate the space into cells of volume $(2/\varepsilon)n^{-\alpha-\beta}$ (same size as the small cells in the out-domination part, but now we have the whole space partitioned into cells of that size). Recall that, for each node $i \in D_1$, the ball B_i has length $b_i^{1/m} \geq n^{-(\alpha+\beta)/m}$. Therefore, the volume of the set of cells intersected by B_i is at most

$$\left(b_i^{1/m} + 2\left((2/\varepsilon)n^{-\alpha-\beta}\right)^{1/m}\right)^m \leq \left(1 + 2(2/\varepsilon)^{1/m}\right)^m b_i = \left(\mu^{1-\alpha}/\lambda\right)^m b_i.$$

Combining this and (5), a.a.s. and regardless of the adversary's choice, the total volume of the cells intersected by the balls of the nodes in D_1 is at most

$$\sum_{i \in D_1} \left(\mu^{1-\alpha}/\lambda\right)^m b_i = \left(\mu^{1-\alpha}/\lambda\right)^m n^{-\beta} \sum_{i \in D_1} r_i^{-\alpha} \leq (1 + o(1))\frac{\xi^{1-\alpha}}{1-\alpha}\lambda^{-m}. \tag{6}$$

Let $\mathcal{Y}$ be the set of cells intersected by the balls of the nodes in D_1, and put $Y = |\mathcal{Y}|$. By (6), a.a.s.

$$Y \leq \frac{\varepsilon\xi^{1-\alpha}}{2(1-\alpha)}\lambda^{-m}n^{\alpha+\beta}.$$

Thus, in view of (4) we just need to make ξ small enough so that $|\mathcal{X} \setminus \mathcal{Y}|$ is larger than $|D_2| = \lfloor \xi \mu^{-m} n^{\alpha+\beta} \rfloor$. That is because dangerous cells in $\mathcal{X} \setminus \mathcal{Y}$ contain nodes in T' that are not in-dominated by D_1, and each one of these cells requires one different node in D_2 to out-dominate its nodes. Recall that our choice of $\varepsilon \in (0,1)$ depends only on α. Then picking ξ sufficiently small so that $\frac{\varepsilon \xi^{1-\alpha}}{2(1-\alpha)} + \xi < \frac{\varepsilon}{8}$, we get

$$|\mathcal{X} \setminus \mathcal{Y}| \geq X - Y \geq \left(\frac{\varepsilon}{8} - \frac{\varepsilon \xi^{1-\alpha}}{2(1-\alpha)} \right) \lambda^{-m} n^{\alpha+\beta} > \xi \mu^{-m} n^{\alpha+\beta} \geq \lfloor \xi \mu^{-m} n^{\alpha+\beta} \rfloor = |D_2|,$$

where we also used that $\mu > 3\lambda > \lambda$. Finally, distinguishing the cases $m = O(1)$ and $m \to \infty$, we observe that

$$\mu^{-m} = \left(\lambda (1 + 2(2/\varepsilon)^{1/m}) \right)^{-m/(1-\alpha)} = \begin{cases} \Theta(1) & \text{for } m = O(1), \\ (6 + o(1))^{-m/(1-\alpha)} & \text{for } m \to \infty, \end{cases}$$

so the claimed lower bound follows.

3. Domination in Facebook 100 graphs

Facebook distributed 100 samples of social networks from universities within the United States measured as of September 2005 [35], which range in size from 700 nodes to 42,000 nodes. We call these networks the *Facebook 100* (or simply FB100) graphs. As the domination number is sensitive to nodes of low degree, we used the k-core of the network, where $1 \leq k \leq 5$; see [33]. For $k \in \mathbb{N}$, the *k-core* of a graph is the largest induced subgraph of minimum degree at least k. The k-core can be found by a simple node deletion algorithm that repeatedly deletes nodes with degree less than k. This algorithm always terminates with the k-core of the graph, which is possibly empty.

Several algorithms were used to bound the domination number of the FB100 graphs, but one providing the smallest dominating sets is an adaptation of the *DS-DC* algorithm [28]. In the algorithm, initially all nodes V are in the dominating set S. It then selects a node u of minimum degree in S, and deletes it only if the set $S \setminus \{u\}$ remains dominating. The algorithm then repeats these steps for all nodes in S in order of their increasing degrees. We considered other algorithms, such as greedy algorithms where high degree nodes are added to an empty dominating set sequentially, or by choosing a random dominating set, but DS-DC outperformed these algorithms. We omit a detailed discussion of the performance of other algorithms owing to space.

