1 Introduction

The Element you are about to read tells a tale that shrank in the writing. The opposite is true for its companion volume [190]. When we were asked to write a contribution to what would become the Elements in Non-local Data Interactions: Foundations and Applications series in April 2021,[1] quickly the idea took hold to provide an overview, a snapshot of the field of differential equations and variational methods on graphs and their applications to machine learning, image processing, and image analysis. But this is a very active field, being developed in a variety of different directions, and so the story we wanted to tell outgrew the format of a contribution to the Cambridge Elements series. We are grateful to Cambridge University Press and the editors for deeming our full contribution worthy enough of its own publication [190] and allowing us to extract and adapt some introductory sections for publication in the Elements series. This Element provides an introduction to differential equations and variational methods on graphs seen through the lenses of the authors. It focuses on some areas to which we ourselves have actively contributed, but at the same time aims to give sufficient background to equip the interested reader to explore this fascinating topic. We have aimed to make this Element a satisfyingly comprehensive read in itself, while also allowing it to be a teaser for its more extensive companion volume. Those who compare both volumes will find that the companion volume contains, besides additional details in some of the sections that correspond to those in the current Element, chapters on applications of differential equations on graphs and their computational implementation, as well as chapters on further theoretical explorations of the relationships between various graph-based models and of the continuum limits of such models when the number of nodes of the underlying graph is taken to infinity.

Differential equations, both ordinary differential equations (ODEs) and partial differential equations (PDEs), have a long history in mathematics and we assume they need no introduction to the reader. Slightly younger, yet still of a venerable age, is the study of graphs.[2] By a differential equation on a graph

[1] The idea for this Element and its companion book owes a great debt to the Nonlocal Methods for Arbitrary Data Sources (NoMADS) project funded by the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement No. 777826).

[2] In this Element, by 'graph' we mean the discrete objects consisting of vertices and (potentially weighted) edges connecting the vertices that are studied in graph theory. Where we want to talk about the graph of a function, this is made explicit. As is quite common, we tend to use 'graph' and 'network' [78] interchangeably, although some might prefer to distinguish between networks in the real world and the mathematical graphs that can be used to model them. We also use related terminology, such as vertices and nodes [15, box 2.1], interchangeably.

we mean a discretization of a differential equation, usually a PDE, on a graph: if we can write the PDE as $F(u(x,t)) = 0$, for a differential operator F and a function u defined on (a subset of) $\mathbb{R}^m \times \mathbb{R}$, then we obtain a differential equation[3] on a graph by replacing the (spatial) derivatives with respect to x in F by finite difference operators based on the structure of the graph and replacing u by a function defined on (a subset of) $V \times \mathbb{R}$, where V is the set of nodes of the graph.

Variational models in the calculus of variations are typically formulated in terms of minimization of a function(al). To consider such a model on a graph, we discretize the functional by replacing integrals with the appropriate sums and differential operators with the corresponding finite difference operators on graphs. This process also turns finite-dimensional the space of admissible functions over which the functional is minimized. The line between calculus of variations and mathematical optimization becomes blurry here, or is even crossed. Because the authors of this Element approach these models from the point of view of variational calculus, we will include them under the umbrella of variational models (on graphs), even if the originators of any particular model may have had a different inspiration when proposing that model.

The field of machine learning is concerned with the development of methods and algorithms to analyse data sets. ‘Learning’ in this context refers to leveraging the properties of some collection of ‘training data’ (which may or may not be a part of the data set which is to be analysed) to draw conclusions about the data set. Machine learning has undertaken an enormous flight in the twenty-first century. Terms like ‘big data’, ‘machine learning’, and ‘artificial intelligence’ are now commonplace for many people, both because of the commercial successes of the many tech companies that exist by the grace of data availability and the methods to learn from the data, and because of the enormous speed with which deep-learned algorithms have transformed many areas of science, industry, and public and private life. Scientific curiosity goes hand in hand with a societal need to understand the methods that play such a big role in so many sectors.

