*The Annals of Applied Probability*
2018, Vol. 28, No. 4, 2003–2062
https://doi.org/10.1214/17-AAP1313


# CONSISTENCY OF MODULARITY CLUSTERING ON RANDOM GEOMETRIC GRAPHS

BY ERIK DAVIS[1] AND SUNDER SETHURAMAN[1]

*University of Arizona*

Given a graph, the popular "modularity" clustering method specifies a partition of the vertex set as the solution of a certain optimization problem. In this paper, we discuss scaling limits of this method with respect to random geometric graphs constructed from i.i.d. points $\mathcal{X}_n = \{X_1, X_2, \dots, X_n\}$, distributed according to a probability measure $\nu$ supported on a bounded domain $D \subset \mathbb{R}^d$. Among other results, we show, via a Gamma convergence framework, a geometric form of consistency: When the number of clusters, or partitioning sets of $\mathcal{X}_n$ is a priori bounded above, the discrete optimal modularity clusterings converge in a specific sense to a continuum partition of the underlying domain $D$, characterized as the solution to a "soap bubble" or "Kelvin"-type shape optimization problem.

## CONTENTS



---

Received April 2016; revised February 2017.
[1]Supported in part by ARO W911NF-14-1-0179.
*MSC2010 subject classifications.* Primary 60D05; secondary 62G20, 05C82, 49J55, 49J45, 68R10.
*Key words and phrases.* Modularity, community detection, consistency, random geometric graph, Gamma convergence, Kelvin's problem, scaling limit, shape optimization, optimal transport, total variation, perimeter.

**1. Introduction.** One of the basic tasks in understanding the structure and function of complex networks is the identification of community structure or modular organization, where by a community we mean a subset of densely interconnected nodes, only sparsely connected to outsiders (cf. [32, 63]). A widely popular approach to community detection is the method of modularity clustering, introduced by Newman and Girvan (cf. [55, 57]), which specifies an optimal clustering, that is, a partition of the network as the solution of a certain optimization problem (see Section 2 for precise definitions). In particular, the method is used for a variety of networks arising in scientific contexts, including metabolic networks [43], epigenetic networks [53], brain networks [46], and networks encoding ecological [31] and political interactions [62].

On the other hand, a popular model of complex networks with geometric structure is the random geometric graph, where vertices are sampled from a geometric domain and edge weights are determined by a function of the distance between vertices [34, 52, 60]. We note, a well-studied case is the unweighted version, when the connectivity function is a threshold function of distance. These graphs are well-established mathematical models of various physical phenomena, such as continuum percolation. They have also found use in a number of applied settings, including the modeling of ad hoc wireless networks [10, 29, 45], protein–protein interaction networks [64], as well as the study of combinatorial optimization problems [25, 26]. In clustering studies of these graphs and their variants, the modularity functional is often used to assess the quality of the clustering obtained [5, 24].

In these contexts, it is a natural question to ask about the consistency of modularity clustering with respect to random geometric graphs. That is, one would like to know how the modularity clusterings converge or stabilize as the number of sampled data points or vertices grow, and how to characterize geometrically any limit clusterings. From the point of view of applications, establishing consistency is relevant to benchmarking performance. We will focus on a class of random geometric graphs, where a length scale is introduced in the connectivity function, so that distances between adjacent vertices shrink as the number of data points grow, and spatial scaling limits can be taken.

In this article, we study two questions: The first asks about the large limit behavior of the modularity functional on these random geometric graphs, evaluated on a fixed partition of the underlying geometric domain, a sort of "fixed-point" limit. The second question asks whether the optimal modularity clusterings for the discrete graphs converge, and if so to what geometric limit clustering.

With respect to the first question, when the fixed partition of the domain involves at most $K \geq 1$ sets, we show that the limit of the modularity functional, known to be a priori bounded between $-1$ and 1, as the sample size grows, is of the form $1 - 1/K$ plus a term involving the partition, which vanishes when the partition is "balanced", that is, when each set in the partition has the same volume with respect to an underlying measure (Theorem 2.1). As a corollary, we obtain, for these random geometric graphs, that the maximum of the modularity functional taken over all partitions, as the sample size grows, tends to 1 (Corollary 2.2). These limits prove, in the context of random geometric graphs, some heuristics given in the literature (cf. Section 2.4).

With respect to the second question, we show that, given a fixed upper bound $K$ on the number of clusters, the optimal modularity clusterings of the discrete random graphs converge in a distributional sense to an optimal clustering, satisfying a certain shape optimization problem (Theorem 2.3). This geometric continuum optimization problem, of interest in itself, is a form of Kelvin's problem: Informally, find a partition of the domain, where each set has the same volume, but where the perimeters between sets is minimized. Noting the first result above, it seems natural that a constraint specifying equal volumes would emerge in the limiting shape problem. Nonetheless, it seems intriguing that a form of Kelvin's shape optimization problem (cf. Section 2.2) would appear.

Previously, in the literature, a type of statistical consistency of modularity clustering for stochastic block models has been considered. In the stochastic block model, each data point is assigned a group from $K$ groups according to a probability $\pi$. Then, if the two points belong to groups $a$ and $b$, respectively, one puts an edge between them with probability $F_{a,b}$. Modularity clustering now gives an empirical group assignment to each data point. One of the main results shown is that the proportion of misclassification of empirical group assignments, with respect to the a priori assignment, vanishes in probability, Bickel and Chen [11] for the model above, Zhao, Levina and Zhu [80] for degree-corrected models, Rohe, Chatterjee and Yu [66] for high-dimensional models and Le, Levina and Vershynin [51] via low rank approximation.

In this context, the main focus of the article is to understand the geometry of optimal partitions, and other asymptotics with respect to the modularity functional. Our results represent limit results on a "geometric" form of consistency of the discrete modularity clusterings, which seem not to have been considered before. Moreover, as a corollary of this type of consistency, we show that the proportion of misclassification of the data points into sets given by the modularity functional, with respect to the limit optimization problem, vanishes a.s. (Corollary 2.4).

We use the recent framework of optimal transport introduced by García Trillos and Slepčev in [38], in the context of continuum limits of graph variational problems, to help relate point cloud empirical measures to absolutely continuous ones. We rewrite, after some manipulation, the modularity functional in terms of a "graph total variation" term and a "balance" term (Section 4). The proofs, with respect to the first question on asymptotics of the modularity functional of a particular clustering, make use of concentration estimates for these terms through $U$-statistics and other bounds.

On the other hand, with respect to the second question, we observe that maximizers of the modularity functional are the same as minimizers of an energy functional built as the sum of the "graph total variation" and "balance" terms (cf. Section 6.1). In this energy functional, the coefficient in front of the "balance" term diverges as the reciprocal of the length scale when the number of data points grows. In the limit, the soft penalty "balance" term becomes a hard constraint. We show convergence of a subsequence of minimizers to an optimizer of the limit problem by use a modified notion of Gamma convergence that we formulate (cf. Section 3.2), together with a compactness principle. For the "liminf" part of Gamma convergence, although we have to treat the balance constraint, the analysis of the graph total variation term follows from work in [38], which handles a similar expression.

However, dealing with the "balance" constraint represents a serious difficulty with respect to the "limsup" or "recovery sequence" part of the Gamma convergence setup. Without the constraint, one can form the recovery sequence by approximating with piecewise smooth partitions. However, such approximations become more complicated when also enforcing the "balance" constraint. But, the probabilistic result for the first question (Theorem 2.1), given for general measurable clusterings, already can be seen to yield a recovery sequence, with respect to the modified notion of Gamma convergence for random functionals. Interestingly, this notion of Gamma convergence has the same strength, in terms of yielding subsequential convergence, as the more usual formulation (cf. Remark 3.10). Perhaps of use in other problems, we observe that this more probabilistic approach offers a different perspective.

With respect to previous work on statistical clustering methods, consistency of $K$-means methods have been considered by Pollard in [61], and more recently, via Gamma convergence, by Thorpe, Theil, Johansen and Cade in [71]. On single linkage hierarchical clustering, consistency has been shown by Hartigan in [47]. On Fuzzy C-Means clustering, consistency has been considered by Sabin in [67]. With respect to spectral clustering, consistency has been considered by Belkin and Niyogi in [9], Von Luxburg, Belkin and Bousquet in [77] and García Trillos and Slepčev in [35], the last article, employing the framework of [38], as in this paper. We also mention, using Gamma convergence (or "epi-convergence") techniques, consistency of maximum likelihood and other estimators was studied by Wets in [78] and references therein.

Also, related, Arias-Castro and Pelletier [6] considered the consistency of the dimension reduction algorithm "maximum variance unfolding", and Arias-Castro, Pelletier and Pudlo [7] and García Trillos, Slepčev, von Brecht, Laurent and Bresson [39] studied the consistency of Cheeger and ratio cuts. Pointwise estimates between graph Laplacians and their continuum counterparts were considered by Belkin and Niyogi [8], Coifman and Lafon [20], Giné and Koltchinskii [40], Hein, Audibert and Von Luxburg [48] and Singer [69]. Also, spectral convergence in more general contexts was considered by Ting, Huang and Jordan [72] and Singer and Wu [70].

In addition, we mention there is a large literature on Gamma convergence of discrete lattice based variational expressions to continuum optimization problems (cf. Braides [13], Braides and Gelli [14] and references therein). More recently, see van Gennip and Bertozzi [73] which studies Gamma convergence of the Ginzburg–Landau graph based functionals to continuum limits.

The plan of the paper is the following. In Section 2, we define the random geometric graph model and state and discuss the two main results (Theorems 2.1 and 2.3) on modularity clustering; a brief outline of the proof the theorems is given in Section 2.5. In Section 3, we discuss preliminaries with respect to weak convergence topologies, optimal transport distances and Gamma convergence—we remark that the proof of Theorem 2.3 relies on this section, but the proof of Theorem 2.1 does not. In Section 4, we develop the modularity functional into a convenient form amenable to later analysis. In Sections 5 and 6, the proofs of the two main theorems are given respectively. Finally, in the Appendix, some technical calculations and proofs, referenced in previous sections, are collected.

## 2. Model and results.

We first introduce in Section 2.1 the modularity functional on graphs with weighted edges. In Section 2.2, we discuss a form of Kelvin's continuum shape optimization problem. In Sections 2.3, 2.4 and 2.5, we state our main theorems, make some remarks and give a brief outline of the proofs.

### 2.1. *Graph partitioning by modularity maximization.*

Let $\mathcal{G} = (X, W)$ be a weighted graph with vertex set $X := \{x_1, x_2, \dots, x_n\}$ and weight matrix $W$, where the entry $W_{ij} \geq 0$ denotes the weight of the (undirected) edge between vertex $x_i$ and $x_j$. The degree of vertex $x_i$ is $d_i := \sum_j W_{ij}$, and the total weight of the graph is $m := \frac{1}{2}\sum_i d_i = \frac{1}{2}\sum_{i,j} W_{ij}$. When convenient, we will refer to the vertex $x_i$ simply as "vertex $i$".

Let $\mathcal{U} = \{U_k\}_{k=1}^K$ be a partition of the vertex set $X$. We shall sometimes refer to the nontrivial sets $U_k$ of $\mathcal{U}$ as "clusters", and we denote by $|\mathcal{U}|$ the number of clusters. We may have $|\mathcal{U}| < K$, if one of the sets $U_k$ is empty. The partition $\mathcal{U}$ is therefore equivalent to an assignment $\{c_i\}_{i=1}^n$ of labels $c_i \in \{1, \dots, K\}$ to vertices, where $U_k = \{x_i : c_i = k\}$ for $1 \leq k \leq K$.

Modularity was originally introduced as a quantitative measure of the bias towards of vertices in a given cluster to be connected to other vertices in the same cluster [57]. Informally, the idea is to measure the proportion of edge weight between vertices in the same cluster, and compare it to the expected proportion if the edge weights were redistributed at random, according to a "null" model.

The total proportion of edge weight between vertices in the same cluster is

$$\frac{1}{2m}\sum_{i,j} W_{ij}\delta(c_i, c_j),$$

where $\delta(c_i, c_j)$ is the indicator that $c_i = c_j$, and the factor of $1/2$ arises because distinct vertex pairs appear twice in the sum.

Let $E_{ij}$ denote the expected edge weight between vertices $i$ and $j$ under a random redistribution model, which we specify below in different forms. Then the expected proportion of edge weight between vertices in the same cluster is

$$\frac{1}{2m}\sum_{i,j} E_{ij}\delta(c_i, c_j).$$

Then the modularity $Q$ is defined to be

$$Q(\mathcal{U}) = \frac{1}{2m}\sum_{i,j}(W_{ij} - E_{ij})\delta(c_i, c_j),$$

and one has $-1 \leq Q \leq 1$. The guiding thought is that a partition $\mathcal{U}$ with large modularity $Q(\mathcal{U})$ would represent a good clustering of the graph.

The most popular choice of null model, and the one originally introduced in [57], specifies $E_{ij} = \frac{d_i d_j}{2m}$. For unweighted graphs, where $W$ is the adjacency matrix, this choice corresponds to the *configuration model*. Namely, with respect to the vertex degrees $d_1, \dots, d_n$, consider the following distribution over graphs with $n$ vertices and $m$ edges: At each vertex $i$, place $d_i$ half-edges. Then, successively choose a pair of half-edges at random without replacement and connect them to form an edge with unit weight. Then, to dominant order, the expected number of edges between vertices $i$ and $j$ behaves as $\frac{d_i d_j}{2m}$.

Another natural choice corresponds to the *Erdős–Rényi model*, when $E_{ij} = \frac{2m}{n^2}$. Namely, for unweighted graphs, when $W$ again is the adjacency matrix, if $m$ edges are distributed uniformly, the expected number of edges between vertices $i$ and $j$ is $E_{ij} = \frac{2m}{n^2}$ to dominant order.

More generally, we may define $E_{ij}$ which interpolates in and among these two possibilities, weighting the degree structure in various ways. For $\alpha \in \mathbb{R}$, let $S = \sum_{i=1}^n d_i^\alpha$, and $E_{ij} = 2m\frac{d_i^\alpha d_j^\alpha}{S^2}$. Define the "$\alpha$"-modularity $Q$ to be

$$Q(\mathcal{U}) := \frac{1}{2m}\sum_{i,j}\left(W_{ij} - 2m\frac{d_i^\alpha d_j^\alpha}{S^2}\right)\delta(c_i, c_j). \tag{2.1}$$

When $\alpha = 0$ or $1$, this reduces to the modularity corresponding to the Erdős–Rényi model or the configuration model, respectively.

The *modularity maximization* problem is the following: Given a graph $\mathcal{G}$, find the partition for which the modularity $Q$ is maximized. In other words, one solves the following optimization problem:

$$\underset{\mathcal{U}}{\text{maximize}}\, Q(\mathcal{U}), \tag{2.2}$$

where the maximization occurs over all partitions $\mathcal{U}$ of the vertex set. One is particularly interested in optimal partitions $\mathcal{U}^* \in \arg\max_{\mathcal{U}} Q(\mathcal{U})$, representing a division of the graph into natural communities.

We also consider the following variant of problem (2.2), in which we restrict the partitions to have at most $K$ classes:

$$\begin{aligned} &\underset{\mathcal{U}}{\text{maximize}}\, Q(\mathcal{U}), \\ &\quad\text{subject to } |\mathcal{U}| \le K. \end{aligned} \tag{2.3}$$

We remark that finding a global maximum of the modularity is known to be NP-hard ([17]). Nonetheless, there are various algorithms to compute approximate maximizers, among them greedy algorithms (cf. [12, 19]), and relaxation methods (cf. [49, 56, 58, 79]). There is also a Potts model interpretation of modularity which offers another computational perspective (cf. [44, 65]).

2.2. *Geometric partitioning.* We now discuss a continuum partitioning problem, which will emerge as a scaled limit of the discrete graph partitioning optimization. Consider a domain $D \subset \mathbb{R}^d$, and a partition $\mathcal{U} = \{U_k\}_{k=1}^K$ of $D$. For what follows, we take $K \ge 1$ to be fixed.

Suppose that we have a probability measure $\mu$ on $D$. We say that a partition $\mathcal{U}$ is balanced with respect to the measure $\mu$ if each $U_k$ has equal $\mu$-measure, that is,

$$\mu(U_k) = \frac{1}{K}, \qquad k = 1, \dots, K.$$

Because $\mathcal{U}$ is a partition, any two distinct sets $U_k$ and $U_l$ are disjoint. However, if they are adjacent their boundaries will intersect nontrivially as a $d-1$ dimensional surface. When the boundaries of $\{U_k\}_{k=1}^K$ are piecewise-smooth, we may measure the size of the interface between $U_k$ and $U_l$, with respect to a density $\phi$ on $D$, by

$$\int_{\partial U_k \cap \partial U_l \cap D} \phi(x)\, d\mathcal{H}^{(d-1)}(x),$$

where $d\mathcal{H}^{(d-1)}$ denotes the Euclidean $(d-1)$-dimensional surface measure. The measure of the total interface or perimeter between the sets of $\mathcal{U}$ is therefore

$$\frac{1}{2} \sum_{1 \le k \ne l \le K} \int_{\partial U_k \cap \partial U_l \cap D} \phi(x)\, d\mathcal{H}^{(d-1)}(x), \tag{2.4}$$

noting the factor $1/2$ accounts for the fact that each distinct pair $k, l$ contributes twice to the sum.

Because $\partial U_k \cap D = \bigcup_{l \neq k} \partial U_k \cap \partial U_l \cap D$, the quantity (2.4) is equal to

$$\frac{1}{2} \sum_{1 \leq k \leq K} \int_{\partial U_k \cap D} \phi(x)\, d\mathcal{H}^{(d-1)}(x). \tag{2.5}$$

More generally, for partitions $\mathcal{U}$ consisting of measurable sets $\{U_k\}_{k=1}^K$, we may define the weighted perimeter of $U_k$ as follows:

$$\mathrm{Per}(U_k; \phi) := \mathrm{TV}(\mathbb{1}_{U_k}; \phi),$$

where $\mathrm{TV}(\mathbb{1}_{U_k}; \phi)$, defined below, is the weighted total variation of the indicator function $\mathbb{1}_{U_k}$. For sufficiently regular sets, this definition agrees with the informal notion of perimeter (2.5).

Let $L^\infty(D, \theta)$ and $L^p(D, \theta)$ be the usual spaces of functions $u$ where $\inf\{C : |u(x)| \leq C \text{ for } \theta\text{-a.e. } x\} < \infty$ and $\int_D |u(x)|^p \, d\theta(x) < \infty$ for $1 \leq p < \infty$, respectively. When $\theta$ is Lebesgue measure on $D$ and the underlying domain $D$ is understood, we abbreviate $L^p := L^p(D, \theta)$. The weighted total variation of a function $u \in L^1$ is

$$\mathrm{TV}(u; \phi) := \sup_{\substack{\Phi \in C_c^1(D; \mathbb{R}^d) \\ |\Phi(x)| \leq \phi(x)}} \int_D u(x) \,\mathrm{div}\, \Phi(x)\, dx,$$

where $C_c^1(D; \mathbb{R}^d)$ is the space of continuously differentiable, compactly supported vector fields on $D$, and $\mathrm{div}\, \Phi(x) = \sum_{i=1}^d \frac{\partial \Phi_i}{\partial x_i}$. In our later applications, the density $\phi$ will be bounded above and below on $D$ by positive constants. In this case, the weighted total variation has many of the same properties as the total variation with respect to the uniform density $\mathbb{1}_D$, discussed in detail in Chapter 3 of [3].

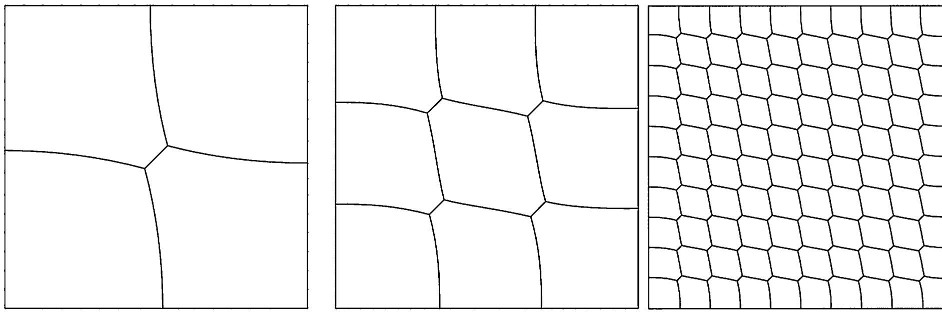
Consider now the following geometric partitioning problem. Among all balanced $K$-partitions of $D$, choose that which minimizes the total perimeter of its sets:

$$\begin{aligned} &\underset{\mathcal{U}}{\text{minimize}} \frac{1}{2} \sum_{k=1}^K \mathrm{Per}(U_k; \phi) \\ &\text{subject to } \mu(U_k) = \frac{1}{K}, k = 1, \ldots, K. \end{aligned} \tag{2.6}$$

See Figure 1 for the behavior when $D$ is a square.

It is clear, by dividing $D$ in terms of a moving hyperplane, that balanced $K$-partitions with finite perimeter exist. Also, depending on $D$, the problem may or may not have a unique solution (modulo relabeling of sets): Suppose $\phi = 1$ and $d\mu = dx$. When $D$ is a long, thin rectangle in $\mathbb{R}^2$, there is only one perimeter minimizing division into two sets of equal area. However, when $D$ is a disc in $\mathbb{R}^2$, there are infinitely many such divisions, given by cuts of the disc along a diameter.

The general problem (2.6) can be seen as a type of "soap bubble" problem. It can also be seen as a bounded domain form of Kelvin's problem: Find a tiling



(a) $K = 4$ (b) $K = 9$. (c) $K = 100$.

FIG. 1. *Local minimizers of problem* (2.6) *on* $D = (0, 1)^2$, *with* $\phi = 1$, $d\mu = dx$, *produced using* [16].

of $\mathbb{R}^d$ where each tile has unit $\mu$-volume so that the $\phi$-perimeter between tiles is minimized. We note in passing, when $\phi = 1$ and $d\mu = dx$, in $d = 2$, this problem has been solved in terms of hexagonal tiles, and is the subject of much research in $d \geq 3$. See Morgan [54], which discusses these and other related optimizations.

We remark, when $\phi = 1$ and $d\mu = dx$, the problem (2.6) has been considered in the literature. Cañete and Ritore [18] have studied minimal partitions of the disc into three regions, and in this context proved that the regions are bounded by circular arcs making perpendicular contact with the boundary of the disc and meeting at a 120 degree triple point within. Some conjectures about minimizers for larger values of $K$ and other domains are presented in Cox and Flikkema [21]. Oudet [59] derives a numerical algorithm, via Gamma convergence, for computing approximate solutions.

2.3. *Results*. Before stating the two main theorems, we first define the random geometric graphs under consideration. We make the following standing assumptions throughout on the domain $D \subset \mathbb{R}^d$, ground measure $\nu$ on $D$, and underlying edge weight structure.

(D) $D$ is a bounded, open, connected subset of $\mathbb{R}^d$ with Lipschitz boundary.

(M) In $d \geq 1$, $\nu$ is a probability measure on $D$ with density $\rho$ such that $\rho$ is Lipschitz and bounded above and below by positive constants.

Further, in $d = 1$, $\rho$ satisfies additional conditions: (a) $\rho$ is differentiable on $D := (c, d)$ and (b) $\rho$ is increasing in some interval with left endpoint $c$ and decreasing in some interval with right endpoint $d$.

Let $\{X_i\}_{i \in \mathbb{N}}$ be a collection of i.i.d. samples from $\nu$, and define $\mathcal{X}_n = \{X_i\}_{i=1}^n$. We will denote by $\mathbb{P}$ and $\mathbb{E}$ the probability and expectation with respect to the underlying probability space.

The random geometric graph is constructed from the points $\mathcal{X}_n$ through a kernel function $\eta : \mathbb{R}^d \to \mathbb{R}$, where $\eta$ satisfies:

(K1) $\int_{\mathbb{R}^d} \eta(x)\,dx = 1$.
(K2) $\eta$ is radial and nonincreasing, that is, $\eta(z) \le \eta(z_0)$ if $\|z\| \ge \|z_0\|$.
(K3) $\eta(0) > 0$ and $\eta$ is continuous at 0.
(K4) $\eta$ is compactly supported.

There is a large class of kernels admitted under assumptions (K1)–(K4), including the kernel associated with the well-known random geometric graph, where $\eta$ is the indicator function of a ball.

Between vertices $X_i, X_j$, in terms of a parameter $\varepsilon_n$, we put an edge with weight

$$W_{ij} = \begin{cases} \dfrac{1}{\varepsilon_n^d}\eta\left(\dfrac{X_i - X_j}{\varepsilon_n}\right) =: \eta_{\varepsilon_n}(X_i - X_j) & \text{if } i \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

The parameter $\varepsilon_n > 0$ serves as a length scale. For example, if the support of $\eta$ is contained in a ball of radius one, then two vertices $X_i$ and $X_j$ are connected by an edge only if they are separated by a distance no more than $\varepsilon_n$ (cf. Figure 2). Since the modularity functional $Q$ is the same under weights $W$ and $cW$, for $c > 0$, the factor $\varepsilon_n^{-d}$ in (2.7) was chosen so that the expected value of $W_{ij}$ is of order 1.

