A Sublinear Bipartiteness Tester for Bounded Degree Graphs

Oded Goldreich* Dana Ron†

February 5, 1998

Abstract

We present a sublinear-time algorithm for testing whether a bounded degree graph is bipartite or far from being bipartite. Graphs are represented by incidence lists of bounded length d, and the testing algorithm can perform queries of the form: "who is the $i\mathrm{th}$ neighbor of vertex v". The tester should determine with high probability whether the graph is bipartite or ϵ-far from bipartite for any given distance parameter ϵ. Distance between graphs is defined to be the fraction of entries on which the graphs differ in their incidence-lists representation. Our testing algorithm has query complexity and running time $\mathrm{poly}((\log N)/\epsilon) \cdot \sqrt{N}$ where N is the number of graph vertices. In previous work [GR96] we showed that $\Omega(\sqrt{N})$ queries are necessary (for constant ϵ), and hence the performance of our algorithm is tight (in its dependence on N), up to polylogarithmic factors.

In our analysis we use techniques that were previously applied to prove fast convergence of random walks on expander graphs. Here we use the counter-positive statement that slow convergence implies small cuts in the graph, and further show that these cuts have certain additional properties. This implication is applied in showing that for any graph, the graph vertices can be divided into disjoint subsets such that: (1) the total number of edges between the different subsets is small; and (2) each subset itself exhibits a certain mixing property that is useful in our analysis.

Keywords: Approximation Algorithms, Graph Algorithms, Property Testing, Random Walks on Graphs, Expansion of Graphs.

*Department of Computer Science, Weizmann Institute of Science, Rehovot, ISRAEL. E-mail: `oded@wisdom.weizmann.ac.il`. On sabbatical leave at LCS, MIT.

†Laboratory for Computer Science, MIT, 545 Technology Sq., Cambridge, MA 02139. E-mail: `danar@theory.lcs.mit.edu`. Supported by a Bunting fellowship.

1 Introduction

Property Testing as formulated in [RS96] and [GGR96][1] is the study of the following family of tasks: Given oracle access to an unknown function, determine whether the function has a certain predefined property or is far from any function having that property. Distance between functions is measured in terms of the fraction of the domain-elements on which the two functions have different values. Thus, testing a property is a relaxation of *deciding* that property, and it suggests a certain notion of *approximation*. In particular, in applications where functions close to having the property are almost as good as ones having the property, a testing algorithm, which is faster than the corresponding decision procedure, is a very valuable alternative to the latter. The same holds in applications where one encounters functions that either have the property or are far from having it.

Testing algebraic properties (e.g., linearity or being a polynomial of low-degree) plays an important role in the settings of Program Testing (e.g., [BLR93, RS96, Rub94]) and Probabilistically-Checkable Proof systems (e.g., [BFL91, BFLS91, FGL+91, AS92b, ALM+92]). Recently, the applicability of property testing has been extended to the domain of combinatorial optimization and the context of approximation algorithms (rather than inapproximability results via PCP). In particular, fast property testers for a variety of standard graph theoretic problems such as 3-Colorability, Max-CUT and edge-connectivity, have been presented [GGR96, GR96], and applications to the standard notion of approximation have been suggested (e.g., to approximating max-CUT in dense graphs [GGR96]).

The complexity and applicability of property testing depends very much on the representation of the objects being tested. Two models, corresponding to the two standard representations of graphs, were suggested for testing graph properties. In the first model, most appropriate to the study of dense graphs, graphs are represented by their *adjacency-matrix* (equivalently, *adjacency predicate*) [GGR96]. This means that the tester may make queries of the form "are u and v adjacent in the graph". Moreover, the distance between two N-vertex graphs is defined as the fraction of vertex-pairs on which the graphs disagree over the total of N^2 possible vertex-pairs (i.e., elements in the domain of the adjacency predicate). In the second model, most appropriate to the study of bounded-degree graphs, graphs are represented by their *incidence-lists* [GR96]: That is, an N-vertex graph of degree bound d is represented by a function from $\{1, 2, ..., N\} \times \{1, 2, ..., d\}$ to $\{0, 1, 2, ..., N\}$. This means that the tester may make queries of the form "who is the i^{th} neighbor of v" (and the answer may be a vertex or 0 indicating that v has less than i neighbors). In this model, the distance between N-vertex graphs of degree bound d is defined as the fraction of vertex-pairs on which they disagree over the total dN pairs in the domain of the function.

It is not surprising that property testing in the above two models has different flavor and complexity, and requires different techniques. A natural graph property exhibiting such a difference is bipartiteness. In the *first model* (adjacency-matrix representation), a simple algorithm of complexity independent of the size of the graph was shown to be a good tester of bipartiteness [GGR96]: Given a distance parameter ϵ, the algorithm uniformly selects a set of $\tilde{O}(\epsilon^{-2})$ vertices and accepts if and only if the subgraph induced by these vertices is bipartite. Clearly, each bipartite graph is accepted, and it was shown that any graph which is ϵ-far from bipartite is rejected with high probability. Under the distance metric of the first model, this means that graphs for which every 2-partition has ϵN^2 bipartite-violating edges, are rejected with high probability – a statement which is meaningful for dense graphs. On the other hand, it was shown that in the *second model* (incidence-lists representation), $\Omega(\sqrt{N})$ queries are required for testing bipartiteness (for constant d and ϵ such as $d = 3$ and $\epsilon = 0.01$) [GR96].

In this work we show that bipartiteness can be tested in the *second model* (incidence-lists representation)

[1] In [GGR96] Property Testing was given a broader definition. Here we restrict ourselves to the special case of testing using queries under the uniform distribution as defined already in [RS96].

in time $\tilde{O}(\text{poly}(1/\epsilon) \cdot \sqrt{N})$. This result is quite tight in light of the above cited lower bound. Furthermore, it enriches the study of combinatorial property testing in two ways:

1. The graph testing algorithms presented in both [GGR96] and [GR96] have complexity bounded by a function of the distance parameter ϵ (independent of the size of the graph). As shown in [GR96], such complexity can not be achieved for some natural properties. Our result demonstrates that property testing may have something to offer also in such a case. In general, we believe that a property testing algorithm is of interest if its complexity (for, say, constant ϵ) is lower than the complexity of deciding the property. We have demonstrated a natural problem for which property testing requires and can be done in time which is approximately the square root of the time required for deciding.

2. The graph testing algorithms presented in [GGR96] operate by uniformly selecting a small sample of vertices and inspecting the subgraph induced by them. This is certainly an important paradigm, but limited in scope to dense graphs and furthermore to cases where random subgraphs inherit properties of the graph. The algorithms in [GR96] operate by uniformly selecting a vertex and inspecting its close neighborhood. This paradigm seems restricted to bounded-degree graphs and to properties which are "approximately local". The algorithm presented in this paper can be viewed as a combination of both paradigms. In a way, we select a random sample of vertices together with random paths connecting them. Certainly, we cannot just select random vertices and then try to find paths among them. Instead, we take many random walks starting at (few) uniformly selected vertices.

Techniques. The algorithm presented in this paper is fairly simple. The algorithm uniformly selects $O(1/\epsilon)$ *starting* vertices, and from each starting vertex it performs $\text{poly}((\log N)/\epsilon) \cdot \sqrt{N}$ *random walks*, each of length $\text{poly}((\log N)/\epsilon)$. If for any starting vertex s, it detects that s lies on an odd-length cycle, then it rejects the graph. Otherwise it accepts. It is clear that if the graph is bipartite, then it is always accepted. The main thrust of our analysis is in proving that if the graph is far from bipartite then an odd-length cycle is detected with high probability. More precisely, we prove the counter-positive of that statement: If the acceptance probability is not too small then there exists a partition of the graph vertices that does not cause many violation (i.e. edges between vertices that belong to the same side of the partition).

To prove the existence of such a *good* partition, we use combinatorial techniques that were previously applied to prove fast convergence of random walks on expanders [Mih89]. Whereas Mihail [Mih89] showed that if there are no small cuts in the graph then convergence must be rapid, we show that too slow of a convergence implies the existence of small cuts *with certain additional properties* needed for the rest of our analysis. In particular, we show that for any graph, the graph vertices can be divided into disjoint subsets such that: (1) the total number of edges between the different subsets is small, and (2) each subset S exhibits certain mixing properties. Namely, there exists a vertex s such that for every vertex v in S, a short walk from s ends at v with probability approximately $\sqrt{\frac{1}{N \cdot |\mathrm{S}|}}$. This mixing property is used to show that either the vertices in S can be 2-partitioned without causing many violations, or an odd-length cycle (containing s) is detected with high probability. Hence, if the graph is accepted with high enough probability, then we can deduce that almost all of these subsets can be 2-partitioned without having many internal violations. Adding the (relatively few) edges between the subset, we end up with a good partition of the whole graph. As a corollary to our analysis, we obtain several lemmas which may be of independent interest. In particular, a drastic "degeneration" of our analysis yields the following combinatorial proposition (whose proof is given in Appendix C).

Proposition 1 *Let* G *be an undirected graph having* N *vertices and degree at most* d. *If* G *is* ϵ*-far from bipartite then it contains an odd-length cycle of length* $L = O(\epsilon^{-1} \log N)$. *Furthermore, such a cycle can be found in* *time linear in N. On the other hand, if G has no odd-cycle of length at most L then it can be 2-partitioned in linear time so that there are at most $\epsilon \cdot dN$ violating edges.*

2 Preliminaries

Let $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ be an undirected simple graph with N vertices where each vertex has degree at most d. For a vertex v, let $\Gamma(v)$ be the set of neighbors of v. We think of G as being represented by a two-dimensional array of size $N \times d$, where for each vertex v and integer $i \in \{1, \ldots, d\}$ the value of the corresponding entry is the i^{th} neighbor of v. If v has less than d neighbors then this value may be 0 (where $0 \notin \mathrm{V}$). For any subgraph H of G let the size of H, denoted $|\mathrm{H}|$, be the number of vertices in H.

Let $\mathrm{P} = (\mathrm{V}_1, \mathrm{V}_2)$ be a partition of V. We say that an edge $(v, u) \in \mathrm{E}$ is a *violating* edge with respect to P, if v and u belong to the same subset V_b, (for some $b \in \{1, 2\}$). A partition P is said to be ϵ-*good*, where $0 \leq \epsilon \leq 1$, if the number of violating edges in G with respect to P is at most $\epsilon \cdot dN$. We say that G is ϵ-*far* from being bipartite, if there is no ϵ-good partition of V. In other words, G is ϵ-far from being bipartite if the fraction of entries in its array representation that need to be modified in order to make it bipartite is greater than 2ϵ.[2]

An algorithm for *testing* bipartiteness is given a size parameter, N, a degree parameter, d, and a distance parameter ϵ. It is then given *oracle access* to an unknown graph G (with N vertices and maximum degree d). That is, the algorithm may ask queries of the form "who is the i^{th} neighbor of vertex v" (i.e., make probes into the array representation of G). If G is bipartite then with probability at least $\frac{2}{3}$ the algorithm should accept it, and if G is ϵ-far from bipartite, then with probability $\frac{2}{3}$ it should reject it.

3 The Algorithm

In this section we present our algorithm for testing bipartiteness. Since the algorithm has oracle access to G, as defined in Section 2, it can be viewed as performing *walks* on G, starting from vertices of its choice. In particular, our algorithm (described in Figure 1), performs *random* walks on G: At each step, if the degree of the current vertex v is $d' \leq d$, then the walk *remains* at v with probability $1 - \frac{d'}{2d} \geq \frac{1}{2}$, and for each $u \in \Gamma(v)$, the walk *traverses* to u with probability $\frac{1}{2d}$. Thus, the stationary distribution over the vertices is uniform. If we consider only steps in which the walk continued to a new vertex, then each random walk corresponds to a path in the graph. This path is not necessarily simple, but does not contain self loops. Note that when referring to the length of the walk, we mean the total number of steps taken, including steps in which the walk remains at the current vertex, while the length of the corresponding path does not include these steps.

Theorem 2 *The algorithm* Test-Bipartite *constitutes a tester for bipartiteness with complexity* $\mathrm{poly}((\log N)/\epsilon) \cdot \sqrt{N}$. *Specifically,*

- *If G is bipartite then the algorithm always accepts.*
- *If G is ϵ-far from being bipartite then the algorithm rejects with probability at least $\frac{2}{3}$. Furthermore, whenever the algorithm rejects a graph it outputs a* certificate *to the non-bipartiteness of the graph in form of an odd-length cycle of length* $\mathrm{poly}(\epsilon^{-1} \log N)$.

[2] We note that, for sake of simplicity, this definition slightly differs from that discussed in the Introduction and in [GR96]. There, ϵ is the fraction of entries that should be modified in the graph representation. This means that each (undirected) edge (v, u) in G is counted twice - once as an entry $[v, i]$ and once as an entry $[u, j]$.

Algorithm Test-Bipartite

- Repeat $T = \Theta(\frac{1}{\epsilon})$ times:
  1. Uniformly select s in V.
  2. If odd-cycle(s) returns found then reject.
- In case the algorithm did not reject in any one of the above iterations, it accepts.

odd-cycle(s)

1. Let $K \stackrel{\text{def}}{=} \mathrm{poly}((\log N)/\epsilon) \cdot \sqrt{N}$, and $L \stackrel{\text{def}}{=} \mathrm{poly}((\log N)/\epsilon)$;
2. Perform K random walks starting from s, each of length L;
3. If some vertex v is reached (from s) both on a prefix of a random walk corresponding to an even-length path and on a walk-prefix corresponding to an odd-length path then return found. Otherwise, return not-found.

Figure 1: Algorithm **Test-Bipartite** and Procedure odd-cycle.

4 Analysis of the Algorithm

The completeness part of Theorem 2 (i.e., accepting bipartite graphs) is straightforward. We focus on proving the soundness of the algorithm (i.e., that ϵ-far graphs are rejected with probability $\frac{2}{3}$). What we eventually show (in Subsection 4.6) is the counterpositive. Namely, that if the test accepts G with probability greater than $\frac{1}{3}$ then there exists an ϵ-good partition of G. We start with an overview of our analysis.

