New Bounds for Approximating Extremal Distances in Undirected Graphs

Massimo Cairo, Roberto Grossi, Romeo Rizzi

New Bounds for Approximating Extremal Distances in Undirected Graphs

Massimo Cairo* Roberto Grossi† Romeo Rizzi‡

Abstract

We provide new bounds for the approximation of extremal distances (the diameter, the radius, and the eccentricities of all nodes) of an undirected graph with n nodes and m edges. First, we show under the Strong Exponential Time Hypothesis (SETH) of Impagliazzo, Paturi and Zane [JCSS01] that it is impossible to get a $(3/2-\varepsilon)$-approximation of the diameter or a $(5/3-\varepsilon)$-approximation of all the eccentricities in $O(m^{2-\delta})$ time for any $\varepsilon, \delta > 0$, even allowing for a constant additive term in the approximation. Second, we present an algorithmic scheme that gives a $(2-1/2^k)$-approximation of the diameter and the radius and a $(3-4/(2^k+1))$-approximation of all eccentricities in $\widetilde{O}(mn^{\frac{1}{k+1}})$ expected time for any $k \geq 0$. For $k \geq 2$, this gives a family of previously unknown bounds, and approaches near-linear running time as k grows. Third, we observe a connection between the approximation of the diameter and the h-dominating sets, which are subsets of nodes at distance $\leq h$ from every other node. We give bounds for the size of these sets, related with the diameter.

1 Introduction

The diameter, the radius and the eccentricities of nodes are well-known extremal distances in graphs [7]. In an undirected graph $G = (V, E)$, letting $d(\cdot, \cdot)$ denote the distance among the nodes of G, the eccentricities are $\epsilon(v) = \max_{u \in V} d(v, u)$ for all nodes $v \in V$, the diameter is $D = \max_{v \in V} \epsilon(v) = \max_{u,v \in V} d(v, u)$, and the radius is $r = \min_{v \in V} \epsilon(v)$. (We postpone the discussion of directed graphs.) Their efficient computation is a basic problem in graphs [1, 3, 5, 8, 11, 12, 16, 22, 24, 28, 30].

Let $n = |V|$ be the number of nodes and $m = |E|$ be the number of edges. With $n-1$ graph searches (i.e. BFS traversals for unweighted graphs or Dijkstra searches for weighted graphs) of cost $\widetilde{O}(m)$ time each, where $\widetilde{O}(\cdot)$ notation neglects poly-log factors, the above distances can be computed as a variant of the all-pairs shortest paths problem [14, 32]. Many faster solutions have been proposed [4, 9, 10, 15, 17, 21, 23, 25, 26, 27, 30, 31, 33], but no $O(m^{2-\varepsilon})$-time algorithm is known for sparse graphs. On the other hand, a single graph search provides approximations with bounded error in $\widetilde{O}(m)$ time. Approximate solutions lying between these two extremes are interesting as they bring to light useful combinatorial properties [3].

We consider algorithms for α-multiplicative β-additive approximations $\widetilde{D}$, $\widetilde{r}$ and $\widetilde{\epsilon}_v$ that take $\widetilde{O}(m^\gamma)$ expected time.[1] Specifically, it is required that $\frac{1}{\alpha}D - \beta \leq \widetilde{D} \leq D$ for the diameter, $r \leq \widetilde{r} \leq \alpha r + \beta$ for the radius, and $\frac{1}{\alpha}\epsilon(v) - \beta \leq \widetilde{\epsilon}_v \leq \epsilon(v)$ for the eccentricities. Under these requirements we investigate the interplay among the parameters α, β and γ (the lower, the better), giving upper and lower bounds for the approximation quality using α, β and for the running time using γ. To motivate our work, we give an overview of the previous results in terms of α, β and γ, illustrating them for the diameter (Fig. 1a) and the eccentricities (Fig. 1c). Several other results exist but they address different aspects from what depicted here.

The upper bounds are reported as bullets in the upper envelopes of Figs. 1a,c. As discussed above, all the known exact algorithms achieve $\alpha = 1$, $\beta = 0$ and $\gamma = 2$ (leftmost bullet). Also, a single graph search yields $\alpha = 2$ for the diameter and the radius and $\alpha = 3$ for the eccentricities, all with $\beta = 0$ and $\gamma = 1$ (rightmost bullet). Roditty and Vassilevska W. [24] obtain $\alpha = 3/2$, $\beta < 1$ and $\gamma = 3/2$ for D and r (middle bullet in Fig. 1a): they employ Las Vegas randomization to reduce the $\widetilde{O}(m\sqrt{n} + n^2)$ time achieved in the seminal work by Aingworth, Chekuri, Indyk and Motwani [3] to expected $\widetilde{O}(m\sqrt{n})$ time. Chechik, Larkin, Roditty, Schoenebeck, Tarjan and Vassilevska W. [11] present new deterministic algorithms, addressing also eccentricities, and obtain $\alpha = 3/2$ for D, r and $\alpha = 5/3$ for $\epsilon(v)$, all with $\beta = 0$ and $\gamma = 3/2$ (middle bullet in Fig. 1c).

Some conditional lower bounds are known, represented as shaded zones in Fig. 1, under the Strong Exponential Time Hypothesis (SETH) of Impagliazzo, Paturi and Zane [20] stating that for every $\varepsilon > 0$ there is an integer k such that k-SAT cannot be solved in time $O(2^{(1-\varepsilon)n})$: any algorithm with α, β and γ below these bounds implies that SETH is false. Roditty and Vassilevska W. [24] give a reduction from SAT to

*Scuola Normale Superiore
†Università di Pisa
‡Università di Verona

[1] To give a uniform treatment in terms of γ, we will say that a cost of $\widetilde{O}(m^a n^b)$ has $\gamma = a + b$.

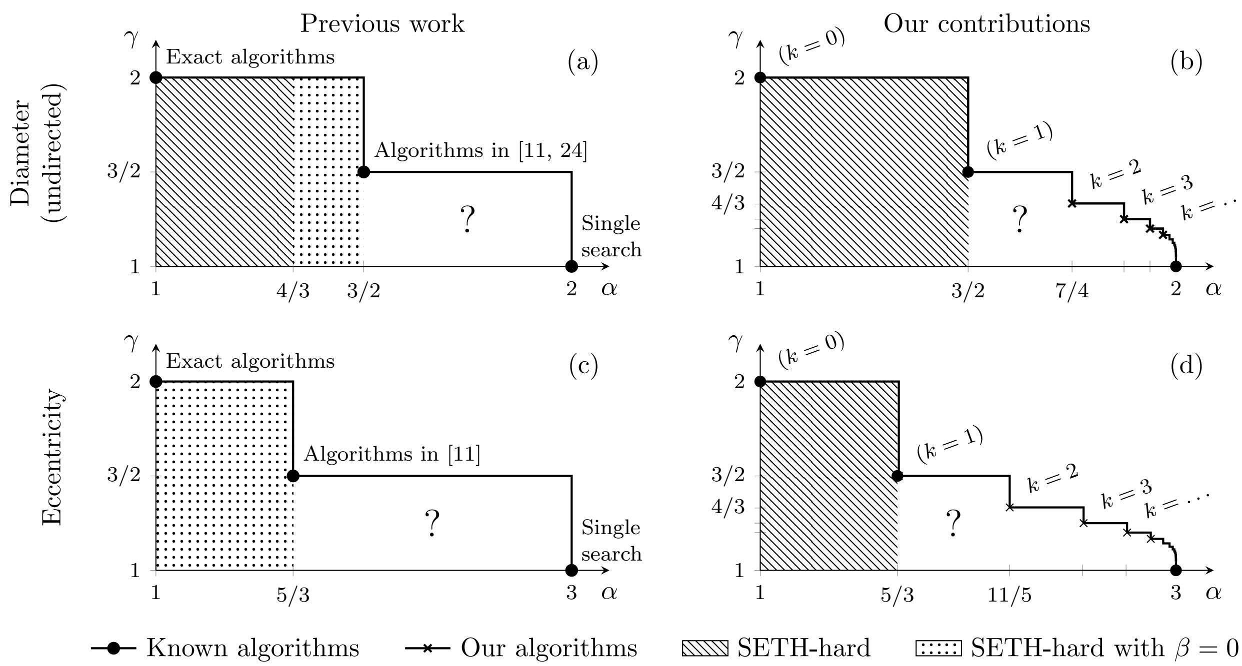

Figure 1: α-multiplicative $O(1)$-additive approximations in $\widetilde{O}(m^{\gamma})$ expected time.

the problem of distinguishing between diameter 2 and 3 in graphs implying that, under SETH, no algorithm can solve it in $O(m^{2-\delta})$ time: in particular, it is impossible to get $\alpha < 3/2$, $\beta = 0$ and $\gamma < 2$. With their construction, the possibility of $\alpha < 3/2$ with $\gamma < 2$ remains open if one allows for $\beta \geq 1$. Chechik et al. [11] reduce SAT to the problem of distinguishing between graphs of diameter $3(\ell+1)$ and $4(\ell+1)$, where $\ell \geq 0$ is a given parameter, showing that it is impossible to get $\alpha < 4/3$, $\beta = O(m^{\delta})$, $\gamma < 2-2\delta$ for any $\delta \geq 0$ under SETH. However, as noted by the authors, this construction still leaves open the possibility of $4/3 \leq \alpha < 3/2$ with $\gamma < 2$ and $\beta \geq 1$ (lighter shaded zone in Fig. 1a). Independently of this paper, Abboud, Vassilevska W. and Wang [2] give a conditional lower bound for the eccentricities. They exclude $\alpha < 5/3$ with $\beta = 0$ and $\gamma < 2$ (shaded zone in Fig. 1c) under the Orthogonal Vectors conjecture (OV), which is implied by SETH as shown by Williams [29].

In this paper, we make three further steps in the study of the problem of approximating the extremal distances mentioned before. In particular, we obtain new bounds illustrated in Figs. 1b,d.

First, we describe a reduction from SAT to two problems on graphs: deciding whether the diameter is $3t$ or $2t$, and deciding whether the maximum eccentricity over a given subset of nodes $X \subseteq V$ is $5t$ or $3t$. Under SETH, none of them can be solved when $t > c \cdot m^{\delta}$ in $O(m^{2-2\delta-\zeta})$ time (for any $c > 0$, $0 \leq \delta < 1$ and $\zeta > 0$): hence it is impossible to get $\alpha < 3/2$ for the diameter and $\alpha < 5/3$ for the eccentricities with $\beta = O(m^{\delta})$ and $\gamma < 2 - 2\delta$ for any $\delta \geq 0$ under SETH. In this way we tighten the known bounds for the diameter and the eccentricities (shaded zones in Fig. 1b,d). In particular, the recently achieved approximation factors $\alpha = 3/2$ for D and $\alpha = 5/3$ for $\epsilon(v)$ cannot be improved in truly subquadratic time ($\gamma < 2$) even with an additive term $\beta = O(1)$ under SETH. This also indicates that the eccentricities are more difficult to approximate than the diameter. Notice that our lower bound is incomparable with that in [2], as the former holds with the more general condition $\beta = O(m^{\delta})$ under SETH, while the latter holds under the weaker condition OV.

