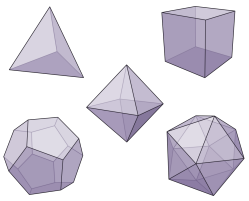
these architectures, a canonical mesh (of the body, face, or hand) was assumed to be known and the synthesis task consisted of regressing the 3D coordinates of the nodes (the embedding of the surface, using the jargon of differential geometry). Kulon et al. (2020) showed a hybrid pipeline for 3D hand pose estimation with an image CNN-based encoder and a geometric decoder. A demo of this system, developed in collaboration with a British startup company Ariel AI and presented at CVPR 2020, allowed to create realistic body avatars with fully articulated hands from video input on a mobile phone faster than real-time. Ariel AI was acquired by Snap in 2020, and at the time of writing its technology is used in Snap's augmented reality products.

Examples of complex 3D hand poses reconstructed from 2D images in the wild (Kulon et al., 2020).

7 Historic Perspective

"Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection." This somewhat poetic definition of symmetry is given in the eponymous book of the great mathematician Hermann Weyl (2015), his *Schwanengesang* on the eve of retirement from the Institute for Advanced Study in Princeton. Weyl traces the special place symmetry has occupied in science and art to the ancient times, from Sumerian symmetric designs to the Pythagoreans who believed the circle to be perfect due to its rotational symmetry. Plato considered the five regular polyhedra bearing his name today so fundamental that they must be the basic building blocks shaping the material world. Yet, though Plato is credited with coining the term συμμετρία, which literally translates as 'same measure', he used it only vaguely to convey the beauty of proportion in art and harmony in music. It was the astronomer and mathematician Johannes Kepler to attempt the first rigorous analysis of the symmetric shape of water crystals. In his treatise ('On the Six-Cornered Snowflake'), he attributed the six-fold dihedral structure of snowflakes to hexagonal packing of particles – an idea that though preceded the clear understanding of how matter is formed, still holds today as the basis of crystallography (Ball, 2011).



The tetrahedron, cube, octahedron, dodecahedron, and icosahedron are called *Platonic solids*.

Fully titled *Strena, Seu De Nive Sexangula* ('New Year's gift, or on the Six-Cornered Snowflake') was, as suggested by the title, a small booklet sent by Kepler in 1611 as a Christmas gift to his patron and friend Johannes Matthäus Wackher von Wackenfels.

Symmetry in Mathematics and Physics In modern mathematics, symmetry is almost univocally expressed in the language of group theory. The origins of this theory are usually attributed to Évariste Galois, who coined

the term and used it to study solvability of polynomial equations in the 1830s. Two other names associated with group theory are those of Sophus Lie and Felix Klein, who met and worked fruitfully together for a period of time (Tobies, 2019). The former would develop the theory of continuous symmetries that today bears his name; the latter proclaimed group theory to be the organising principle of geometry in his Erlangen Program, which we mentioned in the beginning of this text. Riemannian geometry was explicitly excluded from Klein's unified geometric picture, and it took another fifty years before it was integrated, largely thanks to the work of Élie Cartan in the 1920s.

Emmy Noether, Klein's colleague in Göttingen, proved that every differentiable symmetry of the action of a physical system has a corresponding conservation law (Noether, 1918). In physics, it was a stunning result: beforehand, meticulous experimental observation was required to discover fundamental laws such as the conservation of energy, and even then, it was an empirical result not coming from anywhere. Noether's Theorem — "a guiding star to 20th and 21st century physics", in the words of the Nobel laureate Frank Wilczek — showed that the conservation of energy emerges from the translational symmetry of time, a rather intuitive idea that the results of an experiment should not depend on whether it is conducted today or tomorrow.

The symmetry associated with charge conservation is the global *gauge invariance* of the electromagnetic field, first appearing in Maxwell's formulation of electrodynamics (Maxwell, 1865); however, its importance initially remained unnoticed. The same Hermann Weyl who wrote so dithyrambically about symmetry is the one who first introduced the concept of gauge invariance in physics in the early 20th century, emphasizing its role as a principle from which electromagnetism can be *derived*. It took several decades until this fundamental principle — in its generalised form developed by Yang and Mills (1954) — proved successful in providing a unified framework to describe the quantum-mechanical behavior of electromagnetism and the weak and strong forces, finally culminating in the Standard Model that captures all the fundamental forces of nature but gravity. We can thus join another Nobel-winning physicist, Philip Anderson (1972), in concluding that "it is only slightly overstating the case to say that physics is the study of symmetry."

Weyl first conjectured (incorrectly) in 1919 that invariance under the change of scale or "gauge" was a local symmetry of electromagnetism. The term *gauge*, or *Eich* in German, was chosen by analogy to the various track gauges of railroads. After the development of quantum mechanics, Weyl (1929) modified the gauge choice by replacing the scale factor with a change of wave phase. See Straumann (1996).

Early Use of Symmetry in Machine Learning In machine learning and its applications to pattern recognition and computer vision, the importance of symmetry has long been recognised. Early work on designing equivariant feature detectors for pattern recognition was done by Amari (1978), Kanatani (2012), and Lenz (1990). In the neural networks literature, the famous Group Invariance Theorem for Perceptrons by Minsky and Papert (2017) puts fundamental limitations on the capabilities of (single-layer) perceptrons to learn invariants. This was one of the primary motivations for studying multi-layer architectures (Sejnowski et al., 1986; Shawe-Taylor, 1989, 1993), which ultimately led to deep learning.

Shun'ichi Amari is credited as the creator of the field of *information geometry* that applies Riemannian geometry models to probability. The main object studied by information geometry is a *statistical manifold*, where each point corresponds to a probability distribution.

In the neural network community, *Neocognitron* (Fukushima and Miyake, 1982) is credited as the first implementation of shift invariance in a neural network for "pattern recognition unaffected by shift in position". His solution came in the form of hierarchical neural network with local connectivity, drawing inspiration from the receptive fields discovered in the visual cortex by the neuroscientists David Hubel and Torsten Wiesel two decades earlier (Hubel and Wiesel, 1959). These ideas culminated in Convolutional Neural Networks in the seminal work of Yann LeCun and co-authors (LeCun et al., 1998). The first work to take a representation-theoretical view on invariant and equivariant neural networks was performed by Wood and Shawe-Taylor (1996), unfortunately rarely cited. More recent incarnations of these ideas include the works of Makadia et al. (2007); Esteves et al. (2020) and one of the authors of this text (Cohen and Welling, 2016).

This classical work was recognised by the Nobel Prize in Medicine in 1981, which Hubel and Wiesel shared with Roger Sperry.

Graph Neural Networks It is difficult to pinpoint exactly when the concept of Graph Neural Networks began to emerge—partly due to the fact that most of the early work did not place graphs as a first-class citizen, partly since GNNs became practical only in the late 2010s, and partly because this field emerged from the confluence of several research areas. That being said, early forms of graph neural networks can be traced back at least to the 1990s, with examples including Alessandro Sperduti's Labeling RAAM (Sperduti, 1994), the "backpropagation through structure" of Goller and Kuchler (1996), and adaptive processing of data structures (Sperduti and Starita, 1997; Frasconi et al., 1998). While these works were primarily concerned with operating over "structures" (often trees or directed acyclic graphs), many of the invariances preserved in their architectures are reminiscent of the GNNs more commonly in use today.

The first proper treatment of the processing of generic graph structures (and the coining of the term *"graph neural network"*) happened after the turn of the 21st century. Within the Artificial Intelligence lab at the Università degli Studi di Siena (Italy), papers led by Marco Gori and Franco Scarselli have proposed the first "GNN" (Gori et al., 2005; Scarselli et al., 2008). They relied on recurrent mechanisms, required the neural network parameters to specify *contraction mappings*, and thus computing node representations by searching for a fixed point—this in itself necessitated a special form of backpropagation (Almeida, 1990; Pineda, 1988) and did not depend on node features at all. All of the above issues were rectified by the Gated GNN (GGNN) model of Li et al. (2015). GGNNs brought many benefits of modern RNNs, such as gating mechanisms (Cho et al., 2014) and backpropagation through time, to the GNN model, and remain popular today.

Concurrently, Alessio Micheli had proposed the *neural network for graphs* (NN4G) model, which focused on a *feedforward* rather than recurrent paradigm (Micheli, 2009).