The Rapidly–Mixing Case. To gain intuition, consider first the following "ideal" case: From each starting vertex s in G, and for every $v \in \mathrm{V}$, the probability that a random walk of length $L = \mathrm{poly}((\log N)/\epsilon)$ ends at v is at least $\frac{1}{2N}$ and at most $\frac{2}{N}$ – i.e., approximately the probability assigned by the stationary distribution. (Note that this ideal case occurs when G is an expander). Let us fix a particular starting vertex s. For each vertex v, let p_v^0 be the probability that a random walk (of length L) starting from s, ends at v and corresponds to an even-length path. Define p_v^1 analogously for odd-length paths. Then, by our assumption on G, for every v, $p_v^0 + p_v^1 \geq \frac{1}{2N}$. We consider two cases regarding the sum $\sum_{v \in \mathrm{V}} p_v^0 \cdot p_v^1$ — In case the sum is (relatively) "small", we show that there exists a partition $(\mathrm{V}_0, \mathrm{V}_1)$ of V that is ϵ-good, and so G is ϵ-close to being bipartite. Otherwise (i.e., when the sum is not "small"), we show that $\Pr[\text{odd-cycle}(s) = \text{found}]$ is constant. This implies that in case G is accepted with probability at least $\frac{1}{3}$ then G is ϵ-close to being bipartite. In what follows we give some intuition concerning the two cases.

Consider first the case in which $\sum_{v \in \mathrm{V}} p_v^0 \cdot p_v^1$ is smaller than $c \cdot \frac{\epsilon}{N}$ for some suitable constant $c < 1$. Let the partition $(\mathrm{V}_0, \mathrm{V}_1)$ be defined as follows: $\mathrm{V}_0 = \{v : p_v^0 \geq p_v^1\}$ and $\mathrm{V}_1 = \{v : p_v^1 > p_v^0\}$. Consider a particular vertex $v \in \mathrm{V}_0$. By definition of V_0 and our rapid-mixing assumption, $p_v^0 \geq \frac{1}{4N}$. Assume v has neighbors in V_0. Then for each such neighbor u, $p_u^0 \geq \frac{1}{4N}$ as well. However, since there is a probability of $\frac{1}{2d}$ of taking a transition from u to v in walks on G, we can infer that each neighbor u contributes $\Omega(\frac{1}{2d} \cdot \frac{1}{4N})$ to the probability p_v^1. Thus, if there are many violating edges with respect to $(\mathrm{V}_0, \mathrm{V}_1)$, then the sum $\sum_{v \in \mathrm{V}} p_v^0 \cdot p_v^1$ is large, contradicting our case hypothesis.

We now turn to the second case ($\sum_{v \in \mathrm{V}} p_v^0 \cdot p_v^1 \geq c \cdot \frac{\epsilon}{N}$). For every fixed pair $i, j \in \{1, \ldots, K\}$, (recall that $K = \Omega(\sqrt{N})$ is the number of walks taken from s), consider the 0/1 random variable that is 1 if and only if both the i^{th} and the j^{th} walk end at the same vertex v but correspond to paths with different parity. Then the expected value of each random variable is $\sum_{v \in \mathrm{V}} 2 \cdot p_v^0 \cdot p_v^1$. Since there are $K^2 = \Omega(N)$ such variables, the expected value of their sum is greater than 1. These random variables are not pairwise independent, nonetheless we can obtain a constant bound on the probability that the sum is 0 using Chebyshev's inequality (cf., [AS92a, Sec. 4.3]).

The General Case. Unfortunately, we may not assume in general that for every (or even some) starting vertex, all (or even almost all) vertices are reached with probability $\Theta(1/N)$. Instead, for each vertex s, we may consider the set of vertices that are reached from s with relatively high probability on walks of length $L = \mathrm{poly}((\log N)/\epsilon)$. As was done above, we could try and partition these vertices according to the probability that they are reached on random walks corresponding to even-length and odd-length paths, respectively. The difficulty that arises is how to combine the different partitions induces by the different starting vertices, and how to argue that there are few violating edges between vertices partitioned according to one starting vertex and vertices partitioned according to another (assuming they are exclusive).

To overcome this difficulty, we proceed in a slightly different manner. Let us call a vertex s *good*, if the probability that odd-cycle(s) returns found is at most 0.1. Then, assuming G is accepted with probability greater than $\frac{1}{3}$, all but at most $\frac{\epsilon}{16}$ of the vertices are *good*. We define a partition in stages as follows. In the first stage we pick any *good* vertex s. What we can show is that not only is there a set of vertices S that are reached from s with high probability and can be partitioned without many violations (due to the goodness of s), but also that there is a small cut between S and the rest of the graph. Thus, no matter how we partition the rest of the vertices, there cannot be many violating edges between S and $\mathrm{V} \setminus \mathrm{S}$. We therefore partition S (as above), and continue with the rest of the vertices in G.

In the next stage, and those that follow, we consider the subgraph H induces by the yet "unpartitioned" vertices. If $|\mathrm{H}| < \frac{\epsilon}{4}N$ then we can partition H arbitrarily and stop since the total number of edges adjacent to vertices in H is less than $\frac{\epsilon}{4} \cdot dN$. If $|\mathrm{H}| \geq \frac{\epsilon}{4}N$ then we can show that any *good* vertex s in H that has a certain additional property (which at least half of the vertices in H have), determines a set S (whose vertices are reached with high probability from s) with the following properties: S can be partitioned without having many violating edges among vertices in S; and there is a small cut between S and the rest of H. Thus, each such set S accounts for the violating edges between pairs of vertices that both belong to S as well as edges between pairs of vertices such that one vertex belongs to S and one to $\mathrm{V}(\mathrm{H}) \setminus \mathrm{S}$. Adding it all together, the total number of violating edges with respect to the final partition is at most $\epsilon \cdot dN$.

THE SET S. To prove the existence of such sets S, consider first the initial stage in the partition process (i.e., here $\mathrm{H} = \mathrm{G}$). Recall that in this stage we are looking for a subset of vertices $\mathrm{S} \subseteq \mathrm{V}$, all reached with relatively high probability from some good vertex s, that are separated from the rest of G by relatively few edges. From the previous discussion we know that if for all (or almost all) vertices v in G, a random walk of length $\mathrm{poly}((\log N)/\epsilon)$ starting from s ends at v with probability $\Theta(1/N)$ then we can define a good partition of all of G and be done. Thus assume we are not in this case. Namely, there is a significant fraction of vertices that are reached from s with probability that differs significantly from $1/N$. In other words, the distribution on the ending vertices (when starting from s) is far from stationary. What we can show (using techniques of Mihail [Mih89]) is that this implies the existence of a small cut between some set of vertices S that are each reached from s with probability that is roughly $1/\sqrt{|S| \cdot N}$ and the rest of G. Furthermore, we can show that S has an additional property that combined with the fact that s is good implies that it can be partitioned without having many violating edges.

In the next stages of the partition process, we would have liked to apply the same techniques to determine small cuts (with other desired properties) in subgraphs H of G. If we could at each stage "cut-away" the subgraph H from the rest of G and perform walks only inside H then we would have proceeded as in the first stage. However, these subgraphs H are only determined by the analysis while the algorithm, oblivious to the analysis, always performs random walks on all of G. Therefore we would like to have a way to map walks in G to walks in H so that probabilities of events occurring in imaginary walks on H can be related to events occurring in the real walks on G. Consider a walk of length L in G that starts at s in H. Suppose we remove from this walk all steps outside of H and refer to the remaining sequence of steps as the *restriction* of the walk to H. If the walk never takes long excursions outside of H, then for sufficiently large L, the restriction of the walk to H is sufficiently long for our purposes (i.e. proving the existence of a set S with the desired properties). However, if the walk does take long excursions (and in particular if it exits H and does not return within L steps) then it is not useful for our purposes.

THE MARKOV CHAIN. To model both the undesired long excursions, and the fact that we want to disregard (or contract to one step) the short excursions, we define, for any given subgraph H of G, an auxiliary Markov Chain. The states of the Markov Chain are the vertices of H and some additional auxiliary states. We prove several claims concerning the chain, and in particular relate random walks on the chain to random walks on G. The basic idea is that short excursions out of H starting at $v \in \mathrm{H}$ and ending at $u \in \mathrm{H}$ (in walks on G), are translated (in the Markov Chain) to a single transition between v and u. On the other hand, long excursions are translated to walks outside of H (on auxiliary paths) that effectively do not return to H (when performing walks of a particular length on the Markov Chain). We then show that for a suitable choice of "long" and "short", for at least half of the starting vertices in H, (which we refer to as *useful* vertices) the probability of entering an auxiliary path in the Markov Chain (which corresponds to exiting H for a long excursion in $\mathrm{G} \setminus \mathrm{H}$) is small.

Armed with this property of the Markov Chain, we prove that for every useful starting vertex s in H there exists a subset of vertices S in H that are all reached with high probability from s and are separated from the rest of H by a small cut. We then give sufficient conditions (on s and S) under which the set S can be partitioned without many violations. In case these conditions are not satisfied then we show that a sufficient number of walks starting from s in the Markov Chain, will detect an odd cycle with probability greater than 0.1. Based on the definition of the Markov Chain, these conditions (for the same s and S) also imply that (slightly longer) walks on G will detect an odd cycle in G with probability greater than 0.1. Combining all the above we prove Theorem 2.

Organization. In Subsection 4.1 we define the Markov Chain discussed above. In Subsection 4.2 we bound the probability of entering auxiliary paths in the Markov Chain (i.e., taking long excursions outside of H) for most starting vertices. In Subsection 4.3 we determine the set S (discussed above). Subsections 4.4 and 4.5 present a dichotomy: Either S can be partitioned without many violations, or an odd cycle is detected with non-negligible probability. The proof is wrapped up in Subsection 4.6.

4.1 The Markov Chain $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$

Let H be a subgraph of G. For any given pair of lengths, ℓ_1 and ℓ_2, we define a Markov Chain $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. Roughly speaking, $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ captures random walks of length at most $\ell_1 \cdot \ell_2$ in G that do not exit H for (sub)walks of length ℓ_2 or more. The states of the chain consist of the vertices of H and some additional auxiliary states. For vertices that do not have neighbors outside of H, the transition probabilities in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ are exactly as in walks on G. However, for vertices v that have neighbors outside of H there are two modifications: (1) For each vertex u, the transition probability from v to u, denoted $q_{u,v}$, is the probability of a walk (in G) starting from v and ending at u after less than ℓ_2 steps (without passing through any other vertex in H). Thus, walks of length less than ℓ_2 out of H (and in particular the walk $v - u$ in case $(v, u) \in \mathrm{E}$), are contracted into single transitions. (2) There is an auxiliary path of length ℓ_1 emitting from v. The transition probability from v to the first auxiliary vertex on the path equals the probability that a walk starting from v exits H and does not return in less than ℓ_2 steps. From the last auxiliary vertex on the auxiliary path there are transitions to vertices in H with the corresponding conditional probabilities of reaching them after such a walk.

A more formal definition of $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ follows. For every vertex v in H we have a state v in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. For simplicity, we shall continue referring to these states as vertices. Let the *border* of H, denoted $\mathrm{B}(\mathrm{H})$, be the set of vertices in H that have at least one neighbor in G that is not in H. Then, for every vertex $v \in \mathrm{B}(\mathrm{H})$, we have a set $a_{v,1}, \ldots, a_{v,\ell_1}$ of *auxiliary* states. Let $p_{v,u}^{\mathrm{H}}(t)$ denote the probability of a walk of length t that starts at v and ends at u without passing through any other vertex in H. Namely, it is the sum over all such walks w, of the product, taken over all steps in w, of the transition probabilities of these steps. In particular, $p_{v,v}^{\mathrm{H}}(1) \geq \frac{1}{2}$ (where equality holds in case v has degree d), and for every $u \in \Gamma(v)$, $p_{v,u}^{\mathrm{H}}(1) = \frac{1}{2d}$. The transition probabilities, $q_{x,y}$, in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ are defined as follows:

- For every v and u in H, $q_{v,u} = \sum_{t=1}^{\ell_2-1} p_{v,u}^{\mathrm{H}}(t)$.

  Thus, $q_{v,u}$ is a sum of $p_{v,u}^{\mathrm{H}}(1)$ and $\sum_{t=2}^{\ell_2-1} p_{v,u}^{\mathrm{H}}(t)$. The first term implies that for every v in H, $q_{v,v} \geq \frac{1}{2}$ and for every pair of neighbors v and u, $q_{v,u} \geq \frac{1}{2d}$. The second term, which we refer to as the *excess* probability is due to walks of length less than ℓ_2 (from v to u) passing through vertices outside of H, and can be viewed as *contraction* of these walks.

  Note that for every pair of vertices v and u, $q_{v,u} = q_{u,v}$.

- For every $v \in \mathrm{B}(\mathrm{H})$, $q_{v,a_{v,1}} = \sum_{u \in \mathrm{H}} \sum_{t \geq \ell_2} p_{v,u}^{\mathrm{H}}(t)$; for every ℓ, $1 \leq \ell < \ell_1$, $q_{a_{v,\ell},a_{v,\ell+1}} = 1$; and for every $u \in \mathrm{H}$, $q_{a_{v,\ell_1},u} = \frac{1}{q_{v,a_{v,1}}} \cdot \sum_{t \geq \ell_2} p_{v,u}^{\mathrm{H}}(t)$.

  In other words, $q_{v,a_{v,1}}$ is the probability that a random walk in G that starts from v takes at least ℓ_2 steps outside of H before returning to H, and $q_{a_{v,\ell_1},u}$ is the conditional probability of reaching u in such a walk. Thus, the auxiliary states form auxiliary paths in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, where these paths correspond to walks of length at least ℓ_2 outside of H.

We shall restrict our attention to walks of length at most ℓ_1 in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, and hence any walk that starts at a vertex of H and enters an auxiliary path never returns to vertices of H. For any two states y, z in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ let $q_{y,z}(t)$ be the probability that a walk of length t starting from y ends at z. We further let the parity of the lengths of paths corresponding to walks in G be carried on to $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. That is, each transition between vertices v and u that corresponds to walks outside of H consists of two transitions – one due to even-length paths corresponding to walks from v to u outside of H, and one to odd-length paths. For any two vertices in H we let $q_{v,u}^{\sigma}(t)$ denote the probability in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ of a walk of length t starting from v, ending at u, and corresponding to a path whose length has parity σ.

In all that follows we assume that G is connected. Our analysis can easily be modified to deal with the case in which G is not connected, simply by treating separately each of its connected components. Under the assumption that G is connected, for every v and u in H, there exists a t such that $q_{u,v}(t) > 0$, and hence $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ is irreducible. Furthermore, because for each $v \in \mathrm{H}$ $q_{v,v} \geq \frac{1}{2}$, $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ is also aperiodic. Thus it has a unique stationary distribution.

