We would like to thank all reviewers for taking the time to provide our work with extensive and constructive feedback, which will be very useful for improving our work for the final version.

- **R1 and R2: expressiveness** R1 and R2 requested a demonstration that the resulting GCN^2 method is indeed more expressive than a GCN. Contrary to the global equivariant methods, our method can not implement a Weisfeiler-Lehman test, making theoretical analysis more difficult. Instead, we use an empirical evaluation, similar to [1]. We use a neural network with random weights on a graph and compute a graph embedding by mean-pooling. Then we say that the neural network finds two graphs in a set of graphs to be different if the graph embeddings differ by an L2 norm of more then a multiple of $\epsilon = 10^{-3}$ of the mean L2 norms of the embeddings of the graphs in the set. The networks is most expressive if it only finds isomorphic graphs to be not different. We test this on (A) a set 100 of random non-isomorphic, non-regular graphs, (B) a set of 100 non-isomorphic regular graphs, (C) a set 15 of non-isomorphic strongly regular graphs (see `http://users.cecs.anu.edu.au/~bdm/data/graphs.html`) and (D) a set of 100 isomorphic graphs, where all graphs have 25 nodes and average of degree 6. We measure average difference rate between pairs of graphs in the sets over 100 different weight initialisations. We compare the simple invariant message passing (GCN), PPGN [2], and our GCN^2. We see that only our GCN^2 can disambiguate between the strongly regular graphs, showing the expressivity of GCN^2. A version of PPGN that uses higher order tensors should also be able to discriminate strongly regular graphs, but at even higher computational cost.
- **R1 and R3: runtime cost?** As an additional experiment we show the runtime cost of one forward-pass of GCN, PPGN and our GCN^2. The models have three layers and 32 dimensional activations. For simplicity, we use a square lattice as graph, in which the number of edges is proportional to the number of nodes. In the results below, we observe that GCN^2 has indeed a linear scaling and a multiplicative constant about 2x compared to GCN. If the average degree of the graph is higher, this constant may be higher. The global PPGN methods scales superlinearly. We will add more extensive experiments to the final version.
- **R1, R2 and R6: figures unclear.** We will add extensive captions to the figures to clarify.
- **R1: local symmetries superset?** We indeed mean that global syms restrict to local syms and will clarify.
- **R1: equivariance necessary for expressiveness?** In its simplest linear form, an invariant message passing network equals a GCN with limited expressivity, making more general equivariance necessary for equivariance. We agree with R1 that for more general non-linear kernels, such a general statement does not hold. We will clarify this.
- **R1: kernel shared?** Indeed, we mean that the parameters are shared, but the resulting kernels differ by a basis transformation. This is implied by Eq. (4), when the isomorphism is between different edges neighbourhoods. We used edge isoms and edge neighbourhood isoms interchangeably. We will clarify these points.
- **R1: graph feature?** By graph feature, we mean some feature over an entire graph, e.g. the stacking of all node features for a conventional GCN or the matrix feature of [2]. The idea of Sec 4 and Fig 7 is to embed a node feature (in the vector representation) into a feature of the edge neighbourhood and then to re-interpret the the edge feature of one edge as an invariant GCN-like graph feature over the edge neighbourhood graph, having a scalar feature at each node of the edge neighbourhood. Please see App. C for more details. We'll clarify this in Sec 4.
- **R1: corrections.** Thank you for finding the typos. We will correct them.
- **R2: discriminate Fig 2?** We ran the experiment mentioned above additionally on the graphs of Fig 2 and found that a conventional GCN always returns the same embedding, while our GCN^2 is able to discriminate.
- **R6: overloading of symbols.** Thank you for pointing out the fact that our re-use of the symbol ϕ for distinct related concepts is unclear. We will fix that.
- **R6: technicality.** We think that the use of gauges are a necessity to precisely describe the feature spaces and equivariant maps. Also, we agree with R1 and R3 that the categorical perspective is a useful ways of thinking about these models. However, we agree with R6 that besides the technical sections, we should explain better intuitively how the resulting model works. We will fix this in the final version.
- **R6: references.** We thank R6 for providing three interesting and relevant references. We will relate our model to these in the final version.

Model	Random	Regular	Str. Regular	Isom.
GCN	1	6E-8	0	0
PPGN	1	0.97	0	6E-8
GCN^2	1	1	1	6E-8

Table 1: Rate of pairs of graphs in set found dissimilar in expressiveness experiment. An ideal method finds only isomorphic graphs not dissimilar.

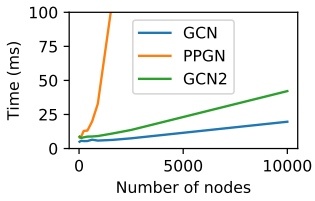

Figure 1: Runtime cost of one forward-pass on square lattices.

[1] Giorgos Bouritsas et al. Improving graph neural network expressivity via subgraph isomorphism counting. 2020.

[2] Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. 2019.