But what is the role of differential equations in all of this? After all, many of the advances in machine learning, both in terms of development of new methods and analysis of existing ones, come from statistics, computer science, and the specific application fields – scientific or industrial – where the methods are used. One might argue for the general notion that increased diversity in the points of view from which a particular topic is scrutinized leads to different and complementary insights that all strengthen the full picture. There certainly

[3] Or a difference equation, if u does not depend on the variable t.

is validity in that notion when it comes to the study of machine learning; in this case, though, there are stronger ties that go beyond generalities.

A substantial part of the root system of differential equations in machine learning lies in the field(s) of mathematical image processing and image analysis.[4] Not only do (the ingredients of) many differential-equation-based machine learning methods have roots in the mathematical imaging literature and many machine learning methods have applications in imaging problems, but there is also a substantial overlap in the communities active in these fields.

Despite the success of artificial neural networks, they are not central objects in this Element. The main focus in the current Element is on those methods that represent the data (e.g., the image) in terms of a graph in order to apply a graph-based variational model or differential equation.

These methods have many desirable properties, which have made them popular in recent years. Because the graph-based models and equations have close connections to well-established models and equations in the continuum setting, there is a wealth of ideas to pursue and techniques to apply to study and understand them. Moreover, the development of numerical methods for differential equations has a long and rich history, yielding many algorithms that can be adapted to the graph-based setting. Another key difference between most machine learning methods discussed in this Element compared to deep learning methods is that the latter are mostly data-driven,[5] which usually means many training data are required to obtain well-performing networks, while the former are explicitly model-driven and so tend to require fewer training data (but also are less likely to discover patterns for which the models are not built to look).

The scope of this Element is broad in some parts and narrow in others. On the one hand, we wish to provide an overview of an exciting, ever-broadening, and expansive field. Thus, in the general literature overview in Section 2 we have taken a very broad view of the topic of differential equations on graphs and their applications with the aim of placing it in its historical context and pointing the reader to the many aspects and fields that are closely related to it. On the other hand, in the remainder of this Element, we focus on very specific models with a certain amount of bias in the direction of those models with which the authors have close experience themselves, yet not to the complete exclusion of other models. These models revolve around the graph Ginzburg–Landau functional,

[4] We sometimes refer to these two fields collectively as ‘imaging’ or ‘mathematical imaging’, or just ‘image analysis’ for brevity.

[5] Although there are recent trends to incorporate (physical) models into the data-driven machinery.

which is introduced in this Element in Section 5 and which plays a central role in a number of graph-based clustering and classification methods.

Graph clustering and classification are similar tasks that both aim to use the structure of the graph to group its nodes into subsets called clusters or classes (or sometimes communities or phases, depending on the context or application). In most of the settings that we discuss here, these subsets are required to be pairwise disjoint, so that mathematically we can speak of a partition of the node set (assuming nonempty subsets) and we obtain non-overlapping clusters or classes. A key guiding principle of both tasks is to have strong connections (i.e., many edges or, in an edge-weighted graph, highly weighted edges) between nodes in the same class or cluster and few between nodes in different clusters or classes.[6] This is not the only requirement for a good clustering or classification – in the absence of any other desiderata, a trivial partition of the node set into one cluster would maximize intra-cluster connectivity and minimize inter-cluster connectivity – and so additional demands are typically imposed. Two types of constraints that will receive close attention in this Element are constraints on cluster sizes and constraints that encourage fidelity to a priori known class membership of certain nodes. The presence of such a priori known labels is what sets classification apart from clustering.

There are mathematical imaging tasks that can be formulated in terms of graph clustering or classification, most notably image segmentation, which can be viewed as the task of clustering or classifying the pixels of a digital image based on their contribution to (or membership of) objects of interest in the image. Other imaging tasks, such as image denoising, reconstruction, or inpainting, can be formulated in ways that are mathematically quite closely related to graph clustering and classification.