4.2 Probability of Long walks Outside of H

In our first lemma we show that the probability of entering an auxiliary path while taking walks of length at most ℓ_1 in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, starting from a uniformly chosen vertex in H, is small, provided $\ell_1 \ll \ell_2$. This implies that

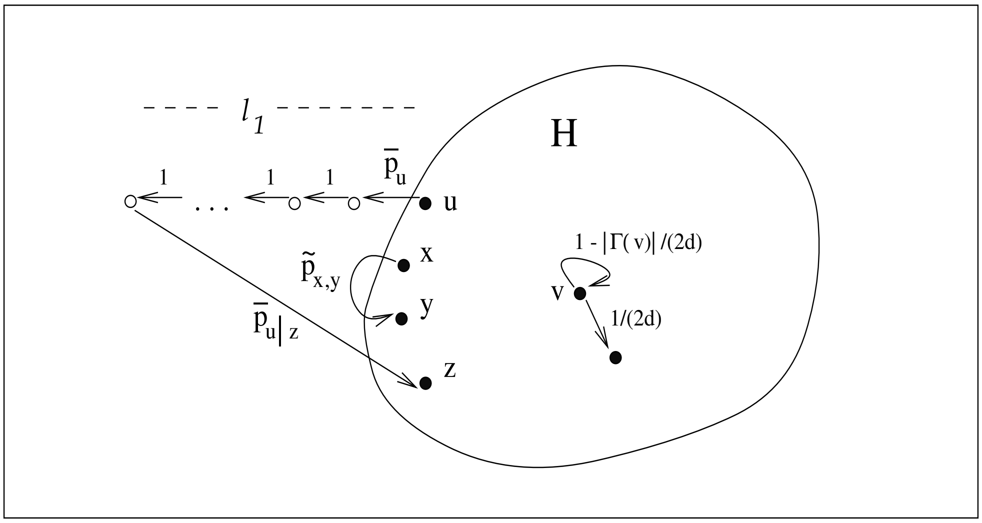

Figure 2: The structure of $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. The states corresponding to vertices of H are depicted as black dots, and the auxiliary states as white ones. Here $\tilde{p}_{x,y} \stackrel{\rm def}{=} \sum_{t=1}^{\ell_2-1} p_{x,y}^{\mathrm{H}}(1)$, $\bar{p}_u \stackrel{\rm def}{=} \sum_{z\in\mathrm{H}} \sum_{t\geq\ell_2} p_{u,z}^{\mathrm{H}}(t)$, and $\bar{p}_{u|z} \stackrel{\rm def}{=} \frac{1}{\bar{p}_u} \cdot \sum_{t\geq\ell_2} p_{u,z}^{\mathrm{H}}(t)$.

for $L = \ell_1 \cdot \ell_2$, with high probability, a random walk of length L in G (starting from a uniformly chosen vertex in H), will perform at least ℓ_1 steps in H.

Lemma 4.1 *Let* H *be a subgraph of* G*, and* ℓ_1 *and* ℓ_2 *be integers. The probability that a walk in* $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ *starting from a uniformly chosen vertex of* H *enters an auxiliary path after at most* ℓ_1 *steps, is at most* $\frac{\ell_1}{\ell_2} \cdot \frac{|\mathrm{G}|}{|\mathrm{H}|}$.

We first establish the following related lemma that refers to random walks in G (as opposed to random walks in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, which are considered in Lemma 4.1). Phrased slightly differently, Lemma 4.2 says that if we uniformly choose a vertex in G, then the probability that in the next step we start a walk that exits H and does not return to H in less than ℓ_2 steps, is at most $\frac{1}{\ell_2}$. (In particular, for every starting vertex $v \notin \mathrm{H}$ the contribution to this probability is 0.)

Lemma 4.2 $\sum_{v,u\in\mathrm{H}} \sum_{t\geq\ell_2} p_{v,u}^{\mathrm{H}}(t) \quad \leq \quad \frac{|\mathrm{G}|}{\ell_2}$.

Proof: To prove the lemma we define an additional Markov Chain, which we denote by $\mathrm{M}(\mathrm{H})$. The chain $\mathrm{M}(\mathrm{H})$ is used to describe random walks in G (of any length), where the parts of the walks that are outside of H pass through auxiliary states. For each vertex v in H we have a state in $\mathrm{M}(\mathrm{H})$. For every pair of vertices v and u in $\mathrm{B}(\mathrm{H})$, and for every $t \geq 2$ such that there exists a walk of length t between v and u outside of H, we have two sets of $t-1$ auxiliary states — one set creates a path of length t from v to u, and one set creates a path from u to v.

The transition probabilities $q_{v,u}$ in $\mathrm{M}(\mathrm{H})$ are defined as follows. For every $v \in \mathrm{H}$, $q_{v,v} = 1 - \frac{|\Gamma(v)|}{2d}$. For every $v, u \in \mathrm{H}$ such that $u \in \Gamma(v)$, $q_{v,u} = \frac{1}{2d}$. For every pair of vertices v and u in $\mathrm{B}(\mathrm{H})$ and for every $t \geq 2$ (such that u can be reached from v in a walk of length t outside of H), the probability of entering the auxiliary path connecting u to v is $p_{v,u}^{\mathrm{H}}(t)$; for each auxiliary state on the path, the transition probability to the next state is 1, and the last state goes with probability 1 to u. Let $\pi_s^{\mathrm{M}(\mathrm{H})}$ be the probability assigned to state s by the stationary distribution of $\mathrm{M}(\mathrm{H})$. The following claim, whose proof is provided in Appendix A, says that for every vertex v in H, the stationary probability of v is the same as in walks on G.

Claim 1: *For every* $v \in \mathrm{H}$, $\pi_v^{\mathrm{M(H)}} = \frac{1}{N}$.

By construction of $\mathrm{M(H)}$, for every pair of vertices v and u in $\mathrm{B(H)}$, and for every $t \geq 2$, the stationary probability of the first auxiliary state on the corresponding auxiliary path is $\pi_v^{\mathrm{M(H)}} \cdot p_{v,u}^{\mathrm{H}}(t)$. This is true since this state has only one incoming transition, and this transition is from v. By definition of the transition probabilities on auxiliary paths, for every $2 \leq \ell < t-1$, the stationary probability of the ℓ^{th} auxiliary state on the path is $\pi_v^{\mathrm{M(H)}} \cdot p_{v,u}^{\mathrm{H}}(t)$ as well. Let $\Pi_{v,u,t}$ denote the total stationary distribution on the auxiliary path of length t from v to u. Then, on one hand $\Pi_{v,u,t} = t \cdot \pi_v^{\mathrm{M(H)}} \cdot p_{v,u}^{\mathrm{H}}(t)$, and on the other hand, since all paths are disjoint, $\sum_{v,u \in \mathrm{H},\, t \geq 2} \Pi_{v,u,t} < 1$. It follows that

$$\sum_{v,u \in \mathrm{B(H)},\, t \geq \ell_2} \pi_v^{\mathrm{M(H)}} \cdot p_{v,u}^{\mathrm{H}}(t) \;=\; \sum_{v,u \in \mathrm{B(H)},\, t \geq \ell_2} \frac{1}{t} \cdot \Pi_{v,u,t} \;\leq\; \sum_{v,u \in \mathrm{B(H)},\, t \geq \ell_2} \frac{1}{\ell_2} \cdot \Pi_{v,u,t} \;<\; \frac{1}{\ell_2}\,.$$

Since by Claim 1, for every $v \in \mathrm{H}$, $\pi_v^{\mathrm{M(H)}} = \frac{1}{N}$, Lemma 4.2 follows. ∎

Proof of Lemma 4.1: Let $\mathrm{M} \stackrel{\mathrm{def}}{=} \mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, and recall that $|\mathrm{G}| = N$. Observe that in case $\ell_2 < \frac{N}{|\mathrm{H}|} \cdot \ell_1$ then the claim holds trivially. Thus, assume $\ell_2 \geq \frac{N}{|\mathrm{H}|} \cdot \ell_1$. We first prove that the probabilities assigned by the stationary distribution to all vertices in H are the same, and each is bounded below by $\frac{1}{2} \cdot \frac{1}{|H|}$. Let π_s^{M} denote the probability assigned to state s by the stationary distribution of M. We first show that a distribution that assigns the same probability, π to each vertex is stationary.

Consider any vertex v. Then $\pi_v^{\mathrm{M}} = \sum_{z \in \mathrm{M}} \pi_z^{\mathrm{M}} \cdot q_{z,v}$. We need to show that this sum is in fact π. For each of the neighbors u of v in H, there is a contribution of $\pi_u^{\mathrm{M}} \cdot \frac{1}{2d}$, which by our assumption is $\pi \cdot \frac{1}{2d}$. Hence, the neighbors of v in H contribute a total of $\pi \cdot \frac{|\Gamma(v) \cap \mathrm{H}|}{2d}$. The transition from v to itself contributes an additional term of $\pi \cdot (1 - \frac{|\Gamma(v)|}{2d})$. In case $v \notin \mathrm{B(H)}$ we are done since all of v's neighbors are in H (and for every other state z, $q_{z,v} = 0$). Otherwise, there are two additional contributions. The first is due to walks of length less than ℓ_2 outside of H that start at some u in H and end at v, which are translated in M into a transition from u to v with probability $\sum_{t=2}^{\ell_2 - 1} p_{u,v}^{\mathrm{H}}(t)$. (In case there is an edge between u and v, this is the excess probability between u and v.) Since $p_{u,v}^{\mathrm{H}}(t) = p_{v,u}^{\mathrm{H}}(t)$, the total contribution of these transitions is $\pi \cdot \sum_{u \in \mathrm{H}} \sum_{t=2}^{\ell_2 - 1} p_{v,u}^{\mathrm{H}}(t)$. The other contribution is due to walks of length at least ℓ_2 outside of H that start at some u in H and end at v, which are translated into a transition from the auxiliary state a_{u,ℓ_1} to v.

By construction of the chain, for every auxiliary path emitting from a vertex u, all states on the path have equal stationary probability, and this probability is $\pi_u^{\mathrm{M}} \cdot q_{u,a_{u,1}}$ Since the transition probability from a_{u,ℓ_1} to v is $\frac{1}{q_{u,a_{u,1}}} \cdot \sum_{t \geq \ell_2} p_{u,v}^{\mathrm{H}}(t)$, (and $p_{u,v}^{\mathrm{H}}(t) = p_{v,u}^{\mathrm{H}}(t)$), the total contribution from these transitions is is $\pi \cdot \sum_{u \in \mathrm{H}} \sum_{t \geq \ell_2} p_{v,u}^{\mathrm{H}}(t)$. Together, the contribution of transitions that are due to walks outside of H is $\pi \cdot \sum_{u \in \mathrm{H}} \sum_{t \geq 2} p_{v,u}^{\mathrm{H}}(t)$. This expression equals to π times the probability of taking a transition from v to some vertex outside of H and is thus $\pi \cdot \frac{|\Gamma(v) \setminus \mathrm{H}|}{2d}$. Summing all contributions, we get that for every $v \in \mathrm{H}$,

$$\pi_v^{\mathrm{M}} \;=\; \pi \cdot \frac{|\Gamma(v) \cap \mathrm{H}|}{2d} + \pi \cdot \left(1 - \frac{|\Gamma(v)|}{2d}\right) + \pi \cdot \frac{|\Gamma(v) \setminus \mathrm{H}|}{2d} \;=\; \pi$$

Next we prove that $\pi \geq \frac{1}{2}\frac{1}{|\mathrm{H}|}$. We use the fact that the probabilities assigned by the stationary distribution must sum to 1. The contribution of the vertices of H is $|\mathrm{H}| \cdot \pi$. The total probability assigned by the stationary distribution to auxiliary states is

$$\sum_{v \in \mathrm{H}} \ell_1 \cdot \pi \cdot \sum_{u \in \mathrm{H}} \sum_{t \geq \ell_2} p_{v,u}^{\mathrm{H}}(t)$$

which by Lemma 4.2 is at most $\pi \cdot \frac{\ell_1 \cdot N}{\ell_2}$, and by our assumption that $\ell_2 \geq \ell_1 \frac{N}{|\mathrm{H}|}$, is bounded by $\pi \cdot |\mathrm{H}|$. Thus, $\pi \geq \frac{1}{2|\mathrm{H}|}$.

For any state s, let Ψ_s denote the event that a walk starting from s enters an auxiliary path in at most ℓ_1 steps. Let $s \sim U_H$ denote choosing s uniformly in H, and let $s \sim \pi^{\mathrm{M}}$ denote choosing s according to the stationary distribution of M. Then, from what we have shown concerning the stationary distribution of the vertices of H, it follows that

$$\Pr_{s\sim U_H}[\Psi_s] \;=\; \Pr_{s\sim\pi^{\mathrm{M}}}[\Psi_s \mid s \in \mathrm{H}] \;=\; \frac{\Pr_{s\sim\pi^{\mathrm{M}}}[\Psi_s \text{ and } s \in \mathrm{H}]}{\Pr_{s\sim\pi^{\mathrm{M}}}[s \in \mathrm{H}]} \;\leq\; \frac{\Pr_{s\sim\pi^{\mathrm{M}}}[\Psi_s]}{\Pr_{s\sim\pi^{\mathrm{M}}}[s \in \mathrm{H}]} \leq 2 \cdot \Pr_{s\sim\pi^{\mathrm{M}}}[\Psi_s]$$

But

$$\begin{aligned}
\Pr_{s\sim\pi^{\mathrm{M}}}[\Psi_s] &= \sum_{t=1}^{\ell_1} \Pr_{s\sim\pi^{\mathrm{M}}}[\text{ a walk starting from } s \text{ enters an aux. path at step } t] \\
&= \ell_1 \cdot \sum_{v\in \mathrm{B(H)}} (\text{ stationary prob. on aux. edge from } v \text{ to } a_{v,1}) \\
&= \ell_1 \cdot \sum_{v\in\mathrm{H}} \pi_v^{\mathrm{M}} \cdot \sum_{u\in\mathrm{H}} \sum_{t\geq \ell_2} p_{v,u}^{\mathrm{H}}(t) \;\leq\; \frac{\ell_1}{\ell_2} \cdot \frac{N}{|\mathrm{H}|}
\end{aligned}$$

where the last inequality follows from Lemma 4.2 and the fact that $\pi_v^{\mathrm{M}} \leq \frac{1}{|\mathrm{H}|}$. The lemma follows. ■

Definition 4.1 *We say that a vertex s is* useful *with respect to $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ if the probability that a walk in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ starting from s enters an auxiliary path after at most ℓ_1 steps, is at most $\frac{2\ell_1}{\ell_2} \cdot \frac{|\mathrm{G}|}{|\mathrm{H}|}$.*

As a direct corollary to Lemma 4.1 (using Markov's inequality), we obtain

Corollary 3 *Let H be a subgraph of G, and ℓ_1 and ℓ_2 be integers. Then at least half of the vertices s in H are useful with respect to $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$.*

4.3 Determining the Set S

In the following lemma we adapt techniques used by Mihail [Mih89]. While Mihail showed that high expansion leads to fast convergence of random walks to the stationary distribution, we show that too slow of a convergence implies small cuts that have certain additional properties. In particular, the vertices on one side of the cut can be reached with roughly the same, relatively high probability from some vertex s. The places where we diverge from Mihail's analysis, (which in parts we follows quite closely), are when we use the specific properties of the Markov Chain $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, in order to obtain the additional properties of the cut.

