On Testing Bipartiteness in the Bounded-Degree Graph Model

Oded Goldreich*

September 8, 2017

Summary

This memo presents a variant of the known tester of Bipartiteness in the bounded-degree graph model, which is presented in Section 9.4.1 of my book on Property Testing (hereafter referred to as *the book*). The purpose of this variation is to show that, *when the graph is rapid mixing, Bipartiteness can be tested in $O(\sqrt{k})$ time, rather than in $\widetilde{O}(\sqrt{k})$ time.*

Much of the following text is reproduced from Section 9.4.1 of the book, and the essence of the improvement is in capitalizing on half of the vertices that appear on each (2ℓ-step long) random walk rather than using only the last vertex in each of the m walks. This is reflected in the proof of Claim 3.2, where we consider $\ell^2 \cdot (m^2 - m)$ collision events (rather than the $\cdot(m^2 - m)$ events considered in the book).

The tester

This memo refers to testing Bipartiteness in the bounded-degree graph model. It shows that, for constant $\epsilon > 0$, Bipartiteness can be tested in $O(\sqrt{k})$ time, *provided that the graph is rapid mixing* (in a natural sense that is slightly stronger than that used in the book).

Algorithm 1 (testing `Bipartiteness` (the rapid mixing case)): *On input d, k, ϵ and oracle access to an incidence function of an k-vertex graph, $G = ([k], E)$, of degree bound d, proceed as follows*

1. *Pick an arbitrary vertex s in $[k]$.*
2. (Try to find an odd-length cycle through vertex s):
   (a) *Perform $m \stackrel{\rm def}{=} O(\sqrt{k/\epsilon}/\log k)$ random walks starting from s, each of length 2ℓ, where $\ell \stackrel{\rm def}{=} O(\log k)$.*
   (b) *Let R_0* (respectively, R_1) *denote the set of vertices reached from s in an even* (respectively, odd) *number of steps in any of these walks. That is, for every such walk $(s = v_0, v_1, ..., v_{2\ell})$, place $v_0, v_2, ..., v_{2\ell}$ in R_0 and place $v_1, v_3, ..., v_{2\ell-1}$ in R_1.*
   (c) *If $R_0 \cap R_1$ is not empty, then* reject. *Otherwise,* accept.

*Department of Computer Science, Weizmann Institute of Science, Rehovot, ISRAEL.

Algorithm 1 is a degenerate version of the bipartite tester presented in the book (i.e., Algorithm 9.20). Specifically, we perform only one iteration of the main loop (rather than $O(1/\epsilon)$ iterations), start it at an arbitrary vertex (rather than at a random one), and, most importantly, we take $O(\sqrt{k/\epsilon}/\log k)$ random walks of length $O(\log k)$ (rather than $\widetilde{O}(\sqrt{k})$ random walks of length $\text{poly}(\epsilon^{-1}\log k)$). Hence, the time (and query) complexity of Algorithm 1 is $m \cdot \ell \cdot \log d = O(\sqrt{k/\epsilon})$, where the $\log d$ factor is due to determining the degree of each vertex encountered in the random walk (before selecting a neighbor at random) and it can be avoided when taking a lazy random walk (as performed in the analysis anyhow).

The analysis

As in the book, we actually consider lazy random walks rather than genuine random walks. Recall that a lazy random walk (as in Definition 2) traverses each of the incident edges with probability $1/2d$, and stays in the current vertex otherwise.

The rapid mixing feature that we assume here is natural but somewhat stronger than the one used in the book. It asserts that for every vertex $v \in [k]$ and number of steps $t \in [2\ell]$, a lazy random walk of length t starting at s reaches v with probability at most $\frac{1}{k} + \exp(-\Omega(t))$, which implies the original rapid mixing condition for $\ell = O(\log k)$.[1]