We hope this Element may serve as an overview and an inspiration, both for those who already work in the area of differential equations on graphs and for those who do not (yet) and wish to familiarize themselves with a field rich with mathematical challenges and abundant applications.

1.1 Outline of This Element

We give a brief overview of the history of and literature in the field of differential equations and variational methods on graphs with applications in machine learning and image analysis in Section 2. For a more extensive overview, we refer to section 1.2 of the companion volume [190]. In Section 3 we lay the

[6] There may be deviations from or additions to this general requirement. For example, in Macgregor and Sun [135] (see also Macgregor [133, chapter 6]) the goal is to find two clusters (with an algorithm that is local, in the sense that its run time is independent of the size of the graph) that are densely connected to each other, but weakly to the rest of the graph.

mathematical foundations that we require to formulate differential equations and variational models on graphs. For example, we define spaces of node functions and important operators and functionals on these spaces, such as the graph Laplacian operators and the graph total variation functional.

In most of this Element we consider undirected graphs, but in Section 4 we discuss very briefly works that generalize some of the concepts from Section 3 to directed graphs, in particular graph Laplacians.

Besides the graph total variation functional, another very important graph-based functional, which we have already mentioned, is the graph Ginzburg–Landau functional. It deserves its own place in the spotlight; thus, in Section 5 we introduce it and discuss some of its variants and properties, including its connection to the graph total variation functional.

The spectrum of an operator gives insight into its behaviour. In Section 6 indeed we see that the spectra of graph Laplacians shed some light on their usefulness in graph clustering and classification problems.

The role of functionals in variational models is as ‘energy’ or an ‘objective function’ that needs to be minimized. Two important dynamical systems that (approximately) accomplish this minimization for the graph Ginzburg–Landau functional are described by the graph Allen–Cahn equation and the graph Merriman–Bence–Osher scheme, respectively. The former, which is an ordinary differential equation obtained as a gradient flow of the graph Ginzburg–Landau functional, is the topic of Section 7, while the latter is the focus of Section 8.

Closely related to the graph Allen–Cahn and Merriman–Bence–Osher dynamics are the graph mean curvature flow dynamics that are described in Section 9. Although exactly how closely these are related is still an open question, one property most of them have in common is a threshold on the parameter of the model below which the dynamics trivializes. This freezing phenomenon is discussed in Section 10.

The main focus in this Element is on models that cluster or classify the node set of the graph into two subsets. In Section 11 we take a look at multiclass extensions of some of these models.

In Section 12 we discuss some finite difference methods on graphs, namely Laplacian learning and Poisson learning. Finally, Section 13 provides a brief conclusion to this Element as we look forward (or sideways) to additional topics that appear in the companion volume [190].

2 History and Literature Overview

Any area of mathematical enquiry will have its roots in what came before. The field of differential equations on graphs for clustering and classification problems is no exception, so there is always the risk that the starting point of any historical overview may feel somewhat arbitrary. It is inescapably personal too; the priorities and narratives generated by our own research inevitably have influenced our departure point for this story and the route we will follow afterwards. As such, the references given in this section are not meant to be exhaustive, nor are all contributions from the references that are given always exhaustively described. One reason this field has generated such enthusiastic interest is that it brings together ideas from many different directions, from statistics and machine learning to discrete mathematics and mathematical analysis. We encourage any attempts to understand the history of this field from perspectives different from the one we provide here, shining more light on the exciting diversity of viewpoints differential equations on graphs have to offer.

For a fuller overview of the literature, we refer to section 1.2 of the companion volume [190]. Neither the current section nor the literature section in the companion volume are exhaustive overviews, if such a thing is even a possibility, but they should provide enough starting points for someone who is eager to learn about this active field of research. We apologize for the undoubtedly many important works in their respective areas that are missing from our bibliography.