Second, we present an algorithmic scheme that gives nontrivial approximations with exponent $\gamma = 1 + \varepsilon$ arbitrarily close to the minimum 1. Specifically, for any integer $k \geq 0$, we present a randomized algorithm hinging on a novel iterative procedure that selects $O(n^{1/(k+1)})$ nodes from which to launch the graph searches (for all pairs shortest paths, a different iterative sampling has been presented in [14]). The properties of the selected nodes allow us to get $\alpha = 2 - 1/2^k$ for diameter and radius and $\alpha = 3 - 4/(2^k + 1)$ for eccentricities, with $\beta < 1$, in $\widetilde{O}(mn^{\frac{1}{k+1}})$ expected time (thus $\gamma = 1 + \frac{1}{k+1}$). For $k \geq 2$, this gives a family of previously unknown bounds (crosses in

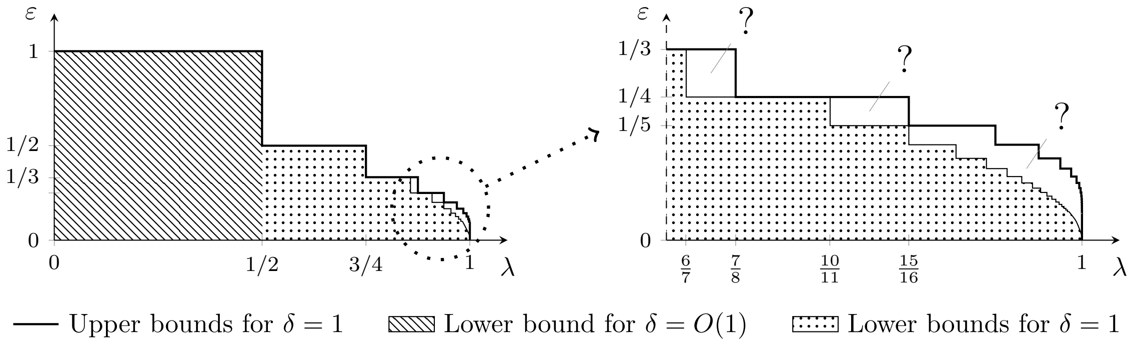

Figure 2: h-dominating sets of size $\widetilde{O}(n^{\varepsilon})$ for $h = \lambda(D+\delta)$.

Figs. 1b,d), and approaches near-linear time as k grows. For example, when $k = 2$ we obtain $\alpha = 7/4$ for diameter and radius and $\alpha = 11/5$ for the eccentricities, in $\widetilde{O}(mn^{1/3})$ expected time. Looking at Fig. 1, we can observe that previous work focused on bounds for α in $[1, 3/2]$ (diameter and radius) and $[1, 5/3]$ (eccentricities); instead, our results explore for the first time the ranges $(3/2, 2]$ and $(5/3, 3]$ (as we could not find any counterargument). Also, approaching $\gamma \approx 1$ for larger values of α is crucial to analyze massive networks.

Third, we study the size of distance h-dominating sets in undirected graphs of diameter D. An h-dominating set X is a subset of the nodes at distance $\leq h$ from every other node in the graph. Our study is closely related to the possibility of fast diameter approximations, but could be of independent interest for other graph problems (e.g. facility location). We first show that an h-dominating set X can be used to produce an upper bound $\overline{D}$ of the diameter D in $\widetilde{O}(|X| \cdot m)$ time, such that $D \leq \overline{D} \leq D + h$. Upper bounds are crucial for approximations, as any approximation algorithm with bounded error must (implicitly or explicitly) provide one: in particular an α-multiplicative β-additive approximation $\widetilde{D}$ is equivalent to an upper bound $\overline{D}$ satisfying $D \leq \overline{D} \leq \alpha(D+\beta)$, by choosing $\overline{D} = \alpha(\widetilde{D}+\beta)$. A closer look at the algorithms in [3, 24], when applied to undirected graphs, shows that they implicitly rely on the existence of an h-dominating set of size $\widetilde{O}(\sqrt{n})$ for $h = (D+1)/2$. In our algorithmic scheme we explicitly find h-dominating sets of size $\widetilde{O}(n^{1/(k+1)})$ for $h = \frac{2^k-1}{2^k}(D+1)$. To obtain faster approximation algorithms, one way is to find h-dominating sets X of smaller size: thus, it is interesting to study the size of X in the worst case, throwing a bridge between diameter approximation and extremal graph theory [6, 18]. We analyze the upper bounds given by our algorithms and show several lower bounds through explicit constructions.

In particular, we consider the worst-case size of h-dominating sets for $h = \lambda(D+\delta)$, for some $0 \leq \lambda < 1$ and $\delta \geq 0$, because this is related to diameter approximations with $\alpha = \lambda + 1$ and $\beta = \lambda\delta/\alpha$. If there is an h-dominating set X of size $\widetilde{\Theta}(n^{\varepsilon})$, then it is possible to use it to get $\gamma = 1 + \varepsilon$. Our results are illustrated in Fig. 2. First, we rule out the possibility to get $\varepsilon < 1$ for $\lambda < 1/2$ and $\delta = O(1)$. Second, we use a family of constructions G_t^{ℓ} to exclude $\varepsilon < 1/\ell$ for $\lambda < 1 - \frac{2}{\ell(\ell+1)+2}$ and $\delta \leq 1$, for any chosen $\ell \geq 2$. For some values of λ the bounds obtained by our algorithms are optimal (up to logarithmic factors). For these values, a faster algorithm for diameter approximation with $\alpha = \lambda + 1$ would need new techniques to bound the diameter from above as it cannot rely on the existence of small h-dominating sets.

Finally, some interesting questions concern directed graphs. Previous work on diameter approximation does not distinguish between directed and undirected graphs, sharing the same algorithmic techniques. Here our results are tailored for undirected graphs, except for our lower bound on the diameter, which holds also for directed graphs. One of the reasons lies in a central lemma (Lemma 4.2) that does not hold for directed graphs as the inferred distances in the proof are not symmetrical (i.e. it can be $d(x, y) \neq d(y, x)$). We do not see this as a limitation of our results. Actually we leave open the possibility that the problem of approximating the diameter in undirected vs directed graphs could require different techniques and values of β, γ when $3/2 < \alpha < 2$ (Fig. 1b), while this situation does not seem to emerge for $\alpha = 3/2$. It could be also interesting to investigate generalizations of h-dominating sets for directed graphs.

The paper is organized as follows. After some preliminaries in Section 2, we give the lower bound under SETH in Section 3. We present our approximation scheme in Section 4, and introduce our framework for distance dominating sets in Section 5.

2 Preliminaries

We consider undirected graphs $G = (V, E)$, with $n = |V|$ nodes and $m = |E|$ nodes (wlog $m \geq n$). Each edge $(u, v) \in E$ is associated with a real positive weight $w(u, v) > 0$, where conventionally $w(u, v) = 1$ for unweighted graphs. The maximum edge weight is denoted by $M = \max_{(u,v)\in E} w(u, v)$, where $M = 1$ on unweighted graphs. We denote distances between any two nodes by $d(u, v)$ and use the shorthand $d(S, v) = \min_{u\in S} d(u, v)$. The eccentricity of a node $v \in V$ is $\epsilon(v) = \max_{u\in V} d(u, v)$. The diameter of G is $D = D(G) = \max_{v\in V} \epsilon(v) = \max_{u,v\in V} d(u, v)$ and the radius is $r = r(G) = \min_{v\in V} \epsilon(v)$. We use the definitions of α-multiplicative β-additive approximations $\widetilde{D}$, $\widetilde{r}$ and $\widetilde{\epsilon}_v$ given in Section 1, and simply say (almost) α-approximation when $\beta = O(M)$.

To work uniformly on weighted and unweighted graphs, we use the term *graph search* to indicate either a BFS traversal for unweighted graphs or a Dijkstra search for weighted graphs. We denote by $C = C(n, m)$ the cost of this search: for example, $C = O(m)$ for BFS and $C = O(m + n \log n)$ for Dijkstra implemented with Fibonacci heaps. We always assume $C = \widetilde{O}(m)$ and $C = \Omega(m)$. Given $S \subseteq V$, the value of $d(S, v)$ for all the nodes $v \in V$ can be computed in time $O(C)$ by starting a graph search from a dummy node connected to every node in S: we call it *multi-source* graph search.

We define $N_\ell(v)$ as the first ℓ nodes (v inclusive) discovered during a graph search launched from node v. In other words, $N_\ell(v)$ contains the nearest ℓ nodes to v, breaking ties arbitrarily. We use the graphical notation $u \to v$ to denote an edge (u, v) traversed in a path, and $u \rightsquigarrow v$ to denote any shortest path from u to v (possibly empty if $u = v$).

Lemma 2.1. *For any pair of nodes $u, v \in V$, we have $|\epsilon(u) - \epsilon(v)| \leq d(u, v)$.*

Proof. Wlog $\epsilon(u) \geq \epsilon(v)$. Take $x \in V$ so that $\epsilon(u) = d(u, x) \leq d(u, v) + d(v, x) \leq d(u, v) + \epsilon(v)$. □

We need the following known result on uniform random sampling, also used in [3, 24].

Lemma 2.2. *Given a family $\mathcal{H}$ of at most n sets each of size ℓ over a universe U of size L, a random sampling of $\Theta(L/\ell \cdot \log n)$ elements hits all the sets in $\mathcal{H}$ with high probability.*

Proof. For any $H \in \mathcal{H}$ the probability that an element $u \in U$ sampled uniformly at random is not in H is $\frac{L-\ell}{L}$. If $S \subseteq U$ contains s elements sampled independently and uniformly at random, then $\mathbb{P}[S \cap H = \emptyset] = \mathbb{P}[u \notin H]^s = (\frac{L-\ell}{L})^s$. By the union bound, the probability p that $S \cap H = \emptyset$ for some $H \in \mathcal{H}$ is at most: $|\mathcal{H}| \cdot \mathbb{P}[S \cap H = 0] = n \cdot \left(\frac{L-\ell}{L}\right)^s = n \cdot \left[\left(1 - \frac{\ell}{L}\right)^{\frac{L}{\ell}}\right]^{s\cdot\ell/L} = O(n \cdot \mathrm{e}^{-s\cdot\ell/L})$. If $s = \alpha \cdot L/\ell \cdot \log n$ for some constant α, then $p = O(n \cdot \mathrm{e}^{-\alpha\cdot\log n}) = 1/n^{\Omega(\alpha)}$. □

3 Hardness of improved approximations

We transform an instance φ of k-SAT into an undirected, unweighted graph G^φ_t (of exponential size) of diameter $2t$ or $3t$, where the latter occurs iff φ is satisfiable. Our transformation is based on the construction of Roditty and Vassilevska W. [24] which produces graphs of diameter 2 and 3. Chechik et al. [11] showed a related construction that gives diameter $3t$ or $4t$ respectively, for any $t \geq 1$. Still, it is challenging to obtain $3t$ and $2t$. We need the gadget graph $T^t[B]$ described below.