Definition 2 (the rapid mixing feature): *Let $(v_1, ..., v_t) \leftarrow \mathcal{RW}_t(v_0)$ be an t-step lazy random walk* (on $G = ([k], E)$) *starting at v_0; that is, for every $\{u, v\} \in E$ and every $i \in [t]$, it holds that*

$$\mathbf{Pr}_{(v_1,\ldots,v_t)\leftarrow\mathcal{RW}_t(v_0)}[v_i = v | v_{i-1} = u] = \frac{1}{2d} \tag{1}$$

$$\mathbf{Pr}_{(v_1,\ldots,v_t)\leftarrow\mathcal{RW}_t(v_0)}[v_i = u | v_{i-1} = u] = 1 - \frac{d_G(u)}{2d} \tag{2}$$

where $d_G(u)$ denotes the degree of u in G. Then, the graph G is said to be rapidly mixing *if, for every $t \in \mathbb{N}$ and $v_0, v \in [k]$, it holds that*

$$\mathbf{Pr}_{(v_1,\ldots,v_t)\leftarrow\mathcal{RW}_t[v_0]}[v_t = v] < \frac{1}{k} + \exp(-\Omega(t)). \tag{3}$$

Note that if the graph is an expander, then it is rapidly mixing.

Lemma 3 (analysis of Algorithm 1): *Suppose that $m \cdot \ell = \Omega(\sqrt{k/\epsilon})$ and $m = \Omega(1/\epsilon)$. If the graph G is rapidly mixing and ϵ-far from Bipartite, then Algorithm 1 rejects with probability at least* $2/3$.

The key quantities in the analysis are the following probabilities that refer to the parity *of the length of a path obtained from the lazy random walk by omitting the self-loops* (transitions that remain at the current vertex). Let $p_{t,0}(v)$ (respectively, $p_{t,1}(v)$) denote the probability that a *lazy random walk of length t, starting at s, reaches v while making an even* (respectively, *odd*) *number of real* (i.e., non-self-loop) *steps.* That is, for every $t \in \mathbb{N}$, $\sigma \in \{0, 1\}$ and $v \in [k]$,

$$p_{t,\sigma}(v) \stackrel{\text{def}}{=} \mathbf{Pr}_{(v_1,\ldots,v_t)\leftarrow\mathcal{RW}_t(s)}[v_t = v \ \wedge \ |\{i \in [t] : v_i \neq v_{i-1}\}| \equiv \sigma \pmod 2)] \tag{4}$$

[1] The original rapid mixing condition is stated in Eq. (9.5). Actually, it suffices to strengthen the original rapid mixing condition by augmenting it with the requirement that, for any $v \in [k]$ and $t \in [2\ell]$, a lazy random walk of length t starting at s reaches v with probability at most $O(1/t^3)$.

The path-parity of the walk $(v_1, ..., v_t)$ is defined as $|\{i \in [t] : v_i \neq v_{i-1}\}| \bmod 2$. (Note that $p_{\ell,\sigma}$ coincides with p_σ as defined in the book.)

By the rapid mixing assumption (for every $v \in [k]$), it holds that for every $t \in [\ell+1, 2\ell]$ it holds that

$$\frac{1}{2k} \;<\; p_{t,0}(v) + p_{t,1}(v) \;<\; \frac{2}{k}. \tag{5}$$

Letting $p_\sigma(v) \stackrel{\rm def}{=} \sum_{t=\ell+1}^{2\ell} p_{t,\sigma}$, we consider two cases regarding the sum $\sum_{v\in[k]} p_0(v)p_1(v)$: If the sum is (relatively) "small", then we show that $[k]$ can be 2-partitioned so that there are relatively few edges between vertices that are placed in the same side, which implies that G is close to being bipartite. Otherwise (i.e., when the sum is not "small"), we show that, with significant probability, when Step 2 is started at vertex s, it is completed by rejecting G. These two cases are analyzed in the following two (corresponding) claims.

