Testing properties of graphs and functions

László Lovász*
Institute of Mathematics, Eötvös Loránd University, Budapest
and
Balázs Szegedy
Department of Mathematics, University of Toronto

January 2008

Contents

1 Introduction **2**

2 Preliminaries **4**
2.1 Homomorphisms 4
2.2 Functions and graphons 4
2.3 Distances of graphs and functions 5
2.4 W-random graphs 7
2.5 Relating different norms 10

3 Main results **12**
3.1 Property testing for functions 12
3.2 Graph properties vs. function properties 17
3.2.1 Closure of graph properties 17
3.2.2 Closure and distance 20
3.2.3 Monotone closure 22
3.2.4 Robustness 24
3.2.5 Characterizing testability of graph properties 24
3.3 Property testing vs. parameter testing 27
3.4 Flexible properties 30

4 Concluding remarks **33**

*Research supported by OTKA grant No. 67867

Abstract

We define an analytic version of the graph property testing problem, which can be formulated as studying an unknown 2-variable symmetric function through sampling from its domain and studying the random graph obtained when using the function values as edge probabilities. We give a characterization of properties testable this way, and extend a number of results about "large graphs" to this setting.

These results can be applied to the original graph-theoretic property testing. In particular, we give a new combinatorial characterization of testable graph properties. Furthermore, we define a class of graph properties (flexible properties) which contains all the hereditary properties, and generalize various results of Alon, Shapira, Fischer, Newman and Stav from hereditary to flexible properties.

1 Introduction

Graph property testing is a very active area in computer science. In its most restricted form (and this will be our concern in this paper), it studies properties of (very large) graphs that can be tested by studying a randomly chosen induced subgraph of bounded size.

To be more precise, we have to describe what kind of error is allowed. In this paper, by *graph* we always mean a finite simple graph. A *graph property* is a class of graphs invariant under isomorphism. The *edit distance* of two graphs G_1, G_2 on the same node set is $|E(G_1)\triangle E(G_2)|$. The *edit distance* of a graph G from a graph property is the minimum number of edges we have to change (add or delete) to obtain a graph with the property. If no graph with the same number of nodes has the property, then this distance is infinite.

Definition 1.1 A graph property $\mathcal{P}$ is **testable**, if there exists another property $\mathcal{P}'$ (called a *test property*) satisfying the following conditions:

(a) if a graph G has property $\mathcal{P}$, then for all $1 \le k \le |V(G)|$ at least a fraction of $2/3$ of its k-node induced subgraphs have property $\mathcal{P}'$, and

(b) for every $\varepsilon > 0$ there is a $k_\varepsilon \ge 1$ such that if G is a graph whose edit distance from $\mathcal{P}$ is at least $\varepsilon|V(G)|^2$, then for all $k_\varepsilon \le k \le |V(G)|$ at most a fraction of $1/3$ of the k-node induced subgraphs of G have property $\mathcal{P}'$.

The notion of testability has other variations: we may also know the number of nodes of G, or we can take a sample whose size is growing slowly with the size of G, etc. The definition above is in a sense the most restrictive, and it has often been referred to with adjectives like "oblivious testing" and "order independent testing". Since this is the only version we consider in this paper, we simplify terminology by calling it simply "testable".

The definition here for graph property testing is a non-standard one, and in fact corresponds to what was referred to as "oblivious testing" in [2] and "order independent testing" in some other works. Please explain these issues after Definition 1.1.

We could strengthen this definition by requiring a fraction of $1-\varepsilon$ instead of $2/3$ and a fraction of ε instead of $1/3$. We could also weaken it by allowing the test property $\mathcal{P}'$ to depend on ε. It can be seen that neither of these modifications would change the notion of testability.

A surprisingly general sufficient condition for testability was proved by Alon and Shapira [2]: *Every hereditary graph property is testable.* (A graph property is *hereditary*, if whenever a graph has the property, then all its induced subgraphs also have the property.) Alon, Fischer, Newman and Shapira [1] gave a characterization of testable graph properties in terms of Szemerédi partitions (which is quite involved and we don't quote it here). In fact, Szemerédi partitions play a central role in most results of this theory.

One of the main graph theoretic results in this paper is to give another combinatorial characterization of testable properties (Theorem 3.20). It says that a graph property is testable if and only if for every graph with the property, a sufficiently large "typical" induced subgraph is "close" to having the property.

Our main goal is, however, to treat property testing in terms of the theory of convergent (dense) graph sequences and graph limits [10, 11, 7]. A sequence of graphs (G_n) is *convergent* if the density of copies of any fixed graph F in G_n tends to a limit. It turns out that the limit of a convergent graph sequence can be represented by a symmetric measurable function $W:\ [0,1]^2 \to [0,1]$, and that many problems and constructions in graph theory have a simpler and cleaner formulation when extended to this limit (see [6] for a survey).

Parameter testing (or estimation) is closely related to property testing, but is in many respects simpler. This area has a very natural treatment in the framework of graph limits [5, 7]. The two theories are connected by a result of Fischer and Newman [8], who proved that the edit distance from a testable property is a testable parameter. The analytic theory of property testing is more involved than the analytic theory of parameter testing, mainly because of the different type of error that is permitted.

Above, we used the "edit distance" of two graphs in the definition of testable properties. However, there is a different distance, called the "cut distance", which plays a central role in graph convergence; for example, a sequence of graphs is convergent if and only if it is Cauchy in an appropriately normalized cut distance. The main technical issue in the analytic theory of property testing is the interplay between these two distances; see Section 2.5 for some auxiliary results of this nature that might be interesting on their own right.

The space of limit objects (two-variable functions) with the "cut distance" is compact, a fact which is essentially equivalent to various (weak and strong) versions of Szemerédi's Regularity Lemma [11]. So while we do not explicitly use the Regularity Lemma, it is implicit in the utilization of the compactness of this space.

A further surprisingly general result using the edit distance is the theorem of Alon and Stav [3], proving that for every hereditary property, a random graph with appropriate density is the farthest from the property in edit distance. The analytic results developed in this paper allow us to state and prove a simple analytic analogue of this fact, from which the original result follows along with generalizations. Similar analytic analogues are derived for the other above mentioned results.

2 Preliminaries

2.1 Homomorphisms

For two graphs F and G, a *homomorphism* from F to G is an adjacency preserving map $V(F) \to V(G)$. The number of such homomorphisms is denoted by $\hom(F,G)$. We'll almost always use the normalized version of this number,

$$t(F,G) = \frac{\hom(F,G)}{|V(G)|^{|V(F)|}},$$

which can be interpreted as the probability that a random map $V(F) \to V(G)$ is a homomorphism. We denote by $\mathrm{ind}(F,G)$ the number of those injective homomorphisms that also preserve non-adjacency (in other words, the number of induced copies of F in G). The normalized version of this number is

$$t_{\mathrm{ind}}(F,G) = \frac{\mathrm{ind}(F,G)}{(|V(G)|)_{|V(F)|}}$$

(here $(n)_k = n(n-1)\cdots(n-k+1)$).

2.2 Functions and graphons

Let $\mathcal{W}$ denote the space of all symmetric measurable functions $W : [0,1]^2 \to \mathbb{R}$ (i.e., $W(x,y) = W(y,x)$ for all $x,y \in [0,1]$). Let $\mathcal{W}_0$ denote the set of all functions $W \in \mathcal{W}$ such that $0 \le W \le 1$.

A function $W \in \mathcal{W}$ is called a *stepfunction*, if there is a partition $S_1 \cup \cdots \cup S_k$ of $[0,1]$ into measurable sets such that W is constant on every product set $S_i \times S_j$. The number k is the *number of steps* of W.

Let $\mathcal{F}_k$ denote the set of all graphs on node set $[k] = \{1,\dots,k\}$. For $W \in \mathcal{W}_0$ and $F \in \mathcal{F}_k$, define

$$t(F,W) = \int_{[0,1]^k} \prod_{ij \in E(F)} W(x_i,x_j)\,dx$$

and

$$t_{\mathrm{ind}}(F,W) = \int_{[0,1]^k} \prod_{ij \in E(F)} W(x_i,x_j) \prod_{\substack{ij \notin E(F) \\ i \ne j}} \bigl(1 - W(x_i,x_j)\bigr)\,dx\,.$$

Two functions $W_1, W_2 \in \mathcal{W}_0$ are *isomorphic*, in notation $W_1 \cong W_2$, if $t(G,W_1) = t(G,W_2)$ for every graph G. It was proved by Borgs, Chayes and Lovász [4] that two functions are isomorphic if and only if there is a third function $U \in \mathcal{W}_0$ and two measure preserving maps $\phi_1, \phi_2 : [0,1] \to [0,1]$ such that $W_i(x,y) = U(\phi_i(x), \phi_i(y))$ for $i = 1,2$ and almost all $x,y \in [0,1]$. An isomorphism class of functions in $\mathcal{W}_0$ is called a *graphon* [7].

A sequence of graphs (G_n) with $|V(G_n)| \to \infty$ is called *convergent*, if $t(F,G_n)$ tends to a limit for every fixed graph F. (This is equivalent with $t_{\mathrm{ind}}(F,G_n)$ tending to a limit for every F.) It was proved in [10] that for every convergent sequence of graphs (G_n) there is a function $W \in \mathcal{W}_0$ such that $t(F,G_n) \to t(F,W)$ for every graph F. We call W the *limit* of the sequence, and write $G_n \to W$. Every function in $\mathcal{W}_0$ arises as the limit of a convergent graph sequence. Furthermore, the limit is unique up to isomorphism.

For every graph G, we define a function $W_G \in \mathcal{W}_0$ as follows. Let $V(G) = \{1, \dots, n\}$ and consider a point $(x, y) \in [0, 1]^2$. Define integers i and j such that $x \in ((i-1)/n, i/n]$ and $y \in ((j-1)/n, j/n]$ (if $x = 0$ we define $i = 0$, and similarly for j). Then we set

$$W(x,y) = \begin{cases} 1, & \text{if } ij \in E(G), \\ 0, & \text{otherwise.} \end{cases}$$

(informally, we consider the adjacency matrix $A = (a_{ij})$ of G, and replace each entry a_{ij} by a square of size $(1/n) \times (1/n)$ with the constant function a_{ij} on this square). Note that W_G depends on the labeling of the nodes of G (but only up to a measure preserving transformation).

2.3 Distances of graphs and functions

As mentioned in the introduction, our results concern the interaction of two distances between graphs, the edit distance and the cut distance. Let G_1 and G_2 be two graphs with a common node set V. Instead of the edit distance mentioned in the introduction, we shall use its normalized version

$$d_1(G_1, G_2) = \frac{|E(G_1) \triangle E(G_2)|}{|V|^2},$$

and the normalized cut distance

$$d_\square(G_1, G_2) = \max_{S,T \subset V} \frac{\left| e_{G_1}(S,T) - e_{G_2}(S,T) \right|}{|V|^2}$$

(here $e_{G_1}(S,T)$ denotes the number of edges of G with one endpoint in S and the other endpoint in T).

We consider on $\mathcal{W}$ the *cut norm*

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y)\, dx\, dy \right|$$

where the supremum is taken over all measurable subsets S and T. (See [7] for several useful properties of this norm.) We will also use the standard L_1 norm

$$\|W\|_1 = \int_{[0,1]^2} |W(x,y)|\, dx\, dy.$$

This defines two metrics on $\mathcal{W}_0$ by

$$d_1(U, W) = \|U - W\|_1, \qquad \text{and} \qquad d_\square(U, W) = \|U - W\|_\square .$$

For every set $S \subseteq \mathcal{W}_0$ and every $c > 0$, we define, as usual, the balls

$$B_1(S, c) = \{W \in \mathcal{W}_0 : \; d_1(W, S) < c\} \qquad \text{and} \qquad B_\square(S, c) = \{W \in \mathcal{W}_0 : \; d_\square(W, S) < c\}.$$

Clearly $d_\square \le d_1$, and hence $B_1(S, c) \subseteq B_\square(S, c)$. So $d_\square$ is continuous with respect to d_1. In general, d_1 is not continuous w.r.t. $d_\square$, but see Theorem 3.4 for a weaker statement.

For $W \in \mathcal{W}_0$ and $\phi : \; [0,1] \to [0,1]$, set $W^\phi(x,y) = W(\phi(x), \phi(y))$. We define

$$\delta_1(U, W) = \inf_\phi d_1(U, W^\phi),$$

where ϕ ranges over all invertible measure preserving maps from $[0,1]$ to $[0,1]$. This is a quasi-metric on $\mathcal{W}_0$, in which U and W are distance 0 if and only if they are isomorphic. The metric $\delta_\square(U,W)$ is defined analogously.

The main advantage of $\delta_\square$ over $d_\square$ is that the space $(W,\delta_\square)$ is compact, as was proved in [11].

We can use this distance to define yet another distance between graphs:

$$\delta_\square(G_1,G_2)=\delta_\square(W_{G_1},W_{G_2}).$$

Note that in this definition the graphs G_1 and G_2 do not need to have the same number of nodes, and their distance is independent of the labeling of their nodes. If it happens that $V(G_1)=V(G_2)$, then clearly

$$\delta_\square(G_1,G_2)\le d_\square(G_1,G_2)\le d_1(G_1,G_2).$$

The paper [7] contains a more explicit description of this distance, and its relation to combinatorially defined distances. One of the main conclusions is that a sequence of graphs is convergent if and only if it is Cauchy in this metric. So $(\mathcal{W}_0,\delta_\square)$ is the completion of the set of graphs with distance $\delta_\square$.

We summarize some further facts about homomorphism densities and distances, mostly from [10]. Let F be a graph with k nodes and m edges. For every graph G, we have

$$t(F,G)=t(F,W_G),$$

but for the "induced" versions we only have the following approximate equality:

$$\big|t_{\rm ind}(F,G)-t_{\rm ind}(F,W_G)\big|\le\frac{\binom{k}{2}}{|V(G)|-\binom{k}{2}}. \tag{1}$$

Lemma 4.1 in [10] asserts that for any two functions $U,W\in\mathcal{W}_0$,

$$|t(F,U)-t(F,W)|\le m\|U-W\|_\square. \tag{2}$$

This implies (via the functions W_G and W_H) a similar inequality for any two graphs G and H:

$$|t(F,G)-t(F,H)|\le m\delta_\square(G,H). \tag{3}$$

An analogue of inequality (2) for induced densities in functions can be proved by essentially the same argument:

$$|t_{\rm ind}(F,U)-t_{\rm ind}(F,W)|\le\binom{k}{2}\cdot\|U-W\|_\square. \tag{4}$$

Using the easy inequality (1), this implies for the induced densities in graphs that

$$|t_{\rm ind}(F,H)-t_{\rm ind}(F,G)|\le\binom{k}{2}\delta_\square(G,H)+\frac{\binom{k}{2}}{|V(G)|-\binom{k}{2}}+\frac{\binom{k}{2}}{|V(H)|-\binom{k}{2}} \tag{5}$$

(assuming that $|V(G)|,|V(H)|>\binom{k}{2}$).