Computational chemistry It is also very important to note an independent and concurrent line of development for GNNs: one that was entirely driven by the needs of computational chemistry, where molecules are most naturally expressed as graphs of atoms (nodes) connected by chemical bonds (edges). This invited computational techniques for molecular property prediction that operate directly over such a graph structure, which had become present in machine learning in the 1990s: this includes the ChemNet model of Kireev (1995) and the work of Baskin et al. (1997). Strikingly, the "molecular graph networks" of Merkwirth and Lengauer (2005) explicitly proposed many of the elements commonly found in contemporary GNNs—such as edge type-conditioned weights or global pooling—as early as 2005. The chemical motivation continued to drive GNN development into the 2010s, with two significant GNN advancements centered around improving molecular fingerprinting (Duvenaud et al., 2015) and predicting quantum-chemical properties (Gilmer et al., 2017) from small molecules. At the time of writing this text, molecular property prediction is one of the most successful applications of GNNs, with impactful results in virtual screening of new antibiotic drugs (Stokes et al., 2020).

Node embeddings Some of the earliest success stories of deep learning on graphs involve learning representations of nodes in an unsupervised fashion, based on the graph structure. Given their structural inspiration, this direction also provides one of the most direct links between graph representation learning and network science communities. The key early

approaches in this space relied on *random walk*-based embeddings: learning node representations in a way that brings them closer together if the nodes co-occur in a short random walk. Representative methods in this space include DeepWalk (Perozzi et al., 2014), node2vec (Grover and Leskovec, 2016) and LINE (Tang et al., 2015), which are all purely self-supervised. Planetoid (Yang et al., 2016) was the first in this space to incorporate supervision label information, when it is available.

Recently, a theoretical framework was developed by Srinivasan and Ribeiro (2019) in which the equivalence of structural and positional representations was demonstrated. Additionally, Qiu et al. (2018) have demonstrated that all random-walk based embedding techniques are equivalent to an appropriately-posed matrix factorisation task.

Unifying random walk objectives with GNN encoders was attempted on several occasions, with representative approaches including Variational Graph Autoencoder (VGAE, Kipf and Welling (2016b)), embedding propagation (García-Durán and Niepert, 2017), and unsupervised variants of GraphSAGE (Hamilton et al., 2017). However, this was met with mixed results, and it was shortly discovered that pushing neighbouring node representations together is already a key part of GNNs' inductive bias. Indeed, it was shown that an *untrained* GNN was already showing performance that is competitive with DeepWalk, in settings where node features are available (Veličković et al., 2019; Wu et al., 2019). This launched a direction that moves away from combining random walk objectives with GNNs and shifting towards *contrastive* approaches inspired by mutual information maximisation and aligning to successful methods in the image domain. Prominent examples of this direction include Deep Graph Informax (DGI, Veličković et al. (2019)), GRACE (Zhu et al., 2020), BERT-like objectives (Hu et al., 2020) and BGRL (Thakoor et al., 2021).

Probabilistic graphical models Graph neural networks have also, concurrently, resurged through embedding the computations of *probabilistic graphical models* (PGMs, Wainwright and Jordan (2008)). PGMs are a powerful tool for processing graphical data, and their utility arises from their probabilistic perspective on the graph's edges: namely, the nodes are treated as *random variables*, while the graph structure encodes *conditional independence* assumptions, allowing for significantly simplifying the calculation and sampling from the joint distribution. Indeed, many algorithms for (exactly or approximately) supporting learning and inference on PGMs rely on forms of passing messages over their edges (Pearl, 2014), with examples including variational mean-field inference and loopy belief propagation (Yedidia et al., 2001; Murphy et al., 2013).

This connection between PGMs and message passing was subsequently

developed into GNN architectures, with early theoretical links established by the authors of structure2vec (Dai et al., 2016). Namely, by posing a graph representation learning setting as a Markov random field (of nodes corresponding to input features and latent representations), the authors directly align the computation of both mean-field inference and loopy belief propagation to a model not unlike the GNNs commonly in use today.

The key "trick" which allowed for relating the latent representations of a GNN to probability distributions maintained by a PGM was the usage of *Hilbert-space embeddings* of distributions (Smola et al., 2007). Given ϕ, an appropriately chosen embedding function for features $\mathbf{x}$, it is possible to embed their probability distribution $p(\mathbf{x})$ as the *expected* embedding $\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\phi(\mathbf{x})$. Such a correspondence allows us to perform GNN-like computations, knowing that the representations computed by the GNN will always correspond to an embedding of *some* probability distribution over the node features.

The structure2vec model itself is, ultimately, a GNN architecture which easily sits within our framework, but its setup has inspired a series of GNN architectures which more directly incorporate computations found in PGMs. Emerging examples have successfully combined GNNs with conditional random fields (Gao et al., 2019; Spalević et al., 2020), relational Markov networks (Qu et al., 2019) and Markov logic networks (Zhang et al., 2020).

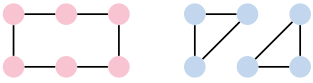
The Weisfeiler-Lehman formalism The resurgence of graph neural networks was followed closely by a drive to understand their fundamental limitations, especially in terms of expressive power. While it was becoming evident that GNNs are a strong modelling tool of graph-structured data, it was also clear that they wouldn't be able to solve *any* task specified on a graph perfectly. A canonical illustrative example of this is deciding *graph isomorphism*: is our GNN able to attach different representations to two given non-isomorphic graphs? This is a useful framework for two reasons. If the GNN is unable to do this, then it will be hopeless on any task requiring the discrimination of these two graphs. Further, it is currently not known if deciding graph isomorphism is in P, the complexity class in which all GNN computations typically reside.

Due to their permutation invariance, GNNs will attach identical representations to two isomorphic graphs, so this case is trivially solved.

The best currently known algorithm for deciding graph isomorphism is due to Babai and Luks (1983), though a recent (not fully reviewed) proposal by Babai (2016) implies a quasi-polynomial time solution.

The key framework which binds GNNs to graph isomorphism is the *Weisfeiler-Lehman* (WL) graph isomorphism test (Weisfeiler and Leman, 1968). This test generates a graph representation by iteratively passing node features

along the edges of the graph, then *randomly hashing* their sums across neighbourhoods. Connections to *randomly-initialised* convolutional GNNs are apparent, and have been observed early on: for example, within the GCN model of Kipf and Welling (2016a). Aside from this connection, the WL iteration was previously introduced in the domain of *graph kernels* by Shervashidze et al. (2011), and it still presents a strong baseline for unsupervised learning of whole-graph representations.



One simple example: the WL test cannot distinguish a *6-cycle* from *two triangles*.

While the WL test is conceptually simple, and there are many simple examples of non-isomorphic graphs it cannot distinguish, its expressive power is ultimately strongly tied to GNNs. Analyses by Morris et al. (2019) and Xu et al. (2018) have both reached a striking conclusion: *any* GNN conforming to one of the three flavours we outlined in Section 5.3 cannot be more powerful than the WL test!

Popular aggregators such as maximisation and averaging fall short in this regard, because they would not be able to distinguish e.g. the neighbour multisets $\{\!\{\mathbf{a}, \mathbf{b}\}\!\}$ and $\{\!\{\mathbf{a}, \mathbf{a}, \mathbf{b}, \mathbf{b}\}\!\}$.

In order to exactly reach this level of representational power, certain constraints must exist on the GNN update rule. Xu et al. (2018) have shown that, in the discrete-feature domain, the aggregation function the GNN uses must be *injective*, with *summation* being a key representative. Based on the outcome of their analysis, Xu et al. (2018) propose the Graph Isomorphism Network (GIN), which is a simple but powerful example of a maximally-expressive GNN under this framework. It is also expressible under the convolutional GNN flavour we propose.

One example of such aggregators are the *moments* of the multiset of neighbours.

Lastly, it is worth noting that these findings do not generalise to *continuous* node feature spaces. In fact, using the Borsuk-Ulam theorem (Borsuk, 1933), Corso et al. (2020) have demonstrated that, assuming real-valued node features, obtaining injective aggregation functions requires *multiple* aggregators (specifically, equal to the *degree* of the receiver node). Their findings have driven the Principal Neighbourhood Aggregation (PNA) architecture, which proposes a multiple-aggregator GNN that is empirically powerful and stable.

Higher-order methods The findings of the previous paragraphs do not contradict the practical utility of GNNs. Indeed, in many real-world applications the input features are sufficiently *rich* to support useful discriminative computations over the graph structure, despite of the above limitations.

Which, in contrast, almost always consider *featureless* or *categorically-featured* graphs.

However, one key corollary is that GNNs are relatively quite weak at detecting some rudimentary *structures* within a graph. Guided by the specific

limitations or failure cases of the WL test, several works have provided *stronger* variants of GNNs that are *provably* more powerful than the WL test, and hence likely to be useful on tasks that require such structural detection.

One prominent example is computational chemistry, wherein a molecule's chemical function can be strongly influenced by the presence of aromatic *rings* in its molecular graph.