Claim 3.1 (a small sum implies closeness to being bipartite): *Suppose* $\sum_{v\in[k]} p_0(v)p_1(v) \;\leq\; 0.01\ell^2 \cdot \epsilon/k$. *Let* $V_1 \stackrel{\rm def}{=} \{v \in [k] : p_0(v) < p_1(v)\}$ *and* $V_2 = [k] \setminus V_1$. *Then, the number of edges with both end-points in the same* V_σ *is bounded above by* $\epsilon dk/2$, *which implies that* G *is* ϵ*-close to being bipartite.*

Proof Sketch: Consider an edge $\{u, v\}$ such that both u and v are in the same V_σ, and assume, without loss of generality, that $\sigma = 1$. Then, by the (lower bound of the) *rapid mixing hypothesis*, both $p_1(v)$ and $p_1(u)$ are greater than $\frac{1}{2} \cdot \frac{\ell}{2k}$ (since $p_{t,0}(v) + p_{t,1}(v) \geq \frac{2}{k}$ for every $t \in [\ell+1, 2\ell]$). Using observations as in the book, we infer that $p_0(v) > \frac{1}{3d} \cdot p_1(u)$ (since $p_{t,0}(v) > \frac{1}{3d} \cdot p_{t,1}(v)$ for every $t \in [\ell+1, 2\ell]$). Thus, the edge $\{u, v\}$ contributes at least $\frac{p_1(u)}{3d} \cdot p_1(v) \geq \frac{(\ell/4k)^2}{3d}$ to the sum $\sum_{w\in[k]} p_0(w)p_1(w)$. It follows that we can have at most $\frac{0.01\ell^2 \cdot \epsilon/k}{\ell^2/(48dk^2)} < \epsilon dk/2$ such edges, and the claim follows. ■

Claim 3.2 (a large sum implies high rejection probability): *Suppose* $\sum_{v\in[k]} p_0(v)p_1(v) \;\geq\; 0.01\ell^2 \cdot \epsilon/k$, *and that Step 2 is executed with vertex* s. *If* $m \cdot \ell = \Omega(\sqrt{k/\epsilon})$ *and* $m = \Omega(1/\epsilon)$, *then, with probability at least* $2/3$, *the set* $R_0 \cap R_1$ *is not empty* (and rejection follows).

Proof: Consider the probability space defined by an execution of Step 2 (with start vertex s). For every $i \neq j$ such that $i, j \in [m]$ and every $t_0, t_1 \in [\ell+1, 2\ell]$, we define an indicator random variable $\zeta_{(i,t_0),(j,t_1)}$ representing *the event that the vertex encountered in the* $t_0^{\rm th}$ *step of the* $i^{\rm th}$ *walk equals the vertex encountered in the* $t_1^{\rm th}$ *step of the* $j^{\rm th}$ *walk, and that the* t_0*-step prefix of the* $i^{\rm th}$ *walk has an even path-parity whereas the* t_1*-step prefix of the* $j^{\rm th}$ *walk has an odd path-parity.* Recalling the definition of the $p_{t,\sigma}(v)$'s, observe that $\mathbf{Pr}[\zeta_{(i,t_0),(j,t_1)}{=}1] = \sum_{v\in[k]} p_{t_0,0}(v)p_{t_1,1}(v)$. Hence,

$$\begin{aligned}
\sum_{i\neq j} \sum_{t_0,t_1\in[\ell+1,2\ell]} \mathbb{E}[\zeta_{(i,t_0),(j,t_1)}] &= m(m-1) \cdot \sum_{t_0,t_1\in[\ell+1,2\ell]} \sum_{v\in[k]} p_{t_0,0}(v)p_{t_1,1}(v) \\
&= m(m-1) \cdot \sum_{v\in[k]} p_0(v)p_1(v) \\
&> \frac{600k}{\ell^2 \cdot \epsilon} \cdot \sum_{v\in[k]} p_0(v)p_1(v) \\
&\geq 6
\end{aligned}$$

where the first inequality is due to the setting of m, and the second inequality is due to the claim's hypothesis. Note that $\mathbf{Pr}[|R_0 \cap R_1| > 0] \geq \mathbf{Pr}[\sum_{i \neq j} \sum_{t_0,t_1 \in [\ell+1,2\ell]} \zeta_{(i,t_0),(j,t_1)} > 0]$, since whenever the event captured by the $\zeta_{(i,t_0),(j,t_1)}$ holds it is the case that the t_0^{th} vertex of the i^{th} walk (which equals the t_1^{th} vertex of the j^{th} walk) resides in $R_0 \cap R_1$.