The following result from [7] (Theorem 4.10) provides a converse to (2):

Theorem 2.1 *Let $U, W \in \mathcal{W}_0$ and let $k > 1$ be a positive integer. Assume that for every simple graph F on k nodes, we have*

$$|t(F,U) - t(F,W)| \le 3^{-k^2}.$$

Then

$$\delta_\square(U,W) \le \frac{22}{\sqrt{\log k}}.$$

We conclude with a lemma showing that convergence in the cut norm has good analytic properties.

Lemma 2.2 *Suppose that $\|W_n - W\|_\square \to 0$ as $n \to \infty$ ($W, W_n \in \mathcal{W}_0$). Then for every $Z \in \mathcal{W}_0$*

$$\|Z(W_n - W)\|_\square \to 0 \qquad (n \to \infty).$$

In particular,

$$\int_{[0,1]^2} Z(x,y)W_n(x,y)\,dx\,dy \to \int_{[0,1]^2} Z(x,y)W_n(x,y)\,dx\,dy$$

and

$$\int_S W_n \to \int_S W$$

for every measurable set $S \subseteq [0,1]^2$.

Proof. If Z is the indicator function of a rectangle, these conclusions follow from the definition of the $\|.\|_\square$ norm. Hence the conclusion follows for stepfunctions, since they are linear combinations of a finite number of indicator functions of rectangles. Then it follows for all integrable functions, since they are approximable in $L_1([0,1]^2)$ by stepfunctions. $\square$

2.4 W-random graphs

For every function $W \in \mathcal{W}_0$ and every finite set $X \subseteq [0,1]$, we define a graph $\mathbb{G}(X,U)$ on $V(G) = X$ by connecting $x, y \in X$, $x \ne y$ with probability $U(x,y)$ (making independent decisions for different pairs $\{x,y\}$). If G is a simple graph and $X \subseteq V(G)$, we denote by $G[X]$ the subgraph induced by X.

For every function $W \in \mathcal{W}_0$, we define the *W-random graph* $\mathbb{G}(n,W) = \mathbb{G}(X,W)$, where $X \subseteq [0,1]$ consists of n independent, uniformly distributed random points in $[0,1]$. Note that for every $F \in \mathcal{F}_n$,

$$\mathsf{P}(\mathbb{G}(n,W) = F) = t_{\text{ind}}(F,W).$$

Clearly the distribution of $\mathbb{G}(n,W)$ is invariant under isomorphism of functions, i.e., it only depends on W as a graphon.

We need an analogous notation $\mathbb{G}(k,G) = G[X]$, where G is a finite graph and X is a random subset of $V(G)$ chosen uniformly from all k-element subsets.

It was proved in [10] and with probability 1, $\mathbb{G}(n,W) \to W$, and in fact, the convergence is quite fast, as shown by the following concentration inequality:

Theorem 2.3 *Let $W \in \mathcal{W}_0$ and let F be a graph with k nodes. Then for every $0 < \varepsilon < 1$,*

$$\mathsf{P}\Big(|t(F,\mathbb{G}(n,W)) - t(F,W)| > \varepsilon\Big) \le 2\exp\left(-\frac{\varepsilon^2}{18k^2}n\right). \tag{6}$$

The following bound on the distance of a W-random graph from W was proved in [7] (Theorem 4.9(ii)):

Theorem 2.4 *Let $U \in \mathcal{W}_0$ and let $k > 1$ be a positive integer. Then with probability at least $1 - e^{-k^2/(2\log k)}$, we have*

$$\delta_\square(U, \mathbb{G}(k,U)) \le \frac{10}{\sqrt{\log k}}.$$

We'll use a related construction of a random graph associated with W. Let X_i be a uniform random element of the interval $L_i = [(i-1)/n, i/n]$, and $X = \{X_1, \dots, X_n\}$. Let $\mathbb{G}'(n,W) = \mathbb{G}(X,W)$. (Note that the distribution of $\mathbb{G}(n,W)$ only depends on the isomorphism type of W, but $\mathbb{G}'(n,W)$ does depend on how the points of $[0,1]$ are ordered.) For the cut distance of $\mathbb{G}'(n,W)$ from W we have the following bound:

Lemma 2.5 *Let $W \in \mathcal{W}$ and $G' = \mathbb{G}'(n,W)$. Then with probability more than $1 - 1/\sqrt{n}$,*

$$\delta_\square(W_{G'}, U) < \frac{50}{\sqrt{\log\log n}}.$$

Proof. We use subgraph densities and Theorem 2.1. To estimate $|t(F, W_{\mathbb{G}(n,U)}) - t(F, W_{G'})|$, note that the random variable $t(F, W_{\mathbb{G}(n,U)})$ can be generated as follows: (1) we select k random integers $r_1, \dots, r_k \in [n]$ uniformly and independently (with repetition); (2) we select $Z_i \in [0, 1/n]$ uniformly and independently for $i = 1, \dots, k$, and compute $X_i = Z_i + (r_i - 1)/n$; (3) we create random variables Y_{ij} $(i,j \in [k])$ such that $Y_{ij} = Y_{ji}$ but otherwise they are independent, and

$$Y_{ij} = \begin{cases} 1 & \text{with probability } W(X_i, X_j), \\ 0 & \text{with probability } 1 - W(X_i, X_j). \end{cases}$$

Then $t(F, W_{\mathbb{G}(n,U)})$ is the expectation of $\prod_{ij \in E(F)} Y_{ij}$ over the choice (1). The computation of $t(F, W_{G'})$ is similar, except that if $r_i = r_j$ then we choose $Z_i = Z_j$. It follows that if the integers $r_1, \dots, r_k$ are distinct, then $t(F, W_{\mathbb{G}(n,U)})$ and $t(F, W_{G'})$ are generated by the same procedure, and hence they can be coupled so that they are equal. The probability that there is no repetition among $r_1, \dots, r_k$ is

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} > 1 - \frac{\binom{k}{2}}{n},$$

So it follows that $t(F, W_\mathbb{G}(n,U)) = t(F, W_{G'})$ with probability at least $1 - \frac{\binom{k}{2}}{n}$, and hence

$$\mathsf{E}\big(|t(F, W_{\mathbb{G}(n,U)}) - t(F, W_{G'})|\big) < \frac{k^2}{n}.$$

Summing this over all graphs F on k nodes, we get

$$\mathsf{E}\Big(\sum_F |t(F, W_{\mathbb{G}(n,U)}) - t(F, W_{G'})|\Big) < \frac{2^{k^2}}{n},$$

and so the probability that $\sum_F |t(F, W_{\mathbb{G}(n,U)}) - t(F, W_{G'})| > 3^{-k^2}$ is less than $6^{k^2}/n$. Thus with probability more than $1 - 6^{k^2}/n$, we have

$$|t(F, W_{\mathbb{G}(n,U)}) - t(F, W_{G'})\Big)| < 3^{-k^2}$$

for all graphs F with at most k nodes. Theorem 2.1 implies that in this case

$$\delta_\square(\mathbb{G}(n,U), G') \le \frac{22}{\sqrt{\log k}}.$$

So if we choose $k = \lfloor \sqrt{\log(n/2)}/4 \rfloor$, then

$$\delta_\square(\mathbb{G}(n,U), G') \le \frac{44}{\sqrt{\log\log n}}.$$

Combining with (8), we get that with probability more than $1 - 1/\sqrt{n}$,

$$\delta_\square(W_{G'}, U) \le \delta_\square(G', \mathbb{G}(n,U)) + \delta_\square(W_{\mathbb{G}(n,U)}, U) \le \frac{10}{\sqrt{\log n}} + \frac{44}{\sqrt{\log\log n}} < \frac{50}{\sqrt{\log\log n}}.$$

$\square$

One advantage of using $\mathbb{G}'(n,W)$ over $\mathbb{G}(n,W)$ is that it is easier to handle its d_1 distance from other graphs and functions.

Lemma 2.6 *Let $W \in \mathcal{W}$ and $G' = \mathbb{G}'(n,W)$. Let H be a graph on $[n]$. Then*

$$\mathsf{E}(d_1(G', H)) = \|W_H - W\|_1$$

and

$$\mathsf{E}(\|W_{G'} - W\|_1) \le 2\|W_H - W\|_1.$$

Proof. For the first formula, note that for $1 \le i, j \le n$, the probability that a pair $i \ne j$ contributes to $d_1(H, G')$ is

$$\begin{cases} W(X_i, X_j), & \text{if } ij \notin E(H), \\ 1 - W(X_i, X_j), & \text{if } ij \in E(G_n). \end{cases}$$

Summing over all $i \ne j$, and taking expectation, the equality follows. The second inequality is an easy consequence:

$$\mathsf{E}(\|W_{G'} - W\|_1) \le \mathsf{E}(\|W_{G'} - W_H\|_1) + \|W_H - W\|_1 = \mathsf{E}(d_1(G', H)) + \|W_H - W\|_1 \le 2\|W_H - W\|_1.$$

$\square$

As a useful consequence, we obtain the following fact. The Regularity Lemma implies (see e.g. [11]) that functions in $\mathcal{W}_0$ can be approximated by functions of the form W_G, so that the number of nodes of G can be bounded uniformly if the error is measured in the cut distance. Obviously, one cannot approximate all functions by functions W_G in the L_1-norm. But for every n there is graph on n nodes that approximates W so that the cut distance tends to 0 uniformly, and at the same time the approximation in the L_1 norm is almost as good as possible.

Corollary 2.7 *Let G be a simple graph and $U \in \mathcal{W}_0$. Then there exists a simple graph G' such that*

$$\|W_{G'} - U\|_1 \le 4\|W_G - U\|_1 \qquad \text{and} \qquad \delta_\square(W_{G'}, U) \le \frac{50}{\sqrt{\log\log n}}.$$

Proof. Let $V(G) = [n]$. We are going to show that the random graph $G' = \mathbb{G}'(n, W)$ satisfies the conditions with positive probability. By Lemma 2.6, $\mathsf{E}\big(d_1(W'_G, W)\big) \le 2\|W_G - W\|_1$, and so with probability at least $1/2$,

$$d_1(W_{G'}, W) \le 4\|W_G - W\|_1. \tag{7}$$

On the other hand, Lemma 2.5 implies that with probability at least $3/4$, we have

$$\delta_\square(G', U) \le \frac{50}{\sqrt{\log\log n}}. \tag{8}$$

So with positive probability, both (7) and (8) hold. □

2.5 Relating different norms

As pointed out in the introduction, the analytic problem behind property testing is to relate the cut norm and the L_1 norm. In this section, which contains our main technical tools, we study this connection. Some of the lemmas below are purely analytic in nature, and may be of interest on their own right.

Lemma 2.8 *Suppose that $\|U_n - U\|_\square \to 0$ and $\|W_n - W\|_\square \to 0$ as $n \to \infty$ ($U, W, U_n, W_n \in \mathcal{W}_0$). Then*

$$\liminf_{n\to\infty} \|W_n - U_n\|_1 \ge \|W - U\|_1.$$

Proof. Let $\sigma(x, y) = \mathrm{sgn}(U(x, y) - W(x, u))$. Then

$$\begin{aligned} \|U_n - W_n\|_1 &\ge \int_{[0,1]^2} \sigma(x,y)(U_n(x,y) - W_n(x,y))\,dx\,dy \\ &\longrightarrow \int_{[0,1]^2} \sigma(x,y)(U(x,y) - W(x,y))\,dx\,dy = \|U - W\|_1 \end{aligned}$$

by Lemma 2.2. □

There are easy examples showing that $\|W_n - U_n\|_1 \to \|W - U\|_1$ does not hold in general: for example, let $W_n = W_{\mathbb{G}(n,1/2)}$ and $W = U = U_n = 1/2$. But we can formulate two statements that provide partial converses to this lemma.

Lemma 2.9 *Suppose that $\|U_n - U\|_\square \to 0$ and $\|W_n - W\|_\square \to 0$ as $n \to \infty$ ($U, W, U_n, W_n \in \mathcal{W}_0$). Suppose further that U is $0-1$ valued. Then*

$$\|U_n - W_n\|_1 \to \|U - W\|_1.$$

Proof. By Lemma 2.2,

$$\begin{aligned}\|U_n - W_n\|_1 &\le \|U_n - U\|_1 + \|U - W_n\|_1 \\ &= \int_{U=0} U_n + \int_{U=1} (1-U_n) + \int_{U=0} W_n + \int_{U=1} (1-W_n) \\ &\to \int_{U=0} W + \int_{U=1} (1-W) = \|U - W\|_1 .\end{aligned}$$

Combined with Lemma 2.8, the assertion follows. □

Lemma 2.10 *Suppose that $\|U_n - U\|_\square \to 0$ as $n \to \infty$ ($U, U_n \in \mathcal{W}_0$). Then for every $W \in \mathcal{W}_0$ there is a sequence of functions $W_n \in \mathcal{W}_0$ such that $\|W_n - W\|_\square \to 0$ and*

$$\|U_n - W_n\|_1 \to \|U - W\|_1.$$

Proof. First we consider the case when $U \ge W$. Let

$$Z(x,y) = \begin{cases} W(x,y)/U(x,y) & \text{if } U(x,y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

and $W_n = ZU_n$. Trivially $W_n \in \mathcal{W}_0$, $W = ZU$, and

$$\|W - W_n\|_\square = \|Z(U - U_n)\|_\square \to 0$$

by Lemma 2.2. Furthermore,

$$\|U_n - W_n\|_1 = \|(U - W) + Z(U - U_n)\|_1 \le \|U - W\|_1 + \|Z(U - U_n)\|_1 \to \|U - W\|_1$$

(using Lemma 2.2 again). Combining this with Lemma 2.8 we get that $\|U_n - W_n\|_1 \to \|U - W\|_1$.

The case when $U \le W$ follows by a similar argument, replacing $U, W, \dots$ by $1-U, 1-W, \dots$.

Finally, in the general case, consider the function $V = \max(U, W)$. Then clearly $\|U - V\|_1 + \|V - W\|_1 = \|U - W\|_1$. Since $U \le V$, there exists a sequence (V_n) of functions such that $\|V_n - V\|_\square \to 0$ and $\|V_n - U_n\|_1 \to \|V - U\|_1$. Since $V \ge W$, there is a sequence (W_n) of functions such that $\|W_n - W\|_\square \to 0$ and $\|W_n - V_n\|_1 \to \|W - V\|_1$. Hence

$$\begin{aligned}\limsup_{n\to\infty} \|U_n - W_n\|_1 &\le \limsup_{n\to\infty} \|U_n - V_n\|_1 + \limsup_{n\to\infty} \|V_n - W_n\|_1 \\ &= \|U - V\|_1 + \|V - W\|_1 = \|U - W\|_1.\end{aligned}$$

Using Lemma 2.8 again, the lemma follows. □

A fact similar to Lemma 2.8 holds for the distances δ_1 and $\delta_\square$ replacing the norms $\|.\|_1$ and $\|.\|_\square$. This does not seem to follow directly from Lemma 2.8, and the proof is more involved.