Perhaps the most direct place to hunt for more expressive GNNs is the WL test itself. Indeed, the strength of the original WL test can be enhanced by considering a *hierarchy* of WL tests, such that k-WL tests attach representations to k-*tuples* of nodes (Morris et al., 2017). The k-WL test has been directly translated into a *higher-order* k-GNN architecture by Morris et al. (2019), which is provably more powerful than the GNN flavours we considered before. However, its requirement to maintain tuple representations implies that, in practice, it is hard to scale beyond $k = 3$.

There have been efforts, such as the δ-k-LGNN (Morris et al., 2020), to sparsify the computation of the k-GNN.

Concurrently, Maron et al. (2018, 2019) have studied the characterisation of invariant and equivariant graph networks over k-tuples of nodes. Besides demonstrating the surprising result of *any* invariant or equivariant graph network being expressible as a linear combination of a finite number of generators—the amount of which only depends on k—the authors showed that the expressive power of such layers is equivalent to the k-WL test, and proposed an empirically scalable variant which is provably 3-WL powerful.

Besides generalising the domain over which representations are computed, significant effort had also went into analysing specific failure cases of 1-WL and augmenting GNN *inputs* to help them distinguish such cases. One common example is attaching *identifying features* to the nodes, which can help detecting structure. Proposals to do this include *one-hot* representations (Murphy et al., 2019), as well as purely *random* features (Sato et al., 2020).

For example, if a node sees its own identifier k hops away, it is a direct indicator that it is within a k-cycle.

More broadly, there have been many efforts to incorporate *structural* information within the message passing process, either by modulating the message function or the graph that the computations are carried over. Several interesting lines of work here involve sampling *anchor node sets* (You et al., 2019), aggregating based on *Laplacian eigenvectors* (Stachenfeld et al., 2020; Beaini et al., 2020; Dwivedi and Bresson, 2020), or performing *topological data analysis*, either for positional embeddings (Bouritsas et al., 2020) or driving message passing (Bodnar et al., 2021).

In the computational chemistry domain, it is often assumed that molecular function is driven by substructures (the *functional groups*), which have directly inspired the modelling of molecules at a *motif* level. For references, consider Jin et al. (2018, 2020); Fey et al. (2020).

Signal processing and Harmonic analysis Since the early successes of Convolutional Neural Networks, researchers have resorted to tools from harmonic analysis, image processing, and computational neuroscience trying to

provide a theoretical framework that explains their efficiency. M-theory is a framework inspired by the visual cortex, pioneered by Tomaso Poggio and collaborators (Riesenhuber and Poggio, 1999; Serre et al., 2007), based on the notion of templates that can be manipulated under certain symmetry groups. Another notable model arising from computational neuroscience were *steerable pyramids*, a form of multiscale wavelet decompositions with favorable properties against certain input transformations, developed by Simoncelli and Freeman (1995). They were a central element in early generative models for textures (Portilla and Simoncelli, 2000), which were subsequently improved by replacing steerable wavelet features with deep CNN features Gatys et al. (2015). Finally, Scattering transforms, introduced by Stéphane Mallat (2012) and developed by Bruna and Mallat (2013), provided a framework to understand CNNs by replacing trainable filters with multiscale wavelet decompositions, also showcasing the deformation stability and the role of depth in the architecture.

Signal Processing on Graph and Meshes Another important class of graph neural networks, often referred to as *spectral*, has emerged from the work of one of the authors of this text (Bruna et al., 2013), using the notion of the *Graph Fourier transform*. The roots of this construction are in the signal processing and computational harmonic analysis communities, where dealing with non-Euclidean signals has become prominent in the late 2000s and early 2010s. Influential papers from the groups of Pierre Vandergheynst (Shuman et al., 2013) and José Moura (Sandryhaila and Moura, 2013) popularised the notion of "Graph Signal Processing" (GSP) and the generalisation of Fourier transforms based on the eigenvectors of graph adjacency and Laplacian matrices. The graph convolutional neural networks relying on spectral filters by Defferrard et al. (2016) and Kipf and Welling (2016a) are among the most cited in the field and can likely be credited) as ones reigniting the interest in machine learning on graphs in recent years.

It is worth noting that, in the field of computer graphics and geometry processing, non-Euclidean harmonic analysis predates Graph Signal Processing by at least a decade. We can trace spectral filters on manifolds and meshes to the works of Taubin et al. (1996). These methods became mainstream in the 2000s following the influential papers of Karni and Gotsman (2000) on spectral geometry compression and of Lévy (2006) on using the Laplacian eigenvectors as a non-Euclidean Fourier basis. Spectral methods have been used for a range of applications, most prominent of which is the construction

Learnable shape descriptors similar to spectral graph CNNs were proposed by Roee Litman and Alex Bronstein (2013), the latter being a twin brother of the author of this text.

of shape descriptors (Sun et al., 2009) and functional maps (Ovsjanikov et al., 2012); these methods are still broadly used in computer graphics at the time of writing.

Computer Graphics and Geometry Processing Models for shape analysis based on intrinsic metric invariants were introduced by various authors in the field of computer graphics and geometry processing (Elad and Kimmel, 2003; Mémoli and Sapiro, 2005; Bronstein et al., 2006), and are discussed in depth by one of the authors in his earlier book (Bronstein et al., 2008). The notions of intrinsic symmetries were also explored in the same field Raviv et al. (2007); Ovsjanikov et al. (2008). The first architecture for deep learning on meshes, Geodesic CNNs, was developed in the team of one of the authors of the text (Masci et al., 2015). This model used local filters with shared weights, applied to geodesic radial patches. It was a particular setting of gauge-equivariant CNNs developed later by another author of the text (Cohen et al., 2019). A generalisation of Geodesic CNNs with learnable aggregation operations, MoNet, proposed by Federico Monti et al. (2017) from the same team, used an attention-like mechanism over the local structural features of the mesh, that was demonstrated to work on general graphs as well. The graph attention network (GAT), which technically speaking can be considered a particular instance of MoNet, was introduced by another author of this text (Veličković et al., 2018). GATs generalise MoNet's attention mechanism to also incorporate node feature information, breaking away from the purely structure-derived relevance of prior work. It is one of the most popular GNN architectures currently in use.

In the context of computer graphics, it is also worthwhile to mention that the idea of learning on sets (Zaheer et al., 2017) was concurrently developed in the group of Leo Guibas at Stanford under the name PointNet (Qi et al., 2017) for the analysis of 3D point clouds. This architecture has lead to multiple follow-up works, including one by an author of this text called Dynamic Graph CNN (DGCNN, Wang et al. (2019b)). DGCNN used a nearest-neighbour graph to capture the local structure of the point cloud to allow exchange of information across the nodes; the key characteristic of this architecture was that the graph was constructed on-the-fly and updated between the layers of the neural network in relation to the downstream task. This latter property made DGCNN one of the first incarnations of 'latent graph learning', which in its turn has had significant follow up. Extensions to DGCNN's k-nearest neighbour graph proposal include more explicit control

over these graphs' edges, either through bilevel optimisation (Franceschi et al., 2019), reinforcement learning (Kazi et al., 2020) or direct supervision (Veličković et al., 2020). Independently, a variational direction (which probabilistically samples edges from a computed *posterior* distribution) has emerged through the NRI model (Kipf et al., 2018). While it still relies on quadratic computation in the number of nodes, it allows for explicitly encoding uncertainty about the chosen edges.

Another very popular direction in learning on graphs without a provided graph relies on performing GNN-style computations over a *complete* graph, letting the network infer its own way to exploit the connectivity. The need for this arisen particularly in natural language processing, where various words in a sentence interact in highly nontrivial and non-sequential ways. Operating over a complete graph of words brought about the first incarnation of the Transformer model (Vaswani et al., 2017), which de-throned both recurrent and convolutional models as state-of-the-art in neural machine translation, and kicked off an avalanche of related work, transcending the boundaries between NLP and other fields. Fully-connected GNN computation has also concurrently emerged on simulation (Battaglia et al., 2016), reasoning (Santoro et al., 2017), and multi-agent (Hoshen, 2017) applications, and still represents a popular choice when the number of nodes is reasonably small.

Algorithmic reasoning For most of the discussion we posed in this section, we have given examples of *spatially* induced geometries, which in turn shape the underlying domain, and its invariances and symmetries. However, plentiful examples of invariances and symmetries also arise in a *computational* setting. One critical difference to many common settings of Geometric Deep Learning is that links no longer need to encode for any kind of similarity, proximity, or types of relations—they merely specify the "recipe" for the dataflow between data points they connect.

