

DETECTING TAMPERING IN RANDOM GRAPHS

ROSS G. PINSKY

ABSTRACT. Let $\mathcal{G}_n = (V_n, E_n)$ be a growing sequence of deterministic finite graphs, with V_n denoting the vertices and E_n denoting the edges. Consider the random graph $\mathcal{G}_n(p_n) = (V_n, E_n(p_n))$ obtained by including any given edge with probability p_n , independent of other edges, and let $P_n^{p_n}$ denote the corresponding probability measure on \mathcal{G}_n . Now tamper with the random graph in some regular way. For example, if \mathcal{G}_n is the complete graph on n vertices, so that $\mathcal{G}_n(p_n)$ is the Erdos-Renyi graph, then one might tamper with it by disconnecting all the edges of a randomly chosen vertex, or by adding all the edges of a randomly chosen Hamiltonian path from \mathcal{G}_n , or by adding all the edges of a randomly chosen clique of order k_n from \mathcal{G}_n . Denote the resulting induced measure on \mathcal{G}_n by $P_n^{p_n, \text{tamper}}$. The tampering is called *detectable* if $\lim_{n \rightarrow \infty} \|P_n^{p_n, \text{tamper}} - P_n^{p_n}\|_{\text{TV}} = 1$, *strongly undetectable* if the above limit is 0, and *weakly undetectable* if $\{\|P_n^{p_n, \text{tamper}} - P_n^{p_n}\|_{\text{TV}}\}_{n=1}^{\infty}$ is bounded away from 0 and 1. We study the tampering problem for a variety of examples.

1. INTRODUCTION AND STATEMENT OF RESULTS

The motivation for this paper comes from the following example. Consider a random permutation $\sigma \in S_n$ as a row of n cards labeled from 1 to n and laid out from left to right in random order. Now tamper with the cards as follows. Select k_n of the cards at random, remove them from the row, and then replace them in the vacated spaces in increasing order. Let U_n denote the uniform measure on S_n , that is, the measure corresponding to a “random permutation,” and let $U_n^{\text{incsubseq}, k_n}$ denote the measure

1991 *Mathematics Subject Classification.* 05C80, 60C05 .

Key words and phrases. random graph, Erdos-Renyi graph, random hypercube, total variation norm, Hamiltonian path.

on S_n induced from U_n by the above tampering. Can one detect the tampering asymptotically as $n \rightarrow \infty$? Let $\|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}}$ denote the total variation norm between the probability measures U_n and $U_n^{\text{incsubseq}, k_n}$. If $\lim_{n \rightarrow \infty} \|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}} = 1$, we call the tampering detectable. If $\lim_{n \rightarrow \infty} \|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}} = 0$, we call the tampering strongly undetectable, while if $\{\|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}}\}_{n=1}^{\infty}$ is bounded away from 0 and 1, we call the tampering weakly undetectable. Note that by construction, a permutation $\sigma \in S_n$ will have an increasing sequence of length k_n with $U_n^{\text{incsubseq}, k_n}$ -probability 1. On the other hand, the celebrated result concerning the length of the longest increasing subsequence in a random permutation ([3], [9], [1]) states that the U_n probability of there being an increasing subsequence of length $cn^{\frac{1}{2}}$ goes to 0 as $n \rightarrow \infty$, if $c > 2$. Thus, one certainly has $\lim_{n \rightarrow \infty} \|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}} = 1$, if $k_n \geq cn^{\frac{1}{2}}$, with $c > 2$. The above-mentioned result also states that the U_n -probability of there being an increasing subsequence of length $cn^{\frac{1}{2}}$ goes to 1 as $n \rightarrow \infty$, if $c < 2$. From this it follows that for $k_n \leq cn^{\frac{1}{2}}$, $c < 2$, the distribution of the *number* of increasing subsequences of length k_n , which we denote by $N_{S_n}^{\text{incr}, k_n}$, converges to the δ -distribution at ∞ as $n \rightarrow \infty$. This might suggest intuitively that one can tamper on the order k_n without detection, if $k_n \leq cn^{\frac{1}{2}}$ with $c < 2$; after all, how much can one additional increasing subsequence be felt in such a situation? However, this turns out to be false. In [7], it was shown that $\lim_{n \rightarrow \infty} \|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}} = 0$, if $k_n \leq n^l$ with $l < \frac{2}{5}$ and in [8] it was shown that $\lim_{n \rightarrow \infty} \|U_n - U_n^{\text{incsubseq}, k_n}\|_{\text{TV}} = 1$, if $k_n \geq n^l$ with $l > \frac{4}{9}$.

In this paper, we study the above type of phenomenon for random graphs. Take a growing sequence of deterministic finite graphs $\{\mathcal{G}_n\}_{n=1}^{\infty}$, where $\mathcal{G}_n = (V_n, E_n)$, with V_n denoting the vertices and E_n denoting the edges. Let $p_n \in (0, 1)$ and let $E_n(p_n)$ denote the random set of edges obtained by including any particular edge from E_n with probability p_n and excluding it with probability $1 - p_n$, independently for all of the edges. We consider the

random graph $\mathcal{G}_n(p_n) = (V_n, E_n(p_n))$. Denote the corresponding probability measure by $P_n^{p_n}$.

We begin with two examples in order to motivate and make easily understood the abstract construction that follows. For the first example, let $\mathcal{G}_n \equiv G(n)$ be the complete graph on n vertices. The random graph $\mathcal{G}_n(p_n) \equiv G(n; p_n)$ is the so-called Erdos-Renyi graph with edge probabilities p_n . Consider the random graph with $p_n = p$ fixed for all n . For some choice of k_n , consider the class of cliques of size k_n in $G(n)$; that is, the class of complete subgraphs of $G(n)$ of order k_n . Randomly select one of the cliques of order k_n and “add” it to the random graph; that is, add to the random graph every edge of this clique that is not already in the random graph. Call the induced measure $P_n^{p, \text{clique}, k_n}$. For which choices of k_n is the tampering detectable, for which choices is it strongly undetectable, and for which choices is it weakly undetectable? That is, for which choices of k_n does $\|P_n^{p, \text{clique}, k_n} - P_n^p\|_{\text{TV}}$ converge to 1, for which choices does it converge to 0 and for which choices is it bounded away from 0 and 1?

For the second example, again let $\mathcal{G}_n = G(n)$, the complete graph on n vertices. Consider the random graph $\mathcal{G}_n(p_n) = G(n, p_n)$, this time with edge probability p_n depending on n . Choose a vertex at random from the random graph and “disconnect” it; that is, delete all of the edges that it possesses. Call the induced measure $P_n^{p_n, \text{disconnect}}$. For which choices of p_n is the tampering detectable, for which choices is it strongly undetectable, and for which choices is it weakly undetectable?

We now consider the general abstract set up. Let $\mathcal{E}_n \equiv 2^{E_n}$ denote the set of all possible subsets of E_n . Of course, for any particular realization of $\mathcal{G}_n(p_n)$, one has $E_n(p_n) \in \mathcal{E}_n$. Thus, we will consider $P_n^{p_n}$ as a probability measure on \mathcal{E}_n . We will refer to each element of \mathcal{E}_n as an edge configuration.

Now let $\{O_{n,j}\}_{j=1}^{m_n}$ be subsets of \mathcal{E}_n of equal $P_n^{p_n}$ -probability, and let $\mathcal{C}_n \subset \mathcal{E}_n$ be defined by $\mathcal{C}_n = \cup_{j=1}^{m_n} O_{n,j}$. Define the tampered random graph to be

the one corresponding to the probability measure on \mathcal{E}_n defined by

$$(1.1) \quad P_n^{p_n, \mathcal{C}_n}(\cdot) \equiv \frac{1}{m_n} \sum_{j=1}^{m_n} P_n^{p_n}(\cdot | O_{n,j}).$$

In the first example above, \mathcal{C}_n is the set of edge configurations which contain at least one clique of order k_n . We have $m_n = \binom{n}{k_n}$, the number of cliques of order k_n in $G(n)$, and numbering these cliques from 1 to m_n , then $O_{n,j}$ is the set of edge configurations which contain the j -th clique. In the second example, \mathcal{C}_n is the set of edge configurations which possess at least one disconnected vertex. We have $m_n = n$, the number of vertices in $G(n)$, and numbering the vertices from 1 to n , then $O_{n,j}$ is the set of edge configurations for which the j -th vertex is disconnected. A little thought should convince the reader that the measures $P_n^{p, \text{clique}, k_n}$ and $P_n^{p_n, \text{disconnect}}$ from the two examples above are in fact the measure $P_n^{p_n, \mathcal{C}_n}$ for the two cases of \mathcal{C}_n and $O_{n,j}$ noted in this paragraph.

Remark. In all of our examples, we will employ the generic notation E_n for the edge set and \mathcal{E}_n for the set of edge configurations. The generic notation for the graph itself will be changed corresponding to the particular example—e.g., $G(n)$ for the complete graph.

Define $N_{\mathcal{G}_n}^{\mathcal{C}_n} : \mathcal{E}_n \rightarrow \{0, 1, 2, \dots, m_n\}$ by

$$(1.2) \quad N_{\mathcal{G}_n}^{\mathcal{C}_n}(\omega) = \sum_{j=1}^{m_n} 1_{O_{n,j}}(\omega).$$

So $N_{\mathcal{G}_n}^{\mathcal{C}_n}(\omega)$ counts the number of $j \in \{1, \dots, m_n\}$ for which the edge configuration $\omega \in \mathcal{E}_n$ belongs to $O_{n,j}$. We can consider $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ as a random variable with respect to either the measure $P_n^{p_n}$ or the measure $P_n^{p_n, \mathcal{C}_n}$. In our first example above, we will denote the corresponding random variable $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ by $N_{G(n)}^{\text{clique}, k_n}$; so

$$N_{G(n)}^{\text{clique}, k_n}(\omega) = \text{the number of cliques of order } k_n \text{ in } \omega \in \mathcal{E}_n.$$

In our second example, we will denote the corresponding random variable $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ by $N_{G(n)}^{\text{disconnect}}$; so

$$N_{G(n)}^{\text{disconnect}}(\omega) = \text{the number of disconnected vertices in } \omega \in \mathcal{E}_n.$$

We have the following simple result.

Proposition 1. *i. If the distribution of $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ under $P_n^{p_n}$ converges to the δ -distribution at 0 ($\lim_{n \rightarrow \infty} P_n^{p_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n} \geq 1) = 0$), then one is in the detectable case;*

ii. If one is in the strongly undetectable case, then the distribution of $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ under $P_n^{p_n}$ converges to the δ -distribution at ∞

$$(\lim_{M \rightarrow \infty} \liminf_{n \rightarrow \infty} P_n^{p_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n} \geq M) = 1).$$

Proof. By construction, $P_n^{p_n, \mathcal{C}_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n} \geq 1) = 1$. Part (i) then follows immediately since

$$\lim_{n \rightarrow \infty} \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{\text{TV}} \geq \lim_{n \rightarrow \infty} |P_n^{p_n, \mathcal{C}_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n} \geq 1) - P_n^{p_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n} \geq 1)| = 1.$$

We could give a simple, direct proof of part (ii), but refrain from doing so since part (ii) follows from Proposition 2 below. \square

Naive intuition would suggest that parts (i) and (ii) of Proposition 1 should be if and only if statements, or at least “almost” if and only if statements, where by “almost” we mean that we are willing to exclude a narrow borderline bifurcation region between the two regimes. But, as was noted in the first paragraph of this paper, neither part (i) nor part (ii) is “almost” an if and only if statement in the context of tampering by adding an increasing subsequence to a random permutation in S_n . Indeed, for increasing sequences of length $k_n = cn^l$ with $l \in (\frac{4}{9}, \frac{1}{2})$, or $k_n = cn^{\frac{1}{2}}$, with $c < 2$, the tampering is detectable and the distribution of $N_{S_n}^{\text{incsubseq}, k_n}$, the number of increasing subsequences of length k_n , converges to the δ -distribution at ∞ . The main thrust of this paper is to bring examples in the context of random graphs in order to examine to what extent the above naive intuition holds.

