

Centrality measures for graphons: Accounting for uncertainty in networks

Marco Avella-Medina, Francesca Parise, Michael T. Schaub, and Santiago Segarra, *Member, IEEE*

Abstract—As relational datasets modeled as graphs keep increasing in size and their data-acquisition is permeated by uncertainty, graph-based analysis techniques can become computationally and conceptually challenging. In particular, node centrality measures rely on the assumption that the graph is perfectly known — a premise not necessarily fulfilled for large, uncertain networks. Accordingly, centrality measures may fail to faithfully extract the importance of nodes in the presence of uncertainty. To mitigate these problems, we suggest a statistical approach based on graphon theory: we introduce formal definitions of centrality measures for graphons and establish their connections to classical graph centrality measures. A key advantage of this approach is that centrality measures defined at the modeling level of graphons are inherently robust to stochastic variations of specific graph realizations. Using the theory of linear integral operators, we define degree, eigenvector, Katz and PageRank centrality functions for graphons and establish concentration inequalities demonstrating that graphon centrality functions arise naturally as limits of their counterparts defined on sequences of graphs of increasing size. The same concentration inequalities also provide high-probability bounds between the graphon centrality functions and the centrality measures on any sampled graph, thereby establishing a measure of uncertainty of the measured centrality score.

Index Terms—Random graph theory, Networks, Graphons, Centrality measures, Stochastic block model

1 INTRODUCTION

MANY biological [1], social [2], and economic [3] systems can be better understood when interpreted as networks, comprising a large number of individual components that interact with each other to generate a global behavior. These networks can be aptly formalized by graphs, in which nodes denote individual entities, and edges represent pairwise interactions between those nodes. Consequently, a surge of studies concerning the modeling, analysis, and design of networks have appeared in the literature, using graphs as modeling devices.

A fundamental task in network analysis is to identify salient features in the underlying system, such as key nodes or agents in the network. To identify such important agents, researchers have developed *centrality measures* in various contexts [4], [5], [6], [7], [8], each of them capturing different aspects of node importance. Prominent examples for the utility of centrality measures include the celebrated PageRank algorithm [9], [10], employed in the search of relevant sites on the web, as well as the identification of influential agents in social networks to facilitate viral marketing campaigns [11].

A crucial assumption for the applicability of these centrality measures is that the observation of the underlying network is complete and noise free. However, for many sys-

tems we might be unable to extract a complete and accurate graph-based representation, e.g., due to computational or measurement constraints, errors in the observed data, or because the network itself might be changing over time. For such reasons, some recent approaches have considered the issue of robustness of centrality measures [12], [13], [14], [15] and general network features [16], as well as their computation in dynamic graphs [17], [18], [19]. The closest work to ours is [20], where convergence results are derived for eigenvector and Katz centralities in the context of random graphs generated from a stochastic block model.

As the size of the analyzed systems continues to grow, traditional tools for network analysis have been pushed to their limit. For example, systems such as the world wide web, the brain, or social networks can consist of billions of interconnected agents, leading to computational challenges and the irremediable emergence of uncertainty in the observations. In this context, *graphons* have been suggested as an alternative framework to analyze large networks [21], [22]. While graphons have been initially studied as limiting objects of large graphs [23], [24], [25], they also provide a rich non-parametric modeling tool for networks of any size [26], [27], [28], [29]. In particular, graphons encapsulate a broad class of network models including the stochastic block model [30], [31], random dot-product graphs [32], the infinite relational model [33], and others [34]. A testament of the practicality of graphons is their use in applied disciplines such as signal processing [35], collaborative learning [36], and control [37].

In this work we aim at harnessing the additional flexibility provided by the graphon framework to suggest a statistical approach to agents' centralities that inherently accounts for network uncertainty, as detailed next.

Authors are ordered alphabetically. All authors contributed equally. M. Avella-Medina is with the Department of Statistics, Columbia University. F. Parise is with the Laboratory for Information and Decision Systems, MIT. M. Schaub is with the Institute for Data, Systems, and Society, MIT and with the Department of Engineering Science, University of Oxford, UK. S. Segarra is with the Department of Electrical and Computer Engineering, Rice University. Emails: marco.avella@columbia.edu, {parise, mschaub}@mit.edu, segarra@rice.edu. Funding: This work was supported by the Swiss National Science Foundation grants P2EGP1_168962, P300P1_177746 (M. Avella-Medina) and P2EZZP2_168812, P300P2_177805 (F. Parise); the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 702410 (M. Schaub); the Spanish MINECO TEC2013-41604-R and the MIT IDSS Seed Fund Program (S. Segarra).

1.1 Motivation

Most existing applications of network centrality measures follow the paradigm in Fig. 1a: a specific graph — such as a social network with friendship connections — is observed, and conclusions about the importance of each agent are then drawn based on this graph, e.g., which individuals have more friends or which have the most influential connections. Mathematically, these notions of importance are encapsulated in a centrality measure that ranks the nodes according to the observed network structure. For instance, the idea that importance derives from having the most friends is captured by degree centrality. Since the centrality of any node is computed solely from the network structure [4], [5], [6], [7], [8], a crucial assumption hidden in this analysis is that the empirically observed network captures all the data we care about.

However, in many instances in which centrality measures are employed, this assumption is arguably not fulfilled: we typically do not observe the complete network at once. Further, even those parts we observe contain measurement errors, such as false positive or false negative links, and other forms of uncertainty. The key question is therefore how to identify crucial nodes via network-based centrality measures without having access to an accurate depiction of the ‘true’ latent network.

One answer to this problem is to adopt a statistical inference-based viewpoint towards centrality measures, by assuming that the observed graph is a specific realization of an underlying stochastic generative process; see Fig. 1b. In this work, in particular, we use graphons to model such underlying generative process, because they provide a rich non-parametric statistical framework. Further, it has been recently shown that graphons can be efficiently estimated from one (or multiple) noisy graph observations [28], [38]. Our main contribution is to show that, based on the inferred graphon, one can compute a latent centrality profile of the nodes that we term *graphon centrality function*. This graphon centrality may be seen as a fundamental measure of node importance, irrespective of the specific realization of the graph at hand. This leads to a robust estimate of the centrality profiles of all nodes in the network. In fact, we provide high-probability bounds between the distance of such latent graphon centrality functions and the centrality profiles in any realized network, in terms of the network size.

To illustrate the dichotomy of the standard approach towards centrality and the one outlined here, let us consider the graphon in Fig. 1c. Graphons will be formally defined in Section 2.2, but the fundamental feature is that it defines a random graph model from where graphs of any pre-specified size can be obtained. If we generate one of these graphs with 100 nodes, we can apply the procedure in Fig. 1a to obtain a centrality value for each agent, as shown in the red curve in Fig. 1c. In the standard paradigm we would then sort the centrality values to find the most central nodes, which in this case would correspond to the node marked as v_1 in Fig. 1c. On the other hand, if we have access to the generative graphon model (or an estimate thereof), then we can compute the continuous graphon centrality function and compare the deviations from it in the specific graph realization; see blue and red curves in Fig. 1c.

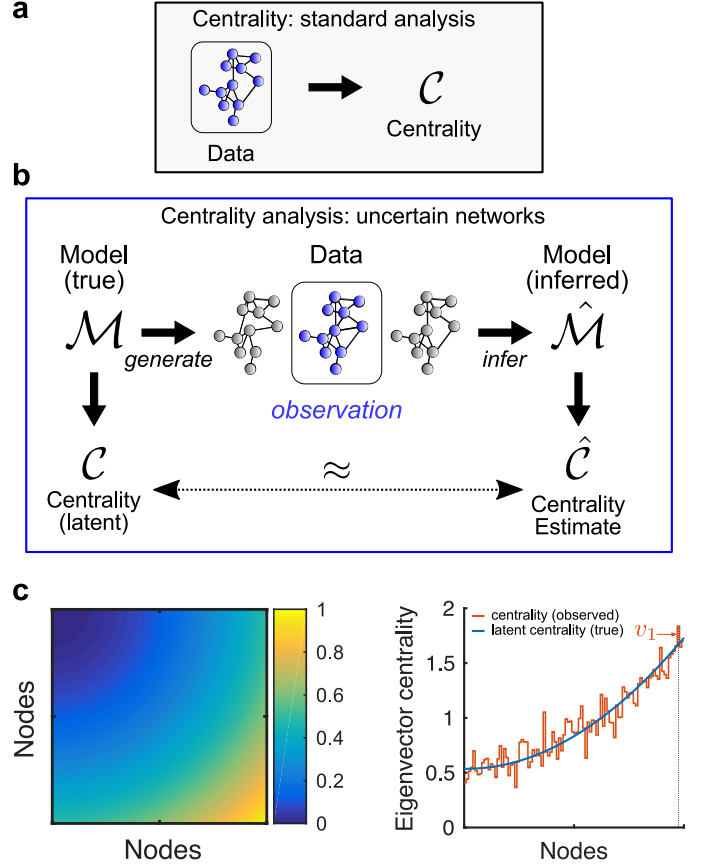


Fig. 1: Schematic — Network centrality analysis. **(a)** Classical centrality analysis computes a centrality measure purely based on the observed network. **(b)** If networks are subject to uncertainty, we may adopt a statistical perspective on centrality, by positing that the observed network is but one realization of a true, unobserved latent model. Inference of the model then would lead to a centrality estimate that accounts for the uncertainty in the data in a well-defined manner. **(c)** Illustrative example. Left: A network of 100 nodes is generated according to a graphon model with a well-defined increasing connectivity pattern. Right: This graphon model defines a latent (expected) centrality for each node (blue curve). The centralities of a single realization of the model (red curve) will in general not be equivalent to the latent centrality, but deviate from it. Estimating the graphon-based centrality thus allows us to decompose the observed centrality into an expected centrality score (blue), and a fluctuation that is due to randomness.

The result is that while v_1 is the most central node in the specific graph realization (see red curve), we would expect it to be less central within the model-based framework since the two nodes to its right have higher latent centrality (see blue curve). Stated differently, in this specific realization, v_1 benefited from the random effects in terms of its centrality. If another random graph is drawn from the same graphon model, the rank of v_1 might change, e.g., node v_1 might become less central. Based on a centrality analysis akin to Fig. 1a, we would conclude that the centrality of this node decreased relative to other agents in the network. However, this difference is exclusively due to random variations and thus not statistically significant. The approach outlined in Fig. 1b and, in particular, centrality measures defined on graphons thus provide a statistical framework to analyze centralities in the presence of uncertainty, shielding us from making the wrong conclusion about the change in centrality of node v_1 if the network is subject to uncertainty.

A prerequisite to apply the perspective outlined above is to have a consistent theory of centrality measures for graphons, with well-defined limiting behaviors and well-understood convergence rates. Such a theory is developed in this paper, as we detail in the next section.

1.2 Contributions and article structure

Our contributions are listed below.

- 1) We develop a theoretical framework and definitions for centrality measures on graphons. Specifically, using the existing spectral theory of linear integral operators, we define the degree, eigenvector, Katz and PageRank centrality functions (see Definition 3).
- 2) We discuss and illustrate three different analytical approaches to compute such centrality functions (see Section 4).
- 3) We derive concentration inequalities showing that our newly defined graphon centrality functions are natural limiting objects of centrality measures for finite graphs. These concentration inequalities improve the current state of the art and constitute the main technical results of this paper (see Theorems 1 and 2).
- 4) We illustrate how such bounds can be used to quantify the distance between the latent graphon centrality function and the centrality measures of finite graphs sampled from the graphon.

The remainder of the article is structured as follows. In Section 2 we review preliminaries regarding graphs, graph centralities, and graphons. Subsequently, in Section 3 we recall the definition of the graphon operator and use it to introduce centrality measures for graphons. Section 4 discusses how centrality measures for graphons can be computed using different strategies, and provides some detailed numerical examples. Thereafter, in Section 5, we derive our main convergence results. Section 6 provides concluding remarks. Appendix A contains proofs omitted in the paper. Appendix B provided in the supplementary material presents some auxiliary results and discussions.

Notation: The entries of a matrix \mathbf{X} and a (column) vector \mathbf{x} are denoted by X_{ij} and x_i , respectively; however, in some cases $[\mathbf{X}]_{ij}$ and $[\mathbf{x}]_i$ are used for clarity. The notation T stands for transpose. $\text{diag}(\mathbf{x})$ is a diagonal matrix whose i th diagonal entry is x_i . $\lceil x \rceil$ denotes the ceiling function that returns the smallest integer larger than or equal to x . Sets are represented by calligraphic capital letters, and $1_{\mathcal{B}}(\cdot)$ denotes the indicator function over the set \mathcal{B} . $\mathbf{0}$, $\mathbf{1}$, \mathbf{e}_i , and \mathbf{I} refer to the all-zero vector, the all-one vector, the i -th canonical basis vector, and the identity matrix, respectively. The symbols \mathbf{v} , φ , and λ are reserved for eigenvectors, eigenfunctions, and eigenvalues, respectively. Additional notation is provided at the beginning of Section 3.

2 PRELIMINARIES

In Section 2.1 we introduce basic graph-theoretic concepts as well as the notion of node centrality measures for finite graphs, emphasizing the four measures studied throughout the paper. A brief introduction to graphons and their relation to random graph models is given in Section 2.2.

2.1 Graphs and centrality measures

An undirected and unweighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set \mathcal{V} of N nodes or vertices and an edge set \mathcal{E} of unordered pairs of elements in \mathcal{V} . An alternative representation of such a graph is through its adjacency matrix $\mathbf{A} \in \{0, 1\}^{N \times N}$, where $A_{ij} = A_{ji} = 1$ if $(i, j) \in \mathcal{E}$ and $A_{ij} = 0$ otherwise. In this paper we consider simple graphs (i.e., without self-loops), so that $A_{ii} = 0$ for all i .

Node centrality is a measure of the importance of a node within a graph. This importance is not based on the intrinsic nature of each node, but rather on the location that the nodes occupy within the graph. More formally, a centrality measure assigns a nonnegative centrality value to every node such that the higher the value, the more central the node is. The centrality ranking imposed on the node set \mathcal{V} is in general more relevant than the absolute centrality values. Here, we focus on four centrality measures, namely, the degree, eigenvector, Katz and PageRank centrality measures overviewed next; see [8] for further details.

Degree centrality is a local measure of the importance of a node within a graph. The degree centrality c_i^d of a node i is given by the number of nodes connected to i , that is,

$$\mathbf{c}^d := \mathbf{A}\mathbf{1}, \quad (1)$$

where the vector \mathbf{c}^d collects the values of c_i^d for all $i \in \mathcal{V}$.

Eigenvector centrality, just as degree centrality, depends on the neighborhood of each node. However, the centrality measure c_i^e of a given node i does not depend only on the number of neighbors, but also on how important those neighbors are. This recursive definition leads to an equation of the form $\mathbf{A}\mathbf{c}^e = \lambda\mathbf{c}^e$, i.e., the vector of centralities \mathbf{c}^e is an eigenvector of \mathbf{A} . Since \mathbf{A} is symmetric, its eigenvalues are real and can be ordered as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$. The eigenvector centrality \mathbf{c}^e is then defined as the principal eigenvector \mathbf{v}_1 , associated with λ_1 :

$$\mathbf{c}^e := \sqrt{N} \mathbf{v}_1. \quad (2)$$

For connected graphs, the Perron-Frobenius theorem guarantees that λ_1 is a simple eigenvalue, and that there is a unique associated (normalized) eigenvector \mathbf{v}_1 with positive real entries. As will become apparent later, the \sqrt{N} normalization introduced in (2) facilitates the comparison of the eigenvector centrality on a graph to the corresponding centrality measure defined on a graphon.

Katz centrality measures the importance of a node based on the number of immediate neighbors in the graph as well as the number of two-hop neighbors, three-hop neighbors, and so on. The effect of nodes further away is discounted at each step by a factor $\alpha > 0$. Accordingly, the vector of centralities is computed as $\mathbf{c}_\alpha^k = \mathbf{1} + (\alpha\mathbf{A})\mathbf{1} + (\alpha\mathbf{A})^2\mathbf{1} + \dots$, where we add the number of k -hop neighbors weighted by α^k . By choosing α such that $0 < \alpha < 1/\lambda_1(\mathbf{A})$, the above series converges and we can write the Katz centrality compactly as

$$\mathbf{c}_\alpha^k := (\mathbf{I} - \alpha\mathbf{A})^{-1}\mathbf{1}. \quad (3)$$

Notice that if α is close to zero, the relative weight given to neighbors further away decreases fast, and \mathbf{c}_α^k is driven mainly by the one-hop neighbors just like degree centrality. In contrast, if α is close to $1/\lambda_1(\mathbf{A})$, the solution to (3) is almost a scaled version of \mathbf{c}^e . Intuitively, for intermediate values of

α , Katz centrality captures a hybrid notion of importance by combining elements from degree and eigenvector centralities. We remark that Katz centrality is sometimes defined as $c_\alpha^k - 1$. Since a constant shift does not alter the centrality ranking, we here use formula (3). We also note that Katz centrality is sometimes referred to as Bonacich centrality in the literature.

PageRank measures the importance of a node in a recursive way based on the importance of the neighboring nodes (weighted by their degree). Mathematically, the PageRank centrality of node is given by

$$c_\beta^p := (1 - \beta)(\mathbf{I} - \beta \mathbf{A} \mathbf{D}^{-1})^{-1} \mathbf{1}, \quad (4)$$

where $0 < \beta < 1$ and \mathbf{D} is the diagonal matrix of the degrees of the nodes. Note that the above formula corresponds to the stationary distribution of a random ‘surfer’ on a graph, who follows the links on the graph with probability β and with probability $(1 - \beta)$ jumps (‘teleports’) to a uniformly at random selected node in the graph. See [10] for further details on PageRank.

2.2 Graphons

A *graphon* is the limit of a convergent sequence of graphs of increasing size, that preserves certain desirable features of the graphs contained in the sequence [21], [22], [23], [24], [25], [39], [40], [41]. Formally, a graphon is a measurable function $W : [0, 1]^2 \rightarrow [0, 1]$ that is symmetric $W(x, y) = W(y, x)$. Intuitively, one can interpret the value $W(x, y)$ as the probability of existence of an edge between x and y . However, the ‘nodes’ x and y no longer take values in a finite node set as in classical finite graphs but rather in the continuous interval $[0, 1]$. Based on this intuition, graphons also provide a natural way of generating random graphs [39], [42], as introduced in the seminal paper [23] under the name W -random graphs. In this paper we will make use of the following model, in which the symmetric adjacency matrix $\mathbf{S}^{(N)} \in \{0, 1\}^{N \times N}$ of a simple random graph of size N constructed from a graphon is such that for all $i, j \in \{1, \dots, N\}$

$$\Pr[S_{ij}^{(N)} = 1 | u_i, u_j] = \kappa_N W(u_i, u_j), \quad (5)$$

where u_i and u_j are latent variables selected uniformly at random from $[0, 1]$, and κ_N is a constant regulating the sparsity of the graph (see also Definition 7)¹

This means that, when conditioned on the latent variables (u_1, u_2, \dots, u_N) , the off-diagonal entries of the symmetric matrix $\mathbf{S}^{(N)}$ are independent Bernoulli random variables with success probability given by $\kappa_N W$. In this sense, when $\kappa_N = 1$, the constant graphon $W(x, y) = p$ gives rise to Erdős-Rényi random graphs with edge probability p . Analogously, a piece-wise constant graphon gives rise to stochastic block models [30], [31]; for more details see Section 4.1. Interestingly, it can be shown that the distribution of any simple exchangeable random graph [34], [39] is characterized by a function W as discussed above [39], [47], [48]. Finally, observe that for any measure preserving map $\pi : [0, 1] \rightarrow [0, 1]$, the graphons $W(x, y)$ and

1. Throughout this paper we adopt the terminology of sparse graphs for graphs generated following (5) with parameter $\kappa_N \rightarrow 0$ and $N\kappa_N \rightarrow \infty$ as $N \rightarrow \infty$, even though this does not imply a bounded degree. This terminology is consistent with common usage in the literature [43], [44], [45]. Note also that [46] proposed an interesting graph limit framework for graph sequences with bounded degree.