Under all circumstances, we have $\varepsilon_n \to 0$, although the specific rate at which $\varepsilon_n$ vanishes will depend on the dimension $d$, as well as the parameter $\alpha$. There are in fact two different sets of assumptions, (I1) being more restrictive than (I2), corresponding to our two main results. We first mention typical examples, satisfying the assumptions. Taking $\varepsilon_n = n^{-\beta}$, with $\beta > 0$, conditions (I1) and (I2) will hold if respectively

$$\beta < \begin{cases} 2/(d+1) & \text{if } \alpha = 0, 1, \\ 1/(d+1) & \text{if } \alpha \neq 0, 1, \end{cases} \quad \text{and} \quad \beta < \begin{cases} \min(1/d, 1/2), & \text{if } \alpha = 0, 1, \\ 1/(d+1), & \text{if } \alpha \neq 0, 1. \end{cases}$$

More precisely, (I1) and (I2) are the following:

(I1) When $\alpha = 0, 1$, we suppose that $\sum_{n=1}^{\infty} \exp(-n\varepsilon_n^{(d+1)/2}) < \infty$. However, when $\alpha \neq 0, 1$, we suppose that $\sum_{n=1}^{\infty} n \exp(-n\varepsilon_n^{d+1}) < \infty$.

FIG. 2. *Random geometric graph, constructed from* $n = 200$ *uniformly distributed points on the strip* $D = (0, 4) \times (0, 1)$ *with connectivity function* $\eta = \mathbb{1}_{|x|<1}$ *and* $\varepsilon = n^{-0.3}$.

(I2) When $\alpha = 0$ or $\alpha = 1$, we suppose that

$$\lim_{n\to\infty} \frac{(\log\log n)^{1/2}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \qquad \text{if } d = 1,$$

$$\lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \qquad \text{if } d = 2,$$

$$\lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} = 0 \qquad \text{if } d \geq 3.$$

However, when $\alpha \neq 0, 1$, we suppose that $\sum_{n=1}^{\infty} n \exp(-n\varepsilon_n^{d+1}) < \infty$.

In Section 2.4, we discuss motivations behind these assumptions in more detail.

Now, given the set $\mathcal{X}_n$ and a sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$, denote by $W_n$ the weight matrix with entries given by (2.7), and denote by $\mathcal{G}_n = (\mathcal{X}_n, W_n)$ the corresponding weighted random geometric graph. We let

$Q_n$ denote the modularity functional (2.1) corresponding to $\mathcal{G}_n$.

Consider a partition $\mathcal{U} = \{U_k\}_{k=1}^K$ of the domain $D$. For any $n \geq 1$, this induces a partition $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ of the sample $\mathcal{X}_n$, where for $1 \leq k \leq K$,

$$U_{n,k} = U_k \cap \mathcal{X}_n.$$

THEOREM 2.1 (Asymptotics). *Suppose* $\{\varepsilon_n\}_{n\in\mathbb{N}}$ *satisfies condition* (I1). *Fix* $K \geq 1$ *and let* $\mathcal{U} = \{U_k\}_{k=1}^K$ *be a partition of* $D$ *where each* $U_k$ *is a subset of finite perimeter*, $\mathrm{Per}(U_k; \rho^2) < \infty$. *Let* $\mathcal{U}_n$ *be the induced partition of* $\mathcal{X}_n$ *for* $n \geq 1$. *Then the modularity* $Q_n(\mathcal{U}_n)$ *satisfies*

$$1 - 1/K - Q_n(\mathcal{U}_n) \xrightarrow{a.s.} \sum_{k=1}^{K} \big(\mu(U_k) - 1/K\big)^2, \tag{2.8}$$

*as* $n \to \infty$, *where* $d\mu(x) = \frac{\rho^{1+\alpha}(x)\,dx}{\int_D \rho^{1+\alpha}(x)\,dx}$.

*Further, if* $\sum_{k=1}^K (\mu(U_k) - 1/K)^2 = 0$, *we have*

$$\frac{1 - 1/K - Q_n(\mathcal{U}_n)}{\varepsilon_n} \xrightarrow{a.s.} C_{\eta,\rho} \sum_{k=1}^{K} \mathrm{Per}\big(U_k; \rho^2\big), \tag{2.9}$$

*as* $n \to \infty$, *where*

$$C_{\eta,\rho} = \frac{\int_{\mathbb{R}^n} \eta(x)|x_1|\,dx}{2\int_D \rho^2(x)\,dx} \quad \textit{and} \quad x = (x_1, \ldots, x_d).$$

One way to interpret these law of large numbers limits is that the nonnegative quantity, a "balance" term,

$$\sum_{k=1}^{K} \left(\mu(U_k) - \frac{1}{K}\right)^2,$$

is a measure of how balanced the partition $\mathcal{U}$ is with respect to the measure $\mu$. In our model, the limiting modularity of a partition is predominantly determined by the number of clusters and the extent to which they are balanced.

We state a corollary of Theorem 2.1, which follows by considering balanced partitions where $K$ is not restricted.

COROLLARY 2.2. *Suppose the assumptions of Theorem 2.1 are satisfied. Let $Q_n^* = \max_{\mathcal{U}_n} Q_n(\mathcal{U}_n)$ denote the maximum modularity associated to $\mathcal{G}_n$, where the maximum is over all partitions $\mathcal{U}_n$ of $\mathcal{X}_n$. Then we have, as $n \to \infty$,*

$$Q_n^* \xrightarrow{a.s.} 1.$$

Our second main result is a characterization of the behavior of optimal clusterings $U_n^* \in \arg\max_{|\mathcal{U}_n| \le K} Q_n(\mathcal{U}_n)$, as $n \to \infty$. To this end, we introduce a suitable notion of convergence.

To a sequence of sets $\{U_n\}_{n\in\mathbb{N}}$ with $U_n \subset \mathcal{X}_n$, we associate the "graph measures" $\{\gamma_n\}_{n\in\mathbb{N}}$, where $\gamma_n = \frac{1}{n}\sum_{i=1}^n \upsilon_{(X_i, \mathbb{1}_{U_n}(X_i))}$, and $\upsilon_{\cdot}$ denotes a point mass. In words, the measure $\gamma_n$ is the distribution of the graph of $\mathbb{1}_{U_n}$ under the empirical measure $\nu_n$ on $\mathcal{X}_n$. Let also $U \subset D$ and define $\gamma$ as the distribution of the graph of $\mathbb{1}_U$ under $\nu$. We will write, with respect to a realization of $\{X_i\}_{i\in\mathbb{N}}$, that, as $n \to \infty$,

$$U_n \text{ converges } \textit{weakly} \text{ (denoted } \xrightarrow{w}\text{) to } U \text{ if } \gamma_n \xrightarrow{w} \gamma. \tag{2.10}$$

Correspondingly, consider a sequence of partitions $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ of $\mathcal{X}_n$, and a partition $\mathcal{U} = \{U_k\}_{k=1}^K$ of $D$. Let $\mathrm{Sym}(K)$ denote the permutations of $\{1, \dots, K\}$. Since the sets in the collections $\mathcal{U}_n$ and $\mathcal{U}$ are unordered, we say that, as $n \to \infty$,

$$\mathcal{U}_n \text{ converges } \textit{weakly} \text{ (denoted } \xrightarrow{w}\text{) to } \mathcal{U}$$

if there exists a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ of permutations in $\mathrm{Sym}(K)$ such that

$$\gamma_{n,\pi_n k} \xrightarrow{w} \gamma_k \qquad \text{for } k = 1, \dots, K. \tag{2.11}$$

THEOREM 2.3 (Optimal Clusterings). *Suppose $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfies condition* (I2). *Let $\mathcal{U}_n^* \in \arg\max_{|\mathcal{U}_n| \le K} Q_n(\mathcal{U}_n)$ be an optimal partition of $\mathcal{X}_n$, for $n \ge 1$. Let also, with respect to problem* (2.6), *$\phi = \rho^2$ and $d\mu = \rho^{1+\alpha}\,dx / \int_D \rho^{1+\alpha}(x)\,dx$.*

*If $\mathcal{U}^*$ is the unique solution* (*modulo reordering of its constituent sets*) *of problem* (2.6), *then a.s. $\mathcal{U}_n^*$ converges weakly to $\mathcal{U}^*$, in the sense of* (2.11).

*If there is more than one solution to the problem* (2.6), *then a.s. $\mathcal{U}_n^*$ converges weakly along a subsequence, in the sense of* (2.11), *to a solution $\mathcal{U}^*$ of* (2.6).

Figure 3 illustrates this result. We remark that, in the proof of Theorem 2.3, we will in fact show an equivalent form of convergence, in terms of Wasserstein, Kantorovich–Rubenstein-type $(TL^1)^K$ distances, via a Gamma convergence statement (Theorem 6.1) for certain energy functionals.

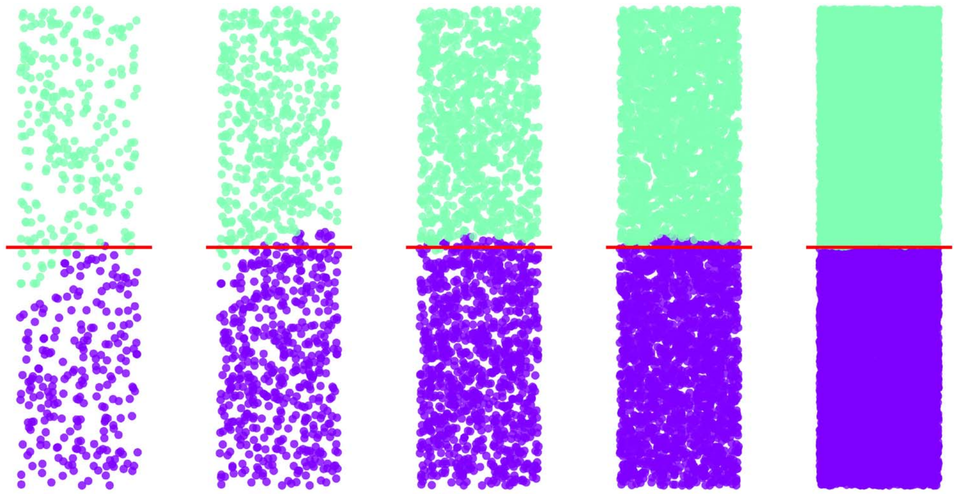



FIG. 3. *Binary partitions produced by modularity clustering of random geometric graphs with various values of n on the domain* $D = (0, 1) \times (0, 4)$, *using the algorithm of* [58]. *Here* $\alpha = 1$, *the density* $\rho$ *is uniform*, $\eta = \mathbb{1}_{|x|<1}$, *and* $\varepsilon_n = n^{-0.3}$. *The red lines indicate the optimal cut associated with the continuum partitioning problem for* $K = 2$.

We also note that the parameter $\alpha$, which parametrizes a family of null models in the modularity functional (2.1), appears in the balancing measure $\mu$ associated with the continuum problem (2.6). As $\alpha$ increases, the measure $\mu$ puts more mass near the modes of the density $\rho$, altering the optimal partitions as illustrated in Figure 4. For the choice $\alpha = -1$, the balancing measure $\mu$ is independent of the density $\rho$. Still, the weighted perimeter depends on $\rho$. So, in part (B) of Figure 4, both sets have area $3/2$, but the optimal interface is curved, due to the form of $\rho$.

We also observe that the notion of convergence in Theorem 2.3 allows to capture various statistics on the discrete partitions; for instance, we show that the proportion of misclassified points vanishes almost surely.

COROLLARY 2.4. *Consider the setting of Theorem* 2.3. *Let* $\{\mathcal{U}^*_{n_m}\}_{m\in\mathbb{N}}$ *be a subsequence of optimal* $Q_{n_m}$*-modularity partitions of the sample space which a.s. converges weakly, in the sense of* (2.11), *to a solution* $\mathcal{U}^*$ *of* (2.6). *Then the proportion of correctly classified points converges to* 1. *That is, a.s. as* $m \to \infty$,

$$p_{n_m} := \min_{1\le k\le K} \frac{|\{x \in \mathcal{X}_{n_m} : x \in U^*_k \cap U^*_{n_m,\pi_{n_m}(k)}\}|}{|\{x \in \mathcal{X}_{n_m} : x \in U^*_k\}|} \to 1.$$

PROOF. It is enough to show, with respect to a sequence of sets $V_m \subset \mathcal{X}_{n_m}$ and a set $V \subset D$, of positive Lebesgue measure and finite perimeter, where a.s.

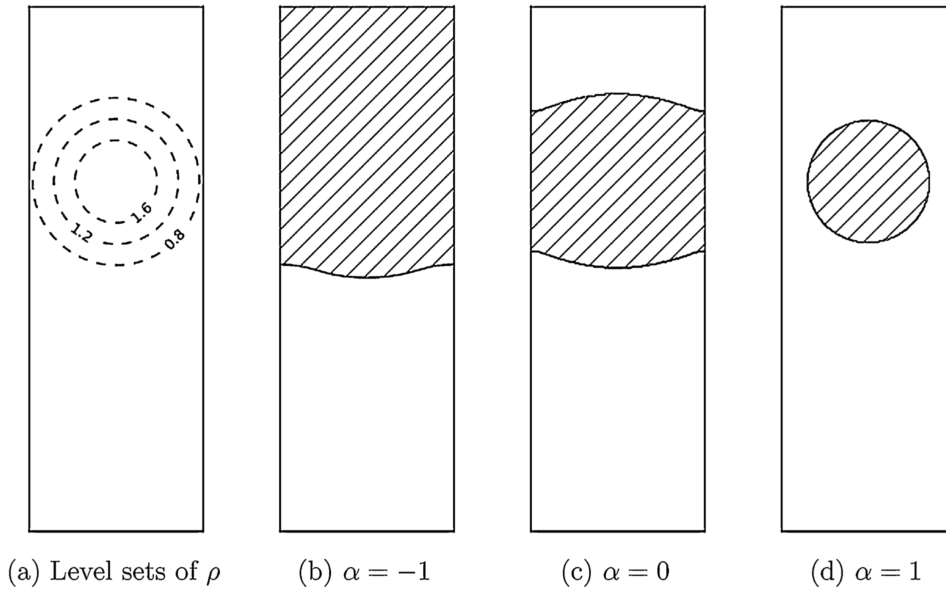



FIG. 4. *Behavior of problem* (2.6) *for* $K = 2$ *and various* $\alpha$ *on* $D = (0, 1) \times (0, 3)$, *computed using a variant of the algorithm in* [59]. *Here*, $\phi = \rho^2$, $d\mu = \rho^{1+\alpha}(x)\,dx / \int_D \rho^{1+\alpha}(x)\,dx$, *for* $\rho(x) \propto \min(2\exp(-4\|x - z\|^2), 1/2)$, *with* $z = (1/2, 2)$.

$V_m \stackrel{w}{\to} V$, in the sense of (2.10), that

$$\frac{|\{x \in \mathcal{X}_{n_m} : x \in V \cap V_m\}|}{|\{x \in \mathcal{X}_{n_m} : x \in V\}|} \to 1 \qquad \text{a.s.}$$

as $m \to \infty$. Then the statement in the corollary would follow by application of this limit with $V_m = U^*_{n_m, \pi_{n_m}(k)}$ and $V = U^*_k$ for $1 \le k \le K$.

In terms of the measures $\gamma_m$ and $\gamma$, which govern the graphs of $\mathbb{1}_{V_m}$ and $\mathbb{1}_V$ under $\nu_n$ and $\nu$ respectively, we can write

$$\frac{|\{x \in \mathcal{X}_{n_m} : x \in V \cap V_m\}|}{|\{x \in \mathcal{X}_{n_m} : x \in V\}|} = \frac{\int_{D\times\mathbb{R}} \mathbb{1}_V(x) y \, d\gamma_m(x, y)}{\int_{D\times\mathbb{R}} \mathbb{1}_V(x) \, d\gamma_m(x, y)}.$$

Since a.s. $\gamma_m \stackrel{w}{\to} \gamma$ as $m \to \infty$, by approximating $(x, y) \mapsto \mathbb{1}_V(x) y$ and $(x, y) \mapsto \mathbb{1}_V(x)$ by bounded, continuous functions, we have

$$\frac{\int_{D\times\mathbb{R}} \mathbb{1}_V(x) y \, d\gamma_m(x, y)}{\int_{D\times\mathbb{R}} \mathbb{1}_V(x) \, d\gamma_m(x, y)} \stackrel{a.s.}{\longrightarrow} \frac{\int_{D\times\mathbb{R}} \mathbb{1}_V(x) y \, d\gamma(x, y)}{\int_{D\times\mathbb{R}} \mathbb{1}_V(x) \, d\gamma(x, y)} = 1,$$

as $m \to \infty$, concluding the argument. □

2.4. *Discussion*. 1. As alluded to in the Introduction, the phenomena shown in Corollary 2.2 for random geometric graphs has been considered before in other models. Indeed, in [44] the authors provide heuristic arguments for the limiting behavior $Q_n^* \to 1$ under two regimes: (i) when the graphs are regular lattices, and (ii) when the graphs are Erdős–Rényi graphs with edge probability $p = 2/n$. In [42], the authors derive the limiting behavior $Q_n^* \to 1$ under a sparse graph model, in which modules of some characteristic size are adjoined to the graph. Further, these asymptotics are consistent with the empirical results associated with large real-world graphs [12].

2. It has been observed in the literature that modularity optimization may fail to identify clusters smaller than a certain level, depending on the total size and interconnectedness of the graph. In other words, modularity possesses a "resolution limit" in terms of its clustering (cf. [33, 42]). An extreme example is when the graph contains a pair of cliques (complete subgraphs) connected by a single edge, but modularity would lump them into a common cluster (cf. Figure 3 of [33]).

In [65], the authors consider a variant of the modularity given by

$$Q^\lambda = \frac{1}{2m}\sum_{i,j}\left(W_{ij} - \lambda\frac{d_i d_j}{2m}\right)\delta(c_i, c_j), \tag{2.12}$$

where $\lambda$ is a parameter. In [50], the parameter $\lambda$ is related to the resolution limit phenomena: Namely, higher values of $\lambda$ allow for smaller cluster sizes.

The methods used to prove Theorem 2.3 give the following asymptotic behavior of optimal $Q^\lambda$ modularity clusterings: When $\lambda = \lambda_n := \kappa\varepsilon_n^\gamma$ is scaled with $n$, for $\gamma \geq 0$ and $\varepsilon_n$ satisfying (I2), three distinct possibilities arise for the limiting problem. When $0 \leq \gamma < 1$, the continuum partitioning problem remains as it is in (2.6). When $\gamma = 1$, the hard constraints $\mu(U_k) = 1/K$ for $1 \leq k \leq K$ of the limiting problem get replaced by a soft balancing condition, resulting in

$$\underset{\mathcal{U}}{\text{minimize}}\frac{1}{2}\sum_{k=1}^{K}\text{Per}(U_k;\phi) + \kappa\sum_{k=1}^{K}\big(\mu(U_k) - 1/K\big)^2. \tag{2.13}$$

When $\gamma > 1$, the continuum problem degenerates to a perimeter minimization problem with no balancing condition, which has as its solution a single global cluster $D$ (and the other $K-1$ sets being empty).

3. One can ask also about the reasons behind assumptions (I1) and (I2). With respect to Theorem 2.1, a lower bound for $\varepsilon_n$ should be informed by the fluctuations of the functional. In fact, the variance of $Q_n(\mathcal{U}_n)$ can be seen to be of order $(n^2\varepsilon_n^{d+1})^{-1}$ when $\alpha = 0, 1$ [by a computation with formula (4.7)]; so, under condition (I1), the variance vanishes. However, when $\alpha \neq 0, 1$, a worse bound is useful to control the nonlinearity of the functional. In particular, an univariate Bernstein concentration bound is applied twice when $\alpha \neq 0, 1$, whereas a stronger (in this context) $U$-statistics bivariate concentration bound is only used once when $\alpha = 0, 1$.

On the other hand, assumption (I2) in Theorem 2.3 is informed by the connectivity radius of the random geometric graphs. For instance, if $\varepsilon_n$ were to vanish too quickly, the underlying graphs would contain $O(n)$ disconnected components (cf. Theorem 13.25 in [60]). Then, presumably, one would be able to find a $\mathcal{U}_n^*$ such that $(1/2 - Q_n(\mathcal{U}_n^*))/\varepsilon_n \xrightarrow{a.s.} 0$ and consequently obtain a continuum cluster point $\mathcal{U}^*$. This is a contradiction, as the resulting partition $\mathcal{U}^*$ would have zero perimeter—in other words, one of the sets in $\mathcal{U}^*$ would be $D$ itself—and so could not satisfy the balance conditions. This is a version of the argument in Remark 1.6 of [38].

The threshold that $\varepsilon_n$ should be larger than, for the graphs to be connected, is known: In $d \geq 1$, it is of order $(\log(n))^{1/d}/n^{1/d}$ (cf. [60]). Viewed from this lens, condition (I2) is more optimal when $\alpha = 0, 1$ than when $\alpha \neq 0, 1$—we remark the $\alpha = 0, 1$ conditions in fact reflect the bounds on the optimal transport maps in Proposition 3.2, as can be seen in the argument of Theorem 2.3. As alluded above, when $\alpha \neq 0, 1$, the difficulty is in analyzing the nonlinear "balance" term in the functional, whereas the "total variation" term is handled more optimally.

It is not clear if (I1), (I2), in the case $\alpha \neq 0, 1$, are close to optimal or artifacts of the technical estimates. It would be of interest to investigate further the optimality of these conditions.

4. We briefly discuss the assumptions on $\rho$, $D$, and $\eta$. The proof of Theorem 2.3 makes use of certain "transport maps" (cf. Proposition 3.2). For $d \geq 2$, we use the optimal transport results of [37], which require that $\rho$ be bounded above and below by positive constants, and that $D$ is sufficiently nice. On the other hand, for $d = 1$, it is not necessary that $\rho$ be bounded below to define a suitable transport map; here, the technical condition required is (A.6). However, in all dimensions, we require a lower bound on $\rho$, as this enables us to handle the general $\alpha \neq 0, 1$ case via a Lipschitz inequality for the map $x \mapsto x^\alpha$ (cf. Lemma 5.6). The boundedness of $D$ is also used in several intermediate technical results. The Lipschitz continuity of $\rho$ is used for handling the "balance" term in the proof of Theorem 2.3 (principally in Lemma 5.3, by way of Lemma A.13). We remark that, by a simple modification of the proof, this condition could be weakened to Hö continuity, with exponent greater than $1/2$.

With respect to $\eta$, the radial and monotone assumptions are convenient in relating certain graph functionals to their nonlocal analogues (cf. Lemma 6.3). The continuity at zero is used in the proof of the compactness property, Lemma 6.11. Finally, we remark that the compact support of $\eta$ allows to analyze behavior near the boundary of $D$, although this assumption could be weakened to a suitable condition on the decay of $\eta$ at infinity.

2.5. *Brief outline of the proofs of Theorems* 2.1 *and* 2.3. The arguments for both the main theorems rely on a decomposition of the modularity functional into "graph total variation" and "quadratic balance" terms, done in Section 4. We remark the identification of the "quadratic balance" term seems new (cf. the different but related decomposition in [49]), and its analysis will be an integral part of later asymptotics.

To prove Theorem 2.1, in Section 5, we use concentration bounds for $U$-statistics to compare the random "graph total variation" and "quadratic balance" parts of the modularity functional to their mean-values, which are separately analyzed. The argument of Theorem 2.1 does not rely on Section 3, which are preliminaries for Theorem 2.3.

The approach to prove Theorem 2.3 in Section 6 is to formulate both the modularity clustering problem (2.3) and the continuum partitioning problem (2.6) as optimization problems on a common metric space $(TL^1(D))^K$. In Section 3, we discuss the $TL^1(D)$ topology and framework of García Trillos and Slepčev in [38] to study weak convergence of "graph measures". Connections with optimal transportation maps are also made. In addition, a modified form of Gamma convergence for random functionals, different from other versions in the literature, is introduced (cf. Remark 3.10), which is a main vehicle behind the argument of Theorem 2.3, and which may be of its own interest.

Although we wish to maximize the modularity, it will be convenient to pose an equivalent problem of minimizing a related energy $E_n$ in Section 6.1. In particular, the maximum modularity clusterings of the graph $\mathcal{G}_n = (\mathcal{X}_n, W_n)$ will be related to the solution of

$$\underset{\mathcal{V}\in(TL^1(D))^K}{\text{minimize}}\ E_n(\mathcal{V}),$$

and the optimal partitions of Problem (2.6) will be related to the solution of

$$\underset{\mathcal{V}\in(TL^1(D))^K}{\text{minimize}}\ E(\mathcal{V}).$$

We state in Theorem 6.1 that the random functionals $E_n$ Gamma converge to $E$, in the modified sense as alluded to above. The two limits in the Gamma convergence formulation, "liminf" and "recovery", are shown in Sections 6.2 and 6.3. We note that the "recovery" limit relies on the convergence results (5.29), (5.30) and (5.31); this is the only dependence on Section 5 in Section 6.