Lemma 2.11 *Suppose that $\delta_\square(U_n, U) \to 0$ and $\delta_\square(W_n, W) \to 0$ as $n \to \infty$ ($U, W, U_n, W_n \in \mathcal{W}_0$). Then*

$$\liminf \delta_1(W_n, U_n) \ge \delta_1(W, U). \tag{9}$$

Proof. Let d denote the lim inf on the left hand side of (9). Let ε be an arbitrary positive number. There is a number k (depending on U, W and ε), a partition $[0,1] = \cup_{i=1}^k S_i$ of the unit interval into k measurable pieces and two functions $W', U' \in \mathcal{W}_0$ such that both W' and U' are constant on every rectangle $S_i \times S_j$ and furthermore that $\|W - W'\|_1,\ \|U - U'\|_1 \leq \varepsilon$. Let n be a natural number such that $\|W_n - W\|_\square\ ,\ \|U_n - U\|_\square \leq \varepsilon/k^4$ and $\delta_1(W_n, U_n) \leq d + \varepsilon$. There are measure preserving transformations $\rho, \pi : [0,1] \mapsto [0,1]$ such that $\|W_n^\rho - U_n^\pi\|_1 \leq d + 2\varepsilon$. Let $S_{i,j}$ denote the set $S_i^\rho \cap S_j^\pi$ for every $1 \leq i, j \leq k$. It is clear that $S_{i,j}$ is a partition of the unit interval and that both W'^ρ and U'^π are constant on each rectangle $S_{i_1,j_1} \times S_{i_2,j_2}$. We have that

$$d + 2\varepsilon \geq \|W_n^\rho - U_n^\pi\|_1 \geq \sum_{1 \leq i_1, j_1, i_2, j_2 \leq k} \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W_n^\rho - U_n^\pi) \Big|.$$

Using that

$$\Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W^\rho - W_n^\rho) \Big| \leq \frac{\varepsilon}{k^4} \ \text{ and } \ \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (U^\pi - U_n^\pi) \Big| \leq \frac{\varepsilon}{k^4},$$

we get that

$$\sum_{1 \leq i_1, j_1, i_2, j_2 \leq k} \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W^\rho - U^\pi) \Big| \leq d + 4\varepsilon.$$

Writing $W^\rho - U^\pi = (W^\rho - W'^\rho) + (W'^\rho - U'^\pi) + (U'^\pi - U^\pi)$ we obtain that

$$\sum_{1 \leq i_1, j_1, i_2, j_2 \leq k} \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W^\rho - U^\pi) \Big| \geq \sum_{1 \leq i_1, j_1, i_2, j_2 \leq k} \Big(\Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W'^\rho - U'^\pi) \Big| - \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W^\rho - W'^\rho) \Big| - \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (U^\pi - U'^\pi) \Big| \Big).$$

This implies that

$$\sum_{1 \leq i_1, j_1, i_2, j_2 \leq k} \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W^\rho - U^\pi) \Big| \geq \sum_{1 \leq i_1, j_1, i_2, j_2 \leq k} \Big| \int_{S_{i_1,j_1} \times S_{i_2,j_2}} (W'^\rho - U'^\pi) \Big| - 2\varepsilon.$$

Using that both W'^ρ and U'^π are constant on the sets $S_{i_1,j_1} \times S_{i_2,j_2}$ we obtain that the right side of the above inequality is equal to $\|W'^\rho - U'^\pi\|_1 - 2\varepsilon$. Consequently

$$\|W'^\rho - U'^\pi\|_1 \leq d + 6\varepsilon.$$

Using that $\|W^\rho - W'^\rho\|_1,\ \|U^\pi - U'^\pi\|_1 \leq \varepsilon$ we get that

$$\|W^\rho - U^\pi\|_1 \leq d + 10\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, (9) follows. □

3 Main results

3.1 Property testing for functions

We define a notion of testability for properties of functions in $\mathcal{W}_0$ and for graphons. Formally, a *function property* is a subset $\mathcal{R} \subseteq \mathcal{W}_0$; a *graphon property* is a function property that is invariant under isomorphism. A function property is *closed* if it is closed in the $\|.\|_\square$ norm.

Example 3.1 The function property of being 0-1 valued is a graphon property by the characterization of isomorphism, but it is not closed, since for a sequence of random graphs G_n with edge probability $1/2$, the functions W_{G_n} are 0-1 valued, but their limit in the $\|.\|_\square$ norm, namely the identically $1/2$ function, is not.

The following definition of testability of a function property is analogous to the testability of a graph property. The framework is that we study a function $W \in \mathcal{W}_0$ by observing a W-random graph $\mathbb{G}(k, W)$.

Definition 3.2 A function property $\mathcal{R}$ is *testable* if there is a graph property $\mathcal{R}'$ (called a *test property* for $\mathcal{R}$) such that

(a) $\mathsf{P}(\mathbb{G}(k, W) \in \mathcal{R}') \geq 2/3$ for every function $W \in \mathcal{R}$ and every $k \geq 1$, and

(b) for every $\varepsilon > 0$ there is a $k_\varepsilon \geq 1$ such that $\mathsf{P}(\mathbb{G}(k, W) \in \mathcal{R}') \leq 1/3$ for every $k \geq k_\varepsilon$ and every function $W \in \mathcal{W}_0$ with $d_1(W, \mathcal{R}) \geq \varepsilon$.

Similarly as for graph properties, the constants $1/3$ and $2/3$ are arbitrary, but it would not change the property if we replaced them by any two real numbers $0 < a < b < 1$:

Lemma 3.3 *Let $0 < a < b < 1$. A graphon property $\mathcal{R}$ is testable if and only if there is a graph property $\mathcal{R}''$ such that for every $\varepsilon > 0$ there is a constant $k(\varepsilon)$ such that for every function $W \in \mathcal{W}_0$ and $k \geq k(\varepsilon)$,*

(a) if $W \in \mathcal{R}$, then $\mathsf{P}(\mathbb{G}(k, W) \in \mathcal{R}'') \geq b$, and

(b) if $d_1(W, \mathcal{R}) > \varepsilon$ then $\mathsf{P}(\mathbb{G}(k, W) \in \mathcal{R}'') \leq a$.

Proof. Suppose that $\mathcal{R}$ is testable, and let $\mathcal{R}'$ be the graph property in the definition of testability. For every simple graph F and $k \leq |V(F)|$, define

$$q_k(F) = \mathsf{P}(\mathbb{G}(k, F) \in \mathcal{R}').$$

Define the graph property

$$\mathcal{R}'' = \{G : \; q_k(F) \geq \frac{1}{2}\}.$$

Let $W \in \mathcal{R}$ and let $n > k$ be large enough. Then

$$\mathsf{E}\big(q_k(\mathbb{G}(n, W))\big) = \mathsf{P}(\mathbb{G}(k, W) \in \mathcal{R}') \geq \frac{2}{3}.$$

We use Azuma's inequality to show that $q_k(\mathbb{G}(n, W))$ is highly concentrated. Indeed, let $X_1, \dots, X_n \in [0, 1]$ be independent uniform samples, and let $Z_t = \mathsf{E}_t\big(q_k(\mathbb{G}(n, W))\big)$, where E_t means conditional expectation with respect to the choice of $X_1, \dots, X_t$ and the randomization of the edges between nodes $\{1, \dots, t\}$. Then $Z_0, \dots, Z_n$ is a martingale, with expectation $\mathsf{E}(Z_0) = \mathsf{E}\big(q_k(\mathbb{G}(n, W))\big) \geq 2/3$. Furthermore, $|Z_{t+1} - Z_t|$ is bounded by the probability that a random k-subset of $\{1, \dots, n\}$ contains $t + 1$, which is k/n. Thus by Azuma's Inequality,

$$\mathsf{P}\Big(q_k(\mathbb{G}(n, W)) \leq \frac{1}{2}\Big) \leq \mathsf{P}\Big(q_k(\mathbb{G}(n, W)) \leq \mathsf{E}(q_k(\mathbb{G}(n, W))) - \frac{1}{6}\Big) \leq \exp\Big(-\frac{n}{72k^2}\Big).$$

Choosing n large enough, this probability will be less than $1-b$, and hence

$$\mathsf{P}\big(\mathbb{G}(n,W)\in\mathcal{R}''\big)>b.$$

So (a) is satisfied. The proof of (b) is analogous, and so is the proof of the converse. $\square$

It follows from the definition that a graphon property is testable if and only if its closure in the $\|.\|_1$ norm is testable. Furthermore, the closure in the $\|.\|_\square$ norm of a testable property is testable (but not the other way around, see example 3.6(d) below). It follows from Theorem 3.4 below (but it is not hard to see directly too) that if $\mathcal{R}$ is testable, then its closures in the $\|.\|_1$ norm and $\|.\|_\square$ norm coincide.

Since the distribution of $\mathbb{G}(k,W)$ is preserved under isomorphism of W, every closed testable function property is a graphon property. In our applications to graph theory only closed graphon properties will play a role, and we'll focus on characterizing testability for such properties.

It is trivial that $B_1(\mathcal{R},\varepsilon)\subseteq B_\square(\mathcal{R},\varepsilon)$ for every $\mathcal{R}\subseteq\mathcal{W}$ and $\varepsilon>0$. A reverse containment characterizes testable graphon properties.

Theorem 3.4 *A graphon property $\mathcal{R}$ is testable if and only if for every $\varepsilon>0$ there is an $\varepsilon'>0$ such that $B_\square(\mathcal{R},\varepsilon')\subseteq B_1(\mathcal{R},\varepsilon)$.*

Proof. Suppose that $\mathcal{R}$ is testable with test property $\mathcal{R}'$. Let $\varepsilon>0$, let $k=k(\varepsilon)$ be the constant in the definition, and let $\varepsilon'=2^{-k^2}$. Suppose that for some $W\in\mathcal{W}_0$, we have $d_\square(W,\mathcal{R})<\varepsilon'$. Then there is a $U\in\mathcal{R}$ such that $d_\square(W,U)<\varepsilon'$. By (4), we have for every graph F on k nodes

$$|t_{\mathrm{ind}}(F,W)-t_{\mathrm{ind}}(F,U)|\le\binom{k}{2}\varepsilon'.$$

Thus

$$|\mathsf{P}(\mathbb{G}(k,W)\notin\mathcal{R}')-\mathsf{P}(\mathbb{G}(k,U)\notin\mathcal{R}')|\le\sum_{F\in\mathcal{F}_n\setminus\mathcal{R}'}|t_{\mathrm{ind}}(F,W)-t_{\mathrm{ind}}(F,U)|\le 2^{\binom{k}{2}}\binom{k}{2}\varepsilon'<\frac{1}{3}.$$

Hence $\mathsf{P}(\mathbb{G}(n,W)\notin\mathcal{R}')<\mathsf{P}(\mathbb{G}(n,U)\notin\mathcal{R}')+\frac{1}{3}\le\frac{2}{3}$, which proves that $d_1(W,\mathcal{R})\le\varepsilon$.

Conversely, suppose that $\mathcal{R}$ satisfies the condition in the proposition. Choose

$$k_n=\frac{1}{2}\sqrt{\log n},\qquad \varepsilon_n=n^{-1/3}.$$

Let $\mathcal{R}'$ be the graph property that a graph G with $|V(G)|=n$ has if and only if there exists a $U\in\mathcal{R}$ such that $|t(F,U)-t(F,G)|\le\varepsilon_n$ for all graphs F with $|V(F)|\le k_n$. We show that (a) and (b) are satisfied.

First, suppose that $W\in\mathcal{R}$. Let n be large, and let F be a graph with at most k_n nodes. By Theorem 2.3, we have $|t(F,\mathbb{G}(n,W))-t(F,W)|\le\varepsilon_n$ with probability at least

$$1-2\exp\big(-\frac{n\varepsilon_n^2}{18k_n^2}\big).$$

Since the number of graphs on at most k_n nodes is at most $\exp(k_n^2/2)$, the probability that $|t(F,\mathbb{G}(n,W))-t(F,W)|\le\varepsilon_n$ holds for every graph F on at most k_n nodes is at least

$$1-2\exp\big(\frac{k_n^2}{2}-\frac{n\varepsilon_n^2}{18k_n^2}\big)>\frac{2}{3}$$

if n is large enough. Since $W \in \mathcal{R}$, in every such case $\mathbb{G}(n,W) \in \mathcal{R}'$. This proves that $\mathcal{R}'$ satisfies (a).

Second, let $\varepsilon > 0$ and suppose that $d_1(W,\mathcal{R}) > \varepsilon$. By hypothesis, there is an $\varepsilon' > 0$ (depending only on ε) such that $d_\square(W,\mathcal{R}) > \varepsilon'$.

Let n be large, and suppose, by way of contradiction, that with probability larger than 1/3, we have $\mathbb{G}(n,W) \in \mathcal{R}'$. In every such case, there exists a function $U \in \mathcal{R}$ such that

$$|t(F,U) - t(F,\mathbb{G}(n,W))| \leq \varepsilon_n \tag{10}$$

for all graphs F with at most k_n nodes. Similarly as above, Theorem 2.5 in [10] implies that

$$|t(F,\mathbb{G}(n,W)) - t(F,W)| \leq \varepsilon_n \tag{11}$$

for all F with at most k_n nodes with probability at least 2/3. There is an outcome for $\mathbb{G}(n,W)$ for which both (10) and (11) occur, and so there always exists a function $U \in \mathcal{R}$ such that

$$|t(F,W) - t(F,U)| \leq 2\varepsilon_n$$

for all graphs F with at most k_n nodes. By Theorem 3.6 in [7], it follows that

$$d_\square(U,W) \leq \frac{22}{\sqrt{\log_2 k_n}}.$$

If n is large enough, this is less that ε', contradicting the definition of ε'. $\square$

We also prove the following characterization, which is a functional analogue of the characterization of Alon, Fischer, Newman and Shapira [1].

Theorem 3.5 *A graphon property $\mathcal{R}$ is testable if and only if for every $\varepsilon > 0$ there is an $\varepsilon' > 0$ and a finite set S of stepfunctions such that $\mathcal{R} \subseteq B_\square(S,\varepsilon') \subseteq B_1(\mathcal{R},\varepsilon)$.*

This theorem gives a "constructive" method of testing a testable graphon property: for every fixed error bound, it suffices to compute the $d_\square$ distance from a finite number of stepfunctions to separate the case when $W \in \mathcal{R}$ from the case when $d_1(W,\mathcal{R}) \geq \varepsilon$.