This will then lead to a discussion of the extent to which the graph structure abets the detection.

Let $E_n^{p_n}$ denote the expectation with respect to $P_n^{p_n}$. Part (i) of the above proposition along with the first moment method allow us to conclude that

$$(1.3) \quad \text{if } \lim_{n \rightarrow \infty} E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n} = 0, \text{ then the tampering is detectable.}$$

The following fundamental proposition, which does give an if and only if statement with regard to strong undetectability, will play a key role.

Proposition 2. *Let $P_n^{p_n}$, $P_n^{p_n, \mathcal{C}_n}$ and $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ be as described above. Then*

$$(1.4) \quad \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{TV} = \frac{1}{2} \sum_{\omega \in \mathcal{E}_n} \left| \frac{N_{\mathcal{G}_n}^{\mathcal{C}_n}(\omega)}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}} - 1 \right| P_n^{p_n}(\omega).$$

Thus, in particular,

$$\lim_{n \rightarrow \infty} \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{TV} = 0$$

if and only if the weak law of numbers holds for $N_{\mathcal{G}_n}^{\mathcal{C}_n}$ under $P_n^{p_n}$; that is, if and only if

$$\lim_{n \rightarrow \infty} P_n^{p_n} \left(\left| \frac{N_{\mathcal{G}_n}^{\mathcal{C}_n}}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}} - 1 \right| > \epsilon \right) = 0, \text{ for all } \epsilon > 0.$$

The second moment method then yields the following corollary. Let Var_{n, p_n} denote the variance with respect to $P_n^{p_n}$. (The notation $f = \Theta(g)$ means that $f = O(g)$ and $g = O(f)$.)

Corollary 1. *i. If $\text{Var}_{n, p_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n}) = o((E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n})^2)$, then*

$\lim_{n \rightarrow \infty} \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{TV} = 0$ *and the tampering is strongly undetectable;*

ii. If $\text{Var}_{n, p_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n}) = \Theta((E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n})^2)$, then $\{\|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{TV}\}_{n=1}^{\infty}$ is bounded away from 1; thus the tampering is not detectable.

Remark. Part (i) of the corollary is of course just Chebyshev's inequality (or Jensen's inequality). We will prove part (ii). Note that Jensen's inequality applied to (1.4) gives

$$(1.5) \quad \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{TV} \leq \frac{1}{2} \frac{\sqrt{\text{Var}_{n, p_n}(N_{\mathcal{G}_n}^{\mathcal{C}_n})}}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}}.$$

In the examples treated in this paper, strong undetectability, when it occurs, will be proven via the second moment method. We will calculate the right hand side of (1.5) rather explicitly, and thus also give information on the rate at which the total variation goes to zero. However, these rates will not be stated in our results.

We now present results for six examples, the first four involving the complete graph $G(n)$ and the last two involving the hypercube. The first two examples on the complete graph were introduced above. Consider first the cliques in the Erdos-Renyi graph $G(n, p)$. Using (1.3) and part (i) of Corollary 1, we will prove the following result. (Let $\log_{\frac{1}{p}}^{(k)}$ denote the k -th iterate of $\log_{\frac{1}{p}}$. Also, let $\sum_{j=2}^1(\cdot) \equiv 0$.)

Theorem 1. *Consider the Erdos-Renyi graph $G(n, p)$ with $p \in (0, 1)$ and tamper with it by adding a random clique of order k_n .*

i. Assume that for some $m \geq 1$ and some $c \in (0, 2)$, one has

$$k_n \geq 2 \log_{\frac{1}{p}} n - 2 \sum_{j=2}^m \log_{\frac{1}{p}}^{(j)} n - c \log_{\frac{1}{p}}^{(m+1)} n.$$

Then the tampering is detectable; furthermore, the distribution of $N_{G(n)}^{\text{clique}, k_n}$ under P_n^p converges to the δ -distribution at 0;

ii. Assume that for some $m \geq 1$ and some $c > 2$, one has

$$k_n \leq 2 \log_{\frac{1}{p}} n - 2 \sum_{j=2}^m \log_{\frac{1}{p}}^{(j)} n - c \log_{\frac{1}{p}}^{(m+1)} n.$$

Then the tampering is strongly undetectable; furthermore, the distribution of $N_{G(n)}^{\text{clique}, k_n}$ under P_n^p satisfies the law of large numbers, so in particular it converges to the δ -distribution at ∞ .

Remark 1. The theorem demonstrates that for this model, both parts of Proposition 1 are at least “almost” if and only if statements.

Remark 2. The first and second moment calculations along with estimates to distinguish between the cases $k_n \geq c \log_{\frac{1}{p}} n$ with $c > 2$ and $k_n \leq c \log_{\frac{1}{p}} n$ with $c < 2$ can be found, for example, in [2, chapter 7]. There the motivation

was to show that in the former case the probability of there being at least one clique of order k_n goes to 0, and that in the latter case it goes to 1. We will present, an analysis of the second moment which seems to us a bit simpler, and we will analyze the two moments more precisely in order to get the tighter dichotomy stated above in the theorem.

Now consider the second example from above—disconnecting a random vertex in the Erdos-Renyi graph $G(n, p_n)$. The calculation and analysis of the first and second moments of $N_{G(n)}^{\text{disconnect}}$ is very easy; see [2, chapter 7]. Using those estimates, we will give an easy proof of the following result.

Theorem 2. *Consider the Erdos-Renyi graph $G(n, p_n)$ and tamper with it by disconnecting a randomly selected edge.*

- i. If $p_n = \frac{\log n + \omega_n}{n}$ with $\lim_{n \rightarrow \infty} \omega_n = \infty$, then the tampering is detectable; furthermore, the distribution of $N_{G(n)}^{\text{disconnect}}$ under $P_n^{p_n}$ converges to the δ -distribution at 0;*
- ii. If $p_n = \frac{\log n + \omega_n}{n}$ with $\lim_{n \rightarrow \infty} \omega_n = -\infty$, then the tampering is strongly undetectable; furthermore, the distribution of $N_{G(n)}^{\text{disconnect}}$ under $P_n^{p_n}$ satisfies the law of large numbers, so in particular it converges to the δ -distribution at ∞ ;*
- iii. If $p_n = \frac{\log n + \omega_n}{n}$ with $\{\omega_n\}_{n=1}^{\infty}$ bounded, then the tampering is weakly undetectable; furthermore, the distribution of $N_{G(n)}^{\text{disconnect}}$ under $P_n^{p_n}$ converges neither to the δ -distribution at ∞ nor to the δ -distribution at 0.*

Remark. The theorem demonstrates that for this model, both parts of Proposition 1 are if and only if statements.

We now consider a third example with the complete graph $G(n)$ and the corresponding Erdos-Renyi graph $G(n, p_n)$. Recall that a *Hamiltonian path* in $G(n)$ is a path that traverses each of the vertices of the graph exactly once; that is, a path of the form $x_1 x_2 \cdots x_n$, where the x_i are all distinct. Choose a Hamiltonian path from $G(n)$ at random and “add” it to $G(n, p_n)$; that is, add to the random graph every edge of this Hamiltonian path that is not already in the random graph. Call the induced measure $P_n^{p_n, \text{Ham}}$. In the notation

of the abstract formulation above, \mathcal{C}_n is the set of edge configurations which contain at least one Hamiltonian path. We have $m_n = \frac{1}{2}n!$, the number of Hamiltonian paths in $G(n)$, and numbering these paths from 1 to m_n , then $O_{n,j}$ is the set of edge configurations which contain the j -th Hamiltonian path. Let

$$N_{G(n)}^{\text{Ham}}(\omega) = \text{the number of Hamiltonian paths in } \omega \in \mathcal{E}_n.$$

Quite sophisticated graph theoretical techniques along with probabilistic analysis have yielded the following beautiful result: if $p_n = \frac{\log n + \log \log n + \omega_n}{n}$, then

$$(1.6) \quad \lim_{n \rightarrow \infty} P_n^{p_n}(N_{G(n)}^{\text{Ham}} \geq 1) = \begin{cases} 1, & \text{if } \lim_{n \rightarrow \infty} \omega_n = \infty; \\ 0, & \text{if } \lim_{n \rightarrow \infty} \omega_n = -\infty. \end{cases}$$

(See [6] and [2, chapter 7 and references].) So from Proposition 1-i, it follows that the tampering is detectable if $p_n = \frac{\log n + \log \log n + \omega_n}{n}$ with $\lim_{n \rightarrow \infty} \omega_n = -\infty$. However, we have $E_n^{p_n} N_{G(n)}^{\text{Ham}} = \frac{1}{2}n!p_n^n$; thus, this expectation converges to ∞ if $p_n \geq \frac{e}{n}$. Consequently, the first moment method fails to give the correct threshold for detectability. As we will show, the second moment method will allow us to conclude that the tampering is strongly undetectable if $\lim_{n \rightarrow \infty} p_n = 1$, but the method will not work otherwise. In light of (1.6), it would seem that the second moment does not give the correct threshold for strong undetectability at all, but in fact it does. We will prove the following result.

Theorem 3. *Consider the Erdos-Renyi graph $G(n, p_n)$ and tamper with it by adding a random Hamiltonian path.*

- i. If $\lim_{n \rightarrow \infty} p_n = 0$, then the tampering is detectable. However, if in addition, $p_n = \frac{\log n + \log \log n + \omega_n}{n}$ with $\lim_{n \rightarrow \infty} \omega_n = \infty$, then the distribution of $N_{G(n)}^{\text{Ham}}$ under $P_n^{p_n}$ converges to the δ -distribution at ∞ ;*
- ii. If $\lim_{n \rightarrow \infty} p_n = 1$, then the tampering is strongly undetectable; furthermore the distribution of $N_{G(n)}^{\text{Ham}}$ under $P_n^{p_n}$ satisfies the law of large numbers, so in particular it converges to the δ -distribution at ∞ ;*

iii. If $p_n \equiv p \in (0, 1)$, then the tampering is weakly undetectable; furthermore, the distribution of $N_{G(n)}^{Ham}$ under P_n^p converges to the δ -distribution at ∞ . However, $N_{G(n)}^{Ham}$ under P_n^p does not satisfy the law of large numbers.

Remark 1. For this model neither part (i) nor part (ii) of Proposition 1 is even almost an if and only if statement. Instead we meet up again with the phenomenon that occurred in the case of increasing subsequences in random permutations. Indeed, for the wide range of p_n satisfying $\lim_{n \rightarrow \infty} p_n = 0$ and $p_n = \frac{\log n + \log \log n + \omega_n}{n}$ with $\lim_{n \rightarrow \infty} \omega_n = \infty$, one has detectability, but the distribution of $N_{G(n)}^{Ham}$ under $P_n^{p_n}$ converges to the δ -distribution at ∞ . Furthermore, for fixed p , the distribution converges to the δ -distribution at ∞ , but the tampering is not strongly undetectable.

