Open Problems: Approximation Power of Invariant Graph Networks

Haggai Maron **Heli Ben-Hamu** **Yaron Lipman**
Weizmann Institute of Science
Rehovot, Israel

Abstract

Learning graph data is of huge interest to the machine learning community. Recently, graph neural network models motivated by algebraic invariance and equivariance principles have been proposed (Ravanbakhsh et al., 2017; Kondor et al., 2018; Maron et al., 2019b). These models were shown to be universal (Maron et al., 2019c; Keriven and Peyré, 2019), in contrast to the popular message passing models Xu et al. (2019); Morris et al. (2018). In this note we formulate several open problems aiming at characterizing the trade-off between expressive power and complexity of these models.

A basic requirement from a hypothesis class is approximation power. Classical results in neural network literature establish approximation power of Multilayer Perceptron (MLP) (Cybenko, 1989). In a nutshell, given an arbitrary continuous function $g : \mathbb{R}^d \to \mathbb{R}$ it can be approximated with an MLP f over a bounded Euclidean domain $\Omega \subset \mathbb{R}^d$. If one desires to learn graph data $G = (V, E)$, $|V| = n$, one approach is to represent G as an Euclidean vector in $\mathbb{R}^{n\times n}$ and use an MLP. This approach enjoys the expressiveness (i.e., approximation power) of MLPs however suffers from three main drawbacks: (i) it can only work on graphs of fixed size, $n = \text{const}$; (ii) it is not well-defined on graphs. That is, encoding a graph G as a vector $x \in \mathbb{R}^{n^2}$ is done once the vertices V are put in order. Two different orders will lead to two different vectors $x \neq x'$ and a general MLP will give $f(x) \neq f(x')$, providing a different output to the *same* graph; (iii) One of the main implications of the previous phenomenon is that MLPs suffer from poor generalization on graph data.

A standard approach to alleviate these issues is to use neural networks that are by construction *invariant* to vertex ordering. Unfortunately, most previous attempts (and all scalable ones) to build such models resulted in a strict decrease in the expressiveness, that is, in the ability of the model to approximate arbitrary invariant graph functions. Therefore, the following remains an open problem:

Meta-problem: *Find a scalable graph neural network model that is invariant and expressive.*

One notion suggested to test expressiveness of graph networks is to compare them to the k-Weisfeiler-Lehman (WL) graph isomorphism tests (Grohe, 2017). The logic is: if an invariant network can distinguish between two non-isomorphic graphs it can attach a different label to each graph.

The popular message passing models (Battaglia et al., 2016; Gilmer et al., 2017) are equivariant or invariant by construction but unfortunately known to have expressive power bounded by the 1-WL test, a.k.a. the color refinement algorithm (Xu et al., 2019; Morris et al., 2018).

Graph Invariant Networks (IGNs) use a 2-dimensional adjacency-like feature tensors, $\mathbb{R}^{n\times n}$ as the graph representation method for neural networks (Kondor et al., 2018; Maron et al., 2019b,a; Chen et al., 2019). IGNs are composed of permutation equivariant layers, L_i, $i = 1, \dots, I$ followed by a permutation-invariant layer H and an fully connected network M. An IGN is defined by

$$F = M \circ H \circ L_d \circ \cdots \circ L_1. \tag{1}$$

We make a distinction between linear equivariant and/or invariant layers and polynomial layers. We accordingly refer to linear IGN and polynomial IGN. Higher-order k-dimensional tensors, that represent high order interactions between graph nodes (e.g., between triplets of vertices), were also

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

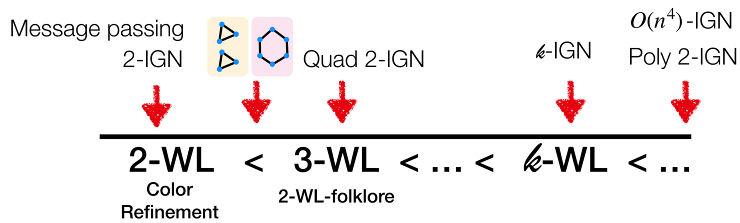

Figure 1: Current results on expressive power of Invariant Graph Networks (IGN).

suggested as intermediate representations in the network. We will denote by k-IGN an IGN with maximal tensor order of k (e.g., the order of an adjacency matrix $\mathbb{R}^{n^2}$ is two).

Known results. Following (Morris et al., 2018; Xu et al., 2019) that analyzed the message passing model, several results regarding the approximation power of IGNs were proved. These results are summarized in Figure 1.