Instead, the computations of the neural network mimic the reasoning process of an *algorithm* (Cormen et al., 2009), with additional invariances induced by the algorithm's control flow and intermediate results. In the space of algorithms the assumed input invariants are often referred to as *preconditions*, while the invariants preserved by the algorithm are known as *postconditions*.

For example, one invariant of the Bellman-Ford pathfinding algorithm (Bellman, 1958) is that, after k steps, it will always compute the shortest paths to the source node that use no more than k edges.

Eponymously, the research direction of *algorithmic reasoning* (Cappart et al., 2021, Section 3.3.) seeks to produce neural network architectures that ap-

propriately preserve algorithmic invariants. The area has investigated the construction of general-purpose neural computers, e.g., the *neural Turing machine* (Graves et al., 2014) and the *differentiable neural computer* (Graves et al., 2016). While such architectures have all the hallmarks of general computation, they introduced several components at once, making them often challenging to optimise, and in practice, they are almost always outperformed by simple relational reasoners, such as the ones proposed by Santoro et al. (2017, 2018).

As modelling complex postconditions is challenging, plentiful work on inductive biases for learning to execute (Zaremba and Sutskever, 2014) has focused on primitive algorithms (e.g. simple arithmetic). Prominent examples in this space include the *neural GPU* (Kaiser and Sutskever, 2015), *neural RAM* (Kurach et al., 2015), *neural programmer-interpreters* (Reed and De Freitas, 2015), *neural arithmetic-logic units* (Trask et al., 2018; Madsen and Johansen, 2020) and *neural execution engines* (Yan et al., 2020).

Emulating combinatorial algorithms of *superlinear* complexity was made possible with the rapid development of GNN architectures. The *algorithmic alignment* framework pioneered by Xu et al. (2019) demonstrated, theoretically, that GNNs *align* with dynamic programming (Bellman, 1966), which is a language in which most algorithms can be expressed. It was concurrently empirically shown, by one of the authors of this text, that it is possible to design and train GNNs that align with algorithmic invariants in practice (Veličković et al., 2019). Onwards, alignment was achieved with *iterative algorithms* (Tang et al., 2020), *linearithmic algorithms* (Freivalds et al., 2019), *data structures* (Veličković et al., 2020) and *persistent memory* (Strathmann et al., 2021). Such models have also seen practical use in *implicit planners* (Deac et al., 2020), breaking into the space of *reinforcement learning* algorithms.

Concurrently, significant progress has been made on using GNNs for *physics simulations* (Sanchez-Gonzalez et al., 2020; Pfaff et al., 2020). This direction yielded much of the same recommendations for the design of generalising GNNs. Such a correspondence is to be expected: given that algorithms can be phrased as discrete-time simulations, and simulations are typically implemented as step-wise algorithms, both directions will need to preserve similar kinds of invariants.

Tightly bound with the study of algorithmic reasoning are measures of *extrapolation*. This is a notorious pain-point for neural networks, given that most of their success stories are obtained when generalising *in-distribution*;

i.e. when the patterns found in the training data properly anticipate the ones found in the test data. However, algorithmic invariants must be preserved irrespective of, e.g., the size or generative distribution of the input, meaning that the training set will likely not cover any possible scenario encountered in practice. Xu et al. (2020b) have proposed a geometric argument for what is required of an extrapolating GNN backed by rectifier activations: its components and featurisation would need to be designed so as to make its constituent modules (e.g. message function) learn only *linear* target functions. Bevilacqua et al. (2021) propose observing extrapolation under the lens of *causal reasoning*, yielding *environment-invariant* representations of graphs.

Geometric Deep Learning Our final historical remarks regard the very name of this text. The term 'Geometric Deep Learning' was first introduced by one of the authors of this text in his ERC grant in 2015 and popularised in the eponymous IEEE Signal Processing Magazine paper (Bronstein et al., 2017). This paper proclaimed, albeit "with some caution", the signs of "a new field being born." Given the recent popularity of graph neural networks, the increasing use of ideas of invariance and equivariance in a broad range of machine learning applications, and the very fact of us writing this text, it is probably right to consider this prophecy at least partially fulfilled. The name "4G: Grids, Graphs, Groups, and Gauges" was coined by Max Welling for the ELLIS Program on Geometric Deep Learning, co-directed by two authors of the text. Admittedly, the last 'G' is somewhat of a stretch, since the underlying structures are manifolds and bundles rather than gauges. For this text, we added another 'G', Geodesics, in reference to metric invariants and intrinsic symmetries of manifolds.

Acknowledgements

This text represents a humble attempt to summarise and synthesise decades of existing knowledge in deep learning architectures, through the geometric lens of invariance and symmetry. We hope that our perspective will make it easier both for newcomers and practitioners to navigate the field, and for researchers to synthesise novel architectures, as instances of our blueprint. In a way, we hope to have presented *"all you need to build the architectures that are all you need"*—a play on words inspired by Vaswani et al. (2017).

The bulk of the text was written during late 2020 and early 2021. As it often happens, we had thousands of doubts whether the whole picture makes sense, and used opportunities provided by our colleagues to help us break our "stage fright" and present early versions of our work, which saw the light of day in Petar's talk at Cambridge (courtesy of Pietro Liò) and Michael's talks at Oxford (courtesy of Xiaowen Dong) and Imperial College (hosted by Michael Huth and Daniel Rueckert). Petar was also able to present our work at Friedrich-Alexander-Universität Erlangen-Nürnberg—the birthplace of the Erlangen Program!—owing to a kind invitation from Andreas Maier. The feedback we received for these talks was enormously invaluable to keeping our spirits high, as well as polishing the work further. Last, but certainly not least, we thank the organising committee of ICLR 2021, where our work will be featured in a keynote talk, delivered by Michael.

We should note that reconciling such a vast quantity of research is seldom enabled by the expertise of only four people. Accordingly, we would like to give due credit to all of the researchers who have carefully studied aspects of our text as it evolved, and provided us with careful comments and references: Yoshua Bengio, Charles Blundell, Andreea Deac, Fabian Fuchs, Francesco di Giovanni, Marco Gori, Raia Hadsell, Will Hamilton, Maksym Korablyov, Christian Merkwirth, Razvan Pascanu, Bruno Ribeiro, Anna Scaife, Jürgen Schmidhuber, Marwin Segler, Corentin Tallec, Ngân Vũ, Peter Wirnsberger and David Wong. Their expert feedback was invaluable to solidifying our unification efforts and making them more useful to various niches. Though, of course, any irregularities within this text are our responsibility alone. It is currently very much a work-in-progress, and we are very happy to receive comments at any stage. Please contact us if you spot any errors or omissions.

Bibliography

Yonathan Aflalo and Ron Kimmel. Spectral multidimensional scaling. *PNAS*, 110(45):18052–18057, 2013.

Yonathan Aflalo, Haim Brezis, and Ron Kimmel. On the optimality of shape and data representation in the spectral domain. *SIAM J. Imaging Sciences*, 8(2):1141–1160, 2015.

Luis B Almeida. A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. In *Artificial neural networks: concept learning*, pages 102–111. 1990.

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. *arXiv:2006.05205*, 2020.

Sl Amari. Feature spaces which admit and detect invariant signal transformations. In *Joint Conference on Pattern Recognition*, 1978.

Brandon Anderson, Truong-Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. *arXiv:1906.04015*, 2019.

Philip W Anderson. More is different. *Science*, 177(4047):393–396, 1972.

Mathieu Andreux, Emanuele Rodola, Mathieu Aubry, and Daniel Cremers. Anisotropic Laplace-Beltrami operators for shape analysis. In *ECCV*, 2014.

Salim Arslan, Sofia Ira Ktena, Ben Glocker, and Daniel Rueckert. Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity. In *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities*, pages 3–13. 2018.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv:1607.06450*, 2016.

László Babai. Graph isomorphism in quasipolynomial time. In *ACM Symposium on Theory of Computing*, 2016.

László Babai and Eugene M Luks. Canonical labeling of graphs. In *ACM Symposium on Theory of computing*, 1983.

Francis Bach. Breaking the curse of dimensionality with convex neural networks. *JMLR*, 18(1):629–681, 2017.

Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. Agent57: Outperforming the atari human benchmark. In *ICML*, 2020.

Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *Trans. PAMI*, 39(12):2481–2495, 2017.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv:1409.0473*, 2014.

Philip Ball. In retrospect: On the six-cornered snowflake. *Nature*, 480(7378): 455–455, 2011.

Bassam Bamieh. Discovering transforms: A tutorial on circulant matrices, circular convolution, and the discrete fourier transform. *arXiv:1805.05533*, 2018.

Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922.

