Preface for the special issue *Big Data and Partial Differential Equations*

Yves van Gennip and Carola-Bibiane Schönlieb

September 21, 2017

Partial differential equations (PDEs) are expressions involving an unknown function in many independent variables and their partial derivatives up to a certain order. Since PDEs express continuous change, they have long been used to formulate a myriad of dynamical physical and biological phenomena: heat flow, optics, electrostatics and -dynamics, elasticity, fluid flow and many more. Many of these PDEs can be derived in a variational way, i.e. via minimization of an 'energy' functional. In this globalized and technologically advanced age, PDEs are also extensively used for modeling social situations (e.g. models for opinion formation, mathematical finance, crowd motion) and tasks in engineering (such as models for semiconductors, networks, and signal and image processing tasks).

In particular in recent years there has been increasing interest from applied analysts in applying the models and techniques from variational methods and PDEs to tackle problems in data science. This issue of the European Journal of Applied Mathematics highlights some recent developments in this young and growing area. It gives a taste of endeavours in this realm in two exemplary contributions on PDEs on graphs [1, 2] and one on probabilistic domain decomposition for numerically solving large-scale PDEs [3].

The graph framework Applied mathematics research on graphs in the context of data science starts with the observation that many kinds of discrete data can be represented as a weighted graph or network. This representation is convenient when developing data processing methods as it provides a mathematical structure that one can work with. In 2012 a paper by Andrea Bertozzi and Arjuna Flenner kicked off a boom in applied mathematics research on this topic [4]. In this paper the authors use graph versions of the Ginzburg-Landau functional for data clustering, data classification and image segmentation. Minimization of the classical continuum Ginzburg-Landau functional,

$$F(u) := \varepsilon \int_{\Omega} |\nabla u|^2 \, dx + \frac{1}{\varepsilon} \int_{\Omega} W(u) \, dx,$$

provides a model for phase separation. Here $W(u) = u^2(1 - u)^2$ is a double well potential with minima at $u = 0$ and $u = 1$, and u describes the relative presence of the two phases $\{u \approx 0\}$ and $\{u \approx 1\}$ in the domain Ω. When F is minimized under some suitable constraints on u (e.g. a mass constraint of the form $\int_{\Omega} u \, dx = M$) and for small values of the parameter ε, u will take values close to 0 and 1, with transitions between those values occurring in small regions of width $\mathcal{O}(\varepsilon)$.

In [4] the graph functional

$$f(u) := \sum_{i,j \in V} \omega_{ij}(u_i - u_j)^2 + \frac{1}{\varepsilon} \sum_{i \in V} W(u_i)$$

was introduced. This is a functional whose input argument u is a function on the nodes of a graph, instead of on a continuum set $\Omega \subset \mathbb{R}^n$ and which serves as a graph counterpart to F. Here V is the node set of the graph, ω_{ij} is a nonnegative weight on the edge between nodes i and j in a finite, simple, undirected graph and u_i is the value of the function u on node i. In [4] this functional was used in combination with either a

mass constraint or an additional data fidelity term of the form $\sum_{i:\text{ training data}}(u_i - u_i^{\text{training}})^2$ to cluster or classify the nodes of a graph into two groups ('phases' where $u \approx 0$ and $u \approx 1$) based on the pairwise node similarity encoded in the edge weights ω_{ij}. By treating the pixels of an image as nodes in a graph, data classifcation can be used for image segmentation as well.

Interesting mathematical questions that could arise from such a model are:

1. Can we find graph analogues of properties of the continuum functional?

2. Is the continuum functional a limit of the graph functionals in some sense?

3. What can we say about the resulting algorithm and its usage for data analysis/image processing?

4. Are there other network problems that can be tackled by a PDE inspired approach?

5. Are there other PDE/variational systems that have interesting network analogues? And if the inspiring PDEs are related, are their graph analogues related?

Some of these questions have been considered in the state-of-the-art literature, with some highlights reported in the following.

