We thank the reviewers for their detailed reviews and suggestions. We address the most relevant concerns below.

Performance of LPCA embedding in downstream tasks (e.g., link predication and node clustering): While beyond the scope of our current work, the performance of our embeddings in downstream tasks is a very interesting question. In our work we show affirmatively that low-dimensional embeddings can in fact capture structural information like the triangle density and degree sequence. A broad next question we seek to answer is: how does the ability to preserve such structural information impact an embedding's usefulness in downstream application?

Comparison to baselines beyond TSVD (e.g., DeepWalk, Node2Vec): The prior work for Seshadhri et al. gives a detailed comparison between node2vec and TSVD, which are comparable on triangle density reconstruction. We will clarify this in the final version. It would be interesting to expand this comparison to other methods, although, related to the question above, such methods may perform worse than LPCA on graph construction but still perform well on other downstream tasks. We are actively working on follow up work which seeks to 'invert' modern embeddings to reveal graph structure such as triangle density, the degree sequence, and edges.

Reviewer 1:

The analysis is performed on low-rank factorizations of the adjacency matrix of the graph. Although very interesting, this is quite limited; the majority of recent embedding models factorize more 'informative' node proximity matrices.

As the reviewer suggests, capturing higher order connectivity relates closely to today's most popular embedding methods. We view the submission as a step in this direction. Before our work, even for the simple adjacency matrix, the power of low-dimensional embeddings was previously unclear, and the large gap between using one embedding per node (a PSD embedding as in Seshadhri et al.) vs. two (a non-PSD embedding as in our work) was unknown.

In Theorem 5, how arbitrarily large c (maximum degree) can be?...shouldn't we focus on the expected degree instead?

Theorem 5 holds for any c, from 1 to $n-1$. The expected degree of our construction is the same as the maximum degree (the graph is regular), so the same theorem holds with 'maximum degree' replaced by 'expected degree'.

In Theorem 6, I missed the connection... to sign-rank... is $\sigma()$ used as shown in the paper or the sign function $s()$?

$\sigma()$ is used as shown. If we scale the entries of X, Y by a large enough value then $\sigma(XY^T)$ exactly equals $s(XY^T)$, so any results for sign rank directly apply in our model. We will expand the discussion of this in the final version.

Reviewer 2:

Result 2 seems to suggest the optimization for LPCA has guaranteed global convergence...Could the authors confirm...

The low-rank LPCA objective is non-convex and convergence to a global minimum is not *guaranteed*. However, for all graphs tested, we achieved exact recovery and so did in fact converge to a global minimum. Understanding this theoretically would be very interesting, as would be applying relaxation methods as suggested by the reviewer, which may give provable convergence and/or not require setting the rank k a priori.

Does LPCA have superior performance in terms of other substructures or proximity metrics? (outside degree/triangles)

We also performed experiments on 5-cycle density and other graph motifs, showing superior performance of LPCA. We will add to the supplemental material. When LPCA gives an exact factorization, all substructures are exactly preserved.

Reviewer 3:

The minor relaxation for the model in [SSSG20] by introducing two embedding X and Y for the node embedding, may be not exactly the original problem in [SSSG20], which consider the PSD low-rank embedding for the graph.

This is correct. However, we argue that our relaxation is in fact more natural. Many modern embedding methods do not produce a PSD factorization, and thus fall into our model, but not the one of [SSSG20].

Reviewer 4:

In the last paragraph of page 6, it is hard to understand why 'EFD is consistently higher for the random networks' is equivalent to "the embeddings capture structure inherent to real-world networks outside just the degree sequence'.

Thanks – we will clarify the discussion. At a high level, higher EFD for the random networks suggests that real-world networks must have some structure that can be compressed, leading to lower EFD. More explicitly understanding this is a key direction for future work, related to understanding gaps between real world networks and random graph models.

Theorem 6 presents an important result but the proof seems to be missing (not found in the supplement).

We will give a detailed proof in the final version. See our answer to Reviewer 1 above which explains the connection to sign rank and gives a proof sketch.