(i) Maron et al. (2019b) proved that linear 2-IGN is at-least as powerful as the general message passing model in Gilmer et al. (2017). Chen et al. (2019) proved that linear 2-IGNs are not universal, namely, cannot approximate arbitrary permutation invariant graph functions. Maron et al. (2019a) proved that a linear k-IGN can implement the k-Weisfeiler-Lehman (WL) graph isomorphism test. This implies that any increase in the tensor order, for $k > 2$, results in additional representation power over message passing.

(ii) Maron et al. (2019c) and Keriven and Peyré (2019) proved that linear k-IGNs are universal approximators of permutation invariant functions on graph data for large enough k. For example the k currently known is polynomial in n, i.e., $k = O(n^4)$, producing intractable size tensors.

(iii) Maron et al. (2019a) proved that quadratic 2-IGN, in particular linear 2-IGN with the addition of quadratic equivariant layer (i.e., matrix multiplication), is as powerful as the folklore 2-WL test which is as powerful as the 3-WL test (Grohe and Otto, 2015). This implies that quadratic 2-IGN is strictly more expressive than message passing. Surprisingly, just using a single equivariant linear operator (scaled identity) and single equivariant quadratic operator (matrix multiplication) is already enough to achieve 3-WL expressiveness.

(iv) Yarotsky (2018) proved that polynomial IGNs (i.e., the layers are equivariant polynomial maps) are universal.

Open problems. As a possible road map toward our meta-problem and to better understand the expressiveness landscape, we offer the following open problems:

We suspect that k-IGNs cannot implement higher WL tests. Proving the following problem will show (along with (i) above) that the k-IGN hierarchy exactly matches the k-WL hierarchy, which is a nice theoretical result.

Problem 1. ***Upper bounds on linear IGN expressive power.*** *Can linear k-IGN implement $k+1$-WL test for some k?*

A natural extension of linear k-IGNs is polynomials k-IGN, or more specifically (k, l)-IGNs, where using polynomial equivariant/invariant layers of degree up to l and tensors of order of up to k is permitted. We conjecture that in general incorporating polynomial layers in IGNs allows us to get more expressive models while staying with a low tensor order representation for memory and computational efficiency. One result in this direction is (iii) above.

Problem 2. ***Polynomial IGN expressive power.*** *Can (k, l)-IGN implement $(k + s)$-WL, $s \geq 1$?*

Result (iii) also demonstrates that not all polynomial layers are needed for achieving better expressiveness. Another practically important problem is

Problem 3. ***Characterize minimal polynomial bases*** *of layers of (k, l)-IGN (in terms of number of elements or some computational complexity measure) that achieve k-WL expressiveness.*

References

Battaglia, P. W., Pascanu, R., Lai, M., Rezende, D., and Kavukcuoglu, K. (2016). Interaction Networks for Learning about Objects, Relations and Physics.

Chen, Z., Villar, S., Chen, L., and Bruna, J. (2019). On the equivalence between graph isomorphism testing and function approximation with gnns. *arXiv preprint arXiv:1905.12560*.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pages 1263–1272.

Grohe, M. (2017). *Descriptive complexity, canonisation, and definable graph structure theory*, volume 47. Cambridge University Press.

Grohe, M. and Otto, M. (2015). Pebble games and linear equations. *The Journal of Symbolic Logic*, 80(3):797–844.

Keriven, N. and Peyré, G. (2019). Universal invariant and equivariant graph neural networks. *CoRR*, abs/1905.04943.

Kondor, R., Son, H. T., Pan, H., Anderson, B., and Trivedi, S. (2018). Covariant compositional networks for learning graphs. *arXiv preprint arXiv:1801.02144*.

Maron, H., Ben-Hamu, H., Serviansky, H., and Lipman, Y. (2019a). Provably powerful graph networks. *arXiv preprint arXiv:1905.11136*.

Maron, H., Ben-Hamu, H., Shamir, N., and Lipman, Y. (2019b). Invariant and equivariant graph networks. In *International Conference on Learning Representations*.

Maron, H., Fetaya, E., Segol, N., and Lipman, Y. (2019c). On the universality of invariant networks. In *International conference on machine learning*.

Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., and Grohe, M. (2018). Weisfeiler and leman go neural: Higher-order graph neural networks. *arXiv preprint arXiv:1810.02244*.

Ravanbakhsh, S., Schneider, J., and Poczos, B. (2017). Equivariance through parameter-sharing. *arXiv preprint arXiv:1702.08389*.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2019). How powerful are graph neural networks? In *International Conference on Learning Representations*.

Yarotsky, D. (2018). Universal approximations of invariant maps by neural networks. *arXiv preprint arXiv:1804.10306*.