In Section 6.4, we state a compactness property for the functionals $E_n$, which together with Theorem 6.1, will imply that the minimizers of $E_n$ converge subsequentially in $(TL^1)^K$ to a minimizer of $E$, and thereby prove Theorem 2.3 at the end of the section.

## 3. Preliminaries for Theorem 2.3.

As a reference for later use in Section 6, we discuss in Section 3.1 the $TL^1$ topology and framework, introduced by García Trillos and Slepčev in [38], and connections to weak convergence of graph measures and optimal transportation maps. Then, in Section 3.2, we define a variant of Gamma convergence for random energy functionals that we will use to prove Theorem 2.3.

3.1. *$TL^1$ topology and framework.* Given a measurable space $S \subset \mathbb{R}^d$, we let $\mathcal{B}(S)$ denote the Borel $\sigma$-algebra on $S$, and similarly let $\mathcal{P}(S)$ denote the set of Borel probability measures on $S$. Also, given two spaces, $S_1$ and $S_2$, a measurable map $T : S_1 \to S_2$, and a measure $\mu \in \mathcal{P}(S_1)$, we define the *push-forward* $T_\sharp \mu \in \mathcal{P}(S_2)$ by

$$T_\sharp \mu(A) = \mu\big(T^{-1}(A)\big) \qquad \text{for } A \in \mathcal{B}(S_2).$$

In particular, $T_\sharp \mu$ is the distribution of $TX$ where $X$ has distribution $\mu$.

Given measures $\mu, \theta \in \mathcal{P}(S)$, recall that a coupling between $\mu$ and $\theta$ is a probability measure $\pi$ on $S \times S$ such that the marginal with respect to the first variable is $\mu$, and the marginal with respect to the second variable is $\theta$. Consider the set of couplings $\Gamma(\mu, \theta) := \{\pi \in \mathcal{P}(S \times S) : \pi(U \times S) = \mu(U) \text{ and } \pi(S \times U) = \theta(U)\ \forall U \in \mathcal{B}(S)\}$. Define the distance on $\mathcal{P}(S)$ by

$$d_{1,S}(\mu, \theta) := \inf_{\pi \in \Gamma(\mu,\theta)} \int_S |x - y| \, d\pi(x, y).$$

This is a metric on $\mathcal{P}_1(S)$, the subset of probability measures in $\mathcal{P}(S)$ with finite first moment. This metric is known as "earth mover's" distance or the 1-Wasserstein distance.

When $S$ is complete, a case of a more general result (see Theorem 6.9 of [76]) is the following: Let $\{\mu_n\}_{n \in \mathbb{N}}$ and $\mu$ be measures in $\mathcal{P}_1(S)$. Then, as $n \to \infty$,

$$\mu_n \xrightarrow{w} \mu \text{ and first moments converge} \quad \text{iff} \quad d_{1,S}(\mu_n, \mu) \to 0. \tag{3.1}$$

Now, as in [38], to understand weak convergence of "graph measures", define the space $TL^1(S)$ by

$$TL^1(S) := \big\{(\mu, f) : \mu \in \mathcal{P}(S), \|f\|_{L^1(S,\mu)} < \infty\big\},$$

and, for $(\mu, f)$ and $(\theta, g)$ in $TL^1(S)$, define the distance

$$d_{TL^1,S}\big((\mu, f), (\theta, g)\big) := \inf_{\pi \in \Gamma(\mu,\theta)} \int \int_{S \times S} |x - y| + \big|f(x) - g(y)\big| \, d\pi(x, y).$$

One may identify an element $(\mu, f) \in TL^1(S)$ with a graph measure $(Id \times f)_\sharp \mu \in \mathcal{P}(S \times \mathbb{R})$, whose support is contained in the graph of $f$. Consider now, with respect to $(\mu, f), (\theta, g) \in TL^1(S)$, the graph measures $\gamma = (Id \times f)_\sharp \mu$, $\widetilde{\gamma} = (Id \times g)_\sharp \theta \in \mathcal{P}(S \times \mathbb{R})$. It may be seen that

$$d_{TL^1,S}\big((\mu, f), (\theta, g)\big) = \inf_{\pi \in \Gamma(\gamma,\widetilde{\gamma})} \iint_{(S\times\mathbb{R})\times(S\times\mathbb{R})} |x-y| + |s-t| \, d\pi\big((x, s), (y, t)\big),$$

and hence

$$d_{1,S\times\mathbb{R}}(\gamma, \widetilde{\gamma}) \text{ is equivalent to } d_{TL^1,S}\big((\mu, f), (\theta, g)\big), \tag{3.2}$$

in the sense that $d_{1,S\times\mathbb{R}}$ is bounded above and below by positive multiples of $d_{TL^1,S}$ (cf. Remark 3.2 and Proposition 3.3 of [38]).

We now restrict $S$ to be the bounded domain $D \subset \mathbb{R}^d$ introduced in Section 2.3. We will abbreviate $TL^1 := TL^1(D)$. Then, for $(\mu, f) \in TL^1$, the graph measure $\gamma = (Id \times f)_\sharp \mu$ belongs to $\mathcal{P}_1(D \times \mathbb{R})$ in that it has a finite first moment since $\int_D |x| + |f(x)|\, d\mu(x) < \infty$. Hence, by (3.2), $TL^1$ can be viewed as a metric space with metric $d_{TL^1,D}$.

With respect to graph measures $\gamma', \gamma'' \in \mathcal{P}_1(D \times \mathbb{R})$, consider their extensions $\bar{\gamma}'$ and $\bar{\gamma}''$ to $\overline{D \times \mathbb{R}}$ by setting $\bar{\gamma}'(\partial D \times \mathbb{R}) = \bar{\gamma}''(\partial D \times \mathbb{R}) = 0$. Then the distance

$$d_{1,\overline{D\times\mathbb{R}}}(\bar{\gamma}', \bar{\gamma}'') = d_{1,D\times\mathbb{R}}(\gamma', \gamma''). \tag{3.3}$$

Suppose now $(\mu_n, f_n) \xrightarrow{TL^1} (\theta, g)$, and $\gamma_n$ and $\gamma$ are the associated graph measures on $D \times \mathbb{R}$ for $n \geq 1$. Then, as $\lim_{n\to\infty} d_{1,D\times\mathbb{R}}(\gamma_n, \gamma) = 0$, we have, by (3.2) and (3.3), that $\lim_{n\to\infty} d_{1,\overline{D\times\mathbb{R}}}(\bar{\gamma}_n, \bar{\gamma}) = 0$. Since $\overline{D \times \mathbb{R}}$ is complete, by (3.1), $\bar{\gamma}_n \xrightarrow{w} \bar{\gamma}$ in $\mathcal{P}(\overline{D \times \mathbb{R}})$ and associated first moments converge, and so, equivalently, $\gamma_n \xrightarrow{w} \gamma$ in $\mathcal{P}(D \times \mathbb{R})$ and associated first moments converge, as $n \to \infty$.

We now make a remark on definition (2.11) with respect to the product space $(TL^1)^K$, equipped with the product topology. Fix a realization $\{X_i\}_{i\in\mathbb{N}}$. Recall the empirical measures $\nu_n$ and probability measure $\nu$ on $D$ from the beginning of Section 2.3. Let $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ be a partition of $\mathcal{X}_n$ for $n \geq 1$, and $\mathcal{U} = \{U_k\}_{k=1}^K$ be a partition of $D$. We say the sequence $(\nu_n, \mathcal{U}_n) := ((\nu_n, \mathbb{1}_{U_{n,k}}))_{k=1}^K$ converges in $(TL^1)^K$ to $(\nu, \mathcal{U}) := ((\nu, \mathbb{1}_{U_k}))_{k=1}^K$ if $(\nu_n, \mathbb{1}_{U_{n,k}}) \xrightarrow{TL^1} (\nu, \mathbb{1}_{U_k})$ for $1 \leq k \leq K$. Now, by the comment below (3.3), convergence in the metric $d_{TL^1,D}$ implies weak convergence in $\mathcal{P}(D \times \mathbb{R})$. But, since indicators of sets are uniformly bounded, weak convergence of graph measures of indicators in $\mathcal{P}(D \times \mathbb{R})$ implies convergence of first moments, and so is equivalent, by (3.1) and (3.2), to convergence with respect to $d_{TL^1,D}$. Hence, noting definition (2.10), we obtain

$$U_{n,k} \xrightarrow{w} U_k \quad \text{for } 1 \leq k \leq K \quad \text{if and only if} \quad (\nu_n, \mathcal{U}_n) \xrightarrow{(TL^1)^K} (\nu, \mathcal{U}). \tag{3.4}$$

This convergence is certainly sufficient for $\mathcal{U}_n \xrightarrow{w} \mathcal{U}$ in the sense of definition (2.11), by choosing the identity permutations. However, we observe an equivalent condition is the following: $\mathcal{U}_n \xrightarrow{w} \mathcal{U}$ if and only if there exists a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ of permutations in $\mathrm{Sym}(K)$ such that $((\nu_n, U_{n,\pi_n(k)}))_{k=1}^K \xrightarrow{(TL^1)^K} ((\nu, U_k))_{k=1}^K$ for $1 \leq k \leq K$.

We now discuss when this convergence may be formulated in terms of transportation maps. We say that a measurable function $T : D \to D$ is a *transportation map* between the measures $\mu \in \mathcal{P}(D)$ and $\theta \in \mathcal{P}(D)$ if $\theta = T_\sharp \mu$. In this context, for $f \in L^1(\theta)$, the change of variables formula holds:

$$\int_D f(y)\, d\theta(y) = \int_D f\big(T(x)\big)\, d\mu(x).$$

A transportation map $T$ yields a coupling $\pi_T \in \Gamma(\mu, \theta)$ defined by $\pi_T := (Id \times T)_\sharp \mu$ where $(Id \times T)(x) = (x, T(x))$. It is well known, when $\theta$ is absolutely continuous with respect to Lebesgue measure on $D$, that the infimum $d_{1,D}(\mu, \theta)$ can be achieved by a coupling $\pi_T$ induced by a transportation map $T$ between $\mu$ and $\theta$. Indeed, we note briefly that this is only one result among many others which relate various "Monge" and "Kantorovich" distances via optimal transport theory. See [76] and references therein; see also [4, 75].

We will say that a sequence $\{T_n\}_{n\in\mathbb{N}}$ of transportation maps, with $T_{n\sharp}\theta = \theta_n$, with respect to a sequence of measures $\{\theta_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(D)$, is *stagnating* if

$$\lim_{n\to\infty} \int_D \left|x - T_n(x)\right| d\theta(x) = 0.$$

The following is Proposition 3.12 in [38].

LEMMA 3.1. *Consider a measure $\theta \in \mathcal{P}(D)$ which is absolutely continuous with respect to the Lebesgue measure. Let $(\theta, f) \in TL^1(D)$ and let $\{(\theta_n, f_n)\}_{n\in\mathbb{N}}$ be a sequence in $TL^1(D)$. The following statements are equivalent*:

(i) $(\theta_n, f_n) \xrightarrow{TL^1} (\theta, f)$.

(ii) $\theta_n \xrightarrow{w} \theta$ *and there exists a stagnating sequence of transportation maps $T_{n\sharp}\theta = \theta_n$ such that*

$$\lim_{n\to\infty} \int_D \left|f(x) - f_n\big(T_n(x)\big)\right| d\theta(x) = 0. \tag{3.5}$$

(iii) $\theta_n \xrightarrow{w} \theta$ *and for any stagnating sequence of transportation maps $T_{n\sharp}\theta = \theta_n$, the convergence* (3.5) *holds*.

In order to make use of the above result on $TL^1$ convergence, we will need to find a suitable stagnating sequence $\{T_n\}_{n\in\mathbb{N}}$ of transportation maps.

PROPOSITION 3.2. *Recall, from the beginning of Section 2.3, the assumptions on the probability measure $\nu$ on $D$, and that $\nu_n$ denotes the empirical measure corresponding to i.i.d. samples drawn from $\nu$.*

*Then there is a constant $C > 0$ such that, with respect to realizations of $\{X_i\}_{i\in\mathbb{N}}$ in a probability 1 set $\Omega_0$, a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ exists where $T_{n\sharp}\nu = \nu_n$ and*

$$\limsup_{n\to\infty} \frac{n^{1/2}\|Id - T_n\|_{L^\infty}}{(\log\log n)^{1/2}} \le C \qquad \text{if } d = 1,$$

$$\limsup_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_{L^\infty}}{(\log n)^{3/4}} \le C \qquad \text{if } d = 2,$$

$$\limsup_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_{L^\infty}}{(\log n)^{1/d}} \le C \qquad \text{if } d \ge 3.$$

PROOF. We prove the $d = 1$ case in the Appendix (Proposition A.18), as a consequence of quantile transform results for the empirical measure, making use of the technical conditions assumed on $\rho$. In García Trillos and Slepčev [37], the $d = 2$ and $d \geq 3$ cases are first discussed, in the context of concentration estimates in the literature when $D$ is a cube and $\nu$ is the uniform measure, and then proved for general $D$ and nonuniform $\nu$. □

Although a result of Varadarajan (cf. Theorem 11.4.1 in [27]) implies that a.s. $\nu_n \xrightarrow{w} \nu$, Proposition 3.2 gives a way to specify the probability 1 set on which the weak convergence holds.

COROLLARY 3.3. *On the probability* 1 *set* $\Omega_0$ *of Proposition* 3.2, *the empirical measures* $\nu_n$ *converge weakly to* $\nu$ *as* $n \to \infty$.

PROOF. Let $f : D \to \mathbb{R}$ be a bounded, Lipschitz continuous function. Since

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) = \int_D f(T_n x)\, d\nu(x),$$

we may write

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_D f(x)\, d\nu(x)\right| &\leq \int_D \left|f(T_n x) - f(x)\right| d\nu(x)\\ &\leq C\int_D |x - T_n x|\, d\nu(x)\\ &\leq C\|Id - T_n\|_{L^\infty},\end{aligned}$$

where $C$ is a Lipschitz constant for $f$. By Proposition 3.2, for each realization of $\{X_i\}_{i\in\mathbb{N}}$ in $\Omega_0$, we have $\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \int_D f(x)\, d\nu(x)$ as $n \to \infty$. Hence, by the Portmanteau theorem (Theorem 3.9.1 in [28]), we have the weak convergence $\nu_n \xrightarrow{w} \nu$ as $n \to \infty$. □

3.2. *On Gamma convergence of random functionals.* Here, we introduce a type of $\Gamma$-convergence, with respect to random functionals, which will be an important tool in the proof of Theorem 2.3 in Section 6, and may be of interest in its own right. For what follows, let $X$ denote a metric space with metric $d$ and let $F_n : X \to [0, \infty]$ be functionals on this space.

We first state the definition with respect to deterministic functionals.

DEFINITION 3.4. The sequence $\{F_n\}_{n\in\mathbb{N}}$ $\Gamma$-converges with respect to the topology on $X$ if the following conditions hold:

1. *Liminf inequality*: For every $x \in X$ and every sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to $x$,

$$F(x) \le \liminf_{n\to\infty} F_n(x_n).$$

2. *Limsup inequality*: For every $x \in X$, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to $x$ satisfying

$$\limsup_{n\to\infty} F_n(x_n) \le F(x).$$

The function $F$ is called the $\Gamma$-limit of $\{F_n\}_{n\in\mathbb{N}}$, and we write $F_n \xrightarrow{\Gamma} F$.

When we wish to make the dependence on the metric $d$ explicit, we say that $\{F_n\}_{n\in\mathbb{N}}$ $\Gamma(d)$-converges to $F$, or $F$ is the $\Gamma(d)$-limit of $\{F_n\}_{n\in\mathbb{N}}$, etc.

REMARK 3.5. If the liminf inequality holds, the limsup inequality is equivalent to the following condition: For every $x \in X$, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ with $x_n \to x$ and $\lim_{n\to\infty} F_n(x_n) = F(x)$. The sequence $\{x_n\}_{n\in\mathbb{N}}$ is referred to as a *recovery sequence* for $x$.

A basic consequence of Definition 3.4 is the following (cf. [13], Theorem 1.21).

THEOREM 3.6. *Let $F_n : X \to [0,\infty]$ be a sequence of functionals $\Gamma$-converging to $F$. Suppose $\{x_n\}_{n\in\mathbb{N}}$ is a relatively compact sequence in $X$ with*

$$\lim_{n\to\infty}\Big(F_n(x_n) - \inf_{x\in X} F_n(x)\Big) = 0. \tag{3.6}$$

*Then*:

1. *$F$ attains its minimum value and*

$$\min_{x\in X} F(x) = \lim_{n\to\infty} \inf_{x\in X} F_n(x).$$

2. *The sequence $\{x_n\}_{n\in\mathbb{N}}$ has a cluster point, and every cluster point of the sequence is a minimizer of $F$.*

For this theorem to be applicable, it is standard to put some condition on $\{F_n\}_{n\in\mathbb{N}}$ so that (3.6) implies that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is relatively compact in $X$.

DEFINITION 3.7. We say that the sequence of nonnegative functionals $\{F_n\}_{n\in\mathbb{N}}$ has the compactness property, if when a sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfies the following two conditions:

(i) $\{x_n\}_{n\in\mathbb{N}}$ is bounded in $X$,
(ii) the energies $\{F_n(x_n)\}_{n\in\mathbb{N}}$ are bounded,

then $\{x_n\}_{n\in\mathbb{N}}$ is relatively compact in $X$.

We now extend the above notions to the random setting. Here, we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of functionals $F_n : X \times \Omega \to [0, \infty]$.

DEFINITION 3.8. We say the (random) sequence $\{F_n\}_{n\in\mathbb{N}}$ $\Gamma$-converges to the deterministic functional $F : X \to [0, \infty]$ if:

1. *Liminf inequality* With probability 1, the following statement holds: For any $x \in X$ and any sequence $\{x_n\}_{n\in\mathbb{N}}$ with $x_n \to x$,

$$F(x) \le \liminf_{n\to\infty} F_n(x_n).$$

2. *Recovery sequence* For any $x \in X$, there exists a (random) sequence $\{x_n\}_{n\in\mathbb{N}}$ with $x_n \xrightarrow{a.s.} x$ and $F_n(x_n) \xrightarrow{a.s} F(x)$.

DEFINITION 3.9. We say the (random) sequence $\{F_n\}_{n\in\mathbb{N}}$ has the compactness property if with probability 1, the sequence $\{F_n(\cdot, \omega)\}_{n\in\mathbb{N}}$ has the compactness property in Definition 3.7.

REMARK 3.10. The definition for $\Gamma$-convergence of random functionals, Definition 3.8, is weaker than the one in [38], which prescribes that Definition 3.4 holds with probability 1. However, in our Definition 3.8, with respect to the recovery sequence, the probability 1 set may depend on the sequence, and therefore is an easier condition to verify, say with probabilistic arguments. Interestingly, this weaker definition has the same strength in terms of its application in the following Gamma convergence statement, Theorem 3.11, a main vehicle in the proof of Theorem 2.3.

In passing, we also note that the compactness criterion of random functionals, Definition 3.9, can also be weakened, without altering the statement of the Gamma convergence Theorem 3.11, to the following: $\{F_n\}_{n\in\mathbb{N}}$ has the compactness property if when a sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfies (i) and (ii) in Definition 3.7 on a probability 1 set, there is a probability 1 subset (both sets may depend on $\{x_n\}_{n\in\mathbb{N}}$) on which $\{x_n\}_{n\in\mathbb{N}}$ is relatively compact. We remark parenthetically in the proof of Theorem 3.11 where a small change would be made if the weakened criterion were used. We also note that we do not use this weaker condition in the arguments of this article.

THEOREM 3.11. *Let $F_n : X \times \Omega \to [0, \infty]$ be a sequence of random functionals $\Gamma$-converging to a limit $F : X \to [0, \infty]$, in the sense of Definition 3.8, which is not identically equal to $\infty$. Suppose that $\{F_n\}_{n\in\mathbb{N}}$ has the compactness property, in the sense of Definition 3.9, and also the following condition holds: For $\omega$ in a probability 1 set, there exists a bounded sequence, $x_n = x_n(\omega)$, whose bound may depend on $\omega$, such that*

$$\lim_{n\to\infty} \Big( F_n(x_n) - \inf_{x\in X} F_n(x) \Big) = 0.$$

*Then, with probability 1:*

1. *F attains its minimum value and*

$$\min_{x\in X} F(x) = \lim_{n\to\infty} \inf_{x\in X} F_n(x).$$

2. *The sequence $\{x_n\}_{n\in\mathbb{N}}$ has a cluster point, and every cluster point of the sequence is a minimizer of $F$.*

PROOF. Pick $\tilde{x} \in X$, along with a recovery sequence $\{\tilde{x}_n\}_{k\in\mathbb{N}}$, so that on a probability 1 set $\Omega_1$ we have $\lim_{n\to\infty} \tilde{x}_n = \tilde{x}$ and $\lim_{n\to\infty} F_n(\tilde{x}_n) = F(\tilde{x})$. Let $\Omega_2$ be a probability 1 set on which $x_n = x_n(\omega)$ is a bounded sequence where $\lim_{n\to\infty}(F_n(x_n) - \inf_{x\in X} F_n(x)) = 0$. Hence, on the probability 1 set $\Omega_1 \cap \Omega_2$, we obtain

$$\limsup_{n\to\infty} F_n(x_n) \le F(\tilde{x}). \tag{3.7}$$

Applying the argument for (3.7) with respect to a countable collection $\{\tilde{x}^{(m)}\}_{m\in\mathbb{N}}$ with $\lim_{m\to\infty} F(\tilde{x}^{(m)}) = \inf_{x\in X} F(x)$, we obtain on a probability 1 set $\Omega_3 \subset \Omega_2$ that

$$\limsup_{n\to\infty} F_n(x_n) \le \inf_{x\in X} F(x). \tag{3.8}$$

Now, because $F$ is not identically equal to $\infty$, the right-hand side of the above inequality is finite. Then, on the probability 1 set $\Omega_3$, the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{F_n(x_n)\}_{n\in\mathbb{N}}$ are bounded. Let $\Omega_4$ be the probability 1 set on which the compactness property for $\{F_n\}_{n\in\mathbb{N}}$ holds. [If instead the weakened compactness criterion mentioned in Remark 3.10 is used, with respect to sequence $\{x_n\}_{n\in\mathbb{N}}$, then $\Omega_4$ would be the probability 1 subset of $\Omega_3$ on which $\{x_n\}_{n\in\mathbb{N}}$ is relatively compact.] In particular, on $\Omega_5 = \Omega_3 \cap \Omega_4$, the bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ is relatively compact. With respect to the set $\Omega_5$, let $\{x_{n_k}\}_{k\in\mathbb{N}}$ be a subsequence converging to a cluster point $x^*$, that is, $\lim_{k\to\infty} x_{n_k} = x^*$.

Let $\Omega_6$ be a probability 1 set on which the liminf inequality holds. Then, on $\Omega_7 = \Omega_5 \cap \Omega_6$, we have

$$\inf_{x\in X} F(x) \le F(x^*) \le \liminf_{k\to\infty} F_{n_k}(x_{n_k}). \tag{3.9}$$

Combining (3.8) and (3.9) shows, since $\Omega_7 \subset \Omega_3$, that on the set $\Omega_7$ we have

$$\limsup_{k\to\infty} F_{n_k}(x_{n_k}) \le \limsup_{n\to\infty} F_n(x_n) \le \inf_{x\in X} F(x) \le F(x^*) \le \liminf_{k\to\infty} F_{n_k}(x_{n_k}). \tag{3.10}$$

Hence, we conclude that $F$ attains its minimum value, $F(x^*) = \inf_{x\in X} F(x)$ and $x^*$ is a minimizer of $F$, proving part of the first statement. In fact, the second statement also follows: With respect to the probability 1 set $\Omega_3 \cap \Omega_6$, every cluster point of $\{x_n\}_{n\in\mathbb{N}}$ is a minimizer of $F$.

We now show the remaining part of the first statement. With respect to the set $\Omega_7$, let $\{x_{m_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{x_n\}_{n\in\mathbb{N}}$ where, for some cluster point $x^{**}$,

$$\lim_{k\to\infty} F_{m_k}(x_{m_k}) = \liminf_{n\to\infty} F_n(x_n) \quad \text{and} \quad \lim_{k\to\infty} x_{m_k} = x^{**}.$$

Then, by (3.10), we conclude on $\Omega_7$ that $\lim_{n\to\infty} F_n(x_n) = \inf_{x\in X} F(x)$. Since $\Omega_7 \subset \Omega_2$, we have on $\Omega_7$ that $\lim_{n\to\infty}(F_n(x_n) - \inf_{x\in X} F_n(x)) = 0$. Hence, we conclude on $\Omega_7$ that $\lim_{n\to\infty} \inf_{x\in X} F_n(x_n) = \inf_{x\in X} F(x)$. □

**4. Reformulation of the modularity functional.** In this section, we write the modularity functional as a sum of a "graph total variation" term and a "quadratic balance" term, which will aid in its subsequent analysis.