Figure 1 presents the DS-DC predicted upper bounds on $\gamma(G)$, where G is a graph in the FB100 data set. We plotted the upper bound predicted by the MGEO-P model in Theorem 1, and we note the close similarity between that bound and ones for FB100. Note that we ignore constants in the big Oh term in the upper bound from the model, and simply plot the bound generated by $n^{\alpha+\beta} \log n$. The values for α, β, and the dimension parameter m for each of the FB100 graphs are taken from tables provided in [4]. (For example, in order to determine the power-law exponent, the Clauset-Shalizi-Newman power law exponent estimator was used; see [4] for more details.) The MGEO-P bound seems well-correlated with the bounds provided in the k-core, especially where $k = 3, 4, 5$. See Table 1, which fits the domination number of the FB100 graphs to the curve $y = n^x \log n$.
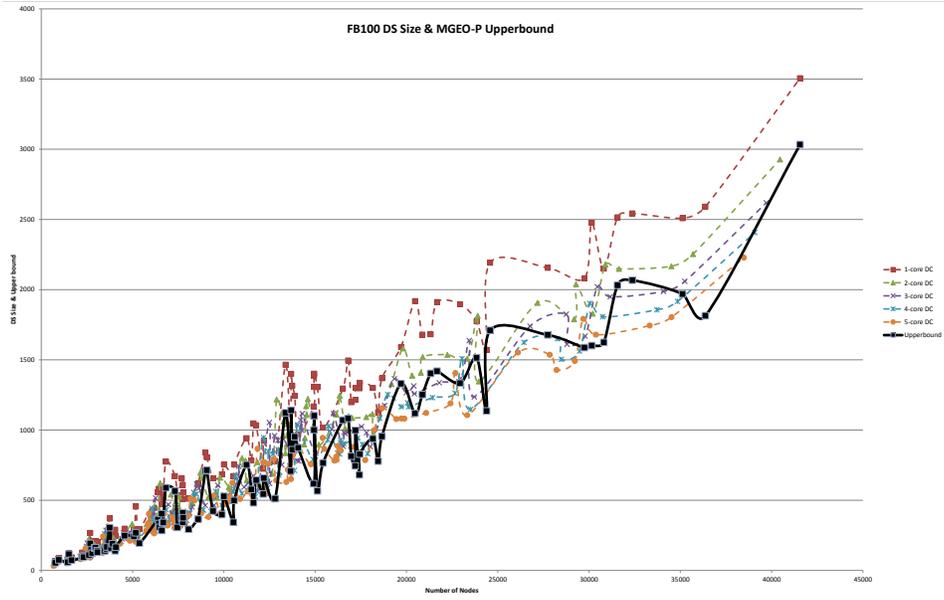
Fig. 1. Upper bounds on the domination number of the FB100 networks vs MGEO-P.

k	x	R^2
1	0.509	0.8472
2	0.492	0.8292
3	0.4818	0.8179
4	0.4741	0.8093
5	0.4677	0.803

Table 1. Fitting the domination number of the k-cores of FB100 to $y = n^x \log n$, suggesting a sub-linear trend.
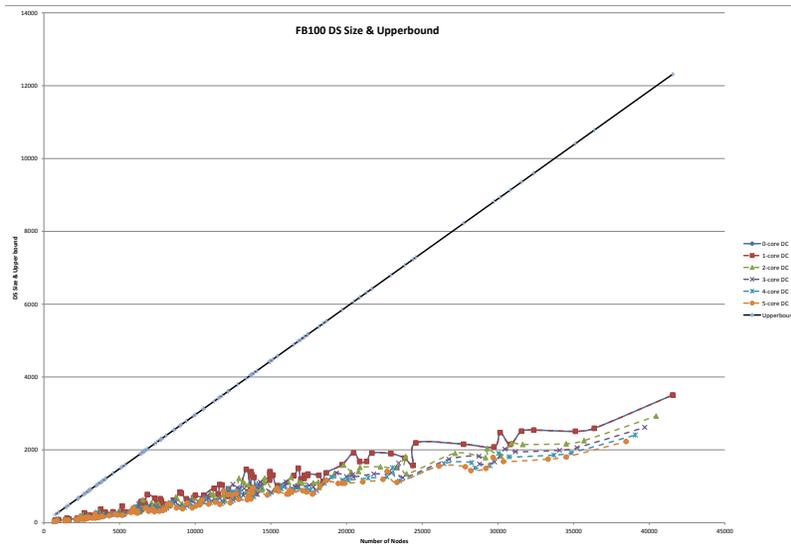