For a given node set B and $t \geq 1$, the gadget $T^t[B]$ is a graph $T = (V, E)$ on node set $V = B \cup Q$, where Q is a set of additional *private* nodes. The purpose of this graph is to make the nodes in B at distance exactly t from each other. For $t = 1$, $T^1[B]$ is the complete graph on B, with no private node. For $t = 2$, we introduce a private node $c \in Q$ and define $T^2[B]$ as the star with center in c and tips in B. To construct $T^{t+2}[B]$, we first introduce a distinct private node $u' \in Q$ for every $u \in B$. Then, we build the gadget $T^t[B'] = (B' \cup Q', E')$ on the node set $B' = \{u' : u \in B\}$ and add an edge (u', u) for every $u \in B$. Specifically, we define $Q = B' \cup Q'$ and $E = E' \cup \{(u', u) : u \in B\}$. The result is a star-like structure, illustrated in Fig. 3 where private nodes are colored black.

Proposition 3.1. *$T^t[B]$ contains $O(t \cdot |B|)$ nodes and $O(t \cdot |B| + |B|^2)$ edges. For any $a, b \in V$ we have $d(a, b) \leq t$ with equality iff $a, b \in B$ and $a \neq b$. In $T^{2z}[B]$, for any $a \in V$ we have $d(c, a) \leq z$ with equality iff $u \in B$. The only shortest path between $u, v \in B$ contains c.*

Proof. The statements are trivially true for $T^1[B]$ and $T^2[B]$. In the general case, they can be shown by induction noticing that any shortest path in $T^{t+2}[B]$ is a shortest path in $T^t[B']$ possibly concatenated with an edge (u, u') at each endpoint u, if $u \in B$. □

3.1 Construction of G^φ_t

Let φ be an instance of k-SAT on an even number d of variables $D = \{x_1, \ldots, x_d\}$ and c clauses $C = \{\gamma_1, \ldots, \gamma_c\}$. We first preprocess φ to remove duplicated clauses, so we assume the γ_i's are distinct. Then, we divide the variables in two sets D_1 and D_2, each of size $d/2$, and construct the sets P_1 and P_2, containing the $2^{d/2}$ partial assignments on the variables in D_1 and D_2,

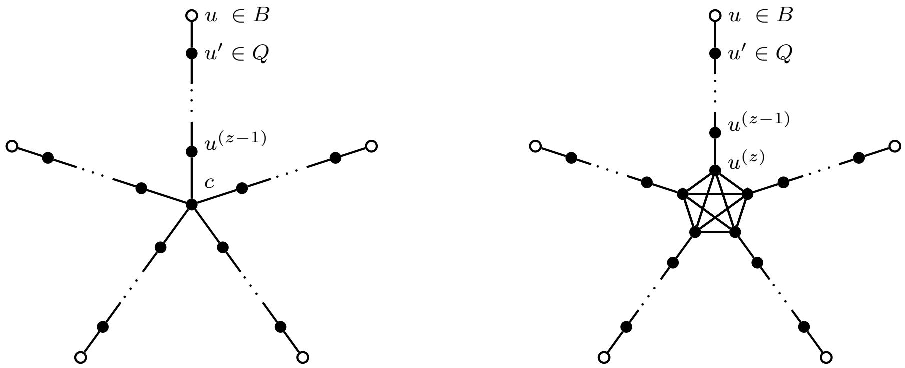

Figure 3: Gadget $T^t[B]$ for $|B| = 5$ and $t = 2z$ (left) or $t = 2z + 1$ (right).

respectively. We say that a partial assignment $p \in P_i$ *satisfies* a clause $\gamma \in C$ if γ contains at least one literal evaluating to *true* under p. We extend the set of clauses C to $C^* = C \cup \{\delta_1, \delta_2\}$, where δ_i is a dummy clause satisfied by the partial assignments $p \in P_i$ only.[2] Then, we build the set $P^* = P_1 \cup P_2 \cup \{\pi\}$, where π is an empty partial assignment that does not satisfy any clause.

The graph $G_t^\varphi = (V, E)$ is defined as follows. We start with a node set U containing a node u_γ for every clause $\gamma \in C^*$. Then, for each partial assignment $p \in P^*$ we introduce a new node w_p and define the gadget $T_p = (V_p, E_p) = T^t[\{w_p\} \cup U_p]$, where $U_p = \{u_\gamma \in U : \gamma \text{ is } \textit{not} \text{ satisfied by } p\}$. The graph G_t^φ is the union of the gadgets T_p: specifically, $V = \bigcup_{p \in P^*} V_p$ and $E = \bigcup_{p \in P^*} E_p$, with $U \subseteq V$ as $V_\pi \supset U_\pi = U$ and $V_p \cap V_q \subseteq U$ for $p \neq q$ as private nodes of different gadgets are distinct.

Proposition 3.2. *G_t^φ contains $t \cdot 2^{d/2+o(d)}$ edges and can be constructed in $t \cdot 2^{d/2+o(d)}$ time.*

Proof. Observe that $|P^*| = 2 \cdot 2^{d/2} + 1 = 2^{d/2+o(d)}$ and $|C^*| = O((2d)^k) = 2^{O(k \cdot \log d)} = 2^{o(d)}$ as we do not have duplicated clauses and k is constant. Computing U_p for every $p \in P^*$ requires $O(|P^*| \cdot |C^*| \cdot k) = 2^{d/2+o(d)}$ time. Each of the $|P^*|$ gadgets has $O(t \cdot |C^*| + |C^*|^2) = t \cdot 2^{o(d)}$ edges by Proposition 3.1 and can be constructed trivially knowing U_p. Thus, the total time and size is at most $t \cdot 2^{d/2+o(d)}$. □

Proposition 3.3. *Any path in G_t^φ that does not contain a node in U between its endpoints is also a path in T_p for some $p \in P^*$ and in particular both its endpoints belong to V_p.*

Proof. Otherwise, take two consecutive edges $x \to v \to y$ along the path which belong to distinct gadgets T_p and T_q. They must be incident to a node $v \in V_p \cap V_q \subseteq U$. □

Proposition 3.4. *For any two distinct non-private nodes $a, b \in U \cup \{w_p : p \in P^*\}$, we have $d(a, b) \geq t$.*

Proof. Take a shortest path $a \rightsquigarrow b$ and apply Proposition 3.3. If it is a path on T_p, then $d(a, b) = t$ by Proposition 3.1. Otherwise, it is of the form $a \rightsquigarrow v \rightsquigarrow b$ with $v \in U$ distinct from a and b, hence $d(a, b) = d(a, v) + d(v, b) \geq 2t$ by induction. □

Lemma 3.1. *G_t^φ has diameter either $2t$ or $3t$. It is $3t$ iff φ is satisfiable.*

Proof. We first show that φ is satisfiable iff there exist two disjoint sets U_p and U_q for some $p, q \in P^*$. Observe that the sets U_p and U_q are disjoint iff every clause in $C^* = C \cup \{\delta_1, \delta_2\}$ is satisfied by either p or q. The dummy clauses δ_1 and δ_2 are both satisfied iff p and q belong respectively to P_1 and P_2, and thus form a valid total assignment $p \cup q$. Moreover, $p \cup q$ satisfies φ iff each clauses in C is satisfied by either p or q (as the clauses are disjunctive). As any total assignment satisfying φ can be written as $p \cup q$ for $p \in P_1 \subseteq P^*$ and $q \in P_2 \subseteq P^*$, the claim is proven.

The following four facts hold.

(i) We have $d(w_p, w_q) \geq 2t$ for any $p \neq q$.

Take any shortest path from w_p to w_q. By Proposition 3.3 it contains a node $v \in U$, since w_p and w_q belong to distinct gadgets. Thus $d(w_p, w_q) = d(w_p, v) + d(v, w_q) \geq 2t$ by Proposition 3.4.

(ii) If $d(a, b) > 2t$ for some nodes a and b, then some sets U_p and U_q are disjoint.

Take $p, q \in P^*$ such that $a \in V_p$ and $b \in V_q$. If U_p and U_q are not disjoint there is a node

[2]For example $\delta_i = x \vee \overline{x}$, where x is any variable in the group D_i.

$z \in U_p \cap U_q \subseteq V_p \cap V_q$. As $a, z \in V_p$ and $z, b \in V_q$ we get $d(a,b) \leq d(a,z) + d(z,b) \leq 2t$ by Proposition 3.1.

(iii) If U_p and U_q are disjoint, then $d(w_p, w_q) \geq 3t$.

Notice that $p \neq q$ as U_p and U_q are non-empty (they contain u_{δ_1} or u_{δ_2}). Any shortest path from w_p to w_q contains a node $v \in U$ by point (i) above. If there is no other $v' \in U$ on the path, then $w_p, v \in V_p$ and $v, w_q \in V_q$ by Proposition 3.3, and in particular $v \in V_p \cap V_q \cap U \subseteq U_p \cap U_q$ contradicting our assumption. So, the path is of the form $w_p \rightsquigarrow v \rightsquigarrow v' \rightsquigarrow w_q$ with distinct $v, v' \in U$. Hence $d(w_p, w_q) = d(w_p, v) + d(v, v') + d(v', w_q) \geq 3t$ by Proposition 3.4.

(iv) For any two nodes a and b, we have $d(a,b) \leq 3t$.

Take $p, q \in P^*$ such that $a \in V_p$ and $b \in V_q$. Consider a path $a \rightsquigarrow u_{\delta(p)} \rightsquigarrow u_{\delta(q)} \rightsquigarrow b$ where $\delta(s)$ is the clause δ_i *not* satisfied by s. Since $a, u_{\delta(p)} \in V_p$, $u_{\delta(p)}, u_{\delta(q)} \in V_\pi$ and $u_{\delta(q)}, b \in V_q$, we have $d(a,b) \leq 3t$ by Proposition 3.1.

The four facts above imply that the diameter is either $2t$ or $3t$, and the latter occurs iff φ is satisfiable. $\square$

For even $t = 2z$, all the gadgets T_p appearing in G^φ_{2z} have a center, say $c_p \in V_p$.

Lemma 3.2. *In G^φ_{2z} the maximum eccentricity $\overline{\epsilon} = \max_{p \in P^*} \epsilon(c_p)$ among the centers c_p of all the gadgets T_p is either $3z$ or $5z$. It is $5z$ iff φ is satisfiable.*

Proof. By Lemma 3.1 it is sufficient to prove $\overline{\epsilon} = D - z$.

We first prove $\epsilon(c_p) \leq D - z$ for any $p \in P^*$. For nodes $v \in V_p$, we have $d(c_p, v) \leq z \leq D - z$ by Proposition 3.1. If $v \notin V_p$, any shortest path from w_p to v has to pass through c_p. Indeed, consider the first node $u \in U$ on the shortest path, which must exist by Proposition 3.3: the sub-path $w_p \rightsquigarrow u$ contains c_p by Proposition 3.1. Write the path as $w_p \rightsquigarrow c_p \rightsquigarrow v$: we have $d(c_p, v) = d(w_p, v) - d(w_p, c_p) \leq D - z$ as $d(w_p, v) \leq D$ and $d(w_p, c_p) = z$ by Proposition 3.1.