Lemma 4.3 *Let H be a subgraph of G with at least $\frac{\epsilon}{4}N$ vertices, and let $\ell_1 = \Theta\left(\left(\frac{\log(N/\epsilon)}{\epsilon^3}\right)^3\right)$, $\ell_2 = \Theta\left(\frac{\ell_1}{\epsilon^2}\right)$, and $F = O\left(\frac{1}{\epsilon}\right)$. Then for every vertex s that is* useful *with respect to $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, there exists a subset of vertices S in H an integer t, $\ell_1/2 \leq t \leq \ell_1$, and a value $\beta = \Omega\left(\frac{\epsilon^2}{\log(N/\epsilon)}\right)$, such that:*

1. *The number of edges between S and the rest of H is at most $\frac{\epsilon}{2} \cdot d \cdot |\mathrm{S}|$.*
2. *For every $v \in \mathrm{S}$, $\sqrt{\frac{1}{|\mathrm{S}|} \cdot \frac{\beta}{|\mathrm{H}|}} \leq q_{s,v}(t) \leq F \cdot \sqrt{\frac{1}{|\mathrm{S}|} \cdot \frac{\beta}{|\mathrm{H}|}}$;*

We start with an overview of this rather technically involved proof. Let $\mathrm{M} \stackrel{\mathrm{def}}{=} \mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, and fix a useful starting vertex s in H. In the proof we consider two cases. In the first (easy) case, there exists t, $\ell_1/2 \leq t \leq \ell_1$, such that for all by at most $\frac{\epsilon}{8}|\mathrm{H}|$ of the vertices v in H, $q_{s,v}(t) \geq \frac{1}{2}\pi_v^{\mathrm{M}}$, where π_v^{M} is the probability assigned by the stationary distribution of M to v. In other words, in this case almost all vertices in H are reached with probability that is not much smaller than that assigned by the stationary distribution. Here we let S be the subset of these vertices that are not reached with much higher probability as well.

In the second (and main) case, we have that for every t between $\ell_1/2$ and ℓ_1, for at least $\frac{\epsilon}{8}|\mathrm{H}|$ of the vertices v in H, $q_{s,v}(t) < \frac{1}{2}\pi_v^{\mathrm{M}}$. This means that the walk on M is not rapidly mixing. Using the counterpositive of the standard rapid mixing analysis, one may infer that there is a relatively small "cut" in M. However, this is not sufficient for our goal for several reasons. Firstly, we are interested in a small cut in H (while a small cut in M might involve auxiliary states). Secondly, we are interested in a cut that has the additional property stated in the lemma. Fortunately, we are able to adapt the specific analysis of Mihail [Mih89] to overcome both problems. Building on Mihail's formulation, we first restrict our attention to the states of M that correspond to vertices in H, where here we use the hypothesis that s is useful (see Definition 4.1). Furthermore, we consider as candidates for the set S only those vertices that are reached from s with probability that is greater than the stationary probability. We can then obtain a relatively small cut for which all vertices v's with $q_{s,v}(t)$ above some value are on one side and the rest on the other. Using a more careful analysis we determine a cut, $(\mathrm{S}, \mathrm{V}(\mathrm{H}) \setminus \mathrm{S})$, with the extra properties required in the lemma. In particular, for each $v \in \mathrm{S}$, $q_{s,v}(t)$ is relatively big, and all these values are of about the same size.

Proof: By the lemma's hypotheses concerning the size of H and the ratio between ℓ_1 and ℓ_2, and by the definition of a *useful* vertex (Definition 4.1), for every *useful* vertex s, the probability that a walk starting from s will enter an auxiliary path in at most ℓ_1 steps is less than $\epsilon/256$ (for the appropriate choice of constants in the $\Theta(\cdot)$ notation of ℓ_2). In other words, for each *useful* s, and for every $t \leq \ell_1$, the sum over all auxiliary states a, of $q_{s,a}(t)$, is bounded above by $\epsilon/256$.

Fix a *useful* vertex s. For every step $t \leq \ell_1$, and for each state z in M, let $e_z(t) \stackrel{\mathrm{def}}{=} q_{s,z}(t) - \pi_z$ where for notational convenience we let $\pi_z = \pi_z^{\mathrm{M}}$ denote the probability assigned by the stationary distribution of M to z. That is, $e_z(t)$ measures the difference between the probability of being at state z at time t (when starting from s) and the stationary probability of z. Recall (from the proof of Lemma 4.1), that for every vertex $v \in \mathrm{H}$, π_v has the same value, and this value is at least $\frac{1}{2\cdot|\mathrm{H}|}$ and at most $\frac{1}{|\mathrm{H}|}$. By the above definition, for every t, $\sum_z e_z(t) = 0$, and $\vec{e}(t+1) = \vec{e}(t) \cdot \mathrm{M}$, where we use the same notation, M, for the Markov Chain and its transition matrix. Let $\|\vec{e}(t)\| \stackrel{\mathrm{def}}{=} \sum_z e_z^2$ denote the Euclidean norm (squared) of the discrepancy vector $\vec{e}$. and let $\|\vec{e}^{\mathrm{H}}(t)\| \stackrel{\mathrm{def}}{=} \sum_{v\in\mathrm{H}} e_v^2$ be the contribution to the norm from vertices in H.

Case 1 (easy): Suppose that there exists t, $\ell_1/2 \leq t \leq \ell_1$, such that for all by at most $\frac{\epsilon}{8}|\mathrm{H}|$ of the vertices v in H, $e_v(t) \geq -\frac{1}{2}\pi_v$ (i.e., $q_{s,v}(t) \geq \frac{1}{2}\pi_v$). In other words, almost all vertices in H are reached with probability that is not much smaller than that assigned by the stationary distribution. Denote the set of these vertices by W. Thus, for each v in W,

$$q_{s,v}(t) \;\geq\; \frac{1}{2}\pi_v \;=\; \Omega\left(\frac{1}{|\mathrm{H}|}\right) \;=\; \Omega\left(\sqrt{\frac{1}{|\mathrm{W}|\cdot|\mathrm{H}|}}\right)$$

Set S to be the subset of vertices v in W for which $q_{s,v}(t)$ is at most $F = O(1/\epsilon)$ times this lower bound. Since $\sum_{v\in\mathrm{W}} q_{s,v}(t) \leq 1$, and $|\mathrm{W}| > (1-\frac{\epsilon}{8})|\mathrm{H}|$, for the appropriate constant in the $O(\cdot)$ notation, $|\mathrm{W} \setminus \mathrm{S}| \leq \frac{\epsilon}{4}\cdot|\mathrm{S}|$. Furthermore, by definition of W, (and the lower bounds on the sizes of W and S), $|\mathrm{V}(\mathrm{H}) \setminus \mathrm{W}| \leq \frac{\epsilon}{4}\cdot|\mathrm{S}|$. Therefore, $|\mathrm{V}(\mathrm{H}) \setminus \mathrm{S}| \leq \frac{\epsilon}{2}\cdot|\mathrm{S}|$ and so the number of edges between S and the rest of H is at most $\frac{\epsilon}{2}\cdot d|\mathrm{S}|$ as required.

Case 2 (main case): We turn to the case in which for every t between $\ell_1/2$ and ℓ_1, for at least $\frac{\epsilon}{8}|\mathrm{H}|$ of the vertices v in H, $e_v(t) < -\frac{1}{2}\pi_v$. We prove the lemma for this case by a series of claims, (all using the same hypotheses as the lemma, and the case hypothesis). We first note that under the case hypothesis and the fact that for every $v \in$H, $\pi_v \geq \frac{1}{2|\mathrm{H}|}$,

$$\|\vec{e}^{\mathrm{H}}(t)\| \;>\; \frac{\epsilon}{8}\cdot|\mathrm{H}|\cdot\left(-\frac{1}{4|\mathrm{H}|}\right)^2 \;=\; \frac{\epsilon}{128\cdot|\mathrm{H}|}$$

for every $\ell_1/2 \leq t \leq \ell_1$. In particular this is true for $\|\vec{e}^{\mathrm{H}}(\ell_1)\|$.

Claim 1: *There exists $\bar{t}$, $\ell_1/2 \leq \bar{t} < \ell_1$, such that*

$$\|\vec{e}^{\mathrm{H}}(\bar{t})\| - \|\vec{e}^{\mathrm{H}}(\bar{t}+1)\| \;<\; \delta^2\cdot\|\vec{e}^{\mathrm{H}}(\bar{t})\|$$

where $\delta = O\left(\sqrt{\frac{\log(N/\epsilon)}{\ell_1}}\right)$.

Proof: Assume in contradiction that $\|\vec{e}^{\mathrm{H}}(t+1)\| \leq (1-\delta^2)\cdot\|\vec{e}^{\mathrm{H}}(t)\|$ for all $t = \ell_1/2,\ldots,\ell_1-1$. Since $\|\vec{e}^{\mathrm{H}}(\ell_1/2)\| \leq 1$, we would get that

$$\|\vec{e}^{\mathrm{H}}(\ell_1)\| \leq (1-\delta^2)^{\frac{\ell_1}{2}}\cdot\|\vec{e}^{\mathrm{H}}(\ell_1/2)\| \;\leq\; \exp(-\log(128N/\epsilon))\cdot 1 \;<\; \frac{\epsilon}{128N} \;\leq\; \frac{\epsilon}{128|\mathrm{H}|}\,.$$

□

Let $\bar{t}$ be as determined by Claim 1. We next obtain a lower bound on $\|\vec{e}^{\mathrm{H}}(\bar{t})\| - \|\vec{e}^{\mathrm{H}}(\bar{t}+1)\|$. (This bound actually holds for every $t < \ell_1$ but we will use it only for $t = \bar{t}$.)

Claim 2:

$$\|\vec{e}^{\mathrm{H}}(\bar{t})\| - \|\vec{e}^{\mathrm{H}}(\bar{t}+1)\| \;\geq\; \sum_{v,u\in\mathrm{H}}\frac{1}{2}q_{v,u}\cdot(e_v(\bar{t})-e_u(\bar{t}))^2 + \sum_{v\in\mathrm{H}}\frac{1}{2}q_{v,a_{v,1}}\cdot\left((3e_v(\bar{t}))^2 + 2e_v\pi_v - (\pi_v)^2\right)$$

Let us ignore momentarily the second term in the inequality of Claim 2 (which is due to the auxiliary paths of M and is bounded in the proof of the next claim). Then we see that the contribution to the difference between $\vec{e}^{\mathrm{H}}(\bar{t})$ and $\vec{e}^{\mathrm{H}}(\bar{t}+1)$, is mainly due to significant differences between $e_v(\bar{t})$ and $e_u(\bar{t})$ (equivalently, differences between $q_{s,v}(\bar{t})$ and $q_{s,u}(\bar{t})$) for vertices v and u in H that have an edge between them. We later relate this term more precisely to cuts in H.

Proof of Claim 2: For simplicity, in what follows, we shall think of there being exactly $|\mathrm{H}|\cdot\ell_1$ auxiliary vertices, that is $a_{v,1},\ldots,a_{v,\ell_1}$ for each vertex v in H, where for $v \notin \mathrm{B}(\mathrm{H})$, $q_{v,a_{v,1}} = 0$. For technical convenience, for every $v \in \mathrm{H}$, we define $\bar{q}_{v,v} \stackrel{\text{def}}{=} q_{v,v} - \frac{1}{2}$, (which by definition of $q_{v,v}$ is always non-negative.) For very pair of different states z, y, $\bar{q}_{z,y} \stackrel{\text{def}}{=} q_{z,y}$. Note that for every vertex v, the sum over all states z (including v itself) of $\bar{q}_{v,z}$ is $\frac{1}{2}$. In the equation below we perform an algebraic manipulation on $\|\vec{e}^{\mathrm{H}}(\bar{t})\|$ that brings it to a convenient form

$$\begin{aligned}\|\vec{e}^{\mathrm{H}}(\bar{t})\| \;=\; \sum_{v\in\mathrm{H}}e_v(\bar{t})^2 \;&=\; \sum_{v\in\mathrm{H}}\left(\sum_{u\in\mathrm{H}}\left(\bar{q}_{v,u}\cdot e_v(\bar{t})^2 + \bar{q}_{u,v}\cdot e_v(\bar{t})^2\right) \;+\; 2\bar{q}_{v,a_{v,1}}e_v(\bar{t})^2\right)\\ &=\; \sum_{v,u\in\mathrm{H}}\bar{q}_{v,u}\cdot\left(e_v(\bar{t})^2 + e_u(\bar{t})^2\right) + \sum_{v\in\mathrm{H}}2\bar{q}_{v,a_{v,1}}\cdot e_v(\bar{t})^2 \qquad (1)\end{aligned}$$

Next we bound $||\vec{e}^{\mathrm{H}}(\bar{t}+1)||$. Note that since $\bar{t} < \ell_1$, for each of the auxiliary states a_{v,ℓ_1} (i.e. on the end of the auxiliary path of length ℓ_1 from v), the probability of reaching a_{v,ℓ_1} in $\bar{t}$ steps from s is 0, and hence $e_{a_{v,\ell_1}}(\bar{t}) = -\pi_{a_{v,\ell_1}}$. As we have noted before, the stationary distribution of all auxiliary vertices on the auxiliary path emitting from v is the same, and since the only transition entering the first state on the path is from v, $\pi_{a_{v,\ell_1}} = \pi_{a_{v,1}} = \pi_v \cdot q_{v,a_{v,1}}$. By definition of M, this implies that for every $u \in \mathrm{H}$,

$$
\begin{aligned}
\sum_{v\in\mathrm{H}} \bar{q}_{a_{v,\ell_1},u} \cdot e_{a_{v,\ell_1}}(\bar{t}) &= \sum_{v\in\mathrm{H}} \left(\frac{1}{q_{v,a_{v,1}}} \cdot \sum_{t\geq\ell_2} p^{\mathrm{H}}_{v,u}(t)\right) \cdot (-\pi_v \cdot q_{v,a_{v,1}}) \\
&= -\sum_{v\in\mathrm{H}} \pi_v \cdot \sum_{t\geq\ell_2} p^{\mathrm{H}}_{v,u}(t) \\
&= -\pi_u \cdot \sum_{v\in\mathrm{H}} \sum_{t\geq\ell_2} p^{\mathrm{H}}_{u,v}(t) \\
&= -\pi_u \cdot q_{u,a_{u,1}} \\
&= -\pi_u \cdot \bar{q}_{u,a_{u,1}} \qquad (2)
\end{aligned}
$$

