On Constant-Size Graphs That Preserve the Local Structure of High-Girth Graphs*

Hendrik Fichtenberger[1], Pan Peng[1,2], and Christian Sohler[1]

1 **Department of Computer Science, TU Dortmund, Germany**
`{firstname.lastname}@tu-dortmund.de`

2 **State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, China**

Abstract

Let $G = (V, E)$ be an undirected graph with maximum degree d. The k-disc of a vertex $v \in V$ is defined as the rooted subgraph that is induced by all vertices whose distance to v is at most k. The k-disc frequency vector of G, $\mathrm{freq}_k(G)$, is a vector indexed by all isomorphism types of k-discs. For each such isomorphism type Γ, the k-disc frequency vector counts the fraction of vertices that have k-disc isomorphic to Γ. Thus, the frequency vector $\mathrm{freq}_k(G)$ of G captures the *local structure* of G. A natural question is whether one can construct a much smaller graph H such that H has a similar local structure. N. Alon proved that for any $\epsilon > 0$ there always exists a graph H whose size is independent of $|V|$ and whose frequency vector satisfies $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \le \epsilon$. However, his proof is only existential and neither gives an explicit bound on the size of H nor an efficient algorithm. He gave the open problem to find such explicit bounds [9]. In this paper, we solve this problem for the special case of high girth graphs. We show how to efficiently compute a graph H with the above properties when G has girth at least $2k+2$ and we give explicit bounds on the size of H.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases local graph structure, k-disc frequency vector, graph property testing

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2015.786

1 Introduction

Given a graph $G = (V, E)$, the problem to find a smaller graph H that approximates G with respect to some of its properties is a basic problem in the area of graph algorithms. For example, spanner graphs [4] approximate G with respect to the shortest path structure, combinatorial sparsifiers [2] approximate G with respect to the cut structure, spectral sparsifiers [14] approximate G with respect to the spectral structure, and for a dense graph G the regularity lemma [15] may be thought of as providing a constant size weighted graph that captures an important part of the combinatorial structure of G.

In this paper we consider a different type of approximation. We study the problem of constructing a small graph H that has approximately the same *local structure* as G, where G is assumed to be undirected and to have a maximum degree bounded by d. The motivation to consider such an approximation is that any algorithm that only uses local information will behave similarly on inputs G and H. This is, for example, interesting in the context of property testing in the bounded degree graph model introduced by Goldreich and Ron [7],

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 307696.

18th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'15) / 19th Int'l Workshop on Randomization and Computation (RANDOM'15).
Editors: Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim; pp. 786–799
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

where we are given oracle access to the adjacency lists of a graph G with maximum degree d and the goal is to distinguish graphs with a given property Π from graphs that are ϵ-far from Π, that is, graphs that have to be changed in more than $\epsilon d|V|$ edges to obtain a graph with property Π. It is known that many constant time property testers in this model depend only on the local structure of the input graph. For example, all minor-closed properties can be tested in this way [3, 8]. If one allows the property testing algorithm to be (non-uniformly) depending on n (and ϵ and d) then every hyperfinite property is testable [13], that is, every graph property that contains only graphs that can be partitioned into connected components of constant size by removing an ϵ-fraction of the edges.

We now continue to make the problem more precise. Given a vertex $v \in V$, the k-disc of v is defined as the rooted subgraph that is induced by all vertices whose distance to v is at most k. The k-disc frequency vector of G, $\mathrm{freq}_k(G)$, is an L-dimensional vector indexed by all isomorphism types of k-discs, where L is the number of such isomorphism types. For each isomorphism type Γ, the k-disc frequency vector counts the fraction of vertices that have k-disc Γ. In other words, $\mathrm{freq}_k(G)$ is the frequency distribution of local neighborhoods over the vertices of G. Given G with maximum degree d and a parameter ϵ our problem is to compute a smaller graph H such that $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \epsilon$, where $\mathrm{freq}_k(G), \mathrm{freq}_k(H)$ denote the frequency vectors of G and H, respectively.

1.1 Previous Work

There is a surprisingly simple proof by Alon showing that for every $\epsilon > 0$ and constants d and k, there is an $M(\epsilon)$ such that for every d-bounded degree graph G, there is a graph H of size $M(\epsilon)$ such that the ℓ_1-norm distance of $\mathrm{freq}_k(G)$ and $\mathrm{freq}_k(H)$ is bounded by ϵ (see [11, Proposition 19.10] for the proof). In other words, for every d-bounded graph G of arbitrary size, there exists a small graph H of constant size that approximates the local neighborhood distribution of G. This result may be viewed as an analogue to a weak version of Szemerédi's regularity lemma for dense graphs [15] (see [10, Section 5.5] for more details).

The proof by Alon is based on a compactness argument and does not give explicit bound on $M(\epsilon)$. Obtaining such a bound was suggested by Alon as an open problem [9].

The problem is also related to the theory of graph limits and may be viewed as a finite version of the Aldous-Lyons conjecture [1]. A special case of this conjecture was solved by Elek [5]: He proved that every involution-invariant probability measure on the space of d-bounded trees arises as the local limit of some (infinite) sequence of d-bounded graphs.

1.2 Our Results

In this paper, we give a bound on $M(\epsilon)$ for the special case when the input graph has high girth, where the girth of a graph G is defined as the length of the shortest cycle in G. In other words, we focus on the class of graphs where all k-discs are trees. This class contains some very interesting graphs already. For example, it is known that a random regular graph with high girth is an expander graph with high probability (cf. [6, 12]).

We develop an algorithm that, given oracle access to a graph G with maximum degree d, computes in constant time and with a constant number of queries a small graph H such that $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \epsilon$. Here, a query asks for the adjacency list of a vertex $v \in V(G)$.

► **Theorem 1.** *Let $d \geq 2$, $k \geq 1$, $\epsilon, \delta \in (0,1)$ and define $\varphi := \frac{300 d^{3k+2} L^3}{\epsilon^2 \delta}$. Let $G = (V, E)$ be a d-bounded degree graph of size $|G| \geq 2\varphi^2/\delta$ with $\mathrm{girth}(G) \geq 2k+2$. Then, there is an* *algorithm that outputs, with probability* $1-\delta$*, a graph* H *such that*

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \le \epsilon \text{ and } |V(H)| \le \varphi\,.$$

The algorithm has time and query complexity $\mathcal{O}(1)$ *for constant* d*,* k*,* ϵ *and* δ*.*

If we allow the algorithm to be less efficient (but deterministic), the size of H can be reduced by a factor of L^2/ϵ.

▶ **Theorem 2.** *Let* $d \ge 2$*,* $k \ge 1$*,* $\epsilon \in (0,1)$ *and let* $G = (V,E)$ *be a* d*-bounded graph with* $\mathrm{girth}(G) \ge 2k+2$*. Then, there is a deterministic algorithm that outputs a graph* H *such that*

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \le \epsilon \text{ and } |V(H)| \le 36\frac{d^{3k+2}L}{\epsilon}\,.$$

The algorithm has time complexity $\mathcal{O}(|V(G)|)$*.*

We remark that our results can be directly generalized to graphs that are close to having high girth. For any $\varepsilon > 0$ and integer k, two d-bounded graphs G and G' are called to be *ε-close* to each other if one can obtain G' by inserting/deleting at most εdn edges to/from G. By noting that the ℓ_1-norm distance of the frequency vectors of two graphs G and G' is small if they are close to each other, we have the following corollary.