1. Does f have similar properties to F? In [5] the authors proved that f Γ-converges [6, 7], when $\varepsilon \to 0$, to the graph total variation functional

$$TV(u) := \frac{1}{2} \sum_{i,j \in V} \omega_{ij} |u_i - u_j|,$$

which has as domain the set of node functions u which take values in $\{0, 1\}$. This mirrors the well-known continuum result [8, 9]. Moreover, for such $\{0,1\}$-valued functions u, $TV(u)$ reduces to the graph cut [10] of the node partition $V_0 = \{i : u_i = 0\}$, $V_1 = \{i : u_i = 1\}$, i.e. the sum of the edge weights ω_{ij} corresponding to edges that have one node in V_0 and the other in V_1.

2. Furthermore, when f or TV are defined on certain graphs for which a sensible continuum limit can be defined, they Γ-converge to the continuum total variation in the continuum limit, e.g. on 4-regular graphs obtained by ever finer discretisations of the flat torus [5] and on point clouds obtained by sampling ever more points from an underlying subset of $\mathbb{R}^n$ [11, 12, 13]. In the latter context these limit results can be interpreted as consistency results that show that the discrete model defined on the samples is asymptotically consistent with a continuum model.

3. Minimization of f is in practice (approximately) achieved either by solving a gradient flow equation of Allen-Cahn type,

$$\frac{du_i}{dt} = -\sum_{j \in V} \omega_{ij}(u_i - u_j) - \frac{1}{\varepsilon} W'(u_i)$$

(plus additional terms coming from a mass constraint or fidelity term) or by a graph version of the threshold dynamics (or MBO) scheme [14]:

$$u^{k+1} = \begin{cases} 0, & \text{if } \tilde{u}(\tau) < \frac{1}{2}, \\ 1, & \text{if } \tilde{u}(\tau) \geq \frac{1}{2}, \end{cases} \quad \text{where } \tilde{u}(t) \text{ solves } \begin{cases} \frac{d\tilde{u}_i}{dt} = -\sum_{j \in V} \omega_{ij}(\tilde{u}_i - \tilde{u}_j), \\ \tilde{u}(0) = 0. \end{cases}$$

In the (spectral) graph theory literature [10, 15] $(\Delta u)_i := \sum_{j \in V} \omega_{ij}(u_i - u_j)$ is known as the unnormalised or combinatorial graph Laplacian of u. The equations above can also be formulated and solved with normalised versions of the graph Laplacian.

On a given graph, these equations can be solved quickly and accurately, for example by using a truncated spectral decomposition based on the eigenfunctions of the graph Laplacian in combination with a convex splitting scheme in the case of the graph Allen-Cahn equation [4, 16].

The construction of the underlying graph in the first place can pose a significant computational problem, especially when the number of data points (and thus nodes in the graph) is very large. Matrix completion techniques such as the Nyström extension [17, 18] and fast eigenvalue computation algorithms such as the Rayleigh-Chebychev algorithm [19] make such computations feasible.

This graph Ginzburg-Landau method has found many applications, for example in data clustering and classification and image segmentation [4, 16, 20] and has also been extended to deal with clustering and classification into more than two classes [21, 22, 23, 24, 25]. Recent papers prove convergence of the graph Allen-Cahn algorithm (both the spectrally untruncated and truncated versions) and extend the method to non-smooth potentials and hypergraphs [26, 27].

This shows that such PDE driven techniques can provide fast approximative alternatives to combinatorial problems whose exact solution is too computationally complex.

4. Another example of such a problem is the computation of a maximum cut in graphs, i.e. to find a partition of the node set into two sets such that the sum of the edge weights corresponding to edges with one node in each set is maximal. If the graph is bipartite, this corresponds to partitioning the node set according to the bipartite structure. The exact solution of this classical problem is known to be computationally unfeasible for large graphs. Work currently in preparation introduces a fast approximate solution method for this problem using an adaptation of the graph Ginzburg-Landau functional f [28].