As a double ignition point for the growing interest in differential equations on graphs by the (applied) mathematical analysis community in the late noughties and early tens of the twenty-first century, especially in relation to image processing and data analysis applications, we mention works by Abderrahim Elmoataz and collaborators such as [63, 75, 131] – which have a strong focus on p-Laplacians, ∞-Laplacians, and morphological operations such as dilation and erosion on graphs, as well as processing of point clouds in three dimensions – and works by Andrea L. Bertozzi and collaborators, like [21, 22, 144, 189, 191] – which deal with the Ginzburg–Landau functional on graphs and derived dynamics such as the Allen–Cahn equation and Merriman–Bence–Osher (MBO) scheme on graphs. This is not to say there were no earlier investigations into the topic of differential operators and equations on graphs, for example in [160, 195] or in the context of consensus problems [150, 164], or variational 'energy' minimization problems on graphs, such as in the context of Markov random fields [181], but the two groups we mentioned earlier provided a sustained drive for the investigation of both the applied and theoretical aspects of differential equations on graphs in the setting on which we are focusing in the current Element.

We note that in the current Element, when we talk about *differential equations on graphs*, we typically mean partial differential equations whose spatial derivatives have been replaced by discrete finite difference operators on graphs (see Section 3.2), leading to ODEs or finite-difference equations. A priori this is different from the systems of PDEs formulated on the edges of a network that are coupled through boundary conditions on the nodes, which are also studied under the name 'differential equations on graphs (or networks)' [192].

New research directions tend not to spring into being fully formed, and also the works mentioned earlier have had the benefit of a rich pre-existing literature in related fields. In particular, many of the variational models and differential equations that are studied on graphs have been inspired by continuum cousins that came before and by the, often sizeable, literature that exists about those models and equations. Examples are the Allen–Cahn equation [5], the Cahn–Hilliard equation [40], the Merriman–Bence–Osher scheme [146, 147], flow by mean curvature [6, 30, 167], total variation flow [9], and the Mumford–Shah and Chan–Vese variational models [47, 159].

Inspiration has also come from other directions, such as discrete calculus [101] and numerical analysis and scientific computing [18]. The latter fields not only provide state-of-the-art methods that allow for fast implementations of the graph methods on graphs with many nodes – something which is very important in modern-day applications that deal with large data sets or high-resolution images – but also offer theoretical tools for dealing with discretizations of continuum equations, even though the specifics may differ substantially between a discretization designed with the goal of approximating a continuum problem as accurately as possible and a discretization which is a priori determined by the graph structure that is given by the problem or inherent in the application at hand.

Some of the most prominent applications that have been tackled by differential equations and variational models on graphs, especially by the models and methods central to the current Element, are graph clustering and classification [174], and other applications that are – or can be formulated to become – related, such as community detection [171], image segmentation [46], and graph learning [200]. For an extensive look at these, and other, applications, we refer to chapter 4 of the companion volume [190]. In the current Element we focus on the core graph clustering and classification applications.

Since the early pioneering works we mentioned at the start of this section, differential equations on graphs have enjoyed a lot of attention from applied analysts and researchers from adjacent fields. As a very rough attempt at classification of these different research efforts, we distinguish between those papers that study differential equations and variational models purely at the discrete-graph level and those that are interested in continuum limits of those discrete equations and models as a way to establish their *consistency*. (Some papers may combine elements of both categories.) The focus of this Element is on the first category. For a closer look at the second category, we refer to chapter 7 of [190].

In this first category, we encounter papers that study particular graph-based differential operators or graph-based dynamics, such as the eikonal equation (and the related eikonal depth), p-eikonal equation, p- and ∞-Laplacians [38, 42, 76, 138, 206], semigroup evolution equations [153], dynamics [121] such as mean curvature flow and morphological evolution equations (related to morphological filtering and graph-based front propagation) [70, 182] or advection [172], or discrete variational models such as trend filtering on graphs [194] and the graph Mumford–Shah model [106, 168].