Proof. First, suppose that $\mathcal{R}$ is testable, and let $\varepsilon > 0$. By Theorem 3.4, there is an $\varepsilon' > 0$ such that $B_\square(\mathcal{R},2\varepsilon') \subseteq B_1(\mathcal{R},\varepsilon)$. For every stepfunction s, consider the open ball $B_\square(s,\varepsilon')$. These balls cover the whole space, so by the compactness of $(\mathcal{W},\delta_\square)$ there is a finite set S_0 of stepfunctions such that the corresponding balls cover the whole space. Let S be the set of those stepfunctions in S_0 for which the corresponding balls intersect $\mathcal{R}$. Then clearly $\mathcal{R} \subseteq B_\square(S,\varepsilon')$. On the other hand, $B_\square(U,\varepsilon')$ intersects $\mathcal{R}$ for every $U \in S$, and hence $B_\square(S,\varepsilon') \subseteq B_\square(\mathcal{R},2\varepsilon') \subseteq B_1(\mathcal{R},\varepsilon)$.

Second, suppose that $\mathcal{R}$ satisfies the condition in the theorem, then for every $\varepsilon > 0$ there exists an $\varepsilon' > 0$ and a finite set S of stepfunctions such that $\mathcal{R} \subseteq B_\square(S,\varepsilon') \subseteq B_1(\mathcal{R},\varepsilon/2)$. Let $\varepsilon'' = \varepsilon\varepsilon'/6$. We claim that

$$B_\square(S,\varepsilon' + \varepsilon'') \subseteq B_1(B_\square(S,\varepsilon'),\varepsilon/2). \tag{12}$$

Indeed, let $W \in B_\square(\mathcal{S}, \varepsilon' + \varepsilon'')$. Then there is a $U \in \mathcal{S}$ such that $\|U - W\|_\square < \varepsilon' + 2\varepsilon''$. Consider $Y = (1 - \frac{1}{3}\varepsilon)W + \frac{1}{3}\varepsilon U$. Then

$$\|Y - U\|_\square = \|(1 - \frac{1}{3}\varepsilon)(U - W)\|_\square < (1 - \frac{1}{3}\varepsilon)(\varepsilon' + 2\varepsilon'') < \varepsilon',$$

so $Y \in B_\square(S, \varepsilon')$. On the other hand,

$$\|W - Y\|_1 = \|\frac{1}{3}\varepsilon(W - U)\|_1 \le \frac{1}{3}\varepsilon < \frac{1}{2}\varepsilon,$$

and so $W \in B_1(B_\square(S, \varepsilon'), \varepsilon/2)$. This proves (12), which in turn implies that

$$B_\square(\mathcal{R}, \varepsilon'') \subseteq B_\square(B_\square(S, \varepsilon'), \varepsilon'') \subseteq B_1(B_\square(S, \varepsilon'), \varepsilon/2) \subseteq B_1(B_1(\mathcal{R}, \varepsilon/2), \varepsilon/2) = B_1(\mathcal{R}, \varepsilon).$$

This proves that $\mathcal{R}$ is testable. $\square$

We conclude this section with some examples of testable and non-testable function properties.

Example 3.6 (a) Let $\mathcal{R} = \{U\}$, where $U \in \mathcal{W}$ is the identically 1/2 function. Clearly this is invariant under isomorphism. Consider a random graph $G_n = \mathbb{G}(n, 1/2)$; then $\|W_{G_n} - U\|_\square \to 0$ with probability 1, but $\|W_{G_n} - U\|_1 = 1/2$ for every n. So this property is not testable by Theorem 3.4.

On the other hand, the complementary property $\mathcal{R}^c = \mathcal{W}_0 \setminus \{U\}$ is testable; indeed, its closure is $\mathcal{W}_0$ (either in the $\|.\|_\square$ norm or the $\|.\|_1$ norm), which is trivially testable.

(b) Let $\mathcal{S} \subseteq \mathcal{W}_0$ be an arbitrary graphon property and let $a > 0$ be an arbitrary number. Then $\mathcal{R} = B_\square(\mathcal{S}, a)$ is testable.

Indeed, for $\varepsilon > 0$ define $\varepsilon' = a\varepsilon/(2 - 2\varepsilon)$. Let $W \in B_\square(\mathcal{R}, \varepsilon')$. Then $W \in B_\square(\mathcal{S}, a + \varepsilon')$, and so there is a $U \in \mathcal{S}$ such that $\|U - W\|_\square \le a + 2\varepsilon'$. Consider $Y = (1 - \varepsilon)W + \varepsilon U$. Then $\|Y - U\|_\square = \|(1 - \varepsilon)(U - W)\|_\square \le (1 - \varepsilon)(a + 2\varepsilon') = a$, so $Y \in \mathcal{R}$. On the other hand, $\|W - Y\|_1 = \|\varepsilon(W - U)\|_1 \le \varepsilon$, and so $W \in B_1(\mathcal{R}, \varepsilon)$.

(c) For every fixed graph F and $0 < c < 1$, the property $\mathcal{R}$ that $t(F, W) = c$ is testable; an appropriate test property is

$$\mathcal{R}' = \Big\{G : \ |t(F, G) - c| \le \frac{1}{\log |V(G)|}\Big\}. \tag{13}$$

Indeed, it follows from Theorem 2.3 that with probability $1 - o(1)$,

$$|t(F, W) - t(F, \mathbb{G}(n, W))| \le \frac{1}{\log n}.$$

So if $t(F, W) = c$, then $\mathbb{G}(n, W) \in \mathcal{R}'$ with large probability. Conversely, assume that $d_1(W, \mathcal{R}) > \varepsilon$, and let (say) $t(F, W) > c$. The functions $U_s = (1 - s)W$, $0 \le s \le \varepsilon$, are all in $B_1(W, \varepsilon)$, and hence not in $\mathcal{R}$. It follows that $t(F, U_s) > c$ for all s. On the other hand, $t(F, U_\varepsilon) = (1 - \varepsilon)^{|E(F)|} t(F, W)$, and so $t(F, W) > (1 - \varepsilon)^{-|E(F)|} c$, which implies that $\mathbb{G}(n, W) \notin \mathcal{R}'$ with large probability.

Fixing two subgraph densities, however, may yield a non-testable property: for example, $t(K_2, W) = 1/2$ and $t(C_4, W) = 1/16$ imply that $W \equiv 1/2$ (see [6]).

(d) The graphon property that W is 0-1 valued is not testable. It is closed in the $\|.\|_1$ norm, but its closure in the $\|.\|_\square$ norm is the whole set $\mathcal{W}_0$.

3.2 Graph properties vs. function properties

3.2.1 Closure of graph properties

We want to establish the connection between testability of graph properties and graphon properties. The fact that graphons arise as limits of graph sequences suggests the following definition.

Definition 3.7 If $\mathcal{P}$ is a graph property, then we define its *closure* $\overline{\mathcal{P}}$ as the set of all functions $W \in \mathcal{W}_0$ for which there exists a sequence of graphs $G_n \in \mathcal{P}$ with $|V(G_n)| \to \infty$ such that $G_n \to W$ (i.e., W_{G_n} converges to W in the $\delta_\square$ metric).

Clearly $\overline{\mathcal{P}}$ is closed under isomorphism, i.e. it is a graphon property. The following examples show that $\overline{\mathcal{P}}$ is not necessarily an "extension" of $\mathcal{P}$ in the sense that $\mathcal{P}$ cannot be recovered from it. Intuitively, $\overline{\mathcal{P}}$ is a nice object which is a "clean" version of $\mathcal{P}$; it is an analytic profile of the property $\mathcal{P}$, which eliminates all uncontrollable noise from it.

Example 3.8 (a) Let $\mathcal{P}$ be the graph property that the graph doesn't have a 4-cycle. Then only the 0 function has property $\overline{\mathcal{P}}$. In fact, graphs without 4-cycles are sparse and property testing (in the sense of Definition 1.1) does not distinguish sparse graphs from each other.

(b) Let $\mathcal{P}$ be the graph property that the graph has an even number of edges. Rather counter-intuitively, this property is testable according to Definition 1.1 above, and its closure is the whole set $\mathcal{W}_0$.

(c) Let $\mathcal{P}$ be the graph property that the graph has an even number of nodes. This property is not testable, since adding a single node to a large graph changes the distribution of small induced subgraphs by very little, but changes the property. The closure of this property is again the whole set $\mathcal{W}_0$ (which is testable).

(d) Quasirandomness is defined as a property of a sequence of graphs, but we can make it a graph property $\mathcal{Q}$ (at the cost of a somewhat arbitrary choice of the error bound) as follows: a graph G on n nodes is *quasirandom*, if

$$\left|t(K_2, G) - \frac{1}{2}\right| \le \frac{1}{\log n} \quad \text{and} \quad t(C_4, G) \le \frac{1}{16} + \frac{1}{\log n}.$$

The closure $\overline{\mathcal{Q}}$ of this property consists of only one function, the identically $1/2$ function. This singleton set of functions is not testable, since for any sequence (G_n) of quasirandom graphs, $\|W_{G_n} - \frac{1}{2}\|_\square \to 0$ but $\|W_{G_n} - \frac{1}{2}\|_1 = 1/2$. This implies (by Theorem 3.18 below) that quasirandomness is not a testable property.

The first part of the following fact was stated in [5].

Proposition 3.9 (a) *The closure of a hereditary graph property $\mathcal{P}$ consists of those functions $W \in \mathcal{W}_0$ for which*

$$\mathsf{P}(\mathbb{G}(k, W) \notin \mathcal{P}) = 0$$

for every k. Equivalently, $t_{\text{ind}}(F, W) = 0$ whenever $F \notin \mathcal{P}$.

(b) *The closure of a testable graph property $\mathcal{P}$ consists of those functions $W \in \mathcal{W}_0$ for which*

$$\mathsf{E}(d_1(\mathbb{G}(k,W),\mathcal{P})) \to 0 \qquad (k \to \infty).$$

It follows from Theorem 3.5 that in the last formula, we could replace d_1 by $d_\square$.

Proof. (a) Assume that $\mathsf{P}(\mathbb{G}(k,W) \notin \mathcal{P}) = 0$ for every k. Since $\mathbb{G}(k,W) \to W$ with probability 1 as $k \to \infty$, it follows that $W \in \overline{\mathcal{P}}$.

To show the converse, assume that $W \in \overline{\mathcal{P}}$, and let (G_n) be a sequence of graphs such that $G_n \in \mathcal{P}$ and $G_n \to W$. We can write

$$\mathsf{P}(\mathbb{G}(k,W) \notin \mathcal{P}) = \sum_{F \in \mathcal{F}_k \setminus \mathcal{P}} \mathsf{P}(\mathbb{G}(k,W) = F) = \sum_{F \in \mathcal{F}_k \setminus \mathcal{P}} t_{\mathrm{ind}}(F,W), \tag{14}$$

so it suffices to prove that $t_{\mathrm{ind}}(F,W) = 0$ for $F \notin \mathcal{P}$. By (1), we have

$$|t_{\mathrm{ind}}(F,G_n) - t_{\mathrm{ind}}(F,W_{G_n})| < \frac{\binom{k}{2}}{n - \binom{k}{2}} \tag{15}$$

for every $F \in \mathcal{F}_k$. For $F \in \mathcal{F}_k \setminus \mathcal{P}$, we have $t_{\mathrm{ind}}(F,G_n) = 0$ since $G_n \in \mathcal{P}$ and the property is hereditary. Hence by (15), in this case

$$t_{\mathrm{ind}}(F,W_{G_n}) \to 0 \qquad (n \to \infty).$$

But $G_n \to W$ implies that

$$t_{\mathrm{ind}}(F,W_{G_n}) \to t_{\mathrm{ind}}(F,W),$$

which proves (a).

(b) The argument is similar, but a little more complicated. First, assume that $\mathsf{E}\big(d_1(\mathbb{G}(k,W),\mathcal{P})\big) \to 0$ as $k \to \infty$. Then we can select a sequence $k_1 < k_2 < \dots$ of integers such that

$$\mathsf{P}\Big(d_1(\mathbb{G}(k_m,W),\mathcal{P}) > \frac{1}{m}\Big) \le \frac{1}{2^m}.$$

Since the right hand sides have a finite sum, it follows by the Borel-Cantelli Lemma that $d_1(\mathbb{G}(k_m,W),\mathcal{P}) \to 0$ with probability 1. Hence with probability 1, there are graphs $H_m \in \mathcal{P}$ such that $d_1(\mathbb{G}(k_m,W),H_m) \to 0$, which implies that $\delta_1(\mathbb{G}(k_m,W),H_m) \to 0$. Furthermore, if the sequence (k_m) is sufficiently sparse, we have $\mathbb{G}(k_m,W) \to W$ with probability 1 as $k \to \infty$, and so $H_m \to W$ with probability 1. Thus $W \in \overline{\mathcal{P}}$.

Second, assume that $W \in \overline{\mathcal{P}}$, and let (G_n) be a sequence of graphs such that $G_n \in \mathcal{P}$ and $G_n \to W$. Fix any $\varepsilon > 0$. By the remark after the definition of property testing, there is a graph property $\mathcal{P}'$ and a natural number k_ε such that for all k,n with $k_\varepsilon \le k \le |V(G_n)|$ we have $\mathsf{P}\big(\mathbb{G}(k,G_n) \notin \mathcal{P}'\big) \le \varepsilon$, but $\mathsf{P}\big(\mathbb{G}(k,G) \notin \mathcal{P}'\big) > 1/2$ for all graphs G with $|V(G)| \ge k$ and $d_1(G,\mathcal{P}) > \varepsilon$. By (14) we have

$$\begin{aligned}
\mathsf{P}\big(\mathbb{G}(k,W) \notin \mathcal{P}'\big) &= \sum_{F \in \mathcal{F}_k \setminus \mathcal{P}'} t_{\mathrm{ind}}(F,W) \\
&\le \sum_{F \in \mathcal{F}_k \setminus \mathcal{P}'} t_{\mathrm{ind}}(F,G_n) + \sum_{F \in \mathcal{F}_k \setminus \mathcal{P}'} |t_{\mathrm{ind}}(F,G_n) - t_{\mathrm{ind}}(F,W_{G_n})| \\
&\quad + \sum_{F \in \mathcal{F}_k \setminus \mathcal{P}'} |t_{\mathrm{ind}}(F,W_{G_n}) - t_{\mathrm{ind}}(F,W)|.
\end{aligned}$$

Here the first sum is $\mathsf{P}\big(\mathbb{G}(k, G_n) \notin \mathcal{P}'\big) \le \varepsilon$. The other two sums tend to 0 as $n \to \infty$ by (15) and by $G_n \to W$. So for $k \ge k_0$, we get that $\mathsf{P}\big(\mathbb{G}(k, W) \notin \mathcal{P}'\big) \le \varepsilon$.