Recall that $\vec{e}(\bar{t}+1) = \vec{e}(\bar{t}) \cdot \mathrm{M}$, and that for every $v \in \mathrm{H}$, $\sum_{u\in\mathrm{H}} \bar{q}_{v,u} + \bar{q}_{v,a_{v,1}} = \frac{1}{2}$. Below we use Equation (2) (in the second equality) and the fact that the square of the mean is upper bounded by the mean of the squares (in the third inequality).

$$
\begin{aligned}
||\vec{e}^{\mathrm{H}}(\bar{t}+1)|| \stackrel{\mathrm{def}}{=} \sum_{u\in\mathrm{H}} (e_u(\bar{t}+1))^2 &= \sum_{u\in\mathrm{H}} \left(\frac{1}{2} e_u(\bar{t}) + \sum_{v\in\mathrm{H}} \bar{q}_{v,u} \cdot e_v(\bar{t}) + \sum_{v\in\mathrm{H}} \bar{q}_{a_{v,\ell_1},u} \cdot e_{a_{v,\ell_1}}(\bar{t})\right)^2 \\
&= \sum_{u\in\mathrm{H}} \left(\sum_{v\in\mathrm{H}} 2\bar{q}_{v,u} \cdot \left(\frac{e_u(\bar{t}) + e_v(\bar{t})}{2}\right) + 2\bar{q}_{u,a_{u,1}} \cdot \left(\frac{e_u(\bar{t}) - \pi_u}{2}\right)\right)^2 \\
&\leq \sum_{u\in\mathrm{H}} \left(\sum_{v\in\mathrm{H}} 2\bar{q}_{v,u} \cdot \left(\frac{e_u(\bar{t}) + e_v(\bar{t})}{2}\right)^2 + 2\bar{q}_{u,a_{u,1}} \cdot \left(\frac{e_u(\bar{t}) - \pi_u}{2}\right)^2\right) \\
&= \sum_{v,u\in\mathrm{H}} \frac{1}{2}\bar{q}_{v,u} \cdot (e_v(\bar{t}) + e_u(\bar{t}))^2 + \sum_{v\in\mathrm{H}} \frac{1}{2}\bar{q}_{v,a_{v,1}} \cdot (e_v(\bar{t}) - \pi_v)^2 \qquad (3)
\end{aligned}
$$

By Equations (1) and (3) we have:

$$
\begin{aligned}
||\vec{e}^{\mathrm{H}}(\bar{t})|| \quad - &\quad ||\vec{e}^{\mathrm{H}}(\bar{t}+1)|| \\
\geq &\quad \sum_{v,u\in\mathrm{H}} \frac{1}{2}\bar{q}_{v,u} \cdot (e_v(\bar{t}) - e_u(\bar{t}))^2 + \sum_{v\in\mathrm{H}} \frac{1}{2}\bar{q}_{v,a_{v,1}} \cdot ((3e_v(\bar{t}))^2 + 2e_v\pi_v - (\pi_v)^2) \\
= &\quad \sum_{v,u\in\mathrm{H}} \frac{1}{2}q_{v,u} \cdot (e_v(\bar{t}) - e_u(\bar{t}))^2 + \sum_{v\in\mathrm{H}} \frac{1}{2}q_{v,a_{v,1}} \cdot ((3e_v(\bar{t}))^2 + 2e_v\pi_v - (\pi_v)^2)
\end{aligned}
$$

□

Based on Claims 1 and 2 we prove the following claim. As we noted before, the expression in the left hand side of the inequality stated in Claim 3 will later be related to cuts in H.

Claim 3:

$$
\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v(\bar{t}) - e_u(\bar{t}))^2 \quad \leq \quad 3\delta^2 \cdot ||\vec{e}^{\mathrm{H}}(\bar{t})||
$$

Proof: From Claims 2 and 1 we have that

$$\begin{aligned}\sum_{v,u\in \mathrm{H}} q_{v,u}\cdot(e_v(\bar t)-e_u(\bar t))^2 &\le 2\left(\|\vec e^{\mathrm{H}}(\bar t)\|-\|\vec e^{\mathrm{H}}(\bar t+1)\|\right)-\sum_{v\in\mathrm{H}} q_{v,a_{v,1}}\cdot((3e_v(\bar t))^2+2e_v\pi_v-(\pi_v)^2)\\ &\le 2\delta^2\cdot\|\vec e^{\mathrm{H}}(\bar t)\|-\sum_{v\in\mathrm{H}} q_{v,a_{v,1}}\cdot((3e_v(\bar t))^2+2e_v\pi_v-(\pi_v)^2) \qquad (4)\end{aligned}$$

Let us denote the second term in Equation (4) by X. We next show that $X\ge-\delta^2\cdot\|\vec e^{\mathrm{H}}(\bar t)\|$. The quadratic expression $3e_v(t)^2+2e_v\pi_v-(\pi_v)^2$ has a minimum value of $-4(\pi_v)^2/3$ (obtained at $e_v(t)=-\pi_v/3$). Since $\pi_v\le\frac{1}{|\mathrm{H}|}$, this value is at least $-\frac{4}{3|\mathrm{H}|^2}$. Furthermore, by the definition of $q_{v,a_{v,1}}$ and Lemma 4.2,

$$\sum_{v\in\mathrm{H}} q_{v,a_{v,1}} = \sum_{v,u\in\mathrm{H}}\sum_{t\ge\ell_2} p^{\mathrm{H}}_{v,u}(t) \le \frac{N}{\ell_2}$$

and hence using the lemma's hypothesis concerning the size of H,

$$X \ge \left(-\frac{4}{3|\mathrm{H}|^2}\right)\cdot\frac{N}{\ell_2} \ge -\frac{4N}{3\ell_2}\cdot\frac{4}{\epsilon\cdot N}\cdot\frac{1}{|\mathrm{H}|} = -\frac{8}{3\epsilon\ell_2|\mathrm{H}|}$$

By the case hypothesis, $\|\vec e^{\mathrm{H}}(t)\|>\frac{\epsilon}{128|\mathrm{H}|}$ and so

$$X \ge -O\left(\frac{1}{\epsilon^2\ell_2}\right)\cdot\|\vec e^{\mathrm{H}}(t)\|$$

By the lemma's hypotheses, (and the definition of $\delta=O\left(\sqrt{\frac{\log(N/\epsilon)}{\ell_1}}\right)$), we have that $\ell_1=\Omega\left(\frac{1}{\delta^2}\right)$, and $\ell_2=\Omega\left(\frac{\ell_1}{\epsilon^2}\right)=\Omega\left(\frac{1}{\epsilon^2\cdot\delta^2}\right)$. Therefore, for the appropriate constants in the $O(\cdot)$ and $\Omega(\cdot)$ notation, we have that $X\ge-\delta^2\cdot\|\vec e^{\mathrm{H}}(t)\|$, as required, and the claim follows. □

From this point on, let $e_v\stackrel{\mathrm{def}}{=}e_v(\bar t)$, and define $e_v^+\stackrel{\mathrm{def}}{=}\max(e_v,0)$ and $e_v^-\stackrel{\mathrm{def}}{=}\min(e_v,0)$. Thus, $\vec e=\vec e^{+}+\vec e^{-}$. It will be convenient to deal only with $\vec e^{+}$ (that is, with vertices v such that $e_v>0$, which means that $q_{s,v}(\bar t)\ge\pi_v$). We hence relate $\sum_{v\in\mathrm{H}}(e_v^+)^2$ to $\|\vec e^{\mathrm{H}}\|$.

Claim 4: $\|\vec e^{\mathrm{H}}(\bar t)\|\le\frac{2^{10}}{\epsilon}\cdot\sum_{v\in\mathrm{H}}(e_v^+)^2$.

To prove Claim 4, we shall need the following technical claim whose proof is given in Appendix B.

Claim 5: *Let $x_1,\ldots,x_m$, $-\frac{1}{m}\le x_i\le 1$ be real numbers for which the following holds for some $0<\gamma\le\frac{1}{2}$.*

1. $\sum_{i=1}^m x_i\ge-\gamma$;
2. $\sum_{i=1}^m x_i^2\ge\frac{2\gamma}{m}$.

Then, $\sum_{i,\,x_i>0}x_i^2\ge\frac{\gamma}{4}\cdot\sum_i x_i^2$.

Proof of Claim 4: By the lemma's hypothesis, s is *useful*, and as we have previously shown, this implies that the total probability of being in any auxiliary state at any step $t\le\ell_1$, is at most $\frac{\epsilon}{256}$. Since $\sum_{z\in\mathrm{M}}e_z=0$, and for every state z (and in particular every auxiliary state), $e_z<q_{s,z}(\bar t)$, we get that

$$\sum_{v\in\mathrm{H}}e_v = \sum_{z\in\mathrm{M}}e_z-\sum_{\text{aux. }a}e_a \ge 0-\sum_{\text{aux. }a}q_{s,a}(\bar t) \ge -\frac{\epsilon}{256}$$

Finally, by the case hypothesis, $||\bar{e}^{\mathrm{H}}(\bar{t})|| \geq \frac{\epsilon}{128|\mathrm{H}|}$, and so Claim 4 follows by applying Claim 5 with $\gamma = \frac{\epsilon}{256}$ and $m = |H|$. $\Box$

Claim 6:

$$\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot |(e_v^+)^2 - (e_u^+)^2| \;\; = \;\; O\left(\frac{\delta}{\sqrt{\epsilon}} \sum_{v\in\mathrm{H}} (e_v^+)^2\right) \; .$$

Proof: We first observe that

$$\begin{aligned} \sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v - e_u)^2 \;\; &\geq \;\; \sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ - e_u^+)^2 + \sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^- - e_u^-)^2 \\ &\geq \;\; \sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ - e_u^+)^2 \end{aligned} \tag{5}$$

Combining Equation (5), Claim 3, and Claim 4, we have:

$$\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ - e_u^+)^2 \;\; \leq \;\; 3\delta^2 \cdot \frac{2^{10}}{\epsilon} \cdot \sum_{v\in\mathrm{H}} (e_v^+)^2 \tag{6}$$

On the other hand, using the Cauchy-Schwartz inequality,

$$\begin{aligned} \sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ - e_u^+)^2 \;\; &= \;\; \frac{\left(\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ - e_u^+)^2\right) \cdot \left(\sum_{v\neq u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ + e_u^+)^2\right)}{\sum_{v\neq u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ + e_u^+)^2} \\ &\geq \;\; \frac{\left(\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot |(e_v^+)^2 - (e_u^+)^2|\right)^2}{\sum_{v\neq u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ + e_u^+)^2} \end{aligned} \tag{7}$$

In order to bound the denominator, we perform a similar manipulation to that in Equation (1) and then use the fact that the mean of squares is lower bounded by the square of means (so that $(e_v^+)^2 + (e_u^+)^2 \geq \frac{1}{2}(e_v^+ + e_u^+)^2$). Recall that $\bar{q}_{v,v} = q_{v,v} - \frac{1}{2}$, and for $u \neq v$, $\bar{q}_{v,u} = q_{v,u}$.

$$\begin{aligned} \sum_{v\in\mathrm{H}} (e_v^+)^2 \;\; &= \;\; \sum_{v,u\in\mathrm{H}} \bar{q}_{v,u} \cdot ((e_v^+)^2 + (e_u^+)^2) + \sum_{v\in\mathrm{H}} 2\bar{q}_{v,a_{v,1}} \cdot (e_v^+)^2 \\ &\geq \;\; \frac{1}{2} \sum_{v\neq u\in\mathrm{H}} q_{v,u} \cdot \left(e_v^+ + e_u^+\right)^2 \end{aligned} \tag{8}$$

By combining Equation (7) and (8),

$$\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot (e_v^+ - e_u^+)^2 \;\; \geq \;\; \frac{\left(\sum_{v,u\in\mathrm{H}} q_{v,u} \cdot |(e_v^+)^2 - (e_u^+)^2|\right)^2}{2\sum_{v\in\mathrm{H}} (e_v^+)^2} \tag{9}$$

The Claim follows from Equations (9) and (6). $\Box$

Assume we rename the states in H from '1' to '$|\mathrm{H}|$' so that $e_k^+ \geq e_{k+1}^+$. Let $\mathrm{S}_k \stackrel{\mathrm{def}}{=} \{1, \ldots, k\}$, and let $C(\mathrm{S}_k) \stackrel{\mathrm{def}}{=} \sum_{v\in\mathrm{S}_k, u\notin\mathrm{S}_k} q_{v,u}$ be the probability weight of the corresponding cut. Since for every $v \in \mathrm{H}$, $\sum_{u\in\Gamma(v)} q_{v,u} \leq \frac{1}{2}$, the number of edges between S_k and the rest of H is at most $2d \cdot C(\mathrm{S}_k)$.