▶ **Corollary 3.** *Let* $d \ge 2$*,* $k \ge 1$*,* $\epsilon \in (0,1)$ *and let* $G = (V,E)$ *be a* d*-bounded graph that is* $\frac{\varepsilon}{6d^{k+1}}$*-close to some graph* G' *with* $\mathrm{girth}(G') \ge 2k+2$*. Then, there exists a graph* H *of size at most* $72\frac{d^{3k+2}L}{\epsilon}$ *such that* $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \le \epsilon$*.*

1.3 Proof Overview and Techniques

Our result is based on the following transformation of a graph G that fully preserves the local structure of G: Let $(u_1,v_2),(u_2,v_1) \in E$ be two edges with the properties that (a) the distance from u_1 to v_1 and the distance from v_2 to u_2 in G are large and (b) the local neighborhoods of u_1 and u_2 are isomorphic and (c) the local neighborhoods of v_2 and v_1 are isomorphic. Then one can replace the edges $(u_1,v_2),(u_2,v_1)$ by $(u_1,v_1),(u_2,v_2)$ without changing the local structure of the graph. We believe that this local transformation might be also interesting in the context of lower bounds in property testing, since if we consider sufficiently large local neighborhoods, the behavior of any constant-query property testing algorithm does not change under this transformation.

Our algorithm now works as follows. We use random sampling to identify a subset $U \subseteq V$ of constant size that has approximately the same distribution of neighborhoods (with respect to G) as V. Then we use our transformation to turn G into a graph G' where U has a small cut (relative to the size of U) to $V \setminus U$ and the neighborhood distribution of G is preserved. Then the graph $G'[U]$ has constant size and a similar distribution of neighborhoods as G.

2 Preliminaries

Let $G = (V,E)$ be an undirected graph. We will assume G to be a *d-bounded degree* graph, that is, the maximum degree of a vertex in G is upper bounded by d. Throughout the paper, d is assumed to be a constant. Given two vertices $u,v \in V$, let $\mathrm{dt}_G(u,v)$ be the length of the shortest path between u and v. The girth of G, $\mathrm{girth}(G)$, is defined as the length of the shortest cycle in G. The cut of $V_1,V_2 \subseteq V$ where $V_1 \cap V_2 = \emptyset$ is defined as $E \cap \{(u,v) \mid u \in V_1 \wedge v \in V_2\}$.

For any $v \in V$, the *k-disc* of v, denoted by $\mathrm{disc}_k(G, v)$, is defined as the subgraph that is induced by the vertices that are at distance at most k to v and is rooted at v. Two k-discs are isomorphic if and only if there exists a *root-preserving graph isomorphism*, that is, a graph isomorphism that identifies the roots. For any two k-discs Γ' and Γ'', we write $\Gamma' \simeq \Gamma''$ if Γ' is isomorphic to Γ'', and write $\Gamma' \not\simeq \Gamma''$ otherwise. We denote the number of all non-isomorphic d-bounded degree rooted graphs with radius at most k (that is, k-discs) by $L := L(d, k)$. The set of all such graphs is denoted by $\mathcal{T}_k = (\Gamma_1, \ldots, \Gamma_L)$. Since G has d-bounded degree, the size of each of its k-discs $\Gamma_i \in \mathcal{T}_k$ is bounded by $1 + d + \ldots + d^k \leq 3d^k/2$.

▶ **Fact 4.** *The size of a k-disc $\Gamma \in \mathcal{T}_k$ is at most $3d^k/2$.*

The *k-disc count vector* $\mathrm{cnt}_k(G)$ of a graph G is an L-dimensional vector where the i-th entry counts the number of k-discs in G that are isomorphic to $\Gamma_i \in \mathcal{T}_k$. By Fact 4, L is finite. Note that the total number of k-discs in G is exacly $|V(G)|$. Given a k-disc isomorphism type Γ, $\mathrm{cnt}_k(G)_\Gamma$ is defined as the entry in $\mathrm{cnt}_k(G)$ that corresponds to Γ. Given a subset of vertices $S \subseteq V$, let $\mathrm{cnt}_k(S \mid G)$ be the k-disc count vector such that the ith entry counts the number of k-discs of G with root vertex in S that are isomorphic to Γ_i.

The *k-disc frequency vector* of G, denoted by $\mathrm{freq}_k(G)$, is the vector where the i-th entry counts the fraction of k-discs in G that are isomorphic to $\Gamma_i \in \mathcal{T}_k$, or equivalently, $\mathrm{freq}_k(G) := \mathrm{cnt}_k(G)/|V(G)|$. We define $\mathrm{freq}_k(S \mid G) := \mathrm{cnt}_k(S \mid G)/|S|$ and $\mathrm{freq}_k(G)_\Gamma := \mathrm{freq}_k(G)_\Gamma/|V(G)|$ similarly.

In the following, we consider both k-discs and $(k-1)$-discs. We use Γ to denote k-discs and Δ to denote $(k-1)$-discs.

For any integer k and $\Gamma', \Gamma'' \in \mathcal{T}_k$, we call an edge $(u, v) \in E$ a *(Γ', Γ'')-edge* if $\mathrm{disc}_k(G, u) \simeq \Gamma'$ and $\mathrm{disc}_k(G, v) \simeq \Gamma''$. For any two subsets $V_1, V_2 \subseteq V$ and any two k-disc types Γ', Γ'', we let $\mathrm{e}(\Gamma', \Gamma'' \,|\, V_1, V_2)$ denote the number of (Γ', Γ'')-edges from V_1 to V_2, that is, the number of edges (u, v) such that $u \in V_1$, $v \in V_2$, $\mathrm{disc}_k(G, u) \simeq \Gamma'$ and $\mathrm{disc}_k(G, v) \simeq \Gamma''$.

For any k-disc $\Gamma \in \mathcal{T}_k$ and $(k-1)$-disc $\Delta \in \mathcal{T}_{k-1}$, Γ is called *Δ-extensive* if the $(k-1)$-disc of the root of Γ is isomorphic to Δ. We denote the set of all Δ-extensive k-discs Γ by $\mathrm{ext}(\Delta)$. Given a k-disc $\Gamma \in \mathcal{T}_k$ with root r and a $(k-1)$-disc $\Delta \in \mathcal{T}_{k-1}$, let $\mathrm{neigh}(\Gamma, \Delta)$ be the number of neighbors of r whose $(k-1)$-disc is isomorphic to Δ, that is,

$$\mathrm{neigh}(\Gamma, \Delta) := \left|\{v \mid (r, v) \in E(\Gamma) \wedge \mathrm{disc}_{k-1}(\Gamma, v) \simeq \Delta\}\right|.$$

For any $\Delta_1, \Delta_2 \in \mathcal{T}_{k-1}$, let $\mathrm{neigh}_\Sigma(\Delta', \Delta'')$ be the total number of (Δ', Δ'')-edges starting at the root of any k-disc $\Gamma \in \mathrm{ext}(\Delta')$, that is, $\mathrm{neigh}_\Sigma(\Delta', \Delta'') := \sum_{\Gamma \in \mathrm{ext}(\Delta')} \mathrm{neigh}(\Gamma, \Delta'')$. Note that $\mathrm{neigh}_\Sigma(\cdot, \cdot)$ is not necessarily symmetric. Since $|\mathrm{ext}(\Delta')| \leq |\mathcal{T}_k| \leq L$ and for every $\Gamma \in \mathrm{ext}(\Delta')$, its root's degree is at most d, we get the following bound.

▶ **Fact 5.** *For every pair of $(k-1)$-discs Δ', $\Delta'' \in \mathcal{T}_{k-1}$, we have* $\mathrm{neigh}_\Sigma(\Delta', \Delta'') \leq Ld$.

3 Rewiring Edges

In this section, we show that for every partitioning $V_1 \mathbin{\dot\cup} V_2 = V$ of a graph $G = (V, E)$ with girth at least $2k + 2$ and $\mathrm{freq}_k(V_1 \mid G), \mathrm{freq}_k(V_2 \mid G) \approx \mathrm{freq}_k(G)$, one can reduce the size of the cut of V_1 and V_2 to some constant by rewiring edges without any effect on $\mathrm{freq}_k(V_i \mid G)$, $i \in \{1, 2\}$, and $\mathrm{freq}_k(G)$. Removing the remaining edges in the cut changes the k-disc frequency vectors only slightly. Thus, two smaller graphs with approximately the same k-disc frequency vector as G are obtained.