$W^\pi(x, y) := W(\pi(x), \pi(y))$ define the same probability distribution on random graphs. A precise characterization of the equivalence classes of graphons defining the same probability distribution can be found in [22, Ch. 10].

3 EXTENDING CENTRALITIES TO GRAPHONS

In order to introduce centrality measures for graphons we first introduce a linear integral operator associated with a graphon and recall its spectral properties. From here on, we denote by $L^2([0, 1])$ the Hilbert function space with inner product $\langle f_1, f_2 \rangle := \int_0^1 f_1(x) f_2(x) dx$ for $f_1, f_2 \in L^2([0, 1])$, and norm $\|f_1\| := \sqrt{\langle f_1, f_1 \rangle}$. The elements of $L^2([0, 1])$ are the equivalence classes of Lebesgue integrable functions $f : [0, 1] \rightarrow \mathbb{R}$, that is, we identify two functions f and g with each other if they differ only on a set of measure zero (i.e., $f \equiv g \Leftrightarrow \|f - g\| = 0$). $1_{[0, 1]}$ is the identity function in $L^2([0, 1])$. We use blackboard bold symbols (such as \mathbb{L}) to denote linear operators acting on $L^2([0, 1])$, with the exception of \mathbb{N} and \mathbb{R} that denote the sets of natural and real numbers. The induced (operator) norm is defined as $\|\mathbb{L}\| := \sup_{f \in L^2([0, 1]) \text{ s.t. } \|f\|=1} \|\mathbb{L}f\|$.

3.1 The graphon integral operator and its properties

Following [22], we introduce a linear operator that is fundamental to derive the notions of centrality for graphons.

Definition 1 (Graphon operator). *For a given graphon W , we define the associated graphon operator \mathbb{W} as the linear integral operator $\mathbb{W} : L^2([0, 1]) \rightarrow L^2([0, 1])$*

$$f(y) \rightarrow (\mathbb{W}f)(x) = \int_0^1 W(x, y) f(y) dy.$$

From an operator theory perspective, the graphon W is the integral kernel of the linear operator \mathbb{W} . Given the key importance of \mathbb{W} , we review its spectral properties in the next definition and lemma.

Definition 2 (Eigenvalues and eigenfunctions). *A complex number λ is an eigenvalue of \mathbb{W} if there exists a nonzero function $\varphi \in L^2([0, 1])$, called the eigenfunction, such that*

$$(\mathbb{W}\varphi)(x) = \lambda\varphi(x). \quad (6)$$

It follows from the above definition that the eigenfunctions are only defined up to a rescaling parameter. Hence, from now on we assume all eigenfunctions are normalized such that $\|\varphi\| = 1$.

We next recall some known properties of the graphon operator.

Lemma 1. *The graphon operator \mathbb{W} has the following properties.*

- 1) \mathbb{W} is self-adjoint, bounded, and continuous.
- 2) \mathbb{W} is diagonalizable. Specifically, \mathbb{W} has countably many eigenvalues, all of which are real and can be ordered as $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$. Moreover, there exists an orthonormal basis for $L^2([0, 1])$ of eigenfunctions $\{\varphi_i\}_{i=1}^\infty$. That is, $(\mathbb{W}\varphi_i)(x) = \lambda_i\varphi_i(x)$, $\langle \varphi_i, \varphi_j \rangle = \delta_{i,j}$ for all i, j and any function $f \in L^2([0, 1])$ can be decomposed as $f(x) = \sum_{i=1}^\infty \langle f, \varphi_i \rangle \varphi_i(x)$. Consequently,

$$(\mathbb{W}f)(x) = \sum_{i=1}^\infty \lambda_i \langle f, \varphi_i \rangle \varphi_i(x).$$

If the set of nonzero eigenvalues is infinite, then 0 is its unique accumulation point.

3) Let \mathbb{W}^k denote k consecutive applications of the operator \mathbb{W} . Then, for any $k \in \mathbb{N}$,

$$(\mathbb{W}^k f)(x) = \sum_{i=1}^{\infty} \lambda_i^k \langle f, \varphi_i \rangle \varphi_i(x).$$

4) The maximum eigenvalue λ_1 is positive and there exists an associated eigenfunction φ_1 which is positive, that is, $\varphi_1(x) > 0$ for all $x \in [0, 1]$. Moreover, $\lambda_1 = \|\mathbb{W}\|$.

Points 1 to 3 of the lemma above can be found in [24], while part 4) follows from the Krein-Rutman theorem [49, Theorem 19.2], upon noticing that the graphon operator \mathbb{W} is positive with respect to the cone K defined by the set of nonnegative functions in $L^2([0, 1])$.

3.2 Definitions of centrality measures for graphons

We define centrality measures for graphons based on the graphon operator introduced in the previous section. These definitions closely parallel the construction of centrality measures in finite graphs; see Section 2.1. The main difference is that the linear operator defining the respective centralities is an infinite dimensional operator, rather than a finite dimensional matrix.

Definition 3 (Centrality measures for graphons). *Given a graphon W and its associated operator \mathbb{W} , we define the following centrality functions:*

1) **Degree centrality:** We define $c^d : [0, 1] \rightarrow \mathbb{R}_+$ as

$$c^d(x) := (\mathbb{W}1_{[0,1]})(x) = \int_0^1 W(x, y) dy. \quad (7)$$

2) **Eigenvector centrality:** For \mathbb{W} with a simple largest eigenvalue λ_1 , let φ_1 ($\|\varphi_1\| = 1$) be the associated positive eigenfunction. The eigenvector centrality function $c^e : [0, 1] \rightarrow \mathbb{R}_+$ is

$$c^e(x) := \varphi_1(x). \quad (8)$$

3) **Katz centrality:** Consider the operator \mathbb{M}_α where $(\mathbb{M}_\alpha f)(x) := f(x) - \alpha(\mathbb{W}f)(x)$. For any $0 < \alpha < 1/\|\mathbb{W}\|$, we define the Katz centrality function $c_\alpha^k : [0, 1] \rightarrow \mathbb{R}_+$ as

$$c_\alpha^k(x) := (\mathbb{M}_\alpha^{-1}1_{[0,1]})(x). \quad (9)$$

4) **PageRank centrality:** Consider the operator

$$(\mathbb{L}_\beta f)(x) = f(x) - \beta \int_0^1 W(x, y) f(y) (c^d(y))^\dagger dy,$$

where $(c^d(y))^\dagger = (c^d(y))^{-1}$ if $c^d(y) \neq 0$ and $(c^d(y))^\dagger = 0$ if $c^d(y) = 0$. For any $0 < \beta < 1$, we define $c_\beta^{\text{pr}} : [0, 1] \rightarrow \mathbb{R}_+$ as

$$c_\beta^{\text{pr}}(x) := (1 - \beta)(\mathbb{L}_\beta^{-1}1_{[0,1]})(x). \quad (10)$$

Note that $c^d(y) = \int_0^1 W(y, z) dz = 0$ implies $W(x, y) = 0$ almost everywhere.

Remark 1. The Katz centrality function is well defined, since for $0 < \alpha < 1/\|\mathbb{W}\|$ the operator \mathbb{M}_α is invertible [50, Theorem 2.2]. Moreover, denoting the identity operator by \mathbb{I} , it follows that $\mathbb{M}_\alpha = \mathbb{I} - \alpha\mathbb{W}$. Hence, by using a Neumann series representation

and the properties of the higher order powers of \mathbb{W} we obtain the equivalent representation

$$\begin{aligned} (\mathbb{M}_\alpha^{-1}f)(x) &= ((\mathbb{I} - \alpha\mathbb{W})^{-1}f)(x) = \sum_{k=0}^{\infty} \alpha^k (\mathbb{W}^k f)(x) \\ &= f(x) + \sum_{k=1}^{\infty} \alpha^k \sum_{i=1}^{\infty} \lambda_i^k \langle \varphi_i, f \rangle \varphi_i(x) \\ &= f(x) + \sum_{i=1}^{\infty} \frac{\alpha \lambda_i}{1 - \alpha \lambda_i} \langle \varphi_i, f \rangle \varphi_i(x), \end{aligned}$$

where we used that $|\lambda_i| < \|\mathbb{W}\|$ for all i . Using an analogous series representation it can be shown that PageRank is well defined. Note also that eigenvector centrality is well-defined by Lemma 1, part 4).

Since a graphon describes the limit of an infinite dimensional graph, there is a subtle difference in the semantics of the centrality measure compared to the finite graph setting. Specifically, in the classical setting the network consists of a finite number of nodes and thus for a graph of N nodes we obtain an N -dimensional vector with one centrality value per node. In the graphon setting, we may think of each real $x \in [0, 1]$ as corresponding to one of infinitely many nodes, and thus the centrality measure is described by a function.

4 COMPUTING CENTRALITIES ON GRAPHONS

We illustrate how to compute centrality measures for graphons by studying three examples in detail. The graphons we consider are ordered by increasing complexity of their respective eigenspaces and by the generality of the methods used in the computation of the centralities.

4.1 Stochastic block model graphons

We consider a class of piecewise constant graphons that may be seen as the equivalent of a stochastic block model (SBM). Such graphons play an important role in practice, as they enable us to approximate more complicated graphons in a ‘stepwise’ fashion. This approximation idea has been exploited to estimate graphons from finite data [28], [51], [52]. In fact, optimal statistical rates of convergence can be achieved over smooth graphon classes [45], [53]. The SBM graphon is defined as follows

$$W_{\text{SBM}}(x, y) := \sum_{i=1}^m \sum_{j=1}^m P_{ij} 1_{\mathcal{B}_i}(x) 1_{\mathcal{B}_j}(y), \quad (11)$$

where $P_{ij} \in [0, 1]$, $P_{ij} = P_{ji}$, $\cup_{i=1}^m \mathcal{B}_i = [0, 1]$ and $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ for $i \neq j$. We define the following m dimensional vector of indicator functions

$$\mathbf{1}(x) := [1_{\mathcal{B}_1}(x), \dots, 1_{\mathcal{B}_m}(x)]^T, \quad (12)$$

enabling us to compactly rewrite the graphon in (11) as

$$W_{\text{SBM}}(x, y) = \mathbf{1}(x)^T \mathbf{P} \mathbf{1}(y). \quad (13)$$

We also define the following auxiliary matrices.

Definition 4. Let us define the effective measure matrix $\mathbf{Q}_{\text{SBM}} \in \mathbb{R}^{m \times m}$ and the effective connectivity matrix $\mathbf{E}_{\text{SBM}} \in \mathbb{R}^{m \times m}$ for SBM graphons as follows

$$\mathbf{Q}_{\text{SBM}} := \int_0^1 \mathbf{1}(x) \mathbf{1}(x)^T dx, \quad \mathbf{E}_{\text{SBM}} := \mathbf{P} \mathbf{Q}_{\text{SBM}}. \quad (14)$$

Notice that \mathbf{Q}_{SBM} is a diagonal matrix with entries collecting the sizes of each block. Similarly, the matrix \mathbf{E}_{SBM}

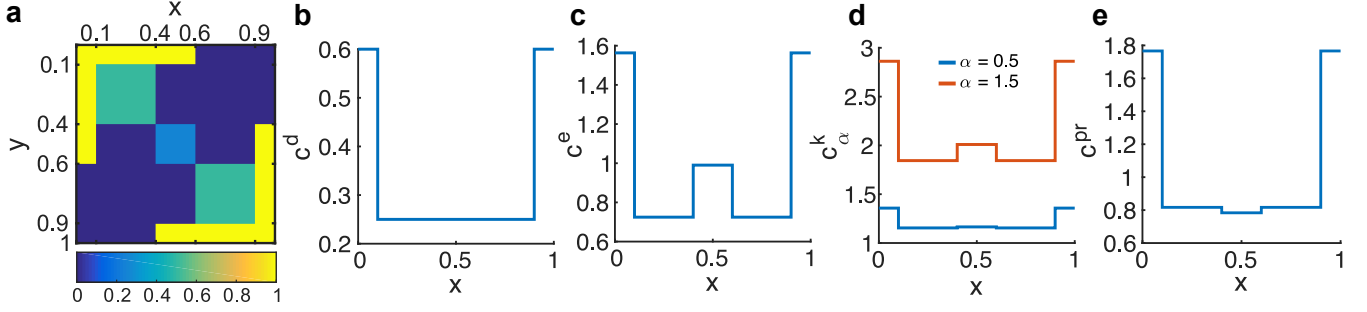


Fig. 2: Illustrative example of a graphon with stochastic block model structure. **(a)** Graphon W_{SBM} with block model structure as described in the text. **(b-d)** Degree, eigenvector, Katz, and PageRank centralities for the graphon depicted in (a).

is obtained by weighting the probabilities in \mathbf{P} by the sizes of the different blocks. Hence, the effective connectivity from block \mathcal{B}_i to two blocks \mathcal{B}_j and \mathcal{B}_k may be equal even if the latter block \mathcal{B}_k has twice the size ($Q_{kk} = 2Q_{jj}$), provided that it has half the probability of edge appearance ($2P_{ik} = P_{ij}$). Notice also that the matrix \mathbf{E}_{SBM} need not be symmetric. As will be seen in Section 4.2, the definitions in (14) are specific examples of more general constructions.

The following lemma relates the spectral properties of \mathbf{E}_{SBM} to those of the operator \mathbb{W}_{SBM} induced by W_{SBM} . We do not prove this lemma since it is a special case of Lemma 3, introduced in Section 4.2 and shown in Appendix A.

Lemma 2. *Let λ_i and \mathbf{v}_i denote the eigenvalues and eigenvectors of \mathbf{E}_{SBM} in (14), respectively. Then, all the nonzero eigenvalues of \mathbb{W}_{SBM} are given by λ_i and the associated eigenfunctions are of the form $\varphi_i(x) = \mathbf{1}(x)^T \mathbf{v}_i$.*

Using the result above, we can compute the centrality measures for stochastic block model graphons based on the effective connectivity matrix.

Proposition 1 (Centrality measures for SBM graphons). *Let λ_i and \mathbf{v}_i denote the eigenvalues and eigenvectors of \mathbf{E}_{SBM} in (14), respectively, and define the diagonal matrix $\mathbf{D}_{\mathbf{E}} := \text{diag}(\mathbf{E}_{\text{SBM}} \mathbf{1})$. The centrality functions c^d , c^e , c_α^k , and c_β^{PR} of the graphon W_{SBM} can be computed as follows*

$$c^d(x) = \mathbf{1}(x)^T \mathbf{E}_{\text{SBM}} \mathbf{1}, \quad c^e(x) = \frac{\mathbf{1}(x)^T \mathbf{v}_1}{\sqrt{\mathbf{v}_1^T \mathbf{Q}_{\text{SBM}} \mathbf{v}_1}}, \quad (15)$$

$$c_\alpha^k(x) = \mathbf{1}(x)^T (\mathbf{I} - \alpha \mathbf{E}_{\text{SBM}})^{-1} \mathbf{1},$$

$$c_\beta^{\text{PR}}(x) = (1 - \beta) \mathbf{1}(x)^T (\mathbf{I} - \beta \mathbf{E}_{\text{SBM}} \mathbf{D}_{\mathbf{E}}^{-1})^{-1} \mathbf{1}.$$

We next illustrate this result with an example. Its proof is given in Appendix A.

4.1.1 Example of a stochastic block model graphon

Consider the stochastic block model graphon W_{SBM} depicted in Fig. 2-(a), with corresponding symmetric matrix \mathbf{P} [cf. (11)] as in (16). Let us define the vector of indicator functions specific to this graphon $\mathbf{1}(x) := [1_{\mathcal{B}_1}(x), \dots, 1_{\mathcal{B}_5}(x)]^T$, where the blocks coincide with those in Fig. 2-(a), that is, $\mathcal{B}_1 = [0, 0.1)$, $\mathcal{B}_2 = [0.1, 0.4)$, $\mathcal{B}_3 = [0.4, 0.6)$, $\mathcal{B}_4 = [0.6, 0.9)$, and $\mathcal{B}_5 = [0.9, 1]$. To apply Proposition 1 we need to compute the effective measure and effective connectivity matrices

[cf. (14)], which for our example are given by

$$\mathbf{Q}_{\text{SBM}} = \text{diag} \begin{pmatrix} 0.1 \\ 0.3 \\ 0.2 \\ 0.3 \\ 0.1 \end{pmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0.5 & 0 & 0 & 0 \\ 1 & 0 & 0.25 & 0 & 1 \\ 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \quad (16a)$$

$$\mathbf{E}_{\text{SBM}} = \begin{bmatrix} 0.1 & 0.3 & 0.2 & 0 & 0 \\ 0.1 & 0.15 & 0 & 0 & 0 \\ 0.1 & 0 & 0.05 & 0 & 0.1 \\ 0 & 0 & 0 & 0.15 & 0.1 \\ 0 & 0 & 0.2 & 0.3 & 0.1 \end{bmatrix}. \quad (16b)$$

The principal eigenvector of \mathbf{E}_{SBM} is given by $\mathbf{v}_1 \approx [0.59, 0.28, 0.38, 0.28, 0.59]^T$. Furthermore, from (15) we can compute the graphon centrality functions to obtain

$$c^d(x) = \mathbf{1}(x)^T [0.6, 0.25, 0.25, 0.25, 0.6]^T,$$

$$c^e(x) \approx \mathbf{1}(x)^T [1.56, 0.72, 0.99, 0.72, 1.56]^T,$$

$$c_\alpha^k(x) \approx \begin{cases} \mathbf{1}(x)^T [1.36, 1.15, 1.16, 1.15, 1.36]^T & \text{if } \alpha = 0.5, \\ \mathbf{1}(x)^T [2.86, 1.84, 2.01, 1.84, 2.86]^T & \text{if } \alpha = 1.5, \end{cases}$$

$$c_{0.85}^{\text{PR}}(x) \approx \mathbf{1}(x)^T [1.77, 0.82, 0.78, 0.82, 1.77]^T,$$

where for illustration purposes we have evaluated the Katz centrality for two specific choices of α , and we have set $\beta = 0.85$ for the PageRank centrality. These four centrality functions are depicted in Fig. 2(b)-(e).

Note that these functions are piecewise constant according to the block partition $\{\mathcal{B}_i\}_{i=1}^m$. Moreover, as expected from the functional form of W_{SBM} in Fig. 2-(a), blocks \mathcal{B}_1 and \mathcal{B}_5 are the most central as measured by any of the four studied centralities. Regarding the remaining three blocks, degree centrality deems them as equally important whereas eigenvector centrality considers \mathcal{B}_3 to be more important than \mathcal{B}_2 and \mathcal{B}_4 . To understand this discrepancy, notice that in any finite realization of the graphon W_{SBM} , most of the edges from a node in block \mathcal{B}_3 will go to nodes in \mathcal{B}_1 and \mathcal{B}_5 , which are the most central ones. On the other hand, for nodes in blocks \mathcal{B}_2 and \mathcal{B}_4 , most of the edges will be contained within their own block. Hence, even though nodes corresponding to blocks \mathcal{B}_2 , \mathcal{B}_3 , and \mathcal{B}_4 have the same expected number of neighbors – thus, same degree centrality – the neighbors of nodes in \mathcal{B}_3 tend to be more central, entailing a higher eigenvector centrality. As expected, an intermediate situation occurs with Katz centrality, whose form is closer to degree centrality for lower values of α (cf. $\alpha = 0.5$) and closer to eigenvector centrality for larger values of this parameter (cf.