Claim 7:

$$\sum_{v,u\in \mathrm{H}} q_{v,u}\cdot |(e_v^+)^2-(e_u^+)^2| \;=\; 2\sum_{k=1}^{|\mathrm{H}|-1}\left((e_k^+)^2-(e_{k+1}^+)^2\right)\cdot C(\mathrm{S}_k)$$

Proof: For brevity, we refer to the vertices according to their new renaming in $\{1,\ldots,|\mathrm{H}|\}$ (e.g., i and j instead of v and u). Using the fact that $q_{i,j}=q_{j,i}$, and that the vertices are ordered according to the value of e_k^+ (and in particular, $e_{|\mathrm{H}|}^+=0$),

$$\begin{aligned}
\sum_{i,j\in \mathrm{H}} q_{i,j}\cdot |(e_i^+)^2-(e_j^+)^2| &= 2\sum_{i,j\in \mathrm{H}, i<j} q_{i,j}\cdot \left((e_i^+)^2-(e_j^+)^2\right)\\
&= 2\sum_{i,j\in \mathrm{H}, i<j} q_{i,j}\cdot \sum_{k=i}^{j-1}\left((e_k^+)^2-(e_{k+1}^+)^2\right)\\
&= 2\sum_{k=1}^{|\mathrm{H}|-1}\left((e_k^+)^2-(e_{k+1}^+)^2)\right)\cdot \sum_{i\le k, j>k} q_{i,j}\\
&= 2\sum_{k=1}^{|\mathrm{H}|-1}\left((e_k^+)^2-(e_{k+1}^+)^2)\right)\cdot C(\mathrm{S}_k)
\end{aligned}$$

□

Claim 8: *There exists k, $1\le k\le |\mathrm{H}|-1$ such that*

1. $C(\mathrm{S}_k) < \frac{\epsilon}{8}|\mathrm{S}_k|$;
2. *For all but at most $\frac{\epsilon}{4}|\mathrm{S}_k|$ of the vertices $i\in \mathrm{S}_k$, $\frac{\beta}{|\mathrm{S}_k|\cdot|\mathrm{H}|} \le (e_i^+)^2 \le F\cdot \frac{\beta}{|\mathrm{S}_k|\cdot|\mathrm{H}|}$, for $\beta=\Omega(\epsilon^2/\log(N/\epsilon))$ and $F=O(1/\epsilon)$.*

Proof: In order to prove the claim, we partition $e_1^+,\ldots,e_{|\mathrm{H}|}^+$ into maximal consecutive *intervals* so that the ratio between the square of the largest e_k^+ in each interval and the square of the smallest e_k^+ in the interval is at most 2. Let $b=O(\log(N/\epsilon))$. Since by the case hypothesis and Claim 4, $\sum_{v\in\mathrm{H}}(e_i^+)^2=\Omega(\epsilon^2/|\mathrm{H}|)$ there must be an interval I, such that $\sum_{i\in I}(e_i^+)^2 \ge \frac{1}{b}\sum_{v\in \mathrm{H}}(e_i^+)^2$. Let $I=\{f,\ldots,\ell\}$ be the first such interval, and let $h\ge\ell$ be the largest index such that $(e_h^+)^2\ge \frac{1}{2}(e_\ell^+)^2$ (thus, $(e_{h+1}^+)^2 < \frac{1}{2}(e_\ell^+)^2$). We claim that for some $\ell\le k\le h$, $C(\mathrm{S}_k)<\frac{\epsilon}{8}\cdot|\mathrm{S}_k|$. Assume, contrary to the claim that all these cuts are large. Then, using the fact that the e_k^+'s are ordered and our choice of h,

$$\begin{aligned}
\sum_{k=1}^{|\mathrm{H}|-1}\left((e_k^+)^2-(e_{k+1}^+)^2\right)C(\mathrm{S}_k) &\ge \sum_{k=\ell}^{h}\left((e_k^+)^2-(e_{k+1}^+)^2\right)C(\mathrm{S}_k)\\
&\ge \frac{\epsilon}{8}\cdot\sum_{k=\ell}^{h} k\cdot\left((e_k^+)^2-(e_{k+1}^+)^2\right)\\
&= \frac{\epsilon}{8}\cdot\left(\ell\cdot(e_\ell^+)^2-h\cdot(e_{h+1}^+)^2+\sum_{k=\ell+1}^{h}(e_k^+)^2\right)\\
&\ge \frac{\epsilon}{8}\cdot\left(\ell\cdot(e_\ell^+)^2-(\ell+(h-\ell))\cdot(e_{h+1}^+)^2+(h-\ell)\cdot(e_h^+)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\epsilon}{8} \cdot \left(\ell \cdot ((e_\ell^+)^2 - (e_{h+1}^+)^2) + (h-\ell) \cdot ((e_h^+)^2 - (e_{h+1}^+)^2)\right) \\
&\geq \frac{\epsilon}{16} \cdot \ell \cdot (e_\ell^+)^2
\end{aligned} \tag{10}
$$

By our choice of I (i.e., $(e_f^+)^2 \leq 2 \cdot (e_\ell^+)^2$, and $\sum_{i \in I} (e_i^+)^2 \geq \frac{1}{b} \sum_{v \in \mathrm{H}} (e_i^+)^2$),

$$
\frac{\epsilon}{16} \cdot \ell \cdot (e_\ell^+)^2 \;\; \geq \;\; \frac{\epsilon}{32} \cdot \ell \cdot (e_f^+)^2 \;\; \geq \;\; \frac{\epsilon}{32} \sum_{i \in I} (e_i^+)^2 \;\; = \;\; \Omega\left(\frac{\epsilon}{\log(N/\epsilon)} \cdot \sum_{v \in \mathrm{H}} (e_v^+)^2\right) \tag{11}
$$

By combing Claim 7 together with Equations (10) and (11) we get

$$
\sum_{v,u \in \mathrm{H}} q_{v,u} \cdot |(e_v^+)^2 - (e_u^+)^2| \;=\; \Omega\left(\frac{\epsilon}{\log(N/\epsilon)} \cdot \sum_{v \in \mathrm{H}} (e_v^+)^2\right)
$$

Since $\delta = O\left(\sqrt{\frac{\log(N/\epsilon)}{\ell_1}}\right)$,

$$
\frac{\epsilon}{\log(N/\epsilon)} \;=\; \Omega\left(\frac{\delta}{\sqrt{\epsilon}} \cdot \sqrt{\ell_1} \cdot \left(\frac{\epsilon}{\log(N/\epsilon)}\right)^{\frac{3}{2}}\right)
$$

And so, by our choice of $\ell_1 = \Theta\left(\frac{\epsilon}{\log(N/\epsilon)}\right)^3$, we get a contradiction to Claim 6 (for the appropriate choice of constants in the $\Theta(\cdot)$ notation). Therefore, for some k, $\ell \leq k \leq h$, $C(\mathrm{S}_k) < \frac{\epsilon}{8} \cdot |\mathrm{S}_k|$. Let us fix this k. By definition of I and h, and using Claim 4 and the case hypothesis, for every $j \in \mathrm{S}_k$,

$$
(e_j^+)^2 \;\; \geq \;\; (e_k)^2 \;\; \geq \;\; \frac{1}{4}(e_f^+)^2 \;\; \geq \;\; \frac{1}{4(\ell - f + 1)} \sum_{i \in I} (e_i^+)^2 \;\; \geq \;\; \frac{1}{k} \cdot \frac{1}{4b} \sum_{v \in \mathrm{H}} (e_v^+)^2 \;\; \geq \;\; \frac{1}{|\mathrm{S}_k|} \cdot \frac{\beta}{N}
$$

for $\beta = \Omega\left(\frac{\epsilon^2}{\log(N/\epsilon)}\right)$. It remains to bound the number of vertices j in S_k for which $(e_j^+)^2$ is much larger than $(e_k^+)^2$. Let $I_1, \ldots, I_r$ be the intervals up to I (i.e., $I_r = I$). For each g, $1 \leq g \leq r$, let f_g and ℓ_g be the first and last elements, respectively, in I_g. Then, by the definition of the intervals, for every g, $(e_{f_g}^+)^2 > 2(e_{f_{g+1}}^+)^2$. Recall that the interval $I = I_r$ was the first interval for which $\sum_{i \in I} (e_i^+)^2 \geq \frac{1}{b} \sum_{v \in \mathrm{H}} (e_v^+)^2$. This implies that for each I_g,

$$
|I_g| \cdot (e_{\ell_g}^+)^2 \;\leq\; |I_r| \cdot (e_{f_r}^+)^2
$$

Let a be the largest index such that $(e_{\ell_a}^+)^2 \geq \frac{32}{\epsilon}(e_k^+)^2$. Then by the above, $\sum_{g \leq a} |I_g| = O(\epsilon \cdot |\mathrm{S}_k|)$, and Claim 8 follows. □

We thus define S to be the subset of vertices v in S_k, for the k implied by Claim 8, for which $(e_v^+)^2$ is within the bounds stated in the claim. The size of the cut between S and the rest of H is hence as desired, and since $q_{s,v}(\bar{t}) > e_v^+$, the lemma follows. ■

4.4 Sufficient Conditions for Good Partitions

In the next lemma we give sufficient conditions under which subsets of vertices can be partitioned without having many violating edges. What the lemma essentially requires is that for some fixed vertex s and subset of vertices S in H, there is a lower bound on the probability that each vertex in S is reached from s (in t steps), and there aren't too many vertices v in the subset such that both $q_{s,v}^0(t)$ and $q_{s,v}^1(t)$ are large (with respect to this lower bound).

Lemma 4.4 *Let* H *be a subgraph of* G*,* s *a vertex in* H*,* S *a subset of vertices in* H *and* ℓ_1 *and* ℓ_2 *integers. Assume that for some* $\alpha > 0$ *and* $t = \Omega(\log(\frac{1}{\alpha}))$*,* $t < \ell_1$*, the following holds in* $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$*:*

1. *For every* $v \in \mathrm{S}$*,* $q_{s,v}(t) \geq \alpha$*;*
2. $\sum_{v \in \mathrm{S}} q^0_{s,v}(t) \cdot q^1_{s,v}(t) < \frac{\epsilon}{c} \cdot |\mathrm{S}| \cdot \alpha^2$ *for some constant c.*

Let $(\mathrm{S}_0, \mathrm{S}_1)$ *be a partition of* S*, where* $\mathrm{S}_0 = \{v : q^0_{s,v}(t) \geq q^1_{s,v}(t)\}$*, and* $\mathrm{S}_1 = \{v : q^1_{s,v}(t) > q^0_{s,v}(t)\}$*. Then the number of violating edges in* G *with respect to* $(\mathrm{S}_0, \mathrm{S}_1)$ *is at most* $2^5 \cdot \frac{\epsilon}{c} \cdot d \cdot |\mathrm{S}|$*.*

Proof: Let $\mathrm{M} \stackrel{\mathrm{def}}{=} \mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. Consider a vertex v and let $v \in \mathrm{S}_\sigma$, for $\sigma \in \{0,1\}$. By definition of the partition $(\mathrm{S}_0, \mathrm{S}_1)$, $q^\sigma_{s,v}(t) \geq \frac{1}{2} q_{s,v}(t) \geq \frac{\alpha}{2}$. By definition of M we have that

$$q^{\bar\sigma}_{s,v}(t) \;\geq\; \sum_{u \in \Gamma(v)} q^\sigma_{s,u}(t-1) \cdot q_{u,v} \;\geq\; \frac{1}{2d} \cdot \sum_{u \in \Gamma(v) \cap \mathrm{S}_\sigma} q^\sigma_{s,u}(t-1) \tag{12}$$

While we know that for every $u \in \mathrm{S}_\sigma$, $q^\sigma_{s,u}(t) \geq \frac{1}{2} q_{s,u}(t) \geq \frac{\alpha}{2}$, we need a lower bound on $q^\sigma_{s,u}(t-1)$.

Claim: *Let* $u \in \mathrm{S}_\sigma$*. If* $t = \Omega(\log(1/\alpha))$*, then* $q^\sigma_{s,u}(t-1) \geq \frac{1}{8} q^\sigma_{s,u}(t)$*.*

We prove the claim momentarily, and first show how the lemma follows from the claim and Equation (12). By combining Equation (12) with the claim, we have that for every vertex v such that $v \in \mathrm{S}_\sigma$,

$$q^{\bar\sigma}_{s,v}(t) \geq \frac{1}{16d} \sum_{u \in \Gamma(v) \cap \mathrm{S}_\sigma} q^\sigma_{s,u}(t)$$

And hence,

$$\begin{aligned} q^0_{s,v}(t) \cdot q^1_{s,v}(t) &\geq q^\sigma_{s,v}(t) \cdot \frac{1}{16d} \sum_{u \in \Gamma(v) \cap \mathrm{S}_\sigma} q^\sigma_{s,u}(t) \\ &\geq \frac{q_{s,v}(t)}{2} \cdot \frac{1}{16d} \sum_{u \in \Gamma(v) \cap \mathrm{S}_\sigma} \frac{q_{s,u}(t)}{2} \end{aligned}$$

Assume, contrary to what is claimed in the lemma that the number of violating edges with respect to $(\mathrm{S}_0, \mathrm{S}_1)$ is more than $2^5 \cdot \frac{\epsilon}{c} \cdot d \cdot |\mathrm{S}|$. Then

$$\begin{aligned} \sum_{v \in \mathrm{S}} q^0_{s,v}(t) \cdot q^1_{s,v}(t) &\geq \sum_{\sigma \in \{0,1\}} \sum_{x,y \in \mathrm{S}_\sigma, (x,y) \in E} 2 \cdot \frac{1}{16d} \cdot \frac{q_{s,x}(t)}{2} \cdot \frac{q_{s,y}(t)}{2} \\ &> 2^5 \cdot \frac{\epsilon}{c} \cdot d \cdot |\mathrm{S}| \cdot \frac{1}{8d} \cdot \frac{\alpha^2}{4} \\ &= \frac{\epsilon}{c} \cdot |\mathrm{S}| \cdot \alpha^2 \end{aligned}$$

where the factor of 2 in the first inequality comes from the contribution of the edge (x,y) both to $q^0_{s,x}(t) \cdot q^1_{s,x}(t)$ and to $q^0_{s,y}(t) \cdot q^1_{s,y}(t)$. But this contradicts the second hypothesis of the lemma.

Proof of Claim: Without loss of generality let $\sigma = 0$. Consider random walks of length t in M that do not enter an auxiliary path (or else they cannot reach u as $t < \ell_1$). In what follows we map walks of length t that end at u and correspond to even length paths, to walks of length $t-1$ that end at u (and have the same parity). We do this by removing a single step in which the walk remained at the current vertex. Intuitively, since the probability of remaining at the current vertex is at least $\frac{1}{2}$, the total probability of the resulting walks (of length $t-1$) is roughly the same as that of the original walks (of length t). In what follows we formalize this.

We associate with each walk a sequence of transition-labels: Transitions that correspond to edges between vertices are given the edge-label, and the self-transition from a vertex v to itself is replaced by $2d - |\Gamma(v)|$ transitions (labeled $|\Gamma(v)|+1,\ldots,2d$), each having probability $\frac{1}{2d}$. Thus each walk of length t (that does not enter an auxiliary path) is uniquely labeled and has exactly the same probability, $\left(\frac{1}{2d}\right)^t$.

Let $v_0 = s$, and $v_1,\ldots,v_t$ be the vertices passed on a random walk of length t. Consider those steps i in which the walk remains at the current vertex. That is, i such that $v_i = v_{i-1}$. Since (conditioned on the event that the walk does not enter an auxiliary path), the probability at each step i that $v_i = v_{i-1}$ is at least $\frac{1}{2}$, the expected number of such steps is at least $\frac{t}{2}$. By a multiplicative Chernoff bound we have that the probability that $|\{i : v_i = v_{i-1}\}| < \frac{t}{4}$, is at most $\exp(-t/12) < \alpha/4$.