Recall the modularity functional in Section 2.3 acting on a partition $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ of the data points $\mathcal{X}_n$ into $K \geq 1$ sets:

$$Q_n(\mathcal{U}_n) = \frac{1}{2m}\sum_{i,j}\left(W_{ij} - 2m\frac{d_i^\alpha d_j^\alpha}{S^2}\right)\delta(c_i, c_j). \tag{4.1}$$

Here, $d_i = \sum_j W_{ij}$, $2m = \sum_{ij} W_{ij}$, and $S = \sum_i (\sum_j W_{ij})^\alpha$ and the weights $W_{ij} = \eta_{\varepsilon_n}(X_i - X_j)$ if $i \neq j$ and equal 0 otherwise. The label $c_i = k$ is assigned to the point $X_i$ if $X_i \in U_{n,k}$ for $1 \leq k \leq K$.

Define $I_n(D)$ as the collection of indicator functions of subsets of $\mathcal{X}_n$. Natural members of $I_n(D)$, in the above context, are $u_{n,k} = \mathbb{1}_{U_{n,k}}$ for $1 \leq k \leq K$. Note that the collection $\{u_{n,k}\}_{k=1}^K$ satisfies $\sum_{k=1}^K u_{n,k} = \mathbb{1}_{\mathcal{X}_n}$.

Observe now that $\delta(c_i, c_j)$, signifying that $X_i$ and $X_j$ have the same label, can be expressed in two ways:

$$\delta(c_i, c_j) = 1 - \frac{1}{2}\sum_{k=1}^K \left|u_{n,k}(X_i) - u_{n,k}(X_j)\right| = \sum_{k=1}^K u_{n,k}(X_i)u_{n,k}(X_j). \tag{4.2}$$

Applying the first identity in (4.2) to the first term in (4.1) gives

$$\begin{aligned}\frac{1}{2m}\sum_{i,j} W_{ij}\delta(c_i, c_j) &= \frac{1}{2m}\sum_{i,j} W_{ij} - \frac{1}{2}\frac{1}{2m}\sum_{k=1}^K\sum_{i,j} W_{ij}\left|u_{n,k}(X_i) - u_{n,k}(X_j)\right| \\ &= 1 - \frac{1}{2}\frac{1}{2m}\sum_{k=1}^K\sum_{i,j} W_{ij}\left|u_{n,k}(X_i) - u_{n,k}(X_j)\right|.\end{aligned}$$

Define the *graph total variation* $\mathrm{GTV}_n(u)$, acting on $u : \mathcal{X}_n \to \mathbb{R}$, to be

$$\mathrm{GTV}_n(u) := \frac{1}{\varepsilon_n}\frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n} \eta_{\varepsilon_n}(X_i - X_j)\left|u(X_i) - u(X_j)\right|. \tag{4.3}$$

Then we may write

$$\frac{1}{2m}\sum_{i,j} W_{ij}\delta(c_i, c_j) = 1 - \varepsilon_n\frac{n(n-1)}{4m}\sum_{k=1}^K \mathrm{GTV}_n(u_{n,k}). \tag{4.4}$$

Similarly, the second relation in (4.2) gives

$$\sum_{i,j}\frac{d_i^\alpha d_j^\alpha}{S^2}\delta(c_i,c_j)=\frac{1}{S^2}\sum_{k=1}^K\sum_{i,j}d_i^\alpha d_j^\alpha u_{n,k}(X_i)u_{n,k}(X_j)$$
$$=\frac{1}{S^2}\sum_{k=1}^K\Bigl(\sum_i d_i^\alpha u_{n,k}(X_i)\Bigr)^2$$
$$=\frac{1}{S^2}\sum_{k=1}^K\Bigl(\sum_i\Bigl(\sum_{\substack{1\le j\le n\\ j\neq i}}\eta_{\varepsilon_n}(X_i-X_j)\Bigr)^\alpha u_{n,k}(X_i)\Bigr)^2.$$

Define $G\Lambda_n(u)$, for $u:D\to\mathbb{R}$, by

$$G\Lambda_n(u):=\frac{1}{n}\sum_{i=1}^n\Bigl(\frac{1}{n-1}\sum_{\substack{1\le j\le n\\ j\neq i}}\eta_{\varepsilon_n}(X_i-X_j)\Bigr)^\alpha u(X_i). \tag{4.5}$$

Then

$$\sum_{i,j}\frac{d_i^\alpha d_j^\alpha}{S^2}\delta(c_i,c_j)=\frac{n^2(n-1)^{2\alpha}}{S^2}\sum_{k=1}^K\bigl(G\Lambda_n(u_{n,k})\bigr)^2.$$

Note that $G\Lambda_n(1)=S/(n(n-1)^\alpha)$. With a bit of algebra, we obtain

$$\sum_{k=1}^K\bigl(G\Lambda_n(u_{n,k})\bigr)^2$$
$$=\sum_{k=1}^K\bigl[\bigl(G\Lambda_n(u_{n,k}-1/K)\bigr)^2$$
$$+2G\Lambda_n(u_{n,k}-1/K)G\Lambda_n(1/K)+\bigl(G\Lambda_n(1/K)\bigr)^2\bigr],$$

which further equals

$$\sum_{k=1}^K\bigl(G\Lambda_n(u_{n,k}-1/K)\bigr)^2$$
$$+2G\Lambda_n\Bigl(\sum_{k=1}^K(u_{n,k}-1/K)\Bigr)G\Lambda_n(1/K)+\frac{1}{K}G\Lambda_n(1)^2$$
$$=\sum_{k=1}^K\bigl(G\Lambda_n(u_{n,k}-1/K)\bigr)^2+\frac{1}{K}G\Lambda_n(1)^2.$$

We have used the relation $\sum_{k=1}^K u_{n,k} = \mathbb{1}_{\mathcal{X}_n}$ in the last equality. Hence,

$$\sum_{i,j} \frac{d_i^\alpha d_j^\alpha}{S^2}\delta(c_i, c_j) = \frac{n^2(n-1)^{2\alpha}}{S^2}\left[\sum_{k=1}^K \bigl(G\Lambda_n(u_{n,k} - 1/K)\bigr)^2\right] + 1/K. \tag{4.6}$$

Combining (4.4) and (4.6) gives

$$\begin{aligned}1 - 1/K - Q_n(\mathcal{U}_n) = {}& \frac{n^2(n-1)^{2\alpha}}{S^2}\left[\sum_{k=1}^K \bigl(G\Lambda_n(u_{n,k} - 1/K)\bigr)^2\right] \\ &+ \varepsilon_n \frac{n(n-1)}{4m}\sum_{k=1}^K \mathrm{GTV}_n(u_{n,k}).\end{aligned} \tag{4.7}$$

**5. Proof of Theorem 2.1: Asymptotic formula.** We analyze the "graph total variation" and "quadratic balance" terms, identified in the decomposition of the modularity functional in Section 4, in the first two subsections. Then, in Section 5.3, we prove Theorem 2.1.

For this section, in accordance with the assumptions of Theorem 2.1, we suppose that the partition $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ of the data points $\mathcal{X}_n$ is induced by a "continuum" partition $\mathcal{U} = \{U_k\}_{k=1}^K$ of $D$ into $K \geq 1$ sets with finite perimeter $\mathrm{Per}(U_k; \rho^2) < \infty$, where $U_{n,k} = \{X_i \in \mathcal{X}_n | X_i \in U_k\}$, for $1 \leq k \leq K$.

Define $I(D)$ as the collection of measurable indicator functions of subsets $U \subset D$. Let $u_k = \mathbb{1}_{U_k}$, and note that $u_k \in I(D)$ is an extension of the indicator $u_{n,k} = \mathbb{1}_{U_{n,k}}$, defined on $\mathcal{X}_n$, for $1 \leq k \leq K$. Of course, the family $\{u_k\}_{k=1}^K$ satisfies $\sum_{k=1}^K u_k = \mathbb{1}_D$.

5.1. *Convergence of graph total variation.* To show a.s. convergence of the graph total variations, we first state that its expectations converge, and then use concentration ideas to elicit convergence of the random quantities.

Let $u \in L^1(D)$. We define the *nonlocal total variation* of $u$ to be

$$\mathrm{TV}_\varepsilon(u; \rho) := \frac{1}{\varepsilon}\int_D \int_D \eta_\varepsilon(x-y)\bigl|u(x) - u(y)\bigr|\rho(x)\rho(y)\,dx\,dy.$$

Note that, if $X$ and $Y$ are independent random variables with density $\rho$, we have

$$\mathbb{E}\left[\frac{1}{\varepsilon}\eta_\varepsilon(X-Y)\bigl|u(X) - u(Y)\bigr|\right] = \mathrm{TV}_\varepsilon(u; \rho).$$

Recalling the definition (4.3) of the graph total variation, we therefore have

$$\mathbb{E}\bigl[\mathrm{GTV}_n(u)\bigr] = \mathrm{TV}_{\varepsilon_n}(u; \rho).$$

Let also $\sigma_\eta := \int_D \eta(x)|x_1|\,dx$, where $x = (x_1, \dots, x_d)$.

LEMMA 5.1. *Let $u \in L^1(D)$ such that* $\mathrm{TV}(u;\rho^2) < \infty$. *Then we have*

$$\lim_{\varepsilon\to 0} \mathrm{TV}_\varepsilon(u;\rho) = \sigma_\eta \,\mathrm{TV}\big(u;\rho^2\big). \tag{5.1}$$

PROOF. For general $\rho$, continuous on $D$ and bounded above and below by positive constants, and $d \ge 2$, the result follows from part of the proof of [38], Theorem 4.1 (see Remark 4.3 in [38]), which is a much more involved result. This proof also holds in $d = 1$. More remarks can be found in the initial arXiv version of this article [22]. □

We now proceed to the almost sure convergence of the graph total variation to its continuum limit.

LEMMA 5.2. *Fix $u \in I(D)$ where* $\mathrm{TV}(u;\rho^2) < \infty$, *and let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence converging to zero such that*

$$\sum_{n=1}^{\infty} \exp\big(-n\varepsilon_n^{(d+1)/2}\big) < \infty. \tag{5.2}$$

*Then, as $n \to \infty$,*

$$\mathrm{GTV}_n(u) \xrightarrow{a.s.} \sigma_\eta \,\mathrm{TV}\big(u;\rho^2\big).$$

PROOF. In revision, we remark it has come to our attention that similar, but different calculations to those we present below are found in [36]; see there also for remarks concerning the optimality of $\varepsilon_n$ in this context.

Let $f_n(x,y) = \frac{1}{\varepsilon_n}\eta_{\varepsilon_n}(x-y)|u(x)-u(y)|$, and $\mathbb{E}f_n = \mathbb{E}[f_n(X_i,X_j)] = \mathrm{TV}_{\varepsilon_n}(u;\rho)$. Then

$$\mathrm{GTV}_n(u) = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \big(f_n(X_i,X_j) - \mathbb{E}f_n\big) + \mathbb{E}f_n.$$

By Lemma 5.1, we have $\lim_{n\to\infty}\mathbb{E}f_n = \sigma_\eta\,\mathrm{TV}(u;\rho^2)$. Therefore, it remains to argue that $\lim_{n\to\infty}\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}(f_n(X_i,X_j) - \mathbb{E}f_n) = 0$ almost surely.

Let now

$$l_n(X_i) = \int_D f_n(X_i,y)\rho(y)\,dy \quad \text{and} \quad m_n(X_j) = \int_D f_n(x,X_j)\rho(x)\,dx,$$

and also $h_n(X_i,X_j) = f_n(X_i,X_j) - l_n(X_i) - m_n(X_j) + \mathbb{E}f_n$, so that

$$f_n(X_i,X_j) - \mathbb{E}f_n = h_n(X_i,X_j) + \big(l_n(X_i) - \mathbb{E}f_n\big) + \big(m_n(X_j) - \mathbb{E}f_n\big).$$

Summing this gives

$$\begin{aligned}&\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}\big(f_n(X_i,X_j)-\mathbb{E}f_n\big)\\ &\quad=\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}h_n(X_i,X_j)\\ &\qquad+\frac{1}{n}\sum_{1\le i\le n}\big(l_n(X_i)-\mathbb{E}f_n\big)+\frac{1}{n}\sum_{1\le j\le n}\big(m_n(X_j)-\mathbb{E}f_n\big).\end{aligned}\tag{5.3}$$

We handle the three terms on the right-hand side of the above equation separately. First, note that $\mathbb{E}l_n=\mathbb{E}f_n$. An application of Bernstein's inequality ([74], Lemma 19.32) yields, for $s>0$, that

$$\mathbb{P}\bigg(\bigg|\frac{1}{n}\sum_i l_n(X_i)-\mathbb{E}f_n\bigg|>s\bigg)\le 2\exp\bigg(-\frac{1}{4}\frac{ns^2}{\mathbb{E}l_n^2+s\|l_n\|_{L^\infty}}\bigg).\tag{5.4}$$

In Lemma A.14 of the Appendix, we prove the upper bounds $\mathbb{E}l_n^2\le C/\varepsilon_n$ and $\|l_n\|_{L^\infty}\le C/\varepsilon_n$ where $C$ is a constant independent of $n$. Hence, given the assumption (5.2), (5.4) is summable and, therefore, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^n l_n(X_i)-\mathbb{E}f_n\xrightarrow{a.s.}0.\tag{5.5}$$

Similarly, we have, as $n\to\infty$,

$$\frac{1}{n}\sum_{j=1}^n m_n(X_j)-\mathbb{E}f_n\xrightarrow{a.s.}0.\tag{5.6}$$

What remains is the double sum $\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}h_n(X_i,X_j)$. Let $\{Y_i\}_{i=1}^n$ be independent copies of $\{X_i\}_{i=1}^n$. By the decoupling inequality of de la Peña and Montgomery-Smith [23], there is a constant $C$ independent of $n$ and $h$ such that

$$\begin{aligned}&\mathbb{P}\bigg(\bigg|\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}h_n(X_i,X_j)\bigg|>s\bigg)\\ &\qquad\le C\mathbb{P}\bigg(C\bigg|\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}h_n(X_i,Y_j)\bigg|>s\bigg).\end{aligned}\tag{5.7}$$

The sum $\sum_{1\le i\ne j\le n}h_n(X_i,Y_j)$ is canonical, that is, $\mathbb{E}_X[h_n(X_i,Y_j)]=0$ a.s. and $\mathbb{E}_Y[h_n(X_i,Y_j)]=0$ a.s., where $\mathbb{E}_X$ and $\mathbb{E}_Y$ denote expectation with respect to the first and second variables, respectively. A general concentration inequality for $U$-statistics given by Giné, Latała and Zinn in Theorem 3.3 of [41], states, for canonical kernels $\{h_{i,j}\}_{1\le i,j\le n}$, that

$$\mathbb{P}\biggl(\biggl|\sum_{1\le i,j\le n} h_{i,j}(X_i,Y_j)\biggr| > s\biggr) \le L\exp\biggl[-\frac{1}{L}\min\biggl(\frac{s^2}{R^2},\frac{s}{Z},\frac{s^{2/3}}{B^{2/3}},\frac{s^{1/2}}{A^{1/2}}\biggr)\biggr],$$

for all $s>0$, where $L$ is a constant not depending on $\{h_{i,j}\}_{1\le i,j\le n}$ or $n$, and

$$A = \max_{i,j}\|h_{i,j}\|_{L^\infty}, \qquad R^2 = \sum_{i,j}\mathbb{E}h_{i,j}^2,$$

$$B^2 = \max_{i,j}\biggl[\biggl\|\sum_i \mathbb{E}_X h_{i,j}^2(X_i,y)\biggr\|_{L^\infty}, \biggl\|\sum_j \mathbb{E}_Y h_{i,j}^2(x,Y_j)\biggr\|_{L^\infty}\biggr],$$

$$Z = \sup\biggl\{\mathbb{E}\sum_{i,j} h_{i,j}(X_i,Y_j)f_i(X_i)g_j(Y_j) : \mathbb{E}\sum_i f_i^2(X_i)\le 1, \mathbb{E}\sum_j g_j^2(Y_j)\le 1\biggr\}.$$

In our context, we take $h_{i,j}=h_n$ for $i\ne j$, and $h_{i,j}=0$ otherwise, which gives the constants $A=\|h_n\|_{L^\infty}$, $B^2=(n-1)\max(\|\mathbb{E}_X h_n^2\|, \|\mathbb{E}_Y h_n^2\|)$, $R^2=n(n-1)\times \mathbb{E}h_n^2$ and, after a manipulation, $Z\le n\|h_n\|_{L^2\to L^2}$, where

$$\|h_n\|_{L^2\to L^2} := \sup\{\mathbb{E}h(X,Y)f(X)g(Y) : \mathbb{E}f^2(X)\le 1, \mathbb{E}g^2(Y)\le 1\}.$$

It follows that

$$\begin{aligned}&\mathbb{P}\biggl(\biggl|\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n} h_n(X_i,Y_j)\biggr| > s\biggr)\\ &\quad\le L\exp\biggl[-\frac{1}{L'}\min\biggl(\frac{n^2s^2}{\mathbb{E}h_n^2},\frac{ns}{\|h_n\|_{L^2\to L^2}},\\ &\qquad \frac{n^{2/3}s^{2/3}}{[\max(\|\mathbb{E}_X h_n^2\|_{L^\infty},\|\mathbb{E}_Y h_n^2\|_{L^\infty})]^{1/3}},\frac{ns^{1/2}}{\|h_n\|_{L^\infty}^{1/2}}\biggr)\biggr]\end{aligned} \tag{5.8}$$

for some constant $L'$ independent of $n$ and $h$.

In Corollary A.15 of the Appendix, we prove the upper bounds

$$\mathbb{E}h_n^2 \le C/\varepsilon_n^{d+1}, \qquad \|\mathbb{E}_Y h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2}, \qquad \|\mathbb{E}_X h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2},$$

$$\|h_n\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \|h_n\|_{L^2\to L^2} \le C/\varepsilon_n.$$

Hence, the minimum in the right-hand side of (5.8) simplifies to

$$C\min(n^2\varepsilon_n^{d+1}s^2, n\varepsilon_n s, n\varepsilon_n^{(d+2)/3}s^{2/3}, n\varepsilon_n^{(d+1)/2}s^{1/2}).$$

We claim, for sufficiently large $n$, this minimum will be attained by $n\varepsilon_n^{(d+1)/2}\times s^{1/2}$: Indeed, by (5.2), $n\varepsilon_n^{(d+1)/2}\to\infty$, and so $n\varepsilon_n^{(d+1)/2}$ is smaller than $n^2\varepsilon_n^{d+1}$. Also, $n\varepsilon_n$ is larger than $n\varepsilon_n^{(d+1)/2}$ since $\varepsilon_n\to 0$ and $d\ge 1$. In addition, $n\varepsilon_n^{(d+2)/3}$ is larger than $n\varepsilon_n^{(d+1)/2}$ as $d\ge 1$.

Hence, by assumption (5.2), the right-hand side of (5.8) converges, yielding as $n \to \infty$,

$$\frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} h_n(X_i, X_j) \xrightarrow{a.s.} 0. \tag{5.9}$$

Applying (5.5), (5.6) and (5.9) to (5.3) completes the proof. □

5.2. *Quadratic balance term.* We first consider convergence of certain "mean-values", and then treat the random expressions, for various values of $\alpha$, in the subsequent subsections.

For $u \in L^1(D)$ and $\varepsilon > 0$, define $\Lambda_\varepsilon(u)$ by

$$\Lambda_\varepsilon(u) := \int_D \left( \int_D \eta_\varepsilon(x-y)\rho(y)\,dy \right)^\alpha u(x)\rho(x)\,dx.$$

Let

$$\rho_\varepsilon(x) := \int_D \eta_\varepsilon(x-y)\rho(y)\,dy, \tag{5.10}$$

and write, with this notation,

$$\Lambda_\varepsilon(u) = \int_D u(x)\big(\rho_\varepsilon(x)\big)^\alpha \rho(x)\,dx. \tag{5.11}$$

Define also

$$\Lambda(u) := \int_D u(x)\rho^{1+\alpha}(x)\,dx. \tag{5.12}$$

LEMMA 5.3. *Let $g$ be a bounded, measurable function on the domain $D$. Then there exists a constant $C$, independent of $g$, such that*

$$\big|\Lambda_\varepsilon(g) - \Lambda(g)\big| \le C\|g\|_{L^\infty(D)}\varepsilon, \tag{5.13}$$

*for all sufficiently small $\varepsilon$. Further, suppose there is a sequence $\{g_\varepsilon\}_{\varepsilon>0}$ with $g_\varepsilon \xrightarrow{L^1} g$ as $\varepsilon \to 0$. Then we have*

$$\lim_{\varepsilon \to 0} \Lambda_\varepsilon(g_\varepsilon) = \Lambda(g). \tag{5.14}$$

PROOF. We first prove inequality (5.13). By Lemma A.13 in the Appendix, there exist positive constants $A$, $B$ such that, for sufficiently small $\varepsilon$, both $\rho$ and $\rho_\varepsilon$ take values in the interval $[A, B]$. Then we have

$$\begin{aligned}
\big|\Lambda_\varepsilon(g) - \Lambda(g)\big| &= \left| \int_D g(x)\big(\rho_\varepsilon(x)\big)^\alpha \rho(x)\,dx - \int_D g(x)\big(\rho(x)\big)^\alpha \rho(x)\,dx \right| \\
&\le B\|g\|_{L^\infty} \int_D \big|\big(\rho_\varepsilon(x)\big)^\alpha - \big(\rho(x)\big)^\alpha\big|\,dx \\
&\le C\|g\|_{L^\infty} \int_D \big|\rho_\varepsilon(x) - \rho(x)\big|\,dx,
\end{aligned}$$

where the last inequality follows from the observation that $x \mapsto x^\alpha$ is Lipschitz on the interval $[A, B]$. By Lemma A.13 again, where $\int_D |\rho_\varepsilon(x) - \rho(x)|\,dx \le C'\varepsilon$ is proved, we obtain

$$\left|\Lambda_\varepsilon(g) - \Lambda(g)\right| \le C''\|g\|_{L^\infty}\varepsilon.$$

We now prove (5.14). Suppose we have a family $\{g_\varepsilon\}_{\varepsilon>0}$ with $g_\varepsilon \xrightarrow{L^1} g$ as $\varepsilon \to 0$. Then

$$\lim_{\varepsilon\to 0}\left|\Lambda_\varepsilon(g_\varepsilon) - \Lambda(g)\right| \le \lim_{\varepsilon\to 0}\int_D \left|g_\varepsilon(x)\big(\rho_\varepsilon(x)\big)^\alpha - g(x)\big(\rho(x)\big)^\alpha\right|\rho(x)\,dx.$$

Since $\rho$ is bounded, it is sufficient to prove that

$$\lim_{\varepsilon\to 0}\int_D \left|g_\varepsilon(x)\big(\rho_\varepsilon(x)\big)^\alpha - g(x)\big(\rho(x)\big)^\alpha\right|dx = 0.$$

Writing $g_\varepsilon\rho_\varepsilon^\alpha - g\rho^\alpha = g_\varepsilon\rho_\varepsilon^\alpha - g\rho_\varepsilon^\alpha + g\rho_\varepsilon^\alpha - g\rho^\alpha$, one may obtain

$$\begin{aligned}&\int_D \left|g_\varepsilon(x)\rho_\varepsilon^\alpha(x) - g(x)\rho^\alpha(x)\right|dx\\ &\quad\le \int_D \left|g_\varepsilon(x) - g(x)\right|\rho_\varepsilon^\alpha(x)\,dx + \|g\|_{L^\infty}\int_D \left|\rho_\varepsilon^\alpha(x) - \rho^\alpha(x)\right|dx.\end{aligned}$$

Now, since $g_\varepsilon \to g$ in $L^1$, and $\rho_\varepsilon$ is bounded above and below by Lemma A.13, we have that the first term on the right-hand side vanishes in the limit. Likewise, by Lemma A.13, we have also $\rho_\varepsilon \to \rho$ Lebesgue a.e. as $\varepsilon \to 0$. By dominated convergence, then, the second term on the right-hand side vanishes, completing the proof. □

In the following Section 5.2.1, the cases $\alpha = 0, 1$ are considered. Then, in Section 5.2.2, the general $\alpha \ne 0, 1$ case is treated, where different techniques are used as the the functional is nonlinear.