Remark 2. In light of (1.6), part (iii) of the theorem seems surprising; however, it has an easy explanation as will be seen in the discussion after Theorem 6. More precisely, when looked at in an appropriate way, a simple central limit theorem argument shows that when $p \in (0, 1)$ is fixed, the total variation norm cannot converge to 0. But then by Proposition 2, it follows that under P_n^p the law of large numbers does not hold for the number of Hamiltonian paths $N_{G(n)}^{ham}$, a fact that is not obvious at all.

In light of part (iii) of Theorem 3, in the case of fixed $p \in (0, 1)$, it is natural to consider tampering with a *sub-Hamiltonian* path of length $k_n < n$; that is, a path of the form $x_1 x_2 \cdots x_{k_n}$, where the x_i are all distinct. Call the induced measure $P_n^{p, \text{subham}, k_n}$. Let

$N_{G(n)}^{\text{subHam}, k_n}(\omega) =$ the number of sub-Hamiltonian paths in $G(n, p)$ of length k_n in $\omega \in \mathcal{E}_n$.

We will prove the following result.

Theorem 4. Consider the Erdos-Renyi graph $G(n, p)$, with $p \in (0, 1)$ and tamper with it by adding a random sub-Hamiltonian path of length k_n .

- i. If $k_n = o(n)$, then the tampering is strongly undetectable; furthermore, the distribution of $N_{G(n)}^{subHam, k_n}$ under P_n^p satisfies the law of large numbers, so in particular it converges to the δ -distribution at ∞ ;*
- ii. If $k_n = \Theta(n)$, then the tampering is weakly undetectable; furthermore, the distribution of $N_{G(n)}^{subHam, k_n}$ under P_n^p converges to the δ -distribution at ∞ .*

Remark. Note that, unsurprisingly, the behavior in part (ii) of the theorem is the same as the behavior in part (iii) of Theorem 3.

We now consider two examples on the n -dimensional hypercube graph H_2^n with the corresponding random hypercube $H_2^n(p_n)$. Recall that the vertices of H_2^n are identified with $\{0, 1\}^n$, and an edge connects two vertices if and only if they differ in exactly one component. Denote vertices by $\bar{x} = (x_1, \dots, x_n)$. We begin with the hypercube analog of the model in Theorem 2. Choose a vertex at random from $H_2^n(p_n)$ and “disconnect” it; that is, delete all the edges it possesses. Call the induced measure $P_n^{p_n, disconnect}$. (This is the same notation as was used in the complete graph analog, but it should not cause confusion.) Let

$$N_{H_2^n}^{disconnect}(\omega) = \text{the number of disconnected vertices in } \omega \in \mathcal{E}_n.$$

Using the first and second moment method, we will give an easy proof of the following theorem.

Theorem 5. *Consider the random hypercube $H_2^n(p_n)$ and tamper with it by disconnecting a randomly selected vertex.*

- i. If $p_n = \frac{1}{2} + \frac{\gamma_n}{n}$, with $\lim_{n \rightarrow \infty} \gamma_n = \infty$, then the tampering is detectable; furthermore, the distribution of $N_{H_2^n}^{disconnect}$ under P_n^p converges to the δ -distribution at 0;*
- ii. If $p = \frac{1}{2} + \frac{\gamma_n}{n}$ with $\lim_{n \rightarrow \infty} \gamma_n = -\infty$, then the tampering is strongly undetectable; furthermore, the distribution of $N_{H_2^n}^{disconnect}$ under P_n^p satisfies the law of large numbers, so in particular, it converges to the δ -distribution at ∞ ;*

iii. If $p = \frac{1}{2} + \frac{\gamma_n}{n}$, with γ_n bounded, then the tampering is weakly undetectable; furthermore, the distribution of $N_{H_2^n}^{\text{disconnect}}$ under P_n^p converges neither to the δ -distribution at ∞ nor to the δ -distribution at 0.

Remark. The theorem demonstrates that for this model, both parts of Proposition 1 are if and only if statements (as was also the case with disconnecting a randomly selected vertex from the Erdos-Renyi graph.)

Now we consider a more challenging model. A geodesic path from \bar{x} to \bar{y} is a shortest path from \bar{x} to \bar{y} . A diameter path in H_2^n is a longest geodesic path in H_2^n . The set of diameter paths is the set of paths $\bar{x}_0\bar{x}_1\cdots\bar{x}_n$, where $\bar{x}_n = \bar{1} - \bar{x}_0$ ($\bar{1} \equiv (1, \dots, 1)$). Now consider the random hypercube $H_2^n(p_n)$. Choose a diameter path from H_2^n at random and “add” it to $H_2^n(p_n)$; that is, add to the random graph every edge of this diameter path that is not already in the random graph. Call the induced measure $P_n^{p_n, \text{diam}}$. In the notation of the abstract formulation above, \mathcal{C}_n is the set of edge configurations which contain at least one diameter path. We have $m_n = 2^{n-1}n!$, the number of diameter paths, and numbering these paths from 1 to m_n , then $O_{n,j}$ is the set of edge configurations which contain the j -th diameter path. Let

$$N_{H_2^n}^{\text{diam}}(\omega) = \text{the number of diameter paths in } \omega \in \mathcal{E}_n.$$

We will prove the following result.

Theorem 6. Consider the random hypercube $H_2^n(p_n)$ and tamper with it by adding a random diameter path.

- i. If $p_n \leq \frac{\gamma}{n}$, with $\gamma < \frac{e}{2}$, then the tampering is detectable; furthermore, the distribution of $N_{H_2^n}^{\text{diam}}$ under $P_n^{p_n}$ converges to the δ -distribution at 0;
- ii. If $p_n \geq \frac{\gamma}{n}$, with $\gamma > \frac{e}{2}$, then the tampering is strongly undetectable; furthermore, the distribution of $N_{H_2^n}^{\text{diam}}$ under $P_n^{p_n}$ satisfies the law of large numbers, so in particular it converges to the δ -distribution at ∞ .

Remark. The theorem demonstrates that for this model, both parts of Proposition 1 are at least “almost” if and only if statements.

We will prove the above theorem via the first and second moment method. The first moment method will give part (i) immediately. The calculations needed to show that the second moment method gives part (ii) may be interesting in their own right and are the most nontrivial ones in this paper.

We now discuss how the structure of the graph has entered into the results in Theorems 1-6. We begin by using the generic notation in order to cover all of the examples at once. Consider the graph $\mathcal{G}_n = (V_n, E_n)$. Label the edges from 1 to $|E_n|$. The random graph $\mathcal{G}_n(p_n) = (V_n, E_n(p_n))$ with probability measure $P_n^{p_n}$ is constructed by considering a collection $\{e_j\}_{j=1}^{|E_n|}$ of IID Bernoulli random variables taking on the values 1 and 0 with respective probabilities p_n and $1 - p_n$, and declaring the j -th edge to exist in $E_n(p_n)$ if and only if $e_j = 1$. Let $N_{\mathcal{G}_n}^{\text{edges}} : \mathcal{E}_n \rightarrow \{0, 1, \dots, |E_n|\}$ count the number of edges present in an edge configuration. The expected value of $N_{\mathcal{G}_n}^{\text{edges}}$ under the measure $P_n^{p_n}$ is $|E_n|p_n$, and the variance is $|E_n|p_n(1 - p_n)$. Now in the tampering models corresponding to Theorems 1, 3, 4 and 6, a certain number l_n of edges were selected and for each such selected index j , if e_j was equal to 0, it was changed to 1. In the tampering models corresponding to Theorems 2 and 5, the roles of 0 and 1 above are interchanged. The expected value of $N_{\mathcal{G}_n}^{\text{edges}}$ under the tampered measure $P_n^{p_n, \mathcal{C}_n}$ is $(|E_n| - l_n)p_n + l_n = |E_n|p_n + (1 - p_n)l_n$ in the cases where 0's were changed to 1's, and is $(|E_n| - l_n)p_n = |E_n|p_n - p_n l_n$ in the cases where 1's were changed to 0's. The change in the mean of $N_{\mathcal{G}_n}^{\text{edges}}$ when using the tampered measure instead of the original one is thus equal either to $(1 - p_n)l_n$ or $-p_n l_n$, depending on whether we are changing 0's to 1's or 1's to 0's. We will denote this change in mean by ΔExp_n . In either case, the variance of $N_{\mathcal{G}_n}^{\text{edges}}$ under the tampered measure is $(|E_n| - l_n)p_n(1 - p_n)$, and under untampered measure is $|E_n|p_n(1 - p_n)$. Let $\text{SD}_n \equiv \sqrt{|E_n|p_n(1 - p_n)}$ denote the standard deviation under the untampered measure. Using the central limit theorem, it is easy to show that if ΔExp_n is on a larger order than

SD_n , then the tampering is detectable. We will prove this later on. In fact, if there were no graph structure at all, and one tampered with the collection of $|E_n|$ IID random variables by selecting at random l_n indices and changing the corresponding random variables as above, then it is easy to show that $SD_n = o(\Delta \text{Exp}_n)$ is necessary and sufficient for detectability, $\Delta \text{Exp}_n = \Theta(SD_n)$ is necessary and sufficient for weak undetectability, and $\Delta \text{Exp}_n = o(SD_n)$ is necessary and sufficient for strong undetectability. The graph structure can abet detectability, but cannot hinder it.

Now let's consider how much the graph structure has abetted detectability in the models in Theorems 1-6. Let's consider first the models in Theorems 1 and 4, where the probability parameter is fixed at p and the tampered structures contain the size parameter k_n . For the model in Theorem 1, where we have the Erdos-Renyi graph $G(n, p)$ and tamper with it by adding a randomly selected clique of order k_n , we have $|E_n| = \frac{1}{2}n(n-1)$ and $l_n = \frac{1}{2}k_n(k_n-1)$. Thus, $\Delta \text{Exp}_n = \frac{1}{2}(1-p)k_n(k_n-1)$, and $SD_n = \sqrt{\frac{1}{2}p(1-p)n(n-1)}$. Consequently, if we had tampered without any reference to the graph structure, and had selected at random $\frac{1}{2}k_n(k_n-1)$ edges for tampering, then the tampering would have been detectable if k_n were on a larger order than $n^{\frac{1}{2}}$, would have been weakly undetectable if k_n were exactly on the order $n^{\frac{1}{2}}$, and would have been strongly undetectable if k_n were on a smaller order than $n^{\frac{1}{2}}$. Theorem 1 demonstrates that the graph structure has abetted the tamper detection very significantly. Indeed, the threshold size for k_n , which distinguishes between detectability and strong nondetectability, was shown to be around $2 \log_{\frac{1}{p}} n$.

Now consider the model in Theorem 4. For this model, where we have the Erdos-Renyi graph $G(n, p)$ and tamper with it by adding a randomly selected sub-Hamiltonian path of length k_n , we have $|E_n| = \frac{1}{2}n(n-1)$ and $l_n = k_n - 1$. Thus, $\Delta \text{Exp}_n = (1-p)(k_n-1)$, and $SD_n = \sqrt{\frac{1}{2}p(1-p)n(n-1)}$. Consequently, if we had tampered without any reference to the graph structure, and had randomly selected $k_n - 1$ edges for tampering, then the tampering would have been strongly undetectable if k_n were on a smaller order than

n , and weakly undetectable if k_n were exactly on the order n . But Theorem 4 gives the same result. Thus, the graph structure has not abetted the tampering at all in the case of sub-Hamiltonian paths!

