We thank the reviewers for the feedback and comments, in what follows we address specific comments made by the reviewers

Reviewer 1

I do not completely understand (apart for some parts of the proofs) why refer to these functions as Graph-based. Boolean k-ary functions may be thought of as hyper-graphs. To some this point of view may be confusing and to some it may be helpful. Our original intuition was that certain constructions from graph theory will be helpful in the construction of some lower bounds (specifically in the expressiveness section) – hence we defined the model in terms of graphs. Since the submission we have made certain intermediate progress that seem to confirm our original intuition that the graph-based point of view is indeed helpful, from a technical perspective. But we agree that in the current paper this did not come into play, and we might tone down the graph-viewpoint in the final version.

Reviewer 2

I was less certain about the significance of the result for finite domains; if there are natural and important examples of distinguishing classes with sublogarithmic sample complexity, it might be good to mention them somewhere in the paper It is true that in many natural cases the VC dimension scales at least logarithmically in the domain size, nevertheless our point here was to quantitatively assess the previous result, note that we did not prove a lower bound, so apriori it might be that we can give better constructions that yield maybe even linear sample complexity.

The definition of the empirical frequency of an edge (between lines 176 and 177 on page 5) seems a bit unusual if I am interpreting it correctly . . . The definition shouldn't be unusual and it will be clarified to avoid any possible confusion. We count the probability of an edge given k-iid vertices drawn from the empirical sample (with repetitions): $\frac{1}{m^k}\sum_{u_{1:k}\in S^k} g(u_{1:k})$. This is completely analogous to the standard empirical distribution for hypotheses classes.

In the example you write $v_1 = v_2$ but $v_1 \in E_g$ yet $v_2 \notin E_g$. this seems a bit confusing. How can $v_1 = v_2$ but one is in E_g and the other not, can you please elaborate on the example so we may understand the confusion? Again, we will make sure to clarify this point.

It might be helpful to summarise, ..., some basic properties of this new notion of VC dimension... ..., is there a Sauer-Shelah type upper bound on the size of the class in terms of the graph VC dimension? We will be happy to elaborate on these in the main text. We have succinctly discussed the relation to the VC dimension (Here VC dimension refers to the VC dimension of a family of k-ary Boolean functions, considered as Boolean functions over the domain X^k). The VC dimension upper bounds the graph VC dimension (and that is what we mean by weaker: Small VC dimension entail small graph VC dimension). Regarding Sauer Shelah: It is kind of surprising but there is no Sauer Shelah Lemma for graph VC dimension, indeed this is noteworthy and we should discuss this in the main text.

I would have also liked to know more about the impact of this work on the study of GANs Discrimination plays a crucial role in the GANs setting, where a discriminator that observes data from true and synthetic distributions and needs to distinguish. Normally, the discriminator trains a neural net to distinguish the distributions – this is as done in the standard learning setting of classification. This work is the first to, theoretically, study discriminators of higher order types that observe multiple points (such empirical works were in fact considered, see related work).

Reviewer 3

I found the paper difficult to follow and I am concerned by the lack of simulation studies and real application. A possible very real application is GANs, where discriminators are widely used during training. Since our results are mostly theoretical, studying the expressiveness of different models, simulations would not indicate the theoretical results we are searching for.