5.2.1. *Quadratic balance term*: $\alpha = 0$ *or* $\alpha = 1$. The expression (4.5) for $G\Lambda_n$ simplifies to give

$$G\Lambda_n(u) = \begin{cases}\dfrac{1}{n}\displaystyle\sum_{i=1}^n u(X_i) & \text{when } \alpha = 0,\\[2ex] \dfrac{1}{n(n-1)}\displaystyle\sum_{1\le i\ne j\le n}\eta_{\varepsilon_n}(X_i - X_j)u(X_i) & \text{when } \alpha = 1.\end{cases}$$

LEMMA 5.4. *Fix* $\alpha = 0$. *Let $u$ be a bounded, measurable function on the domain $D$, and let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence converging to zero such that*

$$\lim_{n\to\infty}\frac{\log\log n}{n\varepsilon_n} = 0.$$

*Then, as* $n \to \infty$,

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda(u)\big) \xrightarrow{a.s.} 0.$$

PROOF. By the law of the iterated logarithm, and the boundedness of $u$, we have

$$\limsup_{n\to\infty} \frac{n}{\sqrt{2n\log\log n}}\big|G\Lambda_n(u) - \Lambda(u)\big| \le C, \qquad \text{a.s.}$$

We may write

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda(u)\big) = \frac{\sqrt{2n\log\log n}}{n\sqrt{\varepsilon_n}} \frac{n}{\sqrt{2n\log\log n}}\big(G\Lambda_n(u) - \Lambda(u)\big),$$

where by assumption, $\lim_{n\to\infty} \frac{\sqrt{n\log\log n}}{n\sqrt{\varepsilon_n}} = 0$. The lemma follows. □

LEMMA 5.5. *Fix* $\alpha = 1$. *Let* $u$ *be a bounded, measurable function on the domain* $D$, *and let* $\{\varepsilon_n\}_{n\in\mathbb{N}}$ *be a sequence converging to zero such that*

$$\sum_{n=1}^{\infty} \exp\big(-n\varepsilon_n^{(d+1)/2}\big) < \infty. \tag{5.15}$$

*Then, as* $n \to \infty$,

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda(u)\big) \xrightarrow{a.s.} 0.$$

PROOF. We first rewrite

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda(u)\big) = \frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda_{\varepsilon_n}(u)\big) + \frac{1}{\sqrt{\varepsilon_n}}\big(\Lambda_{\varepsilon_n}(u) - \Lambda(u)\big).$$

Here, as $\alpha = 1$, $\Lambda_{\varepsilon_n}(u) = \int_D \int_D \eta_{\varepsilon_n}(x-y)u(x)\rho(x)\rho(y)\,dx\,dy$.

By an application of inequality (5.13), the second term on the right vanishes as $n \to \infty$. Hence, we must show that $\lim_{n\to\infty} \frac{1}{\sqrt{\varepsilon_n}}(G\Lambda_n(u) - \Lambda_{\varepsilon_n}(u)) = 0$ a.s.

Let $f_n(x,y) = \frac{1}{\sqrt{\varepsilon_n}}\eta_{\varepsilon_n}(x-y)u(x)$. For $i \ne j$, we have $\mathbb{E}f_n = \mathbb{E}f_n(X_i, X_j) = \frac{1}{\sqrt{\varepsilon_n}}\Lambda_{\varepsilon_n}(u)$. Then

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda_{\varepsilon_n}(u)\big) = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} f_n(X_i, X_j) - \mathbb{E}f_n.$$

We now follow the structure of the proof of the GTV case (Lemma 5.2). Although the definition of $f_n$ is different, let $l_n$, $m_n$, and $h_n$ be given as in Lemma 5.2

in terms of $f_n$, and write

$$\begin{aligned}\frac{1}{n(n-1)} & \sum_{1\le i\ne j\le n} \big(f_n(X_i,X_j)-\mathbb{E}f_n\big) \\ &= \frac{1}{n(n-1)}\sum_{1\le i\ne j\le n} h_n(X_i,X_j) \\ &\quad + \frac{1}{n}\sum_i\big(l_n(X_i)-\mathbb{E}f_n\big)+\frac{1}{n}\sum_j\big(m_n(X_j)-\mathbb{E}f_n\big).\end{aligned}$$

To complete the proof, we now show that the the three terms on the right- hand side of the above equation vanish.

As in the GTV case, with respect to $\{l_n\}_{n\in\mathbb{N}}$, we may arrive in the same steps to an inequality in form (5.4), whose right-hand side is summable: In Lemma A.16 of the Appendix, we prove the upper bounds $\mathbb{E}l_n^2 \le C/\varepsilon_n$ and $\|l_n\|_{L^\infty} \le C/\varepsilon_n^{1/2}$, where $C$ is a constant independent of $n$. Hence, $\frac{1}{n}\sum_{i=1}^n l_n(X_i) - \mathbb{E}f_n \xrightarrow{a.s.} 0$ as $n\to\infty$.

The same argument also gives that $\frac{1}{n}\sum_{j=1}^n m_n(X_j) - \mathbb{E}f_n \xrightarrow{a.s.} 0$ as $n\to\infty$.

Also, analogous steps, as in the GTV case, allows to derive, for the sequence $\{h_n\}_{n\in\mathbb{N}}$, an inequality in form (5.8). In Corollary A.17 of the Appendix, we prove the upper bounds

$$\mathbb{E}h_n^2 \le C/\varepsilon_n^{d+1}, \qquad \|\mathbb{E}_Y h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \|\mathbb{E}_X h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+1},$$
$$\|h_n\|_{L^\infty} \le C/\varepsilon_n^{d+1/2}, \qquad \|h_n\|_{L^2\to L^2} \le C/\varepsilon_n^{1/2}.$$

Hence, the minimum in the right-hand side of (5.8), in the current context, is bounded below by

$$C\min\big(n^2\varepsilon_n^{d+1}s^2, n\varepsilon_n^{1/2}s, n\varepsilon_n^{(d+1)/3}s^{2/3}, n\varepsilon_n^{(d+1/2)/2}s^{1/2}\big).$$

Note that $\varepsilon_n$ vanishes, and our assumption (5.15) implies that $n\varepsilon_n^{(d+1)/2}\to\infty$ as $n\to\infty$. Therefore, we conclude, for sufficiently large $n$, that

$$\begin{aligned}C\min & \big(n^2\varepsilon_n^{d+1}s^2, n\varepsilon_n^{1/2}s, n\varepsilon_n^{(d+1)/3}s^{2/3}, n\varepsilon_n^{(d+1/2)/2}s^{1/2}\big) \\ &\ge Cn\varepsilon_n^{(d+1)/2}\min\big(s^2,s^{1/2}\big).\end{aligned}$$

Hence, the right-hand side of (5.8), in the current context, is summable, and as $n\to\infty$, we have $\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n} h_n(X_i,X_j) \xrightarrow{a.s.} 0$. □

5.2.2. *Quadratic balance term*: *General* $\alpha$. Recall, from (4.5) and (5.12), the forms of $G\Lambda_n(u)$ and $\Lambda(u)$.

LEMMA 5.6. *Fix a bounded, measurable function $u$ on the domain $D$, and let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence converging to zero such that*

$$\sum_{n=1}^{\infty} n \exp\big(-n\varepsilon_n^{d+1}\big) < \infty. \tag{5.16}$$

*Then, as $n \to \infty$,*

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \Lambda(u)\big) \xrightarrow{a.s.} 0. \tag{5.17}$$

PROOF. Recall the forms of $\rho_\varepsilon$ and $\Lambda_\varepsilon$ in (5.10) and (5.11), respectively. We now introduce the intermediate term:

$$\overline{G\Lambda_n}(u) := \frac{1}{n}\sum_{1\le i\le n} \rho_{\varepsilon_n}(X_i)^\alpha u(X_i). \tag{5.18}$$

Then

$$G\Lambda_n(u) - \Lambda(u) = G\Lambda_n(u) - \overline{G\Lambda_n}(u) + \overline{G\Lambda_n}(u) - \Lambda_{\varepsilon_n}(u) + \Lambda_{\varepsilon_n}(u) - \Lambda(u).$$

The proof proceeds in three steps:

*Step 1.* We first attend to $G\Lambda_n(u) - \overline{G\Lambda_n}(u)$. We claim that

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \overline{G\Lambda_n}(u)\big) \xrightarrow{a.s.} 0, \tag{5.19}$$

as $n \to \infty$. Define

$$Z_i = \frac{1}{n-1}\sum_{\substack{1\le j\le n\\ j\ne i}} \eta_{\varepsilon_n}(X_i - X_j),$$

so that $G\Lambda_n(u) = \frac{1}{n}\sum_{i=1}^n Z_i^\alpha u(X_i)$. Then we have

$$\mathbb{E}[Z_i|X_i] = \rho_{\varepsilon_n}(X_i),$$

and further, an application of Bernstein's inequality (Lemma 19.32 of [74]) gives

$$\mathbb{P}\big(\big|Z_i - \rho_{\varepsilon_n}(X_i)\big| > t|X_i\big) \le 2\exp\bigg(-\frac{1}{4}\frac{t^2(n-1)}{a+tb}\bigg), \tag{5.20}$$

where $a = \mathbb{E}[\eta_{\varepsilon_n}(X_i - X_j)^2|X_i]$ and $b = \sup_{x\in D}|\eta_{\varepsilon_n}(X_i - x)|$. Recalling the definition $\eta_\varepsilon(z) = \eta(z/\varepsilon)/\varepsilon^d$ and the assumptions (K1), (K4), we have $a \le C/\varepsilon_n^d$ and $b \le C/\varepsilon_n^d$. Therefore, inequality (5.20) implies

$$\mathbb{P}\big(\big|Z_i - \rho_{\varepsilon_n}(X_i)\big| > t|X_i\big) \le 2\exp\bigg(-C\frac{t^2(n-1)\varepsilon_n^d}{(t+1)}\bigg)$$

in terms of a constant $C$ not depending on $n$. Applying a union bound gives

$$\begin{aligned}\mathbb{P}\Big(\sup_{1\le i\le n}\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|>t\Big)&\le n\mathbb{P}\big(\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|>t\big)\\&\le Cn\exp\biggl(-C\frac{t^2(n-1)\varepsilon_n^d}{(t+1)}\biggr).\end{aligned}\tag{5.21}$$

By Lemma A.13, there exist positive constants $A$, $B$ such that, for sufficiently small $\varepsilon$, both $\rho$ and $\rho_\varepsilon$ take values in the interval $[A, B]$. Let $J_n$ denote the event that $\sup_{1\le i\le n}|Z_i-\rho_{\varepsilon_n}(X_i)|<A/2$. Then, if $J_n$ holds, the inequality $A/2<Z_i<B+A/2$ is satisfied for all $i$. Since the function $x\mapsto x^\alpha$ is Lipschitz on the interval $[A/2, B+A/2]$, we obtain

$$\big|Z_i^\alpha-\rho_{\varepsilon_n}(X_i)^\alpha\big|\le C\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|.\tag{5.22}$$

Hence, when $J_n$ occurs, inequality (5.22) and $\|u\|_{L^\infty}<\infty$ imply, with respect to another constant $C$ independent of $n$, that

$$\begin{aligned}\big|G\Lambda_n(u)-\overline{G\Lambda_n}(u)\big|&\le\frac{1}{n}\sum_{1\le i\le n}\big|Z_i^\alpha-\rho_{\varepsilon_n}(X_i)^\alpha\big|\big|u(X_i)\big|\\&\le C\frac{1}{n}\sum_{1\le i\le n}\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|\le C\sup_{1\le i\le n}\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|,\end{aligned}$$

and moreover,

$$\frac{1}{\sqrt{\varepsilon_n}}\big|G\Lambda_n(u)-\overline{G\Lambda_n}(u)\big|\le\frac{C}{\sqrt{\varepsilon_n}}\sup_{1\le i\le n}\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|.$$

It follows, by (5.21), that

$$\begin{aligned}&\mathbb{P}\biggl(\biggl\{\frac{1}{\sqrt{\varepsilon_n}}\big|G\Lambda_n(u)-\overline{G\Lambda_n}(u)\big|>t\biggr\}\cap J_n\biggr)\\&\quad\le\mathbb{P}\biggl(\frac{C}{\sqrt{\varepsilon_n}}\sup_{1\le i\le n}\big|Z_i-\rho_{\varepsilon_n}(X_i)\big|>t\biggr)\le Cn\exp\biggl(-C\frac{t^2(n-1)\varepsilon_n^{d+1}}{t+1}\biggr).\end{aligned}\tag{5.23}$$

On the other hand, by (5.21) again, we have

$$\mathbb{P}\big(J_n^c\big)\le Cn\exp\biggl(-C\frac{A^2(n-1)\varepsilon_n^d}{A+1}\biggr).\tag{5.24}$$

Combining (5.23) and (5.24) gives

$$\begin{aligned}&\mathbb{P}\biggl(\frac{1}{\sqrt{\varepsilon_n}}\big|G\Lambda_n(u)-\overline{G\Lambda_n}(u)\big|>t\biggr)\\&\quad\le\mathbb{P}\biggl(\biggl\{\frac{1}{\sqrt{\varepsilon_n}}\big|G\Lambda_n(u)-\overline{G\Lambda_n}(u)\big|>t\biggr\}\cap J_n\biggr)+\mathbb{P}\big(J_n^c\big)\\&\quad\le Cn\exp\big(-C(n-1)\varepsilon_n^{d+1}\big),\end{aligned}$$

for some constant $C$ not depending on $n$. From our assumption (5.16) on $\varepsilon_n$, the right-hand side of the above display is summable and so, as $n \to \infty$,

$$\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(u) - \overline{G\Lambda_n}(u)\big) \xrightarrow{a.s.} 0.$$

*Step 2.* Now, we argue, as $n \to \infty$, that

$$\frac{1}{\sqrt{\varepsilon_n}}\big(\overline{G\Lambda_n}(u) - \Lambda_{\varepsilon_n}(u)\big) \xrightarrow{a.s.} 0. \tag{5.25}$$

Noting (5.11), since

$$\mathbb{E}\big[\rho_{\varepsilon_n}(X_i)^\alpha u(X_i)\big] = \Lambda_{\varepsilon_n}(u),$$

by Bernstein's inequality (Lemma 19.32 of [74]), we have

$$\mathbb{P}\bigg(\frac{1}{\sqrt{\varepsilon_n}}\bigg|\frac{1}{n}\sum_{1\le i\le n}\rho_{\varepsilon_n}(X_i)^\alpha u(X_i) - \Lambda_{\varepsilon_n}(u)\bigg| > t\bigg) \le 2\exp\bigg(-\frac{1}{4}\frac{t^2\varepsilon_n n}{a + t\sqrt{\varepsilon_n}b}\bigg),$$

where $a = \mathbb{E}[(\rho(X_i)^\alpha u(X_i))^2]$ and $b = \sup_{x\in D}|\rho_{\varepsilon_n}(x)^\alpha u(x)|$. Both of these are bounded by a constant $C$, and so by the assumption (5.16) on $\varepsilon_n$, we obtain the last display is summable and, therefore, (5.25) holds.

*Step 3.* By Lemma 5.3, we have $|\Lambda_{\varepsilon_n}(u) - \Lambda(u)| \le C\|u\|_{L^\infty}\varepsilon_n$. It follows that

$$\frac{1}{\sqrt{\varepsilon_n}}\big(\Lambda_{\varepsilon_n}(u) - \Lambda(u)\big) \to 0, \tag{5.26}$$

as $n \to \infty$. Combining (5.19), (5.25) and (5.26) gives (5.17). □

5.3. *Proof of Theorem* 2.1. Recall equation (4.7) which decomposes the modularity $Q_n(\mathcal{U}_n)$ with respect to partitions $\mathcal{U}_n$ of $\mathcal{X}_n$ induced from a partition $\mathcal{U} = \{U_k\}_{k=1}^K$ of $D$, where each of the sets $U_k$ have finite perimeter, $\mathrm{Per}(U_k; \rho^2) < \infty$.

Since $S/(n(n-1)^\alpha) = G\Lambda_n(1)$ and $\int_D \rho^{1+\alpha}(x)\,dx = \Lambda(1)$, by Lemmas 5.4, 5.5, and 5.6, which cover the cases $\alpha = 0$, $\alpha = 1$, and $\alpha \ne 0, 1$, we have

$$\frac{S}{n(n-1)^\alpha} \xrightarrow{a.s.} \int_D \rho^{1+\alpha}(x)\,dx \tag{5.27}$$

as $n \to \infty$. In particular, when $\alpha = 1$, we have, as $n \to \infty$,

$$\frac{2m}{n(n-1)} = \frac{1}{n(n-1)}\sum_{\substack{1\le i,j\le n\\ i\ne j}}\eta_{\varepsilon_n}(X_i - X_j) \xrightarrow{a.s} \int_D \rho^2(x)\,dx. \tag{5.28}$$

Further, these same lemmas, applied to the indicators $\{u_k = \mathbb{1}_{U_k}\}_{k=1}^K$, imply that

$$\sum_{k=1}^K\big(G\Lambda_n(u_k - 1/K)\big)^2 \xrightarrow{a.s.} \sum_{k=1}^K\big(\Lambda(u_k - 1/K)\big)^2$$

as $n \to \infty$. Hence, combining these limits,

$$\frac{n^2(n-1)^{2\alpha}}{S^2}\left[\sum_{k=1}^{K}\big(G\Lambda_n(u_k-1/K)\big)^2\right] \xrightarrow{a.s.} \sum_{k=1}^{K}\big(\mu(U_k)-1/K\big)^2, \tag{5.29}$$

as $n \to \infty$, with $d\mu(x) = \frac{\rho^{1+\alpha}(x)}{\int_D \rho^{1+\alpha}(x)\,dx}$.

By Lemma 5.2, we have, as $n \to \infty$,

$$\sum_{k=1}^{K} \mathrm{GTV}_n(u_k) \xrightarrow{a.s.} \sigma_\eta \sum_{k=1}^{K} \mathrm{TV}\big(u_k;\rho^2\big),$$

where $\sigma_\eta = \int_{\mathbb{R}^d} \eta(x)|x_1|\,dx$. Therefore, as $n \to \infty$,

$$\frac{n(n-1)}{4m}\sum_{k=1}^{K} \mathrm{GTV}_n(u_k) \xrightarrow{a.s.} \frac{\sigma_\eta}{2\int_D \rho^2(x)\,dx}\sum_{k=1}^{K} \mathrm{TV}\big(u_k;\rho^2\big). \tag{5.30}$$

Since $\lim_{n\to\infty} \varepsilon_n = 0$, and $\mathrm{TV}(u_k;\rho^2) < \infty$ for $1 \le k \le K$ as the sets in $\mathcal{U}$ have finite perimeter, noting (5.28) and (5.30), we have that $\frac{\varepsilon_n n(n-1)}{4m}\sum_{k=1}^{K} \mathrm{GTV}_n(u_k)$ vanishes a.s., as $n \to \infty$. Hence, from the limit (5.29), we obtain the first statement (2.8) in Theorem 2.1.

To prove the second statement (2.9), we write, dividing (4.7) by $\varepsilon_n$, that

$$\frac{1-1/K-Q_n(\mathcal{U})}{\varepsilon_n} = \frac{n^2(n-1)^{2\alpha}}{S^2}\left[\sum_{k=1}^{K}\left(\frac{1}{\sqrt{\varepsilon_n}}G\Lambda_n(u_k-1/K)\right)^2\right] + \frac{n(n-1)}{4m}\sum_{k=1}^{K}\mathrm{GTV}_n(u_k).$$

By assumption, the partition $\mathcal{U}$ is balanced with respect to $d\mu$, and so

$$\sum_{k=1}^{K}\big(\mu(U_k)-1/K\big)^2 = 0.$$

Equivalently, recalling the definition (5.12) of $\Lambda$, we have $\Lambda(u_k - 1/K) = 0$ for $1 \le k \le K$.

Hence, writing $g_k = u_k - 1/K$, it follows that

$$\sum_{k=1}^{K}\left(\frac{1}{\sqrt{\varepsilon_n}}G\Lambda_n(u_k-1/K)\right)^2 = \sum_{k=1}^{K}\left(\frac{1}{\sqrt{\varepsilon_n}}\big(G\Lambda_n(g_k)-\Lambda(g_k)\big)\right)^2 \xrightarrow{a.s.} 0, \tag{5.31}$$

as $n \to \infty$, by Lemmas 5.4, 5.5 and 5.6 for the various cases of $\alpha$. Combining (5.30) and (5.31) gives

$$\frac{1-1/K-Q_n(\mathcal{U})}{\varepsilon_n} \xrightarrow{a.s.} \frac{\sigma_\eta}{2\int_D \rho^2(x)\,dx}\sum_{k=1}^{K}\mathrm{TV}\big(u_k;\rho^2\big),$$

as $n \to \infty$. This completes the proof of Theorem 2.1. $\square$

**6. Proof of Theorem 2.3: Optimal clusterings.** Following the approach outlined in Section 2.5, we reformulate the modularity clustering problem on the same space as the continuum partitioning problem and state a Gamma convergence result (Theorem 6.1) relating the two optimizations in Section 6.1. In Sections 6.2 and 6.3, we prove the "liminf" and "recovery" parts of Theorem 6.1. Finally, in Section 6.4, we show a compactness principle and combine previous elements to prove Theorem 2.3.

6.1. *Reformulation as a mimimization problem and Gamma convergence.* Recall the identity (4.7),

$$1 - 1/K - Q_n(\mathcal{U}_n) = \frac{n^2(n-1)^{2\alpha}}{S^2}\left[\sum_{k=1}^{K}\big(G\Lambda_n(u_{n,k} - 1/K)\big)^2\right] + \varepsilon_n \frac{n(n-1)}{4m}\sum_{k=1}^{K}\mathrm{GTV}_n(u_{n,k}),$$

where $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ is a partition of the data points $\mathcal{X}_n$ and $u_{n,k} = \mathbb{1}_{U_{n,k}} \in I_n(D)$ for $1 \le k \le K$. As is our convention, we note that some of the $\{U_{n,k}\}_{k=1}^K$ may be empty sets, and so generally we have $|\mathcal{U}| \le K$.

In a sense, what we have done for the modularity functional in (4.7) is to write it such that its "Γ-development" (cf. [15] and references therein) is explicit. In passing from (4.7) to a Γ-limit directly, the "total variation" term vanishes and one recovers only a coarse-grained description of optimal clusterings, characterized as balanced partitions with no condition on the perimeters.

For a finer description of optimal clusterings, we rescale the energies. Define

$$F_n(\mathcal{U}_n) := \frac{1}{\varepsilon_n}\frac{n^2(n-1)^{2\alpha}}{S^2}\left[\sum_{k=1}^{K}\big(G\Lambda_n(u_{n,k} - 1/K)\big)^2\right] \tag{6.1}$$

and

$$\mathrm{TV}_n(\mathcal{U}_n) := \frac{n(n-1)}{4m}\sum_{k=1}^{K}\mathrm{GTV}_n(u_{n,k}), \tag{6.2}$$

so that the problem of maximizing $Q_n(\mathcal{U}_n)$ over clusterings $\mathcal{U}_n$ of $\mathcal{X}_n$ with $|\mathcal{U}_n| \le K$ is equivalent to that of minimizing $F_n(\mathcal{U}_n) + \mathrm{TV}_n(\mathcal{U}_n)$.

We now formulate the modularity optimization problem on the space $(TL^1(D))^K$. Recall that $\nu_n$ denotes the empirical measure. We define

$$M_n(D) := \left\{\big((\nu_n, u_{n,k})\big)_{k=1}^K : u_{n,k} \in I_n(D), \sum_{k=1}^{K} u_{n,k} = \mathbb{1}_{\mathcal{X}_n}\right\},$$

and note that $M_n(D) \subset (TL^1(D))^K$. We often write elements of $M_n(D)$ as

$$(\nu_n, \mathcal{U}_n) := \big\{(\nu_n, u_{n,k})\big\}_{k=1}^K,$$

where $u_{n,k} = \mathbb{1}_{U_{n,k}}$ and $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$.

Define $E_n : (TL^1(D))^K \times \Omega \to [0,\infty]$ by

$$E_n(\mathcal{V}_n) = \begin{cases} F_n(\mathcal{U}_n) + \mathrm{TV}_n(\mathcal{U}_n) & \text{if } \mathcal{V}_n = (\nu_n, \mathcal{U}_n) \in M_n(D), \\ \infty & \text{otherwise.} \end{cases} \tag{6.3}$$

The energy minimization problem

$$\underset{\mathcal{V}_n \in (TL^1(D))^K}{\text{minimize}} \; E_n(\mathcal{V}_n), \tag{6.4}$$

is equivalent to the $K$-class modularity clustering problem (2.3), in the sense that $\mathcal{U}_n$ is a solution to (2.3) iff $\mathcal{V}_n = (\nu_n, \mathcal{U}_n)$ is a solution to (6.4).