Of special interest in the context of the current Element are the graph Allen–Cahn equation, graph MBO scheme, and graph mean curvature flow [33, 144, 187, 191], which are discussed in much greater detail in Sections 7, 8, and 9. For details about applications in which these graph-based dynamics have been used, we refer to chapter 4 of [190]. Some of the applications that are considered in that volume require an extension of the classical two-phase versions of the Allen–Cahn and MBO dynamics to a multiclass context [94, 95]; other variations on multiclass MBO have been developed, such as an incremental reseeding method [31] and auction dynamics [110].

The publications focusing on the discrete level also include papers that study connections between graph-based differential operators or dynamics on the one hand, and on the other hand graph-based concepts that are useful for studying graph structures, which sometimes already had a history outside of the differential-equations-on-graphs literature. For example: the modulus on graphs [3], Cheeger cuts and ratio cuts [141], ranking algorithms and centrality measures such as heat kernel PageRank [142], nonconservative alpha-centrality (as opposed to conservative PageRank) [98], centrality measures and community structure based on the interplay between dynamics via parameterized (or generalized) Laplacians and the network structure [201], random walks and SimRank on uncertain graphs (i.e., graphs in which each edge has a probability of existence assigned to it) [209], distance and proximity measures on graphs [12, 48], and a hubs-biased resistance distance (based on graph Laplacians) [79].

We also draw attention here to graph-based active learning [148], Laplacian learning [211], Poisson learning [44], and results on the Lipschitz regularity of functions in terms of Laplacians on point clouds [45]. For overview articles about graph-based methods in machine learning and image processing, we refer to [20, 23, 51].

In the continuum setting, the dynamics of both the Allen–Cahn equation and the MBO scheme are known to approximate flow by mean curvature, in a sense that has been made precise through convergence analysis [16, 32, 39, 80, 125, 126]. Rigorous connections between the graph Allen–Cahn equation and graph MBO scheme have been established in [33, 35, 36, 37]; their connections with graph mean curvature flow are open questions. Details about these established connections and open questions, as well as a closer look at the various dynamics in the continuum setting, can be found in chapter 6 of [190].

In many of the works just cited, the graphs under consideration on which the variational models or differential equations are formulated are finite, undirected graphs, with edges that connect nodes pairwise and that, if weighted, have a positive weight. Moreover, the graphs are unchanging – also, for the continuum limits that we briefly mentioned, even though the limit $|V| \to \infty$ is considered, typically at each fixed $|V|$, the graph structure is static.

This leaves a lot of room for generalizations and extensions. In this Element we refrain from delving into these generalizations in too much detail, although in Section 4 we do briefly discuss Laplacians on directed graphs [17, 87, 104, 207]. Other possible generalizations are to multislice networks [151], hypergraphs (in which edges can connect more than two nodes) [26, 108], metric graphs and quantum graphs (in which edges are represented by intervals of the real line) [118], signed graphs that can have positive and negative edge weights [60], metric random walk spaces (of which locally finite positively edge-weighted connected graphs are a special case) [139, 140][7], and graphs changing in time [25].

Of interest also is the connection between methods on graphs and the constructions used to build the graphs, as is considered in, for example, [97]. Some more details about building graph models for specific applications are given in section 4.2 of [190].

3 Calculus on Undirected Edge-Weighted Graphs

3.1 Graphs, Function Spaces, Inner Products, and Norms

Except where explicitly stated otherwise, in this Element we consider finite,[8] simple (i.e., without multi-edges[9] and without self-loops[10]), connected,[11] edge-weighted graphs $G = (V, E, \omega)$ – if a graph G is mentioned without further

[7] In [139] heat flow on metric random walk spaces is studied, in [140] total variation flow.

[8] This means $|V| < \infty$.

[9] Given two nodes $i, j \in V$, there is at most one edge $(i, j) \in E$, as is already implied by $E \subseteq V \times V$.

[10] Each edge connects two distinct nodes, that is, for all $i \in V$, $(i, i) \in (V \times V \setminus E)$. For a preprint discussing Laplacians (which we will introduce later) on graphs with self-loops see Açıkmeşe [2].

[11] The definition of connectivity is given later in this section.