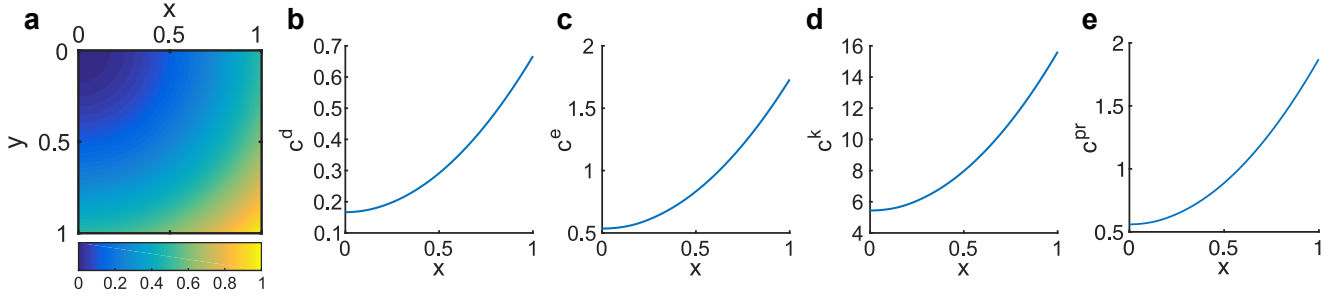


Fig. 3: Illustrative example of a graphon with finite rank. (a) The graphon $W_{FR} = (x^2 + y^2)/2$ is decomposable into a finite number of components, inducing a finite-rank graphon operator. (b-d) Degree, eigenvector, Katz, and PageRank centralities for the graphon depicted in (a).

$\alpha = 1.5$). For the case of PageRank, on the other hand, block B_3 is deemed as less central than B_2 and B_4 . This can be attributed to the larger size of these latter blocks. Indeed, the classical PageRank centrality measure is partially driven by size [10].

4.2 Finite-rank graphons

We now consider a class of finite-rank (FR) graphons that can be written as a finite sum of products of integrable functions. Specifically, we consider graphons of the form

$$W_{FR}(x, y) := \sum_{i=1}^m g_i(x)h_i(y) = \mathbf{g}(x)^T \mathbf{h}(y), \quad (17)$$

where $m \in \mathbb{N}$ and we have defined the vectors of functions $\mathbf{g}(x) = [g_1(x), \dots, g_m(x)]^T$ and $\mathbf{h}(y) = [h_1(y), \dots, h_m(y)]^T$. Observe that $\mathbf{g}(x)$ and $\mathbf{h}(y)$ must be chosen so that W_{FR} is symmetric, and $W_{FR}(x, y) \in [0, 1]$ for all $(x, y) \in [0, 1]^2$. Based on $\mathbf{g}(x)$ and $\mathbf{h}(y)$ we can define the generalizations of \mathbf{Q}_{SBM} and \mathbf{E}_{SBM} introduced in Section 4.1, for this class of finite-rank graphons.

Definition 5. The effective measure matrix \mathbf{Q} and the effective connectivity matrix \mathbf{E} for a finite-rank graphon W_{FR} as defined in (17) are given by

$$\mathbf{Q} := \int_0^1 \mathbf{g}(x)\mathbf{g}(x)^T dx, \quad \mathbf{E} := \int_0^1 \mathbf{h}(x)\mathbf{g}(x)^T dx. \quad (18)$$

The stochastic block model graphon operator introduced in (11) is a special case of the class of operators in (17). More precisely, we recover the SBM graphon by choosing $g_i(x) = 1_{B_i}(x)$ and $h_i(y) = \sum_{j=1}^m P_{ij} 1_{B_j}(y)$ for $i = 1, \dots, m$. The matrices defined in (14) are recovered when specializing Definition 5 to this choice of $g_i(x)$ and $h_i(y)$. We may now relate the eigenfunctions of the FR graphon with the spectral properties of \mathbf{E} , as explained in the following lemma.

Lemma 3. Let λ_i and \mathbf{v}_i denote the eigenvalues and eigenvectors of \mathbf{E} in (18), respectively. Then, all the nonzero eigenvalues of \mathbb{W}_{FR} , the operator associated with (17), are given by λ_i and the associated eigenfunctions are of the form $\varphi_i(x) = \mathbf{g}(x)^T \mathbf{v}_i$.

Lemma 3 is proven in Appendix A and shows that the graphon in (17) is of finite rank since it has at most m non-zero eigenvalues. Notice that Lemma 2 follows from Lemma 3 when specializing the finite rank operator to the SBM case as explained after Definition 5. Moreover, we can leverage the result in Lemma 3 to find closed-form expressions for the centrality functions of FR graphons. To write these expressions compactly, we define the vectors of

integrated functions $\mathbf{g} := \int_0^1 \mathbf{g}(y)dy$ and $\mathbf{h} := \int_0^1 \mathbf{h}(y)dy$, as well as the following normalized versions of \mathbf{h} and \mathbf{E}

$$\mathbf{h}_{nor} = \int_0^1 \frac{\mathbf{h}(y)}{\mathbf{g}^T \mathbf{h}(y)} dy, \quad \mathbf{E}_{nor} = \int_0^1 \frac{\mathbf{h}(y)\mathbf{g}(y)^T}{\mathbf{g}^T \mathbf{h}(y)} dy. \quad (19)$$

With this notation in place, we can establish the following result, which is proven in Appendix A.

Proposition 2 (Centrality measures for FR graphons). *Let \mathbf{v}_1 be the principal eigenvector of \mathbf{E} in (18). Then, the centrality functions c^d , c^e , c_α^k , and c_β^{pr} of the graphon W_{FR} can be computed as follows*

$$\begin{aligned} c^d(x) &= \mathbf{g}(x)^T \mathbf{h}, \quad c^e(x) = \frac{\mathbf{g}(x)^T \mathbf{v}_1}{\sqrt{\mathbf{v}_1^T \mathbf{Q} \mathbf{v}_1}}, \\ c_\alpha^k(x) &= 1 + \alpha \mathbf{g}(x)^T (\mathbf{I} - \alpha \mathbf{E})^{-1} \mathbf{h}, \\ c_\beta^{pr}(x) &= (1 - \beta)(1 + \beta \mathbf{g}(x)^T (\mathbf{I} - \beta \mathbf{E}_{nor})^{-1} \mathbf{h}_{nor}). \end{aligned} \quad (20)$$

In the next subsection we illustrate the use of Proposition 2 for the computation of graphon centralities.

4.2.1 Example of a finite-rank graphon

Consider the FR graphon given by

$$W_{FR}(x, y) = (x^2 + y^2)/2,$$

and illustrated in Fig. 3-(a). Notice that this FR graphon can be written in the canonical form (17) by defining the vectors $\mathbf{g}(x) = [x^2, 1/2]^T$ and $\mathbf{h}(y) = [1/2, y^2]^T$. From (18) we then compute the relevant matrices \mathbf{Q} and \mathbf{E} , as well as the relevant normalized quantities, to obtain

$$\begin{aligned} \mathbf{Q} &= \begin{bmatrix} 1/5 & 1/6 \\ 1/6 & 1/4 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} 1/6 & 1/4 \\ 1/5 & 1/6 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} 1/3 \\ 1/2 \end{bmatrix}, \\ \mathbf{h} &= \begin{bmatrix} 1/2 \\ 1/3 \end{bmatrix}, \quad \mathbf{h}_{nor} \approx \begin{bmatrix} 1.81 \\ 0.79 \end{bmatrix}, \quad \mathbf{E}_{nor} \approx \begin{bmatrix} 0.40 & 0.91 \\ 0.40 & 0.40 \end{bmatrix}. \end{aligned}$$

A simple computation reveals that the principal eigenvector of \mathbf{E} is $\mathbf{v}_1 = [\sqrt{10}/3, 2\sqrt{2}/3]^T$. We now leverage the result in Proposition 2 to obtain

$$\begin{aligned} c^d(x) &= [x^2, 1/2] \begin{bmatrix} 1/2 \\ 1/3 \end{bmatrix} = \frac{x^2}{2} + \frac{1}{6}, \\ c^e(x) &= \frac{3}{2} \sqrt{\frac{3}{3 + \sqrt{5}}} [x^2, 1/2] \begin{bmatrix} \sqrt{10}/3 \\ 2\sqrt{2}/3 \end{bmatrix} \approx 1.07 x^2 + 0.54, \\ c_\alpha^k(x) &= [x^2, 1/2] \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \alpha \begin{bmatrix} 1/6 & 1/4 \\ 1/5 & 1/6 \end{bmatrix} \right)^{-1} \alpha \begin{bmatrix} 1/2 \\ 1/3 \end{bmatrix} + 1 \\ &\approx 10.19 x^2 + 5.44, \\ c_{0.85}^{pr}(x) &\approx 1.31 x^2 + 0.56, \end{aligned}$$

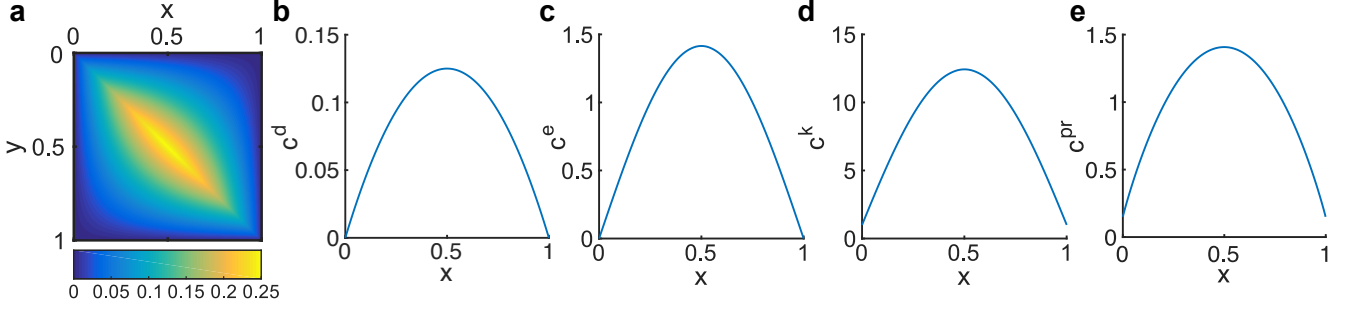


Fig. 4: Illustrative example of a general smooth graphon. (a) The graphon W_G induces an operator that has countably infinite number of nonzero eigenvalues and corresponding eigenfunctions. (b-d) Degree, eigenvector, Katz, and PageRank centralities for the graphon depicted in (a).

where we have set $\beta = 0.85$ in the PageRank centrality. Moreover, we have evaluated the Katz centrality for $\alpha = 0.9/\lambda_1$, where λ_1 is the largest eigenvalue of \mathbb{E} . The four centrality functions are depicted in Fig. 3-(b) through (e). As anticipated from the form of W_{FR} , there is a simple monotonicity in the centrality ranking for all the measures considered. More precisely, highest centrality values are located close to 1 in the interval $[0, 1]$, whereas low centralities are localized close to 0. Unlike in the example of the stochastic block model in Section 4.1.1, all centralities here have the same functional form of a quadratic term with a constant offset.

4.3 General smooth graphons

In general, a graphon W need not induce a finite-rank operator as in the preceding Sections 4.1 and 4.2. However, as shown in Lemma 1, a graphon always induces a diagonalizable operator with countably many nonzero eigenvalues. In most cases, obtaining the degree centrality function is immediate since it entails the computation of an integral [cf. (7)]. On the other hand, for eigenvector, Katz, and PageRank centralities that depend on the spectral decomposition of \mathbb{W} , there is no universal technique available. Nonetheless, a procedure that has shown to be useful in practice to obtain the eigenfunctions φ and corresponding eigenvalues λ of smooth graphons is to solve a set of differential equations obtained by successive differentiation, when possible, of the eigenfunction equation in (6), that is, by considering

$$\frac{d^k}{dx^k} \int_0^1 W(x, y) \varphi(y) dy = \lambda \frac{d^k \varphi(x)}{dx^k}, \quad (21)$$

for $k \in \mathbb{N}$. In the following section we illustrate this technique on a specific smooth graphon that does not belong to the finite-rank class.

4.3.1 Example of a general smooth graphon

Consider the graphon W_G depicted in Fig. 4-(a) and with the following functional form

$$W_G(x, y) = \min(x, y)[1 - \max(x, y)].$$

Specializing the differential equations in (21) for graphon W_G we obtain

$$\frac{d^k}{dx^k} \left[(1-x) \int_0^x y \varphi(y) dy + x \int_x^1 (1-y) \varphi(y) dy \right] = \lambda \frac{d^k \varphi(x)}{dx^k}. \quad (22)$$

First notice that without differentiating (i.e. for $k = 0$) we can determine the boundary conditions $\varphi(0) = \varphi(1) = 0$.

Moreover, by computing the second derivatives in (22), we obtain that $-\varphi(x) = \lambda \varphi''(x)$. From the solution of this differential equation subject to the boundary conditions it follows that the eigenvalues and eigenfunctions of the operator \mathbb{W}_G are

$$\lambda_n = \frac{1}{\pi^2 n^2} \quad \text{and} \quad \varphi_n(x) = \sqrt{2} \sin(n\pi x) \quad \text{for } n \in \mathbb{N}. \quad (23)$$

Notice that \mathbb{W}_G has an infinite — but countable — number of nonzero eigenvalues, with an accumulation point at zero. Thus, \mathbb{W}_G cannot be written in the canonical form for finite-rank graphons (17). Nevertheless, having obtained the eigenfunctions we can still compute the centrality measures for W_G . For degree centrality, a simple integration gives us

$$c^d(x) = (1-x) \int_0^x y dy + x \int_x^1 (1-y) dy = \frac{x(1-x)}{2}.$$

From (23) it follows that the principal eigenfunction is achieved when $n = 1$. Thus, the eigenvector centrality function [cf. (8)] is given by

$$c^e(x) = \sqrt{2} \sin(\pi x).$$

Finally, for the Katz centrality we leverage Remark 1 and the eigenfunction expressions in (23) to obtain

$$\begin{aligned} c_\alpha^k(x) &= 1 + \sum_{k=1}^{\infty} \alpha^k \sum_{n=1}^{\infty} \left(\frac{1}{n^2 \pi^2} \right)^k \frac{1 - (-1)^n}{\pi n} 2 \sin(n\pi x) \\ &= 1 + \sum_{n=1}^{\infty} \frac{2\alpha}{n^2 \pi^2 - \alpha} \frac{1 - (-1)^n}{\pi n} \sin(n\pi x), \end{aligned}$$

which is guaranteed to converge as long as $\alpha < 1/\lambda_1 = \pi^2$. We plot these three centrality functions in Fig. 4-(b) through (d), where we selected $\alpha = 0.9\pi^2$ for the Katz centrality. Also, in Fig. 4-(e) we plot an approximation to the PageRank centrality function c_β^{pr} obtained by solving numerically the integral in its definition. According to all centralities, the most important nodes within this graphon are those located in the center of the interval $[0, 1]$, in line with our intuition. Likewise, nodes at the boundary have low centrality values. Note that while the ranking according to all centrality functions is consistent, unlike in the example in Section 4.2.1, here there are some subtle differences in the functional forms. In particular, degree centrality is again a quadratic function whereas the eigenvector and Katz centralities are of sinusoidal form.

5 CONVERGENCE OF CENTRALITY MEASURES

In this section we derive concentration inequalities relating the newly defined graphon centrality functions with standard centrality measures. To this end, we start by noting that while graphon centralities are functions, standard centrality measures are vectors. To be able to compare such objects, we first show that there is a one to one relation between any finite graph with adjacency matrix \mathbf{A} and a suitably constructed stochastic block model graphon $W_{\text{SBM}|\mathbf{A}}$. Consequently, any centrality measure of \mathbf{A} is in one to one relation with the corresponding centrality function of $W_{\text{SBM}|\mathbf{A}}$. To this end, for each $N \in \mathbb{N}$, we define a partition of $[0, 1]$ into the intervals \mathcal{B}_i^N , where $\mathcal{B}_i^N = [(i-1)/N, i/N]$ for $i = 1, \dots, N-1$, and $\mathcal{B}_N^N = [(N-1)/N, 1]$. We denote the associated indicator-function vector by $\mathbf{1}_N(x) := [1_{\mathcal{B}_1^N}(x), \dots, 1_{\mathcal{B}_N^N}(x)]^T$, consistently with (12).

Lemma 4. For any adjacency matrix $\mathbf{A} \in \{0, 1\}^{N \times N}$ define the corresponding stochastic block model graphon as

$$W_{\text{SBM}|\mathbf{A}}(x, y) = \sum_{i=1}^N \sum_{j=1}^N A_{ij} 1_{\mathcal{B}_i^N}(x) 1_{\mathcal{B}_j^N}(y).$$

Then the centrality function $c_N(x)$ corresponding to the graphon $W_{\text{SBM}|\mathbf{A}}$ is given by

$$c_N(x) = \mathbf{1}_N(x)^T \mathbf{c}_A.$$

where \mathbf{c}_A is the centrality measures of the graph with rescaled adjacency matrix $\frac{1}{N}\mathbf{A}$.

Proof. The graphon $W_{\text{SBM}|\mathbf{A}}$ has the stochastic block model structure described in Section 4.1, with N uniform blocks. By selecting $m = N$ and $\mathcal{B}_i = \mathcal{B}_i^N$ for each $i \in \{1, \dots, N\}$, we obtain $\mathbf{Q}_{\text{SBM}} = \frac{1}{N}\mathbf{I}$. Consequently, the formulas in Proposition 1 simplify as given in the statement of this lemma. \square

Remark 2. Note that the scaling factor $\frac{1}{N}$ does not affect the centrality rankings of the nodes in \mathbf{A} , but only the magnitude of the centrality measures. This re-scaling is needed to avoid diverging centrality measures for graphs of increasing size. Observe further that $c_N(x)$ is the piecewise-constant function corresponding to the vector \mathbf{c}_A . We finally remark that graphons of the type $W_{\text{SBM}|\mathbf{A}}$ have appeared before in the literature using a different notation [24].

By using the previous lemma, we can compare centralities of graphons and graphs by working in the function space. Using this equivalence, we demonstrate that the previously defined graphon centrality functions are not only defined analogously to the centrality measures on finite graphs, but also emerge as the limit of those centrality measures for a sequence of graphs of increasing size. Stated differently, just like the graphon provides an appropriate limiting object for a growing sequence of finite graphs, the graphon centrality functions can be seen as the appropriate limiting objects of the finite centrality measures as the size of the graphs tends to infinity. In this sense, the centralities presented here may be seen as a proper generalization of the finite setting, just like a graphon provides a generalized framework for finite graphs. Most importantly, we show that the distance (in L_2 norm) between the graphon centrality function and the step-wise constant function associated with the centrality vector

of any graph sampled from the graphon can be bounded, with high probability, in terms of the sampled graph size N .

Definition 6 (Sampled graphon). Given a graphon W and a size $N \in \mathbb{N}$ fix the latent variables $\{u_i\}_{i=1}^N$ by choosing either:

- ‘deterministic latent variables’: $u_i = \frac{i}{N}$.
- ‘stochastic latent variables’: $u_i = U_{(i)}$ where $U_{(i)}$ is the i -th order statistic of N random samples from $\text{Unif}[0, 1]$.