5. The continuum counterparts of both the graph Allen-Cahn equation and graph MBO scheme from point 3. can be viewed as approximating mean curvature flow [29, 30, 31, 32, 33]. This suggests that graph curvature and graph mean curvature flow are interesting concepts to consider as well. In [34] the authors introduced both. The graph curvature of a node set S is given by

$$\kappa_i := \begin{cases} \sum_{j \in S^c} \omega_{ij}, & \text{if } i \in S, \\ -\sum_{j \in S} \omega_{ij}, & \text{if } i \in S^c, \end{cases}$$

and the related graph mean curvature flow has a variational formulation along the lines of [35, 36, 37] which leads to a time discrete evolution of node subsets S (given an initial set S_0),

$$S_{n+1} \in \operatorname{argmin}_{\hat{S}} \mathcal{F}(\hat{S}, S_n),$$

where

$$\mathcal{F}(\hat{S}, S_n) := \sum_{i \in S, j \in S^c} \omega_{ij} + \frac{1}{\eth t} \sum_{i \in \hat{S}} d_i s d_i^n.$$

Here $\eth t > 0$ is the time step, d_i is the degree of node i and sd_i^n is the signed graph distance from node i to the boundary of node set S_n. In [34] the authors started studying the very interesting question wether the graph Allen-Cahn equation, graph MBO scheme and graph mean curvature flow are as intimately connected as their continuum counterparts, but establishing such connections is still mostly an open problem.

Other current work studies a graph version of the Ohta-Kawasaki functional [38], which was originally introduced as a variational model for pattern formation in diblock copolymers [39].

The research on these novel methods has shown that new PDE-inspired graph procedures can efficiently (approximately) solve complex graph problems, while at the same time offering fertile ground for proving theoretical connections between the various graph problems (inspired by similar connections their continuum counterparts have) and between the graph problems and their continuum analogues.

Paper [1] relates to question 2. above. Its authors apply similar ideas to those in [11, 12, 13] to prove a consistency result for empirical risk minimization. If a function $u : D \to \{0, 1\}$ acts as a classifier for points that are sampled from $D \subset \mathbb{R}^n$ according to a distribution ν, its empirical risk is

$$R(u) := \int_{D \times \{0,1\}} |u(x) - y| d\nu(x, y).$$

In [1] the authors prove a consistency result for regularized empirical risk functional, which consists of empirical risk regularized by a graph total variation term. They also find different regimes for the regularization parameter associated with the total variation term which relate to the notions of overfitting and underfitting of the data.

Paper [2] is related to question 5. above. It proposes a graph version of the game p-Laplacian [40] as an interpolation between the graph (2-)Laplacian and the graph ∞-Laplacian when $2 \leq p \leq \infty$ and as an interpolation between the graph 1-Laplacian and the graph 2-Laplacian when $1 \leq p < 2$. It proves the existence and uniqueness of game p-harmonic functions with given Dirichlet 'boundary' conditions on a subset of the nodes and relates the graph game p-Laplacian to a tug-of-war game. It also shows the results of numerical experiments in which the graph game p-Laplacian is used for semi-supervised segmentation, clustering and image inpainting.

Numerical solution of large-scale PDEs   When developing algorithms for data-driven applications the scalability of computational methods is essential. In the context of partial differential equations domain decomposition methods are used to divide a large domain into several smaller subdomains in such a way that the solution to the equation on the full domain can be found (or approximated) via the solutions on the subdomains [41, 42, 43, 44, 45, 46]. The initial equation restricted to the subdomains defines a sequence of new local problems which are computationally cheaper to solve. A principal motivation behind this principle is the formulation of PDE solvers which can be easily parallelised. As an example, let us consider a coercive,
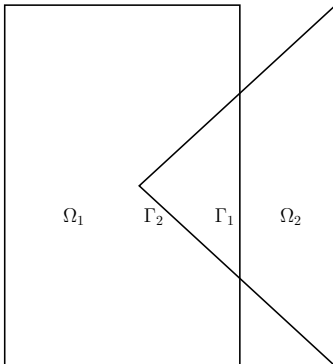

Figure 1: Overlapping domain decomposition. Example setup from [44].