Fig. 2. Upper bounds on the domination number of the FB100 networks vs the bound in (3), showing a substantial overestimation

To contrast the bounds provided in Figure 1 with the bound in (3), we plot them in Figure 2. We plotted the theoretical bound using $\delta = 5$ (that is, the minimum degree of the 5-core). The figure shows a significant over-estimate of the domination number of the bound in (3), further corroborating the claim that the domination numbers of the FB100 graphs are sublinear with respect to the order of the graph.

4.   Proof of Theorem 2

We will relate the domination number to the problem of covering the plane with circles. Given $x \in \mathbb{R}^2$ and $\rho > 0$, we denote by $\mathcal{B}(x, \rho)$ the ball with centre x and of radius ρ. The following theorem is well known [18].

Theorem 3 ([18]). *Given a bounded subset of the plane M, for $\varepsilon > 0$ let $N(\varepsilon)$ be the minimum number of balls of radius ε that can cover M. Then we have that*

$$\lim_{\varepsilon \to 0} \pi \varepsilon^2 N(\varepsilon) = C \operatorname{Area}(\overline{M}),$$

where $\overline{M}$ denotes the closure of M.

Observe that $(C - 1)$ can therefore, be seen as measuring the proportion of unavoidable overlapping. Moreover, [18] shows that an optimal covering of the square $\mathcal{S}$ using balls of radius ε corresponds to arranging the balls in such a way that their centers are the centers of the cells of a hexagonal tiling of length ε. More precisely, consider the lattice

$$\mathcal{L}_\varepsilon = \{ i\varepsilon(\sqrt{3}, 0) + j\varepsilon(\sqrt{3}/2, 3/2) : i, j \in \mathbb{Z} \}. \tag{7}$$

Then the set of balls of radius ε and centre in $\mathcal{L}_\varepsilon$ that intersect $\mathcal{S}$ form a covering of $\mathcal{S}$ that gives the limit in Theorem 3.

Note that for all G with maximum degree Δ, we trivially have $\gamma(G) \geq n/(1 + \Delta(G))$ (for further relations between $\gamma(G)$ and other graph parameters, see, for example, [15]). Given any constant $c > 0$, for $G \in \mathcal{G}(n, r)$ with $r \geq c\sqrt{\log n/n}$, it is easy to show, by Chernoff bounds together with union bounds, that a.a.s. $\Delta(G) = O(r^2 n)$. Therefore, a.a.s. we have $\gamma(G) = \Omega(r^{-2})$. On the other hand, we can trivially construct a dominating set of $\mathcal{G}(n, r)$ by tessellating $\mathcal{S}$ into square cells of side length $r/\sqrt{2}$ and picking one node from each cell (if the cell is not empty). This holds deterministically for any geometric graph (not necessarily random), with no restriction on r, and gives $\gamma(G) = O(r^{-2})$. It follows that, for $G \in \mathcal{G}(n, r)$ with $r \geq c\sqrt{\log n/n}$, a.a.s. $\gamma(G) = \Theta(r^{-2})$.

We first prove the lower bound in part (b). Fix an arbitrarily small constant $\delta > 0$. Tessellate $\mathcal{S}$ into cells of side length $\alpha = \sqrt{\omega \log n/n} = o(r)$. By Chernoff bounds together with a union bound over all cells, we get that a.a.s. each cell contains at least one node. We may condition on this event, and proceed deterministically. For a contradiction, suppose that there exists a dominating set of size $s = \lfloor (C/\pi - \delta) r^{-2} \rfloor$. Consider then s balls whose centers are at the nodes of the dominating set. When using radius r for all balls, each cell is at least touched by some ball, since each cell is non-empty and each node is covered. Hence, by using radius $r' = r + \alpha\sqrt{2}$, each square is totally covered by some ball. Therefore, $\mathcal{S}$ can be covered by $\lfloor (C/\pi - \delta) r^{-2} \rfloor$ balls of radius r'. On the other hand,

$$\pi r'^2 \lfloor (C/\pi - \delta) r^{-2} \rfloor \leq (r^2 + 2\sqrt{2} r\alpha + 2\alpha^2)(C - \pi \delta) r^{-2} = (1 + o(1))(C - \pi\delta),$$

since $\alpha = o(r)$. This contradicts Theorem 3 and so $\gamma(G) > s$. Since the argument holds for any $\delta > 0$, we get the desired lower bound.