Similarly, we define continuum functionals on partitions $\mathcal{U} = \{U_k\}_{k=1}^K$ of $D$, via their indicators $\{u_k\}_{k=1}^K \subset I(D)$, by

$$F(\mathcal{U}) = \begin{cases} 0 & \text{if } \sum_{k=1}^K (\mu(U^{(k)}) - 1/K)^2 = 0, \\ \infty & \text{otherwise,} \end{cases} \tag{6.5}$$

and

$$\mathrm{TV}(\mathcal{U}) = C_{\eta,\rho} \sum_{k=1}^K \mathrm{TV}(u_k; \rho^2), \tag{6.6}$$

where $d\mu = \rho^{1+\alpha} dx / \int_D \rho^{1+\alpha}(x)\, dx$ and

$$C_{\eta,\rho} = \frac{\sigma_\eta}{2\int_D \rho^2(x)\, dx} = \frac{\int_D \eta(x)|x_1|\, dx}{2\int_D \rho^2(x)\, dx}.$$

Define $M(D) \subset (TL^1(D))^K$ by

$$M(D) := \left\{ ((\nu, u_k))_{k=1}^K : u_k \in I(D), \sum_{k=1}^K u_k = \mathbb{1}_D, \sum_{k=1}^K \mathrm{TV}(u_k, \rho^2) < \infty \right\}.$$

As before, we denote elements of $M(D)$ by

$$(\nu, \mathcal{U}) := \{(\nu, u_k)\}_{k=1}^K,$$

where $\mathcal{U} = \{U_k\}_{k=1}^K$ and $u_k = \mathbb{1}_{U_k}$ for $1 \le k \le K$.

Define the energy $E : (TL^1(D))^K \to [0,\infty]$ as

$$E(\mathcal{V}) := \begin{cases} F(\mathcal{U}) + \mathrm{TV}(\mathcal{U}) & \text{if } \mathcal{V} = (\nu, \mathcal{U}) \in M(D), \\ \infty & \text{otherwise.} \end{cases} \tag{6.7}$$

Then, with $\mu$ as above and $\phi = \rho^2$, the continuum partitioning problem (2.6), which does not include the prefactor $C_{\eta,\rho}$, is equivalent to

$$\underset{\mathcal{V} \in (TL^1(D))^K}{\text{minimize}} \; E(\mathcal{V}), \tag{6.8}$$

in the sense that $\mathcal{U}$ is a solution to (2.6) iff $\mathcal{V} = (\nu, \mathcal{U})$ is a solution to (6.8).

As noted in Section 2.2, since there is a solution to (2.6), the problem (6.8) also possesses a solution. In particular, the energy $E$ is not identically infinite.

We now state the Gamma convergence, with respect to the metric space $(TL^1)^K$ equipped with the product topology, used later in the proof of Theorem 2.3.

THEOREM 6.1. *Suppose the assumptions of Theorem 2.3 are satisfied. Then the random functionals $E_n : (TL^1(D))^K \times \Omega \to [0, \infty]$, given in (6.3), $\Gamma$-converge in $(TL^1)^K$ to $E : (TL^1(D))^K \to [0, \infty]$, given in (6.7):*

$$E_n \xrightarrow{\Gamma((TL^1))^K} E,$$

*as $n \to \infty$, in the sense of Definition 3.8.*

PROOF. The proof of Theorem 6.1 is in two steps. In Section 6.2, via Lemma 6.7, we give the "liminf" estimate. In Section 6.3, through Lemma 6.9, we prove the "recovery sequence" property. □

6.2. *Liminf inequality.* We now argue the liminf inequality for the $\Gamma$-convergence in Theorem 6.1, according to Definition 3.8. Recall that $\Omega_0$ denotes the probability 1 set of realizations $\{X_i\}_{i\in\mathbb{N}}$, under which Proposition 3.2 holds.

We first show a closure property of $M_n(D)$ and $M(D)$.

LEMMA 6.2. *On the probability 1 set $\Omega_0$, the following holds: Suppose $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ is a sequence in $M_n(D)$ and $\mathcal{V} = ((\mu_k, u_k))_{k=1}^K \in (TL^1(D))^K$ satisfies $\sum_{k=1}^K \mathrm{TV}(u_k, \rho^2) < \infty$. Then, if $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$, we have $\mathcal{V} \in M(D)$.*

PROOF. Fix a realization in the probability 1 set $\Omega_0$, and let $\mathcal{V}_n = ((\nu_n, u_{n,k}))_{k=1}^K$ and $\mathcal{V} = ((\mu_k, u_k))_{k=1}^K$. By the characterization of $TL^1$ convergence, Lemma 3.1, for each $1 \le k \le K$ we have $\nu_n \xrightarrow{w} \mu_k$, and so by Corollary 3.3 it follows that $\mu_k = \nu$. Further, we have

$$\lim_{n\to\infty} \int_D |u_k(x) - u_{n,k}(T_n x)|\rho(x)\, dx = 0,$$

where $\{T_n\}_{n\in\mathbb{N}}$ is the sequence of transportation maps given in Proposition 3.2.

Hence, as $\rho$ is bounded above and below on $D$, it follows that $u_k$ is the $L^1$ limit of a sequence of indicator functions $\tilde{u}_{n,k}(x) := u_{n,k}(T_n x) \in I(D)$. It follows, by subsequential Lebesgue a.e. convergence, that $u_k \in I(D)$. Similarly, the relation $\sum_{k=1}^K u_k = \mathbb{1}_D$ follows from the corresponding relations for $\{u_{n,k}\}_{k=1}^K$. Thus, given the perimeter assumption on $\mathcal{V}$, we conclude $\mathcal{V} = ((\nu, u_k))_{k=1}^K \in M(D)$. □

We now establish the following technical lemma, which adapts a technique from the proof of Theorem 1.1 in [38] to relate graph functionals with their continuum nonlocal analogues.

Recall that we have defined earlier $\Lambda(g) = \int_D \rho^{1+\alpha}(x)g(x)\, dx$ and $\rho_\varepsilon(x) = \int_D \eta_\varepsilon(x-y)\rho(y)\, dy$ [cf. equations (5.12) and (5.10)].

LEMMA 6.3. *On the probability* 1 *set* $\Omega_0$, *the following statement holds*: *Given any sequence of uniformly bounded*, *nonnegative functions* $\{g_n\}_{n\in\mathbb{N}}$, *and a function* $g$, *if* $(\nu_n, g_n) \xrightarrow{TL^1} (\nu, g)$, *then*

$$\lim_{n\to\infty} G\Lambda_n(g_n) = \Lambda(g). \tag{6.9}$$

PROOF. Fix a realization in the probability 1 set $\Omega_0$. Recall, from (4.5), that

$$\begin{aligned} G\Lambda_n(g_n) &= \frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{1}{n-1}\sum_{\substack{1\le j\le n\\ j\neq i}} \eta_{\varepsilon_n}(X_i - X_j)\Bigg)^{\alpha} g_n(X_i) \\ &= \frac{n^\alpha}{(n-1)^\alpha}\frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{1}{n}\sum_{j=1}^{n}\eta_{\varepsilon_n}(X_i - X_j) - \frac{\eta_{\varepsilon_n}(0)}{n}\Bigg)^{\alpha} g_n(X_i). \end{aligned}$$

Let $\{T_n\}_{n\in\mathbb{N}}$ be the transport maps in Proposition 3.2, and let $R_n(x) := \int_D \eta_{\varepsilon_n}(T_n x - T_n y)\rho(y)\,dy$. Since $T_{n\sharp}\nu = \nu_n$, by a change of variables, we have

$$G\Lambda_n(g_n) = \frac{n^\alpha}{(n-1)^\alpha}\int_D \bigg(R_n(x) - \frac{\eta_{\varepsilon_n}(0)}{n}\bigg)^{\alpha} g_n(T_n x)\rho(x)\,dx.$$

*Step 1.* First, suppose that $\eta$ is of the form $\eta(x) = a$ for $|x| < b$ and $\eta(x) = 0$ for $|x| > b$, with $\int_{\mathbb{R}^d} \eta(x)\,dx = 1$. Define

$$\overline{\varepsilon}_n := \varepsilon_n + 2\frac{\|Id - T_n\|_{L^\infty}}{b}, \tag{6.10}$$

and note that, for Lebesgue a.e. $(x, y) \in D \times D$,

$$|x - y| > b\overline{\varepsilon}_n \quad \text{implies} \quad |T_n x - T_n y| > b\varepsilon_n.$$

By the form of $\eta$, we have the bound

$$\eta\bigg(\frac{T_n x - T_n y}{\varepsilon_n}\bigg) \le \eta\bigg(\frac{x-y}{\overline{\varepsilon}_n}\bigg).$$

Integrating with respect to $\rho(y)\,dy$, and scaling appropriately, we obtain

$$R_n(x) \le (\overline{\varepsilon}_n/\varepsilon_n)^d \rho_{\overline{\varepsilon}_n}(x)$$

for Lebesgue a.e. $x \in D$.

By the assumption (I2) on $\varepsilon_n$, together with the estimates in Proposition 3.2 on $\{T_n\}_{n\in\mathbb{N}}$, it follows that $\varepsilon_n$ vanishes slower than $\|Id - T_n\|_{L^\infty}$, and so for large $n$ we have

$$\tilde{\varepsilon}_n := \varepsilon_n - 2\frac{\|Id - T_n\|_{L^\infty}}{b} > 0.$$

In particular, for Lebesgue a.e. $(x, y) \in D \times D$,

$$|T_n x - T_n y| > b\varepsilon_n \quad \text{implies} \quad |x - y| > b\tilde{\varepsilon}_n$$

and so

$$\eta\left(\frac{x-y}{\tilde{\varepsilon}_n}\right) \le \eta\left(\frac{T_n x - T_n y}{\varepsilon_n}\right). \tag{6.11}$$

Again, integrating with respect to $\rho(y)\,dy$ and scaling appropriately, we obtain a lower bound of $R_n(x)$, and can write, for Lebesgue a.e. $x$,

$$(\tilde{\varepsilon}_n/\varepsilon_n)^d \rho_{\tilde{\varepsilon}_n}(x) \le R_n(x) \le (\overline{\varepsilon}_n/\varepsilon_n)^d \rho_{\overline{\varepsilon}_n}(x). \tag{6.12}$$

By the assumption (I2) on the rate $\varepsilon_n$, we observe that

$$\lim_{n\to\infty} \frac{\tilde{\varepsilon}_n}{\varepsilon_n} = \lim_{n\to\infty} 1 + 2\frac{\|Id - T_n\|_{L^\infty}}{\varepsilon_n} = 1.$$

Similarly, we have $\lim_{n\to\infty} \frac{\overline{\varepsilon}_n}{\varepsilon_n} = 1$.

In light of (6.12), and Lemma A.13 in the Appendix, which bounds $\rho_\varepsilon$ from above and below and shows $\lim_{\varepsilon\to 0} \rho_\varepsilon = \rho$, we make two observations:

(i) For Lebesgue a.e. $x$, we have $R_n(x) \to \rho(x)$ as $n \to \infty$.

(ii) There exist constants $A', B' > 0$ such that, for all large $n$, $A' \le R_n(x) \le B'$ for Lebesgue a.e. $x$.

*Step 2.* Now let $\eta$ be a simple function satisfying assumptions (K1)–(K4), which implies that we may write $\eta$ as a convex combination $\eta = \sum_{l=1}^{L} \lambda_l \eta^l$ for functions $\eta^{(l)}$ satisfying the assumptions of Step 1. We let $R_n^{(l)}(x) := \int_D \eta_{\varepsilon_n}^{(l)}(T_n x - T_n y)\rho(y)\,dy$ so that $R_n(x) = \sum_{l=1}^{(l)} \lambda_l R_n^{(l)}(x)$.

Hence:

(i) For Lebesgue a.e. $x$, each $R_n^{(l)}(x) \to \rho(x)$ as $n \to \infty$. The same holds for the convex combination $R_n$.

(ii) There exist constants $A', B' > 0$ such that, for all large $n$, $A' \le R_n^{(l)}(x) \le B'$ for Lebesgue a.e. $x$. Therefore, the same holds for $R_n$.

Since $\lim_{n\to\infty} n\varepsilon_n^d = \infty$ by the assumption (I2), we have $\eta_{\varepsilon_n}(0)/n \le \|\eta\|_{L^\infty}(n\varepsilon_n^d)^{-1}$ vanishes as $n \to \infty$. Then, by bounded convergence,

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \left|G\Lambda_n(g_n) - \Lambda(g)\right| \\
&\quad = \overline{\lim_{n\to\infty}} \left| \int_D R_n(x)^\alpha g_n(T_n x)\rho(x)\,dx - \int_D g(x)\rho(x)^{1+\alpha}\,dx \right| \\
&\quad \le \overline{\lim_{n\to\infty}} \int_D \left|R_n(x)^\alpha g_n(T_n x) - \rho(x)^\alpha g(x)\right|\rho(x)\,dx.
\end{aligned}$$

Since $\rho$ is bounded, we now argue that

$$\overline{\lim_{n\to\infty}} \int_D \left|R_n(x)^\alpha g_n(T_n x) - \rho(x)^\alpha g(x)\right|dx = 0.$$

By adding and subtracting $\rho(x)^\alpha g_n(T_n x)$, we obtain

$$\int_D |R_n(x)^\alpha g_n(T_n x) - \rho(x)^\alpha g(x)|\, dx$$
$$\leq \int_D |R_n(x)^\alpha - \rho(x)^\alpha| g_n(T_n x)\, dx + \int_D |g_n(T_n x) - g(x)| \rho^\alpha(x)\, dx.$$

With respect to the first term on the right-hand side, by assumption, $g_n(T_n x)$ is uniformly bounded. Also, the sequence $R_n$ is bounded above and below, and converges Lebesgue a.e. to $\rho$, so by dominated convergence the integral vanishes in the limit. With respect to the second term on the right-hand side, suppose $(\nu_n, g_n) \xrightarrow{TL^1} (\nu, g)$ as $n \to \infty$. Since $\rho$ is bounded above and below, we have $\rho^\alpha$ is bounded, and the corresponding integral, by the characterization of $TL^1$ convergence in Lemma 3.1, also vanishes in the limit.

Hence, when the kernel $\eta$ is a simple function, we have that

$$\lim_{n\to\infty} G\Lambda_n(g_n) = \Lambda(g). \tag{6.13}$$

*Step 3.* Now, we consider general $\eta$ satisfying properties (K1)–(K4). We first approximate $\eta$ by simple functions $\eta^{(k)}$, satisfying (K2)–(K4), with $\eta^{(k)} \leq \eta$ and $\eta^{(k)} \to \eta$ pointwise.

Let $G\Lambda_n^{(k)}(g_n) = \frac{1}{n}\sum_{i=1}^n (\frac{1}{n-1}\sum_{j=1}^n \eta_{\varepsilon_n}^{(k)}(X_i - X_j))^\alpha g_n(X_i)$, and $\lambda_k = \int_D \eta^{(k)}(x)\, dx$. Then, by (6.13), we have

$$\lim_{n\to\infty} \frac{1}{\lambda_k^\alpha} G\Lambda_n^{(k)}(g_n) = \Lambda(g).$$

Because $\eta^{(k)} \leq \eta$, and the sequence $\{g_n\}_{n\in\mathbb{N}}$ is assumed nonnegative, we have that

$$\liminf_{n\to\infty} \frac{1}{\lambda_k^\alpha} G\Lambda_n(g_n) \geq \Lambda(g),$$

and taking the limit $\lambda_k \to 1$ as $k \to \infty$ gives

$$\liminf_{n\to\infty} G\Lambda_n(g_n) \geq \Lambda(g). \tag{6.14}$$

Likewise, consider approximating $\eta$ by simple functions $\eta^{(k)}$ satisfying (K2)–(K4), with $\eta^{(k)} \geq \eta$ and $\eta^{(k)} \to \eta$ pointwise. Then, similarly, we obtain that

$$\limsup_{n\to\infty} G\Lambda_n(g_n) \leq \Lambda(g). \tag{6.15}$$

Combining inequalities (6.14) and (6.15) gives (6.9). □

LEMMA 6.4. *On the probability* 1 *set* $\Omega_0$, *the following statement holds*: *Given any sequence* $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ *such that* $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$, *where* $\mathcal{V}_n = (\nu_n, \mathcal{U}_n) \in M_n(D)$ *and* $\mathcal{V} = (\nu, \mathcal{U}) \in M(D)$, *then*

$$F(\mathcal{U}) \leq \liminf_{n\to\infty} F_n(\mathcal{U}_n). \tag{6.16}$$

PROOF. Fix a realization in the probability 1 set $\Omega_0$. If $F(\mathcal{U}) = 0$, the above inequality holds trivially. We now consider the other case when $F(\mathcal{U}) = \infty$. Recalling the definitions of $F$ and $\Lambda$ earlier in the section, this means that there is a $1 \le k \le K$ such that $u_k = \mathbb{1}_{U_k}$ satisfies $\Lambda(u_k - 1/K) \neq 0$. Let $\delta = \Lambda(u_k - 1/K)^2 > 0$.

Now, by Corollary 3.3, $\nu_n \xrightarrow{w} \nu$, and so $(\nu_n, 1/K) \xrightarrow{TL^1} (\nu, 1/K)$ by Lemma 3.1, as $n \to \infty$. Therefore, by Lemma 6.3, as $n \to \infty$ we have

$$G\Lambda_n(1/K) = \frac{1}{K}\frac{S}{n(n-1)^\alpha} \to \frac{1}{K}\Lambda(1) = \frac{1}{K}\int_D \rho^{1+\alpha}(x)\,dx.$$

By decomposing $G\Lambda_n(u_{n,k} - 1/K) = G\Lambda_n(u_{n,k}) - G\Lambda_n(1/K)$ and noting Lemma 6.3 again, it follows that, if $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$, then

$$\lim_{n\to\infty}\sum_{k=1}^{K}\bigl(G\Lambda_n(u_{n,k} - 1/K)\bigr)^2 = \sum_{k=1}^{K}\bigl(\Lambda(u_k - 1/K)\bigr)^2.$$

In particular, there is an $N > 0$, depending on the realization, such that, for $n > N$, we have

$$\sum_{k=1}^{K}\bigl(G\Lambda_n(u_k - 1/K)\bigr)^2 \ge \delta/2.$$

Since

$$F_n(\mathcal{U}_n) = \frac{1}{\varepsilon_n}\frac{n^2(n-1)^{2\alpha}}{S^2}\sum_{k=1}^{K}\bigl(G\Lambda_n(u_{n,k} - 1/K)\bigr)^2,$$

and $\frac{n^2(n-1)^{2\alpha}}{S^2} \to (\int_D \rho^{1+\alpha}(x)\,dx)^{-2}$, it follows that

$$\liminf_{n\to\infty} F_n(\mathcal{U}_n) \ge \liminf_{n\to\infty}\frac{C}{\varepsilon_n}\sum_{k=1}^{K}\bigl(G\Lambda_n(u_k - 1/K)\bigr)^2 \ge \liminf_{n\to\infty}\frac{C\delta}{2\varepsilon_n} = \infty.$$

Hence, in this case also, inequality (6.16) holds. □

LEMMA 6.5. *On the probability* 1 *set* $\Omega_0$, *the following statement holds*: *Given any sequence* $\{u_n\}_{n\in\mathbb{N}}$ *such that* $(\nu_n, u_n) \xrightarrow{TL^1} (\nu, u)$, *then*

$$\sigma_\eta\,\mathrm{TV}(u;\rho^2) \le \liminf_{n\to\infty}\mathrm{GTV}_n(u_n),$$

*where* $\sigma_\eta = \int_{\mathbb{R}^d}\eta(x)|x_1|\,dx$.

PROOF. The desired statement follows the same argument given for the liminf inequality for the Gamma convergence stated in Theorem 1.1 in [38]; see Step 3 of Section 5.1 of [38]. There, the probability 1 set is $\Omega_0$. We note this proof, although stated for $d \ge 2$, also holds in $d = 1$ with the same notation. More remarks can be found in the initial arXiv version of this article [22]. □

LEMMA 6.6. *On the probability* 1 *set* $\Omega_0$, *the following statement holds*: *Given any sequence* $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ *such that* $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$, *where* $\mathcal{V}_n = (\nu_n, \mathcal{U}_n) \in M_n(D)$ *and* $\mathcal{V} = (\nu, \mathcal{U}) \in M(D)$, *then*

$$\mathrm{TV}(\mathcal{U}) \le \liminf_{n\to\infty} \mathrm{TV}_n(\mathcal{U}_n).$$

PROOF. Fix a realization in the probability 1 set $\Omega_0$. Note that $\nu_n \xrightarrow{w} \nu$ by Corollary 3.3, and so $(\nu_n, 1) \xrightarrow{TL^1} (\nu, 1)$ by Lemma 3.1, as $n \to \infty$. Hence, by Lemma 6.3, applied with $\alpha = 1$ and $g_n \equiv 1$, we have that $G\Lambda_n(1) = 2m/(n(n-1)) \to \int_D \rho^2(x)\,dx$, as $n \to \infty$.

Recall that $\mathrm{TV}_n(\mathcal{U}_n) = (n(n-1)/4m)\sum_{k=1}^K \mathrm{GTV}_n(u_{n,k})$. If $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$, by Lemma 6.5, we have

$$\sigma_\eta \mathrm{TV}(u_k; \rho) \le \liminf_{n\to\infty} \mathrm{GTV}_n(u_{n,k})$$

for $1 \le k \le K$. It follows that

$$\mathrm{TV}(\mathcal{U}) = C_{\eta,\rho} \sum_{k=1}^K \mathrm{TV}\big(u_k; \rho^2\big) \le \liminf_{n\to\infty} \frac{n(n-1)}{4m} \sum_{k=1}^K \mathrm{GTV}_n(u_{n,k}),$$

where $C_{\eta,\rho} = \sigma_\eta/(2\int_D \rho^2(x)\,dx)$. □

LEMMA 6.7. *On the probability* 1 *set* $\Omega_0$, *the following statement holds*: *Given any sequence* $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ *in* $(TL^1(D))^K$ *and* $\mathcal{V} \in (TL^1)^K$ *such that* $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$ *as* $n \to \infty$, *then*

$$E(\mathcal{V}) \le \liminf_{n\to\infty} E_n(\mathcal{V}_n).$$

PROOF. Fix a realization in the probability 1 set $\Omega_0$. Without loss of generality, we may assume that $\mathcal{V}_n = (\nu_n, \mathcal{U}_n) \in M_n(D)$, as $E_n(\mathcal{V}_n)$ diverges otherwise. We will also assume $\liminf E_n(\mathcal{V}_n) < \infty$, as otherwise the statement is trivial. Now, if $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$, by Lemma 6.5, we have $\mathcal{V} = ((\mu_k, u_k))_{k=1}^K$ satisfies $\sum_{k=1}^K \mathrm{TV}(u_k, \rho^2) < \infty$. By Lemma 6.2, it follows then that $\mathcal{V} = (\nu, \mathcal{U}) \in M(D)$. Also, by Lemmas 6.4 and 6.6, we have

$$\mathrm{TV}(\mathcal{U}) \le \liminf_{n\to\infty} \mathrm{TV}_n(\mathcal{U}_n) \quad \text{and} \quad F(\mathcal{U}) \le \liminf_{n\to\infty} F_n(\mathcal{U}_n).$$

Adding these two liminf inequalities gives $E(\mathcal{V}) \le \liminf_{n\to\infty} E_n(\mathcal{V}_n)$. □

6.3. *Existence of recovery sequence.* The a.s. recovery sequence associated with $(\nu, \mathcal{U})$ in $M(D)$ will be $\{(\nu_n, \mathcal{U}_n)\}_{n\in\mathbb{N}} \subset M_n(D)$, where $\mathcal{U}_n$ is the partition of

$\mathcal{X}_n$ induced by $\mathcal{U}$. However, before proving this in Lemma 6.9, we first establish a preliminary result.

LEMMA 6.8. *Fix $u \in L^1(D)$, and let $\{T_n\}_{n\in\mathbb{N}}$ be the transport maps given in Proposition 3.2. Then, a.s.,*

$$u \circ T_n \xrightarrow{L^1} u \qquad \text{as } n \to \infty.$$

PROOF. Let $u_\varepsilon$ be a Lipschitz function such that $\int_D |u(x) - u_\varepsilon(x)|\,dx < \varepsilon$. Let $A > 0$ be a lower bound for $\rho$ on $D$. It follows that

$$\begin{aligned} A \int_D \big|u(T_n x) - u(x)\big|\,dx \\ \le \int_D \big|u(T_n x) - u_\varepsilon(T_n x)\big|\rho(x)\,dx \\ + \int_D \big|u_\varepsilon(T_n x) - u_\varepsilon(x)\big|\rho(x)\,dx + \int_D \big|u_\varepsilon(x) - u(x)\big|\rho(x)\,dx. \end{aligned} \tag{6.17}$$

We rewrite the first term in the right-hand side of (6.17) in terms of the data set $\mathcal{X}_n$:

$$\int_D \big|u(T_n x) - u_\varepsilon(T_n x)\big|\rho(x)\,dx = \frac{1}{n}\sum_{i=1}^{n} \big|u(X_i) - u_\varepsilon(X_i)\big|.$$

By the strong law of large numbers, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |u(X_i) - u_\varepsilon(X_i)| = \int_D |u(x) - u_\varepsilon(x)|\rho(x)\,dx < \varepsilon$, almost surely.