Victor Bapst, Thomas Keck, A Grabska-Barwińska, Craig Donner, Ekin Dogus Cubuk, Samuel S Schoenholz, Annette Obika, Alexander WR Nelson, Trevor Back, Demis Hassabis, et al. Unveiling the predictive power of static structure in glassy systems. *Nature Physics*, 16(4):448–454, 2020.

Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011.

Andrew R Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Information Theory*, 39(3):930–945, 1993.

Igor I Baskin, Vladimir A Palyulin, and Nikolai S Zefirov. A neural device for searching direct correlations between structures and properties of chemical compounds. *J. Chemical Information and Computer Sciences*, 37(4): 715–721, 1997.

Peter W Battaglia, Razvan Pascanu, Matthew Lai, Danilo Rezende, and Koray Kavukcuoglu. Interaction networks for learning about objects, relations and physics. *arXiv:1612.00222*, 2016.

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv:1806.01261*, 2018.

Dominique Beaini, Saro Passaro, Vincent Létourneau, William L Hamilton, Gabriele Corso, and Pietro Liò. Directional graph networks. *arXiv:2010.02863*, 2020.

Richard Bellman. On a routing problem. *Quarterly of Applied Mathematics*, 16 (1):87–90, 1958.

Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks*, 5(2):157–166, 1994.

Marcel Berger. *A panoramic view of Riemannian geometry*. Springer, 2012.

Pierre Besson, Todd Parrish, Aggelos K Katsaggelos, and S Kathleen Bandt. Geometric deep learning on brain shape predicts sex and age. *BioRxiv:177543*, 2020.

Beatrice Bevilacqua, Yangze Zhou, and Bruno Ribeiro. Size-invariant graph representations for graph classification extrapolations. *arXiv:2103.05045*, 2021.

Guy Blanc, Neha Gupta, Gregory Valiant, and Paul Valiant. Implicit regularization for deep neural networks driven by an ornstein-uhlenbeck like process. In *COLT*, 2020.

Cristian Bodnar, Fabrizio Frasca, Yu Guang Wang, Nina Otter, Guido Montúfar, Pietro Liò, and Michael Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks. *arXiv:2103.03212*, 2021.

Alexander Bogatskiy, Brandon Anderson, Jan Offermann, Marwah Roussi, David Miller, and Risi Kondor. Lorentz group equivariant neural network for particle physics. In *ICML*, 2020.

Karol Borsuk. Drei sätze über die n-dimensionale euklidische sphäre. *Fundamenta Mathematicae*, 20(1):177–190, 1933.

Davide Boscaini, Davide Eynard, Drosos Kourounis, and Michael M Bronstein. Shape-from-operator: Recovering shapes from intrinsic operators. *Computer Graphics Forum*, 34(2):265–274, 2015.

Davide Boscaini, Jonathan Masci, Emanuele Rodoià, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *NIPS*, 2016a.

Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Michael M Bronstein, and Daniel Cremers. Anisotropic diffusion descriptors. *Computer Graphics Forum*, 35(2):431–441, 2016b.

Sébastien Bougleux, Luc Brun, Vincenzo Carletti, Pasquale Foggia, Benoit Gaüzere, and Mario Vento. A quadratic assignment formulation of the graph edit distance. *arXiv:1512.07494*, 2015.

Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *arXiv:2006.09252*, 2020.

Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *PNAS*, 103(5):1168–1172, 2006.

Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. Springer, 2008.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv:2005.14165*, 2020.

Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *ICLR*, 2013.

Quentin Cappart, Didier Chételat, Elias Khalil, Andrea Lodi, Christopher Morris, and Petar Veličković. Combinatorial optimization and reasoning with graph neural networks. *arXiv:2102.09544*, 2021.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *arXiv:1806.07366*, 2018.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.

Albert Chern, Felix Knöppel, Ulrich Pinkall, and Peter Schröder. Shape from metric. *ACM Trans. Graphics*, 37(4):1–17, 2018.

Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv:1406.1078*, 2014.

Nicholas Choma, Federico Monti, Lisa Gerhardt, Tomasz Palczewski, Zahra Ronaghi, Prabhat Prabhat, Wahid Bhimji, Michael M Bronstein, Spencer R Klein, and Joan Bruna. Graph neural networks for icecube signal classification. In *ICMLA*, 2018.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *ICML*, 2016.

Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In *ICML*, 2019.

Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. *arXiv:1801.10130*, 2018.

Tim Cooijmans, Nicolas Ballas, César Laurent, Çağlar Gülçehre, and Aaron Courville. Recurrent batch normalization. *arXiv:1603.09025*, 2016.

Etienne Corman, Justin Solomon, Mirela Ben-Chen, Leonidas Guibas, and Maks Ovsjanikov. Functional characterization of intrinsic and extrinsic geometry. *ACM Trans. Graphics*, 36(2):1–17, 2017.

Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.

Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal neighbourhood aggregation for graph nets. *arXiv:2004.05718*, 2020.

Luca Cosmo, Anees Kazi, Seyed-Ahmad Ahmadi, Nassir Navab, and Michael Bronstein. Latent-graph learning for disease prediction. In *MICCAI*, 2020.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. *arXiv:2003.04630*, 2020.

Miles D Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning symbolic physics with graph networks. *arXiv:1909.05862*, 2019.

Guillem Cucurull, Konrad Wagstyl, Arantxa Casanova, Petar Veličković, Estrid Jakobsen, Michal Drozdzal, Adriana Romero, Alan Evans, and Yoshua Bengio. Convolutional neural networks for mesh-based parcellation of the cerebral cortex. 2018.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

Hanjun Dai, Bo Dai, and Le Song. Discriminative embeddings of latent variable models for structured data. In *ICML*, 2016.

Jeffrey De Fauw, Joseph R Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, Xavier Glorot, Brendan O'Donoghue, Daniel Visentin, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24 (9):1342–1350, 2018.

Pim de Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs. In *NeurIPS*, 2020.

Andreea Deac, Petar Veličković, and Pietro Sormanni. Attentive cross-modal paratope prediction. *Journal of Computational Biology*, 26(6):536–545, 2019.

Andreea Deac, Petar Veličković, Ognjen Milinković, Pierre-Luc Bacon, Jian Tang, and Mladen Nikolić. Xlvin: executed latent value iteration nets. *arXiv:2010.13146*, 2020.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *NIPS*, 2016.

Austin Derrow-Pinion, Jennifer She, David Wong, Oliver Lange, Todd Hester, Luis Perez, Marc Nunkesser, Seongjae Lee, Xueying Guo, Peter W Battaglia, Vishal Gupta, Ang Li, Zhongwen Xu, Alvaro Sanchez-Gonzalez, Yujia Li, and Petar Veličković. Traffic Prediction with Graph Neural Networks in Google Maps. 2021.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*, 2018.

David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. *NIPS*, 2015.

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *arXiv:2012.09699*, 2020.

Asi Elad and Ron Kimmel. On bending invariant signatures for surfaces. *Trans. PAMI*, 25(10):1285–1295, 2003.

Jeffrey L Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

Carlos Esteves, Ameesh Makadia, and Kostas Daniilidis. Spin-weighted spherical CNNs. *arXiv:2006.10731*, 2020.

Xiaomin Fang, Jizhou Huang, Fan Wang, Lingke Zeng, Haijin Liang, and Haifeng Wang. ConSTGAT: Contextual spatial-temporal graph attention network for travel time estimation at baidu maps. In *KDD*, 2020.

Matthias Fey, Jan-Gin Yuen, and Frank Weichert. Hierarchical inter-message passing for learning on molecular graphs. *arXiv:2006.12179*, 2020.

Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *ICML*, 2020.

Jon Folkman. Regular line-symmetric graphs. *Journal of Combinatorial Theory*, 3(3):215–232, 1967.

Luca Franceschi, Mathias Niepert, Massimiliano Pontil, and Xiao He. Learning discrete structures for graph neural networks. In *ICML*, 2019.

Paolo Frasconi, Marco Gori, and Alessandro Sperduti. A general framework for adaptive processing of data structures. *IEEE Trans. Neural Networks*, 9 (5):768–786, 1998.

Kārlis Freivalds, Emīls Ozoliņš, and Agris Šostaks. Neural shuffle-exchange networks–sequence processing in o (n log n) time. *arXiv:1907.07897*, 2019.

Fabian B Fuchs, Daniel E Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. *arXiv:2006.10503*, 2020.

Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets*, pages 267–285. Springer, 1982.

Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, D Boscaini, MM Bronstein, and BE Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020.

Fernando Gama, Alejandro Ribeiro, and Joan Bruna. Diffusion scattering transforms on graphs. In *ICLR*, 2019.

Fernando Gama, Joan Bruna, and Alejandro Ribeiro. Stability properties of graph neural networks. *IEEE Trans. Signal Processing*, 68:5680–5695, 2020.