For the upper bound, we will show that we can find a covering of $\mathcal{S}$ with $(C/\pi + o(1))r^{-2}$ balls of radius r that are centered at some nodes of G. Again, fix some arbitrarily small constant $\delta > 0$. Let $r' = (1 - \delta)r$, and consider the lattice $\mathcal{L}_{r'}$, as defined in (7). Let $\mathcal{L}'_{r'}$ be the set of all points $x \in \mathcal{L}_{r'}$ such that the ball with centre x and radius r' intersects $\mathcal{S}$. Recall that $\mathcal{L}'_{r'}$ gives the optimal covering of $\mathcal{S}$ with balls of radius r', and therefore, attains the bound given by Theorem 3

$$ s = |\mathcal{L}'_{r'}| = \left(\frac{C}{\pi r'^2}\right)(1 + o(1)) = \left(\frac{C}{\pi}\right)r^{-2}(1 - \delta)^{-2}(1 + o(1)). $$

It might happen that some point $x \in \mathcal{L}'_{r'}$ does not belong to $\mathcal{S}$. In this case, we replace x by the closest point $\hat{x}$ on the boundary of $\mathcal{S}$ (this can be uniquely done, since $\mathcal{S}$ is closed and convex). Note that $\mathcal{B}(x, r') \cap \mathcal{S} \subseteq \mathcal{B}(\hat{x}, r') \cap \mathcal{S}$. We denote $\hat{\mathcal{L}}_{r'}$ the modified set of points that we obtained. By construction, $\hat{\mathcal{L}}_{r'} \subseteq \mathcal{S}$, and we can cover $\mathcal{S}$ using balls with centre in $\hat{\mathcal{L}}_{r'}$ and radius r' or larger. Moreover, $|\hat{\mathcal{L}}_{r'}| = s$. Clearly, if we can guarantee that for each $x \in \hat{\mathcal{L}}_{r'}$ there exists a node of G inside $\mathcal{B}(x, \delta r) \cap \mathcal{S}$, then G is dominated by these nodes, and hence, s yields an upper bound for $\gamma(G)$.

Observe that for any point $x \in \hat{\mathcal{L}}_{r'}$ (and therefore, in $\mathcal{S}$), the area of $\mathcal{B}(x, \delta r) \cap \mathcal{S}$ is at least $(\delta r)^2 \pi / 4$, since at least a quarter of a ball must be inside $\mathcal{S}$. The probability that there is no node of G in $\mathcal{B}(x, \delta r) \cap \mathcal{S}$ is at most

$$ \left(1 - \frac{(\delta r)^2 \pi}{4}\right)^n \leq \exp\left(-\frac{n(\delta r)^2 \pi}{4}\right) \leq \exp\left(-\frac{\omega^2 \delta^2 \pi \log n}{4}\right) = o(n^{-2}). $$

Since there are s events that we need to investigate and clearly $s \leq n$, by a union bound, a.a.s., for every $x \in \hat{\mathcal{L}}_{r'}$, the region $\mathcal{B}(x, \delta r) \cap \mathcal{S}$ contains at least one node of G. It follows that a.a.s. $\gamma(G) \leq s$ and since the argument holds for any $\delta > 0$, we derive the desired upper bound.

For the proof of part (a), note that $N(x)$ is non-decreasing function of x, and $N(x) = 1$ for $x \geq 1/\sqrt{2}$. Fix $r = \Theta(1)$. Tessellate $\mathcal{S}$ into cells of side length $\alpha = \sqrt{(\omega/2)\log n / n}$. For the lower bound, suppose for contradiction that $\gamma(G) \leq N(r + \alpha\sqrt{2}) - 1$. By Chernoff bounds together with a union bound over all cells, a.a.s. there is at least one node in each such cell. Now place $N(r + \alpha\sqrt{2}) - 1$ many balls with centers at the nodes of the dominating set. Since by using radius r each cell is at least touched by some ball, by using radius $r + \alpha\sqrt{2}$ each cell is totally covered by a ball, and therefore $\mathcal{S}$ is covered by $N(r + \alpha\sqrt{2}) - 1$ balls of radius $r + \alpha\sqrt{2}$, contradicting the definition of $N(x)$. Therefore, a.a.s. $\gamma(G) \geq N(r + \alpha\sqrt{2})$.