Now we show $\epsilon(c_p) \geq D - z$ for some $p \in P^*$. Take a diametral node v so that $\epsilon(v) = D$ and $p \in P^*$ such that $v \in V_p$: we have $d(c_p, v) \leq z$ by Proposition 3.1 and $\epsilon(c_p) \geq \epsilon(v) - d(c_p, v) \geq D - z$ by Lemma 2.1. $\square$

Theorem 3.1. *Under SETH, there is no algorithm distinguishing between diameter-$3t$ and diameter-$2t$ graphs when $t > c \cdot m^\delta$ in $O(m^{2-2\delta-\zeta})$ time, for any $c > 0$, $0 \leq \delta < 1$ and $\zeta > 0$.*

Proof. Suppose by contradiction to have a procedure A distinguishing between diameter-$3t$ and diameter-$2t$ graphs for $t > c \cdot m^\delta$ in $O(m^{2-2\delta-\zeta}) = O(m^{(1-\delta)(2-\varepsilon)})$ time for some $\zeta, \varepsilon > 0$, $c > 0$ and $0 \leq \delta < 1$. Fix $\gamma = \varepsilon/4$: under SETH there is a k such that k-SAT cannot be solved in $O(2^{(1-\gamma)d})$ time.

Consider an instance φ of k-SAT in d-variables. By Proposition 3.2, for some $\mu = 2^{d/2+o(d)}$ the graph G^φ_t contains $m \leq \mu t$ edges and can be constructed in $O(\mu t)$ time. As $\delta < 1$, we can pick $t = \Theta(\mu^{\frac{\delta}{1-\delta}})$ such that $t > c \cdot \mu^\delta t^\delta \geq c \cdot m^\delta$ and $m \leq \mu t = O(\mu^{1+\frac{\delta}{1-\delta}}) = O(\mu^{\frac{1}{1-\delta}})$. Notice that for any input graph with $m = \mu t$ edges, the procedure A runs in $O(m^{(1-\delta)(2-\varepsilon)}) = O(\mu^{2-\varepsilon})$ time but takes $\Omega(\mu t)$ time to read the graph, hence we can assume $\mu^{2-\varepsilon} = \Omega(\mu t)$. To solve φ, we construct the graph G^φ_t, which by Lemma 3.1 has diameter $2t$ or $3t$ according to the satisfiability of φ, then we apply the procedure A to distinguish the two cases (as we picked $t > c \cdot m^\delta$). The procedure A runs in $O(m^{(1-\delta)(2-\varepsilon)}) = O(\mu^{2-\varepsilon})$ time where $\mu^{2-\varepsilon} = \Omega(\mu t)$ dominates the $O(\mu t)$ cost of constructing the graph. Thus, the total time is $O(\mu^{2-\varepsilon}) = 2^{(2-\varepsilon)\cdot d/2+o(d)} = 2^{(1-\frac{\varepsilon}{2})d+o(d)} = 2^{(1-2\gamma)d+o(d)} = O(2^{(1-\gamma)d})$ contradicting SETH. $\square$

Corollary 3.1. *Under SETH, there is no algorithm giving a $(\frac{3}{2} - \varepsilon)$-multiplicative $O(m^\delta)$-additive approximation of the diameter in $O(m^{2-2\delta-\zeta})$ time, for any $\varepsilon, \zeta > 0$ and $0 \leq \delta < 1$. In particular, there is no almost $(\frac{3}{2} - \varepsilon)$-approximation in $O(m^{2-\zeta})$ time.*

Proof. Consider a procedure that gives a $(\frac{3}{2} - \varepsilon)$-multiplicative $c \cdot m^\delta$-additive approximation $\widetilde{D}$ of the diameter D. Take a small enough[3] $\gamma > 0$ such that $\widetilde{D}$ satisfies $\widetilde{D} \geq \frac{2}{3}D + \gamma D - c \cdot m^\delta$. For $c' = c/(3\gamma)$ and $t > c' \cdot m^\delta = c \cdot m^\delta/(3\gamma)$, our procedure distinguishes diameter-$2t$ and diameter-$3t$ graphs: in the first case $\widetilde{D} \leq D = 2t$ while in the second case $\widetilde{D} \geq 2t + \gamma \cdot 3t - c \cdot m^\delta > 2t$. Hence, by Theorem 3.1 it cannot run in $O(m^{2-2\delta-\zeta})$ time for any $\zeta > 0$. $\square$

Theorem 3.2. *Under SETH, there is no algorithm deciding whether the maximum eccentricity over a given subset of nodes X is $5z$ or $3z$ when $z > c \cdot m^\delta$ in $O(m^{2-2\delta-\zeta})$ time, for any $c > 0$, $0 \leq \delta < 1$ and $\zeta > 0$.*

Proof. We adapt the proof of Theorem 3.1. Notice that we used our hypothetical procedure distinguishing diameter-$3t$ and diameter-$2t$ graphs only on instances of the form G^φ_t. On such instances having $t = 2z$, this is equivalent to telling whether $\overline{\epsilon} = \max_{p \in P^*}\{\epsilon(c_p)\}$ is $5z$ or $3z$ by Lemma 3.2. Thus, any procedure performing the latter in $O(m^{2-2\delta-\zeta})$ time yields to the same contradiction. $\square$

[3] For approximation factor $\frac{3}{2} - \varepsilon$, choose $\gamma = \frac{4\varepsilon}{3(3-2\varepsilon)}$.

Corollary 3.2. *Under SETH, there is no algorithm giving a $(\frac{5}{3}-\varepsilon)$-multiplicative $O(m^\delta)$-additive approximation of all the eccentricities in a graph in $O(m^{2-2\delta-\zeta})$ time, for any $\varepsilon > 0$, $0 \le \delta < 1$ and $\zeta > 0$. In particular, there is no almost $(\frac{5}{3}-\varepsilon)$-approximation in $O(m^{2-\zeta})$ time.*

Proof. Following the proof of Corollary 3.1, we take a small enough[4] $\gamma > 0$ such that the eccentricity estimation $\widetilde{\epsilon}_v$ satisfies $\widetilde{\epsilon}_v \ge \frac{3}{5}\epsilon(v) + \gamma\epsilon(v) - c \cdot m^\delta$. For $c' = c/(5\gamma)$ and $z > c' \cdot m^\delta = c \cdot m^\delta/5\gamma$, our procedure can decide whether the maximum eccentricity over a given subset of nodes X is $3z$ or $5z$. In the first case $\widetilde{\epsilon}_v \le \epsilon(v) \le 3z$ for any $v \in X$. In the second case, for some $v \in X$ we have $\epsilon(v) = 5z$ and $\widetilde{\epsilon}_v \ge 3z + \gamma \cdot 5z - c \cdot m^\delta > 3z$. By Theorem 3.2, any procedure that performs this cannot run in $O(m^{2-2\delta-\zeta})$ time for any $\zeta > 0$. □

Remark 1. *Under SETH, Theorem 3.1 excludes the possibility of a truly sub-quadratic algorithm that takes a graph of constant diameter $D = 3t$ and produces an approximation $\widetilde{D} > \frac{2}{3}D = 2t$. This result is surprisingly tight: when D is constant but* not *divisible by 3, an $O(m^{2-\delta})$-time algorithm that produces a value $\widetilde{D} > \frac{2}{3}D$ is possible, as shown by Roditty and Vassilevska W. [24].*

4 Fast approximation algorithms

We begin with a motivating example: we are given in input an undirected unweighted graph of diameter 8 and we want to output a pair of nodes at distance 5 or more from each other. By performing a BFS from an arbitrary node in the graph, we are guaranteed to find another node at distance at least 4 from it, but this is not sufficient. On the other hand, by running a 3/2-approximation algorithm for the diameter, we are guaranteed to find two nodes at distance 6 from each other in $\widetilde{O}(m\sqrt{n})$ time. We show that it is possible to obtain distance 5 in $\widetilde{O}(mn^{1/3})$ time, with high probability.

(1) Suppose that every node has $\ge n^{2/3}$ nodes at distance ≤ 3. (Call "ℓ-neighborhood" of a node the set of nodes at distance $\le \ell$ from it.) In this case, a random sampling of $\widetilde{\Theta}(n^{1/3})$ nodes hits the 3-neighborhood of every node with high probability by Lemma 2.2: in particular one of the sampled nodes is at distance ≤ 3 from a diametral node and by Lemma 2.1 it has eccentricity $\ge 8-3 = 5$. We run a BFS from each of the sampled nodes (they are only $\widetilde{O}(n^{1/3})$) so we are guaranteed to find a pair of nodes at distance ≥ 5 from each other.

(2) Suppose that there is a node z of degree $< n^{1/3}$. We launch a BFS from z, hence we can assume that $\epsilon(z) \le 4$ as otherwise the BFS finds two nodes at distance ≥ 5 from each other. We also run a BFS from each of the neighbors of z (they are less than $n^{1/3}$), which together are at distance ≤ 3 from the rest of the graph, since $\epsilon(z) \le 4$. In particular, there is a neighbor of z at distance ≤ 3 from a diametral node, and by Lemma 2.1 it has eccentricity $\ge 8-3 = 5$. As we run a BFS from all the neighbors of z, we are guaranteed to find two nodes at distance ≥ 5 from each other.

(3) We are left with the hard case: there is a node w which has few ($< n^{2/3}$) nodes at distance ≤ 3 *and* every node has many (degree $\ge n^{1/3}$) neighbors at distance 1. Even if the node w is not known, it can be found easily by sampling $\widetilde{\Theta}(n^{1/3})$ nodes at random and then picking the node farthest from the sampled set, as noted by Roditty and Vassilewska W. [24]. We run a BFS from w so we can assume $\epsilon(w) \le 4$. Then, we exploit the fact that w has few nodes at distance ≤ 3 and that every node has many neighbors at distance 1 in order to find a small set at distance ≤ 3 from all the nodes in the graph. Once obtained this set, we can run a BFS from each of its elements and find a node of eccentricity $\ge 8-3 = 5$. In order to obtain this set we proceed as follows. Consider the nodes in the 2-neighborhood of w and observe that their respective neighbors are all within distance 3 from w. Thus, the 1-neighborhoods of every node in the 2-neighborhood of w form a family of sets, each of size $\ge n^{1/3}$, over a universe of size $< n^{2/3}$ (the 3-neighborhood of w). By sampling $\widetilde{O}(n^{1/3})$ nodes uniformly at random from the 3-neighborhood of w, we hit all the sets in the family with high probability by Lemma 2.2. That is, with high probability the set of sampled nodes is at distance ≤ 1 from every node in the 2-neighborhood of w. Notice that the 2-neighborhood of w is in turn at distance ≤ 2 from the rest of the graph, as $\epsilon(w) \le 4$. Hence, with high probability the sampled nodes are at distance $\le 2+1 = 3$ from every node the graph.

The ideas employed in cases 1 and 2 have been already used in the literature. The approach behind some previous diameter approximation algorithms is to look for a node w such that the set $W = N_\ell(w)$ for some small $\ell > 0$ contains nodes at large distance h from w (this roughly corresponds to case 2). In general, this set may not exist in a graph for a given distance h: this happens if every node v has more than ℓ nodes at distance h. Still, in this case there is a useful

[4] For approximation factor $\frac{5}{3}-\varepsilon$, choose $\gamma = \frac{9\varepsilon}{5(5-3\varepsilon)}$.

hitting set S of size $\widetilde{O}(n/\ell)$ at distance h from every node v (this roughly corresponds to case 1). In the algorithms by Aingworth et al. [3] and by Roditty and Vassilevska W. [24] these ideas are applied to give sets W and S of size $\widetilde{O}(\ell) = \widetilde{O}(n/\ell) = \widetilde{O}(\sqrt{n})$. Our algorithm introduces a novel machinery (based on the idea given in case 3) that proceeds *iteratively* to obtain smaller and smaller sets $W_i = N_{\ell_i}(w_i)$.