We now focus only on those walks that end at u and correspond to even-length paths. Let the set of these walks be denoted U. Recall that since $u \in \mathrm{S}_0$, we have that $q^0_{s,u}(t) \geq \frac{\alpha}{2}$. Let T be the subset of walks in U for which $|\{i : v_i = v_{i-1}\}| \geq \frac{t}{4}$. By what we have shown above, $|\mathrm{T}| \geq |\mathrm{U}|/2$. Let T' is the set of walks of length $t-1$ that end at u and can be obtained from some walk in T by removing a single step i such that $v_i = v_{i-1}$. Consider an auxiliary bipartite graph over $\mathrm{T} \cup \mathrm{T}'$ that has the following edges. There is an edge between a node in T and a node in T' if an only if the latter can be obtained from the former by removing a single step i such that $v_i = v_{i-1}$. We allow for multiple edges in case there is more than one way to perform this transformation (that is, if the walk remained at a particular vertex for more than one step, and furthermore, took the same self-transition in all the corresponding steps). By definition of T, each node in T is incident to at least $\frac{1}{4}t$ edges, while each node in T' is incident to at most $t \cdot (2d-1)$ edges. (The factor of $(2d-1)$ is the result of the multiple self-transitions). Therefore, $|\mathrm{T}| \cdot \frac{t}{4} \leq |\mathrm{T}'| \cdot (2d-1) \cdot t$, and so $|\mathrm{T}'| > \frac{1}{8d} \cdot |\mathrm{T}| \geq \frac{1}{16d} \cdot |\mathrm{U}|$. Since each walk in T' has probability $(2d)^{-(t-1)}$ while each walk in U has probability $(2d)^{-t}$, the claim, and subsequently the lemma, follow. ■

4.5 Sufficient Conditions for Detecting Odd Cycles

In the next lemma we describe sufficient conditions for "detecting" odd cycles when performing walks in $\mathrm{M}^{\ell_2}_{\ell_1}(\mathrm{H})$ starting from some vertex s. What the lemma essentially requires is that there exist a subset S of vertices such that there are both lower and upper bounds on the probability that each vertex in S is reached from s (in $t < \ell_1$ steps), and there are many vertices v in S such that both $q^0_{s,v}(t)$ and $q^1_{s,v}(t)$ are large (with respect to the lower bound). As stated later in Corollary 4, these conditions are sufficient for detecting odd cycles when performing random walks in G of length $\ell_1 \cdot \ell_2$.

Lemma 4.5 *Let* H *be a subgraph of* G, s *a vertex in* H, S *a subset of vertices in* H *and* ℓ_1 *and* ℓ_2 *integers. Assume that for some* $\alpha, F > 0$ *and* $t < \ell_1$, *the following holds in* $\mathrm{M}^{\ell_2}_{\ell_1}(\mathrm{H})$*:*

1. *For every* $v \in \mathrm{S}$, $\alpha \leq q_{s,v}(t) \leq F \cdot \alpha$;
2. $\sum_{v \in \mathrm{S}} q^0_{s,v}(t) \cdot q^1_{s,v}(t) \geq \frac{\epsilon}{c} \cdot |\mathrm{S}| \cdot \alpha^2$ *for some constant c.*

Then with probability at least 0.99, *if we perform* $O\left(\frac{F}{\epsilon \cdot \alpha \sqrt{|\mathrm{S}|}}\right)$ *random walks of length* t *starting from* s *in* $\mathrm{M}^{\ell_2}_{\ell_1}(\mathrm{H})$ *then for some vertex* v *we shall end at* v *both on a walk corresponding to an even-length path and on a walk corresponding to an odd-length path.*

We note that when we apply Lemma 4.5, we set $\alpha = \mathrm{poly}(\epsilon/(\log N))/\sqrt{|\mathrm{S}|\cdot|\mathrm{H}|}$, and $F = O(1/\epsilon)$, so that the number of random walks that should be performed is $\mathrm{poly}((\log N)/\epsilon)\sqrt{N}$.

Proof: Let $\mathrm{M} \stackrel{\rm def}{=} \mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, and $\gamma \stackrel{\rm def}{=} \sum_{v\in\mathrm{S}} q_{s,v}^0(t)\cdot q_{s,v}^1(t)$, so that by the second hypothesis of the lemma $\gamma \geq \frac{\epsilon}{c}\cdot|\mathrm{S}|\cdot\alpha^2$. Consider $m = O\left(\frac{F}{\epsilon\cdot\alpha\cdot\sqrt{|\mathrm{S}|}}\right)$ random walks of length t starting from s. For $1 \leq i,j \leq m$, let $\eta_{i,j}$ be a 0/1 random variable that is 1 if and only if the $i^{\rm th}$ and $j^{\rm th}$ walks correspond to paths whose lengths have different parity, but both end at the same vertex in S. Thus, we would like to bound the probability that $\sum_{i<j}\eta_{i,j} = 0$. The difficulty is that the $\eta_{i,j}$'s are not pairwise independent. Yet, since the sum of the covariances of the dependent $\eta_{i,j}$'s is quite small, Chebyshev's Inequality is still very useful (cf., [AS92a, Sec. 4.3]). Details follow. For every $i \neq j$,

$$\mathrm{Exp}[\eta_{i,j}] \;=\; \sum_{\sigma\in\{0,1\}}\sum_{v\in\mathrm{S}} q_{s,v}^{\sigma}(t)\cdot q_{s,v}^{\bar\sigma}(t) \;=\; 2\gamma$$

By Chebyshev's inequality,

$$\Pr\left[\sum_{i<j}\eta_{i,j} = 0\right] \;=\; \frac{\mathrm{Var}\left[\sum_{i<j}\eta_{i,j}\right]}{\left(\mathrm{Exp}\left[\sum_{i<j}\eta_{i,j}\right]\right)^2} \;<\; \frac{\mathrm{Var}\left[\sum_{i<j}\eta_{i,j}\right]}{\left(\binom{m}{2}\cdot 2\gamma\right)^2} \tag{13}$$

We now bound $\mathrm{Var}[\sum_{i<j}\eta_{i,j}]$. Since the $\eta_{i,j}$'s are not pairwise independent, some care is needed: Let $\bar\eta_{i,j} \stackrel{\rm def}{=} \eta_{i,j} - \mathrm{Exp}[\eta_{i,j}]$.

$$\begin{aligned}
\mathrm{Var}\left[\sum_{i<j}\eta_{i,j}\right] &= \mathrm{Exp}\left[\left(\sum_{i<j}\bar\eta_{i,j}\right)^2\right] \\
&= \sum_{i<j}\sum_{k<\ell}\mathrm{Exp}\left[\bar\eta_{i,j}\cdot\bar\eta_{k,\ell}\right] \\
&= \sum_{i<j}\mathrm{Exp}\left[\bar\eta_{i,j}^2\right] + 4\sum_{i<j<k}\mathrm{Exp}\left[\bar\eta_{i,j}\cdot\bar\eta_{j,k}\right] + 0 \\
&= \binom{m}{2}\cdot\mathrm{Exp}[\bar\eta_{1,2}^2] + 4\cdot\binom{m}{3}\cdot\mathrm{Exp}\left[\bar\eta_{1,2}\cdot\bar\eta_{2,3}\right]
\end{aligned} \tag{14}$$

The factor of 4 in the third equality is the number of possibilities among the four elements i,j,k,ℓ (where $i<j$ and $k<\ell$) that exactly two are equal. The 0 term is due to the fact that for $i \neq j \neq k \neq \ell$, the random variables $\eta_{i,j}$ and $\eta_{k,\ell}$ are independent, and hence $\mathrm{Exp}[\bar\eta_{i,j}\cdot\bar\eta_{k,\ell}] = \mathrm{Exp}[\bar\eta_{i,j}]\cdot\mathrm{Exp}[\bar\eta_{k,\ell}] = 0$. We next bound each of the two terms in Equation (14).

$$\mathrm{Exp}[\bar\eta_{1,2}^2] \;\leq\; \mathrm{Exp}[\eta_{1,2}^2] \;=\; \mathrm{Exp}[\eta_{1,2}] \;=\; 2\gamma \tag{15}$$

Let v_i be a random variable that represents the vertex that the $i^{\rm th}$ walk ends at.

$$\begin{aligned}
\mathrm{Exp}[\bar\eta_{1,2}\cdot\bar\eta_{2,3}] &\leq \mathrm{Exp}[\eta_{1,2}\cdot\eta_{2,3}] \\
&\leq \Pr[(\eta_{1,2}=1) \text{ and } (v_3=v_2)] \\
&= \sum_v \Pr[\eta_{1,2}=1 \text{ and } (v_3=v_2=v)] \\
&= \sum_v \Pr[v_3=v \mid \eta_{1,2}=1 \text{ and } (v_2=v)]\cdot\Pr[\eta_{1,2}=1 \text{ and } (v_2=v)] \\
&= \sum_v \Pr[v_3=v]\cdot\Pr[\eta_{1,2}=1 \text{ and } (v_2=v)]
\end{aligned}$$

$$
\begin{aligned}
&\le \max_v\{\Pr[v_3 = v]\} \cdot \sum_v \Pr[(\eta_{1,2} = 1) \text{ and } (v_2 = v)] \\
&= \max_v\{q_{s,v}(t)\} \cdot 2\gamma \\
&\le 2F \cdot \alpha \cdot \gamma \qquad (16)
\end{aligned}
$$

Since by the Lemma's second hypothesis $\gamma \ge \frac{\epsilon}{c} \cdot |\mathrm{S}| \cdot \alpha^2$, we can replace α in Equation (16) with $\sqrt{\frac{c\cdot\gamma}{\epsilon\cdot|\mathrm{S}|}}$ and get

$$
\mathrm{Exp}[\bar{\eta}_{1,2} \cdot \bar{\eta}_{2,3}] \;\le\; 2F \cdot \gamma \cdot \sqrt{\frac{c \cdot \gamma}{\epsilon \cdot |\mathrm{S}|}} \;=\; 2\sqrt{c} \cdot F \cdot \gamma^{\frac{3}{2}} \cdot \sqrt{\frac{1}{\epsilon \cdot |\mathrm{S}|}} \qquad (17)
$$

Combining Equation (13)–(17) we get

$$
\Pr\left[\sum_{i<j} \eta_{i,j} = 0\right] \;=\; O\left(\frac{m^2 \cdot \gamma + m^3 \cdot F \cdot \gamma^{\frac{3}{2}} \cdot \sqrt{\frac{1}{\epsilon\cdot|\mathrm{S}|}}}{m^4 \cdot \gamma^2}\right) \;=\; O\left(\frac{F}{m \cdot \sqrt{\epsilon \cdot |\mathrm{S}| \cdot \gamma}}\right)
$$

As observed above, by the lemma's hypothesis concerning γ, it holds that $\alpha = O(\sqrt{\gamma/(\epsilon|\mathrm{S}|)})$. Since $m = \Omega\left(\frac{F}{\epsilon\cdot\alpha\cdot\sqrt{|\mathrm{S}|}}\right)$, we have that $m = \Omega\left(F\sqrt{\frac{1}{\epsilon\cdot\gamma}}\right)$, and the lemma follows. ■

Based on the construction of $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ we can map walks of length $\ell_1 \cdot \ell_2$ in G to walks of length ℓ_1 in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$, and obtain as a corollary to Lemma 4.5 –

Corollary 4 *Let* H *be a subgraph of* G *and* S, s, ℓ_1, ℓ_2, t, α *and* F *as in Lemma 4.5. Then with probability at least* 0.99, *if we perform* $O\left(\frac{F}{\epsilon\cdot\alpha\cdot\sqrt{|\mathrm{S}|}}\right)$ *random walks of length* $\ell_1 \cdot \ell_2$ *starting from* s *in* G *then for some vertex* v *in* S *we shall reach* v *both on a prefix of a walk that corresponds to an even-length path and on a prefix that corresponds to an odd-length path.*

Proof: Let $\mathrm{M} \stackrel{\text{def}}{=} \mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$ and $L \stackrel{\text{def}}{=} \ell_1 \cdot \ell_2$. We shall map walks of length L in G (starting from $s \in \mathrm{H}$) to walks of length ℓ_1 in $\mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. In case the walk in G does not perform ℓ_2 or more consecutive steps outside of H before it has made at least ℓ_1 steps (not necessarily consecutive) in H, then it is mapped to that sequence of ℓ_1 steps in H. Otherwise, it is mapped to a sequence of less than ℓ_1 steps in H and the remaining steps on an auxiliary path in M. More precisely, we define a mapping ϕ from walks of length L in G to walks of length ℓ_1 in M as follows.

For a walk $w = v_0, \ldots, v_L$ (in G), where $v_0 = s$, let $i_0, \ldots, i_k$ be exactly those indices such that $v_{i_j} \in \mathrm{H}$. (In particular, $i_0 = 0$.) We consider two cases: (1) $k \ge \ell_1$, and for every $0 \le j \le \ell_1 - 1$, $i_{j+1} - i_j < \ell_2$; (2) either $k < \ell_1$, or for some $0 \le j \le \ell_1 - 1$, $i_{j+1} - i_j \ge \ell_2$; In the first case, $\phi(w) \stackrel{\text{def}}{=} v_{i_0} v_{i_1} \ldots v_{i_{\ell_1}}$. In the second case, let i_r be the first index such that $i_{r+1} - i_r \ge \ell_2$ (if no such index exists, i.e., $k < \ell_1$, let $i_r = i_k$). Then $\phi(w) \stackrel{\text{def}}{=} v_{i_0} \ldots v_{i_r} a_{v_{i_r},1} \ldots a_{v_{i_r},\ell_1 - i_r}$. By the definition of M, the distribution on $\phi(w)$ induced by the distribution on w is exactly the same as the distribution on random walks of length ℓ_1 in M.

Let $\Psi_L(\mathrm{G}, s)$ be the probability, when performing walks of length L on G starting from s that for some vertex v in S we shall reach v both on a prefix of a walk that corresponds to an even-length path and on a prefix that corresponds to an odd-length path. Let $\Psi_{\ell_1}(\mathrm{M}, s)$ be the probability, when performing walks of length ℓ_1 on M starting from s that for some vertex v in S we shall end up at v both on a walk that corresponds to an even-length path and on a walk that corresponds to an odd-length path. Then, by the above mapping and Lemma 4.5, $\Psi_L(\mathrm{G}, s) \ge \Psi_{\ell_1}(\mathrm{M}, s) \ge 0.99$. ■

4.6 Putting it all Together (Proof of Theorem 2)

Recall that we need to show that if the test accepts G with probability greater than $\frac{1}{3}$ then G is ϵ-close to bipartite.

