We would like to thank the reviewers for their thorough evaluation and constructive feedback. They have been really helpful in improving our work. Below, we address the main comments. Our revisions will be incorporated in the camera-ready version along with the additional related work, corrections and comments brought forth by the reviewers.

Implications of equivariance and independence to the initial choice of identifiers (R1) SMP's equivariance properties ensure that a change in the one-hot encoding results in a permutation of the rows of each local context U_i. However, in order to produce a final output (for node or graph classification), the rows of each U_i are pooled into a vector (in an equivariant manner as well). As a result, *the output of each node is independent of the initial one-hot encoding*, and the latter need not be consistent across examples and/or layers. Equivariance has some other useful consequences:

Local isomorphisms: if the subgraphs G_i^k and G_j^k induced by G on the k-hop neighborhoods of v_i and v_j are isomorphic, then on node classification, any k-layer SMP f will yield the same result for v_i and v_j. To prove it, we first observe that $f(G)_{v_i} = f(G_i^k)_{v_i}$ and $f(G)_{v_j} = f(G_j^k)_{v_j}$, and then write the definition of equivariance for an isomorphism π mapping G_i^k to G_j^k and i to j.

Transductivity: since all equivariant functions can take a variable number of inputs, SMP can be used transductively. If a node is added to a graph, a new line is simply created in each local context and there is no need to retrain the model. Experimentally, we observed that SMP managed to generalize to larger graphs than those seen during training: e.g., when we trained SMP to detect 4-cycles on graphs with 20 nodes (where it reaches 100% test accuracy), we obtained 99.05% accuracy when evaluating the same task on graphs with 36 nodes.

Isomorphism testing: To test isomorphism using SMP, one can pool after each layer the $n \times d$ local context of each node into a feature vector in $\mathbb{R}^d$. For this purpose, a universal approximator of functions on sets (such as Deep Sets) can be used. Similarly to MPNNs, isomorphism can then be tested by comparing the multisets of node features after each layer. Equivariance guarantees that these multisets do not depend on the initial choice of the one-hot encoding, so that there is no need to sum over permutations—a key difference between SMP and relational pooling methods from the literature.

Discussion about the theoretical results (R2) We would like to elaborate on three of the reviewer's points:

Universality with features: Theorem 2 can be extended to attributed undirected graphs, but $d_{\text{nodes}} + nd_{\text{edges}}$ more channels are required in this case. For the node features, d_{nodes} channels can be used to store the features of all nodes using a variation of max pooling. For edge features, the same sketch of proof as Theorem 2 can be used: if each node can store tensors of size $n \times n \times d_{\text{edges}}$, they can all recover the edge features. However, another embedding is needed as Lemma 1 does not apply anymore. If the graph is undirected, the square root matrix of each feature (which may be complex-valued) constitutes a valid embedding, as it permutes as desired. However, this embedding does not compress the representation, so that $n \times d_{\text{edges}}$ new channels are required. Corollary 1 follows in the same way as previously.

Expressivity: We do not yet have any formal results stating whether SMP is strictly more expressive than Fast SMP. Still, we observe that the proof that PPGN is at least as expressive as Fast SMP does not apply to SMP. This stems from the fact that SMP computes messages of the form $m(U_i, U_j, e_{ij})$, while PPGN can only store messages of the form $m(U_j)$.

Equality of the embeddings in the limit: This is an interesting question. As Lemma 1 is not constructive, it is indeed unclear at this point whether all node embeddings will become equal at infinite depth. At this point we can only observe that it is a possible scenario.

Scalability and lower bound on the complexity (R3, R4) Although SMP is more efficient than previous powerful equivariant methods, large graphs exhibiting the small-world property indeed constitute a challenge. In this case, the scalability of SMP can be improved by simply using fewer identifiers (and ignoring conflicts), at the cost of breaking the theoretical guarantees of the network. We plan to investigate this extension in our future work.

We also agree that lower bounds on the complexity required for universality would be very valuable to the community. We are aware of two results towards this direction: (*i*) if the feature space is continuous, at least d_{max} width is required to make the aggregation function injective [20]. (*ii*) for all message-passing methods (including SMP), solving some simple combinatorial tasks necessitates depth $\times$ width $= \Omega(n)$ [18].

Additional experiments (R1, R2, R4) Following the reviewers' suggestion, we ran experiments on Ring-GNN and Relational Pooling (RP): (*i*) Ring-GNN could solve the cycle-detection problem up to $k = 8, n = 50$. However, in this configuration ($k = 8, n = 50$), it required $5\times$ more epochs and $10\times$ more time than SMP to converge (*ii*) RP with π-SGD (summing over 8 permutations) obtained 100% accuracy on all training sets, but exhibited overfitting (which was not observed on equivariant methods): for $k = 6$, $n = 56$, test accuracy was 84.1% for RP against 99.8% for SMP.

Finally, we acknowledge the importance of additional benchmarking on tasks where both features and structure play a role. We are currently working on the QM9 and ZINC datasets, and plan to make the method available in Pytorch-geometric.