Now we turn to the models in Theorems 2, 3, 5 and 6 where the tampered structures do not have an external parameter k_n and where the probability parameter p_n is variable. For the model in Theorem 2, we have the Erdos-Renyi graph $G(n, p_n)$, and tamper with it by disconnecting a randomly chosen edge. This tampering involves subtracting edges instead of adding them. We have $l_n = n - 1$ and $|E_n| = \frac{1}{2}n(n - 1)$. Thus, $\Delta\text{Exp}_n = -(n - 1)p_n$, and $\text{SD}_n = \sqrt{\frac{1}{2}p_n(1 - p_n)n(n - 1)}$. Consequently, if we had tampered without any reference to the graph structure, and had selected at random $n - 1$ edges for tampering, then the tampering would have been detectable if $\lim_{n \rightarrow \infty} p_n = 1$, would have been weakly undetectable if p_n remained bounded away from 0 and 1, and would have been strongly undetectable if $\lim_{n \rightarrow \infty} p_n = 0$. Theorem 2 demonstrates that the graph structure has abetted the tampering very significantly. Indeed, the threshold probability that distinguishes between detectability and strong undetectability is around $p_n = \frac{\log n}{n}$.

Now consider the model in Theorem 3. In this model, we tamper with the Erdos-Renyi graph $G(n, p_n)$ by adding a randomly selected Hamiltonian path. We have $|E_n| = \frac{1}{2}n(n - 1)$ and $l_n = n - 1$. Thus, $\Delta\text{Exp}_n = (1 - p_n)(n - 1)$ and $\text{SD}_n = \sqrt{\frac{1}{2}p_n(1 - p_n)n(n - 1)}$. Consequently, if we had tampered without any reference to the graph structure, and had selected at random $n - 1$ edges for tampering, then as in the model of Theorem 2. the tampering would have been detectable if $\lim_{n \rightarrow \infty} p_n = 1$, would have been weakly undetectable if p_n remained bounded away from 0 and 1, and would have been strongly detectable if $\lim_{n \rightarrow \infty} p_n = 0$. But Theorem 3 gives the same result. Thus, as was the case with sub-Hamiltonian paths, the graph structure has not abetted the tampering at all in the case of Hamiltonian paths.

Now consider the model from Theorem 5. In this model we tamper with the random hypercube $H_2^n(p_n)$ by disconnecting a randomly selected vertex. We have $l_n = n$ and $|E_n| = 2^{n-1}n$. Thus, $\Delta\text{Exp}_n = -np_n$, and $\text{SD}_n = \sqrt{p_n(1-p_n)}2^{\frac{n-1}{2}}n^{\frac{1}{2}}$. Consequently, if we had tampered without any reference to the graph structure, and had selected at random n edges for tampering, the tampering would have been detectable if $1-p_n$ were on an order smaller than $\frac{n}{2^n}$, would have been weakly undetectable if $1-p_n$ were exactly on the order $\frac{n}{2^n}$, and would have been strongly undetectable if $1-p_n$ were on a larger order than $\frac{n}{2^n}$. Theorem 5 demonstrates that the graph structure has abetted the tampering in a dramatic fashion. Indeed, the threshold probability that distinguishes between detectability and strong undetectability is around $p_n = \frac{1}{2}$, whereas without the graph structure the threshold occurs with p_n converging to 1 exponentially fast.

Now consider the model from Theorem 6. In this model we tamper with the random hypercube $H_2^n(p_n)$ by adding a randomly selected diameter path. We have $|E_n| = 2^{n-1}n$ and $l_n = n$. Thus, $\Delta\text{Exp}_n = (1-p_n)n$, and $\text{SD}_n = \sqrt{p_n(1-p_n)}2^{\frac{n-1}{2}}n^{\frac{1}{2}}$. Consequently, if we had tampered without any reference to the graph structure, and had selected at random n edges for tampering, then as in the model of Theorem 5, the tampering would have been detectable if $1-p_n$ were on an order smaller than $\frac{n}{2^n}$, would have been weakly undetectable if $1-p_n$ were exactly on the order $\frac{n}{2^n}$, and would have been strongly undetectable if $1-p_n$ were on a larger order than $\frac{n}{2^n}$. Theorem 6 demonstrates that the graph structure has abetted the tampering in an extremely dramatic fashion. Indeed, the threshold probability that distinguishes between detectability and strong undetectability is around $p_n = \frac{e}{2^n}$, whereas without the graph structure the threshold occurs with p_n converging to 1 exponentially fast.

In summary, we have seen that the graph structure does not abet the detection if one tampers with an Erdos-Renyi graph with a random Hamiltonian or sub-Hamiltonian path. The graph structure very significantly

abets detection if one tampers with such a graph by disconnecting a random edge or by adding a random clique. For the random hypercube, the graph structure abets detection dramatically if one tampers by disconnecting a random edge, and abets it even more dramatically if one tampers by adding a random diameter path. We also note that the cases where the graph structure does not abet detection are exactly the ones where the naive intuition fails with regard to parts (i) and (ii) of Proposition 1 being if and only if statements.

We will prove Proposition 2 and part (ii) of Corollary 1 in section 2. Then in sections 3 - 8 we will prove Theorems 1 - 6 respectively.

Note. According to our definitions, a tampering system as above, call it $\{\mathcal{T}_n\}_{n=1}^\infty$, may be neither detectable, strongly undetectable nor weakly undetectable. However, there is always a subsequence $\{\mathcal{T}_{n_k}\}_{k=1}^\infty$ which is either detectable, strongly undetectable or weakly undetectable. In our proofs, we will sometimes prove that the tampering system is not detectable and not strongly undetectable, and then conclude that the system must therefore be weakly undetectable. We are able to make this conclusion because our proofs will show implicitly that in fact no subsystem is detectable or strongly undetectable.

2. PROOF OF PROPOSITION 2 AND COROLLARY 1

Proof of Proposition 2. Let $\omega \in \mathcal{E}_n$. Then we have $P_n^{p_n}(\omega | O_{n,j}) = \frac{1_{\{O_{n,j}\}}(\omega) P_n^{p_n}(\omega)}{P_n^{p_n}(O_{n,j})}$. Also, from (1.2) and the fact that the $O_{n,j}$ have the same $P_n^{p_n}$ -probabilities for all j , we have $E_n^{p_n} N_{\mathcal{G}_n}^{C_n} = m_n P_n^{p_n}(O_{n,1})$. Using these facts along with the definition of $P_n^{p_n, C_n}$ in (1.1) and of $N_{\mathcal{G}_n}^{C_n}$ in (1.2), we have

$$P_n^{p_n, C_n}(\omega) = \frac{1}{m_n} \sum_{j=1}^{m_n} P_n^{p_n}(\omega | O_{n,j}) = \frac{P_n^{p_n}(\omega)}{m_n P_n^{p_n}(O_{n,1})} \sum_{j=1}^{m_n} 1_{\{O_{n,j}\}}(\omega) = \frac{N_{\mathcal{G}_n}^{C_n}}{E_n^{p_n} N_{\mathcal{G}_n}^{C_n}} P_n^{p_n}(\omega).$$

Thus, the Radon-Nikodym derivative of $P_n^{p_n, \mathcal{C}_n}$ with respect to $P_n^{p_n}$ is $\frac{N_{\mathcal{G}_n}^{\mathcal{C}_n}}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}}$. From this, a standard argument gives (1.4), and from (1.4) it follows easily that $\lim_{n \rightarrow \infty} \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{\text{TV}} = 0$ if and only if $\lim_{n \rightarrow \infty} P_n^{p_n}(|\frac{N_{\mathcal{G}_n}^{\mathcal{C}_n}}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}} - 1| > \epsilon) = 0$, for all $\epsilon > 0$. \square

Proof of Corollary 1-ii. Let $Y_n = \frac{N_{\mathcal{G}_n}^{\mathcal{C}_n}}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}}$. An alternative form of (1.4) is

$$\|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{\text{TV}} = \sum_{\omega \in \mathcal{E}_n} (1 - \frac{N_{\mathcal{G}_n}^{\mathcal{C}_n}(\omega)}{E_n^{p_n} N_{\mathcal{G}_n}^{\mathcal{C}_n}})^+ P_n^{p_n}(\omega),$$

where $a^+ = a \vee 0$. From this it follows that $\lim_{n \rightarrow \infty} \|P_n^{p_n, \mathcal{C}_n} - P_n^{p_n}\|_{\text{TV}} = 1$ if and only if $\lim_{n \rightarrow \infty} P_n^{p_n}(Y_n > \epsilon) = 0$, for all $\epsilon > 0$. By the assumption in part (ii) of the corollary, $E_n^{p_n} Y_n^2 \leq M$ for some M and all n . For every $\epsilon > 0$, we have

$$\begin{aligned} 1 = E_n^{p_n} Y_n &\leq \epsilon + E_n^{p_n} Y_n 1_{Y_n > \epsilon} \leq \epsilon + (P_n^{p_n}(Y_n > \epsilon))^{\frac{1}{2}} (E_n^{p_n} Y_n^2)^{\frac{1}{2}} \leq \\ &\epsilon + (M P_n^{p_n}(Y_n > \epsilon))^{\frac{1}{2}}. \end{aligned}$$

From this it is not possible that $\lim_{n \rightarrow \infty} P_n^{p_n}(Y_n > \epsilon) = 0$, if $\epsilon < 1$. \square

3. PROOF OF THEOREM 1

The number of cliques of order k_n is $\binom{n}{k_n}$, and each such clique contains $\binom{k_n}{2}$ edges. Thus,

$$(3.1) \quad E_n^p N_{G(n)}^{\text{clique}, k_n} = \binom{n}{k_n} p^{\binom{k_n}{2}}.$$

For $k_n = o(n^{\frac{1}{2}})$, we have

$$\binom{n}{k_n} p^{\binom{k_n}{2}} \sim \frac{n^{k_n}}{k_n!} p^{\frac{k_n(k_n-1)}{2}} \sim \frac{n^{k_n} p^{\frac{k_n(k_n-1)}{2}}}{k_n^{k_n} e^{-k_n} \sqrt{2\pi k_n}},$$

and thus

$$(3.2) \quad \begin{aligned} \log \binom{n}{k_n} p^{\binom{k_n}{2}} &\sim \\ k_n \log_{\frac{1}{p}} n - \frac{1}{2} k_n^2 + \frac{1}{2} k_n - k_n \log_{\frac{1}{p}} k_n + k_n \log_{\frac{1}{p}} e - \frac{1}{2} \log_{\frac{1}{p}} 2\pi k_n. \end{aligned}$$

It is easy to verify that as $n \rightarrow \infty$, the right hand side of (3.2) converges to $-\infty$ if $k_n \geq 2 \log_{\frac{1}{p}} n - c \log_{\frac{1}{p}}^{(2)} n$ with $c < 2$. In particular, it then follows from (3.1) that $\lim_{n \rightarrow \infty} E_n^p N_{G(n,p)}^{\text{clique}, k_n} = 0$, and thus from (1.3) we obtain detectability. It is easy to continue the above analysis in order to demonstrate the same conclusion if for some $m \geq 2$ and some $c \in (0, 2)$, one has $k_n \geq 2 \log_{\frac{1}{p}} n - 2 \sum_{j=1}^m \log_{\frac{1}{p}}^{(j)} n - c \log_{\frac{1}{p}}^{(m+1)} n$. Thus, we obtain part (i) of the theorem.