To this end, our first lemma shows that the fraction of (Δ', Δ'')-edges that start in an arbitrary subset $V_1 \subseteq V$ is approximately the same as for another arbitrary subset $V_2 \subseteq V$ if the frequency distributions of the k-discs in V_1 and V_2 are close.

▶ **Lemma 6.** *Let $G = (V, E)$ be a d-bounded degree graph, $k \in \mathbb{N}$, $\lambda \in [0,1]$ and let $V_1, V_2 \subseteq V$ be such that $|\mathrm{freq}_k(V_1 \mid G)_\Gamma - \mathrm{freq}_k(V_2 \mid G)_\Gamma| \le \lambda$ for all k-discs $\Gamma \in \mathcal{T}_k$. Then, for all $(k-1)$-discs $\Delta', \Delta'' \in \mathcal{T}_{k-1}$ such that $\Delta' \not\simeq \Delta''$, it holds that*

$$\left| \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V)}{|V_1|} - \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V)}{|V_2|} \right| \le \lambda \cdot \mathrm{neigh}_\Sigma(\Delta', \Delta'') \,.$$

Proof. Consider any Δ'-extensive k-disc $\Gamma \in \mathrm{ext}(\Delta')$. Then the number of (Δ', Δ'')-edges in G such that the root of Δ' belongs to V_1 equals

$$\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V) = \sum_{\Gamma \in \mathrm{ext}(\Delta')} \mathrm{cnt}_k(V_1 \mid G)_\Gamma \cdot \mathrm{neigh}(\Gamma, \Delta'') \,.$$

An analogous equation holds for $\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V)$. Note that since $\Delta' \not\simeq \Delta''$, even edges that start and end in V_1 are counted only once in the right-hand side of the equation because $\mathrm{ext}(\Delta') \cap \mathrm{ext}(\Delta'') = \emptyset$. Therefore,

$$\begin{aligned}
&\left| \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V)}{|V_1|} - \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V)}{|V_2|} \right| \\
&= \left| \frac{\sum_{\Gamma \in \mathrm{ext}(\Delta')} \mathrm{cnt}_k(V_1 \mid G)_\Gamma \cdot \mathrm{neigh}(\Gamma, \Delta'')}{|V_1|} - \frac{\sum_{\Gamma \in \mathrm{ext}(\Delta')} \mathrm{cnt}_k(V_2 \mid G)_\Gamma \cdot \mathrm{neigh}(\Gamma, \Delta'')}{|V_2|} \right| \\
&\le \sum_{\Gamma \in \mathrm{ext}(\Delta')} |\mathrm{freq}_k(V_1 \mid G)_\Gamma - \mathrm{freq}_k(V_2 \mid G)_\Gamma| \cdot \mathrm{neigh}(\Gamma, \Delta'') \\
&\le \lambda \cdot \sum_{\Gamma \in \mathrm{ext}(\Delta')} \mathrm{neigh}(\Gamma, \Delta'') \\
&= \lambda \cdot \mathrm{neigh}_\Sigma(\Delta', \Delta'') \,.
\end{aligned}$$ ◀

If V_1, V_2 is a partitioning of V, the former result can be improved. In particular, we show that if $\mathrm{freq}_k(V_1 \mid G) \approx \mathrm{freq}_k(V_2 \mid G)$, then for almost every (Δ', Δ'')-edge from V_1 to V_2 there is a counterpart, that is, a (Δ', Δ'')-edge from V_2 to V_1. We will later use this result to reduce the size of the cut without altering the k-disc frequency vector by swapping the endpoints of edges in the cut so that the new edges lie completely in V_1 and V_2, respectively.

▶ **Lemma 7.** *Let $G = (V, E)$ be a d-bounded degree graph, $k \in \mathbb{N}$, $\lambda \in [0,1]$ and let $V_1 \dot\cup V_2 = V$ be a partitioning of V such that $|\mathrm{freq}_k(V_1 \mid G)_\Gamma - \mathrm{freq}_k(V_2 \mid G)_\Gamma| \le \lambda$ for all k-discs $\Gamma \in \mathcal{T}_k$. Then, for all $(k-1)$-discs $\Delta', \Delta'' \in \mathcal{T}_{k-1}$, it holds that*

$$|\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) - \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1)| \le \frac{|V_1||V_2|}{|V|} \cdot \lambda \cdot \left[\mathrm{neigh}_\Sigma(\Delta', \Delta'') + \mathrm{neigh}_\Sigma(\Delta'', \Delta') \right] .$$

Proof. If $\Delta' \simeq \Delta''$, the bound holds trivially because $\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) = \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1)$. Therefore, assume that $\Delta' \not\simeq \Delta''$ now. Note that by symmetry it holds that

$$\mathrm{e}(\Delta', \Delta'' \,|\, V_i, V_j) = \mathrm{e}(\Delta'', \Delta' \,|\, V_j, V_i) \qquad i, j \in \{1, 2\} \,. \tag{1}$$

Furthermore, since $V_1 \dot\cup V_2$ is a partitioning of V, we have

$$\mathrm{e}(\Delta', \Delta'' \,|\, V_i, V) = \mathrm{e}(\Delta', \Delta'' \,|\, V_i, V_1) + \mathrm{e}(\Delta', \Delta'' \,|\, V_i, V_2) \qquad i \in \{1, 2\} \tag{2}$$

and an analogous equation for $\mathrm{e}(\Delta'', \Delta' \,|\, V_i, V)$. Now, we have

$$\begin{aligned}
&|\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) - \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1)| \\
&= \frac{|V_1||V_2|}{|V|} \cdot \frac{(|V_1| + |V_2|)}{|V_1||V_2|} \cdot \Big| \mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) - \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1) \Big| \\
&= \frac{|V_1||V_2|}{|V|} \cdot \left| \left(\frac{1}{|V_1|} + \frac{1}{|V_2|} \right) \cdot \Big(\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) - \mathrm{e}(\Delta'', \Delta' \,|\, V_1, V_2) \Big) + 0 - 0 \right|
\end{aligned}$$

$$
\begin{aligned}
&= \frac{|V_1||V_2|}{|V|} \cdot \Bigg| \left(\frac{1}{|V_1|} + \frac{1}{|V_2|}\right) \cdot \Big(\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) - \mathrm{e}(\Delta'', \Delta' \,|\, V_1, V_2) \Big) \\
&\quad + \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_1) - \mathrm{e}(\Delta'', \Delta' \,|\, V_1, V_1)}{|V_1|} - \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_2) - \mathrm{e}(\Delta'', \Delta' \,|\, V_2, V_2)}{|V_2|} \Bigg| \\
&= \frac{|V_1||V_2|}{|V|} \cdot \Bigg| \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_1) + \mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2)}{|V_1|} - \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_2) + \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1)}{|V_2|} \\
&\quad - \frac{\mathrm{e}(\Delta'', \Delta' \,|\, V_1, V_1) + \mathrm{e}(\Delta'', \Delta' \,|\, V_1, V_2)}{|V_1|} + \frac{\mathrm{e}(\Delta'', \Delta' \,|\, V_2, V_2) + \mathrm{e}(\Delta'', \Delta' \,|\, V_2, V_1)}{|V_2|} \Bigg| \\
&= \frac{|V_1||V_2|}{|V|} \cdot \Bigg| \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V)}{|V_1|} - \frac{\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V)}{|V_2|} - \frac{\mathrm{e}(\Delta'', \Delta' \,|\, V_1, V)}{|V_1|} + \frac{\mathrm{e}(\Delta'', \Delta' \,|\, V_2, V)}{|V_2|} \Bigg| \\
&\le \frac{|V_1||V_2|}{|V|} \cdot \lambda \Big[\mathrm{neigh}_\Sigma(\Delta', \Delta'') + \mathrm{neigh}_\Sigma(\Delta'', \Delta') \Big],
\end{aligned}
$$

where the fourth equation follows from Eq. (1), the fifth equation follows from Eq. (2) and the inequality follows from applying Lemma 6. ◀

The former result enables us to analyze our main technical tool, that is, the rewiring of edges. First, we will prove that under some condition we can rewire two (Δ', Δ'')-edges without altering the k-disc frequency distribution of the graph or the partitions. This part of the proof shows that there exists, for every vertex $v \in V$, an isomorphism function that maps the k-disc of v in the original graph to the k-disc of v in the rewired graph. We then show that if we cannot find such (Δ', Δ'')-edges, the cut of V_1 and V_2 is small. This implies that the removal of the remaining edges changes the k-disc frequency vector of the graph only slightly.

