We thank all the reviewers and the AC for their time, effort and constructive feedback. [W1], [W2] and [W3] are references included in this response.

R1, R2, R3: Suitability of numerical experiments. We appreciate the concern of all reviewers with respect to the numerical simulations. We would like to note that (i) this is mainly a theoretical paper that proves properties of the GST (as R3 remarked) and that (ii) the Diffusion GST is very similar to the method in [W1] which has been compared extensively with other methods (as R1 observed), and therefore we expect similar numerical results as those in [W1]. In any case, we understand, and share, the concerns of the reviewers, so we propose the following changes to the numerical section. First, we will include an explicit comparison with the GST of [W1]. This method will replace the diffusion scattering, since both are very similar (the only differences being the use of the lazy random walk matrix instead of the lazy adjacency matrix, and the use of moments beyond the mean for the low-pass operator ϕ). Second, we will include comparison with a trainable GIN in [W2], in terms of stability of the resulting architectures. We note that comparing performance with trainable GNNs is tricky since it is highly dependent on the size of the available training set and the details of the training stage (number of epochs, learning rate, etc.), which do not occur in GSTs (which are not trainable). Third, we will add clarification and proper links to the Facebook graph [35] and the authorship attribution dataset [36, W3], to emphasize that these are publicly available, while explaining that we are concerned with datasets involving graph signals, since we want to show how changes in the underlying topology affect the processing of the same signals (i.e. datasets involving graph classification, as those in [W1], where changing the underlying graph changes the graph signal are not useful to illustrate Theorem 1 –even though, in practice, they work–). Fourth, as suggested by R1, we will add a clarification and give due credit to the very good work of [W1] to refer to a more exhaustive comparison between GSTs and other state-of-the-art methods. We hope that these changes will address the concerns of the reviewers. **R1:** Structural constraint and recovery of Mallat's scattering result. The structural constraint allows for edge weights dilations or contractions (i.e. all edge weights increase or all edge weights decrease, albeit with different relative changes). This is required to control the impact that topology changes have on the eigenvectors. Changes such as adding or dropping edges incur in a constant value $\varepsilon = \mathcal{O}(1)$, and as such fix a nonzero minimum for the upper bound. In the very limited number of cases when a topology change can be exactly pinpointed to a change in the eigenvectors, the results in this paper can be improved. One such case is that of the line graph, where it is known how the eigenvectors change when dilating and contracting the edge weights (the effect of a diffeomorphism in [10]), and thus recovering the result in [10]. As space allows, the first observations will be added before Remark 2, while the latter observation will be moved from Remark 1 to a new paragraph and expanded. If necessary, further clarifications on these relationships will be discussed in the supplementary material. **R1:** Graph similarity measures. We would like to clarify that the task is not to compare graph structures in terms of their extracted features, but to analyze how features extracted from graph signals change when the underlying support changes (either because it changes with time, or because it is unknown and has to be estimated, among other examples). The measure of similarity we use in this work is reminiscent of the Gromov-Hausdorff distance, albeit using the spectral norm of the GSO, instead of a max-norm. The comparison with Weisfeller-Lehmann test will hopefully be taken into account by the inclusion of the GIN [W2] in the numerical experiments. **R3:** Relation to other GNNs. Most existing GNNs (with the notable exception of GATs) regularize the linear transform of traditional neural networks by using a graph convolution (5). In this respect, the main computational core of doing a graph convolution followed by pointwise nonlinearities, is the same in GSTs than in GNNs. The main exception, though, is that while GNNs learn the filter coefficients h_k (through different parameterizations), GSTs design them using graph wavelets. Likewise, since Prop. 2 shows stability of the graph filters, which are the same as for GNNs, our stability results may be extended to GNNs with appropriate regularization (since trainable parameters will appear in the bound constants) which is the subject of ongoing work. **R3:** Prop. 3. The formal assumption in Prop. 3 indicates that all involved graph filters in the multirresolution wavelet bank have to satisfy the integral Lipschitz continuity. However, this can be inherited directly from the mother wavelet satisfying the requirement. The hypothesis in Prop. 3 will be changed to reflect this. **R3:** Theorem 1. The bound in Theorem 1 depends on difference between the graphs as defined in (16). This difference will certainly depend on the particularities of the graph topologies considered. Theorem 1 states that it does not depend on the spectral norm of the graph. This will be clarified after (19). **R3:** Different number of nodes. As the theorem is stated now, it requires that both graphs have the same number of nodes. The case when they do not, can be addressed by using correspondences in the same manner as Gromov-Hausdorff distance. This case is beyond the scope of this paper and is currently ongoing work. **R2:** Computation of bound in Fig. 2. The bound in Fig. 2 is computed as in (19). The values of all the constants involved are explained in the supplementary material due to lack of space. In any case, we will add a specific clarification pointing out to this fact in the revised version. **R3:** Update of literature review. We thank the reviewer for bringing to our attention this recently published papers. They will be added to the introduction, and discussed.

[W1] F. Gao, G. Wolf, and M. Hirn, "Geometric scattering for graph data analysis", in *ICML 2019*.

[W2] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *ICLR 2019*.

[W3] E. Isufi, F. Gama, and A. Ribeiro, "Generalizing Graph Convolutional Neural Networks with Edge-Variant Recursions on Graphs," in *EUSIPCO 2019*.