Intuitively, the sum of the $\zeta_{(i,t_0),(j,t_1)}$'s should be positive, with high probability, since the expected value of the sum is large enough and the $\zeta_{(i,t_0),(j,t_1)}$'s are "sufficiently independent" (almost all pairs of $\zeta_{(i,t_0),(j,t_1)}$'s are independent). The intuition is indeed correct, but proving it is less straightforward than it seems, since the $\zeta_{(i,t_0),(j,t_1)}$'s are not pairwise independent. Yet, since the sum of the covariances of the dependent $\zeta_{(i,t_0),(j,t_1)}$'s is quite small, Chebyshev's Inequality is still very useful. Specifically, let $\zeta_{i,j} = \sum_{t_0,t_1 \in [\ell+1,2\ell]} \zeta_{(i,t_0),(j,t_1)}$, and note that $\mathbb{E}[\zeta_{i,j}] = \sum_{v \in [k]} p_0(v)p_1(v)$, since

$$\begin{aligned}
\mathbb{E}\left[\sum_{t_0,t_1 \in [\ell+1,2\ell]} \zeta_{(i,t_0),(j,t_1)}\right] &= \sum_{v \in [k]} \sum_{t_0,t_1 \in [\ell+1,2\ell]} p_{t_0,0}(v)p_{t_1,1}(v) \\
&= \sum_{v \in [k]} p_0(v)p_1(v).
\end{aligned}$$

Then, letting $\mu \stackrel{\text{def}}{=} \mathbb{E}[\zeta_{i,j}] = \sum_{v \in [k]} p_0(v)p_1(v)$, and $\overline{\zeta}_{i,j} \stackrel{\text{def}}{=} \zeta_{i,j} - \mu$, we get:

$$\begin{aligned}
\mathbf{Pr}\left[\sum_{i \neq j} \zeta_{i,j} = 0\right] &< \frac{\mathbb{V}\left[\sum_{i \neq j} \zeta_{i,j}\right]}{(m(m-1) \cdot \mu)^2} \\
&= \frac{1}{m^2(m-1)^2\mu^2} \cdot \sum_{i_1 \neq j_1, i_2 \neq j_2} \mathbb{E}\left[\overline{\zeta}_{i_1,j_1}\overline{\zeta}_{i_2,j_2}\right]
\end{aligned}$$

We partition the terms in the last sum according to the number of distinct elements such that, for $t \in \{2,3,4\}$, we let $(i_1,j_1,i_2,j_2) \in S_t \subseteq [m]^4$ if and only if $|\{i_1,j_1,i_2,j_2\}| = t$ (and $i_1 \neq j_1 \wedge i_2 \neq j_2$). Hence,

$$\mathbf{Pr}\left[\sum_{i \neq j} \zeta_{i,j} = 0\right] < \frac{1}{m^2(m-1)^2\mu^2} \cdot \sum_{t \in \{2,3,4\}} \sum_{(i_1,j_1,i_2,j_2) \in S_t} \mathbb{E}\left[\overline{\zeta}_{i_1,j_1}\overline{\zeta}_{i_2,j_2}\right] \tag{6}$$

Now, note that if $i_1 = j_2$ (resp., $i_2 = j_1$), then $\mathbb{E}[\overline{\zeta}_{i_1,j_1}\overline{\zeta}_{i_2,j_2}] \leq \mathbb{E}[\zeta_{i_1,j_1}\zeta_{i_2,j_2}] = 0$, where the equality is due to the fact that in this case $\zeta_{i_1,j_1} = 1$ and $\zeta_{i_2,j_2} = 1$ make conflicting requirements of the path-parity of walk number $i_1 = j_2$ (resp., $i_2 = j_1$). Hence, rather than summing over the S_t's, we can sum over the following S'_t's defined such that $(i_1,j_1,i_2,j_2) \in S'_t \subseteq S_t$ if and only if $i_1 \neq j_2 \wedge i_2 \neq j_1$. Furthermore, the contribution of each element in $S'_4 = S_4$ to the sum is zero, since the four walks are independent and so $\mathbb{E}[\overline{\zeta}_{i_1,j_1}\overline{\zeta}_{i_2,j_2}] = \mathbb{E}[\overline{\zeta}_{i_1,j_1}] \cdot \mathbb{E}[\overline{\zeta}_{i_2,j_2}] = 0$. Plugging all of this into Eq. (6), we get