elliptic and self-adjoint differential operator L and the boundary-value problem

$$Lu = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega,$$

for $\Omega \subset \mathbb{R}^n$, with $n > 1$ and f sufficiently nice. The basic Schwarz alternating algorithm [44] to solve this equation is, starting with an initial guess u^0 for the solution, we iterate for $k = 0, 1, \ldots$, for subdomains Ω_1, Ω_2 with interfaces Γ_1, Γ_2 as in Figure 1:

$$Lu_1^{k+1} = f, \quad \text{in } \Omega_1,$$
$$u_1^{k+1} = u^k|_{\Gamma_1} \quad \text{on } \Gamma_1,$$
$$u_1^{k+1} = 0 \quad \text{on } \partial\Omega_1 \setminus \Gamma_1,$$

4

and

$$Lu_2^{k+1} = f, \quad \text{in } \Omega_2,$$
$$u_2^{k+1} = u^k|_{\Gamma_2} \quad \text{on } \Gamma_2,$$
$$u_2^{k+1} = 0 \quad \text{on } \partial\Omega_2 \setminus \Gamma_2,$$

and the iterate u^{k+1} on the whole domain defined by

$$u^{k+1}(x) = \begin{cases} u_2^{k+1}(x) & \text{if } x \in \Omega_2, \\ u_1^{k+1}(x) & \text{if } x \in \Omega \setminus \Omega_2. \end{cases}$$

Under certain conditions it can be shown that the iterates converge to the true solution u on Ω. Now, the idea is to distribute the solution of the two subproblems on Ω_1 and Ω_2 to two processors and compute u_1^k and u_2^k in parallel, exchanging information of the solution on the interfaces of the two subdomains between the two processors after every iteration.

In [3] the authors review a particular class of domain decomposition methods which is called probabilistic domain decomposition (PDD), pioneered by Acebron et al. in [48]. The main idea is to use a stochastic representation of the PDE (via the so-called Feynman-Kac formula), then compute the solution in a few sampled points on the interfaces between the subdomains via Monte Carlo, and then use an efficient deterministic PDE solver for the solution of the PDE on each subdomain with fixed boundary conditions coming from the previously computed Monte Carlo simulation. This probabilistic setup renders a parallelisation strategy for solving PDEs that shows very good scalability properties. Indeed, since the solutions of the PDE on the subdomains are completely independent of each other the PDD method can solve the subproblems fully in parallel and hence does not require communication between processors. Also, the choice of the PDE solver on each subdomain is flexible and so can potentially be executed very efficiently. The paper [3] also serves as an introduction to the concept of PDDs and their various usage areas in data-driven applications.

Conclusion The contributions in this issue are examples of modern research topics in partial differential equations that arise in data-driven applications. They nicely show the various facets of this topic, from PDEs as inspiration for and tools to prove consistency of graph based data processing methods to the development of efficient and scalable numerical methods for solving PDEs for large-scale and high-dimensional data. We believe we are only at the beginning of this exciting and important new area of research. The increasing number of open questions in all areas of data science will make mathematical frameworks, like the one provided by PDEs, more and more attractive. PDEs and variational methods have an important role to play in finding answers to these questions and in the development of adaptive, rigorous and efficient data processing and machine learning methods.

References

[1] N.G. Trillos and R. Murray, *A new analytical approach to consistency and overfitting in regularized empirical risk minimization*, EJAM 2017.

[2] A. Elmoataz, X. Desquesnes, and M. Toutain, *On the game p-Laplacian on weighted graphs with applications in image processing and data clustering*, EJAM 2017

[3] F. Bernal, G. dos Reis, and G. Smith, *Hybrid PDE solver for data-driven problems and modern branching*, EJAM 2017.

[4] A.L. Bertozzi, and A. Flenner, *Diffuse interface models on graphs for analysis of high dimensional data*, Multiscale Modeling and Simulation **10(3)** (2012), 1090–1118.

[5] Y. van Gennip and A.L. Bertozzi, *Γ-convergence of graph Ginzburg-Landau functionals*, Advances in Differential Equations **17(11/12)** (2012), 1115–1180.