Utilizing such latent variables construct

- the ‘probability’ matrix $\mathbf{P}^{(N)} \in [0, 1]^{N \times N}$

$$P_{ij}^{(N)} := W(u_i, u_j) \text{ for all } i, j \in \{1, \dots, N\}.$$

- the sampled graphon

$$W_N(x, y) := \sum_{i=1}^N \sum_{j=1}^N P_{ij}^{(N)} 1_{\mathcal{B}_i^N}(x) 1_{\mathcal{B}_j^N}(y).$$

- the operator \mathbb{W}_N of the sampled graphon

$$(\mathbb{W}_N f)(x) := \sum_{j=1}^N P_{ij}^{(N)} \int_{\mathcal{B}_j^N} f(y) dy \text{ for any } x \in \mathcal{B}_i^N$$

The sampled graphon W_N obtained when working with deterministic latent variables can intuitively be seen as an approximation of the graphon W by using a stochastic block model graphon with N blocks, as the one described in Section 4.1, and is useful to study graphon centrality functions as limit of graph centrality measures. On the other hand, the sampled graphon W_N obtained when working with stochastic latent variables is useful as an intermediate step to analyze the relation between the graphon centrality function and the centrality measure of graphs sampled from the graphon according to the following procedure.

Definition 7 (Sampled graph). Given the ‘probability’ matrix $\mathbf{P}^{(N)}$ of a sampled graphon we define

- the sampled matrix $\mathbf{S}^{(N)} \in \{0, 1\}^{N \times N}$ as the adjacency matrix of a symmetric (random) graph obtained by taking N isolated vertices $i \in \{1, \dots, N\}$, and adding undirected edges between vertices i and j at random with probability $\kappa_N P_{ij}^{(N)}$ for all $i > j$. Note that $\mathbb{E}[\mathbf{S}^{(N)}] = \kappa_N \mathbf{P}^{(N)}$.

- the associated (random) linear operator

$$(\mathbb{S}_N f)(x) := \sum_{j=1}^N S_{ij}^{(N)} \int_{\mathcal{B}_j^N} f(y) dy \text{ for any } x \in \mathcal{B}_i^N, \quad (24)$$

and its associated (random) graphon:

$$S_N(x, y) = \mathbf{1}_N(x)^T \mathbf{S}^{(N)} \mathbf{1}_N(y). \quad (25)$$

In general, we denote the centrality functions associated with the sampled graphon operator \mathbb{W}_N by $c_N(x)$ whereas the centrality functions associated with the operator $\kappa_N^{-1} \mathbb{S}_N$ are denoted by $\hat{c}_N(x)$. Note that thanks to Lemma 4 such centrality functions are in one to one correspondence with the centrality measures of the finite graphs $\mathbf{P}^{(N)}$ and $\kappa_N^{-1} \mathbf{S}^{(N)}$, respectively. Consequently, studying the relation between $c(x)$, $c_N(x)$ and $\hat{c}_N(x)$ allows us to relate the graphon centrality function with the centrality measure of graphs sampled from the graphon. We note that previous works on graph limit theory imply the convergence of \mathbb{W}_N and $\kappa_N^{-1} \mathbb{S}_N$ to the graphon operator \mathbb{W} [22], [25], [54]. Intuitively, convergence

of $c_N(x)$ and $\hat{c}_N(x)$ to $c(x)$ then follows from continuity of spectral properties of the graphon operator. In the following, we make this argument precise and more importantly we provide a more refined analysis by establishing sharp rates of convergence of the sampled graphs, under the following smoothness assumption on W . A more detailed discussion of related convergence results can be found in Appendix B (see supplementary material).

Assumption 1 (Piecewise Lipschitz graphon). *There exists a constant L and a sequence of non-overlapping intervals $\mathcal{I}_k = [\alpha_{k-1}, \alpha_k]$ defined by $0 = \alpha_0 < \dots < \alpha_{K+1} = 1$, for a (finite) $K \in \mathbb{N}$, such that for any k, l , any set $\mathcal{I}_{kl} = \mathcal{I}_k \times \mathcal{I}_l$ and pairs $(x, y) \in \mathcal{I}_{kl}$, $(x', y') \in \mathcal{I}_{kl}$ we have that*

$$|W(x, y) - W(x', y')| \leq L(|x - x'| + |y - y'|).$$

This assumption has also been used in the context of graphon estimation [28], [53] and is typically fulfilled for most of the graphons of interest.

Theorem 1 (Convergence of graphon operators). *For a graphon fulfilling Assumption 1, it holds with probability $1 - \delta'$ that:*

$$\begin{aligned} \|\mathbb{W}_N - \mathbb{W}\| &:= \sup_{\|f\|=1} \|\mathbb{W}_N f - \mathbb{W} f\| \\ &\leq 2\sqrt{(L^2 - K^2)d_N^2} + Kd_N =: \rho(N), \end{aligned} \quad (26)$$

where $\delta' = 0$ and $d_N = \frac{1}{N}$ in the case of deterministic latent variables and $\delta' = \delta \in (Ne^{-N/5}, e^{-1})$ and $d_N := \frac{1}{N} + \sqrt{\frac{8 \log(N/\delta)}{(N+1)}}$ in the case of stochastic latent variables. Moreover, if N is large enough, as specified in Lemma 5, then with probability at least $1 - \delta - \delta'$

$$\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \leq \sqrt{\frac{4\kappa_N^{-1} \log(2N/\delta)}{N}} + \rho(N). \quad (27)$$

In particular, if $\kappa_N = \frac{1}{N^\tau}$ with $\tau \in [0, 1)$, then:

$$\lim_{N \rightarrow \infty} \|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| = 0, \text{ almost surely.} \quad (28)$$

Remark 3. To control the error induced by the random sampling, we derive a concentration inequality for uniform order statistics, as reported next, that can be of independent interest. As this result is not central to the discussion of our paper, we relegate its proof to Appendix B. In addition, we make use of a lower bound on the maximum expected degree as reported in Lemma 5 and proven in Appendix B.

Proposition 3. Let $U_{(i)}$ be the order statistics of N points sampled from a standard uniform distribution. Suppose that $N \geq 20$ and $\delta \in (Ne^{-N/5}, e^{-1})$. With probability at least $1 - \delta$

$$\left| U_{(i)} - \frac{i}{N+1} \right| \leq \sqrt{\frac{8 \log(N/\delta)}{(N+1)}} \text{ for all } i.$$

Lemma 5. If N is such that

$$2d_N < \Delta_{\min}^{(\alpha)} := \min_{k \in \{1, \dots, K+1\}} (\alpha_k - \alpha_{k-1}), \quad (29a)$$

$$\frac{1}{N} \log \left(\frac{2N}{\delta} \right) + d_N(2K + 3L) < C^d := \max_x c^d(x) \quad (29b)$$

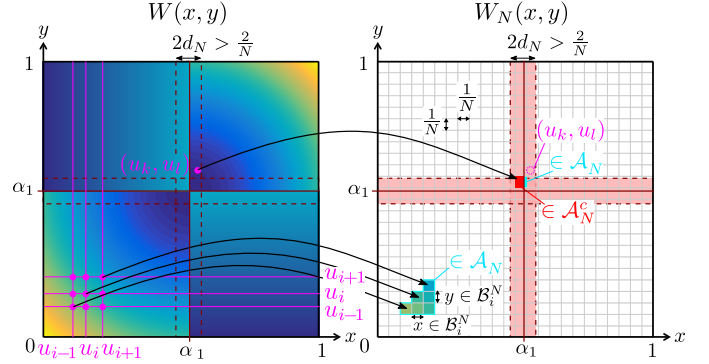


Fig. 5: Schematic for the sets \mathcal{A}_N and \mathcal{A}_N^c in the case of stochastic latent variables for $K = 1$, threshold α_1 and 4 Lipschitz blocks. The plot on the left shows the original graphon $W(x, y)$, the 4 Lipschitz blocks and some representative latent variables. The plot on the right shows the sampled graphon $W_N(x, y)$ which is a piecewise constant graphon with uniform square blocks of side $\frac{1}{N}$. The $\frac{1}{N}$ -grid is illustrated in gray. The constant value in each block $\mathcal{B}_i^N \times \mathcal{B}_j^N$ corresponds to the value in the original graph sampled at the point (u_i, u_j) (as illustrated by the arrows). The set $\mathcal{B}_i^N \times \mathcal{B}_j^N$ in the bottom left is an example where all the points $(x, y) \in \mathcal{B}_i^N \times \mathcal{B}_j^N$ belong to the same Lipschitz block as their corresponding sample, which is (u_i, u_i) , so that $\mathcal{B}_i^N \times \mathcal{B}_j^N = \mathcal{S}_{ii}$ and therein $|D(x, y)| \leq 2Ld_N$. The set $\mathcal{B}_k^N \times \mathcal{B}_l^N$ instead is one of the problematic ones since part of its points (cyan) belong to the same Lipschitz block as (u_k, u_l) and are thus in \mathcal{S}_{kl} but part of its points (red) do not and therefore belong to \mathcal{A}_N^c . Note that by construction with probability $1 - \delta'$ all such problematic points are contained in the set \mathcal{D}_N (which is illustrated in light red). This figure also illustrates that in general \mathcal{D}_N is a strict superset of \mathcal{A}_N^c (hence the bound in (30) is conservative).

$$\text{then } C_N^d := \max_i \left(\sum_{j=1}^N P_{ij}^{(N)} \right) \geq \frac{4}{9} \log \left(\frac{2N}{\delta} \right).$$

Proof of Theorem 1:

We prove the three statements sequentially.

Proof of (26). First of all note that by definition, for any $(x, y) \in \mathcal{B}_i^N \times \mathcal{B}_j^N$ it holds $W_N(x, y) = W(u_i, u_j)$, but it is not necessarily true that $(u_i, u_j) \in \mathcal{B}_i^N \times \mathcal{B}_j^N$. Let us define $k_i, k_j \in \{1, \dots, K+1\}$ such that the point (u_i, u_j) belongs to the Lipschitz block $\mathcal{I}_{k_i k_j}$, as defined in Assumption 1 and illustrated in Fig. 5. We define as \mathcal{S}_{ij} the subset of points in $\mathcal{B}_i^N \times \mathcal{B}_j^N$ that belong to the same Lipschitz block $\mathcal{I}_{k_i k_j}$ as (u_i, u_j) . Mathematically,

$$\mathcal{S}_{ij} = \{(x, y) \in \mathcal{B}_i^N \times \mathcal{B}_j^N \mid (u_i, u_j) \in \mathcal{I}_{k_i k_j} \text{ and } (x, y) \in \mathcal{I}_{k_i k_j}\}.$$

In the following, we partition the set $[0, 1]^2$ into the set $\mathcal{A}_N := \cup_{ij} \mathcal{S}_{ij}$ and its complement $\mathcal{A}_N^c := [0, 1]^2 \setminus \mathcal{A}_N$. In words, \mathcal{A}_N is the set of points, for which (x, y) and its corresponding sample (u_i, u_j) belong to the same Lipschitz block. We now prove that, with probability $1 - \delta'$, \mathcal{A}_N^c has area

$$\text{Area}(\mathcal{A}_N^c) \leq \text{Area}(\mathcal{D}_N) = 4Kd_N - 4K^2d_N^2. \quad (30)$$

To prove the above, we define the set \mathcal{D}_N by constructing a stripe of width $2d_N$ centered at each discontinuity of the graphon, as specified in Assumption 1. This guarantees that any point in $[0, 1]^2 \setminus \mathcal{D}_N$ has distance more than d_N component-wise from a discontinuity.

Note that in the case of deterministic latent variables for any $i \in \{1, \dots, N\}$ and any $x \in \mathcal{B}_i^N$ it holds by construction that $|x - u_i| = |x - \frac{i}{N}| \leq d_N = \frac{1}{N}$ and similarly $|y - u_j| \leq d_N$. In the case of stochastic latent variables, Proposition 3 guarantees that with probability at least $1 - \delta$ for any $i, j \in \{1, \dots, N\}$ and any $x \in \mathcal{B}_i^N, y \in \mathcal{B}_j^N$ it holds $|x - u_i| =$

$|x - U_{(i)}| \leq d_N, |y - u_j| \leq d_N$. In both cases, with probability $1 - \delta'$, all the points in $[0, 1]^2 \setminus \mathcal{D}_N$ are less than d_N close to their sample (u_i, u_j) and more than d_N far from any discontinuity (component-wise) hence they surely belong to \mathcal{A}_N .

Consequently, with probability $1 - \delta'$ we have $\mathcal{A}_N^c \subseteq \mathcal{D}_N$. Each stripe in \mathcal{D}_N has width $2d_N$, length 1, and there are $2K$ stripes in total. Formula (30) is then immediate by noticing that multiplying $2d_N$ times $2K$ counts twice the K^2 intersections between horizontal and vertical stripes.

Consider now any $f \in L^2([0, 1])$ such that $\|f\| = 1$. Let $D(x, y) := W_N(x, y) - W(x, y)$ and note that $|D(x, y)| \leq 1$. Then we get

$$\begin{aligned} \|\mathbb{W}_N f - \mathbb{W} f\|^2 &= \int_0^1 (\mathbb{W}_N f - \mathbb{W} f)^2(x) dx \\ &= \int_0^1 \left(\int_0^1 D(x, y) f(y) dy \right)^2 dx \\ &\leq \int_0^1 \left(\int_0^1 D(x, y)^2 dy \right) \left(\int_0^1 f^2(y) dy \right) dx \quad (31) \\ &= \int_0^1 \left(\int_0^1 D(x, y)^2 dy \right) \|f\|^2 dx = \int_0^1 \int_0^1 D(x, y)^2 dy dx \\ &= \iint_{\mathcal{A}_N} D(x, y)^2 dy dx + \iint_{\mathcal{A}_N^c} D(x, y)^2 dy dx. \quad (32) \end{aligned}$$

Expression (31) follows from the Cauchy-Schwarz inequality; we used $\|f\| = 1$ and, in the last equation, we split the interval $[0, 1]^2$ into the sets \mathcal{A}_N and \mathcal{A}_N^c , as described above and illustrated in Fig. 5.

We can now bound both terms in (32). For the first term, note that for all the points (x, y) in \mathcal{A}_N the corresponding sample (u_i, u_j) belongs to the same Lipschitz block and is at most d_N apart (component-wise). Consequently, for these points $|D(x, y)| \leq 2Ld_N$. Overall, we get

$$\iint_{\mathcal{A}_N} D(x, y)^2 dy dx \leq 4L^2 d_N^2 \iint_{\mathcal{A}_N} 1 dy dx \leq 4L^2 d_N^2.$$

For the second term in (32), we use (30) and the fact that $|D(x, y)| \leq 1$ to get

$$\iint_{\mathcal{A}_N^c} D(x, y)^2 dy dx \leq \iint_{\mathcal{A}_N^c} 1 dy dx = \text{Area}(\mathcal{A}_N^c).$$

Substituting these two terms into (32) yields

$$\|\mathbb{W}_N f - \mathbb{W} f\|^2 \leq (4L^2 d_N^2 + 4Kd_N - 4K^2 d_N^2).$$

Since this bound holds for all functions f with unit norm, we recover (26).

Proof of (27). From the triangle inequality we get that

$$\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \leq \|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}_N\| + \|\mathbb{W}_N - \mathbb{W}\|. \quad (33)$$

We have already bounded the second term on the right hand side of (33), so we now concentrate on the first term.

The operator $\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}_N$ can be seen as the graphon operator of an SBM graphon with matrix $\kappa_N^{-1} \mathbf{S}^{(N)} - \mathbf{P}^{(N)}$. By Lemma 2 we then have that its eigenvalues coincide with the eigenvalues of the corresponding \mathbf{E}_{SBM} matrix which is $\frac{1}{N}(\kappa_N^{-1} \mathbf{S}^{(N)} - \mathbf{P}^{(N)})$ since in this case $\mathbf{Q}_{\text{SBM}} = \frac{1}{N} \mathbf{I}_N$ (given that all the intervals \mathcal{B}_i^N have length $\frac{1}{N}$).² Consequently,

$$\begin{aligned} \|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}_N\| &= \lambda_{\max}(\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}_N) \\ &= \frac{1}{N} \lambda_{\max}(\kappa_N^{-1} \mathbf{S}^{(N)} - \mathbf{P}^{(N)}) = \frac{1}{N} \|\kappa_N^{-1} \mathbf{S}^{(N)} - \mathbf{P}^{(N)}\|. \end{aligned}$$

2. Note that Lemma 2 is formulated for graphon operators (i.e. linear integral operators with nonnegative kernels). An identical proof shows that the result holds also if the kernel assumes negative values.

Hence, to bound the norm of the difference between a random SBM graphon operator $\kappa_N^{-1} \mathbb{S}_N$ based on the graphon $\kappa_N^{-1} S = \kappa_N^{-1} \mathbf{1}(x)^T \mathbf{S}^{(N)} \mathbf{1}(y)$, with $S_{ij}^{(N)} = \text{Ber}(\kappa_N P_{ij}^{(N)})$, and its expectation \mathbb{W}_N defined via the graphon $W_N = \mathbf{1}(x)^T \mathbf{P}^{(N)} \mathbf{1}(y)$, we can employ matrix concentration inequalities. Specifically, we use [55, Theorem 1] in order to bound the deviations of $\|\mathbf{S}^{(N)} - \kappa_N \mathbf{P}^{(N)}\|$.

By Lemma 5, for N large enough, the maximum expected degree $\kappa_N C_N^d := \kappa_N \max_i (\sum_{j=1}^N P_{ij}^{(N)})$ of the random graph represented by $\mathbf{S}^{(N)}$ grows at least as $\frac{4}{9} \kappa_N \log(\frac{2N}{\delta})$. Consequently, all the conditions of [55, Theorem 1] are met and we get that with probability $1 - \delta - \delta'$

$$\begin{aligned} \|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}_N\| &= \frac{\kappa_N^{-1}}{N} \|\mathbf{S}^{(N)} - \kappa_N \mathbf{P}^{(N)}\| \\ &\leq \frac{\kappa_N^{-1}}{N} \sqrt{4\kappa_N C_N^d \log(2N/\delta)} \leq \sqrt{\frac{4\kappa_N^{-1} \log(2N/\delta)}{N}}, \end{aligned}$$

where we used that $C_N^d \leq N$ since each element in $\mathbf{P}^{(N)}$ belongs to $[0, 1]$.

Proof of (28). We finally show that (27) implies almost sure convergence. We start by restating (27) as

$$\Pr \left[\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \leq \sqrt{\frac{4\kappa_N^{-1} \log(2N/\delta)}{N}} + \rho(N) \right] \geq 1 - 2\delta. \quad (34)$$

Further, pick any $\gamma > 0$ and define the infinite sequence of events

$$\mathcal{E}_N := \{ \|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \geq \gamma + \rho(N) \},$$

for each $N \geq 1$. From (34) it follows that $\Pr[\mathcal{E}_N] \leq 4N \exp(-\kappa_N N \gamma^2 / 4)$. Consequently, if $\kappa_N = \frac{1}{N^\tau}$ with $\tau \in [0, 1)$, then:

$$\sum_{N=1}^{\infty} \Pr[\mathcal{E}_N] \leq 4 \sum_{N=1}^{\infty} N \exp\left(-\frac{\kappa_N N \gamma^2}{4}\right) < \infty$$

and by the Borel-Cantelli lemma there exists a positive integer N_γ such that for all $N \geq N_\gamma$, the complement of \mathcal{E}_N , i.e., $\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \leq \gamma + \rho(N)$, holds almost surely. To see that $\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \rightarrow 0$ almost surely we follow the ensuing argument. For any given deterministic sequence $\{a_N\}_{N=1}^{\infty}$ the fact that for each $\gamma > 0$ there is a positive integer N_γ such that for all $N \geq N_\gamma$, $|a_N| \leq \gamma + \rho(N)$ implies that $a_N \rightarrow 0$. In fact for all $\epsilon > 0$, if we set $\gamma = \epsilon/2$ and $N_\epsilon := \max\{N_\gamma, N_\rho\}$ (where N_ρ is the smallest N such that $\rho(N) \leq \epsilon/2$) then we get that for all $N > N_\epsilon$, $\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \leq \epsilon$. Hence, we can conclude that $\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \rightarrow 0$ almost surely. \square

The previous theorem provides us with convergence rates for the graphon operators. Based on these we are able to show a similar convergence result for centrality measures of graphons with a simple dominant eigenvalue.