Note that $\mathbb{G}(k, W)$ has the same distribution as $\mathbb{G}(k, \mathbb{G}(n, W))$ for $n \ge k$. Then for large enough k, n we have

$$\mathsf{P}\big(\mathbb{G}(k, W) \notin \mathcal{P}'\big) \ge \frac{1}{2}\mathsf{P}\big(d_1(\mathbb{G}(n, W), \mathcal{P}) > \varepsilon\big),$$

and so $\mathsf{P}\big(d_1(\mathbb{G}(n, W), \mathcal{P}') > \varepsilon\big) \le 2\varepsilon$. Hence

$$\mathsf{E}\big(d_1(\mathbb{G}(n, W), \mathcal{P})\big) \le (1 - 2\varepsilon) \cdot \varepsilon + (2\varepsilon) \cdot 1 < 3\varepsilon.$$

Since ε was arbitrary, this proves that $\mathsf{E}\big(d_1(\mathbb{G}(n, W), \mathcal{P})\big) \to 0$. $\square$

Part (a) of Proposition 3.9 implies:

Corollary 3.10 *If $\mathcal{P}$ is a hereditary graph property, then $W \in \overline{\mathcal{P}}$ depends only on the support of W.*

We conclude with two lemmas on testability. We'll say more about both (Theorems 3.20 and 3.18), but these simple lemmas will be needed before that.

Lemma 3.11 *Let $\mathcal{P}$ be a testable graph property. Then for every $\varepsilon > 0$ there is an $\varepsilon' > 0$ and a positive integer n' such that if G are two graphs with $G' \in \mathcal{P}$, $\delta_\square(G, G') < \varepsilon'$ and $|V(G)|, |V(G')| \ge n'$, then $d_1(G, \mathcal{P}) < \varepsilon$.*

Proof. Let $\varepsilon > 0$ and let $k = k(\varepsilon)$ in the definition of testability. We show that $n' = 9k^2$ and $\varepsilon' = 1/k^2$ is a good choice. Since G' has property $\mathcal{P}$, we have that at least 2/3 of its k-node induced subgraphs have property $\mathcal{P}'$. It follows by (5) that more than 1/3 of the k-node subgraphs of G have the property $\mathcal{P}'$. Hence $d_1(G, \mathcal{P}) < \varepsilon$ by testability. $\square$

Lemma 3.12 *If $\mathcal{P}$ is testable then $\overline{\mathcal{P}}$ is testable.*

The converse is not true in general, but in Theorem 3.18 we will give a characterization of testable properties in terms of their closure.

Proof. It suffices to prove that if (W_n) is a sequence of functions in $\mathcal{W}_0$ such that $d_\square(W_n, \overline{\mathcal{P}}) \to 0$, then $d_1(W_n, \overline{\mathcal{P}}) \to 0$. We may assume that the sequence W_n is convergent, so $W_n \to U$ for some $U \in \mathcal{W}_0$ (in the $\delta_\square$ distance). Clearly $U \in \overline{\mathcal{P}}$, so by the definition of closure, there are graphs $H_n \in \mathcal{P}$ such that $|V(H_n)| \to \infty$ and $H_n \to U$.

Fix any $\varepsilon > 0$. By Lemma 3.11, there is an $\varepsilon' > 0$ such that if $|V(G)|, |V(H)|$ are large enough, $H \in \mathcal{P}$, and $\delta_\square(G, H) < \varepsilon'$, then $d_1(G, \mathcal{P}) < \varepsilon$. Furthermore, there is an $n_\varepsilon \ge 1$ such that if $n \ge n_\varepsilon$, then $\delta_\square(W_{H_n}, U), \delta_\square(W_n, U) \le \varepsilon'/3$.

Fix any $n \ge n_\varepsilon$, and let $G_{n,m}$ ($m = 1, 2, \dots$) be a sequence of graphs such that $|V(G_{n,m})| \to \infty$ and $G_{n,m} \to W_n$ as $m \to \infty$. Then

$$\delta_\square(H_n, G_{n,m}) \le \delta_\square(W_{H_n}, U) + \delta_\square(U, W_n) + \delta_\square(W_n, W_{G_{n,m}}) < \varepsilon'$$

if m is large enough, hence by the choice of ε', we have $d_1(G_{n,m}, \mathcal{P}) \le \varepsilon$. This means that there are graphs $J_{n,m} \in \mathcal{P}$ with $V(J_{n,m}) = V(G_{n,m})$ such that $d_1(G_{n,m}, J_{n,m}) \le \varepsilon$. By choosing a subsequence, we can assume that $J_{n,m} \to U_n$ as $m \to \infty$ for some $U_n \in \overline{\mathcal{P}}$. Applying Lemma 2.11 we obtain that

$$d_1(W_n, \mathcal{P}) \le \delta_1(W_n, U_n) \le \liminf_{m\to\infty} \delta_1(W_{G_{n,m}}, W_{J_{n,m}}) \le \liminf_{m\to\infty} d_1(G_{n,m}, J_{n,m}) \le \varepsilon.$$

This proves the Lemma. $\square$

3.2.2 Closure and distance

What is the relationship between the d_1 distance from a property and from its closure? The following propositions summarize what we know. We say that a graph G' is an *equitable m-blowup* of G if it is obtained by replacing each node of G by m or $m+1$ twin copies ($m \ge 1$).

Proposition 3.13 (a) *For every hereditary graph property $\mathcal{P}$ and every graph G,*

$$d_1(G, \mathcal{P}) \le d_1(W_G, \overline{\mathcal{P}}).$$

(b) *For every testable graph property $\mathcal{P}$,*

$$|d_1(G, \mathcal{P}) - d_1(W_G, \overline{\mathcal{P}})| \to 0 \qquad (|V(G)| \to \infty).$$

(c) *Let $\mathcal{P}$ be an arbitrary graph property and let $G^1, G^2, \dots$ be all equitable blowups of a graph G. Then*

$$\liminf d_1(G^n, \mathcal{P}) = d_1(W_G, \overline{\mathcal{P}}).$$

Proof. (a) Let $\delta > 0$ and $U \in \overline{\mathcal{P}}$ be such that $\|W_G - U\|_1 \le d_1(W_G, \overline{\mathcal{P}}) + \delta$. Using Corollary 3.10, we may assume that U is a $0-1$-valued function. From the fact that $\mathbb{G}(n, U)$ has property $\mathcal{P}$ with probability 1, it follows that $G' = \mathbb{G}'(n, U)$ has property $\mathcal{P}$ with probability 1. By Lemma 2.6

$$\mathsf{E}(d_1(G, G')) = \|W_G - U\|_1 \le d_1(W_G, \overline{\mathcal{P}}) + \delta.$$

Hence there is an instance of G' for which $G' \in \mathcal{P}$ and $d_1(G, G') \le d_1(W_G, \mathcal{P}) + \delta$. Thus $d_1(G, \mathcal{P}) \le d_1(G, G') \le d_1(W_G, \overline{\mathcal{P}}) + \delta$. Since δ is arbitrary, this proves (a).

(b) Suppose not, then there exists a sequence of graphs (G_n) with $|V(G_n)| \to \infty$ such that $d_1(G_n, \mathcal{P}) \to a$ and $d_1(W_{G_n}, \overline{\mathcal{P}}) \to b$, where $a \ne b$. We may assume that $V(G_n) = [q_n]$, where $q_n \to \infty$.

First, select graphs $H_n \in \mathcal{P}$ such that $V(H_n) = V(G_n)$ and $d_1(G_n, H_n) = d(G_n, \mathcal{P})$. By selecting a subsequence, we may assume that the sequence H_n is convergent; let $U \in \mathcal{W}_0$ be its limit. Clearly $U \in \overline{\mathcal{P}}$. Then $\delta_\square(W_{H_n}, \overline{\mathcal{P}}) \le \delta_\square(W_{H_n}, U) \to 0$, and hence $d_\square(W_{H_n}, \overline{\mathcal{P}}) \to 0$. By Theorem 3.4, this implies that $d_1(W_{H_n}, \overline{\mathcal{P}}) \to 0$. But then in the inequality

$$d_1(W_{G_n}, \overline{\mathcal{P}}) \le d_1(W_{G_n}, W_{H_n}) + d_1(W_{H_n}, \overline{\mathcal{P}})$$

the first term on the right hand side tends to a, while the second tends to 0, showing that $b \le a$.

Second, fix $\varepsilon > 0$, and let $U_n \in \overline{\mathcal{P}}$ be such that $d_1(W_{G_n}, U_n) \le d_1(W_{G_n}, \overline{\mathcal{P}}) + \varepsilon$. By the definition of testability of functions, there is an $m \ge 1$ such that for $m \ge m_0$ and every n, we have that $\mathbb{G}(m, U_n) \in \mathcal{P}'$ with probability at least $2/3$.

Similarly as in the proof of (a), consider the random graph $G' = \mathbb{G}'(q_n, U_n)$. Clearly

$$d_1(G_n, \mathcal{P}) \le d_1(G_n, G') + d_1(G', \mathcal{P}). \tag{16}$$

Here by Lemma 2.6

$$\mathsf{E}(d_1(G_n, G')) = d_1(W_{G_n}, U_n). \tag{17}$$

Next we show that, with probability tending to 1 as $n \to \infty$, we have

$$d_1(G_n, G') \le d_1(W_{G_n}, U_n) + \varepsilon. \tag{18}$$

Indeed, the left hand side is the sum of $q_n/2$ independent random variables, all $0-1$ valued, so this follows by the Law of Large Numbers.

To estimate the other term in (16), let Z be a random m-element subset of X, and let Y be a random m-elements subset of $[0,1]$. The distributions of Z and Y are very close; indeed, we can generate Z by generating Y and accepting it if its elements fall into different intervals $[i/q_n, (i+1)/q_n]$, $i = 0, \dots, q_n - 1$, and only regenerate otherwise. The probability that we need to regenerate tends to 0 as $n \to \infty$, so the total variation distance of Z and Y tends to 0. The probability that $\mathbb{G}(Y, U_n) \in \mathcal{P}'$ is at least $2/3$ as $\overline{\mathcal{P}}$ is testable by Lemma 3.12, and so the probability that $\mathbb{G}(Z, U_n) \in \mathcal{P}'$ is at least $1/2$. But $\mathbb{G}(Z, U_n)$ has the same distribution as a random m-node subgraph of G', and so by testability,

$$\mathsf{P}\big(d_1(G', \mathcal{P}) \le \varepsilon\big) + \frac{1}{3}\mathsf{P}\big(d_1(G', \mathcal{P}) > \varepsilon\big) \ge \frac{1}{2}.$$

This implies that with probability at least $1/4$,

$$d_1(G', \mathcal{P}) \le \varepsilon. \tag{19}$$

With positive probability, both (18) and (19) occur, and so by (16) we have

$$d_1(G_n, \mathcal{P}) \le d_1(G_n, G') + d_1(G', \mathcal{P}) \le d_1(W_{G_n}, U_n) + \varepsilon + \varepsilon \le d_1(W_{G_n}, \overline{\mathcal{P}}) + 3\varepsilon.$$

Sending $n \to \infty$ we see that $a \le b + 3\varepsilon$. Since ε was arbitrary, it follows that $a \le b$, which is a contradiction.

(c) Let $\varepsilon > 0$, and let $U \in \overline{\mathcal{P}}$ be a function such that $\|W_G - U\|_1 \le d_1(W_G, U) + \varepsilon$. Let H_n be a sequence of graphs such that $H_n \to U$ and $H_n \in \mathcal{P}$. Then for an appropriate labeling of the nodes of H_n, we have $\|W_{H_n} - U\|_\square \to 0$. Since W_G is $0-1$ valued, Lemma 2.9 implies that $\|W_{H_n} - W_G\|_1 \to \|U - W_G\|_1$. Let $V(G) = \{1, \dots, k\}$ and $V(H_n) = \{1, \dots, m\}$. Choose n large enough so that $\|W_{H_n} - W_G\|_1 \le \|U - W_G\|_1 + \varepsilon$ and $m \ge k/\varepsilon$.

Let $V_i = \big\{\lfloor (i-1)m/k \rfloor + 1, \dots, \lfloor im/k \rfloor\big\}$ for $i = 1, \dots, k$. Then $(V_1, \dots, V_k)$ is a partition of $V(H_n)$ into k almost equal classes. Define a graph G' on $\{1, \dots, k\}$ by connecting $u \in V_i$ to $v \in V_j$ if and only if $ij \in E(G)$. Then G' is an equitable blowup of G. Furthermore, $W_{G'}$ and

W_G differ only on stripes of width less than $1/m$ along the orders of the squares on which W_G is constant, so $\|W_{G'} - W_G\|_1 \le 2k/m \le 2\varepsilon$. Thus

$$\begin{aligned} d_1(G', \mathcal{P}) &\le d_1(G', H_n) = \|W_{G'} - W_{H_n}\|_1 \le \|W_{G'} - W_G\|_1 + \|W_G - W_{H_n}\|_1 \\ &\le 2\varepsilon + \big(\|U - W_G\|_1 + \varepsilon\big) \le d_1(W_G, \overline{\mathcal{P}}) + 4\varepsilon. \end{aligned}$$

Since ε was arbitrary, this proves that

$$\liminf d_1(G^n, \mathcal{P}) \le d_1(W_G, \overline{\mathcal{P}}).$$

To prove the converse, let $a = \liminf d_1(G^n, \mathcal{P})$, and let $H_n \in \mathcal{P}$ be chosen so that $V(H_n) = V(G^n)$ and $d_1(H_n, G^n) = d_1(G^n, \mathcal{P})$. Select a subsequence such $d_1(H_n - G^n) \to a$, and choose a further subsequence so that H_n is convergent. Let $H_n \to U \in \overline{\mathcal{P}}$. We have $\delta_\square(W_{H_n}, U) \to 0$ and $\delta_\square(W_{G^n}, W_G) \to 0$, hence by Lemma 2.11,

$$\delta_1(W_G, U) \le \liminf \delta_1(W_{H_n}, W_{G_n}) \le a.$$

Hence $d_1(W_G, \overline{\mathcal{P}}) \le \delta_1(W_G, U) \le a$. $\square$

Proposition 3.14 *Let $\mathcal{P}$ be any graph property and $G_n \to W$, a convergent graph sequence. Then*

$$\liminf_{n\to\infty} d_1(G_n, \mathcal{P}) \ge d_1(W, \overline{\mathcal{P}}),$$

Proof. Let $H_n \in \mathcal{P}$ be such that $V(G_n) = V(H_n)$ and $d_1(G_n, H_n) = d_1(G_n, \mathcal{P})$. We may select a subsequence so that $H_n \to U$ for some $U \in \mathcal{W}_0$. Clearly $U \in \overline{\mathcal{P}}$. Furthermore, $\|W_{G_n} - W\|_\square \to 0$ and $\|W_{H_n} - U\|_\square \to 0$, so by Lemma 2.8, we have

$$d_1(W, \mathcal{P}) \le \|W - U\|_1 \le \liminf \|W_{G_n} - W_{H_n}\|_1 = \liminf d_1(G_n, \mathcal{P}),$$

a contradiction. $\square$

3.2.3 Monotone closure

For two functions $U, W \in \mathcal{W}_0$ we write $U \preceq W$ if there exist functions $U', W' \in \mathcal{W}$ such that $U \cong U'$, $W \cong W'$, and $U' \le W'$ almost everywhere.