We now turn to part (ii). Identify the vertices in $G(n)$ by $[n]$. Recall that we have labeled the cliques of order k_n from 1 up to $m_n = \binom{n}{k_n}$, and have defined $O_{n,j}$ to be the collection of edge configurations which contain the j -th clique. We relabel for convenience, letting $O_{j_1, \dots, j_{k_n}}$ denote the collection of edge configurations which contain the clique that involves the vertices $\{j_1, \dots, j_{k_n}\}$. Since $N_{G(n)}^{\text{clique}, k_n} = \sum_{1 \leq j_1 < \dots < j_{k_n} \leq n} 1_{O_{j_1, \dots, j_{k_n}}}$, we have

$$(3.3) \quad E_n^p (N_{G(n)}^{\text{clique}, k_n})^2 = \sum_{1 \leq i_1 < \dots < i_{k_n} \leq n; 1 \leq j_1 < \dots < j_{k_n} \leq n} P_n^p(O_{i_1, \dots, i_{k_n}} \cap O_{j_1, \dots, j_{k_n}}).$$

By symmetry considerations,

$$(3.4) \quad \sum_{1 \leq i_1 < \dots < i_{k_n} \leq n; 1 \leq j_1 < \dots < j_{k_n} \leq n} P_n^p(O_{i_1, \dots, i_{k_n}} \cap O_{j_1, \dots, j_{k_n}}) = \binom{n}{k_n} \sum_{1 \leq j_1 < \dots < j_{k_n} \leq n} P_n^p(O_{1, \dots, k_n} \cap O_{j_1, \dots, j_{k_n}}).$$

Let $l = |[k_n] \cap \{j_1, \dots, j_{k_n}\}|$ denote the number of vertices shared by O_{1, \dots, k_n} and $O_{j_1, \dots, j_{k_n}}$. If $l \geq 2$, then $P_n^p(O_{1, \dots, k_n} \cap O_{j_1, \dots, j_{k_n}}) = p^{2 \binom{k_n}{2} - \binom{l}{2}}$. If $l \leq 1$, then the above formula holds with $\binom{l}{2}$ replaced by 0. Letting the generic E denote the expectation with respect to the uniform measure on k_n -tuples from $[n]$, and defining the random variable W_{n, k_n} on such k_n -tuples by $W_{n, k_n}(j_1, \dots, j_{k_n}) = |[k_n] \cap \{j_1, \dots, j_{k_n}\}|$, it then follows from (3.3) and (3.4) that

$$(3.5) \quad E_n^p (N_{G(n)}^{\text{clique}, k_n})^2 = \binom{n}{k_n}^2 p^{2 \binom{k_n}{2}} \left(E(p^{-\binom{W_{n, k_n}}{2}}; W_{n, k_n} \geq 2) + P(W_{n, k_n} \leq 1) \right).$$

Now $\text{Var}_{n,p}(N_{G(n)}^{\text{clique},k_n}) = o((E_n^p N_{G(n)}^{\text{clique},k_n})^2)$ will follow from (3.1) and (3.5) if we show that

$$(3.6) \quad \lim_{n \rightarrow \infty} E(p^{-\frac{1}{2}} W_{n,k_n}^2; W_{n,k_n} \geq 2) = 0.$$

We have

$$P(W_{n,k_n} = l) = \frac{\binom{k_n}{l} \binom{n-k_n}{k_n-l}}{\binom{n}{k_n}}, \quad l = 0, 1, \dots, k_n.$$

Since $k_n = o(n^{\frac{1}{2}})$, one has $\binom{n}{k_n} \sim \frac{n^{k_n}}{k_n!}$. Also, of course, $\binom{n-k_n}{k_n-l} \leq \frac{n^{k_n-l}}{(k_n-l)!}$ and $\binom{k_n}{l} \leq \frac{k_n^l}{l!}$. Thus, for some $C > 0$,

$$(3.7) \quad P(W_{n,k_n} = l) \leq C \frac{n^{k_n-l} k_n^l k_n!}{(k_n-l)! l! n^{k_n}} \leq \frac{C k_n^{2l}}{n^{l!}}.$$

Using (3.7) we have for some $C_1 > 0$,

$$E(p^{-\frac{1}{2}} W_{n,k_n}^2; W_{n,k_n} \geq 2) \leq C \sum_{l=2}^{k_n} p^{-\frac{l^2}{2}} \frac{k_n^{2l}}{n^{l!}} \leq C_1 \sum_{l=2}^{k_n} \left(\frac{e k_n^2}{l n p^{\frac{1}{2}}} \right)^l.$$

It is easy to check that if k_n is sufficiently large, then $\min_{2 \leq l \leq k_n} l p^{\frac{l}{2}} = k_n p^{\frac{k_n}{2}}$. Thus, from above we have

$$(3.8) \quad E(p^{-\frac{1}{2}} W_{n,k_n}^2; W_{n,k_n} \geq 2) \leq C_1 \sum_{l=2}^{k_n} \left(\frac{e k_n^2}{l n p^{\frac{1}{2}}} \right)^l.$$

Writing $\log_{\frac{1}{p}} \frac{k_n}{n p^{\frac{k_n}{2}}} = \log_{\frac{1}{p}} k_n - \log_{\frac{1}{p}} n + \frac{1}{2} k_n$, it is easy to see that this expression converges to $-\infty$ as $n \rightarrow \infty$, if $k_n \leq 2 \log_{\frac{1}{p}} n - c \log_{\frac{1}{p}}^{(2)} n$, with $c > 2$. Thus for such values of k_n , we have $\lim_{n \rightarrow \infty} \frac{e k_n^2}{l n p^{\frac{k_n}{2}}} = 0$, and it follows from (3.8) that (3.6) holds. It is easy to show that this analysis can be continued in order to demonstrate the same conclusion if for some $m \geq 2$ and some $c > 2$, one has $k_n \leq 2 \log_{\frac{1}{p}} n - 2 \sum_{j=1}^m \log_{\frac{1}{p}}^{(j)} n - c \log_{\frac{1}{p}}^{(m+1)} n$. \square

4. PROOF OF THEOREM 2

Completely straight forward computations in [2, chapter 7] reveal that if $p_n = \frac{\log n + \omega_n}{n}$, then

$$(4.1) \quad \lim_{n \rightarrow \infty} E_n^{p_n} N_{G(n)}^{\text{disconnect}} = \infty \text{ if } \lim_{n \rightarrow \infty} \omega_n = -\infty;$$

$$(4.2) \quad \lim_{n \rightarrow \infty} E_n^{p_n} N_{G(n)}^{\text{disconnect}} = 0 \text{ if } \lim_{n \rightarrow \infty} \omega_n = \infty;$$

$$(4.3) \quad \begin{aligned} & \{E_n^{p_n} N_{G(n)}^{\text{disconnect}}\}_{n=1}^{\infty} \text{ is bounded away from } 0 \text{ and } \infty \\ & \text{if } \{\omega_n\}_{n=1}^{\infty} \text{ is bounded;} \end{aligned}$$

$$(4.4) \quad \begin{aligned} & \text{Var}_{n,p_n}(N_{G(n)}^{\text{disconnect}}) = o((E_n^{p_n} N_{G(n)}^{\text{disconnect}})^2) \text{ if} \\ & \lim_{n \rightarrow \infty} \omega_n = -\infty; \end{aligned}$$

$$(4.5) \quad \begin{aligned} & \text{Var}_{n,p_n}(N_{G(n)}^{\text{disconnect}}) = \Theta((E_n^{p_n} N_{G(n)}^{\text{disconnect}})^2) \text{ if} \\ & \{\omega_n\}_{n=1}^{\infty} \text{ is bounded.} \end{aligned}$$

Thus, part (i) of the theorem follows from (4.2) and (1.3), and part (ii) of the theorem follows from (4.4) and part (i) of Corollary 1. We now prove part (iii). By (4.5) and part (ii) of Corollary 1, the tampering cannot be detectable. By (4.3) and part (ii) of Proposition 1, the tampering cannot be strongly undetectable. Thus, the tampering is weakly undetectable. \square

5. PROOF OF THEOREM 3

For part (i), let $N_{G(n)}^{\text{edges}} : \mathcal{E}_n \rightarrow \{0, 1, \dots, |\mathcal{E}_n|\}$ denote the number of edges in an edge configuration, as was defined at the beginning of the discussion at the end of section 1. As was noted in that discussion, for the complete graph $G(n)$, we have $|E_n| = \frac{1}{2}n(n-1)$, and we have $l_n = n-1$, since a Hamiltonian path has $n-1$ edges. The expected value of $N_{G(n)}^{\text{edges}}$ under $P_n^{p_n}$ is $\frac{1}{2}n(n-1)p_n$ and the variance is $\frac{1}{2}n(n-1)p_n(1-p_n)$. The expected value of $N_{G(n)}^{\text{edges}}$ under the tampered measure $P_n^{p_n, \text{Ham}}$ is $\frac{1}{2}n(n-1)p_n + (1-p_n)(n-1)$,

and the variance is $(\frac{1}{2}n(n-1) - (n-1))p_n(1-p_n)$ which is less than the above variance. Let $\epsilon \in (0, 1)$. Define the events

$$A_{n,\epsilon} = \{|N_{G(n)}^{\text{edges}} - \frac{1}{2}n(n-1)p_n| \geq \frac{1}{\sqrt{2\epsilon}}\sqrt{n(n-1)p_n(1-p_n)}\},$$

$$B_{n,\epsilon} = \{|N_{G(n)}^{\text{edges}} - \frac{1}{2}n(n-1)p_n - (1-p_n)(n-1)| \geq \frac{1}{\sqrt{2\epsilon}}\sqrt{n(n-1)p_n(1-p_n)}\}.$$

Then, by Chebyshev's inequality,

$$(5.1) \quad P_n^{p_n}(A_{n,\epsilon}) \leq \epsilon \text{ and } P_n^{p_n, \text{Ham}}(B_{n,\epsilon}) \leq \epsilon.$$

If $\lim_{n \rightarrow \infty} p_n = 0$, then for sufficiently large n , $A_{n,\epsilon}^c$ and $B_{n,\epsilon}^c$ are disjoint. From this and (5.1), it follows that $\liminf_{n \rightarrow \infty} \|P_n^{p_n} - P_n^{p_n, \text{Ham}}\|_{\text{TV}} \geq 1 - 2\epsilon$. Since ϵ is arbitrary, we conclude that the tampering is detectable. To complete the proof of part (i), we need to show that if $p_n = \frac{\log n + \log \log n + \omega_n}{n}$, with $\lim_{n \rightarrow \infty} \omega_n = \infty$, then the distribution of $N_{G(n)}^{\text{ham}}$ under $P_n^{p_n}$ converges to the δ -distribution at ∞ . For this range of p_n , it is known that not only does the probability of there being at least one Hamiltonian path converge to 1 as $n \rightarrow \infty$, but also the probability of there being at least one Hamiltonian cycle [2]. Any Hamiltonian cycle can be cut open in n possible locations, thereby yielding n Hamiltonian paths.