For the upper bound, consider an optimal arrangement of $N(r - \beta)$ balls of radius $r - \beta$, where $\beta = \omega/\sqrt{n}$. As before, if the centre p of a ball is outside $\mathcal{S}$, but $\mathcal{B}(p, r - \beta) \cap \mathcal{S} \neq \emptyset$, we may shift the centre of the ball towards its closest point p' on the boundary of $\mathcal{S}$. Since $\mathcal{B}(p, r - \beta) \cap \mathcal{S} \subseteq \mathcal{B}(p', r - \beta) \cap \mathcal{S}$, we still preserve the covering property, and therefore, we can obtain an optimal covering of $\mathcal{S}$ with balls of radius $r - \beta$ and centered at points inside of $\mathcal{S}$. As in part (a), it suffices to show the existence of a node $v \in V$ inside $\mathcal{B}(c, \beta) \cap \mathcal{S}$ for any centre c in this optimal arrangement of balls. Since $N(r - \beta) = O(1)$, the probability that there exists a centre c such that $(\mathcal{B}(c, \beta) \cap \mathcal{S}) \cap V = \emptyset$ is at most

$$ O(1)(1 - \beta^2 \pi / 4)^n = O(\exp(-n\beta^2 \pi / 4)) = o(1), $$

and hence, a.a.s. for all centers c, we have that $\mathcal{B}(c, \beta) \cap \mathcal{S}$ contains at least one node of G. These nodes form a dominating set, and so a.a.s. $\gamma(G) \le N(r - \beta)$.

Finally, note that the lower bound in part (b) can be easily adopted to show that a.a.s. $\gamma(G) = \Omega(r^{-2})$ as a.a.s. a positive fraction of cells contain at least one node for the range of r considered in part (c). As already mentioned, the upper bound of $O(r^{-2})$ holds for any (deterministic) geometric graph and any r. Hence, part (c) follows. For part (d), the upper bound is trivial. The lower bound comes from the fact that a.a.s. there will be $\Theta(n)$ isolated nodes, and a dominating set has to contain all of them. The proof of the theorem is finished. $\square$

5. Conclusions and open problems

We considered the domination number of a stochastic model for OSNs, the MGEO-P model. Theorem 1 shows a sublinear bound on the domination number of OSNs, which is well correlated with estimates for the domination number taken for the Facebook 100 data set. In addition, we provided bounds for the domination number of random geometric graphs.

In future work, we would like to broaden our analysis of the domination number to other data sets, and to test larger samples of OSNs. We will contrast the estimates provided by other heuristic algorithms for computing minimum order dominating sets, and provide a fitting of the data to bounds provided by the model.

So-called "elites", those who exert strong influence on the ambient network, are studied extensively in the sociology literature (see [10] for an overview of the literature on this topic). One approach to detecting elites is via their relatively high degree; hence, the use of k-cores in [10]. A different approach to detecting elites is to search for them within a minimum order dominating set, as these sets reach the entire network. Further, if minimum order dominating sets have much smaller order than the network (as we postulate), then that reduces the computational costs of finding elites. We plan on considering this approach to finding elites via dominating sets in future work.

References

1. N. Alon, J. Spencer, *The Probabilistic Method*, Wiley, New York, 2000.
2. A.L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* **286** (1999) 509–512.
3. A. Bonato, *A Course on the Web Graph*, American Mathematical Society Graduate Studies Series in Mathematics, Providence, Rhode Island, 2008.
4. A. Bonato, D.F. Gleich, M. Kim, D. Mitsche, P. Prałat, A. Tian, S.J. Young, Dimensionality matching of social networks using motifs and eigenvalues, *PLOS ONE* **9** (2014) e106052.
5. A. Bonato, J. Janssen, P. Prałat, Geometric protean graphs, *Internet Mathematics* **8** (2012) 2–28.
6. A. Bonato, J. Janssen, P. Prałat, The geometric protean model for on-line social networks, In: *Proceedings of Workshop on Algorithms and Models for the Web Graph (WAW'10)*, 2010.
7. A. Bonato, J. Janssen, P. Prałat, A geometric model for on-line social networks, In: *Proceedings of 3rd Workshop on Online Social Networks (WOSN'10)*, 2010.
8. A. Bonato, A. Tian, Complex networks and social networks, invited book chapter in: *Social Networks*, editor E. Kranakis, Springer, Mathematics in Industry series, pp. 269-285, 2013.
9. C. Cooper, R. Klasing, M. Zito, Lower bounds and algorithms for dominating sets in web graphs, *Internet Mathematics* **2** (2005) 275-300.
10. B. Corominas-Murtra, B. Fuchs, S. Thurner, Detection of the elite structure in a virtual multiplex social system by means of a generalized k-core, Preprint 2014.
11. N.J. Cowan, E.J. Chastain, D.A. Vilhena, J.S. Freudenberg, C.T. Bergstrom, Nodal dynamics, not degree distributions, determine the structural controllability of complex networks, *PLOS ONE* **7** (2012) e38398.
12. E. Estrada, Spectral scaling and good expansion properties in complex networks, *Europhysics Letters* **73** (2006) 649.