[6] G. Dal Maso, *An introduction to Γ-convergence*, Birkhäuser, 1993

[7] A. Braides, *Γ-convergence for beginners*, Oxford University Press, 2002

[8] L. Modica and S. Mortola, *Un esempio di Γ-convergenza*, Bollettino dell'Unione Matematica Italiana **5(14-B)** (1977), 285–299.

[9] L. Modica, *The gradient theory of phase transitions and the minimal interface criterion*, Archive for Rational Mechanics and Analysis **98(2)** (1987), 123–142.

[10] U. von Luxburg, *A tutorial on spectral clustering*, Statistics and Computing **17(4)** (2007), 395–416.

[11] N. Garcia Trillos and D. Slepčev, *Continuum limit of total variation on point clouds*, Archive for Rational Mechanics and Analysis *220(1)* (2016), 193–241.

[12] N. Garcia Trillos, D. Slepčev, J. von Brecht, T. Laurent, and X. Bresson, *Consistency of Cheeger and Ratio Graph Cuts*, Journal of Machine Learning Research **17** (2016), 1–46.

[13] M. Thorpe and F. Theil, *Asymptotic Analysis of the Ginzburg-Landau Functional on Point Clouds*, to appear, arXiv preprint arXiv:1604.04930

[14] B. Merriman, J.K. Bence, and S.J. Osher, *Motion of multiple functions: a level set approach*, Journal of Computational Physics **112(2)** (1994), 334–363.

[15] F. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997.

[16] E. Merkurjev, T. Kostic, and A.L. Bertozzi, *An MBO scheme on graphs for segmentation and image processing*, SIAM Journal on Imaging Sciences **6(4)** (2013), 1903–1930.

[17] E.J. Nyström, *Über die Praktische Auflösung von Linearen Integralgleichungen mit Anwendungen auf Randwertaufgaben der Potentialtheorie*, Commentationes Physico-Mathematicae **4(15)** (1928), 1–52.

[18] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, *Spectral grouping using the Nystrom method*, IEEE Transactions on Pattern Analysis and Machine Intelligence **26(2)** (2004), 214–225.

[19] C.R. Anderson, *A Rayleigh–Chebyshev procedure for finding the smallest eigenvalues and associated eigenvectors of large sparse Hermitian matrices*, Journal of Computational Physics **229(19)** (2010), 7477–7487.

[20] L. Calatroni, Y. van Gennip, C.-B. Schönlieb, H. Rowland, and A. Flenner, *Graph clustering, variational image segmentation methods and Hough transform scale detection for object measurement in images*, Journal of Mathematical Imaging and Vision **57(2)** (2017), 269–291.

[21] C. Garcia-Cardona, A. Flenner, and A.G. Percus, *Multiclass diffuse interface models for semi-supervised learning on graphs*, Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods (ICPRAM 2013) (2013), 78–86.

[22] E. Merkurjev, C. Garcia-Cardona, A.L. Bertozzi, A. Flenner, and A.G. Percus, *Diffuse interface methods for multiclass segmentation of high-dimensional data*, Applied Mathematics Letters **33** (2014), 29–34.

[23] C. Garcia-Cardona, E. Merkurjev, A.L. Bertozzi, A. Flenner, and A.G. Percus, *Multiclass data segmentation using diffuse interface methods on graphs*, IEEE Transactions on Pattern Analysis and Machine Intelligence **36** (2014), 1600–1613.

[24] C. Garcia-Cardona, A. Flenner, and A.G. Percus, *Multiclass semi-supervised learning on graphs using Ginzburg-Landau functional minimization*, Advances in Intelligent Systems and Computing **318** (2015), 119–135.

[25] Z. Meng, E. Merkurjev, A. Koniges, and A.L. Bertozzi, *Hyperspectral Image Classification Using Graph Clustering Methods*, Image Processing On Line **7** (2017), 218–245.

[26] X. Luo and A.L. Bertozzi, *Convergence of the graph Allen–Cahn scheme*, Journal of Statistical Physics **167(3)** (2017), 934–958.