Assumption 2 (Simple dominant eigenvalue). *Let the eigenvalues of \mathbb{W} be ordered such that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$ and assume that $\lambda_1 > \lambda_2$.*

We note that in most empirical studies degeneracy of the dominant eigenvalue is not observed, justifying the above assumption. A noteworthy exception in which a non-unique dominant eigenvalue may arise is if the graph consists of multiple components. In this case, however, one can treat each component separately.

For the proof in case of PageRank we will make the following additional assumption on the graphon.

Assumption 3 (Minimal degree assumption). *There exists $\eta > 0$ such that $W(x, y) \geq \eta$ for all $x, y \in [0, 1]$.*

Note that while this assumption is not fulfilled, e.g., for a SBM graphon with a block of zero connection probability, it can be further relaxed to accommodate such cases as well. However, to simplify the proof and avoid additional technical details, we invoke Assumption 3 in the following theorem.

Theorem 2. (Convergence of centrality measures) *The following statements hold:*

1) For any $N > 1$, the centrality functions $c_N(x)$ and $\hat{c}_N(x)$ corresponding to the operators \mathbb{W}_N and $\kappa_N^{-1}\mathbb{S}_N$, respectively, are in one to one relation with the centrality measures $\mathbf{c}_{\bar{\mathbf{P}}(N)} \in \mathbb{R}^N$, $\mathbf{c}_{\bar{\mathbf{S}}(N)} \in \mathbb{R}^N$ of the graphs with rescaled adjacency matrices $\bar{\mathbf{P}}^{(N)} := \frac{1}{N}\mathbf{P}^{(N)}$ and $\bar{\mathbf{S}}^{(N)} := \frac{1}{N\kappa_N}\mathbf{S}^{(N)}$, via the formula^a

$$c_N(x) = \mathbf{1}_N(x)^T \mathbf{c}_{\bar{\mathbf{P}}(N)}, \quad \hat{c}_N(x) = \mathbf{1}_N(x)^T \mathbf{c}_{\bar{\mathbf{S}}(N)}.$$

2) Under Assumptions 1, 2 and (for PageRank) 3, and N sufficiently large, with probability at least $1 - \delta'$

$$\|c_N - c\| \leq C\rho(N)$$

for some constant C and $\rho(N)$, δ' defined as in Theorem 1.

3) Under Assumptions 1, 2 and (for PageRank) 3, and N sufficiently large, with probability at least $1 - \delta$

$$\|\hat{c}_N - c\| \leq C' \left(\sqrt{\frac{4\kappa_N^{-1} \log(2N/\delta)}{N}} + \rho(N) \right),$$

for some constant C' .

4) Under Assumptions 1, 2 and (for PageRank) 3, if $\kappa_N = \frac{1}{N^\tau}$ with $\tau \in [0, 1)$, then:

$$\lim_{N \rightarrow \infty} \|\hat{c}_N - c\| = 0, \text{ almost surely.}$$

a. Note that $\bar{\mathbf{P}}^{(N)}, \bar{\mathbf{S}}^{(N)}$ belong to $\mathbb{R}_{\geq 0}^{N \times N}$ as opposed to $\{0, 1\}^{N \times N}$. Nonetheless, the definitions of centrality measures given in Section 2.1 can be extended to the continuous interval case in a straightforward manner.

Proof. 1) Follows immediately from Lemma 4 since \mathbb{W}_N and $\kappa_N^{-1}\mathbb{S}_N$ are the operators of the graphons corresponding to $\mathbf{P}^{(N)}$ and $\kappa_N^{-1}\mathbf{S}^{(N)}$, respectively.

2) We showed in Theorem 1 that, under Assumption 1, $\|\mathbb{W}_N - \mathbb{W}\| \leq \rho(N)$ with probability $1 - \delta'$. This fact can be exploited to prove convergence of the centrality measures c_N to c . All the subsequent statements hold with probability $1 - \delta'$.

For degree centrality: $c_N(x) = (\mathbb{W}_N \mathbf{1}_{[0,1]})(x)$ and $c(x) = (\mathbb{W} \mathbf{1}_{[0,1]})(x)$. Since $\|\mathbf{1}_{[0,1]}\| = 1$ we get

$$\|c_N - c\| = \|(\mathbb{W}_N - \mathbb{W})\mathbf{1}_{[0,1]}\| \leq \|\mathbb{W}_N - \mathbb{W}\| \leq \rho(N). \quad (35)$$

For eigenvector centrality: Let $\{\lambda_k, \varphi_k\}_{k \geq 1}$, $\{\lambda_k^{(N)}, \varphi_k^{(N)}\}_{k \geq 1}$ be the ordered eigenvalues and eigenfunctions of \mathbb{W} and \mathbb{W}_N , respectively. Note that $|\lambda_1^{(N)} - \lambda_1| \leq \rho(N)$ since

$\lambda_1^{(N)} = \|\mathbb{W}_N\| \leq \|\mathbb{W}\| + \|\mathbb{W}_N - \mathbb{W}\| \leq \lambda_1 + \rho(N)$ and $\lambda_1 = \|\mathbb{W}\| \leq \|\mathbb{W}_N\| + \|\mathbb{W}_N - \mathbb{W}\| \leq \lambda_1^{(N)} + \rho(N)$. Furthermore, since by Assumption 2 we have that $\lambda_1 > \lambda_2$, there exists a large enough \bar{N} such that for all $N > \bar{N}$ it holds that $\lambda_1^{(N)} > \lambda_2$ and $|\lambda_1 - \lambda_2| > |\lambda_1^{(N)} - \lambda_1|$. Therefore, by Lemma 8 in Appendix B, we obtain

$$\|\varphi_1^{(N)} - \varphi_1\| \leq \frac{\sqrt{2} \|\mathbb{W}_N - \mathbb{W}\|}{|\lambda_1 - \lambda_2| - |\lambda_1^{(N)} - \lambda_1|}. \quad (36)$$

From the facts that $\lambda_1 \neq \lambda_2$ (by Assumption 2), $|\lambda_1^{(N)} - \lambda_1| \leq \rho(N)$, and $\|\mathbb{W}_N - \mathbb{W}\| \leq \rho(N)$, it follows that (36) implies that for $N > \bar{N}$

$$\|\varphi_1^{(N)} - \varphi_1\| \leq \frac{\sqrt{2}\rho(N)}{|\lambda_1 - \lambda_2| - \rho(N)} = O(\rho(N)). \quad (37)$$

For Katz centrality: Take any value of $\alpha < 1/\|\mathbb{W}\|$, so that $\mathbb{M}_\alpha = \mathbb{I} - \alpha\mathbb{W}$ is invertible and $c(x) = (\mathbb{M}_\alpha^{-1} \mathbf{1}_{[0,1]})(x)$ is well defined. Since $\|\mathbb{W}_N - \mathbb{W}\| \rightarrow 0$ as $N \rightarrow \infty$, there exists $N_\alpha > 0$ such that $\alpha < 1/\|\mathbb{W}_N\|$ for all $N > N_\alpha$. This implies that for any $N > N_\alpha$, $[\mathbb{M}_N]_\alpha := \mathbb{I} - \alpha\mathbb{W}_N$ is invertible and $c_N(x) = ([\mathbb{M}_N]_\alpha^{-1} \mathbf{1}_{[0,1]})(x)$ is well defined. Note that $\|\mathbb{W}_N - \mathbb{W}\| \leq \rho(N)$ implies $\|[\mathbb{M}_N]_\alpha - \mathbb{M}_\alpha\| = O(\rho(N))$. We now prove that

$$\|[\mathbb{M}_N]_\alpha^{-1} - \mathbb{M}_\alpha^{-1}\| = O(\rho(N)). \quad (38)$$

To this end, note that $L^2([0, 1])$ is a Hilbert space, the inverse operator \mathbb{M}_α^{-1} is bounded and for N large enough it holds $\|[\mathbb{M}_N]_\alpha - \mathbb{M}_\alpha\| < 1/\|\mathbb{M}_\alpha^{-1}\|$, since $\|[\mathbb{M}_N]_\alpha - \mathbb{M}_\alpha\| \rightarrow 0$. It then follows by [56, Theorem 2.3.5] with $L := \mathbb{M}_\alpha$, $M := [\mathbb{M}_N]_\alpha$ that

$$\|[\mathbb{M}_N]_\alpha^{-1} - \mathbb{M}_\alpha^{-1}\| \leq \frac{\|[\mathbb{M}_N]_\alpha - \mathbb{M}_\alpha\|}{1 - \|[\mathbb{M}_N]_\alpha - \mathbb{M}_\alpha\| \|\mathbb{M}_\alpha^{-1}\|} = O(\rho(N)),$$

thus proving (38). Finally, since $\|\mathbf{1}_{[0,1]}\| = 1$,

$$\begin{aligned} \|c_N - c\| &= \|[\mathbb{M}_N]_\alpha^{-1} \mathbf{1}_{[0,1]} - \mathbb{M}_\alpha^{-1} \mathbf{1}_{[0,1]}\| \\ &\leq \|[\mathbb{M}_N]_\alpha^{-1} - \mathbb{M}_\alpha^{-1}\| = O(\rho(N)). \end{aligned}$$

For PageRank centrality: Consider $\beta \in (0, 1)$ such that \mathbb{L}_β is invertible and $c^{\text{Pr}}(x)$ is well defined. Similar to the argument used to show (38) in the proof for Katz centrality, it suffices to show that $\|[\mathbb{L}_N]_\beta - \mathbb{L}_\beta\| = O(\rho(N))$. To show this note that under Assumption 3 it holds

$$\begin{aligned} c^d(x) &= \int_0^1 W(x, y) dy \geq \eta, \\ c_N^d(x) &= \int_0^1 W_N(x, y) dy \geq \eta. \end{aligned} \quad (39)$$

Hence $(c^d(x))^\dagger = (c^d(x))^{-1} \leq \frac{1}{\eta}$ and $(c_N^d(x))^\dagger = (c_N^d(x))^{-1} \leq \frac{1}{\eta}$. For any $f \in L^2([0, 1])$ such that $\|f\| = 1$,

$$\begin{aligned} \|[\mathbb{L}_N]_\beta f - \mathbb{L}_\beta f\| &= \beta \|\mathbb{W}_N (f \cdot (c_N^d)^{-1}) - \mathbb{W} (f \cdot (c^d)^{-1})\| \\ &\leq \beta \|(\mathbb{W}_N - \mathbb{W})(f \cdot (c^d)^{-1})\| + \beta \|\mathbb{W}_N (f \cdot (c_N^d)^{-1} - f \cdot (c^d)^{-1})\| \\ &\leq \beta \|\mathbb{W}_N - \mathbb{W}\| \|(c^d)^{-1}\| + \beta \|\mathbb{W}_N\| \|(c_N^d)^{-1} - (c^d)^{-1}\| \\ &\leq \beta \rho(N) \|(c^d)^{-1}\| + 2\beta \|\mathbb{W}\| \|(c_N^d)^{-1} - (c^d)^{-1}\|, \end{aligned} \quad (40)$$

where we used that for N large enough $\|\mathbb{W}_N\| \leq 2\|\mathbb{W}\|$. In (40), the notation $(c^d)^{-1}$ is used to denote the function that

takes values $(c^d(x))^{-1}$ and similarly for $(c_N^d)^{-1}$. Observe that equation (39) implies

$$\|(c^d)^{-1}\| \leq \frac{1}{\eta} \quad (41)$$

and

$$\begin{aligned} \|(c_N^d)^{-1} - (c^d)^{-1}\|^2 &= \int_0^1 \left((c_N^d(y))^{-1} - (c^d(y))^{-1} \right)^2 dy \\ &= \int_0^1 \left(\frac{c_N^d(y) - c^d(y)}{c_N^d(y)c^d(y)} \right)^2 dy \leq \frac{1}{\eta^4} \|c_N^d - c^d\|^2. \end{aligned} \quad (42)$$

Combining (35), (40), (41) and (42) yields the desired result

$$\|[L_N]_\beta f - \mathbb{L}_\beta f\| \leq \frac{\beta}{\eta} \left[1 + \frac{2}{\eta} \|\mathbb{W}\| \right] \rho(N) = O(\rho(N)). \quad (43)$$

3) It suffices to mimic the argument made above for $\|c_N - c\|$ adapting it for the case $\|\hat{c}_N - c\|$ by making use of (27). The proof is omitted to avoid redundancy.

4) By Theorem 1 we have that $\|\kappa_N^{-1} \mathbb{S}_N - \mathbb{W}\| \rightarrow 0$ almost surely. This means that the set of realizations $\{\tilde{\mathbb{S}}_N\}_{N=1}^\infty$ of $\{\mathbb{S}_N\}_{N=1}^\infty$ for which $\|\kappa_N^{-1} \tilde{\mathbb{S}}_N - \mathbb{W}\| \rightarrow 0$ has probability one. For each of these realizations it can be proven (exactly as in part 2)³ that

$$\lim_{N \rightarrow \infty} \|\tilde{c}_N - c\| = 0,$$

where $\tilde{c}_N(x)$ is the deterministic sequence of centrality measures associated with the realization $\{\tilde{\mathbb{S}}_N\}_{N=1}^\infty$. Consequently, $\Pr[\lim_{N \rightarrow \infty} \|\hat{c}_N - c\| = 0] = 1$ and $\lim_{N \rightarrow \infty} \|\hat{c}_N - c\| = 0$ almost surely. \square

To sum up, Theorem 2 shows that, on the one hand, the centrality functions of the finite-rank operators \mathbb{W}_N and $\kappa_N^{-1} \mathbb{S}_N$ can be computed by simple interpolation of the centrality vectors of the corresponding finite-size graphs with adjacency matrices $\mathbf{P}^{(N)}$ and $\kappa_N^{-1} \mathbf{S}^{(N)}$ (suitably rescaled). On the other hand, such centrality functions $c_N(x)$ and $\hat{c}_N(x)$ become better approximations of the centrality function $c(x)$ of the graphon W as N increases. As alluded above, the importance of this result derives from the fact that it establishes that the centrality functions here introduced are the appropriate limits of the finite centrality measures, thus validating the presented framework. We finally note that as immediate corollary of the previous theorem we get the following robustness result for the centrality measures of different realizations.

Corollary 1. Consider two graphs \mathbf{S}^{N_1} and \mathbf{S}^{N_2} sampled from a graphon W satisfying Assumptions 1 and 2. Assume without loss of generality that $N_1 \leq N_2$ and let $\mathbf{c}_{\mathbf{S}^{(N_i)}} \in \mathbb{R}^N$ be the centrality of the graphs with rescaled adjacency matrices $\tilde{\mathbf{S}}^{(N_i)} := \frac{1}{N_i \kappa_{N_i}} \mathbf{S}^{(N_i)}$, for $i \in \{1, 2\}$. Then for N sufficiently large, with probability at least $1 - 4\delta$

$$\begin{aligned} &\left\| \mathbf{1}_{N_1}(x)^T \mathbf{c}_{\tilde{\mathbf{S}}^{(N_1)}} - \mathbf{1}_{N_2}(x)^T \mathbf{c}_{\tilde{\mathbf{S}}^{(N_2)}} \right\| \\ &\leq 2C' \left(\sqrt{\frac{4\kappa_{N_1}^{-1} \log(2N_1/\delta)}{N_1}} + \rho(N_1) \right) \end{aligned}$$

3. In part 2(b) the rate of convergence of \mathbb{W}_N to \mathbb{W} was used. Nonetheless, the same statement holds under the less stringent condition $\|\mathbb{W}_N - \mathbb{W}\| \rightarrow 0$, since this is sufficient to prove that $\lambda_1^{(N)} \rightarrow \lambda_1$.

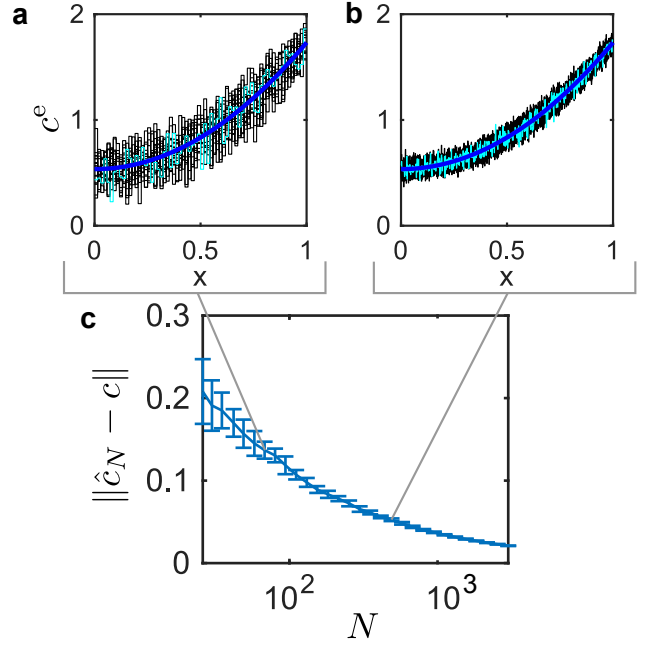


Fig. 6: Convergence of the eigenvector centrality function for the FR graphon in section 4.2.1. (a-b) Eigenvector centrality functions computed from a sampled graphon with deterministic latent variables (black; one realization shown in cyan for visualization purposes) and the eigenvector centrality function of the continuous graphon (blue). The examples shown correspond to a resolution of (a) $N = 68$ grid points, and (b) $N = 489$ grid points. In each case 20 realizations were drawn from the discretized graphon $\mathbf{P}^{(N)}$. (c) Convergence of the error $\|\hat{c}_N - c\|$ as a function of the number of grid points N , corresponding to the number of nodes in the sampled graph. For each point we plot the sample mean \pm one standard deviation.

for some constant C' and $\rho(N)$, δ defined as in Theorem 1.

To check our analytical results, we performed numerical experiments as illustrated in Figs. 6 and 7. In Fig. 6, we consider again the finite-rank graphon from our example in Section 4.2.1 and assess the convergence of the eigenvector centrality function \hat{c}_N^e from the sampled networks (with deterministic latent variables and $\kappa_N = 1$), to the true underlying graphon centrality measure c^e . As this graphon is smooth we have $K = 0$, i.e., there is only a single Lipschitz block, and we observe a smooth decay of the error when increasing the number of grid points N , corresponding to the number of nodes in the sampled graph.