4.1 Sampling procedure

Our approximation algorithms begin with a sampling procedure that outputs a sequence of nodes $w_1, \ldots, w_k \in V$ and node sets $S_0, \ldots, S_k \subseteq V$, which have the property to be sufficiently *close* to all the nodes in the graph (in a way that will be formalized in Lemma 4.3).

We perform k iterations, numbered from 0 to $k-1$, where k is a constant parameter. In the first iteration $i = 0$, we proceed similarly to case 1. We build S_0 by sampling $\Theta(q \cdot \log n)$ nodes uniformly at random, where $q = \widetilde{\Theta}(n^{\frac{1}{k+1}})$. By Lemma 2.2, we have that S_0 hits with high probability the sets $N_{\ell_1}(u)$ for every node u, where $\ell_1 = \lceil n/q \rceil$. We also pick the node w_1 farthest from S_0. Letting $h_0 = d(S_0, w_1)$ and $W_1 = N_{\ell_1}(w_1)$, we have with high probability that W_1 contains a node in S_0 (which is at distance $\geq h_0$ from w_1), so W_1 contains all the nodes at distance $< h_0$ from w_1.

Since W_1 is not small enough (it has size $\ell_1 = O(n/q)$), we do not use it directly. Instead we pass it along to the next iteration $i = 1$. At each iteration $i = 1, \ldots, k-1$, we are given the set $W_i = N_{\ell_i}(w_i)$ containing the nodes at distance $< h_{i-1}$ from w_i, and we look for a *smaller* set $W_{i+1} = N_{\ell_{i+1}}(w_{i+1})$ containing the nodes at distance $< h_i$ from w_{i+1}. To achieve this, we proceed as in case 3. We build a set S_i by sampling $\Theta(q \cdot \log n)$ nodes uniformly at random from the universe W_i of size ℓ_i. By Lemma 2.2, we have that S_i hits with high probability the sets $N_{\ell_{i+1}}(u)$, where $\ell_{i+1} = \lceil \ell_i/q \rceil$, but this time *only* for those nodes u such that $N_{\ell_{i+1}}(u) \subseteq W_i$.

We pick the node w_{i+1} farthest from $Z_i = (V \setminus W_i) \cup S_i$, so that w_{i+1} is far from S_i but its close neighbors do not fall outside W_i. More precisely, letting $h_i = d(Z_i, w_{i+1})$, we will prove that $W_{i+1} = N_{\ell_{i+1}}(w_{i+1})$ contains all the nodes at distance $< h_i$ from w_{i+1} (see fact c in the proof of Lemma 4.2), under the condition $W_{i+1} \cap Z_i \neq \emptyset$. This condition holds with high probability (see Lemma 4.1), yielding a Monte Carlo algorithm. To obtain a Las Vegas one, we check the condition at the end of iteration i: if it does not hold, we repeat the iteration i resampling S_i. The crucial part of our analysis is to generalize the idea given in case 3, to show that if h_{i-1} is large but h_i is small, then S_i is close to every node in the graph (see Lemma 4.2).

Finally, after iteration $k-1$, we are left with a set $W_k = N_{\ell_k}(w_k)$ of size $\ell_k = O(n/q^k)$. Similarly to case 2, we keep the whole set W_k for the approximation.

For uniformity, we define $W_0 = V$, $\ell_0 = n$, and $Z_0 = (V \setminus W_0) \cup S_0 = S_0$ so we do not need to treat the first iteration in a special way. Similarly, we define $S_k = W_k$ and $Z_k = (V \setminus W_k) \cup S_k = V$. By choosing $q = (n/\log n)^{\frac{1}{k+1}}$, we minimize $|S_0 \cup \cdots \cup S_k|$ to $\widetilde{O}(n^{\frac{1}{k+1}})$, as shown in Proposition 4.2.

The whole procedure is summarized in Algorithm 1.

Lemma 4.1. *In Algorithm 1, the condition $W_{i+1} \cap Z_i \neq \emptyset$ in step 2d holds with high probability.*

Proof. Consider the family of sets $\mathcal{H} = \{N_{\ell_{i+1}}(u) : u \in V \text{ and } N_{\ell_{i+1}}(u) \subseteq W_i\}$ on the universe W_i. We apply Lemma 2.2 to show that S_i hits every set in $\mathcal{H}$ with high probability. Indeed, we have $|\mathcal{H}| \leq n$, $|W_i| = \ell_i$, each set in the family $N_{\ell_{i+1}}(u)$ has size ℓ_{i+1}, and S_i contains $\Theta(q \cdot \log n) = \Theta(\ell_i/\ell_{i+1} \cdot \log n)$ nodes sampled uniformly at random from W_i.

If $W_{i+1} \not\subseteq W_i$ then $W_{i+1} \cap Z_i \supseteq W_{i+1} \cap (V \setminus W_i) = W_{i+1} \setminus W_i \neq \emptyset$. Otherwise, $W_{i+1} = N_{\ell_{i+1}}(w_{i+1}) \in \mathcal{H}$. Hence, with high probability S_i hits W_{i+1} and $W_{i+1} \cap Z_i \supseteq W_{i+1} \cap S_i \neq \emptyset$. □

Proposition 4.1. *Algorithm 1 runs in $O(C)$ expected time, where C is the cost of a graph search.*

Proof. Steps 1 and 3 take at most linear time. In each iteration in step 2, the time spent in substeps 2a and 2d is dominated by the $O(C)$ cost of substeps 2b and 2c. As the condition in substep 2d holds with high probability by Lemma 4.1, each iteration is repeated a constant number of times in expectation. As we perform k iterations for constant k, the total cost is $O(C)$ expected time. □

Proposition 4.2. *Algorithm 1 returns sets $S_0, \ldots, S_k$ with $|S_0 \cup \cdots \cup S_k| = O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n)$.*

Proof. Each set $S_0, \ldots, S_{k-1}$ is of size $\Theta(q \cdot \log n) = O((n/\log n)^{\frac{1}{k+1}} \cdot \log n) = O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n)$ by construction. Notice that $\ell_i = O(n/q^i)$, thus $|S_k| = |W_k| = \ell_k = O(n/q^k) = O(n/(n/\log n)^{\frac{k}{k+1}}) = O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n)$. □

For any $0 \leq i \leq k$, let h_i be the maximum distance from Z_i to any node in the graph, where $h_k = 0$ by construction as $Z_k = V$. The set W_{i+1} contains all the nodes at distance $< h_i$ from w_{i+1}. (See fact c in the next proof.) We relate the distances $d(S_i, v)$ with the values h_i. For $i = 0$, we have $Z_0 = (V \setminus W_0) \cup S_0 = S_0$ as $W_0 = V$, thus $d(S_0, v) \leq h_0$. For $i \geq 1$, we obtain the following.

Algorithm 1 Sampling procedure.

Input: Undirected graph $G = (V, E)$ and a *constant* integer parameter $k \geq 0$.

Output: Nodes $w_1, \ldots, w_k \in V$ and node sets $S_0, \ldots, S_k \subseteq V$.

1. Let $W_0 = V$, $\ell_0 = n$ and $q = (n/\log n)^{\frac{1}{k+1}}$.
2. For each $i = 0, 1, \ldots, k-1$:
   (a) Sample $\Theta(q \cdot \log n)$ nodes uniformly at random from W_i.
   Let S_i be the set of sampled nodes.
   (b) Run a multi-source graph search from $Z_i = (V \setminus W_i) \cup S_i$. Let w_{i+1} be the last visited node.
   (c) Perform a graph search from w_{i+1}.
   Let W_{i+1} be the set containing the first $\ell_{i+1} = \lceil \ell_i/q \rceil$ visited nodes.
   (d) Check that $W_{i+1} \cap Z_i \neq \emptyset$. If the check fails, repeat iteration i.
3. Set $S_k = W_k$. *(Also, define $Z_k = (V \setminus W_k) \cup S_k = V$ for uniformity.)*

LEMMA 4.2. *For $1 \leq i \leq k$ and any node $v \in V$, we have either $d(S_i, v) \leq h_i$ or $d(S_i, v) \leq d(w_i, v) - h_{i-1} + 2h_i + M$.*

Proof. We will use the following facts.

(a) For any node $u \in V$ either $d(S_i, u) \leq h_i$ or $d(V \setminus W_i, u) \leq h_i$, as h_i is the maximum distance from $Z_i = (V \setminus W_i) \cup S_i$ to any node.

(b) If $u \in W_i$ then $d(u, w_i) \leq d(V \setminus W_i, w_i)$, as W_i comprises the ℓ_i nearest nodes to w_i.

(c) We have $d(V \setminus W_i, w_i) \geq h_{i-1}$. Take a node $z \in W_i \cap Z_{i-1}$ (as it exists due to the condition in step 2d). By (b), $d(V \setminus W_i, w_i) \geq d(z, w_i) \geq d(Z_{i-1}, w_i) = h_{i-1}$ as w_i is the farthest node from Z_{i-1}.

Fix a node $v \in V$. Consider a shortest path P from w_i to v. Let a be the last node on P such that $d(S_i, a) \leq h_i$. The node a exists since $d(S_i, w_i) \leq h_i$. Indeed, taking any node $u \in S_i \subseteq W_i$, we have $d(S_i, w_i) \leq d(u, w_i) \leq d(V \setminus W_i, w_i)$ by (b): this implies that $d(S_i, w_i) \leq h_i$ by (a).

If $a = v$, then $d(S_i, v) \leq h_i$ and the statement is proven. Otherwise, take the node b which follows a on P, so the path P is of the form $w_i \rightsquigarrow a \to b \rightsquigarrow v$ with $d(S_i, b) > h_i$. By (a), $d(V \setminus W_i, b) \leq h_i$. As $d(V \setminus W_i, w_i) \geq h_{i-1}$ by (c), then $d(w_i, b) \geq h_{i-1} - h_i$ by triangle inequality.[5] Writing P as $w_i \rightsquigarrow b \rightsquigarrow v$ we get $d(b, v) = d(w_i, v) - d(w_i, b) \leq d(w_i, v) - (h_{i-1} - h_i)$. We can bound $d(S_i, v)$ as $d(S_i, v) \leq d(S_i, a) + w(a, b) + d(b, v) \leq h_i + M + d(w_i, v) - (h_{i-1} - h_i) = d(w_i, v) - h_{i-1} + 2h_i + M$. $\square$

[5] Here and in (b)–(c) we are implicitly using the hypothesis that the graph is undirected.

LEMMA 4.3. *Let $v \in V$ be any node. For some $0 \leq \bar{\imath} \leq k$, we have $d(S_{\bar{\imath}}, v) \leq h_{\bar{\imath}} \leq (2^{k-\bar{\imath}} - 1)(\Delta + M)$, where $\Delta = \max_{1 \leq i \leq k} \{d(w_i, v) - d(S_i, v)\}$.*

Proof. As h_0 is the maximum distance from $Z_0 = S_0$, we have $d(S_0, v) \leq h_0$. Let $\bar{\imath} \in \{0, \ldots, k\}$ be the maximum index $\bar{\imath}$ such that $d(S_{\bar{\imath}}, v) \leq h_{\bar{\imath}}$.