13. M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology, *SIGCOMM Comput Commun Rev* (1999) **29** 251–262.
14. E.N. Gilbert, Random plane networks, *J. Soc. Industrial Applied Mathematics* **9** (1961) 533–543.
15. T.W. Haynes, S.T. Hedetniemi, P.J. Slater, *Fundamentals of Domination in Graphs*, CRC Press, 1998.
16. J. Janssen, P. Prałat, Protean graphs with a variety of ranking schemes, *Theoretical Computer Science* **410** (2009) 5491–5504.
17. S. Janson, T. Łuczak and A. Rucinski, *Random Graphs*, Wiley-Intersci. Ser. Discrete Math. Optim., 2000.
18. R. Kershner, The number of circles covering a set, *American Journal of Mathematics* **61** (1939) 665–671.
19. M. Kim, J. Leskovec, Multiplicative attribute graph model of real-world networks, *Internet Mathematics* **8** (2012) 113–160.
20. T.G. Kolda, A. Pinar, T. Plantenga, C. Seshadhri, A scalable generative graph model with community structure, Preprint 2014.
21. R. Kumar, P. Raghavan, S. Rajagopalan, S. Sivakumar, A. Tomkins, Stochastic models for the web graph, In: *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 2000.
22. J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani, Kronecker graphs: An approach to modeling networks, *Journal of Machine Learning Research* **11** (2010) 985–1042.
23. J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, *ACM Trans Knowl Discov Data* **1** (2007) 1–41.
24. J. Leskovec, K.J. Lang, A. Dasgupta, M.W. Mahoney, Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, *Internet Mathematics* **6** (2009) 29–123.
25. T. Luczak, P. Prałat, Protean graphs, *Internet Mathematics* **3** (2006) 21–40.
26. J.M. McPherson, J.R. Ranger-Moore, Evolution on a dancing landscape: Organizations and networks in dynamic blau space, *Social Forces* **70** (1991) 19–42.
27. M. McPherson, L. Smith-Lovin, J.M. Birds of a feather: Homophily in social networks, *Annual Review of Sociology* **27** (2001) 415–444.
28. T. Milenković, V Memišević, A. Bonato, N. Pržulj, Dominating biological networks, *PLOS ONE* **6** (2013) (8), e23016.
29. F. Molnár Jr., N. Derzsy, É. Czabarka, L. Székely, B.K. Szymanski, G. Korniss, Dominating scale-free networks using generalized probabilistic methods, Preprint 2014.
30. P. Prałat, A note on the diameter of protean graphs, *Discrete Mathematics* **308** (2008) 3399–3406.
31. P. Prałat, N. Wormald, Growing protean graphs, *Internet Mathematics* **4** (2009) 1–16.
32. M. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability. Oxford U.P., 2003.
33. S.B. Seidman, Network structure and minimum degree, *Social Networks* **5** (1983) 269–287.
34. I. Stojmenovic, M. Seddigh, J. Zunic, Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks, *IEEE Transactions on Parallel and Distributed Systems* **13** (2002) 14-25.
35. A.L. Traud, P.J. Mucha, M.A. Porter, Social structure of Facebook networks, Preprint 2014.
36. D.J. Watts, S.H. Strogatz, Collective dynamics of "small-world" networks, *Nature* **393** (1998) 440-442.
37. D.B. West, *Introduction to Graph Theory, 2nd edition*, Prentice Hall, 2001.
38. Z. Zhang, F. Comellas, G. Fertin, L. Rong, High-dimensional apollonian networks, Journal of Physics A: Mathematical and General (2006) **39** 1811.