For the stochastic block model graphon W_{SBM} from our example in Section 4.1.1, however, we have $K = 4$ and thus 25 Lipschitz blocks, which are delimited by discontinuous jumps in the value of the graphon. The effect of these jumps is clearly noticeable when assessing the convergence of the centrality measures, as illustrated in Fig. 7 for the example of Katz centrality. If the deterministic sampling grid of the discretized graphon \mathbb{W}_N is aligned with the structure of the stochastic block model \mathbb{W}_{SBM} , there is no mismatch introduced by the sampling procedure and thus the approximation error of the centrality measure c_N is smaller (also in the sampled version \hat{c}_N). Stated differently, if the $\frac{1}{N}$ -grid is exactly aligned with the Lipschitz blocks of the underlying graphon, we are effectively in a situation in which the area \mathcal{A}_N^c is zero, which is analogous to the case of $K = 0$ (see Fig. 5). In contrast, if there is a misalignment between the Lipschitz blocks and the $\frac{1}{N}$ -grid, then additional errors are

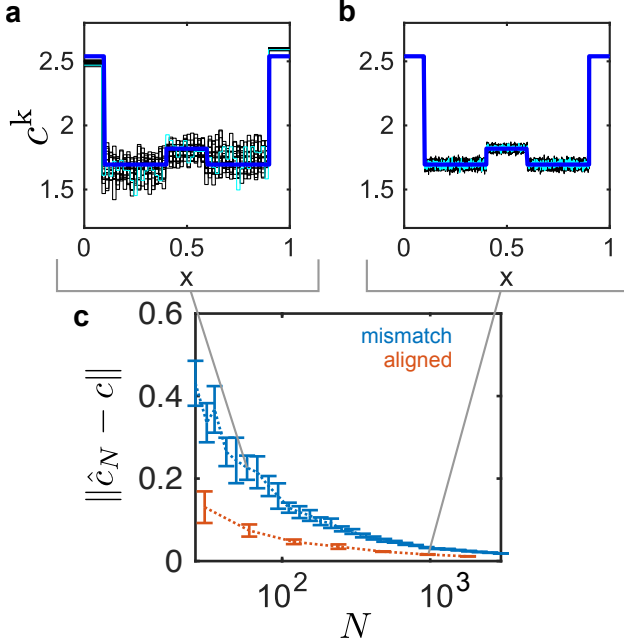


Fig. 7: Convergence of the Katz centrality function for the SBM graphon in section 4.1.1. **(a-b)** Katz centrality functions computed from a sampled graphon with deterministic latent variables (black; one realization shown in cyan for visualization purposes) and the Katz centrality function of the continuous graphon (blue). The examples shown correspond to a resolution of (a) $N = 58$ grid points, and (b) $N = 960$ grid points. In each case 20 realizations were drawn from the discretized graphon $\mathbf{P}^{(N)}$. **(c)** Convergence of the error $\|\hat{c}_N - c\|$ as a function of the number of grid points N , corresponding to the number of nodes in the sampled graph. For each point we plot the sample mean \pm one standard deviation. For visualization purposes we connect the data-points (red) in which the grid approximation was aligned with the piece-wise constant changes of the SBM (i.e. each grid point is within one Lipschitz block of the SBM graphon – c.f. Fig. 5). For this graphon, this happens for all N that are divisible by 10. Likewise, we connected those data points where there was a mismatch between the sampling grid and the SBM graphon structure (blue).

introduced leading to an overall slower convergence, which is consistent with our results above.

6 DISCUSSION AND FUTURE WORK

In many applications of centrality-based network analysis, the system of interest is subject to uncertainty. In this context, a desirable trait for a centrality measure is that the relative importance of agents should be impervious to random fluctuations contained in a particular realization of a network. In this paper, we formalized this intuition by extending the notion of centrality to graphons. More precisely, we proposed a departure from the traditional concept of centrality measures applied to deterministic graphs in favor of a graphon-based, probabilistic interpretation of centralities. Thus, we 1) introduced suitable definitions of centrality measures for graphons, 2) showed how such measures can be computed for specific classes of graphons, 3) proved that the standard centrality measures defined for graphs of finite size converge to our newly defined graphon centralities, and 4) bound the distance between graphon centrality function and centrality measures over sampled graphs.

The results presented here constitute a first step towards a systematic analysis of centralities in graphons and several questions remain unanswered. In particular, we see two main

challenges that need to be addressed to widen the scope of applications of our methods. First, in most practical scenarios the graphon will need to be estimated from (finite) data. The validity (error terms) of the centrality scores will accordingly be contingent on the errors made in this estimation [28], [53]. It would therefore be very interesting to explore how to best leverage existing graphon estimation results in order to estimate our proposed graphon centrality functions. Second, the parameter κ_N allows for the analysis of sparse networks as introduced in [26], [27], [43], [44], [45] but not for networks with asymptotically bounded degrees [46]. An extension of our results to the analysis of networks with finite degrees is thus of future interest.

Additionally, there are a number of other generalizations that are worth investigating.

First, the generalization of centralities for graphons beyond the four cases studied in this paper. In particular, the extension to centralities that do not rely directly on spectral graph properties, such as closeness and betweenness centralities, appears to be a challenging task. Indeed, a suitable notion of path for infinite-size networks needs to be defined first and bounds on the possible fluctuations of such a notion of path would need to be derived. Enriching the class of graphon centrality measures would also contribute to the characterization of the relative robustness of certain classes of centralities, which would allow us to better assess their utility for the analysis of uncertain networks in practice.

Second, the identification of classes of graphons (others than the ones here discussed) for which explicit and efficient formulas of the centrality functions can be derived.

Third, the determination of whether the convergence rates provided in Theorem 2 are also optimal. A related question in this context is to derive convergence results for the exact ordering of the nodes. This is in particular relevant for applications where we would like to know by how much the ranking of an individual node might have changed as a result of the uncertainty of the network. One possible avenue to tackle this kind of question would be to start investigating ℓ_∞ -norm bounds [57] for centralities, which enable us to control the maximal fluctuations of each individual entry of the centrality measures.

Finally, the extension of the centrality definitions from regular graphons to more complex objects, such as asymmetric networks or time-varying graphons [58], [59].

APPENDIX A : OMITTED PROOFS

Proof of Proposition 1

Proof. This proof is a consequence of Proposition 2. We notice that we can specialize the formula therein to the case of block stochastic models to obtain the relations: $\mathbf{h} = \mathbf{E}_{\text{SBM}} \mathbf{1}$, $\mathbf{Q} = \mathbf{Q}_{\text{SBM}}$, $\mathbf{E} = \mathbf{E}_{\text{SBM}}$, $\mathbf{E}_{\text{norm}} = \mathbf{E}_{\text{SBM}} \mathbf{D}_{\mathbf{E}}^{-1}$, $\mathbf{h}_{\text{nor}} = \mathbf{E}_{\text{SBM}} \mathbf{D}_{\mathbf{E}}^{-1} \mathbf{1}$. To see that these equivalencies are true one can check that, for instance:

$$\begin{aligned} [\mathbf{h}_{\text{nor}}]_i &= \sum_{j=1}^m \frac{P_{ij} [\mathbf{Q}_{\text{SBM}}]_{jj}}{\sum_{k=1}^m [\mathbf{Q}_{\text{SBM}}]_{kk} P_{kj}} = \sum_{j=1}^m \frac{P_{ij} [\mathbf{Q}_{\text{SBM}}]_{jj}}{\sum_{k=1}^m [\mathbf{Q}_{\text{SBM}}]_{kk} P_{jk}} \\ &= \sum_{j=1}^m \frac{[\mathbf{E}_{\text{SBM}}]_{ij}}{[\mathbf{D}_{\mathbf{E}}]_{jj}} = \sum_{j=1}^m [\mathbf{E}_{\text{SBM}} \mathbf{D}_{\mathbf{E}}^{-1}]_{ij}. \end{aligned}$$

From the above equivalences, the formulas for $c^d(x)$, $c^e(x)$ follow immediately. For $c_\alpha^k(x)$ we obtain

$$\begin{aligned} c_\alpha^k(x) &= 1 + \alpha \mathbf{1}(x)^T \left(\sum_{k=0}^{\infty} \alpha^k \mathbf{E}_{\text{SBM}}^k \right) \mathbf{E}_{\text{SBM}} \mathbf{1} \\ &= 1 + \mathbf{1}(x)^T \left(\sum_{k=1}^{\infty} \alpha^k \mathbf{E}_{\text{SBM}}^k \right) \mathbf{1} \\ &= \mathbf{1}(x)^T \left(\sum_{k=0}^{\infty} \alpha^k \mathbf{E}_{\text{SBM}}^k \right) \mathbf{1} = \mathbf{1}(x)^T (\mathbf{I} - \alpha \mathbf{E}_{\text{SBM}})^{-1} \mathbf{1}. \end{aligned}$$

Finally, $c_\beta^{\text{pr}}(x)$ can be proven similarly. \square

Proof of Lemma 3

Proof. Assume that \mathbf{v} is an eigenvector of \mathbf{E} such that $\mathbf{E}\mathbf{v} = \lambda\mathbf{v}$ with $\lambda \neq 0$. We now show that this implies that $\varphi(x) = \sum_{j=1}^m v_j g_j(x)$ is an eigenfunction of \mathbb{W} with eigenvalue λ . From an explicit computation of $(\mathbb{W}\varphi)(x)$ we have that

$$\begin{aligned} (\mathbb{W}\varphi)(x) &= \sum_{i=1}^m g_i(x) \int_0^1 h_i(y) \sum_{j=1}^m v_j g_j(y) dy \\ &= \sum_{i=1}^m g_i(x) \sum_{j=1}^m v_j \int_0^1 h_i(y) g_j(y) dy. \end{aligned}$$

Recalling the definition of \mathbf{E} from (18), it follows that

$$(\mathbb{W}\varphi)(x) = \sum_{i=1}^m g_i(x) \sum_{j=1}^m v_j E_{ij} = \sum_{i=1}^m g_i(x) \lambda v_i = \lambda \varphi(x),$$

where we used the fact that \mathbf{v} is an eigenvector of \mathbf{E} for the second equality.

In order to show the converse statement, let us assume that φ is an eigenfunction of \mathbb{W} with associated eigenvalue $\lambda \neq 0$. Then, we may write that

$$(\mathbb{W}\varphi)(x) = \sum_{i=1}^m g_i(x) \int_0^1 h_i(y) \varphi(y) dy = \lambda \varphi(x),$$

from where it follows that

$$\varphi(x) = \lambda^{-1} \sum_{i=1}^m g_i(x) \int_0^1 h_i(y) \varphi(y) dy = \sum_{i=1}^m g_i(x) v_i, \quad (44)$$

where we have implicitly defined $v_i := \lambda^{-1} \int_0^1 h_i(y) \varphi(y) dy$. Substituting (44) into this definition yields, for all $i = 1, \dots, m$,

$$\lambda v_i = \int_0^1 h_i(y) \sum_{j=1}^m g_j(y) v_j dy = \sum_{j=1}^m v_j E_{ij} = [\mathbf{E}\mathbf{v}]_i.$$

By writing the above equality in vector form we get that $\mathbf{E}\mathbf{v} = \lambda\mathbf{v}$, thus completing the proof. \square

Proof of Proposition 2

Proof. The proof for degree centrality follows readily from (7), i.e. $c^d(x) = \int_0^1 \mathbf{g}(x)^T \mathbf{h}(y) dy = \mathbf{g}(x)^T \int_0^1 \mathbf{h}(y) dy = \mathbf{g}(x)^T \mathbf{h}$. Based on Lemma 3, for c^e it is sufficient to prove that $\|c^e\| = 1$. To this end, note that

$$\begin{aligned} \|c^e\|^2 &= \frac{1}{\mathbf{v}_1^T \mathbf{Q} \mathbf{v}_1} \int_0^1 (\mathbf{g}(x)^T \mathbf{v}_1)^2 dx \\ &= \frac{1}{\mathbf{v}_1^T \mathbf{Q} \mathbf{v}_1} \int_0^1 (\mathbf{v}_1^T \mathbf{g}(x)) (\mathbf{g}(x)^T \mathbf{v}_1) dx \\ &= \frac{1}{\mathbf{v}_1^T \mathbf{Q} \mathbf{v}_1} \mathbf{v}_1^T \int_0^1 \mathbf{g}(x) \mathbf{g}(x)^T dx \mathbf{v}_1 = \frac{\mathbf{v}_1^T \mathbf{Q} \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{Q} \mathbf{v}_1} = 1. \end{aligned}$$

For Katz centrality, we first prove by induction that

$$(\mathbb{W}_{\text{FR}}^k f)(x) = \int_0^1 \mathbf{g}(x)^T \mathbf{E}^{k-1} \mathbf{h}(y) f(y) dy, \quad (45)$$

for all finite-rank operators \mathbb{W}_{FR} . The equality holds trivially for $k = 1$. Now suppose that it holds for $k - 1$, we can then compute

$$\begin{aligned} (\mathbb{W}_{\text{FR}}^k f)(x) &= (\mathbb{W}_{\text{FR}} \mathbb{W}_{\text{FR}}^{k-1} f)(x) \\ &= \int_0^1 \mathbf{g}(x)^T \mathbf{h}(z) \int_0^1 \mathbf{g}(z)^T \mathbf{E}^{k-2} \mathbf{h}(y) f(y) dy dz \\ &= \int_0^1 \mathbf{g}(x)^T \left(\int_0^1 \mathbf{h}(z) \mathbf{g}(z)^T dz \right) \mathbf{E}^{k-2} \mathbf{h}(y) f(y) dy \\ &= \int_0^1 \mathbf{g}(x)^T \mathbf{E}^{k-1} \mathbf{h}(y) f(y) dy, \end{aligned}$$

where we used Definition 5 for the last equality. We leverage (45) to compute c_α^k using the expression in Remark 1,

$$\begin{aligned} c_\alpha^k(x) &= 1 + \sum_{k=1}^{\infty} \alpha^k (\mathbb{W}^k \mathbf{1}_{[0,1]})(x) \\ &= 1 + \sum_{k=1}^{\infty} \alpha^k \int_0^1 \mathbf{g}(x)^T \mathbf{E}^{k-1} \mathbf{h}(y) dy \\ &= 1 + \sum_{k=1}^{\infty} \alpha^k \mathbf{g}(x)^T \mathbf{E}^{k-1} \mathbf{h} = 1 + \alpha \sum_{k=0}^{\infty} \alpha^k \mathbf{g}(x)^T \mathbf{E}^k \mathbf{h} \\ &= 1 + \alpha \mathbf{g}(x)^T \left(\sum_{k=0}^{\infty} \alpha^k \mathbf{E}^k \right) \mathbf{h} = 1 + \alpha \mathbf{g}(x)^T (\mathbf{I} - \alpha \mathbf{E})^{-1} \mathbf{h}, \end{aligned}$$

as we wanted to show.

Finally, the methodology to prove the result for PageRank is similar to the one used for Katz, thus we sketch the proof to avoid redundancy. First, we recall the definition of \mathbb{G} from the proof of Proposition 1 and use induction to show that for every finite-rank graphon $(\mathbb{G}^k \mathbf{1}_{[0,1]})(x) = \mathbf{g}(x)^T \mathbf{E}_{\text{nor}}^{k-1} \mathbf{h}_{\text{nor}}$ for all integer $k \geq 1$. We then compute the inverse in the definition of PageRank (10) via an infinite sum as done above for Katz but using the derived expression for $(\mathbb{G}^k \mathbf{1}_{[0,1]})(x)$. \square

REFERENCES

- [1] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, and N. Zhang, "Topological structure analysis of the protein-protein interaction network in budding yeast," *Nucleic acids research*, vol. 31, no. 9, pp. 2443–2450, 2003.
- [2] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.
- [3] A. Garas, P. Argyrakis, C. Rozenblat, M. Tomassini, and S. Havlin, "Worldwide spreading of economic crisis," *New Journal of Physics*, vol. 12, no. 11, p. 113043, 2010.
- [4] N. E. Friedkin, "Theoretical foundations for centrality measures," *American Journal of Sociology*, vol. 96, no. 6, pp. 1478–1504, 1991.
- [5] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.
- [6] —, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [7] P. Bonacich, "Power and centrality: A family of measures," *American journal of sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.
- [8] S. P. Borgatti and M. G. Everett, "A graph-theoretic perspective on centrality," *Social Networks*, vol. 28, no. 4, pp. 466–484, 2006.
- [9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, November 1999.
- [10] D. F. Gleich, "Pagerank beyond the web," *SIAM Review*, vol. 57, no. 3, pp. 321–363, 2015.
- [11] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, 2003, pp. 137–146.
- [12] E. Costenbader and T. W. Valente, "The stability of centrality measures when networks are sampled," *Social Networks*, vol. 25, no. 4, pp. 283–307, 2003.
- [13] S. P. Borgatti, K. M. Carley, and D. Krackhardt, "On the robustness of centrality measures under conditions of imperfect data," *Social Networks*, vol. 28, no. 2, pp. 124–136, 2006.
- [14] M. Benzi and C. Klymko, "On the limiting behavior of parameter-dependent network centrality measures," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 2, pp. 686–706, 2015.
- [15] S. Segarra and A. Ribeiro, "Stability and continuity of centrality measures in weighted graphs," *IEEE Transactions on Signal Processing*, vol. 64, no. 3, pp. 543–555, Feb 2016.
- [16] P. Balachandran, E. Airolidi, and E. Kolaczyk, "Inference of network summary statistics through network denoising," *arXiv preprint arXiv:1310.0423*, 2013.
- [17] K. Lerman, R. Ghosh, and J. H. Kang, "Centrality metric for dynamic networks," in *Proceedings of the Workshop on Mining and Learning with Graphs (MLG)*, 2010, pp. 70–77.
- [18] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks," *Physical Review E*, vol. 84, no. 1, p. 016105, 2011.