We show $h_i \leq (2^{k-i} - 1)(\Delta + M)$ for $\bar{\imath} \leq i \leq k$, proving the lemma. Start with $i = k$: we have $(2^{k-i} - 1)(\Delta + M) = 0$ and $h_k = 0$ by construction. Now suppose inductively $h_i \leq (2^{k-i} - 1)(\Delta + M)$ for some $\bar{\imath} < i \leq k$. We need to show $h_{i-1} \leq (2^{k-(i-1)} - 1)(\Delta + M)$. By the assumption $i > \bar{\imath}$, it cannot be $d(S_i, v) \leq h_i$ (as $\bar{\imath}$ is the maximum index satisfying this condition), so it must be that $d(S_i, v) \leq d(w_i, v) - h_{i-1} + 2h_i + M$ by Lemma 4.2. Therefore $h_{i-1} \leq 2h_i + d(w_i, v) - d(S_i, v) + M \leq 2h_i + \Delta + M \leq 2 \cdot (2^{k-i} - 1)(\Delta + M) + \Delta + M = (2^{k-(i-1)} - 1)(\Delta + M)$. $\square$

4.2 Approximations

We obtain the approximations by exploiting our sampling procedure as shown in Algorithm 2. Notice that we obtain $\epsilon(w_i)$ during the sampling as we run a graph search from w_i.

PROPOSITION 4.3. *Algorithm 2 takes $O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n \cdot C) = \widetilde{O}(mn^{\frac{1}{k+1}})$ expected time.*

Proof. Step 1 runs in $O(k \cdot C)$ expected time by Proposition 4.1. Step 2 performs $|S_0 \cup \cdots \cup S_k|$ graph searches

Algorithm 2 Approximation of the diameter, the radius and all the eccentricities.

Input: Undirected graph $G = (V, E)$ and an integer $k \geq 0$.

Output: Approximations $\widetilde{D}$ (diameter), $\widetilde{r}$ (radius) and $\widetilde{\epsilon}_v$ (eccentricity of every node $v \in V$).

1. Execute Algorithm 1 and obtain $w_1, \dots, w_k \in V$ and $S_0, \dots, S_k \subseteq V$.
2. Run a graph search from every node $x \in S_0, \dots, S_k$.
3. Return:

$$\begin{aligned}\widetilde{D} &= \max\{\max_{1\leq i\leq k} \epsilon(w_i), \max_{x\in S_0\cup\cdots\cup S_k} \epsilon(x)\}\\ \widetilde{r} &= \min_{x\in S_0\cup\cdots\cup S_k} \epsilon(x)\\ \widetilde{\epsilon}_v &= \max\{\max_{1\leq i\leq k} d(w_i, v), \max_{x\in S_0} d(x, v), \max_{x\in S_1\cup\cdots\cup S_k} \epsilon(x) - d(x, v)\}\end{aligned}$$

and by Proposition 4.2 runs in $O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n \cdot C)$ time. Finally, step 3 does not increase the asymptotical time complexity as the values $\widetilde{D}$, $\widetilde{r}$ and $\widetilde{\epsilon}_v$ can be be computed during the previous steps at no extra cost. □

Lemma 4.4. *Algorithm 2 returns $\widetilde{D}$ such that $\frac{2^k}{2^{k+1}-1} D - \frac{2^k-1}{2^{k+1}-1} M \leq \widetilde{D} \leq D$.*

Proof. Note that $\widetilde{D} = \epsilon(u) \leq D$ for some $u \in V$. Take v such that $\epsilon(v) = D$. For any i, we have $d(w_i, v) \leq \epsilon(w_i) \leq \widetilde{D}$ and $d(S_i, v) \geq D - \widetilde{D}$. Take $x \in S_i$ such that $d(S_i, v) = d(x, v)$: we have $\widetilde{D} \geq \epsilon(x) \geq \epsilon(v) - d(x, v) \geq D - d(S_i, v)$ by Lemma 2.1.

Apply Lemma 4.3 to v, where by the facts shown above $\Delta \leq 2\widetilde{D} - D$ and $d(S_{\bar{\imath}}, v) \geq D - \widetilde{D}$. We get $D - \widetilde{D} \leq (2^{k-\bar{\imath}} - 1)(\Delta + M) \leq (2^k - 1)(2\widetilde{D} - D + M)$, hence $\widetilde{D} \geq \frac{2^k}{2^{k+1}-1} D + \frac{2^k-1}{2^{k+1}-1} M$. □

Lemma 4.5. *Algorithm 2 returns $\widetilde{r}$ such that $r \leq \widetilde{r} \leq \frac{2^{k+1}-1}{2^k} r + \frac{2^k-1}{2^k} M$.*

Proof. By definition $\widetilde{r} = \min_{x\in S_0\cup\cdots\cup S_k} \epsilon(x) \geq r$. Take v such that $\epsilon(v) = r$. For every i, we have $d(w_i, v) \leq \epsilon(v) = r$ and $d(S_i, v) \geq \widetilde{r} - r$. Indeed, take $x \in S_i$ such that $d(S_i, v) = d(x, v)$: we have $\widetilde{r} \leq \epsilon(x) \leq \epsilon(v) + d(x, v) = r + d(S_i, v)$.

Apply Lemma 4.3 to v, where $\Delta \leq 2r - \widetilde{r}$ and $d(S_{\bar{\imath}}, v) \geq \widetilde{r} - r$. We get $\widetilde{r} - r \leq (2^{k-\bar{\imath}} - 1)(\Delta + M) \leq (2^k - 1)(2r - \widetilde{r} + M)$ hence $\widetilde{r} \leq \frac{2^{k+1}-1}{2^k} r + \frac{2^k-1}{2^k} M$. □

Remark 2. *A single graph search gives a 3-approximation of all the eccentricities in the graph. Suppose we start the search from $x \in V$: it is sufficient to pick $\widetilde{\epsilon}_v = \max\{d(x, v), \epsilon(x) - d(x, v)\}$. First, $\epsilon(v) \geq d(x, v)$ and $\epsilon(v) \geq \epsilon(x) - d(x, v)$ hence $\widetilde{\epsilon}_v \leq \epsilon(v)$. Second, $\widetilde{\epsilon}_v \geq \epsilon(x) - d(x, v) \geq \epsilon(v) - 2d(x, v) \geq \epsilon(v) - 2\widetilde{\epsilon}_v$ hence $\widetilde{\epsilon}_v \geq \epsilon(v)/3$.*

Lemma 4.6. *Algorithm 2 returns $\widetilde{\epsilon}_v$ such that $\frac{2^k+1}{3\cdot 2^k-1}\epsilon(v) - \frac{2^k-1}{3\cdot 2^k-1} M \leq \widetilde{\epsilon}_v \leq \epsilon(v)$.*

Proof. Consider any node v and notice that the value $\widetilde{\epsilon}_v$ is obtained either as $d(u, v)$ or as $\epsilon(u) - d(u, v)$ for some node $u \in V$. As $\epsilon(u) \leq \epsilon(v) + d(u, v)$ by Lemma 2.1, in either case we get $\widetilde{\epsilon}_v \leq \epsilon(v)$.

For every i, we have $d(w_i, v) \leq \widetilde{\epsilon}_v$ and $d(S_i, v) \geq \frac{\epsilon(v) - \widetilde{\epsilon}_v}{2}$. Indeed, take $x \in S_i$ such that $d(S_i, v) = d(x, v)$: we have $\widetilde{\epsilon}_v \geq \epsilon(x) - d(x, v) \geq \epsilon(v) - 2d(x, v) = \epsilon(v) - 2d(S_i, v)$ by Lemma 2.1.

Apply Lemma 4.3 to the node v, where $\Delta \leq \widetilde{\epsilon}_v - \frac{\epsilon(v)-\widetilde{\epsilon}_v}{2} = \frac{3\widetilde{\epsilon}_v - \epsilon(v)}{2}$ and $d(S_{\bar{\imath}}, v) \geq \frac{\epsilon(v)-\widetilde{\epsilon}_v}{2}$. We get $\frac{\epsilon(v)-\widetilde{\epsilon}_v}{2} \leq h_{\bar{\imath}} \leq (2^{k-\bar{\imath}} - 1)(\Delta + M) \leq (2^k - 1)(\frac{3\widetilde{\epsilon}_v - \epsilon(v)}{2} + M)$. This inequality already gives us a lower bound on $\widetilde{\epsilon}_v$. However, we obtain a better bound treating the case $\bar{\imath} = 0$ separately.

When $\bar{\imath} = 0$, we have $h_{\bar{\imath}} = h_0 \leq (2^k - 1)(\Delta + M)$. Take a node u such that $\epsilon(v) = d(v, u)$. As h_0 is the maximum distance from $Z_0 = S_0$, we have $d(S_0, u) \leq h_0$. Now take $x \in S_0$ such that $d(x, u) = d(S_0, u) \leq h_0$: we have $\epsilon(v) = d(v, u) \leq d(v, x) + d(x, u) \leq \widetilde{\epsilon}_v + h_0$ and $\epsilon(v) - \widetilde{\epsilon}_v \leq h_0 \leq (2^k - 1)(\Delta + M)$.

When $\bar{\imath} \geq 1$, we obtain $\frac{\epsilon(v)-\widetilde{\epsilon}_v}{2} \leq (2^{k-\bar{\imath}} - 1)(\Delta + M) \leq (2^{k-1} - 1)(\Delta + M) \leq \frac{1}{2}(2^k - 1)(\Delta + M)$.

In either case, we have $\epsilon(v) - \widetilde{\epsilon}_v \leq (2^k - 1)(\Delta + M) \leq (2^k - 1)(\frac{3\widetilde{\epsilon}_v - \epsilon(v)}{2} + M)$ hence $\widetilde{\epsilon}_v \geq \frac{2^k+1}{3\cdot 2^k-1}\epsilon(v) - \frac{2^k-1}{3\cdot 2^k-1} M$. □

Lemma 4.4, 4.5 and 4.6 show that the values $\widetilde{D}$, $\widetilde{r}$ and $\widetilde{\epsilon}_v$ are respectively $(2 - \frac{1}{2^k})$-approximations of the diameter and the radius and $(3 - \frac{4}{2^k+1})$-approximations of the eccentricities. They are obtained in $O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n \cdot C)$ expected time by Proposition 4.3. We summarize these results in the following theorem.

THEOREM 4.1. *For any $k \geq 0$, Algorithm 2 gives an almost $(2 - \frac{1}{2^k})$-approximation of the diameter and the radius and an almost $(3 - \frac{4}{2^k+1})$-approximation of all the eccentricities of an undirected graph in $O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n \cdot C) = \widetilde{O}(mn^{\frac{1}{k+1}})$ expected time, where C is the cost of a graph search. The additive terms are $\frac{2^k-1}{2^{k+1}-1}M$ (diameter), $\frac{2^k-1}{2^k}M$ (radius) and $\frac{2^k-1}{3 \cdot 2^k-1}M$ (eccentricities).*

COROLLARY 4.1. *For arbitrarily small $\varepsilon > 0$, there is $\delta > 0$ such that in $O(mn^{\varepsilon})$ expected time it is possible to give an almost $(2-\delta)$-approximation of the diameter and the radius and an almost $(3 - \delta)$-approximation of all the eccentricities of an undirected graph with additive error $< M$.*

5 Distance dominating sets

We recall the definition of (distance) h-dominating sets [18, 19].