We now turn to parts (ii) and (iii). We will prove that

$$(5.2) \quad \begin{aligned} \text{Var}_{n,p_n}(N_{G(n)}^{\text{Ham}}) &= o((E_n^{p_n} N_{G(n)}^{\text{Ham}})^2), \text{ if } \lim_{n \rightarrow \infty} p_n = 1; \\ \text{Var}_{n,p}(N_{G(n)}^{\text{Ham}}) &= \Theta((E_n^p N_{G(n)}^{\text{Ham}})^2), \text{ if } p \in (0, 1). \end{aligned}$$

By Corollary 1, this will prove part (ii) and it will show that in part (iii), the tampering is not detectable. We now show that for part (iii), the tampering is not strongly undetectable, thereby proving that the tampering is weakly undetectable. By the central limit theorem,

$$\lim_{n \rightarrow \infty} P_n^{p, \text{Ham}}(N_{G(n)}^{\text{edges}} \geq \frac{1}{2}n(n-1) + (1-p)(n-1)) = \frac{1}{2},$$

and

$$\lim_{n \rightarrow \infty} P_n^p(N_{G(n)}^{\text{edges}} \geq \frac{1}{2}n(n-1) + (1-p)(n-1)) = 1 - \Phi\left(\left(\frac{2(1-p)}{p}\right)^{\frac{1}{2}}\right) < \frac{1}{2}.$$

It remains to prove (5.2).

Identify vertices in $G(n)$ by $[n]$. There is a two-to-one correspondence between S_n , the permutations of $[n]$, and the set of Hamiltonian paths (two-to-one since the Hamiltonian paths are not oriented). The permutation $\sigma \in S_n$ will correspond to the Hamiltonian path $x_{\sigma_1} \cdots, x_{\sigma_n}$. Thus, as was noted in section 1, we have

$$(5.3) \quad E_n^{p_n} N_{G(n)}^{\text{Ham}} = \frac{1}{2} n! p_n^{n-1}.$$

Recall that we have labeled the Hamiltonian paths from 1 up to $m_n = \frac{1}{2}n!$, and have defined $O_{n,j}$ to be the set of edge configurations which contain the j -th Hamiltonian path. We relabel for convenience. Let $O_{n,\sigma}$ denote the set of edge configurations which contain the Hamiltonian path corresponding to the permutation $\sigma \in S_n$. Then, taking into account the two-to-one correspondence, we have $N_{G(n)}^{\text{Ham}} = \frac{1}{2} \sum_{\sigma \in S_n} 1_{O_{n,\sigma}}$. So

$$(5.4) \quad E_n^{p_n} (N_{G(n)}^{\text{Ham}})^2 = \frac{1}{4} \sum_{\sigma, \tau \in S_n} P_n^{p_n}(O_\sigma \cap O_\tau).$$

By symmetry considerations, letting id denote the identity permutation,

$$(5.5) \quad \sum_{\sigma, \tau \in S_n} P_n^{p_n}(O_\sigma \cap O_\tau) = n! \sum_{\sigma \in S_n} P_n^{p_n}(O_\sigma \cap O_{\text{id}}).$$

An adjacent pair in a permutation $\sigma \in S_n$ is a pair $(j, j+1)$ such that for some i one has $\{\sigma_i, \sigma_{i+1}\} = \{j, j+1\}$. Let $W_n(\sigma)$ denote the number of adjacent pairs in $\sigma \in S_n$. Note that the Hamiltonian path corresponding to σ has $W_n(\sigma)$ edges in common with the Hamiltonian path corresponding to id . Thus, $P_n^{p_n}(O_\sigma \cap O_{\text{id}}) = p_n^{2n-2-W_n(\sigma)}$. Letting the generic E denote the expectation with respect to the uniform measure on S_n , it then follows from (5.4), (5.5) and (5.3) that

$$(5.6) \quad E_n^{p_n} (N_{G(n)}^{\text{Ham}})^2 = (E_n^{p_n} N_{G(n)}^{\text{Ham}})^2 E p_n^{-W_n}.$$

From (5.6), the first line of (5.2) will hold if we show that

$$(5.7) \quad \lim_{n \rightarrow \infty} E p_n^{-W_n} = 1, \text{ if } \lim_{n \rightarrow \infty} p_n = 1,$$

and the second line of (5.2) will hold if we show that

$$(5.8) \quad \limsup_{n \rightarrow \infty} E p^{-W_n} < \infty, \text{ for } p \in (0, 1).$$

The distribution of W_n can be found in [5]; it converges to the Poisson distribution with parameter 2. Let $W'_n(\sigma)$ denote the number of successions in $\sigma \in S_n$, where a succession is an adjacent pair that is in the correct order; that is, a pair $(j, j+1)$ such that for some i one has $(\sigma_i, \sigma_{i+1}) = (j, j+1)$. The distribution of W'_n of course converges to the Poisson distribution with parameter 1, and in [4] it is shown that the exponential moments of W'_n converge to those of the Poisson distribution. A similar calculation shows that this also holds for W_n . In light of this, (5.7) and (5.8) clearly hold. This completes the proof of the theorem. \square

6. PROOF OF THEOREM 4

Part (ii) essentially follows from part (iii) of Theorem 3. Indeed, the central limit theorem argument there shows that the tampering cannot be strongly undetectable. Also, it is clear that for sub-Hamiltonian paths, the tampering is no more detectable than it was for full Hamiltonian paths; thus, the tampering cannot be detectable. We conclude that the tampering is weakly undetectable.

We now turn to part (i). By Corollary 1, it suffices to show that

$$(6.1) \quad \text{Var}_{n,p}(N_{G(n)}^{\text{subHam},k_n}) = o((E_n^{p_n} N_{G(n)}^{\text{subHam},k_n})^2).$$

Let $S_{n;k_n}$ denote the space of permutations of k_n out of n . That is, an element $\sigma \in S_{n;k_n}$ is obtained by choosing in order k_n elements from $[n]$. We will write $\sigma = x_1 \cdots x_{k_n}$. There is a two-to-one correspondence between $S_{n;k_n}$ and the set of sub-Hamiltonian paths of length k_n . We choose a random sub-Hamiltonian path of length k_n by constructing a random element in $S_{n;k_n}$ as follows. One by one, choose k_n points from $[n]$. If the

points x_1, \dots, x_{k_n} were chosen in that order, then we associate the sub-Hamiltonian path $x_1 \cdots x_{k_n}$ to that choice. (The same notation has been used for elements of $S_{n;k_n}$ and sub-Hamiltonian paths of length k_n .)

An adjacent pair for an element $\sigma = x_1 \cdots x_{k_n} \in S_{n;k_n}$ is a pair $(j, j+1)$ such that for some $l \in [k_n - 1]$, one has $\{x_l, x_{l+1}\} = \{j, j+1\}$. When this happens we will say that an adjacent pair occurs at the location $x_l x_{l+1}$. Define $W_{n;k_n}(\sigma)$ to be the number of adjacent pairs in σ . Let the generic E denote the expectation with respect to the uniform measure on $S_{n;k_n}$. The same type of calculation used in the proof of Theorem 3 shows that

$$(6.2) \quad E_n^{p_n} (N_{G(n)}^{\text{subHam}, k_n})^2 = (E_n^p N_{G(n)}^{\text{subHam}, k_n})^2 E p^{-W_{n;k_n}}.$$

Thus, (6.1) will hold if

$$(6.3) \quad \lim_{n \rightarrow \infty} E(p^{-W_{n;k_n}}; W_{n;k_n} \geq 1) = 0.$$

We now estimate $P(W_{n;k_n} \geq m)$, for $1 \leq m < k_n$, where P denotes the probability corresponding to the expectation E . Let $l_1 < \dots < l_m$, with $\{l_r\}_{r=1}^m \subset [k_n - 1]$, and let $A_{l_1, \dots, l_m} \subset S_{n;k_n}$ denote those elements of $S_{n;k_n}$ for which adjacent pairs occur at the m locations $x_{l_1} x_{l_1+1}, \dots, x_{l_m} x_{l_m+1}$. (Note that there can be overlaps in locations; that is, for example, one might have $x_{l_1+1} = x_{l_2}$.) Recall from above how we generated a random Hamiltonian path; namely by selecting k_n points, one by one, from $[n]$. There are $n(n-1) \cdots (n-k_n+1)$ ways of making such a selection. Now we derive an upper bound on the number of ways which will yield an element which belongs to A_{l_1, \dots, l_m} . Furthermore, the upper bound will be independent of the particular choice of $\{l_r\}_{r=1}^m$. We select the k_n points, one by one. For $l \notin \{l_i + 1\}_{i=1}^m$, there are at most $n-l+1$ ways to select the l -th point, and for $l = l_i + 1$, for some i , there are at most 2 ways to select the l -th point. Thus, $n(n-1) \cdots (n-k_n+m+1)2^m$ constitutes an upper bound on the number of ways which yield an element belonging to A_{l_1, \dots, l_m} . Since there

are $\binom{k_n}{m}$ different possible m -tuples $\{l_r\}_{r=1}^m$, we conclude that

$$(6.4) \quad P(W_{n;k_n} \geq m) \leq \binom{k_n}{m} \frac{2^m}{(n - k_n + 1) \cdots (n - k_n + m)}.$$

Since by assumption, $k_n = o(n)$, there exists a $C > 0$ such that $(n - k_n + 1) \cdots (n - k_n + m) \geq \frac{1}{C} \left(\frac{n}{2}\right)^m$. Thus, from (6.4) we have

$$(6.5) \quad P(W_{n;k_n} \geq m) \leq C \frac{\left(\frac{4k_n}{n}\right)^m}{m!}.$$

Now (6.5) gives

$$(6.6) \quad E(p^{-W_{n;k_n}}; W_{n;k_n} \geq 1) \leq C \sum_{m=1}^{k_n-1} \frac{\left(\frac{4k_n}{n}\right)^m}{m!}.$$

Since $k_n = o(n)$, it follows from (6.6) that (6.3) holds, and consequently, that (6.1) holds. \square

7. PROOF OF THEOREM 5

Write p_n in the form $p_n = \frac{1}{2} + \frac{\gamma_n}{n}$. The probability that a given edge in $H_2^n(p_n)$ is disconnected is $(1 - p_n)^n$. Since $|H_2^n| = 2^n$, it follows that

$$(7.1) \quad E_n^{p_n} N_{H_2^n}^{\text{disconnect}} = 2^n (1 - p_n)^n = \left(1 - \frac{2\gamma_n}{n}\right)^n.$$

Thus, if $\lim_{n \rightarrow \infty} \gamma_n = \infty$, then $\lim_{n \rightarrow \infty} E_n^{p_n} N_{H_2^n}^{\text{disconnect}} = 0$, and part (i) of the theorem follows from (1.3).