Hongchang Gao, Jian Pei, and Heng Huang. Conditional random field enhanced graph convolutional neural networks. In *KDD*, 2019.

Alberto García-Durán and Mathias Niepert. Learning graph representations with embedding propagation. *arXiv:1710.03059*, 2017.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *arXiv preprint arXiv:1505.07376*, 2015.

Thomas Gaudelet, Ben Day, Arian R Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy BR Hayter, Richard Vickers, Charles Roberts, Jian Tang, et al. Utilising graph machine learning within drug discovery and development. *arXiv:2012.05716*, 2020.

Felix A Gers and Jürgen Schmidhuber. Recurrent nets that time and count. In *IJCNN*, 2000.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. *arXiv:1704.01212*, 2017.

Ross Girshick. Fast R-CNN. In *CVPR*, 2015.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

Vladimir Gligorijevic, P Douglas Renfrew, Tomasz Kosciolek, Julia Koehler Leman, Daniel Berenberg, Tommi Vatanen, Chris Chandler, Bryn C Taylor, Ian M Fisk, Hera Vlamakis, et al. Structure-based function prediction using graph convolutional networks. *bioRxiv:786236*, 2020.

Christoph Goller and Andreas Kuchler. Learning task-dependent distributed representations by backpropagation through structure. In *ICNN*, 1996.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv:1406.2661*, 2014.

Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *IJCNN*, 2005.

Alex Graves. Generating sequences with recurrent neural networks. *arXiv:1308.0850*, 2013.

Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *arXiv:1410.5401*, 2014.

Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538(7626):471–476, 2016.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv:2006.07733*, 2020.

Mikhael Gromov. *Structures métriques pour les variétés riemanniennes*. Cedic, 1981.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *KDD*, 2016.

Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit regularization in matrix factorization. In *NIPS*, 2017.

Deisy Morselli Gysi, Ítalo Do Valle, Marinka Zitnik, Asher Ameli, Xiao Gan, Onur Varol, Helia Sanchez, Rebecca Marlene Baron, Dina Ghiassian, Joseph Loscalzo, et al. Network medicine framework for identifying drug repurposing opportunities for COVID-19. *arXiv:2004.07229*, 2020.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NIPS*, 2017.

Junheng Hao, Tong Zhao, Jin Li, Xin Luna Dong, Christos Faloutsos, Yizhou Sun, and Wei Wang. P-companion: A principled framework for diversified complementary product recommendation. In *Information & Knowledge Management*, 2020.

Moritz Hardt and Tengyu Ma. Identity matters in deep learning. *arXiv:1611.04231*, 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *CVPR*, 2017.

Claude Adrien Helvétius. *De l'esprit*. Durand, 1759.

R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. In *ICLR*, 2019.

Sepp Hochreiter. *Untersuchungen zu dynamischen neuronalen Netzen*. PhD thesis, Technische Universität München, 1991.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

Yedid Hoshen. Vain: Attentional multi-agent predictive modeling. *arXiv:1706.06122*, 2017.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *ICLR*, 2020.

David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J. Physiology*, 148(3):574–591, 1959.

Michael Hutchinson, Charline Le Lan, Sheheryar Zaidi, Emilien Dupont, Yee Whye Teh, and Hyunjik Kim. LieTransformer: Equivariant self-attention for Lie groups. *arXiv:2012.10885*, 2020.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

Haris Iqbal. Harisiqbal88/plotneuralnet v1.0.0, December 2018. URL https://doi.org/10.5281/zenodo.2526396.

Sarah Itani and Dorina Thanou. Combining anatomical and functional networks for neuropathology identification: A case study on autism spectrum disorder. *Medical Image Analysis*, 69:101986, 2021.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *ICML*, 2018.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *ICML*, 2020.

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):1–9, 2016.

Michael I Jordan. Serial order: A parallel distributed processing approach. In *Advances in Psychology*, volume 121, pages 471–495. 1997.

Chaitanya Joshi. Transformers are graph neural networks. *The Gradient*, 2020.

Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. An empirical exploration of recurrent network architectures. In *ICML*, 2015.

Łukasz Kaiser and Ilya Sutskever. Neural GPUs learn algorithms. *arXiv:1511.08228*, 2015.

Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. Neural machine translation in linear time. *arXiv:1610.10099*, 2016.

Nal Kalchbrenner, Erich Elsen, Karen Simonyan, Seb Noury, Norman Casagrande, Edward Lockhart, Florian Stimberg, Aaron van den Oord, Sander Dieleman, and Koray Kavukcuoglu. Efficient neural audio synthesis. In *ICML*, 2018.

Ken-Ichi Kanatani. *Group-theoretical methods in image understanding*. Springer, 2012.

Zachi Karni and Craig Gotsman. Spectral compression of mesh geometry. In *Proc. Computer Graphics and Interactive Techniques*, 2000.

Anees Kazi, Luca Cosmo, Nassir Navab, and Michael Bronstein. Differentiable graph module (DGM) graph convolutional networks. *arXiv:2002.04999*, 2020.

Henry Kenlay, Dorina Thanou, and Xiaowen Dong. Interpretable stability bounds for spectral graph filters. *arXiv:2102.09587*, 2021.

Ron Kimmel and James A Sethian. Computing geodesic paths on manifolds. *PNAS*, 95(15):8431–8435, 1998.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv:1312.6114*, 2013.

Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. Neural relational inference for interacting systems. In *ICML*, 2018.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv:1609.02907*, 2016a.

Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv:1611.07308*, 2016b.

Dmitry B Kireev. Chemnet: a novel neural network based method for graph/property mapping. *J. Chemical Information and Computer Sciences*, 35(2):175–180, 1995.

Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv:2003.03123*, 2020.

Iasonas Kokkinos, Michael M Bronstein, Roee Litman, and Alex M Bronstein. Intrinsic shape context descriptors for deformable shapes. In *CVPR*, 2012.

Patrick T Komiske, Eric M Metodiev, and Jesse Thaler. Energy flow networks: deep sets for particle jets. *Journal of High Energy Physics*, 2019(1):121, 2019.

Ilya Kostrikov, Zhongshi Jiang, Daniele Panozzo, Denis Zorin, and Joan Bruna. Surface networks. In *CVPR*, 2018.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

Sofia Ira Ktena, Sarah Parisot, Enzo Ferrante, Martin Rajchl, Matthew Lee, Ben Glocker, and Daniel Rueckert. Distance metric learning using graph convolutional networks: Application to functional brain networks. In *MICCAI*, 2017.

Dominik Kulon, Riza Alp Guler, Iasonas Kokkinos, Michael M Bronstein, and Stefanos Zafeiriou. Weakly-supervised mesh-convolutional hand reconstruction in the wild. In *CVPR*, 2020.

Karol Kurach, Marcin Andrychowicz, and Ilya Sutskever. Neural random-access machines. *arXiv:1511.06392*, 2015.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998.

Reiner Lenz. *Group theoretical methods in image processing*. Springer, 1990.

Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

Ron Levie, Federico Monti, Xavier Bresson, and Michael M Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Trans. Signal Processing*, 67(1):97–109, 2018.

Ron Levie, Elvin Isufi, and Gitta Kutyniok. On the transferability of spectral graph filters. In *Sampling Theory and Applications*, 2019.

Bruno Lévy. Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry. In *Proc. Shape Modeling and Applications*, 2006.

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *arXiv:1511.05493*, 2015.

Or Litany, Alex Bronstein, Michael Bronstein, and Ameesh Makadia. Deformable shape completion with graph convolutional autoencoders. In *CVPR*, 2018.

Roee Litman and Alexander M Bronstein. Learning spectral descriptors for deformable shape correspondence. *Trans. PAMI*, 36(1):171–180, 2013.

Hsueh-Ti Derek Liu, Alec Jacobson, and Keenan Crane. A Dirac operator for extrinsic shape analysis. *Computer Graphics Forum*, 36(5):139–149, 2017.

Siwei Lyu and Eero P Simoncelli. Nonlinear image representation using divisive normalization. In *CVPR*, 2008.

Richard H MacNeal. *The solution of partial differential equations by means of electrical networks*. PhD thesis, California Institute of Technology, 1949.

Andreas Madsen and Alexander Rosenberg Johansen. Neural arithmetic units. *arXiv:2001.05016*, 2020.

Soha Sadat Mahdi, Nele Nauwelaers, Philip Joris, Giorgos Bouritsas, Shunwang Gong, Sergiy Bokhnyak, Susan Walsh, Mark Shriver, Michael Bronstein, and Peter Claes. 3d facial matching by spiral convolutional metric learning and a biometric fusion-net of demographic properties. *arXiv:2009.04746*, 2020.