[27] J. Bosch, S. Klamt, and M. Stoll, *Generalizing diffuse interface methods on graphs: non-smooth potentials and hypergraphs*, arXiv preprint arXiv:1611.06094.

[28] B. Keetch and Y. van Gennip, *A Max-Cut approximation using a graph based MBO scheme*, in preparation

[29] K.A. Brakke, *The motion of a surface by its mean curvature*, Princeton University Press (1978)

[30] L. Bronsard and R.V. Kohn, *Motion by mean curvature as the singular limit of Ginzburg-Landau dynamics*, Journal of Differential Equations **90(2)** (1991), 211–237.

[31] L.C. Evans, *Convergence of an algorithm for mean curvature motion*, Indiana University Mathematics Journal **42(2)** (1993), 533-557.

[32] G. Barles, H.M. Soner, and P.E. Souganidis, *Front propagation and phase field theory*, SIAM Journal on Control and Optimization **31(2)** (1993), 439–469.

[33] G. Barles and C. Georgelin, *A simple proof of convergence for an approximation scheme for computing motions by mean curvature*, SIAM Journal of Numerical Analysis **32(2)** (1995), 484–500.

[34] Y. van Gennip, N. Guillen, B. Osting, and A.L. Bertozzi, *Mean Curvature, Threshold Dynamics, and Phase Field Theory on Finite Graphs*, Milan Journal of Mathematics **82(1)** (2014), 3–65.

[35] F. Almgren, J.E. Taylor, and L. Wang, *Curvature-driven flows: a variational approach*, SIAM Journal on Control and Optimization **31(2)** (1993), 387–438.

[36] S. Luckhaus and T. Sturzenhecker, *Implicit time discretization for the mean curvature flow equation*, Calculus of Variations and Partial Differential Equations **3(2)** (1995), 253–271.

[37] J.E. Taylor, *Anisotropic interface motion*, Mathematics of Microstructure Evolution **4** (1996), 135–148.

[38] T. Ohta and K. Kawasaki, *Equilibrium Morphology of Block Copolymer Melts*, Macromolecules **19** (1986), 2621–2632.

[39] Y. van Gennip, *An MBO scheme for minimizing the graph Ohta-Kawasaki functional*, in preparation

[40] Y. Peres and S. Sheffield, *Tug-of-war with noise: A game-theoretic view of the p-Laplacian*, Duke Mathematical Journal **145(1)** (2008), 91–120.

[41] P. L. Lions, *On the Schwarz alternating method*, I. In First international symposium on domain decomposition methods for partial differential equations (1988), 1–42.

[42] J. H. Bramble, J. E. Pasciak, J. P. Wang, and J. Xu, *Convergence estimates for product iterative methods with applications to domain decomposition*, Mathematics of Computation, **57(195)** (1991), 1–21.

[43] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM review, **34(4)** (1992), 581–613.

[44] T. F. Chan, and T. P. Mathew, *Domain decomposition algorithms*, Acta Numerica, **3** (1994), 61–143.

[45] A. Quarteroni, and A. Valli, *Domain decomposition methods for partial differential equations*, Numerical Mathematics and Scientific Computation, The Clarendon Press Oxford University Press, New York (1999), Oxford Science Publications.

[46] B. Smith, P. Bjorstad, P., and W. Gropp, *Domain decomposition: parallel multilevel methods for elliptic partial differential equations*, Cambridge University Press (2004).

[47] Y. Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM Journal on Optimization, **22(2)** (2012), 341–362.

[48] J.A. Acebron, M.P. Busico, P. Lanucara, and R. Spigler, *Domain decomposition solution of elliptic boundary-value problems via Monte Carlo and quasi-Monte Carlo methods*, SIAM J. Sci. Comput. **27(2)** (2005).

[49] Y. van Gennip, *Using evolving interface techniques to solve network problems*, in Mathematisches Forschungsinstitut Oberwolfach Report: Emerging Developments in Interfaces and Free Boundaries **6** (2017), 24–29.