DEFINITION 1. *In a graph $G = (V, E)$, a subset of the nodes $X \subseteq V$ is a (distance) h-dominating set if $d(X, u) \leq h$ for every node $u \in V$.*

We start by showing that Algorithm 1 relies on h-dominating sets.

PROPOSITION 5.1. *The set $X = S_0 \cup \cdots \cup S_k$ generated by Algorithm 1 is a $\frac{2^k-1}{2^k}(D+M)$-dominating set of size $|X| = O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n)$.*

Proof. Let $v \in V$ be any node. By Lemma 4.3, we have $d(X, v) \leq d(S_{\bar{\imath}}, v) \leq (2^{k-\bar{\imath}} - 1)(\Delta + M)$ where $\Delta \leq D - d(X, v)$, as $d(w_i, v) \leq D$ and $d(S_i, v) \geq d(X, v)$. Thus, $d(X, v) \leq (2^k - 1)(D - d(X, v) + M)$, hence $d(X, v) \leq \frac{2^k-1}{2^k}(D + M)$. The bound $|X| = O(n^{\frac{1}{k+1}} \log^{\frac{k}{k+1}} n)$ is given by Proposition 4.2. □

This observation implicitly leads to the following purely combinatorial result.

THEOREM 5.1. *Every undirected graph of n nodes admits a $\frac{2^k-1}{2^k}(D + M)$-dominating set of size $\widetilde{O}(n^{\frac{1}{k+1}})$, for any constant $k \geq 0$.*

This is shown in Fig. 2, for $\lambda = \frac{2^k-1}{2^k}$ and $\varepsilon = \frac{1}{k+1}$.

We now show our general approach to obtain an estimation of the diameter from an h-dominating set. First, we produce an upper bound $\overline{D}$ as described next.

PROPOSITION 5.2. *Given a h-dominating set $X \subseteq V$, it is possible to produce an upper bound $\overline{D}$ of the diameter D such that $D \leq \overline{D} \leq D + h$ in $O(|X| \cdot C)$ time.*

Proof. Compute the value $h_0 = \max_{u \in V} d(X, u)$ with a multi-source graph search, taking $O(C)$ time. As X is a h-dominating set, we have $h_0 \leq h$.[6] Then, compute $D_0 = \max_{x \in X} \epsilon(x) \leq D$ in $O(|X| \cdot C)$ time running a graph search from each $x \in X$. Return the value $\overline{D} = D_0 + h_0 \leq D + h$. To prove that $D \leq \overline{D}$, take $v \in V$ and $x \in X$ such that $\epsilon(v) = D$ and $d(x, v) = d(X, v) \leq h_0$. We obtain $D = \epsilon(v) \leq \epsilon(x) + d(x, v) \leq D_0 + h_0 = \overline{D}$. □

If $h = \lambda(D + \delta)$ for some known constants λ and δ, then the upper bound $\overline{D}$ obtained with Proposition 5.2 can be transformed into a $(\lambda+1)$-multiplicative $\frac{\lambda\delta}{\lambda+1}$-additive approximation $\widetilde{D}$. It is sufficient to choose as estimation of the diameter the value $\widetilde{D} = \frac{\overline{D}-\lambda\delta}{\lambda+1}$. Indeed, we have $\widetilde{D} = \frac{\overline{D}-\lambda\delta}{\lambda+1} \leq \frac{D+h-\lambda\delta}{\lambda+1} = \frac{D+\lambda(D+\delta)-\lambda\delta}{\lambda+1} = D$ and $\widetilde{D} \geq \frac{D-\lambda\delta}{\lambda+1} = \frac{1}{\lambda+1}D - \frac{\lambda\delta}{\lambda+1}$ as required by the definition of α-multiplicative β-additive approximation. In particular, applying Proposition 5.2 to the $\frac{2^k-1}{2^k}(D + M)$-dominating set X produced by Algorithm 1 in $O(C)$ expected time (by Proposition 5.1), we obtain an alternative proof of our bounds for diameter approximation.

5.1 Lower bounds on h-dominating sets

We presented a general algorithmic approach to approximate the diameter by finding small-size h-dominating sets. To better understand the properties of this approach and its limitations, it is natural to provide lower bounds on the size of these sets in the worst case and in relation to the diameter D.

The specific case $h = 1$ and $D = 2$ has been already studied in the literature since a distance 1-dominating set is a classical dominating set. Desormeaux et al. [13] prove that in undirected graphs of diameter 2 the smallest 1-dominating set has size $\Theta(\sqrt{n \log n})$ in the worst case.[7] We provide several lower bounds, with focus on the coefficients λ and δ that relate h and D as $h = \lambda(D + \delta)$. Our bounds are illustrated as shaded zones in Fig. 2.

We first obtain a lower bound from the gadget graph $T^t[B]$ defined in Section 3.

THEOREM 5.2. *For any integer constant $D \geq 1$ and infinite values of n, there exists a family of undirected unweighted graphs of diameter D and number of nodes n where any h-dominating set for $h < D/2$ has size $\Theta(n)$.*

Proof. Consider the gadget graph $T^t[B]$ on a set B of size $s \to \infty$. The number of nodes in is $n = \Theta(s)$ and the diameter is $D = t$ by Proposition 3.1. Consider any

[6] In fact, here we discover that X is an h_0-dominating set, but in general it can be $h_0 < h$.

[7] Desormeaux et al. [13] consider *total* dominating sets. On graphs without isolated nodes, a dominating set of size t can be transformed into a total dominating set of size $2t$ [19], hence their $\Omega(\sqrt{n \log n})$ lower bound still holds.

h-dominating set X for $h < D/2$: it must contain at least one node per branch, in order to be at distance $\leq h$ from each tip $u \in B$. Hence, $|X| \geq s = \Theta(n)$. $\square$

COROLLARY 5.1. *For any constants $\lambda < 1/2$ and $\delta = O(1)$, and infinite values of n, there exists a family of undirected unweighted graphs of diameter D and number of nodes n where any h-dominating set for $h = \lambda(D+\delta)$ has size $\Theta(n)$.*

Proof. For $\lambda < 1/2$ and $\delta = O(1)$, we can pick a large enough D such that $h = \lambda(D + \delta) < D/2$. $\square$

We now provide a family of constructions for every integer $\ell \geq 2$. Each construction for a fixed value of ℓ produces a family of graphs G_t^ℓ parameterized by an integer $t \geq \ell$. The graph G_t^ℓ contains $n = t^\ell$ nodes and has diameter $D = \frac{\ell(\ell+1)}{2}$. Moreover, we prove that every h-dominating set for $h < D$ is of size $\Omega(t) = \Omega(n^{1/\ell})$.

5.2 Construction G_t^2

For simplicity, we first describe the construction in the special case $\ell = 2$. We then generalize the construction for other values of ℓ and rigorously prove our claims.

The node set of G_t^2 contains the pairs of natural numbers $xy \in \{1,\dots,t\}^2$. For any node $ab \in \{1,\dots,t\}^2$, there is a "write" edge $ab \xrightarrow{\mathsf{W}} xb$ for any $x \in \{1,\dots,t\}$, and a "swap" edge $ab \xrightarrow{\mathsf{S}} ba$.

Note that for any edge $xy \to x'y'$ there is an edge $x'y' \to xy$, hence the graph is undirected. The diameter is at most 3 as for any two nodes ab and xy there is a path
$$ab \xrightarrow{\mathsf{W}} ya \xrightarrow{\mathsf{S}} ay \xrightarrow{\mathsf{W}} xy.$$
Any path starting from a node $ab \in \{1,\dots,t\}^2$ matches the following pattern
$$\varepsilon\varepsilon \to \delta\varepsilon \to \delta\varepsilon|\varepsilon\delta \to \delta\delta \to \cdots$$
where $\varepsilon ::= a|b$ and $\delta ::= 1|\dots|t$. Hence, the nodes at distance 2 (or less) from any node ab match the pattern $\delta\varepsilon|\varepsilon\delta$, thus their number is $O(t)$. This implies that any 2-dominating set of G_t^2 needs $\Omega(t) = \Omega(\sqrt{n})$ nodes to reach all the $n = t^2$ nodes in the graph.

5.3 A family of constructions G_t^ℓ

We now fully describe our family of constructions (thus a family of families of graphs). For a given $\ell \geq 2$, we define the graphs G_t^ℓ parametrized by $t \geq \ell$. The node set of G_t^ℓ contains the sequences of ℓ coordinates over the natural numbers $x_1 \dots x_\ell \in \{1,\dots,t\}^\ell$. For a node $a_1 \cdots a_\ell$ and any $x \in \{1,\dots,t\}$, we have a "write" edge that writes the first coordinate:
$$a_1 a_2 \cdots a_\ell \xrightarrow{\mathsf{W}} x a_2 \cdots a_\ell.$$
Then, for $i \in \{1,\dots,\ell-1\}$, we have an edge which swaps the coordinates i and $i+1$, thus for a node $a_1 \cdots a_\ell$ we have
$$a_1 \cdots a_i a_{i+1} \cdots a_\ell \xrightarrow{\mathsf{S}_i} a_1 \cdots a_{i+1} a_i \cdots a_\ell.$$
The resulting graph is clearly undirected.

LEMMA 5.1. *The diameter of G_t^ℓ is at most $\frac{\ell(\ell+1)}{2}$.*

Proof. Let $a_1 \cdots a_\ell$ and $x_1 \cdots x_\ell$ be any two nodes. For every $j \in \{1,\dots,\ell\}$ there is a path
$$\begin{aligned} &x_{\ell-j+2} \cdots x_\ell a_j \cdots a_\ell \xrightarrow{\mathsf{S}_{j-1}} \xrightarrow{\mathsf{S}_{j-2}} \cdots \xrightarrow{\mathsf{S}_1} \\ &a_j x_{\ell-j+2} \cdots x_\ell a_{j+1} \cdots a_\ell \xrightarrow{\mathsf{W}} \\ &x_{\ell-j+1} \cdots x_\ell a_{j+1} \cdots a_\ell \end{aligned}$$
of length j. Concatenating these paths we get
$$\begin{aligned} &a_1 \cdots a_\ell \longrightarrow \\ &x_1 a_2 \dots a_\ell \longrightarrow \cdots \longrightarrow \\ &x_{\ell-j+2} \cdots x_\ell a_j \cdots a_\ell \longrightarrow \cdots \longrightarrow \\ &x_1 \cdots x_\ell, \end{aligned}$$
which is a path from $a_1 \cdots a_\ell$ to $x_1 \cdots x_\ell$ of length $\sum_{j=1}^{\ell} j = \frac{\ell(\ell+1)}{2}$. $\square$

LEMMA 5.2. *In the graph G_t^ℓ, the nodes at distance $\leq p$ from any given node are $O(t^q)$, where q is the largest integer such that $\frac{q(q+1)}{2} \leq p$.*

Proof. Let $\overline{a} = a_1 \dots a_\ell$ be a given node. Consider any node $\overline{x} = x_1 \dots x_\ell$ and define $b_i = 0$ if $x_i \in \{a_1,\dots,a_\ell\}$ and $b_i = 1$ otherwise. We define the *size* of $\overline{x}$ as $s(\overline{x}) = \sum_{i=1}^{\ell} b_i$. We give a weight i to the i-th coordinate, and define the *potential* of $\overline{x}$ as $p(\overline{x}) = \sum_{i=1}^{\ell} i \cdot b_i$. The following facts hold.