For the second term in (6.17), let $C$ be the Lipschitz constant for $u_\varepsilon$. Then a.s., by Proposition 3.2,

$$\limsup_{n\to\infty} \int_D \big|u_\varepsilon(T_n x) - u_\varepsilon(x)\big|\rho(x)\,dx \le \limsup_{n\to\infty} C\|\rho\|_{L^\infty}\|T_n - Id\|_{L^\infty} = 0.$$

Taking limits in (6.17) therefore gives a.s. that

$$\limsup_{n\to\infty} \int_D \big|u(T_n x) - u(x)\big|\,dx \le 2A^{-1}\varepsilon.$$

Letting $\varepsilon$ go to zero along a countable sequence establishes the lemma. □

LEMMA 6.9. *Let $\mathcal{V} \in (TL^1(D))^K$. If $\mathcal{V} = (\nu, \mathcal{U}) \in M(D)$, let $\mathcal{V}_n = (\nu_n, \mathcal{U}_n) \in M_n(D)$, where $\mathcal{U}_{n,k} = \mathcal{U}_k \cap \mathcal{X}_n$ for $1 \le k \le K$ and $n \ge 1$. On the other hand, if $\mathcal{V} \notin M(D)$, let $\mathcal{V}_n = \mathcal{V}$ for $n \ge 1$.*

*Then, a.s., as $n \to \infty$,*

$$\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V} \quad \text{and} \quad E_n(\mathcal{V}_n) \to E(\mathcal{V}).$$

PROOF. In the case that $\mathcal{V} \notin M(D)$, since the sequence $\{\nu_n\}_{n\in\mathbb{N}}$ is composed of distinct elements, for all large $n$, $\mathcal{V} \notin M_n(D)$, and hence $E_n(\mathcal{V}) = E(\mathcal{V}) = \infty$.

Suppose now that $\mathcal{V} = (\nu, \mathcal{U}) \in M(D)$. In the following, we will use the fact that $u_{n,k}(x) = \mathbb{1}_{U_{n,k}}(x) = \mathbb{1}_{U_k}(x) = u_k(x)$ when $x \in \mathcal{X}_n$. To show a.s. that $\mathcal{V}_n \xrightarrow{(TL^1)^K} \mathcal{V}$ as $n \to \infty$, by Lemma 3.1, it is enough to show a.s. that $\nu_n \xrightarrow{w} \nu$ and, for $1 \le k \le K$, that

$$\int_D |u_{n,k}(T_n x) - u_k(x)|\, d\nu(x) = \int_D |u_k(T_n x) - u_k(x)|\rho(x)\, dx \to 0, \tag{6.18}$$

as $n \to \infty$, since $T_n x \in \mathcal{X}_n$ implies $u_{n,k}(T_n x) = u_k(T_n x)$.

The a.s. convergence $\nu_n \xrightarrow{w} \nu$ follows, for instance, by Corollary 3.3. On the other hand, the limit (6.18) follows by Lemma 6.8.

To show that a.s. $E_n(\mathcal{V}_n) \to E(\mathcal{V})$, we need to show that

$$\mathrm{TV}_n(\mathcal{U}_n) \xrightarrow{a.s.} \mathrm{TV}(\mathcal{U}) \quad \text{and} \quad F_n(\mathcal{U}_n) \xrightarrow{a.s.} F(\mathcal{U}), \tag{6.19}$$

as $n \to \infty$. Since condition (I2) on $\{\varepsilon_n\}_{n\in\mathbb{N}}$ implies condition (I1), we shall see that these limits in fact follow from three statements in the proof of Theorem 2.1.

In particular, recall the definitions (6.2) and (6.6) of $\mathrm{TV}_n$ and TV respectively. Then, since $u_{n,k} = u_k$ on $\mathcal{X}_n$, the limit $\mathrm{TV}_n(\mathcal{U}_n) \xrightarrow{a.s.} \mathrm{TV}(\mathcal{U})$, as $n \to \infty$, follows from (5.30).

With regards to the $F_n$ convergence, we consider two possibilities. First, suppose that $\mathcal{U}$ is balanced. Then, recalling the definition (6.1), and again noting that $u_{n,k} = u_k$ on $\mathcal{X}_n$, it follows from (5.31) that $F_n(\mathcal{U}_n) \xrightarrow{a.s.} F(\mathcal{U}) = 0$ as $n \to \infty$.

Suppose now that $\mathcal{U}$ is not balanced, so that $\sum_{k=1}^K (\mu(U_k) - 1/K)^2 \neq 0$. Then (5.29) implies, as $n \to \infty$, that $F_n(\mathcal{U}_n) \xrightarrow{a.s.} F(\mathcal{U}) = \infty$. Having considered all cases, (6.19) is established. □

6.4. *Compactness and proof of Theorem* 2.3. After a few preliminary estimates, we supply the needed compactness property for the graph energies $\{E_n\}_{n\in\mathbb{N}}$ in Theorem 6.12. Then we prove Theorem 2.3 at the end of the section.

LEMMA 6.10. *Let $\{u_n\}_{n\in\mathbb{N}}$ be a sequence of indicator functions on $D$, $u_n \in I(D)$, and $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers with $\lim_{n\to\infty} \varepsilon_n = 0$. If*

$$\sup_{n\in\mathbb{N}} \mathrm{TV}_{\varepsilon_n}(u_n; \mathbb{1}_D) < \infty,$$

*then $\{u_n\}_{n\in\mathbb{N}}$ is relatively compact with respect to the $L^1$ topology.*

PROOF. This result is a special case of Proposition 4.6 of [38] and Theorem 3.1 of [2], which treat more involved settings. See, however, the initial arXiv version of this article [22] for a streamlined argument in our situation, which makes use of the assumptions that the functions $\{u_n\}_{n\in\mathbb{N}}$ are $\{0,1\}$-valued, and that the kernel $\eta$ is compactly supported. □

Recall that $\Omega_0$ denotes the probability 1 set of realizations of $\{X_i\}_{i\in\mathbb{N}}$ under which Proposition 3.2 holds.

LEMMA 6.11. *Suppose $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfies condition* (I2). *On the probability* 1 *set* $\Omega_0$, *the following holds: Given any sequence $\{u_n\}_{n\in\mathbb{N}}$ of indicator functions on the data points, $u_n \in I_n(D)$, if*

$$\sup_{n\in\mathbb{N}} \mathrm{GTV}_n(u_n) < \infty,$$

*then $\{(\nu_n, u_n)\}_{n\in\mathbb{N}}$ is relatively compact with respect to the $TL^1$ topology.*

PROOF. We begin as in the proof of Lemma 6.3. Fix a realization in the probability 1 set $\Omega_0$. Suppose that $\eta$ is of the form $\eta(x) = a$ for $|x| < b$ and $\eta(x) = 0$ for $|x| > b$. Let $\tilde{\varepsilon}_n := \varepsilon_n - 2\frac{\|Id-T_n\|_{L^\infty}}{b}$, with respect to the transport maps $\{T_n\}_{n\in\mathbb{N}}$. Then, for all large $n$, $\tilde{\varepsilon}_n > 0$, and we have inequality (6.11),

$$\eta\left(\frac{x-y}{\tilde{\varepsilon}_n}\right) \le \eta\left(\frac{T_n x - T_n y}{\varepsilon_n}\right) \qquad \text{Lebesgue a.e. } (x,y) \in D \times D$$

Let $A > 0$ be a lower bound for $\rho$ on $D$. Then

$$\begin{aligned} A^2 &\int_D \eta\left(\frac{x-y}{\tilde{\varepsilon}_n}\right)\big|u_n(T_n x) - u_n(T_n y)\big|\,dx\,dy \\ &\le \int_D \eta\left(\frac{x-y}{\tilde{\varepsilon}_n}\right)\big|u_n(T_n x) - u_n(T_n y)\big|\rho(x)\rho(y)\,dx\,dy \\ &\le \int_D \eta\left(\frac{T_n x - T_n y}{\varepsilon_n}\right)\big|u_n(T_n x) - u_n(T_n y)\big|\rho(x)\rho(y)\,dx\,dy \\ &= \varepsilon_n^{d+1}\,\mathrm{GTV}_n(u_n). \end{aligned}$$

The above inequality is equivalent to

$$(\tilde{\varepsilon}_n/\varepsilon_n)^{d+1}\,\mathrm{TV}_{\tilde{\varepsilon}_n}(u_n \circ T_n; \mathbb{1}_D) \le \mathrm{GTV}_n(u_n).$$

Since $\lim_{n\to\infty} \tilde{\varepsilon}_n/\varepsilon_n = 1$, the bound $\sup_n \mathrm{GTV}_n(u_n) < \infty$ implies that

$$\sup_n \mathrm{TV}_{\tilde{\varepsilon}_n}(u_n \circ T_n; \mathbb{1}_D) < \infty.$$

It follows, by Lemma 6.10, that the family $\{u_n \circ T_n\}_{n\in\mathbb{N}}$ is relatively compact with respect to the $L^1$ topology. Since, by Corollary 3.3, $\nu_n \xrightarrow{w} \nu$, we conclude by Lemma 3.1 that $\{(\nu_n, u_n)\}_{n\in\mathbb{N}}$ is relatively compact in $TL^1$.

Suppose now $\eta$ is an arbitrary kernel satisfying assumptions (K1)–(K4). Since $\eta$ is continuous at zero, and $\eta(0) > 0$, there is some radius $R$ such that $\tilde{\eta} = \frac{\eta(0)}{2}\mathbb{1}_{|x|<R}$

satisfies $\tilde{\eta} \le \eta$. Let $c = \int_{\mathbb{R}^d} \tilde{\eta}(x)\,dx$. Then, if $\widetilde{\mathrm{GTV}}_n$ denotes the graph total variation associated to the kernel $\tilde{\eta}/c$ (instead of $\eta$), we have

$$\mathrm{GTV}_n(u_n) \ge c\widetilde{\mathrm{GTV}}_n(u_n).$$

Since $\sup_n \mathrm{GTV}_n(u_n) < \infty$ implies $\sup_n \widetilde{\mathrm{GTV}}_n(u_n) < \infty$, it follows from our previous discussion that the sequence $\{(\nu_n, u_n)\}_{n\in\mathbb{N}}$ is relatively compact in $TL^1(D)$. □

THEOREM 6.12. *Suppose* $\{\varepsilon_n\}_{n\in\mathbb{N}}$ *satisfies condition* (I2). *On the probability* 1 *set* $\Omega_0$, *the following holds*: *Given any sequence* $\{\mathcal{V}_n\}_{n\in\mathbb{N}} \subset (TL^1)^K$, *if*

$$\sup_{n\in\mathbb{N}} E_n(\mathcal{V}_n) < \infty,$$

*then* $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ *is relatively compact with respect to the* $(TL^1)^K$ *topology.*

PROOF. Fix a realization in the probability 1 set $\Omega_0$. Let $\mathcal{V}_n$ be a sequence with $\sup_n E_n(\mathcal{V}_n) < \infty$. By definition of $E_n$, it follows that $\mathcal{V}_n = (\nu_n, \mathcal{U}_n) \in M_n(D)$ where $\mathcal{U}_n = \{U_{n,k}\}_{k=1}^K$ for $n \in \mathbb{N}$. By Corollary 3.3, we have $\nu_n \xrightarrow{w} \nu$. Since $(\nu_n, \mathbb{1}_D) \xrightarrow{TL^1} (\nu, \mathbb{1}_D)$ by Lemma 3.1, we have, by Lemma 6.3, that $G\Lambda_n(1) = 2m/(n(n-1)) \to \int_D \rho^2(x)\,dx$. Recall now that $E_n(\mathcal{V}_n) = \mathrm{TV}_n(\mathcal{U}_n) + F_n(\mathcal{U}_n)$ and

$$\mathrm{TV}_n(\mathcal{U}_n) = \frac{n(n-1)}{4m} \sum_{k=1}^K \mathrm{GTV}_n(u_{n,k}),$$

where $u_{n,k} = \mathbb{1}_{U_{n,k}}$ for $1 \le k \le K$. Hence, given that $\sup_{n\in\mathbb{N}} E_n(\mathcal{V}_n) < \infty$, we have

$$\sup_{n\in\mathbb{N}} \mathrm{GTV}_n(u_{n,k}) < \infty,$$

for $1 \le k \le K$. Thus, by Lemma 6.11, the collection $\{(\nu_n, u_{n,k})\}_{n\in\mathbb{N}}$ is relatively compact in $TL^1$ for $1 \le k \le K$. Thus, $\{\mathcal{V}_n = ((\nu_n, u_{n,k}))_{k=1}^K = (\nu_n, \mathcal{U}_n)\}_{n\in\mathbb{N}}$ is relatively compact in $(TL^1)^K$. □

PROOF OF THEOREM 2.3. We have seen in Theorem 6.1 that

$$E_n \xrightarrow{\Gamma((TL^1)^K)} E,$$

in the sense of Definition 3.8. By Theorem 6.12, the graph energies $E_n$ have the compactness property according to Definition 3.9. Also, as noted in Section 6.1, the energy $E$ is not identically infinite.

For each realization $\{X_i\}_{i\in\mathbb{N}}$, let $\mathcal{U}_n^* \in \arg\max_{|\mathcal{U}_n| \le K} Q_n(\mathcal{U}_n)$ be an optimal partition. Then, by the discussion in Section 6.1, $\mathcal{U}_n^*$ is a minimizer of $F_n + \mathrm{TV}_n$, and so $\mathcal{V}_n = (\nu_n, \mathcal{U}_n^*) \in M_n(D)$ is a minimizer of $E_n$. The sequence $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ is also bounded in $(TL^1)^K$: Indeed, we have $\int_D |x|\, d\nu_n(x) \le \sup_{x\in D}|x|$ and, for $u^*_{n,k} = \mathbb{1}_{U^*_{n,k}}$, $\|u^*_{n,k}\|_{L^1} \le \mathrm{vol}(D)\|u^*_{n,k}\|_{L^\infty} \le \mathrm{vol}(D)$.

Hence, on the full set of realizations $\{X_i\}_{i\in\mathbb{N}}$, denoted as $\Omega$, the sequence $x_n = \mathcal{V}_n$, in the metric space $(TL^1)^K$, satisfies the hypotheses of Theorem 3.11. Therefore, with respect to realizations $\{X_i\}_{i\in\mathbb{N}}$ on a probability 1 set $\Omega^*$, $\mathcal{V}_n$ converges in $(TL^1)^K$, perhaps along a subsequence, to a limit $\mathcal{V}$, which is a minimizer of $E$, and is therefore of the form $\mathcal{V} = (\nu, \mathcal{U}^*)$. Since the "liminf" inequality, Lemma 6.7, holds on $\Omega_0$, we note that $\Omega^* \subset \Omega_0$. Moreover, by Corollary 3.3, $\nu_n \xrightarrow{w} \nu$ on $\Omega^*$. Therefore, by (3.4), on $\Omega^*$, $\mathcal{U}^*_n$ converges weakly, perhaps along a subsequence, to the limit $\mathcal{U}^*$, which is an optimal partition of the continuum problem (2.6) with $\phi = \rho^2$ and $d\mu = \rho^{1+\alpha}/\int_D \rho^{1+\alpha}(x)\, dx$.

In fact, on $\Omega^*$, the distances $\beta_n := \inf\{d_{(TL^1)^K}(\mathcal{V}_n, \mathcal{V}) : \mathcal{V} \in \arg\min E\}$, where $d_{(TL^1)^K}$ is the product metric for $(TL^1)^K$ convergence, satisfy $\beta_n \to 0$ as $n \to \infty$. For if not, there is a subsequence $\{n_m\}_{m\in\mathbb{N}}$ with $\beta_{n_m} \to \beta > 0$. However, by the above discussion one may find a further subsequence $\{n'_m\}_{m\in\mathbb{N}}$ with $\beta_{n'_m} \to 0$, a contradiction.

Moreover, if problem (2.6) has a unique solution $\mathcal{U}^* = \{U^*_k\}_{k=1}^K$, modulo permutations, then $\arg\min E = \{((\nu, u^*_{\pi(k)}))_{k=1}^K : \pi \in \mathrm{Sym}(K)\}$, where we recall $\mathrm{Sym}(K)$ denotes the permutations of $\{1, \dots, K\}$ and $u^*_k = \mathbb{1}_{U^*_k}$ for $1 \le k \le K$. Thus, on the probability 1 set $\Omega^*$, since $\beta_n \to 0$, one may construct a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ of permutations such that, as $n \to \infty$, $((\nu_n, u^*_{n,\pi_n(k)}))_{k=1}^K$ converges in $(TL^1)^K$ to $(\nu, \mathcal{U}^*) = ((\nu, u^*_k))_{k=1}^K$. Hence, by (3.4), $\mathcal{U}^*_n \xrightarrow{w} \mathcal{U}^*$, in the sense of (2.11). □

## APPENDIX

**A.1. Approximation lemma.** Recall that $\rho_\varepsilon(x) := \int_D \eta_\varepsilon(x-y)\rho(y)\, dy$.

LEMMA A.13. *Under the standing assumptions on $\rho$, $D$ and $\eta$ in Section 2.3, we have the following*:

(i) *$\rho_\varepsilon$ converges pointwise to $\rho$ as $\varepsilon \downarrow 0$.*

(ii) *There exists a constant $C$ such that, for sufficiently small $\varepsilon$,*

$$\int_D |\rho_\varepsilon(x) - \rho(x)|\, dx \le C\varepsilon. \tag{A.1}$$

(iii) *There exist constants $a, b$ such that, for sufficiently small $\varepsilon$,*

$$0 < a \le \rho_\varepsilon(x) \le b \qquad \textit{for all } x \in D.$$

PROOF. The pointwise convergence in item (i) follows from continuity of $\rho$.

We now focus attention on item (ii), inequality (A.1). For the moment, fix $x \in D$. Since $\eta$ is compactly supported, we take $R$ such that $\eta(z) = 0$ for $|z| > R$. Then for $0 < \varepsilon < \mathrm{dist}(x, \partial D)/R$, we have, since $\int_{\mathbb{R}} \eta(x) dx = 1$, that

$$\rho(x) = \int_{x+z\in D} \eta_\varepsilon(z)\rho(x)\,dz \quad \text{and so}$$

$$\rho_\varepsilon(x) - \rho(x) = \int_{x+z\in D} \eta_\varepsilon(z)\big(\rho(x+z) - \rho(x)\big)\,dz.$$

Let $L$ be a Lipschitz constant for $\rho$. Then

$$\begin{aligned} \big|\rho_\varepsilon(x) - \rho(x)\big| &\le L \int_{x+z\in D} \eta_\varepsilon(z)|z|\,dz \le L \int_{|z|\le \mathrm{diam}(D)} \eta_\varepsilon(z)|z|\,dz \\ &= L\varepsilon \int_{|z|\le \mathrm{diam}(D)/\varepsilon} \eta(z)|z|\,dz \le L\varepsilon \int_{\mathbb{R}^d} \eta(z)|z|\,dz. \end{aligned}$$

Because $\int_{\mathbb{R}} \eta(x)\,dx = 1$ and $\eta(z) = 0$ for $|z| > R$, the above implies

$$\big|\rho_\varepsilon(x) - \rho(x)\big| \le LR\varepsilon. \tag{A.2}$$

Let $D_{R\varepsilon} = \{x \in D | \,\mathrm{dist}(x, \partial D) < R\varepsilon\}$ and $\partial_{R\varepsilon} D = D \setminus D_{R\varepsilon}$. We write

$$\int_D \big|\rho_\varepsilon(x) - \rho(x)\big|\,dx = \int_{D_{R\varepsilon}} \big|\rho_\varepsilon(x) - \rho(x)\big|\,dx + \int_{\partial_{R\varepsilon} D} \big|\rho_\varepsilon(x) - \rho(x)\big|\,dx,$$

and consider the two terms separately. On $D_{R\varepsilon}$, applying inequality (A.2) yields

$$\int_{D_{R\varepsilon}} \big|\rho_\varepsilon(x) - \rho(x)\big|\,dx \le \mathrm{vol}(D) LR\varepsilon. \tag{A.3}$$

For the second integral, note that because the boundary is Lipschitz, there is a $\varepsilon_0 > 0$ and constant $C$ such that $\mathrm{vol}(\partial D_{R\varepsilon}) \le CR\varepsilon$ for $0 < \varepsilon < \varepsilon_0$. It follows that

$$\int_{\partial_{R\varepsilon} D} \big|\rho_\varepsilon(x) - \rho(x)\big|\,dx \le 2C\|\rho\|_{L^\infty} R\varepsilon. \tag{A.4}$$

Combining (A.3) and (A.4), it follows, for sufficiently small $\varepsilon > 0$, that $\int_D |\rho_\varepsilon(x) - \rho(x)|\,dx \le CR\varepsilon$, where $C$ is a constant independent of $\varepsilon$.

For item (iii) of the lemma, note that because the boundary of $D$ is Lipschitz, there exists constants $r_0, r_1$ such that, for any $x \in D$ and $0 < r < r_1$,

$$0 < r_0 < \frac{\mathrm{vol}(D \cap B(x, r))}{\mathrm{vol}(B(x, r))},$$

where $B(x, r)$ denotes the ball of radius $r$ centered at $x$ (cf. the discussion about cone conditions in Section 4.11 of [1]).

By assumption, $\rho$ is bounded above and below: $0 < A \le \rho(\cdot) \le B$. Also, by assumption (K3), $\eta$ is continuous at zero and $\eta(0) > 0$, so we may take $r$ so that $0 < \eta(0)/2 \le \eta(z)$ for $|z| < r$. Let $\tilde{\eta}(x) = \eta(0)\mathbb{1}_{B(x,r)}/2\mathrm{vol}(B(x, r))$. Then

$$0 < Ar_0\eta(0)/2 \le A \int_D \tilde{\eta}(x - y)\,dy \le \int_D \eta_\varepsilon(x - y)\rho(y)\,dy = \rho_\varepsilon(x).$$

Since $\int_{\mathbb{R}} \eta_\varepsilon(x)dx = 1$, the bound $\rho_\varepsilon(x) \le B$ also holds, completing the item. □

**A.2. Estimates for GTV Limsup.** Recalling the notation in the proof of Lemma 5.2, we let $f_n(X_i, X_j) = \frac{1}{\varepsilon_n}\eta_{\varepsilon_n}(X_i - X_j)|u(X_i) - u(X_j)|$, as well as $l_n(X_i) = \frac{1}{\varepsilon_n}\int_D \eta_{\varepsilon_n}(X_i - y)|u(X_i) - u(y)|\rho(y)\,dy$, and $m_n(X_j) = \frac{1}{\varepsilon_n} \times \int_D \eta_{\varepsilon_n}(x - X_j)|u(x) - u(X_j)|\rho(x)\,dx$. Here, $u \in I(D)$ satisfies $\mathrm{TV}(u; \rho^2) < \infty$.

LEMMA A.14. *There exists a constant $C$ such that, for sufficiently large $n$, we have the following bounds*:

$$|\mathbb{E} f_n| \le C, \qquad \|f_n\|_{L^2 \to L^2} \le C/\varepsilon_n, \qquad \mathbb{E} f_n^2 \le C/\varepsilon_n^{d+1}, \qquad \mathbb{E} l_n^2 \le C/\varepsilon_n,$$
$$\|\mathbb{E}_Y f_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2}, \qquad \|l_n\|_{L^\infty} \le C/\varepsilon_n, \qquad \|\mathbb{E}_X f_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2},$$
$$\mathbb{E} m_n^2 \le C/\varepsilon_n, \qquad \|f_n\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \|m_n\|_{L^\infty} \le C/\varepsilon_n.$$

PROOF. Recall that $\mathbb{E} f_n = \mathrm{TV}_{\varepsilon_n}(u; \rho)$. By Lemma 5.1, this converges to $\mathrm{TV}(u; \rho^2) < \infty$ as $n \to \infty$, and hence $|\mathbb{E} f_n| \le C$.

To address $\mathbb{E} f_n^2$, we have

$$\mathbb{E} f_n^2 = \int_D \int_D \frac{1}{\varepsilon_n^2}\bigl(\eta_{\varepsilon_n}(x - y)\bigr)^2 \bigl|u(x) - u(y)\bigr|^2 \rho(x)\rho(y)\,dx\,dy.$$

Since $\eta$ is bounded above, we have $(\eta_{\varepsilon_n}(x - y))^2 \le C\eta_{\varepsilon_n}(x - y)/\varepsilon_n^d$. Because $u \in I(D)$, we have $|u(x) - u(y)|^2 = |u(x) - u(y)|$. Hence,

$$\begin{aligned}\mathbb{E} f_n^2 &\le C \int_D \int_D \frac{1}{\varepsilon_n^2}\frac{1}{\varepsilon_n^d}\eta_{\varepsilon_n}(x - y)\bigl|u(x) - u(y)\bigr|\rho(x)\rho(y)\,dx\,dy \\ &\le C\,\mathrm{TV}_{\varepsilon_n}(u; \rho)/\varepsilon_n^{d+1} \le C'/\varepsilon_n^{d+1}.\end{aligned}$$

Likewise, one may get the bound, using that $\rho$ is bounded,

$$\begin{aligned}\mathbb{E}_Y f_n^2 &= \int_D \frac{1}{\varepsilon_n^2}\bigl(\eta_{\varepsilon_n}(x - y)\bigr)^2\bigl|u(x) - u(y)\bigr|^2 \rho(y)\,dy \\ &\le C \int_D \eta_{\varepsilon_n}(x - y)\rho(y)\,dy/\varepsilon_n^{d+2} \le C'/\varepsilon_n^{d+2}.\end{aligned}$$

By the symmetry of $f_n$, this also gives $\mathbb{E}_X f_n^2 \le C'/\varepsilon_n^{d+2}$.