▶ **Lemma 8.** *Let $G = (V, E)$ be a d-bounded graph with $\mathrm{girth}(G) \ge 2k + 2$, $k \in \mathbb{N}$, $\lambda \in [0, 1]$ and let $V_1 \,\dot\cup\, V_2 = V$ be a partitioning of V such that $|\mathrm{freq}_k(V_1 \mid G)_\Gamma - \mathrm{freq}_k(V_2 \mid G)_\Gamma| \le \lambda$ for all k-discs $\Gamma \in \mathcal{T}_k$. Then either there exists a graph $H = (V, F)$ such that*

$$\mathrm{girth}(H) \ge 2k + 2 \tag{3}$$

$$|F \cap (V_1 \times V_2)| \le |E \cap (V_1 \times V_2)| - 2 \tag{4}$$

$$\mathrm{disc}_k(H, w) \simeq \mathrm{disc}_k(G, w) \qquad \forall w \in V \tag{5}$$

or the cut between V_1 and V_2 is small:

$$\mathrm{e}(V_1, V_2) \le 6d^{2k+2}L + 2\lambda L d \cdot \min(|V_1|, |V_2|)\,. \tag{6}$$

Proof. Consider the following condition (see Fig. 1):

($\star$) There exist $(u_1, v_2) \in (V_1 \times V_2) \cap E$ and $(u_2, v_1) \in (V_2 \times V_1) \cap E$ such that $\mathrm{dt}_G(u_1, v_1)$, $\mathrm{dt}_G(v_2, u_2) \ge 2k+1$, $\mathrm{disc}_{k-1}(G, u_1) \simeq \mathrm{disc}_{k-1}(G, u_2)$ and $\mathrm{disc}_{k-1}(G, v_2) \simeq \mathrm{disc}_{k-1}(G, v_1)$.

Informally, it states that, for a suitable choice of $(k-1)$-discs Δ' and Δ'', there exists a (Δ', Δ'')-edge from V_1 to V_2 and a (Δ', Δ'')-edge from V_2 to V_1 such that two endpoints of different edges are not too close. We prove in the following that if condition ($\star$) is satisfied, then there exists a graph H with the desired properties, and that Eq. (6) holds otherwise.

▶ **Claim 9.** *If condition ($\star$) is satisfied, there exists a graph $H = (V, F)$ such that Ineq. (3) and (4) and Expr. (5) are satisfied.*

Proof. Suppose that condition ($\star$) is satisfied. We define an intermediate graph $G' := (V, E')$ that is obtained by deleting the edges (u_1, v_2) and (u_2, v_1) from G, that is, $E' := E \setminus$

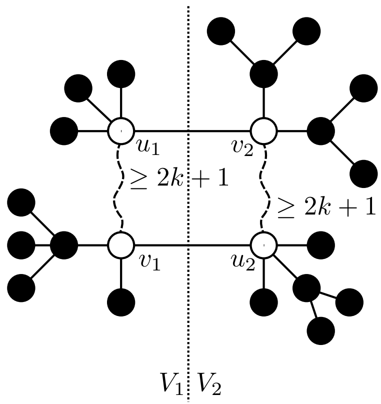

Figure 1 If condition ($\star$) is satisfied (as here for $k = 2$), it is possible to replace the edges (u_1, v_2) and (u_2, v_1) by (u_1, v_1) and (u_2, v_2), respectively, without changing the k-disc vector of G. Otherwise, the cut of V_1 and V_2 is small and removing these edges affects only few k-discs in V_1.

$\{(u_1, v_2), (u_2, v_1)\}$. We further define $H := (V, F)$ to be the graph that is obtained by adding the edges (u_1, v_1) and (u_2, v_2) to G', that is, $F := E' \cup \{(u_1, v_1), (u_2, v_2)\}$.

Observe that since $\mathrm{dt}_G(u_1, v_1), \mathrm{dt}_G(u_2, v_2) \geq 2k + 1$, Ineq. (3) holds, and by definition of H, Ineq. (4) holds. Thus, it remains to prove that Expr. (5) also holds, that is, for any vertex w, the k-discs of w in G and H are isomorphic. In what follows, we carefully construct a root-preserving bijection $f : V(\mathrm{disc}_k(G, w)) \to V(\mathrm{disc}_k(H, w))$ such that for all $x, y \in V(\mathrm{disc}_k(G, w))$, $(x, y) \in E(\mathrm{disc}_k(G, w))$ if and only if $(f(x), f(y)) \in E(\mathrm{disc}_k(H, w))$ to formally prove this somewhat intuitive observation.

Let $w \in V$. We distinguish between the cases that neither (u_1, v_2) nor (u_2, v_1), either one of them, or both are contained in $\mathrm{disc}_k(G, w)$. First, we will specify two isomorphism functions $g_u : V(\mathrm{disc}_k(G, u_1)) \to V(\mathrm{disc}_k(G, u_2))$ and $g_v : V(\mathrm{disc}_k(G, v_2)) \to V(\mathrm{disc}_k(G, v_1))$ for $\mathrm{disc}_k(G, u_1) \simeq \mathrm{disc}_k(G, u_2)$ and $\mathrm{disc}_k(G, v_2) \simeq \mathrm{disc}_k(G, v_1)$, respectively. If there is more than one candidate for g_u and g_v respectively, we make an arbitrary choice unless stated otherwise. We will then define f using these two functions g_u and g_v and prove that f is an isomorphism between $\mathrm{disc}_k(G, w)$ and $\mathrm{disc}_k(H, w)$.

Case 1: $(u_1, v_2) \notin \mathrm{disc}_k(G, w)$, $(u_2, v_1) \notin \mathrm{disc}_k(G, w)$. In this case, we define $f(x) := x$ for all $x \in V(\mathrm{disc}_k(G, w))$. We claim that neither (u_1, v_1) nor (v_2, u_2) belongs to $E(\mathrm{disc}_k(H, w))$. Without loss of generality assume that $(u_1, v_1) \in E(\mathrm{disc}_k(H, w))$. Then $\mathrm{dt}_G(w, u_1), \mathrm{dt}_G(w, v_1) \leq k$, which implies that $\mathrm{dt}_G(u_1, v_1) \leq \mathrm{dt}_G(u_1, w) + \mathrm{dt}_G(w, v_1) \leq 2k$. This is a contradiction to the assumption that $\mathrm{dt}_G(u_1, v_1) \geq 2k + 1$. The same argument shows that $(u_2, v_2) \notin E(\mathrm{disc}_k(G, w))$. Therefore, the k-discs $\mathrm{disc}_k(G, w)$ and $\mathrm{disc}_k(H, w)$ do not contain any of the edges (u_1, v_2), (u_2, v_1), (u_1, v_1), (u_2, v_2) and thus $\mathrm{disc}_k(G, w) \simeq \mathrm{disc}_k(H, w)$ by our definition of H.