1. We have $p(\overline{x}) \leq d(\overline{a}, \overline{x})$. By induction, $p(\overline{a}) = 0$ and $\overline{x} \to \overline{y}$ implies $p(\overline{y}) \leq p(\overline{x}) + 1$. Indeed, if $\overline{x} \xrightarrow{\mathsf{W}} \overline{y}$, then only b_1 can change and it has weight 1. If $\overline{x} \xrightarrow{\mathsf{S}i} \overline{y}$ instead, then the values of b_i and b_{i+1} are swapped (and the others do not change) and their weights differ by one unit only.

2. We have $s(\overline{x}) \leq q$, where q is the largest integer such that $\frac{q(q+1)}{2} \leq p(\overline{x})$. Fixed a maximum potential p, the largest size q is obtained by the greedy choice $b_1,\dots,b_q = 1$ and $b_{q+1},\dots,b_\ell = 0$, where the potential is $\sum_{j=1}^{q} j = \frac{q(q+1)}{2} \leq p$.

Thus, the nodes at distance at most p from a given node $\overline{a}$ have size at most q, and their number is at most

$\sum_{s=0}^{q} \binom{\ell}{s} \cdot (t-\ell)^s \cdot \ell^{\ell-s} = O(t^q)$. Indeed, the value s varies over the possible values of $s(\overline{x})$; the term $\binom{l}{s}$ comes from the choice of which b_i are set to 1; the terms $(t-\ell)^s$ and $\ell^{\ell-s}$ come from the choice of the values for the coordinates x_i where $b_i = 1$ and $b_i = 0$, respectively. $\square$

THEOREM 5.3. *For every $\ell \geq 2$ and infinite values of n, there exists a family of undirected unweighted graphs of diameter $D = \frac{\ell(\ell+1)}{2}$ and number of nodes n where any h-dominating set for $h < D$ has size $\Omega(n^{1/\ell})$.*

Proof. Consider the graphs G_t^ℓ for $t \to \infty$. First, the number of nodes is $n = t^\ell$ by construction. Second, the diameter is at most $\frac{\ell(\ell+1)}{2}$ by Lemma 5.1. Third, by Lemma 5.2 there are at most $O(t^{\ell-1})$ nodes at distance *strictly less* than $\frac{\ell(\ell+1)}{2}$ from any given node. For large enough t, this number is less than the total number of nodes t^ℓ: hence there are some nodes in the graph at distance *at least* $\frac{\ell(\ell+1)}{2}$ from any given node, and in particular the diameter is exactly $D = \frac{\ell(\ell+1)}{2}$.

Since the nodes at distance $h < D$ from any given node are only $O(t^{\ell-1})$, we need at least $\Omega(t)$ different nodes to reach at distance h all the t^ℓ nodes in the graph. Hence, any h-dominating set for $h < D$ has size $\Omega(t) = \Omega(n^{1/\ell})$. $\square$

COROLLARY 5.2. *For $\ell \geq 2$, $\lambda < 1 - \frac{2}{\ell(\ell+1)+2}$, and infinite values of n, there exists a family of undirected unweighted graphs of some diameter D and number of nodes n where any h-dominating set for $h = \lambda(D+1)$ has size $\Omega(n^{1/\ell})$.*

Proof. Observe that $\lambda < 1 - \frac{2}{\ell(\ell+1)+2} = \frac{\ell(\ell+1)}{\ell(\ell+1)+2} = \frac{\frac{\ell(\ell+1)}{2}}{\frac{\ell(\ell+1)}{2}+1} = \frac{D}{D+1}$ and $h = \lambda D + \lambda < \frac{D}{D+1}D + \frac{D}{D+1} = D$. Hence, the family G_t^ℓ is such that every h-dominating set has size $\Omega(t) = \Omega(n^{1/\ell})$. $\square$

References

[1] Amir Abboud, Fabrizio Grandoni, and Virginia Vassilevska Williams. Subcubic equivalences between graph centrality problems, APSP and diameter. In Piotr Indyk, editor, *SODA*, pages 1681–1697. SIAM, 2015.

[2] Amir Abboud, Virginia Vassilevska Williams, and Joshua R. Wang. Approximation and fixed parameter subquadratic algorithms for radius and diameter. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016*. SIAM, 2016.

[3] D. Aingworth, C. Chekuri, P. Indyk, and R. Motwani. Fast estimation of diameter and shortest paths (without matrix multiplication). *SIAM Journal on Computing*, 28(4):1167–1181, August 1999.

[4] Noga Alon, Zvi Galil, and Oded Margalit. On the exponent of the all pairs shortest path problem. *Journal of Computer and System Sciences*, 54(2):255–262, April 1997.

[5] Krists Boitmanis, Karlis Freivalds, Peteris Ledins, and Rudolfs Opmanis. Fast and simple approximation of the diameter and radius of a graph. In Carme Àlvarez and Maria J. Serna, editors, *WEA*, volume 4007 of *Lecture Notes in Computer Science*, pages 98–108. Springer, 2006.

[6] Béla Bollobás. *Extremal graph theory*. Courier Corporation, 2004.

[7] Kellogg S. Booth and Richard J. Lipton. Computing extremal and approximate distances in graphs having unit cost edges. *Acta Informatica*, 15:319–328, 1981.

[8] Michele Borassi, Pierluigi Crescenzi, Michel Habib, Walter A. Kosters, Andrea Marino, and Frank W. Takes. Fast diameter and radius BFS-based computation in (weakly connected) real-world graphs: With an application to the six degrees of separation games. *Theoretical Computer Science*, 586:59–80, June 2015.

[9] Timothy M. Chan. More algorithms for all-pairs shortest paths in weighted graphs. *SIAM Journal on Computing*, 39(5):2075–2089, 2010.

[10] Timothy M. Chan. All-pairs shortest paths for unweighted undirected graphs in $o(mn)$ time. *ACM Transactions on Algorithms*, 8(4):34, 2012.

[11] Shiri Chechik, Daniel Larkin, Liam Roditty, Grant Schoenebeck, Robert Endre Tarjan, and Virginia Vassilevska Williams. Better approximation algorithms for the graph diameter. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014*, pages 1041–1052. SIAM, 2014.

[12] M. Cygan, H. N. Gabow, and P. Sankowski. Algorithmic applications of Baur–Strassen's Theorem: Shortest cycles, diameter and matchings. In *Proceedings of the 2012 IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2012*, pages 531–540. IEEE Computer Society Press, 2012.

[13] Wyatt J. Desormeaux, Teresa W. Haynes, Michael A. Henning, and Anders Yeo. Total domination in graphs with diameter 2. *Journal of Graph Theory*, 75(1):91–103, 2014.

[14] Dorit Dor, Shay Halperin, and Uri Zwick. All-pairs almost shortest paths. *SIAM J. Comput*, 29(5):1740–1759, 2000.

[15] Michael L. Fredman. New bounds on the complexity of the shortest path problem. *SIAM Journal on Computing*, 5(1):83–89, March 1976.

[16] Silvio Frischknecht, Stephan Holzer, and Roger Wattenhofer. Networks cannot compute their diameter in sublinear time. In Yuval Rabani, editor, *SODA*, pages 1150–1162. SIAM, 2012.

[17] Zvi Galil and Oded Margalit. All pairs shortest paths for graphs with small integer length edges. *Journal of Computer and System Sciences*, 54(2):243–254, April 1997.

[18] Teresa W. Haynes, Stephen Hedetniemi, and Peter Slater. *Fundamentals of Domination in Graphs*. Springer monographs in mathematics. CRC Press, Boca Raton, Florida, 1998.

[19] Michael A. Henning and Anders Yeo. *Total domination in graphs*. Springer monographs in mathematics. Springer, New York, 2013.

[20] R. Impagliazzo, R. Paturi, and F. Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63(4):512–530, 2001.

[21] David R. Karger, Daphne Koller, and Steven J. Phillips. Finding the hidden path: Time bounds for all-pairs shortest paths. *SIAM Journal on Computing*, 22(6):1199–1217, December 1993.

[22] Clémence Magnien, Matthieu Latapy, and Michel Habib. Fast computation of empirically tight bounds for the diameter of massive graphs. *ACM Journal of Experimental Algorithmics*, 13, 2008.

[23] Seth Pettie. A new approach to all-pairs shortest paths on real-weighted graphs. *Theoretical Computer Science*, 312(1):47–74, January 2004.

[24] Liam Roditty and Virginia Vassilevska Williams. Fast approximation algorithms for the diameter and radius of sparse graphs. In ACM, editor, *STOC '13: Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, STOCS 2013*, pages 515–524. ACM Press, 2013.

[25] Raimund Seidel. On the all-pairs-shortest-path problem in unweighted undirected graphs. *J. Comput. Syst. Sci.*, 51:400–403, 1995.

[26] A. Shoshan and U. Zwick. All pairs shortest paths in undirected graphs with integer weights. In IEEE, editor, *40th Annual Symposium on Foundations of Computer Science, FOCS 1999*, pages 605–614. IEEE Computer Society Press, 1999.

[27] Tadao Takaoka. An $O(n^3 \log\log n/\log n)$ time algorithm for the all-pairs shortest path problem. *Information Processing Letters*, 96(5):155–161, December 2005.

[28] Oren Weimann and Raphael Yuster. Approximating the diameter of planar graphs in near linear time. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *ICALP (1)*, volume 7965 of *Lecture Notes in Computer Science*, pages 828–839. Springer, 2013.

[29] Ryan Williams. A new algorithm for optimal constraint satisfaction and its implications. In Josep Díaz, Juhani Karhumäki, Arto Lepistö, and Donald Sannella, editors, *Automata, Languages and Programming: 31st International Colloquium, ICALP 2004*, volume 3142 of *Lecture Notes in Computer Science*, pages 1227–1237. Springer, 2004.

[30] Raphael Yuster. Computing the diameter polynomially faster than APSP, January 13 2010.

[31] U. Zwick. All pairs shortest paths in weighted directed graphs-exact and almost exact algorithms. In IEEE, editor, *39th Annual Symposium on Foundations of Computer Science: proceedings, FOCS 1998*, pages 310–319. IEEE Computer Society Press, 1998.

[32] Uri Zwick. Exact and approximate distances in graphs - A survey. In Friedhelm Meyer auf der Heide, editor, *ESA*, volume 2161 of *Lecture Notes in Computer Science*, pages 33–48. Springer, 2001.

[33] Uri Zwick. All pairs shortest paths using bridging sets and rectangular matrix multiplication. *Journal of the ACM*, 49(3):289–317, May 2002.