- [19] P. Grindrod and D. J. Higham, "A matrix iteration for dynamic network summaries," *SIAM Review*, vol. 55, no. 1, pp. 118–128, 2013.
- [20] K. Dasaratha, "Distributions of centrality on networks," *arXiv preprint arXiv:1709.10402*, 2017.
- [21] C. Borgs and J. T. Chayes, "Graphons: A nonparametric method to model, estimate, and design algorithms for massive networks," *arXiv:1706.01143*, 2017.
- [22] L. Lovász, *Large Networks and Graph Limits*. American Mathematical Society Providence, 2012, vol. 60.
- [23] L. Lovász and B. Szegedy, "Limits of dense graph sequences," *Journal of Combinatorial Theory, Series B*, vol. 96, no. 6, pp. 933–957, 2006.
- [24] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi, "Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing," *Advances in Mathematics*, vol. 219, no. 6, pp. 1801–1851, 2008.
- [25] —, "Convergent sequences of dense graphs II: Multiway cuts and statistical physics," *Annals of Mathematics*, vol. 176, no. 1, pp. 151–219, 2012.
- [26] P. J. Bickel and A. Chen, "A nonparametric view of network models and Newman–Girvan and other modularities," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21 068–21 073, 2009.
- [27] P. J. Bickel, A. Chen, and E. Levina, "The method of moments and degree distributions for network models," *The Annals of Statistics*, vol. 39, no. 5, pp. 2280–2301, 2011.
- [28] E. M. Airolidi, T. B. Costa, and S. H. Chan, "Stochastic blockmodel approximation of a graphon: Theory and consistent estimation," in *Advances in Neural Information Processing Systems*, 2013, pp. 692–700.
- [29] J. Yang, C. Han, and E. Airolidi, "Nonparametric estimation and testing of exchangeable graph models," in *Artificial Intelligence and Statistics*, 2014, pp. 1060–1067.
- [30] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.
- [31] T. A. Snijders and K. Nowicki, "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," *Journal of Classification*, vol. 14, no. 1, pp. 75–100, 1997.
- [32] S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," in *International Workshop on Algorithms and Models for the Web-Graph*. Springer, 2007, pp. 138–149.
- [33] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinite relational model," in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2006, pp. 381–388.
- [34] A. Goldenberg, A. X. Zheng, S. E. Fienberg, E. M. Airolidi et al., "A survey of statistical network models," *Foundations and Trends in Machine Learning*, vol. 2, no. 2, pp. 129–233, 2010.
- [35] M. W. Morency and G. Leus, "Signal processing on kernel-based random graphs," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Kos island, Greece, 2017.
- [36] C. E. Lee and D. Shah, "Unifying framework for crowd-sourcing via graphon estimation," *arXiv preprint arXiv:1703.08085*, 2017.
- [37] S. Gao and P. E. Caines, "The control of arbitrary size networks of linear systems via graphon limits: An initial investigation," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 1052–1057.
- [38] Y. Zhang, E. Levina, and J. Zhu, "Estimating network edge probabilities by neighbourhood smoothing," *Biometrika*, vol. 104, no. 4, pp. 771–783, 2017.
- [39] P. Orbanz and D. M. Roy, "Bayesian models of graphs, arrays and other exchangeable random structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 2, pp. 437–461, 2015.
- [40] C. Borgs, J. Chayes, and L. Lovász, "Moments of two-variable functions and the uniqueness of graph limits," *Geometric And Functional Analysis*, vol. 19, no. 6, pp. 1597–1619, 2010.
- [41] L. Lovász and B. Szegedy, "The automorphism group of a graphon," *Journal of Algebra*, vol. 421, pp. 136–166, 2015.
- [42] P. Diaconis and S. Janson, "Graph limits and exchangeable random graphs," *arXiv preprint arXiv:0712.2749*, 2007.
- [43] C. Borgs, J. Chayes, and A. Smith, "Private graphon estimation for sparse graphs," in *Advances in Neural Information Processing Systems*, 2015, pp. 1369–1377.
- [44] C. Gao, Y. Lu, Z. Ma, and H. Zhou, "Optimal estimation and completion of matrices with biclustering structures," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 5602–5630, 2016.
- [45] O. Klopp, A. Tsybakov, and N. Verzelen, "Oracle inequalities for network models and sparse graphon estimation," *The Annals of Statistics*, vol. 45, no. 1, pp. 316–354, 2017.
- [46] V. Veitch and D. M. Roy, "The class of random graphs arising from exchangeable random measures," *arXiv preprint arXiv:1512.03099*, 2015.
- [47] D. J. Aldous, "Representations for partially exchangeable arrays of random variables," *Journal of Multivariate Analysis*, vol. 11, no. 4, pp. 581–598, 1981.
- [48] D. N. Hoover, "Relations on probability spaces and arrays of random variables," *Preprint, Institute for Advanced Study, Princeton, NJ*, vol. 2, 1979.
- [49] K. Deimling, *Nonlinear Functional Analysis*. Dover publications, 1985.
- [50] P. H. Bezandry and T. Diagana, *Almost Periodic Stochastic Processes*. Springer Science & Business Media, 2011.
- [51] P. Latouche and S. Robin, "Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models," *Statistics and Computing*, vol. 26, no. 6, pp. 1173–1185, Nov 2016.
- [52] P. J. Wolfe and S. C. Olhede, "Nonparametric graphon estimation," *arXiv preprint arXiv:1309.5936*, 2013.
- [53] C. Gao, Y. Lu, and H. H. Zhou, "Rate-optimal graphon estimation," *The Annals of Statistics*, vol. 43, no. 6, pp. 2624–2652, 2015.
- [54] B. Szegedy, "Limits of kernel operators and the spectral regularity lemma," *European Journal of Combinatorics*, vol. 32, no. 7, pp. 1156–1167, 2011.
- [55] F. Chung and M. Radcliffe, "On the spectra of general random graphs," *the electronic journal of combinatorics*, vol. 18, no. 1, p. 215, 2011.
- [56] K. Atkinson and W. Han, *Theoretical Numerical Analysis: A Functional Analysis Framework*. Springer, 2009, vol. 39.
- [57] J. Fan, W. Wang, and Y. Zhong, "An l-infinity eigenvector perturbation bound and its application to robust covariance estimation," *arXiv preprint arXiv:1603.03516*, 2016.
- [58] H. Crane, "Dynamic random networks and their graph limits," *The Annals of Applied Probability*, vol. 26, no. 2, pp. 691–721, 2016.
- [59] M. Pensky, "Dynamic network models and graphon estimation," *arXiv preprint arXiv:1607.00673*, 2016.
- [60] R. M. Dudley, *Real Analysis and Probability*. Cambridge University Press, 2002, vol. 74.
- [61] S. Janson, *Graphons, cut norm and distance, couplings and rearrangements*. NYJM Monographs, 2013.
- [62] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation," *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.
- [63] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *arXiv preprint arXiv:1707.09350v3*, 2018.
- [64] S. Boucheron and M. Thomas, "Concentration inequalities for order statistics," *Electronic Communications in Probability*, vol. 17, 2012.
- [65] L. Devroye, *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.
- [66] J. B. Conway, *A Course In Functional Analysis*. Springer Science & Business Media, 1997, vol. 96.
- [67] F. Sauvigny, *Partial Differential Equations 2: Functional Analytic Methods*. Springer Science & Business Media, 2012.



Marco Avella-Medina holds a B.A. degree in Economics (2009), a M.Sc. degree in Statistics (2011) and a Ph.D. in Statistics (2016) from the University of Geneva, Switzerland. From August 2016 to June 2018, he was a postdoctoral researcher with the Statistics and Data Science Center, and the Sloan School of Management at the Massachusetts Institute of Technology. Since July 2018, he is an Assistant Professor in the Department of Statistics at Columbia University. His research interests include robust statistics, high-dimensional statistics and statistical machine learning.



Francesca Parise was born in Verona, Italy, in 1988. She received the B.Sc. and M.Sc. degrees (cum Laude) in Information and Automation Engineering from the University of Padova, Italy, in 2010 and 2012, respectively. She conducted her master thesis research at Imperial College London, UK, in 2012. She graduated from the Galilean School of Excellence, University of Padova, Italy, in 2013. She defended her PhD at the Automatic Control Laboratory, ETH Zurich, Switzerland in 2016 and she is currently

a Postdoctoral researcher at the Laboratory for Information and Decision Systems, M.I.T., USA. Her research focuses on identification, analysis and control of complex systems, with application to distributed multi-agent networks, game theory and systems biology.



Michael T. Schaub obtained a B.Sc. in Electrical Engineering and Information Technology from ETH Zurich (2009), an M.Sc. in Biomedical Engineering (2010) from Imperial College London, and a Ph.D. in Applied Mathematics (2014) from Imperial College. After a Postdoctoral stay at the Université catholique de Louvain and the University of Namur (Belgium), he has been a Postdoctoral Researcher at the Institute of Data, Systems and Society (IDSS) at the Massachusetts Institute of Technology since November 2016. Presently,

he is a Marie-Sklodowska Curie Fellow at IDSS and the Department of Engineering Science, University of Oxford, UK. He is broadly interested in interdisciplinary applications of applied mathematics in engineering, social and biological systems. His research interest include in particular network theory, data science, machine learning, and dynamical systems.



Santiago Segarra (M'16) received the B.Sc. degree in industrial engineering with highest honors (Valedictorian) from the Instituto Tecnológico de Buenos Aires (ITBA), Argentina, in 2011, the M.Sc. in electrical engineering from the University of Pennsylvania (Penn), Philadelphia, in 2014 and the Ph.D. degree in electrical and systems engineering from Penn in 2016. From September 2016 to June 2018 he was a postdoctoral research associate with the Institute for Data, Systems, and Society at the Massachusetts Institute

of Technology. Since July 2018, Dr. Segarra is an Assistant Professor in the Department of Electrical and Computer Engineering at Rice University. His research interests include network theory, data analysis, machine learning, and graph signal processing. Dr. Segarra received the ITBA's 2011 Best Undergraduate Thesis Award in industrial engineering, the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina, the 2017 Penn's Joseph and Rosaline Wolf Award for Best Doctoral Dissertation in electrical and systems engineering as well as four best conference paper awards.

APPENDIX B: ADDITIONAL RESULTS

Invariance of centrality measures under permutations

Just as the topology of a graph is invariant with respect to relabelings or permutations of its nodes, graphons are defined only up to measure preserving transformations. We show in the next lemma that the linear operator \mathbb{W}^π associated with any such ‘permutation’ π (formalized via a measure preserving transformation) of a graphon W shares the same eigenvalues of \mathbb{W} and ‘permuted’ eigenfunctions.

Lemma 6. *Consider the graphon $W^\pi(x, y) := W(\pi(x), \pi(y))$ obtained by transforming W using the measure preserving function $\pi : [0, 1] \rightarrow [0, 1]$. Let \mathbb{W} and \mathbb{W}^π be the associated linear integral operators. If (λ, φ) is an eigenvalue-eigenfunction pair of \mathbb{W} , then $(\lambda, \varphi \circ \pi)$ is an eigenvalue-eigenfunction pair of \mathbb{W}^π .*

Proof. From a direct computation we obtain that

$$\begin{aligned} (\mathbb{W}^\pi(\varphi \circ \pi))(x) &= \int_0^1 W^\pi(x, y) \varphi(\pi(y)) dy \\ &= \int_0^1 W(\pi(x), \pi(y)) \varphi(\pi(y)) dy = \int_0^1 W(\pi(x), y) \varphi(y) dy \\ &= (\mathbb{W}\varphi)(\pi(x)) = \lambda \varphi(\pi(x)). \end{aligned}$$

The third equality uses the fact that π is a measure preserving transformation and the ergodic theorem [60, Ch. 8]. \square

Lemma 6 complements the discussion at the end of Section 2.2 by showing the effect of measure preserving transformations on the spectral properties of the graphon.

Discussion on related graphon convergence results

We introduce some additional definitions in order to compare our work with previous results on graphon convergence. In particular, let us start by introducing the *cut norm* which is typically used for the statement of graphon convergence results. For a graphon W in the graphon space \mathcal{W} , the cut norm is denoted by $\|W\|_\square$ and is defined as

$$\|W\|_\square := \sup_{U, V} \left| \int_U \int_V W(x, y) dx dy \right|,$$

where U and V are measurable subsets of $[0, 1]$. The *cut metric* between two graphons $W, W' \in \mathcal{W}$ is

$$d_\square(W, W') := \inf_{\phi \in \Pi_{[0,1]}} \|W^\phi - W'\|_\square,$$

where $W^\phi(x, y) := W(\phi(x), \phi(y))$ and $\Pi_{[0,1]}$ is the class of measure preserving permutations $\phi : [0, 1] \mapsto [0, 1]$. Intuitively, the function ϕ performs a node relabeling to find the best match between W and W' . Because of such relabeling, $d_\square(W, W')$ is not a well defined metric in \mathcal{W} since we might have that $d_\square(W, W') = 0$ even if $W \neq W'$. To avoid such a problem, we define the space \mathcal{W} as the space where we identify graphons up to measure preserving transformations, so that d_\square is a well defined metric in \mathcal{W} . It can be shown that the metric space (\mathcal{W}, d_\square) is complete [22]. The following lemma is instrumental for our comparison to previous work, as it establishes the equivalence between the $L_{p,q}$ norms and the cut norm. Recall that for any $p, q \geq 1$, the $L_{p,q}$ operator norm is defined as $\|\mathbb{W}\|_{p,q} := \sup_{f \in L_p, \|f\|_p=1} \|\mathbb{W}f\|_q$.

Lemma 7. [61] *For any $W \in \mathcal{W}$ and all $p, 1 \in [1, \infty]$*

$$\|W\|_\square \leq \|\mathbb{W}\|_{p,q} \leq \sqrt{2}(4\|W\|_\square)^{\min(1-1/p, 1/q)}.$$

In particular, for $p = q = 2$ we get

$$\|W\|_\square \leq \|\mathbb{W}\| \leq \sqrt{8\|W\|_\square}.$$

It follows from Lemma 7 that convergence of the graphon operator can be deduced from previous works establishing graphon convergence in cut norm. For example one can use the results of [25], [54] combined with Lemma 7 to easily conclude that the graphon operator converges in operator norm. However taking this approach one typically does not obtain rates of convergence of the sampled graph’s operator to the graphon operator. A convergence rate is instead provided in [22, Lemma 10.16] for general graphons. We next show that, for graphons satisfying Assumption 1 with $K = 0$, the result in [22, Lemma 10.16] leads to a slower rate of convergence than the one provided in Theorem 1. More precisely, combining [22, Lemma 10.16] and Lemma 7, we get that with probability at least $1 - \exp(-\frac{N}{2 \log N})$ it holds

$$\|\mathbb{S}_N - \mathbb{W}\| \leq \sqrt{176}/(\log N)^{1/4}. \quad (46)$$

By defining $\delta = \exp(-\frac{N}{2 \log N})$, the bound provided in Eq. (27) (for $\kappa_N = 1$ and $K = 0$) leads to

$$\|\mathbb{S}_N - \mathbb{W}\| \leq \mathcal{O}\left(1/(\log N)^{1/2}\right), \quad (47)$$

thus proving faster convergence. Finally, we note that the bounds provided in Theorem 1 are not only tighter but also more flexible. In fact, by introducing the parameter δ we establish a trade off between sharper error bounds and the probability that such bounds hold, as typically done in concentration inequality results.

A useful variant of the Davis-Kahan theorem

The following technical lemma is used to prove the convergence of the eigenvector centrality for graphons and is a consequence of the Davis-Kahan $\sin \theta$ theorem for compact operators in Hilbert space [62].

Lemma 8. *Consider two linear integral operators \mathbb{L} and $\hat{\mathbb{L}}$, with ordered eigenvalues $\{\lambda_k\}_{k \geq 1}$, $\{\hat{\lambda}_k\}_{k \geq 1}$. Let $\hat{\varphi}_1, \varphi_1$ be the eigenfunctions associated with the dominant eigenvalues $\hat{\lambda}_1$ and λ_1 (normalized to norm one) and suppose that $|\lambda_1 - \lambda_2| > |\hat{\lambda}_1 - \lambda_1|$. Then*

$$\|\hat{\varphi}_1 - \varphi_1\| \leq \frac{\sqrt{2} \|\hat{\mathbb{L}} - \mathbb{L}\|}{|\lambda_1 - \lambda_2| - |\hat{\lambda}_1 - \lambda_1|}. \quad (48)$$

The proof can be found in [63].

Concentration of uniform order statistics

The goal of this subsection is to derive a uniform deviation bound for order statistics sampled from a standard uniform distribution, as detailed in Proposition 3 in the main text. This result is required in the proof of Theorem 1. Although it is intuitive to expect subgaussian deviations for uniform order statistics, we could not find the desired statement explicitly in the literature and believe it could be of interest on its own right. From a technical point of view, the key ingredient in our argument is to use the exponential Efron-Stein inequality derived in [64].

Let $U_1, \dots, U_N \sim \text{Unif}(0, 1)$ and define their correspondent order statistics $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(N)}$ and spacings

$\Delta_i = U_{(i)} - U_{(i-1)}$ for $i = 1, \dots, N+1$ with the convention $U_{(0)} = 0$ and $U_{(N+1)} = 1$. It is shown in [65] that

- Each $U_{(i)}$ is distributed according to $Beta(i, N+1-i)$ and thus has mean $\frac{i}{N+1}$;
- The joint survival function of the spacings is

$$\mathbb{P}(\Delta_1 > s_1, \dots, \Delta_{N+1} > s_{N+1}) = \left(1 - \sum_{i=1}^{N+1} s_i\right)_+^N.$$

Consequently the spacings are identically (but not independently) distributed with cumulative distribution $F_\Delta(s) = 1 - (1-s)^N$ and marginal density $f_\Delta(s) = N(1-s)^{N-1}$.

The following lemma is a key intermediate step in the derivation of concentration inequalities for order statistics drawn from a uniform distribution.

Lemma 9. *For any $\lambda \geq 0$ it holds*

$$\log \mathbb{E} e^{\lambda|U_{(i)} - \mathbb{E}U_{(i)}|} \leq \lambda \frac{N}{2} \mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)].$$

Proof. We show this result by first proving

- (a) $\log \mathbb{E} e^{\lambda(U_{(i)} - \mathbb{E}U_{(i)})} \leq \lambda \frac{N-i+1}{2} \mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)];$
- (b) $\log \mathbb{E} e^{\lambda(\mathbb{E}U_{(i)} - U_{(i)})} \leq \lambda \frac{i}{2} \mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)].$

To this end, note that the hazard rate of a uniform distribution is increasing since it has the form $h(x) = \frac{1}{1-x}$. Therefore applying Theorem 2.9 of [64] shows (a). Note that therein the result is proven only for $i \geq N/2 + 1$ (equivalently in the notation of [64] $k := N+1-i \leq N/2$) but such condition on i is never used in the proof. Indeed, in order to prove claim (2.1) of Theorem 2.9 in [64] one only needs to show that

$$\text{Ent}[e^{\lambda X_{(k)}}] \leq k \mathbb{E}[e^{\lambda X_{(k+1)}} \psi(\lambda(X_{(k)} - X_{(k+1)}))] \quad (49)$$

for $1 \leq k \leq N$, where $\text{Ent}[Y] = \mathbb{E}[Y \log Y] - \mathbb{E}[Y] \log \mathbb{E}[Y]$ is the entropy of a non-negative random variable Y . This follows easily from the arguments of Proposition 2.3 in [64]. Note that the authors only consider $k := N+1-i \leq N/2$ in this proposition because for $k > N/2$ the bound can be improved.