VE Maiorov. On best approximation by ridge functions. *Journal of Approximation Theory*, 99(1):68–94, 1999.

Ameesh Makadia, Christopher Geyer, and Kostas Daniilidis. Correspondence-free structure from motion. *IJCV*, 75(3):311–327, 2007.

Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

Brandon Malone, Alberto Garcia-Duran, and Mathias Niepert. Learning representations of missing data for predicting patient outcomes. *arXiv:1811.04752*, 2018.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. *arXiv:1812.09902*, 2018.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. *arXiv:1905.11136*, 2019.

Jean-Pierre Marquis. Category theory and klein's erlangen program. In *From a Geometrical Point of View*, pages 9–40. Springer, 2009.

Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on Riemannian manifolds. In *CVPR Workshops*, 2015.

James Clerk Maxwell. A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, (155):459–512, 1865.

Jason D McEwen, Christopher GR Wallis, and Augustine N Mavor-Parker. Scattering networks on the sphere for scalable and rotationally equivariant spherical cnns. *arXiv:2102.02828*, 2021.

Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Learning with invariances in random features and kernel models. *arXiv:2102.13219*, 2021.

Simone Melzi, Riccardo Spezialetti, Federico Tombari, Michael M Bronstein, Luigi Di Stefano, and Emanuele Rodolà. Gframes: Gradient-based local reference frame for 3d shape matching. In *CVPR*, 2019.

Facundo Mémoli and Guillermo Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5(3):313–347, 2005.

Christian Merkwirth and Thomas Lengauer. Automatic generation of complementary descriptors with molecular graph networks. *J. Chemical Information and Modeling*, 45(5):1159–1168, 2005.

Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and Mathematics III*, pages 35–57. 2003.

Alessio Micheli. Neural network for graphs: A contextual constructive approach. *IEEE Trans. Neural Networks*, 20(3):498–511, 2009.

Karla L Miller, Fidel Alfaro-Almagro, Neal K Bangerter, David L Thomas, Essa Yacoub, Junqian Xu, Andreas J Bartsch, Saad Jbabdi, Stamatios N Sotiropoulos, Jesper LR Andersson, et al. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience*, 19(11):1523–1536, 2016.

Marvin Minsky and Seymour A Papert. *Perceptrons: An introduction to computational geometry*. MIT Press, 2017.

Jovana Mitrovic, Brian McWilliams, Jacob Walker, Lars Buesing, and Charles Blundell. Representation learning via invariant causal mechanisms. *arXiv:2010.07922*, 2020.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *ICML*, 2016.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *CVPR*, 2017.

Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. Fake news detection on social media using geometric deep learning. *arXiv:1902.06673*, 2019.

Christopher Morris, Kristian Kersting, and Petra Mutzel. Glocalized Weisfeiler-Lehman graph kernels: Global-local feature maps of graphs. In *ICDM*, 2017.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *AAAI*, 2019.

Christopher Morris, Gaurav Rattan, and Petra Mutzel. Weisfeiler and Leman go sparse: Towards scalable higher-order graph embeddings. In *NeurIPS*, 2020.

Michael C Mozer. A focused back-propagation algorithm for temporal pattern recognition. *Complex Systems*, 3(4):349–381, 1989.

Kevin Murphy, Yair Weiss, and Michael I Jordan. Loopy belief propagation for approximate inference: An empirical study. *arXiv:1301.6725*, 2013.

Ryan Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Relational pooling for graph representations. In *ICML*, 2019.

Ryan L Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. *arXiv:1811.01900*, 2018.

Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

John Nash. The imbedding problem for Riemannian manifolds. *Annals of Mathematics*, 63(1):20—63, 1956.

Behnam Neyshabur, Ryota Tomioka, and Nathan Srebro. Norm-based capacity control in neural networks. In *COLT*, 2015.

Emmy Noether. Invariante variationsprobleme. In *König Gesellsch. d. Wiss. zu Göttingen, Math-Phys. Klassc*, pages 235–257. 1918.

Maks Ovsjanikov, Jian Sun, and Leonidas Guibas. Global intrinsic symmetries of shapes. *Computer Graphics Forum*, 27(5):1341–1348, 2008.

Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Trans. Graphics*, 31(4):1–11, 2012.

Aditya Pal, Chantat Eksombatchai, Yitong Zhou, Bo Zhao, Charles Rosenberg, and Jure Leskovec. Pinnersage: Multi-modal user embedding framework for recommendations at pinterest. In *KDD*, 2020.

Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical Image Analysis*, 48:117–130, 2018.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *ICML*, 2013.

Giuseppe Patanè. Fourier-based and rational graph filters for spectral processing. *arXiv:2011.04055*, 2020.

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

Roger Penrose. *The road to reality: A complete guide to the laws of the universe*. Random House, 2005.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *KDD*, 2014.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W Battaglia. Learning mesh-based simulation with graph networks. *arXiv:2010.03409*, 2020.

Fernando J Pineda. Generalization of back propagation to recurrent and higher order neural networks. In *NIPS*, 1988.

Ulrich Pinkall and Konrad Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics*, 2(1):15–36, 1993.

Allan Pinkus. Approximation theory of the mlp model in neural networks. *Acta Numerica*, 8:143–195, 1999.

Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific Data*, 5 (1):1–13, 2018.

Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017.

Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In *WSDM*, 2018.

H Qu and L Gouskos. Particlenet: jet tagging via particle clouds. *arXiv:1902.08570*, 2019.

Meng Qu, Yoshua Bengio, and Jian Tang. GMNN: Graph Markov neural networks. In *ICML*, 2019.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. Generating 3D faces using convolutional mesh autoencoders. In *ECCV*, 2018.

Dan Raviv, Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Symmetries of non-rigid shapes. In *ICCV*, 2007.

Noam Razin and Nadav Cohen. Implicit regularization in deep learning may not be explainable by norms. *arXiv:2005.06398*, 2020.

Scott Reed and Nando De Freitas. Neural programmer-interpreters. *arXiv:1511.06279*, 2015.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv:1506.01497*, 2015.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *ICML*, 2015.

Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.

AJ Robinson and Frank Fallside. *The utility driven dynamic error propagation network*. University of Cambridge, 1987.

Emma Rocheteau, Pietro Liò, and Stephanie Hyland. Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit. *arXiv:2007.09483*, 2020.

Emma Rocheteau, Catherine Tong, Petar Veličković, Nicholas Lane, and Pietro Liò. Predicting patient outcomes with graph representation learning. *arXiv:2101.03940*, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.

Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv:2006.10637*, 2020.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 115 (3):211–252, 2015.

Raif M Rustamov, Maks Ovsjanikov, Omri Azencot, Mirela Ben-Chen, Frédéric Chazal, and Leonidas Guibas. Map-based exploration of intrinsic shape differences and variability. *ACM Trans. Graphics*, 32(4):1–12, 2013.

Tim Salimans and Diederik P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. *arXiv:1602.07868*, 2016.

Alvaro Sanchez-Gonzalez, Victor Bapst, Kyle Cranmer, and Peter Battaglia. Hamiltonian graph networks with ODE integrators. *arXiv:1909.12790*, 2019.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *ICML*, 2020.

Aliaksei Sandryhaila and José MF Moura. Discrete signal processing on graphs. *IEEE Trans. Signal Processing*, 61(7):1644–1656, 2013.

Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. In *NIPS*, 2017.

Adam Santoro, Ryan Faulkner, David Raposo, Jack Rae, Mike Chrzanowski, Theophane Weber, Daan Wierstra, Oriol Vinyals, Razvan Pascanu, and Timothy Lillicrap. Relational recurrent neural networks. *arXiv:1806.01822*, 2018.

Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry. How does batch normalization help optimization? *arXiv:1805.11604*, 2018.

Ryoma Sato, Makoto Yamada, and Hisashi Kashima. Random features strengthen graph neural networks. *arXiv:2002.03155*, 2020.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. *arXiv:2102.09844*, 2021.

Anna MM Scaife and Fiona Porter. Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 2021.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Trans. Neural Networks*, 20(1):61–80, 2008.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv:1707.06347*, 2017.

Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.

Terrence J Sejnowski, Paul K Kienker, and Geoffrey E Hinton. Learning symmetry groups with hidden units: Beyond the perceptron. *Physica D: Nonlinear Phenomena*, 22(1-3):260–275, 1986.

Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.

Thomas Serre, Aude Oliva, and Tomaso Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the national academy of sciences*, 104(15):6424–6429, 2007.

Ohad Shamir and Gal Vardi. Implicit regularization in relu networks with the square loss. *arXiv:2012.05156*, 2020.

John Shawe-Taylor. Building symmetries into feedforward networks. In *ICANN*, 1989.