$$\begin{aligned}
\mathbf{Pr}\left[\sum_{i \neq j} \zeta_{i,j} = 0\right] &< \frac{1}{m^2(m-1)^2\mu^2} \cdot \sum_{t \in \{2,3\}} \sum_{(i_1,j_1,i_2,j_2) \in S'_t} \mathbb{E}\left[\overline{\zeta}_{i_1,j_1}\overline{\zeta}_{i_2,j_2}\right] \\
&= \frac{1}{m^2(m-1)^2\mu^2} \cdot \left(\sum_{i \neq j} \mathbb{E}\left[\overline{\zeta}^2_{i,j}\right] + \sum_{i_1,i_2,i_3 : |\{i_1,i_2,i_3\}|=3} \left(\mathbb{E}\left[\overline{\zeta}_{i_1,i_2}\overline{\zeta}_{i_1,i_3}\right] + \mathbb{E}\left[\overline{\zeta}_{i_1,i_2}\overline{\zeta}_{i_3,i_2}\right]\right)\right)
\end{aligned}$$

$$< \frac{1}{(m-1)^2\mu^2} \cdot \mathbb{E}[\zeta_{1,2}^2] + \frac{1}{(m-1)\mu^2} \cdot (\mathbb{E}[\zeta_{1,2}\zeta_{1,3}] + \mathbb{E}[\zeta_{1,2}\zeta_{3,2}]) \tag{7}$$

where in the inequality uses $\mathbb{E}[\overline{\zeta}_{i_1,j_1}\overline{\zeta}_{i_2,j_2}] \leq \mathbb{E}[\zeta_{i_1,j_1}\zeta_{i_2,j_2}]$ and $|\{(i_1,i_2,i_3) \in [m^3] : |\{i_1,i_2,i_3\}| = 3\}| < m^2 \cdot (m-1)$. The obvious upper bound on the first term (which uses $\mathbb{E}[\eta_{1,2}^2] \leq \mathbb{E}[\eta_{1,2} \cdot \ell] = \ell \cdot \mu$) is not good enough. Hence, we upper-bound the first term of Eq. (7) by using

$$\begin{aligned}
\mathbb{E}[\zeta_{1,2}^2] &= \sum_{t_0,t_1,t_2,t_3 \in [\ell+1,2\ell]} \mathbb{E}[\zeta_{(1,t_0),(2,t_1)}\zeta_{(1,t_2),(2,t_3)}] \\
&\leq 2 \cdot \sum_{i \geq 0} \sum_{t_0,t_1,t_3 \in [\ell+1,2\ell] : |t_1-t_3| \leq i} \mathbb{E}[\zeta_{(1,t_0),(2,t_1)}\zeta_{(1,t_0+i),(2,t_3)}] \\
&\quad + 2 \cdot \sum_{i \geq 0} \sum_{t_0,t_1,t_2 \in [\ell+1,2\ell] : |t_0-t_2| \leq i} \mathbb{E}[\zeta_{(1,t_0),(2,t_1)}\zeta_{(1,t_2),(2,t_1+i)}] \\
&\leq 2 \cdot \sum_{t_0,t_1 \in [\ell+1,2\ell]} \mathbb{E}[\zeta_{(1,t_0),(2,t_1)}] \\
&\qquad \left(\sum_{i \geq 0} \sum_{t_3 \in [t_1-i,t_1+i]} \mathbf{Pr}[\zeta_{(1,t_0+i),(2,t_3)} = 1 | \zeta_{(1,t_0),(2,t_1)} = 1] \right. \\
&\qquad \left. + \sum_{i \geq 0} \sum_{t_2 \in [t_0-i,t_0+i]} \mathbf{Pr}[\zeta_{(1,t_2),(2,t_1+i)} = 1 | \zeta_{(1,t_0),(2,t_1)} = 1] \right)
\end{aligned}$$