Let us now turn to the proof of (b). Note that the beta distribution is reflection symmetric i.e. if $X \sim Beta(\alpha, \beta)$ then $1-X \sim Beta(\beta, \alpha)$ for $\alpha, \beta > 0$. Therefore $U_{(i)} \sim 1 - U_{(N-i+1)}$ and $\mathbb{E}U_{(i)} - U_{(i)} \sim U_{(N-i+1)} - \mathbb{E}U_{(N-i+1)}$. Hence by (a) we have that

$$\begin{aligned} \log \mathbb{E} e^{\lambda(\mathbb{E}U_{(i)} - U_{(i)})} &= \log \mathbb{E} e^{\lambda(U_{(N-i+1)} - \mathbb{E}U_{(N-i+1)})} \\ &\leq \lambda \frac{N-(N-i+1)+1}{2} \mathbb{E}[\Delta_{N-i+1}(e^{\lambda\Delta_{N-i+1}} - 1)] \\ &= \lambda \frac{i}{2} \mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)], \end{aligned}$$

where in the last step we used the fact that the spacings $\Delta_1, \dots, \Delta_N$ have the same marginal distribution. Finally, the statement of the lemma is an immediate consequence of (a) and (b). \square

Lemma 10. *Suppose that $N > 5$ and $\delta_i \in (e^{-N/5}, e^{-1})$ then, for $i = 1, \dots, N$, with probability at least $1 - \delta_i$*

$$\left|U_{(i)} - \frac{i}{N+1}\right| \leq \sqrt{\frac{8 \log(1/\delta_i)}{(N+1)}}.$$

Proof. We note that by Chernoff's inequality

$$\mathbb{P}\left(\left|U_{(i)} - \frac{i}{N+1}\right| > t\right) \leq \frac{\mathbb{E}[e^{\lambda|U_{(i)} - \frac{i}{N+1}|}]}{e^{\lambda t}} \quad (50)$$

and from Lemma 9 we see that

$$\mathbb{E}[e^{\lambda|U_{(i)} - \frac{i}{N+1}|}] \leq e^{\lambda \frac{N}{2} \mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)]}. \quad (51)$$

From the marginal density of the spacings we get

$$\mathbb{E}[\Delta_i^k] = N \int_0^1 (s_i)^k (1-s_i)^{N-1} ds_i = N \mathcal{B}(k+1, N) = \frac{k!N!}{(N+k)!},$$

where we used the definition of the beta function $\mathcal{B}(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt = \frac{(x-1)!(y-1)!}{(x+y-1)!}$ for integers x, y . Then

$$\begin{aligned} \mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)] &= \mathbb{E}[\Delta_i(\sum_{k=0}^{\infty} \frac{(\lambda\Delta_i)^k}{k!} - 1)] \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}[\Delta_i^k] = \sum_{k=1}^{\infty} \frac{\lambda^k (k+1)N!}{(N+k+1)!} \\ &\leq \sum_{k=1}^{\infty} \frac{\lambda^k (k+1)N!}{N^k (N+1)N!} = \frac{1}{N+1} \sum_{k=1}^{\infty} \left(\frac{\lambda}{N}\right)^k (k+1) \\ &= \frac{1}{N+1} \left[\sum_{k=0}^{\infty} \left(\frac{\lambda}{N}\right)^k (k+1) - 1\right] = \frac{1}{N+1} \left[\frac{1}{(1-\frac{\lambda}{N})^2} - 1\right] \\ &= \frac{1}{N+1} \left[\frac{1-(1-\frac{\lambda}{N})^2}{(1-\frac{\lambda}{N})^2}\right] = \frac{1}{N+1} \left[\frac{1-(1-\frac{\lambda}{N})^2 + 2\frac{\lambda}{N}}{(1-\frac{\lambda}{N})^2}\right] \\ &= \frac{\lambda}{N(N+1)} \left[\frac{2-\frac{\lambda}{N}}{(1-\frac{\lambda}{N})^2}\right] \end{aligned} \quad (52)$$

where we used $\sum_{k=0}^{\infty} \alpha^k (k+1) = \frac{1}{(1-\alpha)^2}$ (obtained by differentiating the geometric sum for $\frac{\lambda}{N} < 1$). If we set $\lambda < 0.35N$ then

$$\mathbb{E}[\Delta_i(e^{\lambda\Delta_i} - 1)] \leq \frac{4\lambda}{N(N+1)}$$

since $y < \frac{7-\sqrt{49-32}}{8} \approx 0.36$ implies $\left[\frac{2-y}{(1-y)^2}\right] < 4$. Combining (50), (51) and (52) yields

$$\mathbb{P}\left(\left|U_{(i)} - \frac{i}{N+1}\right| > t\right) \leq e^{\frac{2\lambda^2}{N+1}} e^{-\lambda t}.$$

Minimizing over λ leads to the choice $\lambda = \frac{t(N+1)}{4}$ and thus

$$\mathbb{P}\left(\left|U_{(i)} - \frac{i}{N+1}\right| > t\right) \leq \exp\left(-\frac{t^2(N+1)}{8}\right).$$

The proof is concluded if we select $t = \sqrt{\frac{8 \log(1/\delta_i)}{(N+1)}}$. Note that for this choice

$$\lambda = \frac{t(N+1)}{4} = \sqrt{\frac{8 \log(1/\delta_i)}{(N+1)}} \frac{(N+1)}{4} = \sqrt{\frac{\log(1/\delta_i)(N+1)}{2}}.$$

We need to verify that $\lambda < 0.35N$ or equivalently that $2(0.35)^2 N^2 - \log(1/\delta_i)N - \log(1/\delta_i) > 0$. A sufficient condition is $N > \frac{\log(1/\delta_i) + \sqrt{\log(1/\delta_i)^2 + 8 \log(1/\delta_i)(0.35)^2}}{4(0.35)^2} =: \bar{N}$.

Note that $\bar{N} < \frac{1+\sqrt{1+8(0.35)^2}}{4(0.35)^2} \log(1/\delta_i) < 5 \log(1/\delta_i)$, since $\log(1/\delta_i) > 1$ for $\delta_i < e^{-1}$. Hence a simpler sufficient condition is $N > 5 \log(1/\delta_i)$. \square

Proof of Proposition 3:

From Lemma 10 we know that for each $i = 1, \dots, N$ if we set $\delta_i = \frac{\delta}{N}$ then with probability at least $1 - \frac{\delta}{N}$

it holds $G_i := \left| U_{(i)} - \frac{i}{N+1} \right| \leq \sqrt{\frac{8 \log(N/\delta)}{(N+1)}} =: t$. It then follows from the union bound that $\mathbb{P}\left(\bigcap_{1 \leq i \leq N} \{G_i \leq t\}\right) = \left(1 - \mathbb{P}\left(\bigcup_{1 \leq i \leq N} \{G_i > t\}\right)\right) \geq 1 - \sum_{i=1}^N \mathbb{P}(G_i > t) \geq 1 - \sum_{i=1}^N \frac{\delta}{N} = 1 - \delta$.

Proof of Lemma 5: A lower bound on the maximum expected degree

Using the definition of W_N and the reverse triangle inequality yields

$$\begin{aligned} \frac{1}{N} C_N^d &= \frac{1}{N} \max_i \left(\sum_{j=1}^N P_{ij}^{(N)} \right) = \frac{1}{N} \max_i \left(\sum_{j=1}^N W(u_i, u_j) \right) \\ &= \max_{x \in [0,1]} \left(\int_0^1 W_N(x, y) dy \right) \geq \max_{x \in \mathcal{C}_N^c} \left(\int_0^1 W_N(x, y) dy \right) \\ &\geq \max_{x \in \mathcal{C}_N^c} \left(\int_0^1 W(x, y) dy - \int_0^1 |D(x, y)| dy \right), \end{aligned} \quad (53)$$

where $\mathcal{C}_N := \{x \in [0, 1] \mid \exists k \in \{1, \dots, K\} \text{ s.t. } |x - \alpha_k| \leq d_N\}$ is the subset of points in $[0, 1]$ that are up to d_N close to a discontinuity. Note that for any $x \in \mathcal{C}_N^c$, with probability $1 - \delta'$ (see part 1 of Theorem 1)

$$\begin{aligned} \int_0^1 |D(x, y)| dy &= \int_{\mathcal{C}_N^c} |D(x, y)| dy + \int_{\mathcal{C}_N} |D(x, y)| dy \\ &\leq 2Ld_N + \text{Area}(\mathcal{C}_N) = 2Ld_N + 2Kd_N. \end{aligned}$$

Substituting in (53) we get

$$\frac{1}{N} C_N^d \geq \max_{x \in \mathcal{C}_N^c} \left(\int_0^1 W(x, y) dy \right) - 2(L + K)d_N.$$

Finally, note that Assumption 1 implies that the degree $c^d(x)$ is piece-wise Lipschitz continuous, that is, for any $k \in \{1, \dots, K+1\}$ and any $x, x' \in \mathcal{I}_k$ it holds $|c^d(x) - c^d(x')| \leq L|x - x'|$. If $\Delta_{\text{MIN}}^{(\alpha)} > 2d_N$ this implies that

$$\left| \max_{x \in \mathcal{C}_N^c} \left(\int_0^1 W(x, y) dy \right) - C^d \right| \leq Ld_N,$$

since there must be at least one point in \mathcal{C}_N^c which belongs to the same Lipschitz block as $\text{argmax } c^d(x)$ and has distance d_N from it. Overall, we have proven

$$C_N^d \geq NC^d - N(3L + 2K)d_N \geq \log\left(\frac{2N}{\delta}\right) \geq \frac{4}{9} \log\left(\frac{2N}{\delta}\right).$$

APPENDIX C: MATHEMATICAL BACKGROUND

For completeness, we provide a self-contained review of the mathematical tools required in the proofs of our results. The subsection on bounded linear operators is a condensed overview of concepts detailed in, e.g., [49], [66], [67]. The subsection on perturbation theory introduces concepts necessary for a formal statement of the $\sin \theta$ theorem of [62] in the case of compact operators.

Bounded linear operators in Hilbert space

Let us start by introducing some basic notions regarding linear operators in metric spaces.

Definition 8. Let \mathcal{X}, \mathcal{Y} be normed linear spaces and let $\mathbb{L} : \mathcal{X} \rightarrow \mathcal{Y}$ be a linear operator.

- (a) \mathbb{L} is continuous at a point $f \in \mathcal{X}$ if $f_n \rightarrow f$ in \mathcal{X} implies $\mathbb{L}f_n \rightarrow \mathbb{L}f$ in \mathcal{Y} .
- (b) \mathbb{L} is continuous if it is continuous at every point, i.e. if $f_n \rightarrow f$ in \mathcal{X} implies $\mathbb{L}f_n \rightarrow \mathbb{L}f$ in \mathcal{Y} for every f .
- (c) \mathbb{L} is bounded if there exists a finite $M \geq 0$ such that, for all $f \in \mathcal{X}$,

$$\|\mathbb{L}f\| \leq M\|f\|.$$

Note that $\|Lf\|$ is the norm of Lf in \mathcal{Y} , while $\|f\|$ is the norm of f in \mathcal{X} .

- (d) The operator norm of \mathbb{L} is $\|\mathbb{L}\| := \sup_{\|f\|=1} \|\mathbb{L}f\|$.
- (e) We let $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ denote the set of all bounded linear operators mapping \mathcal{X} into \mathcal{Y} , that is

$$\mathcal{B}(\mathcal{X}, \mathcal{Y}) = \{\mathbb{L} : \mathcal{X} \rightarrow \mathcal{Y} \mid \mathbb{L} \text{ is bounded and linear}\}.$$

If $\mathcal{X} = \mathcal{Y}$ we write $\mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathcal{X}, \mathcal{X})$

The following Proposition shows that (a), (b) and (c) are equivalent.

Proposition 4. Let $\mathbb{L} : \mathcal{X} \rightarrow \mathcal{Y}$ be a linear operator. Then the following conditions are equivalent.

- (a) \mathbb{L} is continuous at every point of \mathcal{X} .
- (b) \mathbb{L} is continuous at $0 \in \mathcal{X}$.
- (c) $\|\mathbb{L}f\|$ is bounded on the unit ball $\{f \in \mathcal{X} \mid \|f\| \leq 1\}$.

Let us now focus on linear operators acting on Hilbert spaces.

Proposition 5. (Adjoint) Let $\mathbb{L} \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, where \mathcal{X} and \mathcal{Y} are Hilbert spaces. Then there exists a unique bounded linear map $\mathbb{L}^* : \mathcal{Y} \rightarrow \mathcal{X}$ such that

$$\langle \mathbb{L}x, y \rangle = \langle x, \mathbb{L}^*y \rangle \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}.$$

Definition 9. Let \mathcal{X} be a Hilbert space and $\mathbb{L} \in \mathcal{B}(\mathcal{X})$.

- (a) \mathbb{L} is self-adjoint if $\mathbb{L} = \mathbb{L}^*$ i.e. $\langle \mathbb{L}x, y \rangle = \langle x, \mathbb{L}y \rangle$ for all $x, y \in \mathcal{X}$.
- (b) \mathbb{L} is compact if it maps the unit ball in \mathcal{X} to a set with compact closure.

We are now ready to state the spectral theorem for compact operators.

Theorem 1. (Spectral theorem, [67, Theorem 2, Chapter 8 §7]) Let the $\mathbb{L} : \mathcal{X} \rightarrow \mathcal{X}$ be a compact self-adjoint operator on the Hilbert space \mathcal{X} satisfying $\mathbb{L} \neq 0$. Then we have a finite or countably infinite system of orthonormal elements $\{\varphi_k\}_{k \geq 1}$ in \mathcal{X} such that

- (a) The elements φ_k are eigenfunctions associated with the eigenvalues $\lambda_k \in \mathbb{R}$, i.e.

$$\mathbb{L}\varphi_k = \lambda_k \varphi_k, \quad k = 1, 2, \dots$$

If the set of nonzero eigenvalues is infinite, then 0 is the unique accumulation point.

(b) The operator \mathbb{L} has the representation

$$\mathbb{L} = \sum_{k \geq 1} \lambda_k \varphi_k \varphi_k^* \quad \text{i.e.} \quad \mathbb{L}f = \sum_{k \geq 1} \lambda_k \langle \varphi_k, f \rangle \varphi_k \quad \text{for all } f \in \mathcal{X}.$$

The following useful result shows that linear integral operators are compact.

Proposition 6. ([66, Chapter 2, Proposition 4.7]) If $K \in L^2([0, 1]^2)$, then $(\mathbb{K}f)(x) = \int_0^1 K(x, y)f(y)dy$ is a compact operator.

We conclude this subsection with a generalization of the Perron-Frobenius theorem to linear operators in Hilbert space. Let us first introduce some additional notions used in the statement of the result. A closed convex set $K \subset \mathcal{X}$ is called a *cone* if $\lambda K \subset K$ for all $\lambda \geq 0$ and $K \cap (-K) = \{0\}$. If the set $\{u - v : u, v \in K\}$ is dense in \mathcal{X} , then K is called a *total cone*.

Theorem 2. (Krein-Rutman theorem, [49, Theorem 19.2]) Let \mathcal{X} be a Hilbert space, $K \subset \mathcal{X}$ a total cone and $\mathbb{L} : \mathcal{X} \rightarrow \mathcal{X}$ a compact linear operator that is positive (i.e. $\mathbb{L}(K) \subset K$) with positive spectral radius $r(\mathbb{L})$. Then $r(\mathbb{L})$ is an eigenvalue with an eigenvector $\varphi \in K \setminus \{0\} : \mathbb{L}\varphi = r(\mathbb{L})\varphi$.

Perturbation theory for compact self-adjoint operators

The natural definition of the angle between two nonzero vectors φ and $\tilde{\varphi}$ in a Hilbert space \mathcal{X} is the number

$$\Theta(\varphi, \tilde{\varphi}) = \cos^{-1} \left(\frac{\langle \varphi, \tilde{\varphi} \rangle}{\|\varphi\| \|\tilde{\varphi}\|} \right). \quad (54)$$

Note that the above concept is well defined because of Cauchy-Schwartz inequality. Consider now the two subspaces spanned by the two nonzero vectors φ and $\tilde{\varphi}$, that is $[\varphi] := \varphi \varphi^* \mathcal{X}$ and $[\tilde{\varphi}] := \tilde{\varphi} \tilde{\varphi}^* \mathcal{X}$. One can extend (54) to define an angle between the two subspaces $[\varphi]$ and $[\tilde{\varphi}]$ as

$$\Theta([\varphi], [\tilde{\varphi}]) := \inf_{u, v} \left\{ \Theta(u, v) : u \in [\varphi], v \in [\tilde{\varphi}] \right\}.$$

More generally, one can extend this definition of angle to subspaces spanned by eigenfunctions. This will be particularly useful in situations where we are interested in a compact self-adjoint operator \mathbb{L} but we only have access to a modified operator $\tilde{\mathbb{L}} = \mathbb{L} + \mathbb{H}$. Indeed, in this case one way to measure how close these operators are is to measure the angle between subspaces spanned by their eigenfunctions. Let us introduce some notation in order to formalize this. We write the subspace (eigenspace) spanned by the eigenfunctions $\{\varphi_k\}_{k=1}^m$ of \mathbb{L} by $[\mathbb{E}_0] := [\varphi_1 \ \varphi_2 \ \dots \ \varphi_m]$. We denote the projector of $[\mathbb{E}_0]$ by $\mathbb{P}_0 = \mathbb{E}_0 \mathbb{E}_0^* = \sum_{k=1}^m \varphi_k \varphi_k^*$ and its complementary projector by $\mathbb{P}_1 = \mathbb{E}_1 \mathbb{E}_1^*$. Now any vector $x \in \mathcal{X}$ can be written as

$$x = (\mathbb{E}_0 \ \mathbb{E}_1) \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \mathbb{E}_0 x_0 + \mathbb{E}_1 x_1,$$

where $x_0 = \mathbb{E}_0^* x$ and $x_1 = \mathbb{E}_1^* x$. We therefore say that x is represented by $\begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$. The corresponding notation for an operator $\mathbb{L} : \mathcal{X} \rightarrow \mathcal{X}$ is

$$\mathbb{L} = (\mathbb{E}_0 \ \mathbb{E}_1) \begin{pmatrix} \mathbb{L}_0 & 0 \\ 0 & \mathbb{L}_1 \end{pmatrix} \begin{pmatrix} \mathbb{E}_0^* \\ \mathbb{E}_1^* \end{pmatrix} = \mathbb{E}_0 \mathbb{L}_0 \mathbb{E}_0^* + \mathbb{E}_1 \mathbb{L}_1 \mathbb{E}_1^*,$$

where $\mathbb{L}\mathbb{E}_0 = \mathbb{E}_0 \mathbb{L}_0$ and $\mathbb{L}\mathbb{E}_1 = \mathbb{E}_1 \mathbb{L}_1$. Similarly, we can consider the eigenspace $[\mathbb{F}_0]$ spanned by the eigenfunctions $\{\tilde{\varphi}_k\}_{k=1}^m$ of $\tilde{\mathbb{L}} + \mathbb{H}$ and write

$$\tilde{\mathbb{L}} = \mathbb{L} + \mathbb{H} = \mathbb{F}_0 \tilde{\mathbb{L}}_0 \mathbb{F}_0^* + \mathbb{F}_1 \tilde{\mathbb{L}}_1 \mathbb{F}_1^*.$$

The problem of measuring the closeness between the eigenspaces $[\mathbb{E}_0] = \mathbb{P}_0 \mathcal{X}$ and $[\mathbb{F}_0] = \tilde{\mathbb{P}}_0 \mathcal{X}$ can be tackled by looking at the angle between these subspaces. To do so, we can define a diagonal operator Θ_0 using the principal angles between \mathbb{E}_0 and \mathbb{F}_0 , i.e., $(\cos^{-1}(s_1) \ \dots \ \cos^{-1}(s_m))$ where $s_1 \geq \dots \geq s_m$ are the singular values of $\mathbb{E}_0^* \mathbb{F}_0$, or equivalently the square-root of the nonzero eigenvalues of $\mathbb{E}_0^* \mathbb{F}_0 \mathbb{F}_0^* \mathbb{E}_0$. Then, writing $\mathbb{S} := \text{diag}(\cos^{-1}(s_1) \ \dots \ \cos^{-1}(s_m))$, we can define $\Theta_0 = \Theta(\mathbb{E}_0, \mathbb{F}_0)$ as

$$\Theta_0 = \cos^{-1}(\mathbb{S}) \quad \text{i.e.} \quad \Theta_0 f = \sum_{k=1}^m \cos^{-1}(s_k) \langle \phi_k, f \rangle \phi_k$$

for all $f \in \mathcal{X}$ and any basis $\{\phi_k\}_{k=1}^\infty$. We are now ready to state the Davis-Kahan $\sin \theta$ theorem.

Theorem 3. (Davis-Kahan $\sin \theta$ theorem [62]) Let \mathbb{L} and $\tilde{\mathbb{L}} = \mathbb{L} + \mathbb{H}$ be two self-adjoint operators acting on the Hilbert space \mathcal{X} such that $\mathbb{L} = \mathbb{E}_0 \mathbb{L}_0 \mathbb{E}_0^* + \mathbb{E}_1 \mathbb{L}_1 \mathbb{E}_1^*$ and $\mathbb{L} + \mathbb{H} = \mathbb{F}_0 \tilde{\mathbb{L}}_0 \mathbb{F}_0^* + \mathbb{F}_1 \tilde{\mathbb{L}}_1 \mathbb{F}_1^*$ with $[\mathbb{E}_0, \mathbb{E}_1]$ and $[\mathbb{F}_0, \mathbb{F}_1]$ orthogonal. If the eigenvalues of \mathbb{L}_0 are contained in an interval (a, b) , and the eigenvalues of $\tilde{\mathbb{L}}_1$ are excluded from the interval $(a - \delta, b + \delta)$ for some $\delta > 0$, then

$$\|\sin \Theta(\mathbb{E}_0, \mathbb{F}_0)\| = \|\mathbb{F}_1^* \mathbb{E}_0\| \leq \frac{\|\mathbb{F}_1^* \mathbb{H} \mathbb{E}_0\|}{\delta}$$

for any unitarily invariant operator norm $\|\cdot\|$.

Note that the above theorem holds even for non-compact operators. Indeed, one might consider more general orthogonal subspaces defined through their projectors, which in turn might not be written as countable sums of the product of the elements of an orthogonal basis [62].