John Shawe-Taylor. Symmetries and discriminability in feedforward network architectures. *IEEE Trans. Neural Networks*, 4(5):816–826, 1993.

Nino Shervashidze, Pascal Schweitzer, Erik Jan Van Leeuwen, Kurt Mehlhorn, and Karsten M Borgwardt. Weisfeiler-lehman graph kernels. *JMLR*, 12(9), 2011.

Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2 (2):021001, 2020.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

Hava T Siegelmann and Eduardo D Sontag. On the computational power of neural nets. *Journal of Computer and System Sciences*, 50(1):132–150, 1995.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

Eero P Simoncelli and William T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proceedings., International Conference on Image Processing*, volume 3, pages 444–447. IEEE, 1995.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.

Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A Hilbert space embedding for distributions. In *ALT*, 2007.

Stefan Spalević, Petar Veličković, Jovana Kovačević, and Mladen Nikolić. Hierachial protein function prediction with tail-GNNs. *arXiv:2007.12804*, 2020.

Alessandro Sperduti. Encoding labeled graphs by labeling RAAM. In *NIPS*, 1994.

Alessandro Sperduti and Antonina Starita. Supervised neural networks for the classification of structures. *IEEE Trans. Neural Networks*, 8(3):714–735, 1997.

Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv:1412.6806*, 2014.

Balasubramaniam Srinivasan and Bruno Ribeiro. On the equivalence between positional node embeddings and structural graph representations. *arXiv:1910.00452*, 2019.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *JMLR*, 15(1):1929–1958, 2014.

Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv:1505.00387*, 2015.

Kimberly Stachenfeld, Jonathan Godwin, and Peter Battaglia. Graph networks with spectral message passing. *arXiv:2101.00079*, 2020.

Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackerman, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

Heiko Strathmann, Mohammadamin Barekatain, Charles Blundell, and Petar Veličković. Persistent message passing. *arXiv:2103.01043*, 2021.

Norbert Straumann. Early history of gauge theories and weak interactions. *hep-ph/9609230*, 1996.

Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *arXiv:1409.3215*, 2014.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, 2015.

Corentin Tallec and Yann Ollivier. Can recurrent neural networks warp time? *arXiv:1804.11188*, 2018.

Hao Tang, Zhiao Huang, Jiayuan Gu, Bao-Liang Lu, and Hao Su. Towards scale-invariant graph-related problem solving by iterative homogeneous gnns. In *NeurIPS*, 2020.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *WWW*, 2015.

Gabriel Taubin, Tong Zhang, and Gene Golub. Optimal surface smoothing as filter design. In *ECCV*, 1996.

Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Rémi Munos, Petar Veličković, and Michal Valko. Bootstrapped representation learning on graphs. *arXiv:2102.06514*, 2021.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv:1802.08219*, 2018.

Renate Tobies. Felix Klein—-mathematician, academic organizer, educational reformer. In *The Legacy of Felix Klein*, pages 5–21. Springer, 2019.

Andrew Trask, Felix Hill, Scott Reed, Jack Rae, Chris Dyer, and Phil Blunsom. Neural arithmetic logic units. *arXiv:1808.00508*, 2018.

John Tromp and Gunnar Farnebäck. Combinatorics of go. In *International Conference on Computers and Games*, 2006.

Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Springer, 2008.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv:1607.08022*, 2016.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv:1609.03499*, 2016a.

Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *ICML*, 2016b.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *ICLR*, 2018.

Petar Veličković, Rex Ying, Matilde Padovano, Raia Hadsell, and Charles Blundell. Neural execution of graph algorithms. *arXiv:1910.10593*, 2019.

Petar Veličković, Lars Buesing, Matthew C Overlan, Razvan Pascanu, Oriol Vinyals, and Charles Blundell. Pointer graph networks. *arXiv:2006.06380*, 2020.

Petar Veličković, Wiliam Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep Graph Infomax. In *ICLR*, 2019.

Kirill Veselkov, Guadalupe Gonzalez, Shahad Aljifri, Dieter Galea, Reza Mirnezami, Jozef Youssef, Michael Bronstein, and Ivan Laponogov. Hyperfoods: Machine intelligent mapping of cancer-beating molecules in foods. *Scientific Reports*, 9(1):1–12, 2019.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. *arXiv:1506.03134*, 2015.

Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. In *ICLR*, 2016.

Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with lipschitz functions. *JMLR*, 5:669–695, 2004.

Martin J Wainwright and Michael Irwin Jordan. *Graphical models, exponential families, and variational inference*. Now Publishers Inc, 2008.

Yu Wang and Justin Solomon. Intrinsic and extrinsic operators for shape analysis. In *Handbook of Numerical Analysis*, volume 20, pages 41–115. Elsevier, 2019.

Yu Wang, Mirela Ben-Chen, Iosif Polterovich, and Justin Solomon. Steklov spectral geometry for extrinsic shape analysis. *ACM Trans. Graphics*, 38(1): 1–21, 2018.

Yu Wang, Vladimir Kim, Michael Bronstein, and Justin Solomon. Learning geometric operators on meshes. In *ICLR Workshops*, 2019a.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graphics*, 38(5):1–12, 2019b.

Max Wardetzky. Convergence of the cotangent formula: An overview. *Discrete Differential Geometry*, pages 275–286, 2008.

Max Wardetzky, Saurabh Mathur, Felix Kälberer, and Eitan Grinspun. Discrete Laplace operators: no free lunch. In *Symposium on Geometry Processing*, 2007.

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *arXiv:1807.02547*, 2018.

Boris Weisfeiler and Andrei Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI Series*, 2(9):12–16, 1968.

Paul J Werbos. Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356, 1988.

Hermann Weyl. Elektron und gravitation. i. *Zeitschrift für Physik*, 56(5-6): 330–352, 1929.

Hermann Weyl. *Symmetry*. Princeton University Press, 2015.

Marysia Winkels and Taco S Cohen. Pulmonary nodule detection in ct scans with equivariant cnns. *Medical Image Analysis*, 55:15–26, 2019.

Jeffrey Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete Applied Mathematics*, 69(1-2):33–60, 1996.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *ICML*, 2019.

Yuxin Wu and Kaiming He. Group normalization. In *ECCV*, 2018.

Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs. *arXiv:2002.07962*, 2020a.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv:1810.00826*, 2018.

Keyulu Xu, Jingling Li, Mozhi Zhang, Simon S Du, Ken-ichi Kawarabayashi, and Stefanie Jegelka. What can neural networks reason about? *arXiv:1905.13211*, 2019.

Keyulu Xu, Jingling Li, Mozhi Zhang, Simon S Du, Ken-ichi Kawarabayashi, and Stefanie Jegelka. How neural networks extrapolate: From feedforward to graph neural networks. *arXiv:2009.11848*, 2020b.

Yujun Yan, Kevin Swersky, Danai Koutra, Parthasarathy Ranganathan, and Milad Heshemi. Neural execution engines: Learning to execute subroutines. *arXiv:2006.08084*, 2020.

Chen-Ning Yang and Robert L Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical Review*, 96(1):191, 1954.

Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *ICML*, 2016.

Jonathan S Yedidia, William T Freeman, and Yair Weiss. Bethe free energy, kikuchi approximations, and belief propagation algorithms. *NIPS*, 2001.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *KDD*, 2018.

Jiaxuan You, Rex Ying, and Jure Leskovec. Position-aware graph neural networks. In *ICML*, 2019.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *NIPS*, 2017.

Wojciech Zaremba and Ilya Sutskever. Learning to execute. *arXiv:1410.4615*, 2014.

Wei Zeng, Ren Guo, Feng Luo, and Xianfeng Gu. Discrete heat kernel determines discrete riemannian metric. *Graphical Models*, 74(4):121–129, 2012.

Jiani Zhang, Xingjian Shi, Junyuan Xie, Hao Ma, Irwin King, and Dit-Yan Yeung. Gaan: Gated attention networks for learning on large and spatiotemporal graphs. *arXiv:1803.07294*, 2018.

Yuyu Zhang, Xinshi Chen, Yuan Yang, Arun Ramamurthy, Bo Li, Yuan Qi, and Le Song. Efficient probabilistic logic reasoning with graph neural networks. *arXiv:2001.11850*, 2020.

Rong Zhu, Kun Zhao, Hongxia Yang, Wei Lin, Chang Zhou, Baole Ai, Yong Li, and Jingren Zhou. Aligraph: A comprehensive graph neural network platform. *arXiv:1902.08730*, 2019.

Weicheng Zhu and Narges Razavian. Variationally regularized graph-based representation learning for electronic health records. *arXiv:1912.03761*, 2019.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv:2006.04131*, 2020.

Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34 (13):i457–i466, 2018.