We now turn to parts (ii) and (iii). We may assume now that p_n is bounded away from 1. For $\bar{x} \in H_2^n$, let $D_{\bar{x}}$ denote the event that \bar{x} is disconnected. Then writing $N_{H_2^n}^{\text{disconnect}} = \sum_{\bar{x} \in H_2^n} 1_{D_{\bar{x}}}$, and using symmetry, we have

$$(7.2) \quad E_n^{p_n} (N_{H_2^n}^{\text{disconnect}})^2 = E_n^{p_n} \sum_{\bar{x}, \bar{y} \in H_2^n} 1_{D_{\bar{x}}} 1_{D_{\bar{y}}} = 2^n \sum_{\bar{x} \in H_2^n} P_n^{p_n}(D_{\bar{x}} \cap D_{\bar{0}}).$$

If \bar{x} is a neighbor of $\bar{0}$, then $P_n^{p_n}(D_{\bar{x}} \cap D_{\bar{0}}) = (1 - p_n)^{2n-1}$; if $\bar{x} = \bar{0}$, then $P_n^{p_n}(D_{\bar{x}} \cap D_{\bar{0}}) = (1 - p_n)^n$; otherwise $P_n^{p_n}(D_{\bar{x}} \cap D_{\bar{0}}) = (1 - p_n)^{2n}$. Since $\bar{0}$ has n neighbors, this gives

$$(7.3) \quad \sum_{\bar{x} \in H_2^n} P_n^{p_n}(D_{\bar{x}} \cap D_{\bar{0}}) = (2^n - n - 1)(1 - p_n)^{2n} + n(1 - p_n)^{2n-1} + (1 - p_n)^n.$$

Now (7.1)-(7.3) give

$$(7.4) \quad \text{Var}_{n,p_n}(N_{H_2^n}^{\text{disconnect}}) = (E_n^{p_n} N_{H_2^n}^{\text{disconnect}})^2 \left(\frac{n}{2^n(1-p_n)} + \frac{1}{2^n(1-p_n)^n} - \frac{n+1}{2^n} \right).$$

We have $2^n(1-p_n)^n = (1 - \frac{2\gamma_n}{n})^n$. Thus, if $\lim_{n \rightarrow \infty} \gamma_n = -\infty$, then from (7.4) we have $\text{Var}_{n,p_n}(N_{H_2^n}^{\text{disconnect}}) = o((E_n^{p_n} N_{H_2^n}^{\text{disconnect}})^2)$, and from part (i) of Corollary 1 the tampering is strongly undetectable. This proves part (ii). If $\{\gamma_n\}_{n=1}^\infty$ is bounded, then from (7.4) we have $\text{Var}_{n,p_n}(N_{H_2^n}^{\text{disconnect}}) = \Theta((E_n^{p_n} N_{H_2^n}^{\text{disconnect}})^2)$. Thus, by part (ii) of Corollary 1, the tampering cannot be detectable. By (7.1) and part (ii) of Proposition 1, the tampering cannot be strongly undetectable. Thus, the tampering is weakly undetectable. This proves part (iii). \square

8. PROOF OF THEOREM 6

There is a two-to-one correspondence between $H_2^n \times S_n$ and diameter paths in H_2^n . Indeed, for $\bar{x} \in H_2^n$ and $\sigma \in S_n$, we begin the diameter path at \bar{x} and use the permutation σ to determine the order in which we change the components of \bar{x} . (The correspondence is two to one because the diameter path is not oriented.) In particular there are $2^{n-1}n!$ diameter paths. The probability that any particular diameter path is contained in the random hypercube $H_2^n(p_n)$ is p_n^n ; thus we have

$$(8.1) \quad E_n^{p_n} N_{H_2^n}^{\text{diam}} = 2^{n-1}n!p_n^n.$$

From this it follows that $\lim_{n \rightarrow \infty} E_n^{p_n} N_{H_2^n}^{\text{diam}} = 0$, if $p_n \leq \frac{\gamma}{n}$, with $\gamma < \frac{e}{2}$. Using this with (1.3) shows that the tampering is detectable, and proves part (i).

We now turn to part (ii). The diameter paths are labeled from 1 to $m_n = 2^{n-1}n!$, and we have defined $O_{n,j}$ to be the set of edge configurations which contain the j -th diameter edge. We relabel for convenience. Let $O_{\bar{x},\sigma}$ denote the set of edge configurations which contain the diameter path corresponding to (\bar{x}, σ) in the above two-to-one correspondence. Then we

have $N_{H_2^n}^{\text{diam}} = \frac{1}{2} \sum_{\bar{x} \in H_2^n, \sigma \in S_n} 1_{O_{\bar{x}, \sigma}}$. Thus

$$(8.2) \quad E_n^{p_n} (N_{H_2^n}^{\text{diam}})^2 = \frac{1}{4} \sum_{\bar{x}, \bar{y} \in H_2^n, \sigma, \tau \in S_n} P_n^{p_n} (O_{\bar{x}, \sigma} \cap O_{\bar{y}, \tau}).$$

By symmetry considerations, letting id denote the identity permutation and letting $\bar{0} \in H_2^n$ denote the element with zeroes in all of its coordinates, we have

$$(8.3) \quad \sum_{\bar{x}, \bar{y} \in H_2^n, \sigma, \tau \in S_n} P_n^{p_n} (O_{\bar{x}, \sigma} \cap O_{\bar{y}, \tau}) = 2^n n! \sum_{\bar{x} \in H_2^n, \sigma \in S_n} P_n^{p_n} (O_{\bar{x}, \sigma} \cap O_{\bar{0}, \text{id}}).$$

Let $W_n(\bar{x}, \sigma)$ denote the number of edges that the diameter path corresponding to (\bar{x}, σ) has in common with the diameter path corresponding to $(\bar{0}, \text{id})$.

Then we have

$$(8.4) \quad P_n^{p_n} (O_{\bar{x}, \sigma} \cap O_{\bar{0}, \text{id}}) = p_n^{2n - W_n(\bar{x}, \sigma)}.$$

Letting the generic E denote the expectation with respect to the uniform measure on $H_2^n \times S_n$, it then follows from (8.1)-(8.4) that

$$(8.5) \quad E_n^{p_n} (N_{H_2^n}^{\text{diam}})^2 = (E_n^{p_n} N_{H_2^n}^{\text{diam}})^2 E p_n^{-W_n}.$$

Thus, if we show that

$$(8.6) \quad \lim_{n \rightarrow \infty} E(p_n^{-W_n}; W_n \geq 1) = 0,$$

then it will follow from (8.5) that $\text{Var}_{n, p_n} (N_{H_2^n}^{\text{diam}}) = o((E_n^{p_n} N_{H_2^n}^{\text{diam}})^2)$, and thus by part (i) of Corollary 1, the tampering will be strongly undetectable.

We now estimate $P(W_n \geq m)$, for $m \geq 1$, where P denotes the probability corresponding to the expectation E . In fact, in the quite involved estimate that follows, it will be convenient to assume that $m \geq 2$; one can show that the estimate obtained below in (8.15) also holds for $m = 1$. The diameter path $(\bar{0}, \text{id})$ has n edges, which we label e_1, e_2, \dots, e_n , with e_1 being the edge connecting $\bar{0}$ to $(1, 0 \dots, 0)$, e_2 being the edge connecting $(1, 0 \dots, 0)$ to $(1, 1, 0 \dots, 0)$, etc. Let $A_{l_1, \dots, l_m} \subset H_2^n \times S_n$ denote those diameter paths

which contain the edges e_{l_1}, \dots, e_{l_m} . Then

$$(8.7) \quad P(W_n \geq m) \leq \sum_{1 \leq l_1 < l_2 < \dots < l_m \leq n} P(A_{l_1, \dots, l_m}).$$

We now estimate $P(A_{l_1, \dots, l_m})$.

We first determine for which $\sigma \in S_n$ one has that $(\bar{0}, \sigma) \in A_{l_1, \dots, l_m}$; this result will be needed for the general case of determining which (\bar{x}, σ) belong to A_{l_1, \dots, l_m} . We will say that $[j]$ is a sub-permutation of σ if σ maps $[j]$ onto itself. A moment's thought reveals that the edge e_j belongs to the diameter path $(\bar{0}, \sigma)$ if and only if both $[j-1]$ and $[j]$ are sub-permutations for σ . Thus, $(\bar{0}, \sigma) \in A_{l_1, \dots, l_m}$ if and only if $[l_1-1], [l_1], [l_2-1], [l_2], \dots, [l_m-1]$ and $[l_m]$ are all sub-permutations of σ . The number of permutations $\sigma \in S_n$ for which this holds is easily seen to be $(l_1-1)!(l_2-1-l_1)! \dots (l_m-1-l_{m-1})!(n-l_m)!$. Let $T_{m;l_1, \dots, l_m}^n \subset S_n$ denote those permutations for which $[l_1-1], [l_1], [l_2-1], [l_2], \dots, [l_m-1]$ and $[l_m]$ are all sub-permutations. So we have

$$(8.8) \quad |T_{m;l_1, \dots, l_m}^n| = (l_1-1)!(l_2-1-l_1)! \dots (l_m-1-l_{m-1})!(n-l_m)!.$$

We now consider when $(\bar{x}, \sigma) \in A_{l_1, \dots, l_m}$ for general \bar{x} . It is not hard to see that a necessary condition for $(\bar{x}, \sigma) \in A_{l_1, \dots, l_m}$ is that either $\bar{x} = (x_1, \dots, x_n)$ satisfies $x_j = 0$, for all $l_1 \leq j \leq l_m$, or $x_j = 1$, for all $l_1 \leq j \leq l_m$. We will refer to these two conditions on \bar{x} by $K_{0;l_1, l_m}$ and $K_{1;l_1, l_m}$.

If one of these two conditions on \bar{x} is satisfied, then in order to have $(\bar{x}, \sigma) \in A_{l_1, \dots, l_m}$, the following conditions are required on σ . Recall that σ gives the order in which the n coordinates of \bar{x} are changed so that the diameter path moves from \bar{x} to $\bar{1} - \bar{x}$. So if $\sigma = (\sigma_1, \dots, \sigma_n)$, then the σ_j -th edge in the diameter path will involve changing the σ_j -th coordinate. Let $B_{0;l_1}(\bar{x})$ denote those $j \in \{1, \dots, l_1-1\}$ for which $x_j = 0$, and let $C_{1;l_m}(\bar{x})$ denote those $j \in \{l_m+1, \dots, n\}$ for which $x_j = 1$ ($B_{0;1}(\bar{x}), C_{0;m}(\bar{x}) = \emptyset$). Let $r_{l_1, l_m}(\bar{x}) = |B_{0;l_1}(\bar{x})| + |C_{1;l_m}(\bar{x})|$. Then it is not hard to see that the first $r_{l_1, l_m}(\bar{x})$ coordinates in σ must be reserved for $B_{0;l_1}(\bar{x}) \cup C_{1;l_m}(\bar{x})$; that is, $\{\sigma_1, \dots, \sigma_{r_{l_1, l_m}(\bar{x})}\} = B_{0;l_1}(\bar{x}) \cup C_{1;l_m}(\bar{x})$. Let $x^{1;j}$ denote the vertex in

H_2^n whose first j components are 1 and whose remaining components are 0. Of course, this vertex belongs to the diameter path $(\bar{0}, \text{id})$. If σ is as above, then the $r_{l_1, l_m}(\bar{x})$ -th vertex of the diameter path (\bar{x}, σ) will be $x^{1; l_1 - 1}(\bar{x})$ if \bar{x} satisfies condition $K_{0; l_1, l_m}$, and will be $x^{1; l_m}$ if \bar{x} satisfies condition $K_{1; l_1, l_m}$. In the former case, we must then have $\sigma_{r_{l_1, l_m}(\bar{x})+1} = l_1$, and in the latter case, we must then have $\sigma_{r_{l_1, l_m}(\bar{x})+1} = l_m$. In the former case, the $(r_{l_1, l_m}(\bar{x})+1)$ -th vertex of the diameter path (\bar{x}, σ) will be $x^{1; l_1}(\bar{x})$ and the $r_{l_1, l_m}(\bar{x})$ -th edge will be e_{l_1} , and in the latter case, the $(r_{l_1, l_m}(\bar{x})+1)$ -th vertex of the diameter path (\bar{x}, σ) will be $x^{1; l_m - 1}(\bar{x})$ and the $r_{l_1, l_m}(\bar{x})$ -th edge will be e_{l_m} . (Recall that a diameter path has $n + 1$ vertices.)