Case 2: $(u_1, v_2) \in \mathrm{disc}_k(G, w)$, $(u_2, v_1) \notin \mathrm{disc}_k(G, w)$. In this case it holds that $u_2, v_1 \notin \mathrm{disc}_k(G, w)$, since otherwise, either $\mathrm{dt}_G(u_1, v_1) \leq 2k$ or $\mathrm{dt}_G(v_2, u_2) \leq 2k$, which contradicts condition ($\star$).

Now we observe that since $\mathrm{girth}(G) \geq 2k + 2$, the k-disc $\mathrm{disc}_k(G, w)$ is a tree. This implies that the deletion of the edge (u_1, v_2) will partition $\mathrm{disc}_k(G, w)$ into two connected components, say P_{u_1} and P_{v_2}, which represent the set of vertices in $\mathrm{disc}_k(G, w)$ that are connected to u_1 after deleting (u_1, v_2) and the set of remaining vertices that are connected

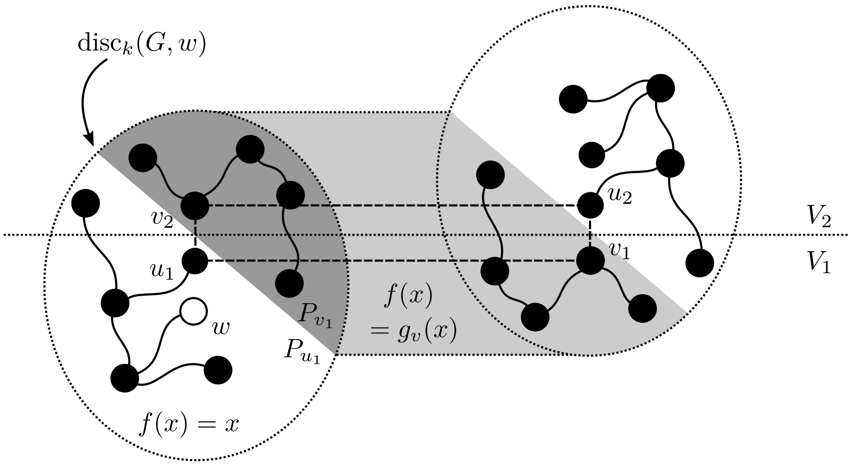

Figure 2 The k-disc of w in G, $\mathrm{disc}_k(G,w)$, is partitioned into two parts P_{u_1} (white background) and P_{v_2} (dark gray background) by the edge (u_1, v_2). The dashed edges are only present in either G or H but not in G'. If $w \in P_{u_1}$ as in the figure, then $f(x) := x$ for any $x \in P_{u_1}$, and $f(x) := g_v(x)$ for any $x \in P_{v_2}$, where g_v is chosen such that the image of P_{v_2} under g_v is a subset of $\mathrm{disc}_k(G', v_1)$ (light gray background).

to v_2, respectively. Without loss of generality assume that $w \in P_{u_1}$. The case that $w \in P_{v_2}$ can be analyzed similarly.

Let $f(x) = x$ if $x \in P_{u_1}$ and $f(x) = g_v(x)$ if $x \in P_{v_2}$. If there is more than one candidate for g_v, we make an arbitrary choice among all isomorphism functions that map P_{v_2} to (a subset of) $\mathrm{disc}_{k-1}(G', v_1)$, that is, the $(k-1)$-disc of v_1 after deleting (u_2, v_1) (see Fig. 2). Since $\mathrm{disc}_{k-1}(G, v_2) \simeq \mathrm{disc}_{k-1}(G, v_1)$ by ($\star$), there is always an isomorphism function that satisfies this condition. Moreover, f is a bijection because g_v is a bijection, and the image of $V(\mathrm{disc}_k(G,w))$ under f is $V(\mathrm{disc}_k(H,w))$ by the construction of H. We now prove that f is an isomorphism function between $\mathrm{disc}_k(G,w)$ and $\mathrm{disc}_k(H,w)$.

First note that $f(w) = w$, $f(u_1) = u_1$ and $f(v_2) = g_v(v_2) = v_1$. Now consider any $x, y \in V(\mathrm{disc}_k(G,w))$. If $x, y \in P_{u_1}$ or $x, y \in P_{v_2}$, then $f(x) = x$, $f(y) = y$ or $f(x) = g_v(x)$, $f(y) = g_v(y)$, respectively. Therefore, $(x, y) \in E(G)$ if and only if $(f(x), f(y)) \in E(H)$.

Now consider the case that $x \in P_{u_1}$ and $y \in P_{v_2}$. If $x = u_1$ and $y = v_2$, then we know that $(x, y) \in E(G)$ and also that $(f(x), f(y)) = (u_1, v_1) \in E(H)$ by the definition of H. Otherwise, either $x \neq u_1$ or $y \neq v_2$. In this case, there is no edge (x, y) in G since $\mathrm{disc}_k(G,w)$ is a tree and x, y lie on different sides of the edge (u_1, v_2). Recall that $f(u_1) = u_1$ and $f(v_2) = g_v(v_2) = v_1$. Since f is a bijection, either $f(x) \neq u_1$ or $f(y) \neq v_1$. Observe that $\mathrm{disc}_k(H,w)$ is a tree by Ineq. (3). Hence there is no edge between $f(x)$ and $f(y)$ in H as they lie on different sides of the edge (u_1, v_1). The case that $x \in P_{v_2}$ and $y \in P_{u_1}$ is symmetric.

Therefore, the function f is a root-preserving isomorphism function between $\mathrm{disc}_k(G,w)$ and $\mathrm{disc}_k(H,w)$.

Case 3: $(u_1, v_2) \notin \mathrm{disc}_k(G,w)$, $(u_2, v_1) \in \mathrm{disc}_k(G,w)$. This case can be analyzed similarly to the foregoing case.

Case 4: $(u_1, v_2) \in \mathrm{disc}_k(G,w), (u_2, v_1) \in \mathrm{disc}_k(G,w)$. Note that this case cannot happen because otherwise we would have $\mathrm{dt}_G(u_1, v_1), \mathrm{dt}_G(u_2, v_2) \leq 2k$, which contradicts the assumption that $\mathrm{dt}_G(u_1, v_1)$, $\mathrm{dt}_G(u_2, v_2) \geq 2k+1$. This completes the case analysis and the proof of Claim 9. ◀

▶ **Claim 10.** *If condition* $(\star)$ *is not satisfied, then Eq.* (6) *holds.*

Proof. Suppose that condition $(\star)$ is not satisfied. First, note that we have

$$\mathrm{e}(V_1, V_2) = \sum_{\Delta' \in \mathcal{T}_{k-1}} \sum_{\Delta'' \in \mathcal{T}_{k-1}} \mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2)\,.$$

Now, let $\Delta', \Delta'' \in \mathcal{T}_{k-1}$ be any two $(k-1)$-disc isomorphism types. The key observation is that if (u_1, v_2) is a (Δ', Δ'')-edge from V_1 to V_2, then for every (Δ', Δ'')-edge (u_2, v_1) from V_2 to V_1 the distance between u_2 and v_2 or the distance between v_1 and u_1 must be smaller than $2k+1$ (otherwise, the edges could be rewired and $(\star)$ would be satisfied). Since the graph is degree-bounded, this implies an upper bound on the number of possible endpoints u_2, v_1 and thus implies an upper bound on $\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1)$. It follows that $\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2)$ is also bounded by Lemma 7. In case there is no (Δ', Δ'')-edge from V_1 to V_2, the number of (Δ', Δ'')-edges from V_2 to V_1 can be bounded directly by Lemma 7.

We proceed to make this precise. For every choice of $\Delta', \Delta'' \in \mathcal{T}_{k-1}$, we distinguish two cases as mentioned before: whether we can find a (Δ', Δ'')-edge from V_1 to V_2 or not.

