We thank the reviewers for their thoughtful feedback and for appreciating the simplicity and potentially wide impact of our results. Due to lack of space, we could only address the major comments, and in this process we add new theoretical and experimental developments which we will add to the paper if accepted.

R1: It seems the main concern is that DBSCAN++ and SNG-DBSCAN are compared using the same sampling rate which may not be fair as they may not necessarily have the same meaning. The reviewer brings up a good point. To this end, we provide Figure 1 which shows that SNG-DBSCAN is still competitive when both algorithms are optimized over both ϵ and sampling rate. We also note that these procedures may outperform DBSCAN simply because the sampling adds an additional degree of freedom and can be interpreted as a regularizer [25].

R2: The main concern is that the theoretical results have strong assumptions. The reviewer is right. Below, we give level-set estimation rates for SNG-DBSCAN under more standard and general non-parametric assumptions. The assumptions are borrowed from other works in level-set estimation (i.e. [26, 44]). Given these results, we can straightforwardly extended them to obtain clustering results with the same convergence rates (i.e. showing that SNG-DBSCAN recovers the connected components of the level-set individually), but omit it here due to space.

R3: The main concern appears to be the novelty of SNG-DBSCAN relative to DBSCAN++. We emphasize that although one samples edges and the other samples vertices, there are still considerable differences: they lead to different theoretical analyses, SNG-DBSCAN appears to perform better, SNG-DBSCAN works for arbitrary distance metrics, and unlike DBSCAN++, SNG-DBSCAN can be easily used in practice by plugging in a subsampled distance matrix into scikit-learn's DBSCAN implementation under the precomputed distance setting.

R5: The true clusters are the connected components of a particular level-set of the density function. We show that SNG-DBSCAN recovers these clusters at rates depending on various properties of the density function. The reviewer is right that since these rates depend on constants that are unknown in practice, they may have little practical use but nonetheless makes the algorithm a principled approach. We will further clarify these constant factor dependencies.

Additional Theory. We show level-set estimation rates for estimating a particular level λ (i.e. $L_f(\lambda) := \{x \in \mathcal{X} : f(x) \geq \lambda\}$) given that hyperparameters of SNG-DBSCAN are set appropriately depending on density f, s, λ and n.

Assumption 1. *f is a uniformly continuous density on compact set $\mathcal{X} \subseteq \mathbb{R}^D$. There exists $\beta, \check{C}, \hat{C}, r_c > 0$ such that the following holds for all $x \in B(L_f(\lambda), r_c) \backslash L_f(\lambda)$: $\check{C} \cdot d(x, L_f(\lambda))^\beta \leq \lambda - f(x) \leq \hat{C} \cdot d(x, L_f(\lambda))^\beta$, where $d(x, A) := \inf_{x' \in A} |x - x'|$, $B(C, r) := \{x \in \mathcal{X} : d(x, C) \leq r\}$.*

where β can be interpreted as the smoothness and curvature of f around the λ-level-set boundary of f. Define $C_{\delta,n} = 16 \log(2/\delta)\sqrt{\log n}$, $\epsilon = (\text{minPts}/(sn \cdot v_D \cdot (\lambda - \lambda \cdot C_{\delta,n}^2/\sqrt{\text{minPts}})))^{1/D}$, and minPts satisfies $C_l \cdot (\log n)^2 \leq \text{minPts} \leq C_u \cdot (\log n)^{\frac{2D}{2+D}} \cdot n^{2\beta/(2\beta+D)}$ where C_l and C_u are positive constants depending on δ, f. Then, the following holds where d_{Haus} is Hausdorff distance:

Theorem 1. *Suppose Assumption 1 holds along with the parameter settings of the above. There exists $C, C_l, C_u > 0$ depending on f, δ such that the following holds with probability at least $1-\delta$. Let $\widehat{L_f(\lambda)}$ be the union of all the clusters returned by SNG-DBSCAN:*

$$d_{\textit{Haus}}(\widehat{L_f(\lambda)}, L_f(\lambda)) \leq C \cdot \left(C_{\delta,n}^{2/\beta} \cdot \textit{minPts}^{-1/2\beta} + C_{\delta,n}^{1/D} \cdot \left(\frac{\sqrt{\log sn}}{sn} \right)^{1/D} \right) \to_{sn/\log(n), n \to \infty} 0.$$

	DBSCAN	DBSCAN++	SNG
Page	0.1118	0.0727	**0.1137**
Blocks	0.0742	0.0586	**0.0760**
kc2	0.3729	0.3621	**0.3747**
	0.1772	0.1780	**0.1792**
Ozone	0.0391	**0.0627**	0.0552
	0.1214	0.1065	**0.1444**
Bank	0.1948	**0.2599**	0.2245
	0.0721	0.0874	**0.0875**
Ionosphere	0.6243	0.1986	**0.6359**
	0.5606	0.2153	**0.5615**
Mozilla	0.1943	0.1213	**0.2791**
	0.1452	0.1589	**0.1806**
Tokyo	0.4204	0.4180	**0.4467**
	0.2830	0.2793	**0.3147**

Figure 1: *DBSCAN tuned over ϵ and SNG-DBSCAN and DBSCAN++ (which uniformly samples the nodes) tuned over ϵ (same grid as in paper for each dataset) and sampling rate (over grid [0.1,0.2,..,0.9]) to maximize ARI and AMI clustering scores. Only some datasets shown.*

Proof Sketch. There are two quantities to bound: (i) $\max_{x \in \widehat{L_f(\lambda)}} d(x, L_f(\lambda))$, and (ii) $\sup_{x \in L_f(\lambda)} d(x, \widehat{L_f(\lambda)})$. The bound for (i) follows by standard uniform kernel density (KDE) estimation bounds with uniform kernel (i.e. [26]) based on the sn samples where the first term in the rate is due to the bias of the smoothing w.r.t. ϵ and the variance term comes from sampling at a rate of s for each estimate. We now turn to the other direction and bound $\sup_{x \in L_f(\lambda)} d(x, \widehat{L_f(\lambda)})$. Let $x \in L_f(\lambda)$. Define $r_0 := ((2C_{\delta,n}\sqrt{D \log sn})/(snv_D \cdot \lambda))^{1/D}$. Using standard concentration inequalities, we show that $B(x, r_0)$ contains at least $1/s$ samples and by standard density estimation guarantees, at least one of them will have sufficiently high KDE with uniform kernel and bandwidth ϵ leading to the conclusion that its ϵ-ball contains at least MinPts edges after subsampling at a rate of $1/s$. Thus, $\sup_{x \in L_f(\lambda)} d(x, \widehat{L_f(\lambda)}) \leq r_0$. $\square$