Let $\mathcal{P}$ a graph property. By its *upward closure* we mean the graph property $\mathcal{P}^\uparrow$ consisting of those graphs that have a spanning subgraph in $\mathcal{P}$. For a function property $\mathcal{R}$, we define its *upward closure* to consist of those functions $W \in \mathcal{W}_0$ for which there exists a function $U \in \mathcal{R}$ such that $U \preceq W$. In both versions, the *downward closure* is defined analogously.

The following theorem, whose proof is surprisingly nontrivial, asserts that closure and upward closure commute.

Theorem 3.15 *For every graph property $\mathcal{P}$,*

$$\overline{\mathcal{P}^\uparrow} = (\overline{\mathcal{P}})^\uparrow.$$

Proof. First, let $W \in \overline{\mathcal{P}^\uparrow}$. Then there exists a graph sequence $G_n \to W \in \mathcal{W}_0$ such that G_n has a spanning subgraph $G'_n \in \mathcal{P}$. We consider the pair (G_n, G'_n) as the graph G_n in which the edges of G'_n colored red, the remaining edges are colored blue. We may choose a subsequence of the indices so that the remaining sequence is convergent as *2-edge-colored graphs*, meaning that for every 2-edge-colored simple graph F, the sequence of densities $t(F, G_n)$ of the densities of color-preserving homomorphisms is convergent. It is shown in [12] that the limit object of a such a sequence can be described by a pair of functions $U, V \in \mathcal{W}_0$, such that V is the limit of the sequence (G_n), U is the limit of the sequence (G'_n), and $U \le V$ almost everywhere. Hence $V \cong W$ and $U \in \overline{\mathcal{P}}$. This proves that $W \in (\overline{\mathcal{P}})^\uparrow$.

Conversely, let $W \in (\overline{\mathcal{P}})^\uparrow$, then there is a $U \in \overline{\mathcal{P}}$ and a $V \cong W$ such that $U \le V$. Let $G_n \in \mathcal{P}$ be such that $G_n \to U$. By Lemma 4.16 in [7], we may label the graphs G_n so that $\|W_{G_n} - U\|_\square \to 0$. Let $\mathcal{L} = \mathcal{L}_n$ denote the partition of $[0,1]$ into $N = |V(G_n)|$ equal intervals $\{I_1, \dots, I_N\}$, and for $W \in \mathcal{W}$, let $W_\mathcal{L}$ denote the function obtained by replacing W by its average on each of the intervals $I_i \times I_j$. Clearly $U_\mathcal{L} \le V_\mathcal{L}$, and

$$\|U - U_\mathcal{L}\|_\square \to 0, \qquad \|V - V_\mathcal{L}\|_\square \to 0 \tag{20}$$

(this holds even with the L_1-norm in place of the cut norm), and so

$$\|U_\mathcal{L} - W_{G_n}\|_\square \to 0. \tag{21}$$

Define

$$W'_n = V_\mathcal{L} + \frac{1 - V_\mathcal{L}}{1 - U_\mathcal{L}}(W_{G_n} - U_\mathcal{L})$$

(where the second term is 0 wherever $U_\mathcal{L} = 1$). Then

$$W'_n \le V_\mathcal{L} + \frac{1 - V_\mathcal{L}}{1 - U_\mathcal{L}}(1 - U_\mathcal{L}) = 1,$$

and

$$W'_n = V_\mathcal{L}\Big(1 - \frac{W_{G_n} - U_\mathcal{L}}{1 - U_\mathcal{L}}\Big) + \frac{W_{G_n} - U_\mathcal{L}}{1 - U_\mathcal{L}} \ge U_\mathcal{L}\Big(1 - \frac{W_{G_n} - U_\mathcal{L}}{1 - U_\mathcal{L}}\Big) + \frac{W_{G_n} - U_\mathcal{L}}{1 - U_\mathcal{L}} = W_{G_n}.$$

By Lemma 2.2,

$$\|W'_n - V_\mathcal{L}\|_\square = \Big\|\frac{1 - V_\mathcal{L}}{1 - U_\mathcal{L}}(W_{G_n} - U_\mathcal{L})\Big\|_\square \to 0 \qquad (n \to \infty). \tag{22}$$

Since W'_n is a stepfunction that is constant on intervals $I \times J$ $(I, J \in \mathcal{L})$, it can be viewed as W_{H_n} for some weighted graph H_n on $[N]$. Create a random graph G'_n as follows: for $1 \le i < j \le N$, connect i and j with probability equal to weight of the edge in H_n. Lemma 4.3 in [7] implies that with probability at least $1 - e^{-N}$,

$$\|W_{G'_n} - W_{H_n}\|_\square \le \frac{10}{\sqrt{N}},$$

and hence

$$\|W_{G'_n} - W'_n\|_\square \to 0 \qquad (n \to \infty). \tag{23}$$

Trivially, G_n is a subgraph of G'_n with probability 1, and (20), (21) and (23) imply that

$$\|W_{G'_n} - V\|_\square \le \|W_{G'_n} - W'_n\|_\square + \|W'_n - V_\mathcal{L}\|_\square + \|V_\mathcal{L} - V\|_\square \to 0.$$

This proves that $V \in \overline{\mathcal{P}^\uparrow}$, and so $W \in \overline{\mathcal{P}^\uparrow}$. $\square$

3.2.4 Robustness

The following definition will be important in our characterization of testable graph properties.

Definition 3.16 A graph property $\mathcal{P}$ is *robust*, if for every $\varepsilon > 0$ there are numbers $n = n(\varepsilon) > 0$ and $\varepsilon' > 0$ such that if G is a graph with $|V(G)| \geq n(\varepsilon)$ and $d_1(W_G, \overline{\mathcal{P}}) \leq \varepsilon'$, then $d_1(G, \mathcal{P}) \leq \varepsilon$.

Proposition 3.13(b) implies that every testable property, in particular every hereditary property, is robust. The following fact, which is an immediate consequence of Proposition 3.13, provides a combinatorial criterion for robustness.

Proposition 3.17 *A graph property $\mathcal{P}$ is robust if and only if for every $\varepsilon > 0$ there is an $\varepsilon' > 0$ and an $n_\varepsilon \geq 1$ such that if G is a graph with $|V(G)| \geq n_\varepsilon$ and G has infinitely many equitable blowups G' with $d_1(G', \mathcal{P}) \leq \varepsilon'$, then $d_1(G, \mathcal{P}) \leq \varepsilon$.*

3.2.5 Characterizing testability of graph properties

Our next main theorem shows the relationship between analytic and graph theoretic testability.

Theorem 3.18 *A graph property $\mathcal{P}$ is testable if and only if it is robust and its closure $\overline{\mathcal{P}}$ is testable.*

Proof. We have seen that every testable graph property is robust (Proposition 3.13(b)), and that its closure is testable (Lemma 3.12). So to complete the proof, it suffices to prove that if $\mathcal{P}$ is robust and $\overline{\mathcal{P}}$ is testable then $\mathcal{P}$ is testable.

Let $\varepsilon > 0$. By the robustness of $\mathcal{P}$, there is an $\varepsilon' > 0$ and $k_\varepsilon \geq 1$ such that if $|V(G)| \geq k_\varepsilon$ and $d_1(W_G, \overline{\mathcal{P}}) < \varepsilon'$, than $d_1(G, \mathcal{P}) < \varepsilon$. By the testability of $\overline{\mathcal{P}}$ and Theorem 3.4, there is an $\varepsilon'' > 0$ such that if $d_\square(W, \overline{\mathcal{P}}) < \varepsilon''$ then $d_1(W, \overline{\mathcal{P}}) < \varepsilon'$. By the definition of $\bar{\mathcal{P}}$ and by Theorem 2.9 in [7], there exists an $n_\varepsilon \geq k_\varepsilon$ such that

(i) for every graph $G \in \mathcal{P}$ with $|V(G)| \geq n_\varepsilon$, we have $d_\square(W_G, \overline{\mathcal{P}}) < \varepsilon''/4$;

(ii) for every graph G with $|V(G)| \geq n_\varepsilon$ and every $n_\varepsilon \leq m \leq |V(G)|$, we have $\delta_\square(\mathbb{G}(m, G), G) < \varepsilon''/4$ with probability at least 2/3.

Let $\mathcal{P}'$ denote the property of a graph G that $d_\square(W_G, \overline{\mathcal{P}}) \leq \varepsilon''/2$ (this depends on ε, but as we remarked after the definition, this is OK). We claim that $\mathcal{P}'$ is a good test property for $\mathcal{P}$ (for the given ε).

Let G be a graph, $n_\varepsilon \leq n \leq |V(G)|$, and let $H = \mathbb{G}(n, G)$. First, suppose that $G \in \mathcal{P}$. By (i), $d_\square(W_G, \overline{\mathcal{P}}) < \varepsilon''/4$. Further by (ii), $\delta_\square(H, G) = \delta_\square(W_H, W_G) \leq \varepsilon''/4$ with probability at least 2/3. In every such case,

$$d_\square(W_H, \overline{\mathcal{P}}) \leq \delta_\square(W_H, W_G) + d_\square(W_G, \overline{\mathcal{P}}) < \varepsilon''/2.$$

Thus H has property $\mathcal{P}'$.

Second, suppose that $d_1(G, \mathcal{P}) \geq \varepsilon$; we want to prove by contradiction that H does not have the property $\mathcal{P}'$ with probability at least 2/3. Assume that it is not true, then with probability larger than 1/3, $d_\square(W_H, \overline{\mathcal{P}}) \leq \varepsilon''/2$. We have $d_\square(W_G, W_H) \leq \varepsilon''/4$ with probability more than

2/3. This implies that there exists at least one induced subgraph H of G with n nodes such that $d_\square(W_G, W_H) < \varepsilon''/4$ and $d_\square(W_H, \overline{\mathcal{P}}) \le \varepsilon''/2$. We obtain that $d_\square(W_G, \overline{\mathcal{P}}) < \varepsilon''$. It follows from the testability of $\overline{\mathcal{P}}$ that $d_1(W_G, \overline{\mathcal{P}}) < \varepsilon'$. It follows from the robustness of $\mathcal{P}$ that $d_1(G, \mathcal{P}) < \varepsilon$, a contradiction. $\square$

A further connection between testable graph properties and graphon properties is the following fact:

Theorem 3.19 *For every closed testable graphon property $\mathcal{R}$ there exists a testable graph property $\mathcal{P}$ such that $\mathcal{R} = \overline{\mathcal{P}}$.*

Proof. Let $\varepsilon > 0$. Since $\mathcal{R}$ is testable, it follows by Theorem 3.4 that there is and $\varepsilon' > 0$ such that $B_\square(\mathcal{R}, \varepsilon') \subseteq B_1(\mathcal{R}, \varepsilon)$. The set $\mathcal{R}$, with the distance function $\delta_\square$, is a compact metric space, and hence it can be covered by a finite number of balls $B_i = B_\square(W_i, \varepsilon'/2)$, $i = 1, \dots, m$. For every $1 \le i \le m$, there is a smallest positive integer n_i such that if $n \ge n_i$ then there is a graph G with n nodes such that $W_G \in B_i$. Let $n_\varepsilon = \max_i n_i$. So for every $U \in \mathcal{R}$, and $n \ge n_\varepsilon$ there is a graph G with n nodes such that (using the i for which $B_i \ni U$),

$$\delta_\square(W_G, U) \le \delta_\square(W_G, W_i) + \delta_\square(W_i, U) \le \frac{\varepsilon'}{2} + \frac{\varepsilon'}{2} = \varepsilon' \le \varepsilon.$$

This also implies that $\delta_1(W_G, \mathcal{R}) \le \varepsilon$.

For every $n \ge 1$, let

$$\varepsilon_n = \max\Big(\frac{50}{\sqrt{\log\log n}},\ \sup\{\varepsilon :\ n_\varepsilon \ge n\}\Big).$$

Clearly $\varepsilon_n \searrow 0$ as $n \to \infty$, and $\varepsilon_{n_\varepsilon} \ge \varepsilon$. We prove that the property $\mathcal{P} = \{G :\ \delta_1(W_G, \mathcal{R}) \le \varepsilon_{|V(G)|}\}$ is robust and its closure is $\mathcal{R}$.

First we show that $\mathcal{R} \supseteq \overline{\mathcal{P}}$. Indeed, if $W \in \overline{\mathcal{P}}$, then there is a sequence $G_n \in \mathcal{P}$ with $|V(G_n)| \to \infty$ such that $\delta_\square(W_{G_n}, W) \to 0$. Since $G_n \in \mathcal{P}$ means that $\delta_\square(W_{G_n}, \mathcal{R}) \le \delta_1(W_{G_n}, \mathcal{R}) \le \varepsilon_{|V(G_n)|} \to 0$, it follows that $\delta_\square(W, \mathcal{R}) = 0$. Since $\mathcal{R}$ is closed, this implies that $W \in \mathcal{R}$.

To show that $\mathcal{R} \subseteq \overline{\mathcal{P}}$, consider any $U \in \mathcal{R}$, and let $\varepsilon > 0$. As we have seen above, there is a graph $G \in \mathcal{P}$ with n_ε nodes such that $\delta_\square(W_G, U) \le \varepsilon$ and $\delta_1(W_G, \mathcal{R}) \le \varepsilon \le \varepsilon_{n_\varepsilon}$. So $G \in \mathcal{P}$, which shows that there is a function W_G with $G \in \mathcal{P}$ arbitrarily close to U.

To show that property $\mathcal{P}$ is robust, consider any graph G with n nodes. Choose $U \in \mathcal{R}$ so that $\|W_G - U\|_1 \le 2d_1(W_G, \mathcal{R})$. By Corollary 2.7, there is a graph $\widetilde{G}$ on n nodes such that $d_1(\widetilde{G}, U) \le 4\|W_G - U\|_1 \le 8d_1(W_G, \mathcal{R})$ and $\|W_{\widetilde{G}} - U\|_\square \le \frac{50}{\sqrt{\log\log n}} \le \varepsilon_n$. Thus $\widetilde{G} \in \mathcal{P}$, and

$$\begin{aligned} d_1(G, \mathcal{P}) &\le d_1(G, \widetilde{G}) = \|W_G - W_{\widetilde{G}}\|_1 \le \|W_G - U\|_1 + \|\widetilde{G} - U\|_1 \\ &\le 2d_1(W_G, \mathcal{R}) + 8d_1(W_G, \mathcal{R}) = 10d_1(W_G, \mathcal{R}). \end{aligned}$$

Thus $\mathcal{P}$ is robust and $\overline{\mathcal{P}} = \mathcal{R}$ is testable. Theorem 3.18 implies that $\mathcal{P}$ is testable. $\square$

Let us say that two graph properties $\mathcal{P}$ and $\mathcal{P}'$ are *equivalent* if their closure is the same. Theorem 3.19 implies that equivalence classes of testable graph properties are in a one to one correspondence with the testable graphon properties.