Case 1: There exist $u_1 \in V_1$ and $v_2 \in V_2$ such that $(u_1, v_2) \in E$, $\mathrm{disc}_{k-1}(G, u_1) \simeq \Delta'$ and $\mathrm{disc}_{k-1}(G, v_2) \simeq \Delta''$, that is, $\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) > 0$. Since condition $(\star)$ is not satisfied, at least one endpoint of every (Δ', Δ'')-edge (u_2, v_1) from V_2 to V_1 must have distance less than $2k+1$ to u_1 or v_2. Without loss of generality, fix such an edge with $\mathrm{dt}_G(u_2, v_2) < 2k+1$. The case $\mathrm{dt}_G(u_1, v_1) < 2k+1$ can be analyzed similarly. There are at most $3d^{2k}/2$ vertices with distance less than $2k+1$ to v_2 by Fact 4. Each of these near vertices can be adjacent to at most d vertices in V_1 whose $(k-1)$-discs are isomorphic to Δ''. Taking the symmetric case $\mathrm{dt}_G(u_1, v_2) < 2k+1$ into account, we have $\mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1) \le 2 \cdot 3d^{2k}/2 \cdot d \le 3d^{2k+1}$. Now by Lemma 7, it holds that

$$\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) + \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1) \le 6d^{2k+1} + \frac{|V_1||V_2|}{|V|} \cdot \lambda \Big[\mathrm{neigh}_\Sigma(\Delta', \Delta'') + \mathrm{neigh}_\Sigma(\Delta'', \Delta')\Big]\,.$$

Case 2: There do not exist $u_1 \in V_1$ and $v_2 \in V_2$ such that $(u_1, v_2) \in E$, $\mathrm{disc}_{k-1}(G, u_1) \simeq \Delta'$ and $\mathrm{disc}_{k-1}(G, v_2) \simeq \Delta''$, that is, $\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) = 0$. By Lemma 7, we have

$$\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) + \mathrm{e}(\Delta', \Delta'' \,|\, V_2, V_1) \le 0 + \frac{|V_1||V_2|}{|V|} \cdot \lambda \Big[\mathrm{neigh}_\Sigma(\Delta', \Delta'') + \mathrm{neigh}_\Sigma(\Delta'', \Delta')\Big]\,.$$

This completes the case analysis. Note that each k-disc $\Gamma \in \mathcal{T}_k$ determines the $(k-1)$-discs of its root and of its at most d neighbors. Moreover, the number of different k-disc isomorphism types in G is at most L. Therefore, the number of pairs Δ', Δ'' such that there exists an edge between a vertex with $(k-1)$-disc Δ' and a vertex with $(k-1)$-disc Δ'' is at most Ld, that is, $\mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) \neq 0$ for at most Ld pairs Δ', Δ'', and we have

$$\begin{aligned}
&\mathrm{e}(V_1, V_2) \\
&= \sum_{\Delta' \in \mathcal{T}_{k-1}} \sum_{\Delta'' \in \mathcal{T}_{k-1}} \mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) \\
&\le Ld \cdot \max_{\Delta', \Delta'' \in \mathcal{T}_{k-1}} \mathrm{e}(\Delta', \Delta'' \,|\, V_1, V_2) \\
&\le Ld \cdot \left(6d^{2k+1} + \frac{\lambda |V_1||V_2|}{|V|} \cdot \max_{\Delta', \Delta'' \in \mathcal{T}_{k-1}} \Big[\mathrm{neigh}_\Sigma(\Delta', \Delta'') + \mathrm{neigh}_\Sigma(\Delta'', \Delta')\Big]\right) \\
&\le 6d^{2k+2}L + \lambda \cdot \min(|V_1|, |V_2|) \cdot 2Ld,
\end{aligned}$$

where the last step follows from Fact 5. This completes the proof of Claim 10 and Lemma 8. ◀

4 Proof of the Main Theorems

We first prove Theorem 1 by arguing along the execution of Algorithm 1.

Algorithm 1

1: **function** PartitionAndRewire($G = (V, E)$, φ)
2: $\quad V_1 \leftarrow$ sample φ vertices from V uniformly at random
3: $\quad V_2 \leftarrow V \setminus V_1$
4: $\quad E' \leftarrow E$, $G' \leftarrow (V, E')$
5: $\quad$ **for all** $(u_1, v_2) \in (V_1 \times V_2) \cap E'$ and $(u_2, v_1) \in (V_2 \times V_1) \cap E'$ **do**
6: $\quad\quad$ **if** $\mathrm{disc}_{k-1}(G', u_1) \simeq \mathrm{disc}_{k-1}(G', u_2) \wedge \mathrm{disc}_{k-1}(G', v_2) \simeq \mathrm{disc}_{k-1}(G', v_1)$ $\wedge\, \mathrm{dt}_{G'}(u_1, v_1) \geq 2k + 1 \wedge \mathrm{dt}_{G'}(u_2, v_2) \geq 2k + 1$ **then**
7: $\quad\quad\quad E' \leftarrow E' \setminus \{(u_1, v_2), (u_2, v_1)\} \cup \{(u_1, v_1), (u_2, v_2)\}$
8: $\quad\quad\quad G' \leftarrow (V, E')$
9: $\quad\quad\quad$ Goto line 5
10: $\quad\quad$ **end if**
11: $\quad$ **end for**
12: $\quad$ **return** $H := G'[V_1]$
13: **end function**

Proof of Theorem 1. We prove that the output of Algorithm 1 is a graph with the desired properties. First, we sample φ vertices $v_1, \ldots, v_\varphi$ from G uniformly at random (cf. line 2). Let E_1 denote the event that all the sampled vertices are different. Note that for any i, j such that $1 \leq i < j \leq \varphi$, the probability that $v_i = v_j$ is at most $1/|V|$, which implies that $\Pr[E_1] \geq 1 - \frac{\varphi^2}{|V|} \geq 1 - \frac{\delta}{2}$ because $|V| \geq 2\varphi^2/\delta$.

Let $V_2 := V \setminus V_1$. For each $i \leq \varphi$, let $\vec{\boldsymbol{f}}_i \in \{0,1\}^L$ denote the random vector that equals the indicator vector $\vec{\mathbf{1}}_\Gamma$ if the k-disc of v_i is isomorphic to Γ. Note that $\mathrm{freq}_k(V_1 \mid G) = \sum_i \vec{\boldsymbol{f}}_i/\varphi$ and that $\Pr[\vec{\boldsymbol{f}}_i = \vec{\mathbf{1}}_\Gamma] = \mathrm{freq}_k(G)_\Gamma$, and thus $\mathrm{E}[\mathrm{freq}_k(V_1 \mid G)] = \mathrm{E}[\vec{\boldsymbol{f}}_i] = \mathrm{freq}_k(G)$. Let $X := \|\mathrm{freq}_k(G) - \mathrm{freq}_k(V_1 \mid G)\|_2^2$. We bound the deviation between $\mathrm{freq}_k(G)$ and $\sum_i \vec{\boldsymbol{f}}_i/\varphi$. It holds that

$$
\begin{aligned}
\mathrm{E}[X] = \mathrm{E}\left[\left\|\mathrm{freq}_k(G) - \frac{\sum_{i=1}^{\varphi} \vec{\boldsymbol{f}}_i}{\varphi}\right\|_2^2\right] &= \mathrm{E}\left[\frac{1}{\varphi^2}\left\|\sum_{i=1}^{\varphi}(\mathrm{freq}_k(G) - \vec{\boldsymbol{f}}_i)\right\|_2^2\right] \\
&= \frac{1}{\varphi^2}\mathrm{E}\left[\sum_{i=1}^{\varphi}\|\mathrm{freq}_k(G) - \vec{\boldsymbol{f}}_i\|_2^2\right] \\
&= \frac{1}{\varphi^2}\sum_{i=1}^{\varphi}\mathrm{E}\left[\|\mathrm{freq}_k(G) - \vec{\boldsymbol{f}}_i\|_2^2\right] \\
&= \frac{1}{\varphi}\mathrm{E}\left[\|\mathrm{freq}_k(G) - \vec{\boldsymbol{f}}_1\|_2^2\right] \\
&\leq \frac{1}{\varphi}\mathrm{E}\left[\|\mathrm{freq}_k(G) - \vec{\boldsymbol{f}}_1\|_1\right] \\
&\leq \frac{2}{\varphi},
\end{aligned}
$$

where the third equation follows from the fact that all $\vec{\boldsymbol{f}}_i$ are independent of each other; the penultimate inequality follows from the fact that the absolute values of all entries of