where the first inequality is justified by letting $i = \max(|t_0-t_2|, |t_1-t_3|)$. For every $t_0, t_1 \in [\ell+1, 2\ell]$ and t_3, the event represented by $[\zeta_{(1,t_0+i),(2,t_3)} = 1 | \zeta_{(1,t_0),(2,t_1)} = 1]$ corresponds to a second collision of the first walk with the second walk, when given that a first collision has occured (at the t_0^{th} step of the first walk). Fixing the second walk and conditioning on the first collision, by the strong rapid mixing feature, the $t_i + i^{\text{th}}$ step of first walk hits the relevant vertex with probability at most $O(1/(i+1)^3)$. The same considerations apply to the events represented by $[\zeta_{(1,t_2),(2,t_1+i)} = 1 | \zeta_{(1,t_0),(2,t_1)} = 1]$ (except that here we fix the first walk and look at a second collision that occurs at the $t_1 + i^{\text{th}}$ step of the second walk). Hence,

$$\begin{aligned}
\mathbb{E}[\zeta_{1,2}^2] &\leq 2 \cdot \sum_{t_0,t_1 \in [\ell+1,2\ell]} \mathbb{E}[\zeta_{(1,t_0),(2,t_1)}] \cdot 2 \sum_{i \geq 0} (2i+1) \cdot O(1/(i+1)^3) \\
&= O(\mu) \cdot \sum_{i \geq 0} (1/(i+1)^2) \\
&= O(\mu),
\end{aligned}$$

where the equality uses $\sum_{t_0,t_1 \in [\ell+1,2\ell]} \mathbb{E}[\zeta_{(1,t_0),(2,t_1)}] = \mu$.

For the second term of Eq. (7), we observe that $\mathbb{E}[\zeta_{1,2}\zeta_{1,3}] = \mathbf{Pr}[\zeta_{1,2} = \zeta_{1,3} = 1]$ is upper-bounded by $\mathbf{Pr}[\zeta_{1,2} = 1] = \mu$ times the probability that the one of the last ℓ vertices of the third walk appears as one of the last ℓ vertices of the first path, since $\zeta_{1,3} = 1$ mandates the latter event. Using the (upper bound of the) *rapid mixing hypothesis*, we upper-bound the latter probability by $2\ell^2/k$, and obtain $\mathbb{E}[\zeta_{1,2}\zeta_{2,3}] \leq \mu \cdot 2\ell^2/k$. (Ditto for $\mathbb{E}[\zeta_{1,2}\zeta_{3,2}]$.) Plugging all of this into Eq. (7), we get

$$
\begin{aligned}
\mathbf{Pr}[|R_0 \cap R_1| = 0] &\le \mathbf{Pr}\left[\sum_{i \neq j} \zeta_{i,j} = 0\right] \\
&< \frac{O(\mu)}{(m-1)^2\mu^2} + \frac{2}{(m-1)\mu^2} \cdot \frac{2\ell^2\mu}{k} \\
&= \frac{O(1)}{(m-1)^2\mu} + \frac{4\ell^2}{(m-1)\mu k} \\
&< \frac{1}{3}
\end{aligned}
$$

where the last inequality uses $\mu \ge 0.01\ell^2 \cdot \epsilon/k$ and $m\ell = \Omega(\sqrt{k/\epsilon})$ (and $m = \Omega(\epsilon^{-1})$), which imply $m^2\mu = \Omega(1)$ and $m \cdot \mu k = \Omega(\epsilon^{-1} \cdot \ell^2\epsilon)$. The claim follows. ■

Acknowledgements

I am grateful to Reut Levi for a discussion that triggered this result as well as for commenting on early drafts of this write-up.