The author(s) would like to thank the reviewers for their time and expertise. This response will begin by discussing points shared across reviews, and will address the most critical reviewer-specific points as space permits.

Infinite rank models. It does not seem to have been made clear enough that Examples 1 (always) and 2 (typically) are models where the rank D (embedding dimension) is *infinite* but the manifold dimension is finite (and equal to the latent position dimension d in Example 1). This miscommunication is apparent, for example, in Reviewer 4's comment "it would be particularly interesting to know examples of classes of infinite rank models for which the authors are able to characterize the Hausdorff dimension from Theorem 3". Inspired particularly by the comments of Reviewers 1 and 4, the revision will contain a larger diversity of examples, extracted from Hoff's and co-authors' work, wherein spectral embeddings almost always fall in the "finite d, infinite D" regime. Reviewer 4 asks "which classes of [these] infinite rank models satisfy assumption 2" and for expanded discussion of the assumption's flexibility. The answer is somewhat hidden in Example 2. *A positive-definite kernel $f(x, y)$ (of arbitrary dimension, and of finite or infinite rank) satisfies Assumption 2 with $\alpha = 1$ if its Hessian exists and is bounded along the diagonal $x = y$.* To say the same for an indefinite kernel, one must replace f with its (operator-sense) absolute value $f_+ + f_-$. This point will be added.

Selection of embedding dimension. The paper does *not* immediately offer an improvement to existing techniques. Nevertheless, in light of the reviewers' comments, there will be added discussion of this topic in the main text, which was previously delegated to the supplementary material and to reference [47]. The method by Zhu & Ghodsi (2006), which uses a profile-likelihood-based analysis of the scree plot, has for a long time provided a functional choice for many practitioners and is easily used within the R package igraph. For a theoretical treatment of dimension selection, the cases $D < \infty$ and $D = \infty$ must be distinguished. In the former, simply finding a consistent estimate of D has limited practical utility: appropriately scaled eigenvalues of the adjacency matrix converge to their population value, and all kinds of unreasonable rank selection procedures are therefore consistent. But, to quote [47], "any quest for a universally optimal methodology for choosing the "best" dimension [...], in general, for finite n, is a losing proposition". In the $D = \infty$ case, reference [33] finds the appropriate rate under which to let $\hat{D} \to \infty$, to achieve consistency in the Wasserstein metric. Unlike the $D < \infty$ case, stronger consistency, i.e., in the maximum latent position error, is not yet available in the $D = \infty$ case — this is an ongoing and nontrivial effort. Reviewer 1 asks how the estimate $\hat{D} = 10$ was selected in Section 5.4. Here, because analysis is partly reproduced from [50], it seemed most expedient to abide by that paper's original choice, which was avowedly arbitrary. After computing the full spectrum overnight, the method of Zhu & Ghodsi (2006) actually returns an estimate $\hat{D} = 6$, so this will be reported.

How theoretical results presented relate to or inform actual practice. Reviewers 2 ("all applications seem to have only loose connections to the main theory"), 3 ("it's not totally clear how the theoretical results presented relate to or inform actual practice.") and 4 ("the paper is limited in the study of the implications of the theoretical results") would have liked to see more direct methodological applications of the theory. This will be addressed through the following revisions.

• Section 5.1 (Graph regression) — added discussion of achievable error rate. It may be conjectured that spectral embedding with $\hat{D} \to \infty$ appropriately slowly followed by neural network regression can achieve the rate $n^{-2\beta/(2\beta+d)}$, where β is the regression function smoothness (in the Hölder norm) and d is the *intrinsic* dimension of the data. This conjecture leverages recent work in reference [40] on the rate of neural networks under low intrinsic (Minkowski) dimension, and hopes of a strong consistency bound (mentioned above) under $D = \infty$.

• Section 5.3 (Visualisation A) — to be removed.

• Section 5.4 (Visualisation B) — to be renamed "Real data", include a new figure and report intrinsic dimension estimates. In anticipation of the comment "why do the main results help make sense of t-SNE for visualization?" (Reviewer 2), there was a brief remark in the original submission "Other methods such as Uniform manifold approximation were tested with comparable results". In retrospect, these results should simply have been presented instead, as they show similar information and are based on a concrete assumption that the data live close to a manifold of low dimension. These will now be included as an additional figure. Instead of selecting the embedding dimension, what the paper *does* allow is estimation of the latent position dimension d, on the basis of the intrinsic dimension of $\hat{X}_i$ (and assuming $\alpha = 1$ in Theorem 3), which can be estimated by several existing techniques (several implemented in Table 2, reference [50]).

Reviewer 2: "I'm not sure the results attract much attention in the NeurIPS community". Graphs have become some of the most studied objects in statistics/machine-learning and von Luxburg's tutorial on spectral clustering has over 8000 citations according to Google Scholar. In a nutshell, the present paper makes a case for "spectral manifold estimation" rather than "spectral clustering".

Reviewer 3: "I do think the motivation is a bit misleading [...]". The second sentence will be changed to "the object of this paper is show that, for a theoretically tractable but rich class of random graph models, such a phenomenon occurs in the spectral embedding of a graph."

Reviewer 3: "does the fact that the generated graphs are almost surely dense matter"? Sparsity *is* handled in the paper, if crudely, with a sparsity factor (Section 4 and Supplementary Material). However, from the suggested references it is clear that the reviewer has more sophisticated sparsity-inducing processes in mind, and these will be discussed.