$\mathrm{freq}_k(G') - \vec{\boldsymbol{f}}_1$ are at most 1. Now by Markov's inequality,

$$\Pr\left[\left\|\mathrm{freq}_k(G) - \frac{\sum_i \vec{\boldsymbol{f}}_i}{\varphi}\right\|_2^2 \geq \frac{2}{\delta}\cdot\frac{2}{\varphi}\right] \leq \Pr\left[X \geq \frac{2}{\delta}\cdot \mathrm{E}[X]\right] \leq \frac{\delta}{2}\,.$$

Therefore, if we let $\lambda = \frac{\varepsilon}{6Ld^{k+1}}$, then with probability at least $1-\delta/2$,

$$\left\|\mathrm{freq}_k(G) - \frac{\sum_i \vec{\boldsymbol{f}}_i}{\varphi}\right\|_1 \leq \sqrt{L}\cdot\left\|\mathrm{freq}_k(G) - \frac{\sum_i \vec{\boldsymbol{f}}_i}{\varphi}\right\|_2 \leq \sqrt{L}\cdot\sqrt{\frac{4}{\delta\varphi}} \leq \lambda/2,$$

where the last inequality follows from our choice of $\varphi = \frac{300d^{3k+2}L^3}{\varepsilon^2\delta} = \frac{300d^kL}{36\lambda^2\delta} \geq \frac{16L}{\lambda^2\delta}$. This further implies that (with probability at least $1-\delta/2$)

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(V_1 \mid G)\|_1 = \left\|\mathrm{freq}_k(G) - \frac{\sum_i \vec{\boldsymbol{f}}_i}{\varphi}\right\|_1 \leq \frac{\lambda}{2}\,. \tag{7}$$

Let E_2 denote the event that $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(V_1 \mid G)\|_1 \leq \frac{\lambda}{2}$. Thus $\Pr[E_2] \geq 1-\delta/2$. If E_2 occurs, then $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(V_2 \mid G)\|_1 \leq \frac{\lambda}{2}$ because $|V_2| \geq |V_1|$, and therefore $\|\mathrm{freq}_k(V_1 \mid G) - \mathrm{freq}_k(V_2 \mid G)\|_1 \leq \lambda$.

Conditioning on both events E_1 and E_2, which occur with probability $\Pr[E_1 \cap E_2] \geq 1 - 2\cdot\frac{\delta}{2} = 1-\delta$, we apply Lemma 8 with G, λ and partition V_1, V_2 as follows: Let $G' = (V, E')$ be a copy of G. As long as condition $(\star)$ is satisfied, we replace G' by the *rewired* graph that satisfies Ineq. (3) and (4) and Expr. (5) (cf. lines 5 to 11). After rewiring, there remain at most $6d^{2k+2}L + 2\lambda Ld\varphi$ edges in the cut of V_1 and V_2, which are (virtually) deleted by returning the graph $H := G'[V_1]$.

The k-disc of a vertex is altered if and only if an edge is inserted to the k-disc or removed from it. The maximum size of a k-disc is at most $3d^k/2$ by Fact 4. Therefore, removing a single edge alters at most $3d^k/2$ k-discs. By Lemma 8 it holds that

$$\|\mathrm{freq}_k(V_1 \mid G) - \mathrm{freq}_k(H)\|_1 \leq \frac{3d^k/2\cdot\left(6d^{2k+2}L + 2\lambda Ld\varphi\right)}{\varphi} \leq \frac{9d^{3k+2}L}{\varphi} + 3Ld^{k+1}\lambda \leq \frac{3\varepsilon}{4}, \tag{8}$$

where the last inequality follows from our choice of $\varphi = \frac{300d^{3k+2}L^3}{\varepsilon^2\delta}$ and $\lambda = \frac{\varepsilon}{6Ld^{k+1}}$. It follows from Eqs. (7) and (8) and the triangle inequality that

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \frac{\lambda}{2} + \frac{3\varepsilon}{4} \leq \varepsilon\,.$$

Now we analyze the query (and time) complexity of the above algorithm. Note that the algorithm only needs to sample φ vertices and query all the $(k+2)$-discs of vertices in V_1. In particular, the rewiring step (cf. line 6) can be performed as follows: we consider all the vertices that are endpoints of some edges leaving V_1 by exploring the neighbors of all vertices in V_1. We want to find $u_1, v_1 \in V_1$ with $\mathrm{dt}_G(u_1, v_1) \geq 2k+1$ such that $(u_1, v_2) \in (V_1 \times V_2) \cap E'$ and $(u_2, v_1) \in (V_2 \times V_1) \cap E'$. To test if we should rewire the corresponding edges or not, we only need to consider the $(k+1)$-discs of $v_2, u_2 \in V_2$ to determine if $\mathrm{dt}_G(v_2, u_2) \geq 2k+1$. This implies that we only need to query the $(k+2)$-discs of all vertices in V_1. It follows that the algorithm makes at most $\varphi \cdot \frac{3d^{k+2}}{2} = \mathcal{O}(1)$ queries for constants d, ϵ and k to the oracle of G. Also note that since $|V_1| \in \mathcal{O}(1)$, the number of rewiring steps as well as the number of eges with at least one end in V_1 is at most $|V_1|d \in \mathcal{O}(1)$. Therefore, the algorithm has constant time complexity. ◀

Algorithm 2

1: **function** $\textsc{RewireAndSplit}(G = (V, E), \varphi)$
2: Partition V into V_1, V_2 such that
$|V_1| = \varphi$ and $|\mathrm{freq}_k(V_1 \mid G)_\Gamma - \mathrm{freq}_k(V_2 \mid G)_\Gamma| \le 1/\varphi$ for all Γ
3: $E' \leftarrow E$, $G' \leftarrow (V, E')$
4: **for all** $(u_1, v_2) \in (V_1 \times V_2) \cap E'$ and $(u_2, v_1) \in (V_2 \times V_1) \cap E'$ **do**
5: **if** $\mathrm{disc}_{k-1}(G', u_1) \simeq \mathrm{disc}_{k-1}(G', u_2) \wedge \mathrm{disc}_{k-1}(G', v_2) \simeq \mathrm{disc}_{k-1}(G', v_1)$
$\wedge\, \mathrm{dt}_{G'}(u_1, v_1) \ge 2k+1 \wedge \mathrm{dt}_{G'}(u_2, v_2) \ge 2k+1$ **then**
6: $E' \leftarrow E' \cap \{(u_1, v_2), (u_2, v_1)\} \cup \{(u_1, v_1), (u_2, v_2)\}$
7: $G' \leftarrow (V, E')$
8: Goto line 4
9: **end if**
10: **end for**
11: **return** $H := G'[V_1]$
12: **end function**

Now, we prove Theorem 2 by arguing along the execution of Algorithm 2.