As an application of these results, we give a purely combinatorial characterization of testability, which generalizes the result of Alon and Shapira on the testability of hereditary properties, and also contains a finite analogue of Theorem 3.5.

Theorem 3.20 *For a graph property $\mathcal{P}$, the following are equivalent:*

(a) *$\mathcal{P}$ is testable;*

(b) *For every $\varepsilon > 0$ there is an $\varepsilon' > 0$ and an $n' > 0$ such that if $G \in \mathcal{P}$, G' is any other graph such that $\delta_\square(G, G') < \varepsilon'$, and $|V(G)|, |V(G')| \geq n'$, then $d_1(G', \mathcal{P}) < \varepsilon$.*

(c) *For every $\varepsilon > 0$ there is an $\varepsilon_0 > 0$ and an $n_0 > 0$ such that if $G \in \mathcal{P}$ and G' is an induced subgraph of G such that $\delta_\square(G, G') < \varepsilon_0$ and $|V(G')| \geq n_0$, then $d_1(G', \mathcal{P}) < \varepsilon$.*

Remark 3.21 Condition (b) says that, roughly speaking, if a graph G is close to a graph $H \in \mathcal{P}$ in the $\delta_\square$ distance, then it is also close to a (possibly different) graph $J \in \mathcal{P}$ in edit distance. But we need to be careful: Let $\mathcal{P}$ be the (trivial) graph property of having at most 1 node, then a large edgeless graph will be close to $\mathcal{P}$ in the $\delta_\square$ distance, but not in d_1. The theorem shows that it is enough to add the assumption that the graphs are large enough.

Condition (c) is a weakening of Alon–Shapira condition that the graph is hereditary. Theorem 2.11 in [7] implies that a randomly chosen k-node induced subgraph of G is closer to G than $10/\sqrt{\log k}$ in the $\delta_\square$ distance, with large probability. So we can think of induced subgraphs G' satisfying $\delta_\square(G, G') < \varepsilon_0$ as "typical".

Proof. Lemma 3.11 says that (a)⇒(b), and (b)⇒(c) is trivial. To prove that (c)⇒(a), we start with proving a version of the condition in the theorem for functions.

Claim 3.21.1 *For every $\varepsilon > 0$ there is an $\varepsilon_1 > 0$ and an $n_1 > 0$ such that if $W \in \overline{\mathcal{P}}$ and G is graph such that $|V(G)| \geq n_1$, $t_{\mathrm{ind}}(G, W) > 0$ and $d_\square(W_G, W) < \varepsilon_1$, then $d_1(G, \mathcal{P}) < \varepsilon$.*

Given $\varepsilon > 0$, let ε_0 and n_0 be as in the condition of the theorem, and set $\varepsilon_1 = \varepsilon_0/2$, $n_1 = n_0$. Let $H_n \in \mathcal{P}$ be a sequence of graphs such that $H_n \to W$. Then $t_{\mathrm{ind}}(G, H_n) \to t_{\mathrm{ind}}(G, W) > 0$, so for large enough n, G is an induced subgraph of H_n. Furthermore, we have $\delta_\square(G, H_n) \leq \delta_\square(W_G, W) + \delta_\square(W, W_{H_n}) \to \delta_\square(W_G, W) \leq \varepsilon_0/2$, and so for large enough n we have $\delta_\square(G, H_n) < \varepsilon_0$. So the condition of the theorem implies that $d_1(G, \mathcal{P}) < \varepsilon$. This proves Claim 3.21.1.

To prove that $\mathcal{P}$ is testable, we use Theorem 3.18: it suffices to prove that $\overline{\mathcal{P}}$ is testable and $\mathcal{P}$ is robust. To prove that $\overline{\mathcal{P}}$ is testable, we use Theorem 3.4: We want to prove that if a function is close to $\overline{\mathcal{P}}$ in the $\|.\|_\square$ norm, then it is also close in the $\|.\|_1$ norm. Our next step is proving a special case of this.

Claim 3.21.2 *For every $\varepsilon > 0$ there is an $\varepsilon_2 > 0$ such that if $W \in \overline{\mathcal{P}}$ and $U \in \mathcal{W}$ is a function such that $U = W$ wherever $W(x, y) \in \{0, 1\}$ and $d_\square(W, U) < \varepsilon_2$, then $d_1(U, \overline{\mathcal{P}}) < \varepsilon$.*

Let ε_1 and n_1 be as in Claim 3.21.1, and set $\varepsilon_2 = \varepsilon_1/2$. Choose a sequence G_n of graphs such that $G_n \to U$ and $t_{\mathrm{ind}}(G_n, U) > 0$. (For example, random graphs $G_n = \mathbb{G}(n, U)$ have this property with large probability.) By the condition on U, we have $t_{\rm ind}(G_n, W) > 0$. Furthermore, $\delta_\square(W_{G_n}, W) \le \delta_\square(W_{G_n}, U) + \delta_\square(U, W) \to \delta_\square(U, W) < \varepsilon_1/2$, and hence for large enough n, we have $\delta_\square(W_{G_n}, W) < \varepsilon_1$. If we also choose n large enough so that $|V(G_n)| \ge n_1$, then $d_1(G_n, \mathcal{P}) < \varepsilon$ by Claim 3.21.1. Thus there exist graphs $H_n \in \mathcal{P}$ with $V(H_n) = V(G_n)$ such that $d_1(G_n, H_n) \le \varepsilon$. We choose a subsequence such that $H_n \to W'$ for some function W', which is then in $\overline{\mathcal{P}}$. Now Lemma 2.11 implies that $\delta_1(U, W') < \varepsilon$, and hence $d_1(U, \overline{\mathcal{P}}) < \varepsilon$.

Claim 3.21.3 *$\overline{\mathcal{P}}$ is testable.*

We use Theorem 3.4. Let (U_n) be a sequence of functions in $\mathcal{W}_0$ such that $d_\square(U_n, \overline{\mathcal{P}}) \to 0$, we want to prove that $d_1(U_n, \overline{\mathcal{P}}) \to 0$. We may assume that (U_n) is convergent in the δ_1 distance, and hence applying appropriate measure preserving transformations, we may assume that there is a function $W \in \mathcal{W}_0$ such that $\|U_n - W\|_\square \to 0$. Let

$$U'_n(x,y) = \begin{cases} W(x,y), & \text{if } W(x,y) \in \{0,1\}, \\ U_n(x,y), & \text{otherwise.} \end{cases}$$

Let $\varepsilon > 0$ and choose n_2, ε_2 as in Claim 3.21.2, with input $\varepsilon/2$. Just as in the proof of Theorem 3.30, we see that $\|U_n - U'_n\|_1 \to 0$, so we can choose a large enough n such that $\|U_n - U'_n\|_1 < \min\{\varepsilon/2, \varepsilon_2/2\}$ and also $\|U_n - W\|_\square < \varepsilon_2/2$. Then $U'_n = W$ wherever $W \in \{0,1\}$ and $\|U'_n - W\|_\square \le \|U'_n - U_n\|_\square + \|U_n - W\|_\square \le \|U'_n - U_n\|_1 + \|U_n - W\|_\square < \varepsilon_2$, and so by Claim 3.21.2 we have $\|U'_n - W\|_1 < \varepsilon/2$. Thus $\|U_n - W\|_1 \le \|U'_n - U_n\|_1 + \|U_n - W\|_1 \le \varepsilon$.

Claim 3.21.4 *$\mathcal{P}$ is robust.*

Let $\varepsilon > 0$, let ε_1 and n_1 be chosen so that they satisfy Claim 3.21.1 with input $\varepsilon/2$, and let $\varepsilon' = \min(\varepsilon/4, \varepsilon_1/3)$. Let G be a graph with $|V(G)| = m \ge n_1$ and $d_1(W_G, \overline{\mathcal{P}}) < \varepsilon'$. Let $W \in \overline{\mathcal{P}}$ be a function such that $\|W_G - W\|_1 < \varepsilon'$. Consider the random graph $G' = \mathbb{G}'(n, W)$. Then $t_{\rm ind}(G', W) > 0$ with probability 1, and by Lemma 2.5, with probability tending to 1, $\|W_{G'} - W\|_\square \le \varepsilon'$ if n is large enough. Furthermore, Lemma 2.6 implies that $\mathsf{E}(d_1(G, G')) = \|W_G - W\|_1 < \varepsilon'$, and so with probability at least 1/2, we have $d_1(G, G') < 2\varepsilon' < \varepsilon/2$. In such a case

$$\|W_G - W\|_\square \le \|W_G - W_{G'}\|_\square + \|W_{G'} - W\|_\square \le \|W_G - W_{G'}\|_1 + \|W_{G'} - W\|_\square < 3\varepsilon' = \varepsilon_1,$$

and so by Claim 3.21.1, we have $d_1(G', \mathcal{P}) < \varepsilon/2$. Hence $d_1(G, \mathcal{P}) \le d_1(G, G') + d_1(G', \mathcal{P}) < \varepsilon$, which proves that $\mathcal{P}$ is robust.

Using Theorem 3.18, this completes the proof of Theorem 3.20. $\square$

3.3 Property testing vs. parameter testing

Parameter testing is a problem related to property testing, and in some respects simpler; but the main facts are often analogous. It was introduced in [7], where a number of different characterizations were also given (see also [5]). We summarize the main results about parameter testing, to point out this analogy; finally, we prove a direct connection between these notions.

A *graph parameter* is a function defined on isomorphism types of graphs. A graph parameter f is *testable*, if for every $\varepsilon > 0$ there is a positive integer $k = k(\varepsilon)$ such that if G is a graph with at least k nodes and S is a random subset of k nodes of G (chosen uniformly over all k-element sets), then

$$\mathsf{P}(|f(G) - f(G[S])| > \varepsilon) < 1/3.$$

Testability of parameters is related to the convergence of graph sequences:

Proposition 3.22 ([7]) *A graph parameter f is testable if and only if $f(G_n)$ converges for every convergent graph sequence (G_n).*

A *graphon functional* is a real valued function defined on graphons; equivalently, a functional defined on $\mathcal{W}_0$ that is invariant under isomorphism. A graphon functional f is *testable*, if there is a graph parameter g such that for every $\varepsilon > 0$ there is a positive integer $k(\varepsilon)$ such that for every $W \in \mathcal{W}_0$ and $k \geq k(\varepsilon)$,

$$\mathsf{P}(|f(W) - g(\mathbb{G}(k, W))| > \varepsilon) < 1/3.$$

Informally, we can estimate the value of $f(W)$ by generating a W-random graph with sufficiently many nodes, and evaluating the graph parameter g on this.

Testable graphon parameters have a simple characterization:

Proposition 3.23 *A graphon parameter is testable if and only if it is continuous in the norm $\|.\|_\square$.*

Proof. Suppose that f is testable. If $\|U - W\|_\square$ is small enough, then the distributions of $\mathbb{G}(k, W)$ and $\mathbb{G}(k, U)$ are close, and so there is a graph F such that $|f(W) - g(F)| \leq \varepsilon$ and $|f(U) - g(F)| \leq \varepsilon$, so $|f(W) - f(U)| \leq 2\varepsilon$. So f is continuous.

Conversely, if f is continuous, then it is uniformly continuous by the compactness of $\mathcal{W}$, and so for every $\varepsilon > 0$ there is an $\varepsilon' > 0$ such that if $\delta_\square(U, W) \leq \varepsilon'$ then $|f(U) - f(W)| \leq \varepsilon'$. Define $g(F) = f(W_F)$ for a graph F. By Theorem 2.4, if k is large enough, then with large probability,

$$\delta_\square(W_{\mathbb{G}(k,W)}, W) \leq \varepsilon',$$

and hence with large probability,

$$|g(\mathbb{G}(k, W)) - f(W)| = |f(W_{\mathbb{G}(k,W)}) - f(W)| \leq \varepsilon.$$

This proves that f is testable. $\square$

Theorem 5.1 in [7] implies the following.

Proposition 3.24 *A graph parameter is testable if and only if there is a testable graphon parameter f such that $f(W_G) - g(G) \to 0$ if $|V(G)| \to \infty$.*

We can view this last fact as follows: Every graphon functional f gives rise to a graph parameter $\widehat{f}$ by $\widehat{f}(G) = f(W_G)$. This graph parameter $\widehat{f}$ is testable if f is testable, and testable graph parameters are exactly those that are "asymptotically equal" to $\widehat{f}$ for some testable graphon functional f.

Alon and Shapira [2] proved that the edit distance from every hereditary graph property is a testable parameter. More generally, Fischer and Newman [8] proved:

Theorem 3.25 (Fischer–Newman) *A graph property $\mathcal{P}$ is testable if and only if the edit distance $d_1(G, \mathcal{P})$ is a testable parameter.*

The following theorem gives an analogue of this result for graphons, from which Theorem 3.25 can be deduced:

Theorem 3.26 *A graphon property $\mathcal{R}$ is testable if and only if the distance $d_1(., \mathcal{R})$ is a testable functional.*

The content of the theorem is that the d_1 distance from a testable property is continuous function in the $\|.\|_\square$ norm. It is trivial that this distance is continuous in the $\|.\|_1$ norm for any graphon property. The second half of the proof below shows that the d_1 distance from *any* graphon property is lower semi-continuous in the $\|.\|_\square$ norm.

Proof. It is easy to see that if the functional $d_1(., \mathcal{R})$ is testable then $\mathcal{R}$ is testable: the set $\{W : d_1(W, \mathcal{R}) < \varepsilon\}$ is open in the $\|.\|_\square$ norm, and contains the compact set $\overline{\mathcal{R}}$, so it contains a neighborhood $B_\square(\mathcal{R}, \varepsilon')$ of $\mathcal{R}$ for some $\varepsilon' > 0$.

Now suppose that $\mathcal{R}$ is testable. Let $W \in \mathcal{W}_0$ and let $W_n \to W$. We claim that $d_1(W_n, \mathcal{R}) \to d_1(W, \mathcal{R})$. We may assume that $\|W_n - W\|_\square \to 0$.