Recall that $\|f_n\|_{L^2 \to L^2}$ is given by

$$\|f_n\|_{L^2 \to L^2} = \sup\biggl\{\int_{D \times D} f_n(x, y)h(x)g(y)d\nu(x)d\nu(y) : \|h\|_{L^2(D,\nu)} \le 1, \|g\|_{L^2(D,\nu)} \le 1\biggr\}.$$

It is straightforward to show $\int_D |f_n(x, y)|\rho(x)\,dx \le C/\varepsilon_n$, and similarly for the integral with respect to $y$, where $C$ is some constant independent of $n$. Thus, with respect to the map $Jg(x) = \int_D f_n(x, y)g(y)\rho(y)\,dy$, by Theorem 6.18 of [30], we have $\|Jg\|_{L^2(D,\nu)} \le C\|g\|_{L^2(D,\nu)}/\varepsilon_n$, which implies $\|f_n\|_{L^2\to L^2} \le C/\varepsilon_n$.

Now considering $l_n$, we have

$$\mathbb{E}l_n^2 = \int_D \left(\frac{1}{\varepsilon_n}\int_D \eta_{\varepsilon_n}(x-y)\big|u(x)-u(y)\big|\rho(y)\,dy\right)^2 \rho(x)\,dx.$$

By Jensen's inequality, it follows that

$$\begin{aligned}\mathbb{E}l_n^2 &\le \frac{1}{\varepsilon_n^2}\int_{D\times D} \eta_{\varepsilon_n}(x-y)\big|u(x)-u(y)\big|^2\rho(y)\rho(x)\,dy\,dx\\ &= \frac{1}{\varepsilon_n^2}\int_{D\times D} \eta_{\varepsilon_n}(x-y)\big|u(x)-u(y)\big|\rho(y)\rho(x)\,dy\,dx \le \frac{1}{\varepsilon_n}\mathbb{E}f_n \le C/\varepsilon_n.\end{aligned}$$

Similarly, as $\rho$ is bounded, we have

$$\begin{aligned}\big|l_n(x)\big| &= \frac{1}{\varepsilon_n}\int_D \eta_{\varepsilon_n}(x-y)\big|u(x)-u(y)\big|\rho(y)\,dy\\ &\le \frac{1}{\varepsilon_n}\int_D \eta_{\varepsilon_n}(x-y)\rho(y)\,dy \le C/\varepsilon_n.\end{aligned}$$

The same argument applied to $m_n$ gives the required inequalities. □

Recall, from the proof of Lemma 5.2, that $h_n(X_i, X_j) = \frac{1}{\varepsilon_n}\eta_{\varepsilon_n}(X_i - X_j) \times |u(X_i) - u(X_j)| - \int_D f_n(X_i, y)\rho(y)\,dy - \int_D f_n(x, X_j)\rho(x)\,dx + \mathrm{TV}_{\varepsilon_n}(u;\rho)$.

COROLLARY A.15. *There exists a constant $C$, such that, for sufficiently large $n$, we have the following bounds*:

$$\mathbb{E}h_n^2 \le C/\varepsilon_n^{d+1}, \qquad \|\mathbb{E}_Y h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2}, \qquad \|\mathbb{E}_X h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2},$$
$$\|h_n\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \|h_n\|_{L^2\to L^2} \le C/\varepsilon_n.$$

PROOF. Since $h_n(X_i, X_j) = f_n(X_i, X_j) - l_n(X_i) - m_n(X_j) + \mathbb{E}f_n$, we have

$$\sqrt{\mathbb{E}h_n^2} \le \sqrt{\mathbb{E}f_n^2} + \sqrt{\mathbb{E}l_n^2} + \sqrt{\mathbb{E}_n m^2} + \sqrt{(\mathbb{E}f_n)^2}.$$

All terms in the right-hand side may be bounded by $\sqrt{C/\varepsilon_n^{d+1}}$, and hence $\mathbb{E}h_n^2 \le C/\varepsilon_n^{d+1}$.

Similarly, in the bound

$$\sqrt{\|\mathbb{E}_X h_n^2\|_{L^\infty}} \le \sqrt{\|\mathbb{E}_X f_n^2\|_{L^\infty}} + \sqrt{\|\mathbb{E}_X l_n^2\|_{L^\infty}} + \sqrt{\|\mathbb{E}_X m_n^2\|_{L^\infty}} + \sqrt{(\mathbb{E}f_n)^2},$$

all terms on the right are dominated by $\sqrt{C/\varepsilon_n^{d+2}}$, so $\|\mathbb{E}_X h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2}$.

By symmetry of $h_n$, this gives $\|\mathbb{E}_Y h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+2}$.

Likewise, a similar triangle inequality gives $\|h_n\|_{L^\infty} \le C/\varepsilon_n^{d+1}$.

For the last bound, we write

$$\begin{aligned}\|h_n\|_{L^2\to L^2} &\le \|f_n\|_{L^2\to L^2} + \|l_n\|_{L^2\to L^2} + \|m_n\|_{L^2\to L^2} + |\mathbb{E} f_n| \\ &\le \|f_n\|_{L^2\to L^2} + \|l_n\|_{L^\infty} + \|m_n\|_{L^\infty} + |\mathbb{E} f_n|,\end{aligned}$$

and note each term in the right-hand side is bounded by $C/\varepsilon_n^d$. □

**A.3. Estimates for GF Limsup.** Recall, from the proof of Lemma 5.5, the notation $f_n(X_i, X_j) = \frac{1}{\sqrt{\varepsilon_n}} \eta_{\varepsilon_n}(X_i - X_j) u(X_j)$, $l_n(X_i) = \frac{1}{\sqrt{\varepsilon_n}} \int_D \eta_{\varepsilon_n}(X_i - y) \times u(X_i)\rho(y)\,dy$, and $m_n(X_j) = \frac{1}{\sqrt{\varepsilon_n}} \int_D \eta_{\varepsilon_n}(x - X_j) u(x)\rho(x)\,dx$. Here, $u \in I(D)$.

LEMMA A.16. *There exists a constant $C$, such that, for sufficiently large $n$, we have the following bounds*:

$$|\mathbb{E} f_n| \le C/\varepsilon_n^{1/2}, \qquad \|f_n\|_{L^2\to L^2} \le C/\varepsilon_n^{1/2}, \qquad \mathbb{E} f_n^2 \le C/\varepsilon_n^{d+1},$$
$$\mathbb{E} l_n^2 \le C/\varepsilon_n, \qquad \|\mathbb{E}_Y f_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \|l_n\|_{L^\infty} \le C/\varepsilon_n^{1/2},$$
$$\|\mathbb{E}_X f_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \mathbb{E} m_n^2 \le C/\varepsilon_n,$$
$$\|f_n\|_{L^\infty} \le C/\varepsilon_n^{d+1/2}, \qquad \|m_n\|_{L^\infty} \le C/\varepsilon_n^{1/2}.$$

PROOF. These inequalities are easier than the ones in Lemma A.14, and follow from the boundedness of $u$ and $\rho_{\varepsilon_n}(x) = \int_D \eta_{\varepsilon_n}(x - y)\rho(y)\,dy$. □

Recall that $h_n(X_i, X_j) = f_n(X_i, X_j) - l_n(X_i) - m_n(X_j) + \mathbb{E} f_n$. The proof of the following is similar to that of Corollary A.15.

COROLLARY A.17. *There exists a constant $C$, such that, for sufficiently large $n$, we have the following bounds*:

$$\mathbb{E} h_n^2 \le C/\varepsilon_n^{d+1}, \qquad \|\mathbb{E}_Y h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+1}, \qquad \|\mathbb{E}_X h_n^2\|_{L^\infty} \le C/\varepsilon_n^{d+1},$$
$$\|h_n\|_{L^\infty} \le C/\varepsilon_n^{d+1/2}, \qquad \|h_n\|_{L^2\to L^2} \le C/\varepsilon_n^{1/2}.$$

**A.4. Transport distance in $d = 1$.** We define the transport maps $\{T_n\}_{n\in\mathbb{N}}$ in $d = 1$ and establish a bound on the rate at which $\|Id - T_n\|_{L^\infty} \to 0$ as $n \to \infty$.

Recall that by assumption $(M)$ of Section 2.3, $\nu$ is a probability measure on $D = (c, d)$ with density $\rho$ that is differentiable, Lipschitz and bounded above and below by positive constants. Further, $\rho$ is increasing in some interval with left endpoint $c$ and decreasing in some interval with right endpoint $d$. Let $F$ denote the distribution function of $\nu$.

Given a sample $\mathcal{X}_n = \{X_1, \dots, X_n\}$, we let $F_n$ denote the distribution function of the empirical measure $\nu_n = \frac{1}{n}\sum_{i=1}^n \upsilon_{X_i}$. We define $T_n$ by

$$T_n x = F_n^{-1}(F(x)), \tag{A.5}$$

where $F_n^{-1}(t) = \inf\{x \in \mathbb{R} : t \le F_n(x)\}$. The map $T_n$ is a valid transport map, that is, $T_{n\sharp}\nu = \nu_n$.

By the assumptions on $\rho$, it follows that

$$\sup_{c<x<d} F(x)\big(1 - F(x)\big)\big|\rho'(x)\big|/\rho^2(x) < \infty. \tag{A.6}$$

It is known (see Theorem 3 on p. 650 of [68]) that when (i) $\rho > 0$ on $(c, d)$, (ii) inequality (A.6) is satisfied, and (iii) $\rho$ is increasing in some interval with left endpoint $c$ and decreasing in some interval with right endpoint $d$, the standardized quantile process

$$\mathcal{Q}_n(t) := g(t)\sqrt{n}\big[F_n^{-1}(t) - F^{-1}(t)\big],$$

with $g(t) = \rho(F^{-1}(t))$, satisfies, almost surely,

$$\limsup_{n\to\infty} \sup_{0<t<1} \big|\mathcal{Q}_n(t)/\sqrt{2\log\log n}\big| \le 1. \tag{A.7}$$

Since $\rho$ is bounded above and below by nonnegative constants, so is $g$, and so we have constants $C, C' > 0$ such that

$$C\big|\mathcal{Q}_n(t)\big| \le \sqrt{n}\big|F_n^{-1}(t) - F^{-1}(t)\big| \le C'\big|\mathcal{Q}_n(t)\big|.$$

Since $\rho$ is positive, $F$ is strictly increasing, and hence we have

$$C \sup_{0<t<1} \big|\mathcal{Q}_n(t)\big| \le \sup_{c<x<d} \sqrt{n}\big|F_n^{-1}\big(F(x)\big) - F^{-1}\big(F(x)\big)\big| \le C' \sup_{0<t<1} \big|\mathcal{Q}_n(t)\big|.$$

Recalling our definition of $T_n$, this may be rewritten as

$$C \sup_{0<t<1} \big|\mathcal{Q}_n(t)\big| \le \sqrt{n}\|Id - T_n\|_{L^\infty} \le C' \sup_{0<t<1} \big|\mathcal{Q}_n(t)\big|.$$

In light of (A.7) and the above inequality, we obtain the following estimate.

PROPOSITION A.18. *There is a constant C such that, almost surely, the transport maps $T_n$, defined by* (A.5), *satisfy* $\lim_{n\to\infty} \frac{\sqrt{n}\|Id-T_n\|_{L^\infty}}{\sqrt{\log\log n}} \le C$.

**Acknowledgments.** We'd like to thank the Editors and referees for their constructive criticism.

## REFERENCES

[1] ADAMS, R. A. and FOURNIER, J. J. F. (2003). *Sobolev Spaces*, 2nd ed. *Pure and Applied Mathematics* (*Amsterdam*) **140**. Elsevier/Academic Press, Amsterdam. MR2424078

[2] ALBERTI, G. and BELLETTINI, G. (1998). A non-local anisotropic model for phase transitions: Asymptotic behaviour of rescaled energies. *European J. Appl. Math.* **9** 261–284. MR1634336

[3] AMBROSIO, L., FUSCO, N. and PALLARA, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. *Oxford Mathematical Monographs*. The Clarendon Press, Oxford Univ. Press, New York. MR1857292

[4] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2005). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, Basel. MR2129498

[5] ANTONIONI, A., EGLOF, M. and TOMASSINI, M. (2013). An energy-based model for spatial social networks. In *Advances in Artificial Life ECAL* 2013 226–231. MIT Press, Cambridge, MA.

[6] ARIAS-CASTRO, E. and PELLETIER, B. (2013). On the convergence of maximum variance unfolding. *J. Mach. Learn. Res.* **14** 1747–1770. MR3104494

[7] ARIAS-CASTRO, E., PELLETIER, B. and PUDLO, P. (2012). The normalized graph cut and Cheeger constant: From discrete to continuous. *Adv. in Appl. Probab.* **44** 907–937. MR3052843

[8] BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.

[9] BELKIN, M. and NIYOGI, P. (2008). Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. System Sci.* **74** 1289–1308. MR2460286

[10] BETTSTETTER, C. (2002). On the minimum node degree and connectivity of a wireless multihop network. In *Proceedings of the* 3rd *ACM International Symposium on Mobile Ad Hoc Networking & Computing* 80–91. ACM, New York.

[11] BICKEL, P. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[12] BLONDEL, V., GUILLAUME, J., LAMBIOTTE, R. and LEFEBVRE, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008** 10008–10020.

[13] BRAIDES, A. (2002). *Γ-Convergence for Beginners*. *Oxford Lecture Series in Mathematics and Its Applications* **22**. Oxford Univ. Press, Oxford. MR1968440

[14] BRAIDES, A. and GELLI, M. S. (2006). From discrete systems to continuous variational problems: An introduction. In *Topics on Concentration Phenomena and Problems with Multiple Scales*. *Lect. Notes Unione Mat. Ital.* **2** 3–77. Springer, Berlin. MR2267880

[15] BRAIDES, A. and TRUSKINOVSKY, L. (2008). Asymptotic expansions by Γ-convergence. *Contin. Mech. Thermodyn.* **20** 21–62. MR2398821

[16] BRAKKE, K. A. (1992). The surface evolver. *Exp. Math.* **1** 141–165. MR1203871

[17] BRANDES, U., DELLING, D., GAERTLER, M., GÖRKE, R., HOEFER, M., NIKOLOSKI, Z. and WAGNER, D. (2008). On modularity clustering. *IEEE Trans. Knowl. Data Eng.* **20** 172–188.

[18] CAÑETE, A. and RITORÉ, M. (2004). Least-perimeter partitions of the disk into three regions of given areas. *Indiana Univ. Math. J.* **53** 883–904. MR2086704

[19] CLAUSET, A., NEWMAN, M. and MOORE, C. (2004). Finding community structure in very large networks. *Phys. Rev. E* **70** 066111.

[20] COIFMAN, R. R. and LAFON, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** 5–30. MR2238665

[21] COX, S. J. and FLIKKEMA, E. (2010). The minimal perimeter for $N$ confined deformable bubbles of equal area. *Electron. J. Combin.* **17** Research Paper 45. MR2607331

[22] DAVIS, E. and SETHURAMAN, S. (2017). Consistency of modularity clustering on random geometric graphs. Available at arXiv:1604.03993v1.

[23] DE LA PEÑA, V. H. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate $U$-statistics. *Ann. Probab.* **23** 806–816. MR1334173

[24] DHARA, M. and SHUKLA, K. K. (2012). Advanced cost based graph clustering algorithm for random geometric graphs. *Int. J. Comput. Appl.* **60** 20–34.

[25] DÍAZ, J., PETIT, J. and SERNA, M. (2002). A survey of graph layout problems. *ACM Comput. Surv.* **34** 313–356.

[26] DÍ AZ, J., PENROSE, M. D., PETIT, J. and SERNA, M. (2001). Approximating layout problems on random geometric graphs. *J. Algorithms* **39** 78–116. MR1823657

[27] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. Revised reprint of the 1989 original. MR1932358

[28] DURRETT, R. (2010). *Probability*: *Theory and Examples*, 4th ed. *Cambridge Series in Statistical and Probabilistic Mathematics* **31**. Cambridge Univ. Press, Cambridge. MR2722836

[29] EL GAMAL, A., MAMMEN, J., PRABHAKAR, B. and SHAH, D. (2004). Throughput-delay trade-off in wireless networks. In *Twenty-Third Annual Joint Conference Proceedings of the IEEE Computer and Communications Societies*.

[30] FOLLAND, G. B. (2013). *Real Analysis*: *Modern Techniques and Their Applications*, 2nd ed. Wiley, New York. MR1681462

[31] FORTUNA, M., STOUFFER, D., OLESEN, J., JORDANO, P., MOUILLOT, D., KRASNOV, B., POULIN, R. and BASCOMPTE, J. (2010). Nestedness versus modularity in ecological networks: Two sides of the same coin? *J. Anim. Ecol.* **79** 811–817.

[32] FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174. MR2580414

[33] FORTUNATO, S. and BARTHÉLEMY, M. (2006). Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA* **104** 36–41.

[34] FRANCESCHETTI, M. and MEESTER, R. (2007). *Random Networks for Communication*: *From Statistical Physics to Information Systems*. *Cambridge Series in Statistical and Probabilistic Mathematics* **24**. Cambridge Univ. Press, Cambridge. MR2398551

[35] GARCÍA TRILLOS, N. and SLEPČEV, D. (2016). A variational approach to the consistency of spectral clustering. *Appl. Comput. Harmon. Anal.*

[36] GARCÍA TRILLOS, N., SLEPČEV, D. and VON BRECHT, J. (2016). Estimating perimeter using graph cuts. Available at arXiv:1602.04102.

[37] GARCÍA TRILLOS, N. and SLEPČEV, D. (2015). On the rate of convergence of empirical measures in $\infty$-transportation distance. *Canad. J. Math.* **67** 1358–1383. MR3415656

[38] GARCÍA TRILLOS, N. and SLEPČEV, D. (2016). Continuum limit of total variation on point clouds. *Arch. Ration. Mech. Anal.* **220** 193–241. MR3458162

[39] GARCÍA TRILLOS, N., SLEPČEV, D., VON BRECHT, J., LAURENT, T. and BRESSON, X. (2016). Consistency of Cheeger and ratio graph cuts. *J. Mach. Learn. Res.* **17** Paper No. 181. MR3567449

[40] GINÉ, E. and KOLTCHINSKII, V. (2006). Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. In *High Dimensional Probability*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 238–259. IMS, Beachwood, OH. MR2387773

[41] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for $U$-statistics. In *High Dimensional Probability*, *II* (*Seattle*, *WA*, 1999). *Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

[42] GOOD, B. H., DE MONTJOYE, Y.-A. and CLAUSET, A. (2010). Performance of modularity maximization in practical contexts. *Phys. Rev. E* (3) **81** 046106. MR2736215

[43] GUIMERA, R. and AMARAL, L. (2005). Functional cartography of complex metabolic networks. *Nature* **433** 895–900.

[44] GUIMERA, R., SALES-PARDO, M. and AMARAL, L. (2004). Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70** 025101.

[45] GUPTA, P. and KUMAR, P. R. (2000). The capacity of wireless networks. *IEEE Trans. Inform. Theory* **46** 388–404. MR1748976

[46] HAGMANN, P., CAMMOUN, L., GIGANDET, X., MEULI, R., HONEY, C. J., WEDEEN, V. J. and SPORNS, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6** e159.

[47] HARTIGAN, J. A. (1981). Consistency of single linkage for high-density clusters. *J. Amer. Statist. Assoc.* **76** 388–394. MR0624340

[48] HEIN, M., AUDIBERT, J.-Y. and VON LUXBURG, U. (2005). From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians. In *Learning Theory*. *Lecture Notes in Computer Science* **3559** 470–485. Springer, Berlin. MR2203281

[49] HU, H., LAURENT, T., PORTER, M. A. and BERTOZZI, A. L. (2013). A method based on total variation for network modularity optimization using the MBO scheme. *SIAM J*. *Appl*. *Math*. **73** 2224–2246. MR3143855

[50] LANCICHINETTI, A. and FORTUNATO, S. (2011). Limits of modularity maximization in community detection. *Phys*. *Rev*. *E* **84** 066122.

[51] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2016). Optimization via low-rank approximation for community detection in networks. *Ann*. *Statist*. **44** 373–400. MR3449772

[52] MEESTER, R. and ROY, R. (1996). *Continuum Percolation*. *Cambridge Tracts in Mathematics* **119**. Cambridge Univ. Press, Cambridge. MR1409145

[53] MILL, J., TANG, T., KAMINSKY, Z., KHARE, T., YAZDANPANAH, S., BOUCHARD, L., JIA, P., ASSADZADEH, A., FLANAGAN, J., SCHUMACHER, A., WANG, S.-C. and PETRONIS, A. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am*. *J*. *Hum*. *Genet*. **82** 696–711.

[54] MORGAN, F. (2009). *Geometric Measure Theory*: *A Beginner's Guide*, 4th ed. Elsevier/Academic Press, Amsterdam. MR2455580

[55] NEWMAN, M. (2006). Modularity and community structure in networks. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **103** 8577–8582.

[56] NEWMAN, M. (2013). Spectral methods for community detection and graph partitioning. *Phys*. *Rev*. *E* **88** 042822.

[57] NEWMAN, M. and GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Phys*. *Rev*. *E* (3) **69** 026113.

[58] NEWMAN, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Phys*. *Rev*. *E* (3) **74** 036104. MR2282139

[59] OUDET, É. (2011). Approximation of partitions of least perimeter by $\Gamma$-convergence: Around Kelvin's conjecture. *Exp*. *Math*. **20** 260–270. MR2836251

[60] PENROSE, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[61] POLLARD, D. (1981). Strong consistency of $k$-means clustering. *Ann*. *Statist*. **9** 135–140. MR0600539

[62] PORTER, M., MUCHA, P., NEWMAN, M. and WARMBRAND, C. (2005). A network analysis of committees in the US House of Representatives. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **102** 7057–7062.

[63] PORTER, M. A., ONNELA, J.-P. and MUCHA, P. J. (2009). Communities in networks. *Notices Amer*. *Math*. *Soc*. **56** 1082–1097. MR2568495

[64] PRZULJ, N., CORNEIL, D. G. and JURISICA, I. (2004). Modeling interactome: Scale-free or geometric? *Bioinformatics* **20** 3508–3515.

[65] REICHARDT, J. and BORNHOLDT, S. (2006). Statistical mechanics of community detection. *Phys*. *Rev*. *E* (3) **74** 016110. MR2276596

[66] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann*. *Statist*. **39** 1878–1915. MR2893856

[67] SABIN, M. (1987). Convergence and consistency of fuzzy c-means/ISODATA algorithms. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **9** 661–668.

[68] SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics*. *Classics in Applied Mathematics* **59**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1986 original. MR3396731

[69] SINGER, A. (2006). From graph to manifold Laplacian: The convergence rate. *Appl*. *Comput*. *Harmon*. *Anal*. **21** 128–134. MR2238670

[70] SINGER, A. and WU, H.-T. (2017). Spectral convergence of the connection Laplacian from random samples. *Inf. Inference* **6** 58–123. MR3636868

[71] THORPE, M., THEIL, F., JOHANSEN, A. M. and CADE, N. (2015). Convergence of the $k$-means minimization problem using Γ-convergence. *SIAM J. Appl. Math.* **75** 2444–2474. MR3422446

[72] TING, D., HUANG, L. and JORDAN, M. I. (2010). An analysis of the convergence of graph Laplacians. In *Proceedings of the 27th International Conference on Machine Learning*.

[73] VAN GENNIP, Y. and BERTOZZI, A. L. (2012). Γ-convergence of graph Ginzburg–Landau functionals. *Adv. Differential Equations* **17** 1115–1180. MR3013414

[74] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[75] VILLANI, C. (2004). *Topics in Optimal Transportation*. American Mathematical Society, Providence, RI.

[76] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

[77] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. MR2396807

[78] WETS, R. J.-B. (1999). Statistical estimation from an optimization viewpoint. *Ann. Oper. Res.* **85** 79–101. MR1685588

[79] ZHANG, X. and NEWMAN, M. (2015). Multiway spectral community detection in networks. *Phys. Rev. E* **92** 052808.

[80] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083



DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721
USA
E-MAIL: edavis@math.arizona.edu
sethuram@math.arizona.edu