Proof of Theorem 2. We prove that the output of Algorithm 2 is a graph with the desired properties. Let $\varphi := 12Ld^{3k+2}/\epsilon$. Without loss of generality assume that $\varphi \le |V(G)|/3$ (otherwise just output $H := G$ directly). First, we partition V into two parts V_1 and V_2 such that $\varphi \le |V_1| \le 2\varphi$ and for any k-disc Γ, $|\mathrm{freq}_k(G)_\Gamma - \mathrm{freq}_k(V_i \mid G)_\Gamma| \le 1/\varphi$ (cf. line 2). Such a partition can be constructed as follows: For each k-disc $\Gamma \in \mathcal{T}_k$, we put $\lceil \varphi \cdot \mathrm{freq}_k(G)_\Gamma \rceil$ vertices v with $\mathrm{disc}_k(G, v) \simeq \Gamma$ into V_1 and the remaining ones into V_2. Thus, $\varphi \le |V_1| \le \varphi + |\mathcal{T}_k| \le 2\varphi$ and we have

$$|\mathrm{freq}_k(G)_\Gamma - \mathrm{freq}_k(V_1 \mid G)_\Gamma| \le \left| \frac{\varphi \cdot \mathrm{freq}_k(G)_\Gamma - \lceil \varphi \cdot \mathrm{freq}_k(G)_\Gamma \rceil}{\varphi} \right| \le 1/\varphi\,.$$

Since $|V_2| = n - 2\varphi \ge \varphi$, we also have $|\mathrm{freq}_k(G)_\Gamma - \mathrm{freq}_k(V_2 \mid G)_\Gamma| \le 1/\varphi$. By the triangle inequality, the partitions V_1 and V_2 satisfy the prerequisite of Lemma 8 with $\lambda = 2/\varphi$. Additionally, it follows that

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(V_1 \mid G)\|_1 \le \frac{L}{\varphi}\,. \tag{9}$$

Let $G' = (V, E')$ be a copy of G. As long as G' and the partition V_1, V_2 satisfy the prerequisite of Lemma 8 and condition ($\star$), we *rewire* the edges of G' according to Lemma 8 so that G' will satisfy the properties given by Ineq. (3) and (4) and Expr. (5), (cf. lines 4 to 10). When G' does not satisfy condition ($\star$) anymore, we let $H := G'[V_1]$ and we are done. Note that at the end of the process, G' satisfies Eq. (6), which implies that the number of edges between V_1 and V_2 in G', that is, the *boundary* of H, is at most

$$6d^{2k+2}L + 2\lambda dL \cdot \min(|V_1|, |V_2|) \le 6d^{2k+2}L + \frac{4dL}{\varphi} \cdot \varphi \le 7d^{2k+2}L.$$

Now note that for any vertex $v \in H$, the k-disc of v in H differs from the k-disc of v in G' only if v is within distance at most k to the boundary of H, which in turn has size at most $7d^{2k+2}L$. By Fact 4, we have that the total number of vertices in H with different k-discs in H and G' is at most $3d^k/2 \cdot 7d^{2k+2}L \le 11d^{3k+2}L$, which implies that

$$\|\mathrm{freq}_k(V_1 \mid G) - \mathrm{freq}_k(H)\|_1 \le \frac{11d^{3k+2}L}{\varphi}\,. \tag{10}$$

It follows from Eqs. (9) and (10) and the triangle inequality that

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \le \frac{L + 11d^{3k+2}L}{\varphi} \le \frac{12d^{3k+2}L}{\varphi} \le \epsilon,$$

where the last inequality follows from our choice of φ.

Finally, we note that the graph H can be constructed by the following deterministic algorithm. We first compute the frequency vector $\mathrm{freq}_k(G)$ of G, which takes time $|V(G)|$. Then we consider all d-bounded graphs of size at most φ and output the graph H with frequency vector that is closest to $\mathrm{freq}_k(G)$ in ℓ_1-norm distance, which can be done in constant time. In total, the running time of the algorithm is $O(|V(G)|)$. ◀

Finally, we give a short proof of Corollary 3.

Proof of Corollary 3. We first note that the number of vertices whose k-discs may be altered by inserting / deleting a single edge $e = (u, v)$ is upper bounded by the number of k-discs that contain this edge. The number of such k-discs is exactly the number of vertices w such that there exists a path of length at most k from u to w and a path of length at most k from v to w, and is thus upper bounded by $1 + d + d(d-1) + \cdots + d(d-1)^{k-1} \le 3d^k/2$.

Let $\delta = \frac{\varepsilon}{6d^{k+1}}$. Since G' is δ-close to G, G' can be obtained from G by inserting / deleting at most $\delta d|V|$ edges, and thus the total number of vertices that may have different k-discs in G and G' is at most $\delta d|V| \cdot 3d^k/2$. Finally, since a vertex that has different k-discs in G and G' may contribute at most $2/|V|$ to the $s\ell_1$-norm distance of $\mathrm{freq}_k(G)$ and $\mathrm{freq}_k(G')$, we have

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(G')\|_1 \le \frac{2}{|V|} \cdot \left(\delta d|V| \cdot \frac{3d^k}{2}\right) = 3\delta d^{k+1} \le \frac{\varepsilon}{2}\,.$$

Now since G' satisfies that $\mathrm{girth}(G') \ge 2k+2$, by Theorem 2, we know that there exists a graph H with size at most $72\frac{d^{3k+2}L}{\varepsilon}$ such that $\|\mathrm{freq}_k(G') - \mathrm{freq}_k(H)\|_1 \le \varepsilon/2$. Therefore, $\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \le \varepsilon$. ◀

Acknowledgements. We thank anonymous referees for their valuable comments concerning the presentation of the paper.

References

1 D. Aldous and R. Lyons. Processes on Unimodular Random Networks. *Electronic Journal of Probability*, 12(54):1454–1508, 2007.

2 A. Benczúr and D. Karger. Approximating s-t Minimum Cuts In $\tilde{O}(n^2)$ Time. In *Proceedings of the 28th annual ACM Symposium on Theory of Computing*, pages 47–55. ACM, 1996.

3 I. Benjamini, O. Schramm, and A. Shapira. Every Minor-Closed Property of Sparse Graphs Is Testable. *Advances in Mathematics*, 223(6):2200–2218, 2010.

4 P. Chew. There Are Planar Graphs Almost as Good as the Complete Graph. *Journal of Computer and System Sciences*, 39(2):205–219, 1989.

5 G. Elek. On the Limit of Large Girth Graph Sequences. *Combinatorica*, 30(5):553–563, 2010.

6 J. Friedman. *A proof of Alon's second eigenvalue conjecture and related problems.* American Mathematical Society, 2008.

7 O. Goldreich and D. Ron. Property Testing in Bounded Degree Graphs. *Algorithmica*, 32:302–343, 2002.

8 A. Hassidim, J. Kelner, H. Nguyen, and K. Onak. Local Graph Partitions for Approximation and Testing. In *Proceedings of the 50th annual IEEE Symposium on Foundations of Computer Science*, pages 22–31. IEEE, 2009.

9 P. Indyk, A. McGregor, I. Newman, and K. Onak. Bertinoro workshop on sublinear algorithms 2011. `http://sublinear.info/42`. In *Open Problems in Data Streams, Property Testing, and Related Topics*, 2011.

10 L. Lovász. Very Large Graphs. *arXiv preprint arXiv:0902.0132*, 2009.

11 L. Lovász. *Large Networks and Graph Limits.* American Mathematical Society, 2012.

12 B. McKay, N. Wormald, and B. Wysocka. Short Cycles in Random Regular Graphs. *Electronic Journal of Combinatorics*, 11(1):66, 2004.

13 I. Newman and C. Sohler. Every Property of Hyperfinite Graphs Is Testable. *SIAM Journal on Computing*, 42(3):1095–1112, 2013.

14 D. Spielman and S. Teng. Spectral Sparsification of Graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

15 E. Szemerédi. Regular Partitions of Graphs. Technical report, DTIC Document, 1975.