Let $\varepsilon > 0$, and let $U \in \mathcal{R}$ be such that $\|W - U\|_1 \le d_1(W, \mathcal{R}) + \varepsilon$. By Lemma 2.10, there is a sequence of functions $U_n \in \mathcal{W}$ such that $\|U_n - U\|_\square \to 0$ and $\|U_n - W_n\|_1 \to \|U - W\|_1$. By the testability of $\mathcal{R}$ and by Theorem 3.4, it follows that $\|U_n - U\|_1 \to 0$, and so

$$\|W_n - U\|_1 \le \|W_n - U_n\|_1 + \|U_n - U\|_1 \to \|U - W\|_1.$$

Hence

$$\limsup_{n\to\infty} d_1(W_n, \mathcal{R}) \le \limsup_{n\to\infty} \|W_n - U\|_1 \le \|U - W\|_1 \le d_1(W, \mathcal{R}) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this implies that

$$\limsup_{n\to\infty} d_1(W_n, \mathcal{R}) \le d_1(W, \mathcal{R}) \tag{24}$$

To prove the reverse, let $V_n \in \mathcal{R}$ be chosen so that $\|W_n - V_n\|_1 \le d_1(W_n, \mathcal{R}) + 1/n$. By selecting a subsequence, we may assume that the sequence (V_n) is convergent in the $\delta_\square$ distance. Let $V \in \mathcal{W}_0$ be its limit. Clearly $V \in \mathcal{R}$. Thus by Lemma 2.11 we have

$$d_1(W, \mathcal{R}) \le \delta_1(W, V) \le \liminf_{n\to\infty} \delta_1(W_n, V_n) \le \liminf_{n\to\infty} d_1(W_n, \mathcal{R}). \tag{25}$$

The relations (24) and (25) prove that $d_1(W_n, \mathcal{R}) \to d_1(W, \mathcal{R})$, and so $d_1(., \mathcal{R})$ is continuous. $\square$

The theorem of Fischer and Newman (Theorem 3.25) is easy to derive from here. By the results of [7] Theorem 6.1(d), it suffices to prove that for every testable property $\mathcal{P}$, $d_1(W,\overline{\mathcal{P}})$ is a continuous function of W in the $\|.\|_\square$ norm, and $d_1(G,\mathcal{P}) - d_1(W_G,\overline{\mathcal{P}}) \to 0$ if $|V(G)| \to \infty$. The first assertion follows from Theorem 3.26, the second, from Proposition 3.13(b).

3.4 Flexible properties

Let $W \in \mathcal{W}_0$. A function $U \in \mathcal{W}_0$ is a *flexing* of W if $U(x,y) = W(x,y)$ for all x,y with $W(x,y) \in \{0,1\}$ (so we may change the values of W that are strictly between 0 and 1; note that we may change these to 0 or 1, so the relation is not symmetric). We say that a function property is *flexible* if it is preserved under flexing. The following Proposition gives some sufficient conditions for flexibility; the proofs are straightforward and omitted.

Proposition 3.27 (a) *Every function property that implies that the function is* $0-1$ *valued is flexible.*

(b) *For every graph* F*, the function property* $\{W \in \mathcal{W}_0 :\ t(F,W) = 0\}$ *is flexible.*

(c) *If* $\mathcal{R}$ *is a flexible function property, then the function properties* $\mathcal{R}^* = \{1-W :\ W \in \mathcal{R}\}$, $\mathcal{R}^\uparrow$ *and* $\mathcal{R}^\downarrow$ *are flexible.*

(d) *The intersection and union of any set of flexible properties is flexible.*

We call a graph property *flexible* if its closure is flexible.

Example 3.28 (a) The graph property that G is clique of size $\lceil |V(G)|/2\rceil$, together with isolated nodes, is flexible, since its closure consists of a single graphon (represented by the function W that is 1 if $x,y \le 1/2$ and 0 otherwise).

(b) The graph property that $\omega(G) \ge |V(G)|/2$ is flexible (where $\omega(G)$ is the size of the largest clique in G), since it is the upward closure of the property in (a). Similarly, the property that $\alpha(G) \ge |V(G)|/2$ is flexible (where $\alpha(G)$ is the size of the largest independent set in G).

(c) The graph property that there is a labeling of the nodes by $\{1,\dots,n\}$ such that two nodes are connected if and only if their labels sum to at most n, is flexible. Indeed, the closure of this graph property consists of a single graphon (represented by the function W that is 1 if $x+y \le 1$ and 0 otherwise). This closure is flexible by Proposition 3.27(a).

(d) The graph property that there is a labeling of the nodes by $\{1,\dots,n\}$ such that all pairs whose labels sum to at most n are connected by an edge, is flexible. Indeed, this is the upward closure of the property in (c).

These and other examples follow from the following proposition:

Proposition 3.29 (a) *Every graph property* $\mathcal{P}$ *for which* $\overline{\mathcal{P}}$ *consists of* $0-1$ *valued functions is flexible.*

(b) *If a graph property is flexible, then so are its upward and downward closures.*

(c) *If a graph property* $\mathcal{P}$ *is flexible, then so is the property obtained by complementing all graphs in* $\mathcal{P}$*.*

(d) *Every hereditary graph property is flexible.*

Proof. (a) is obvious. (b) follows from Theorem 3.15 and Proposition 3.27 (c). Assertion (c) follows from Proposition 3.27 (c). Finally, (d) follows from 3.9 (a), since if $t(F,W)=0$, then $t(F,U)=0$ for every flexing U of W. $\square$

Our main result about flexible properties is the following.

Theorem 3.30 (a) *Every closed flexible graphon property is testable.*

(b) *Every robust flexible graph property is testable.*

Remark 3.31 1. It is important to assume that the property is closed. We have seen that the function property of being 0-1 valued is flexible and it is trivially a graphon property, but it is not testable.

2. We can weaken this notion slightly in a way that we preserve testability. We say that a function property is *weakly flexible* if it is closed under those flexings that do not change the integral of the function. A good example for this is the property which consists of those functions whose integral is 1/2. One can modify the proof of Theorem 3.30 to show that every weakly flexible function property is testable.

Proof. The second assertion is an immediate consequence of the first, so we only prove (a).

Assume that $d_\square(W_n,\mathcal{R})\to 0$ but $d_1(W_n,\mathcal{R})\geq\varepsilon$ for some fixed $\varepsilon>0$. We can assume that W_n converges to some $W\in\mathcal{R}$ in the $\|.\|_\square$ norm. Let $S_0=W^{-1}(0)$, $S_1=W^{-1}(1)$ and let $Z_n\in\mathcal{W}_0$ denote the function which is 1 on S_1, 0 on S_0 and is identical with W_n anywhere else. By flexibility, we have $Z_n\in\mathcal{R}$. By Lemma 2.2,

$$\lim_{n\to\infty}\|W_n-Z_n\|_1=\lim_{n\to\infty}\Big(\int_{S_0}W_n+\int_{S_1}(1-W_n)\Big)=\int_{S_0}W+\int_{S_1}(1-W)=0,$$

which is a contradiction. $\square$

The theorem of Alon and Shapira [2] follows easily. Let $\mathcal{P}$ be a hereditary graph property. By Proposition 3.29 $\mathcal{P}$ is flexible, and so by Theorem 3.30, its closure is testable. So it suffices to prove that $\mathcal{P}$ is robust, which follows from Proposition 3.13(b).

Let $U,W\in\mathcal{W}_0$ and consider a convex combination $Z=\alpha U+(1-\alpha)W$, $0<\alpha<1$. Then both U and W are flexings of Z. This implies the following very useful observation:

Proposition 3.32 *If $\mathcal{R}$ is a flexible function property, then $\mathcal{W}_0\setminus\mathcal{R}$ is a convex set.* $\square$

Corollary 3.33 *The distance $d_1(U,\mathcal{R})$ from a flexible function property is a concave function of U.* $\square$

Alon and Stav [3] proved the surprising fact that for every hereditary graph property $\mathcal{P}$ there is a number p, $0\leq p\leq 1$, such that the random graph $\mathbb{G}(n,p)$ is, asymptotically, at maximum edit distance from the property:

Theorem 3.34 (Alon and Stav) *For every hereditary graph property $\mathcal{P}$ there is a number p, $0 \le p \le 1$, such that for every graph G with $|V(G)| = n$,*

$$d_1(G, \mathcal{P}) \le \mathsf{E}(d_1(\mathbb{G}(n,p), \mathcal{P})) + o(1) \qquad (n \to \infty).$$

The following theorem states a functional version and a generalization of this fact.

Theorem 3.35 (a) *For every flexible graphon property $\mathcal{R}$, the maximum of $d_1(., \mathcal{R})$ is attained by a constant function.*

(b) *For every flexible and robust graph property $\mathcal{P}$ there is a number p, $0 \le p \le 1$, such that for every graph G with $|V(G)| = n$,*

$$d_1(G, \mathcal{P}) \le \mathsf{E}(d_1(\mathbb{G}(n,p), \mathcal{P})) + o(1) \qquad (n \to \infty).$$

By Proposition 3.29(d), Theorem 3.34 is a consequence. A further corollary is that the conclusion of Theorem 3.35 holds for the properties in Example 3.28(a),(b).

Proof. Part (a) of the theorem is an easy consequence of Proposition 3.32 and Corollary 3.33. (The proof in fact works for any norm on $\mathcal{W}$ instead of the norm $\|.\|_1$.)

To prove part (b), let $p \in [0,1]$ maximize $d_1(p, \overline{\mathcal{P}})$ (where p also denotes the constant p function), and let b denote the maximum value. It suffices to prove that

$$\mathsf{E}(d_1(\mathbb{G}(n,p), \mathcal{P})) \ge b + o(1) \qquad (n \to \infty), \tag{26}$$

and

$$d_1(G, \mathcal{P}) \le b + o(1) \qquad (|V(G)| \to \infty). \tag{27}$$

The first bound is quite easy: We can fix a choice G_n for the random graphs so that $d_1(G_n, \mathcal{P}) \le \mathsf{E}(d_1(\mathbb{G}(n,p), \mathcal{P})) + o(1)$ and $G_n \to p$. Then by Proposition 3.14, we have

$$\liminf d_1(G_n, \mathcal{P}) \ge d_1(p, \overline{\mathcal{P}}) = b.$$

Suppose that (27) fails. Then there exists a sequence of graphs G_n with $|V(G_n)| \to \infty$ such that

$$d_1(G_n, \mathcal{P}) \to b' > b.$$

We may assume that $G_n \to W \in \mathcal{W}_0$. By part (a),

$$d_1(W, \overline{\mathcal{P}}) \le d_1(p, \overline{\mathcal{P}}).$$

Since the property $\mathcal{P}$ is robust and flexible, it is testable, and hence by Proposition 3.13(b), we have $d_1(G_n, \mathcal{P}) - d_1(W_{G_n}, \overline{\mathcal{P}}) \to 0$. Thus $d_1(W_{G_n}, \overline{\mathcal{P}}) \to b'$. Theorem 3.26 then implies that $d_1(W, \overline{\mathcal{P}}) = b'$, a contradiction. $\square$

4 Concluding remarks

We have mentioned after the definition of testable properties that some modifications in the definition do not change the notion of testability. Let us discuss some other possible modifications that would lead to a different, generally less interesting notion.

Examples 3.8(b) and (c) suggest that in the definition of of the edit distance, we could allow adding or removing nodes as well as adding or removing edges. This would of course change which graph properties are testable, but would not change the closure of testable properties, due to Theorem 3.20.

Example 3.8(d) is counterintuitive again, since the densities in the definition can be estimated from samples easily. The trouble is that a small error in these densities only implies that the graph is close to a quasirandom graph in the $d_\square$ distance, not in d_1.

Considering how useful the cut distance has been, why don't we measure the error in cut distance? The answer is that this would lead to a trivial notion: *Every graphon property is testable if we measure the error in the distance* $d_\square$. Indeed, given a graphon property $\mathcal{R}$, we can define $\mathcal{R}'$ as the property of the graph G that $d_\square(W_G, \mathcal{R}) \le 10/\sqrt{\log |V(G)|}$. If $W \in \mathcal{R}$, then $G = \mathbb{G}(n, W)$ satisfies $\delta_\square(W_G, W) \le 10/\sqrt{\log |V(G)|}$ with large probability, and in such cases $G \in \mathcal{R}'$. Conversely, if $d_\square(W, \mathcal{R}) > \varepsilon$, then for $G = \mathbb{G}(k, W)$ with $k > 2^{8/\varepsilon^2}$ we have $\delta_\square(W_G, W) < \varepsilon/2$ with large probability; if this happens, then we must have $d_\square(W_G, \mathcal{R}) > \varepsilon/2$ (else it would follow that $d_\square(W, \mathcal{R}) < \varepsilon$), and so if k is large enough, we have $G \in \mathcal{R}'$.

We get a different and potentially interesting notion of testing function properties if instead of $\mathbb{G}(n, W)$, we consider the edge-weighted graph $\mathbb{H}(n, W)$ in which we select n random points $X_1, \dots, X_n$ from $[0, 1]$, but instead of randomizing a second time to get the edges, we keep $W(X_i, X_j)$ as the weight of the edge ij. The singleton set $\mathcal{R} = \{U \equiv 1/2\}$ would become testable in this notion. We don't have a characterization of testability in this sense.

Acknowledgement

We are indebted to the anonymous referee for pointing out several errors and inconsistencies, and for suggesting many improvements to the presentation.

References

[1] N. Alon, E. Fischer, I. Newman and A. Shapira: A Combinatorial Characterization of the Testable Graph Properties: It's All About Regularity, *Proc. of the 38th ACM Symp. Theor. of Comp. (STOC)* (2006) 251–260.

[2] N. Alon and A. Shapira: A Characterization of the (natural) Graph Properties Testable with One-Sided Error, Proc. 46-th IEEE FOCS (2005), 429–438.

[3] N. Alon and U. Stav: What is the furthest graph from a hereditary property? *Random Structures and Algorithms* (to appear).

[4] C. Borgs, J. Chayes, L. Lovász: Moments of Two-Variable Functions and the Uniqueness of Graph Limits, http://www.cs.elte.hu/ lovasz/limitunique.pdf

[5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, B. Szegedy and K. Vesztergombi: Graph limits and parameter testing, *Proc. of the 38th ACM Symp. Theory of Comp.* (2006), 261–270.

[6] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi: Counting graph homomorphisms, in: *Topics in Discrete Mathematics* (ed. M. Klazar, J. Kratochvil, M. Loebl, J. Matoušek, R. Thomas, P. Valtr), Springer (2006), 315–371.

[7] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing (submitted).

[8] Eldar Fischer, Ilan Newman: Testing versus Estimation of Graph Properties, Proc. 37-th ACM STOC (2005), 138–146.

[9] O. Goldreich, S. Goldwasser and D. Ron: Property testing and its connection to learning and approximation, *J. ACM* **45** (1998), 653–750.

[10] L. Lovász and B. Szegedy: Limits of dense graph sequences, *J. Comb. Theory B* **96** (2006), 933–957.

[11] L. Lovász and B. Szegedy: Szemerédi's Lemma for the analyst, *Geom. Func. Anal.* **17** (2007), 252–270.

[12] L. Lovász and B. Szegedy: The moment problem for 2-variable functions and reflection positive graph parameters (manuscript) `http://www.cs.elte.hu/~lovasz/moment.pdf`