We thank the reviewers for reading the paper and for their detailed comments. We are happy that all reviewers agree the problem is important and that the approach is natural and interesting.

Assumption 2 R3 and R4 note that assumption 2—the embedding method extracts the salient predictive information from the network—may be difficult to assess in practice. This is true, and, as both reviewers note, we make this explicit in the paper. We say that such judgments should be based on experience with the embedding methods, and empirical performance in related situations. For instance, if a graph embedding method does well at predicting many features associated with social identity (e.g., age, music preference, political affiliation) in a number of contexts then we can expect that the embedding method will adjust for features associated with social identity. By analogy, it is commonly believed that convolutional neural nets extract semantically-meaningful representations of images, but this belief exists independent of supporting theory, generative models, or clear heuristics for reasoning about when such representations might succeed. Direct judgments about what sort of information a method might extract are often easier to make than judgments about assumptions required for (hypothetical) formal results.

That said, we agree identifying conditions that guarantee good behavior of embedding methods is an important and fascinating topic with application to the causal inference approach here, and is an good direction for future work. However, establishing such results is outside the scope of the paper. We think (and we believe the reviewers agree) that the contribution is appropriately developed for a NeurIPS paper exploring a newly-established and unconventional method.

Partially correcting for confounding R3 and R4 both note that partially correcting for confounding may hurt. This is indeed possible for the reasons the reviewers point out. We have added a discussion of this point and the need for the investigator to consider it. The point about variables that strongly affect treatment but not outcome is well-taken (though note that since the embedding construction extracts information relevant to both treatment and outcome, the method automatically mitigates this somewhat).

R2 We have corrected the typos you point out (1. is anonymized ref, 2. is estimate of all nuisance params, 3. is sigmoid)

R3 *Linear Gaussian Model* You point out that it is possible for a linear Gaussian model with the wrong (mediator) causal structure to satisfy Assumption 1. We agree; this was a minor error on our part. We have followed your suggestion and replaced assumption 1 with the (slightly stronger) assumption of the DAG such that the network and Z are not affected by the treatment. Beyond the fact that it fixes the bug, we think this condition is more interpretable—thank you!

X is not a 'noisy' observation. We have clarified this point. This is an important way in which the network setting deviates from the iid proxy setting. In the iid setting, each x_i is a noisy version of the associated x_i. In the network setting, we expect to get more and more information about each node as the network grows. For example, in stochastic block models, community identities can be recovered exactly under very weak conditions (Bickle and Chen 2009). Intuitively, your social identity can be fully inferred in the limit of infinite possible possible friendships. The assumption that the network asymptotically fully reveals z is appropriate (or, at least, standard) in this setting.

Text Embeddings for Causal Inference You write that you read the text embeddings paper, and suggest merging it with the networks paper. We are gratified you enjoyed it. In our view, the papers are complementary, and should be published independently. Indeed, the first version of this work combined both ideas in a generic 'embeddings for causal inference' paper. Reviewers for ICML (including R4) encouraged us to separate the papers. That was the correct choice—each paper is much improved in clarity and development in this way. The results from each paper do not anticipate the results from the other—the fact that we can use graph embeddings to correct for unobserved confounding in networks does not make it obvious that we can use document embeddings for identification-preserving dimension reduction in text. Beyond that, the technical development is quite different: the text paper does not address non-iid data, and does not require the extension of double-ML methods to handle this. Further, the experimental setup and implementation (data, methods, and challenges) are disjoint for the two papers.

That said, if you are committed to only supporting one of the two papers, we ask that you make it the networks one. This paper handles a more unusual learning situation, and has a more substantial technical development.

(For the benefit of R4, we note that although the network paper is a small revision of the earlier version, the text paper was hugely reworked, and now focuses on dimension reduction and applications. Unfortunately, NeurIPS policy does not allow external links in reviewer response.)

R4 Thank you for the second round of careful comments! We are gratified you agree that the paper is much improved. *Why is the parametric model given 128 blocks?* So that the dimension of the community assignment vector matches the capacity of the 128-dimensional embedding model. 128 was chosen because it is the default setting for the graph embedding method we used. We chose to match to avoid hyperparameter tuning issues, but note that (i) the network is very large, so the high capacity should not disadvantage the blockmodel, and (ii) informal testing suggested results remain poor irrespective of number of blocks.