We say that a vertex s in G is *good* (for defining a partition) if the probability that odd-cycle(s) returns found is at most 0.1. Otherwise it is *bad*. Since the test rejects G with probability less than $\frac{2}{3}$, and $T = \Omega(1/\epsilon)$, the fraction of *bad* vertices in G is at most $\frac{\epsilon}{16}$. We now show that in such a case we can find a partition of the graph vertices that has at most ϵdN violating edges. We shall do so in steps, where in each step we partition a new set of vertices S until we are left with at most $\frac{\epsilon}{4}N$ vertices. For each partitioned set S we show that: (1) there are few (at most $\frac{\epsilon}{4}d|\mathrm{S}|$) violating edges between pairs of vertices in S; and (2) there are few (at most $\frac{\epsilon}{2}d|\mathrm{S}|$) edges between S and the yet "unpartitioned" vertices R so that no matter how the vertices in R are partitioned, the number of violating edges between S and R is small.

At each step, let D be the set of vertices we have already partitioned, and let H be the subgraph induced by $\mathrm{V} \setminus \mathrm{D}$. Initially, $\mathrm{D} = \emptyset$, and $\mathrm{H} = \mathrm{G}$. Let ℓ_1 and ℓ_2 be as required by Lemma 4.3, and let the length L of the walks we perform on G be $\ell_1 \cdot \ell_2$. Since $\ell_1 = O\left(\left(\frac{\log(N/\epsilon)}{\epsilon}\right)^3\right)$, and $\ell_2 = O\left(\frac{\ell_1}{\epsilon^2}\right)$, we get that $L = O\left(\frac{\log^6(N/\epsilon)}{\epsilon^8}\right)$. Let $\mathrm{M} \stackrel{\mathrm{def}}{=} \mathrm{M}_{\ell_1}^{\ell_2}(\mathrm{H})$. While $|\mathrm{H}| \geq \frac{\epsilon}{4}N$ we do the following. We select any vertex s in H that is both *good* and *useful* with respect to M (see Definition 4.1). By Corollary 3, at least half of the vertices in H are *useful*. Since $|\mathrm{H}| \geq \frac{\epsilon}{4}N$ and the total number of *bad* vertices is $\frac{\epsilon}{16}N < \frac{\epsilon}{8}N$, there exist *good* and *useful* vertices.

We next apply Lemma 4.3 to determine a set S, and an integer t, $\ell_1/2 \leq t \leq \ell_1$, with the properties stated in the lemma. In particular, the number of vertices between S and the rest of H is at most $\frac{\epsilon}{2}d|\mathrm{S}|$, and for every $v \in \mathrm{S}$, $\sqrt{\frac{\beta}{|\mathrm{S}|\cdot|\mathrm{H}|}} \leq q_{s,v}(t) \leq F \cdot \sqrt{\frac{\beta}{|\mathrm{S}|\cdot|\mathrm{H}|}}$, where $F = O\left(\frac{1}{\epsilon}\right)$, and $\beta = \Omega\left(\frac{\epsilon^2}{\log(N/\epsilon)}\right)$. We claim that it must be the case that $\sum_{v\in\mathrm{S}} q_{s,v}^0(t) \cdot q_{s,v}^1(t) \leq \frac{\epsilon\cdot\beta}{2^{10}|\mathrm{H}|}$. This claim, (which we establish momentarily) implies that we can apply Lemma 4.4 (with $\alpha = \sqrt{\frac{\beta}{|\mathrm{S}|\cdot|\mathrm{H}|}}$ (note that $t \geq \ell_1/2 = \Omega(\log(1/\alpha))$ as required)) to show that S can be partitioned so that there are at most $\frac{\epsilon}{4}d|\mathrm{S}|$ violating edges with respect to this partition. The claim holds since otherwise, we could apply Lemma 4.5, or, more precisely Corollary 4, and by letting the number of walks perform from each starting vertex be

$$O\left(\frac{F}{\epsilon\cdot\alpha\cdot\sqrt{|S|}}\right) = O\left(\frac{\sqrt{|H|}}{\epsilon^2\cdot\sqrt{\beta}}\right) = O\left(\frac{\log^{1/2}(N/\epsilon)\cdot\sqrt{N}}{\epsilon^4}\right) = K$$

(where F, α and β are as set above), obtain a contradiction to our assumption the s is *good*.

Thus, as long as $|\mathrm{H}| \geq \frac{\epsilon}{4}N$, each set S contributed at most $\frac{\epsilon}{4} \cdot |\mathrm{S}| \cdot d + \frac{\epsilon}{2} \cdot |\mathrm{S}| \cdot d$ violating edges to the partition. Since these sets are disjoint, all these violating edges sum up to $\frac{3\epsilon}{4} \cdot d \cdot N$. The final H contributes at most $\frac{\epsilon}{4} \cdot N \cdot d$, and so G is ϵ-close to Bipartite.

Verifying that indeed $T = O(1/\epsilon)$, $K = \mathrm{poly}((\log N)/\epsilon)) \cdot \sqrt{N}$, and $L = \mathrm{poly}((\log N)/\epsilon))$, and that the odd-cycle procedure can be implemented in time $\tilde{O}(K \cdot L)$, the theorem follows.

Acknowledgments

Thanks to Nati Linial for helpful discussions.

References

[ALM+92] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. In *Proceedings of the Thirty-Third Annual Symposium on Foundations of Computer Science*, pages 14–23, 1992.

[AS92a] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., 1992.

[AS92b] S. Arora and S. Safra. Probabilistic checkable proofs: A new characterization of NP. In *Proceedings of the Thirty-Third Annual Symposium on Foundations of Computer Science*, pages 1–13, 1992.

[BFL91] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[BFLS91] L. Babai, L. Fortnow, L. Levin, and M. Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, pages 21–31, 1991.

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993.

[FGL+91] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. In *Proceedings of the Thirty-Second Annual Symposium on Foundations of Computer Science*, pages 2–12, 1991.

[GGR96] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. In *Proceedings of the Thirty-Seventh Annual Symposium on Foundations of Computer Science*, pages 339–348, 1996.

[GR96] O. Goldreich and D. Ron. Testing properties of bounded-dgree graphs. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 339–348, 1996.

[Mih89] M. Mihail. Conductance and convergence of Markov chains - A combinatorial treatment of expanders. In *Proceedings 30th Annual Conference on Foundations of Computer Science*, pages 526–531, 1989.

[RS96] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[Rub94] R. Rubinfeld. Robust functional equations and their applications to program testing. In *Proceedings of the Thirty-Fifth Annual Symposium on Foundations of Computer Science*, 1994.

A Proof of Claim 1 in Lemma 4.2

Consider first an even more detailed Markov Chain, denoted $\overline{\mathrm{M}}(\mathrm{H})$. As in $\mathrm{M}(\mathrm{H})$, there is a state in $\overline{\mathrm{M}}(\mathrm{H})$ for every vertex in H, and the transitions between vertices in H are as in $\mathrm{M}(\mathrm{H})$ (i.e., as in walks on G). However, between each u and v in $\mathrm{B}(\mathrm{H})$, there is an auxiliary path for *every walk* from u to v that passes only through vertices not in H (rather than for every walk-length as in $\mathrm{M}(\mathrm{H})$). Each such walk is determined by a sequence of *transition-labels*. A transition from x to y, where x and y are neighbors in G, is given the label of the edge from x to y. As for self-transitions from x to itself, we think of there being $2d - |\Gamma(x)|$ transitions, labeled

$|\Gamma(x)|+1,\ldots,2d$. Each of these self-transitions has probability $\frac{1}{2d}$. By this definition, for any integer ℓ, a walk of length ℓ between any two vertices has probability $\left(\frac{1}{2d}\right)^{\ell}$.

In view of the above, the probability of entering an auxiliary path in $\overline{\mathrm{M}}(\mathrm{H})$ from $u \in \mathrm{B}(\mathrm{H})$ to $v \in \mathrm{B}(\mathrm{H})$, corresponding to a walk w outside of H, is $\left(\frac{1}{2d}\right)^{|w|}$. The transition probabilities between each auxiliary state on an auxiliary path and the next state on the path (or the vertex reached in $\mathrm{B}(\mathrm{H})$), is 1. Note that for each auxiliary path from u to v that corresponds to a walk w, there is an auxiliary path from v to u (corresponding to the reverse of w), where both are entered with exactly the same probability.

Given the definition of $\overline{\mathrm{M}}(\mathrm{H})$, we see that $\overline{\mathrm{M}}(\mathrm{H})$ can be transformed into $\mathrm{M}(\mathrm{H})$ as follows. For every pair of vertices $u, v \in \mathrm{B}(\mathrm{H})$, and for each length ℓ, all auxiliary paths of length ℓ between u and v in $\overline{\mathrm{M}}(\mathrm{H})$ are merged into a single auxiliary path in $\mathrm{M}(\mathrm{H})$. The probability of entering the resulting path in $\mathrm{M}(\mathrm{H})$ is the sum over the probabilities of entering the corresponding paths in $\mathrm{M}(\mathrm{H})$. It follows that the stationary probability of each auxiliary state in $\mathrm{M}(\mathrm{H})$ is the sum of the stationary probabilities of the auxiliary states in $\overline{\mathrm{M}}(\mathrm{H})$ that were merged into it, while the stationary probability of vertices in H remains the same. However, it is not hard to verify that the stationary probability in $\overline{\mathrm{M}}(\mathrm{H})$ of each vertex in H, is the same as in walks on G, i.e., it is $\frac{1}{N}$. This follows from the correspondence between walks on G and walks on $\overline{\mathrm{M}}(\mathrm{H})$. Stated slightly differently, it follows from the fact that $\overline{\mathrm{M}}(\mathrm{H})$ can be transformed into the Markov chain defined by walks on G by merging, for each vertex $v \in \mathrm{G} \setminus \mathrm{H}$, all auxiliary states in $\overline{\mathrm{M}}(\mathrm{H})$ that correspond to that vertex. □

B Proof of Claim 5 in Lemma 4.3

Let $X \stackrel{\rm def}{=} \sum_i x_i^2$, and $m_+ \stackrel{\rm def}{=} |\{i : x_i > 0\}|$. Assume in contradiction that $\sum_{i,\, x_i>0} x_i^2 < \frac{\gamma}{4} \cdot X$. Conditioned on this bound on the sum of their squares, the sum of the positive x_i's is maximized when they are all equal, i.e., when each x_i is $\sqrt{\frac{\gamma X}{4m_+}}$. Hence,

$$\sum_{i,\, x_i>0} x_i \le m_+ \cdot \sqrt{\frac{1}{m_+} \cdot \frac{\gamma}{4} \cdot X} < \sqrt{m \cdot \frac{\gamma}{4} \cdot X} \tag{18}$$

We next observe that the Claim's first hypothesis implies that

$$\sum_{i,\, x_i<0} |x_i| \;=\; \sum_{i,\, x_i>0} x_i - \sum_i x_i \;\le\; \sum_{i,\, x_i>0} x_i + \gamma \tag{19}$$

By Equations (19) and (18),

$$\sum_{i,\, x_i<0} |x_i| \;\le\; \sqrt{m \cdot \frac{\gamma}{4} \cdot X} + \gamma \;\le\; \sqrt{m \cdot \frac{\gamma}{4} \cdot X} + \frac{1}{2} \cdot m \cdot X \tag{20}$$

where the second inequality follows from the second hypothesis of the claim (and the definition of X). Since for every negative x_i, $|x_i| \le \frac{1}{m}$, Equation (20) implies that

$$\sum_{i,\, x_i<0} x_i^2 \;\le\; \frac{1}{m} \cdot \sum_{i,\, x_i<0} |x_i| \;\le\; \frac{1}{m} \cdot \left(\sqrt{m \cdot \frac{\gamma}{4} \cdot X} + \frac{1}{2} \cdot m \cdot X \right) \;=\; \sqrt{\frac{\gamma}{4} X/m} + \frac{1}{2} \cdot X \tag{21}$$

Putting together the initial contrary assumption that $\sum_{i,\, x_i>0} x_i^2 < \frac{\gamma}{4} \cdot X$ with Equation (21), we get that

$$\begin{aligned} X &= \sum_{i,\, x_i>0} x_i^2 + \sum_{i,\, x_i<0} x_i^2 \\ &< \frac{\gamma}{4} \cdot X + \sqrt{\frac{\gamma}{4m}} \cdot \sqrt{X} + \frac{1}{2} X \end{aligned}$$

But this implies that

$$X < \frac{1}{(\frac{1}{2} - \frac{\gamma}{4})^2} \cdot \frac{\gamma}{4m}$$

which for $\gamma \leq 1/2$ is less than $2\gamma/m$, and we have reached a contradiction to the second hypothesis of the Claim. ■

C Proof of Proposition 1

We show the counterpositive of the claim. Namely, if there are no odd-cycles in G of length at most L then G is ϵ-close to bipartite.

Consider first the (simple) case in which all vertices in G are reachable from some vertex s by paths of length $L/2$. Consider a breadth-first-search (BFS) tree rooted at s, and the partition induced by putting odd-level vertices on one side and the rest on the other. By our hypothesis (non-existence of short odd-cycles), there can be no edges between vertices of the same level (and by the properties of a BFS tree there can be no edges between vertices which differ in levels by more than 1). Thus, the above partition demonstrates that G is bipartite.

In the more general case, we start an *iterative* process by which we partition the vertices in the graph. In each iteration, let D be the set of vertices that have already been assigned a side in the partition. Initially, $\mathrm{D} = \emptyset$. Consider a BFS tree in the subgraph induced by $\mathrm{V} \setminus \mathrm{D}$ starting from some vertex $s \in \mathrm{V} \setminus \mathrm{D}$. Let i be the first level such that the number of vertices in level $i+1$ is smaller than ϵ times the number of vertices in all i first levels. The existence of such an $i \leq L$ follows from our choice of $L \geq \log_{1+\epsilon} N$. Denote the nodes in the first i levels by D'. Then, the number of edges between D' and the rest of $\mathrm{V} - \mathrm{D}$ is at most $\epsilon d|\mathrm{D}'|$, where d is the degree bound. As for D' itself, the subgraph induced by it is bipartite (by an argument as in the simple case). Thus, we set $\mathrm{D} = \mathrm{D} \cup \mathrm{D}'$ and proceed. Each D' accounts for at most $\epsilon d|\mathrm{D}'|$ potentially violating edges (between D' and the yet unpartitioned part of G), totaling to an ϵ fraction of dN. ■