If \bar{x} satisfies condition $K_{0; l_1, l_m}$, then the next $l_m - l_1$ coordinates of σ must involve the numbers $(l_1 + 1, l_1 + 2, \dots, l_m)$, and must move the diameter path (\bar{x}, σ) from the vertex $x^{1; l_1}(\bar{x})$ to the vertex $x^{1; l_m}(\bar{x})$ while passing through the edges e_{l_2}, \dots, e_{l_m} . Based on our analysis above, for this to happen one requires that $(\sigma_{l_1+1} - l_1, \sigma_{l_1+2} - l_1, \dots, \sigma_{l_m} - l_1)$ belong to $T_{m-2; l_2 - l_1, \dots, l_{m-1} - l_1}^{l_m - l_1} \subset S_{l_m - l_1}$. Similarly, if \bar{x} satisfies condition $K_{1; l_1, l_m}$, then the next $l_m - l_1$ coordinates of σ must involve the numbers $(l_1 + 1, l_1 + 2, \dots, l_m)$, and must move the diameter path (\bar{x}, σ) from the vertex $x^{1; l_m}(\bar{x})$ to the vertex $x^{1; l_1}(\bar{x})$ while passing through the edges $e_{l_{m-1}}, \dots, e_{l_1}$. Inverting the direction of our analysis above, for this to happen one requires that $(\sigma_{l_1+1} - l_1, \sigma_{l_1+2} - l_1, \dots, \sigma_{l_m} - l_1)$ belong to $T_{m-2; l_m - l_{m-1}, \dots, l_m - l_2}^{l_m - l_1} \subset S_{l_m - l_1}$. Then finally, the last $n - r_{l_1, l_m}(\bar{x}) - 1 - (l_m - l_1)$ coordinates of σ can be chosen arbitrarily from the remaining numbers. Putting the above all together, we obtain

(8.9)

$$\begin{aligned}
P(A_{l_1, \dots, l_m}) = & \\
& \frac{1}{2^n n!} \sum_{c=0}^{n-l_m} \sum_{b=0}^{l_1-1} \binom{n-l_m}{c} \binom{l_1-1}{b} (b+c)! (n-b-c-1-(l_m-l_1))! \times \\
& (|T_{m-2; l_2-l_1, \dots, l_{m-1}-l_1}^{l_m-l_1}| + |T_{m-2; l_m-l_{m-1}, \dots, l_m-l_2}^{l_m-l_1}|).
\end{aligned}$$

(Given that \bar{x} satisfies condition $K_{0;l_1,l_m}$ or condition $K_{1;l_1,l_m}$, there are $\binom{n-l_m}{c} \binom{l_1-1}{b}$ ways to choose \bar{x} so that $b = |B_{0;l_1}(\bar{x})|$ and $c = |C_{1;l_m}(\bar{x})|$. And given this, there are $(b+c)!(n-b-c-1-(l_m-l_1))!(|T_{m-2;l_2-l_1,\dots,l_{m-1}-l_1}^{l_m-l_1}|$ ways to choose σ if condition $K_{0;l_1,l_m}$ was satisfied, and $(b+c)!(n-b-c-1-(l_m-l_1))!(|T_{m-2;l_m-l_{m-1},\dots,l_m-l_2}^{l_m-l_1}|$ ways to choose σ if condition $K_{1;l_1,l_m}$ was satisfied.)

We have

$$(8.10) \quad \begin{aligned} & \sum_{c=0}^{n-l_m} \sum_{b=0}^{l_1-1} \binom{n-l_m}{c} \binom{l_1-1}{b} (b+c)!(n-b-c-1-(l_m-l_1))! = \\ & \sum_{c=0}^{n-l_m} \sum_{b=0}^{l_1-1} \frac{\binom{n-l_m}{c} \binom{l_1-1}{b}}{\binom{n-1-l_m+l_1}{b+c}} (n-1-(l_m-l_1))! \leq n^2(n-1-(l_m-l_1))!, \end{aligned}$$

where the last inequality follows from the fact that the fraction in the sum above is always less than 1. After completing the current proof, we will prove the following proposition.

Proposition 3. *For every $\delta > 0$, there exist a $c_\delta > 0$ and an $r_\delta \geq 0$ such that $|T_{m;l_1,\dots,l_m}^n| \equiv (l_1-1)!(l_2-1-l_1)! \cdots (l_m-1-l_{m-1})!(n-l_m)!$ satisfies*

$$\sum_{1 \leq l_1 < l_2 < \cdots < l_m \leq n} |T_{m;l_1,\dots,l_m}^n| \leq c_\delta m^{r_\delta} (1+\delta)^m (n-m)!, \quad 1 \leq m \leq n < \infty.$$

From Proposition 3, it follows that for any $\delta > 0$, there exists a $c_\delta > 0$ and an $r_\delta \geq 0$ such that

$$(8.11) \quad \begin{aligned} & \sum_{l_1 < l_2 < \cdots < l_{m-1} < l_m} (|T_{m-2;l_2-l_1,\dots,l_{m-1}-l_1}^{l_m-l_1}| + |T_{m-2;l_m-l_{m-1},\dots,l_m-l_2}^{l_m-l_1}|) \leq \\ & 2c_\delta m^{r_\delta} (1+\delta)^m (l_m-l_1-m+2)!. \end{aligned}$$

(Note that in the sum above, the last subscript, $l_{m-1}-l_1$ in $T_{m-2;l_2-l_1,\dots,l_{m-1}-l_1}^{l_m-l_1}$ and l_m-l_2 in $T_{m-2;l_m-l_{m-1},\dots,l_m-l_2}^{l_m-l_1}$, is strictly less than the superscript l_m-l_1 , whereas in the sum in Proposition 3 the last subscript, l_m in $T_{m;l_1,\dots,l_m}^n$, can attain the value n of the superscript; however, this is no problem since the

inequality goes in the right direction.) Now (8.9), (8.10) and (8.11) give

$$(8.12) \quad \sum_{l_1 < l_2 < \dots < l_{m-1} < l_m} P(A_{l_1, \dots, l_m}) \leq \frac{2c_\delta m^{r_\delta} n^2}{2^n} (1 + \delta)^m \frac{(n-1-(l_m-l_1))!(l_m-l_1-m+2)!}{n!}.$$

Now summing over l_1 and l_m , and denoting $k = l_m - l_1 + 1$, we have

$$(8.13) \quad \sum_{1 \leq l_1 < l_2 < \dots < l_{m-1} < l_m \leq n} P(A_{l_1, \dots, l_m}) \leq \frac{2c_\delta m^{r_\delta} n^3}{2^n} (1 + \delta)^m \sum_{k=m}^n \frac{1}{\binom{n}{k}} \frac{(k-m+1)!}{k!}.$$

Let $\rho(k) \equiv \frac{1}{\binom{n}{k}} \frac{(k-m+1)!}{k!} = \frac{(k-m+1)!}{n(n-1)\dots(n-k+1)}$, $m \leq k \leq n$, and let $h(k) = \frac{\rho(k+1)}{\rho(k)}$. It is easy to check that h is increasing, which implies that ρ is convex. Thus, ρ attains its maximum at an endpoint. We conclude that the maximum of $\rho(k)$ is $\rho(n) = \frac{(n-m+1)!}{n!}$. Using Stirling's formula, it is easy to check that there exists a K such that $\frac{(n-m+1)!}{n!} \leq K(\frac{e}{n})^{m-1}$. Using these facts in (8.13), we obtain

$$(8.14) \quad \sum_{1 \leq l_1 < l_2 < \dots < l_{m-1} < l_m \leq n} P(A_{l_1, \dots, l_m}) \leq \frac{2Kc_\delta m^{r_\delta} n^5}{2^n e} \left(\frac{(1+\delta)e}{n}\right)^m.$$

Using (8.14) in (8.7) now gives

$$(8.15) \quad P(W_n \geq m) \leq \frac{2Kc_\delta m^{r_\delta} n^5}{2^n e} \left(\frac{(1+\delta)e}{n}\right)^m.$$

Thus, if $p_n = \frac{\gamma}{n}$, then from (8.15) we have

$$(8.16) \quad E(p_n^{-W_n}; W \geq 1) \leq \sum_{m=1}^n \left(\frac{n}{\gamma}\right)^m P(W_n \geq m) \leq \frac{2Kc_\delta n^{r_\delta+5}}{2^n e} \sum_{m=1}^n \left(\frac{(1+\delta)e}{\gamma}\right)^m.$$

We may choose $\delta > 0$ as small as we like in (8.16). For $\gamma > \frac{e}{2}$, choose δ so that $\frac{(1+\delta)e}{\gamma} < 2$. Then it follows from (8.16) that (8.6) holds for $p_n = \frac{\gamma}{n}$ with $\gamma > \frac{e}{2}$. \square

We now return to prove Proposition 3.

Proof of Proposition 3. Define

$$C_j^{(0)} = \sum_{i=0}^j \frac{1}{\binom{j}{i}}, \quad j \geq 0,$$

and then define by induction the iterates

$$C_j^{(m)} = \sum_{i=0}^j \frac{C_i^{(m-1)}}{\binom{j}{i}}, \quad j \geq 0, m \geq 1.$$

We have

$$(8.17) \quad \sum_{1 \leq l_1 < l_2} (l_1 - 1)!(l_2 - 1 - l_1)! = (l_2 - 2)! \sum_{l_1=1}^{l_2-1} \frac{1}{\binom{l_2-2}{l_1-1}} = (l_2 - 2)! C_{l_2-2}^{(0)},$$

and then using (8.17),

$$\begin{aligned} & \sum_{1 \leq l_1 < l_2 < l_3} (l_1 - 1)!(l_2 - 1 - l_1)!(l_3 - 1 - l_2)! = \\ & \sum_{2 \leq l_2 < l_3} (l_2 - 2)! C_{l_2-2}^{(0)} (l_3 - 1 - l_2)! = (l_3 - 3)! \sum_{l_2=2}^{l_3-1} \frac{C_{l_2-2}^{(0)}}{\binom{l_3-3}{l_2-2}} = (l_3 - 3)! C_{l_3-3}^{(1)}. \end{aligned}$$

Continuing in this vein, we obtain

$$\sum_{1 \leq l_1 < l_2 < \dots < l_m} (l_1 - 1)!(l_2 - 1 - l_1)! \dots (l_m - 1 - l_{m-1})! = (l_m - m)! C_{l_m-m}^{(m-2)},$$

and

$$(8.18) \quad \sum_{1 \leq l_1 < l_2 < \dots < l_m < n} (l_1 - 1)!(l_2 - 1 - l_1)! \dots (l_m - 1 - l_{m-1})!(n - l_m)! = (n - m)! C_{n-m}^{(m-1)}.$$

In light of (8.18), to complete the proof of Proposition 3, it suffices to show that for every $\delta > 0$, there exist a $c_\delta > 0$ and an $r_\delta \geq 0$ such that

$$(8.19) \quad \sup_{n \geq 1} C_n^{(k)} \leq c_\delta k^{r_\delta} (1 + \delta)^k, \quad k \geq 1.$$

Let $n_0 \geq 1$, and for $n > n_0$ write

$$(8.20) \quad C_n^{(k)} = \sum_{i=0}^{n_0} \frac{C_i^{(k-1)}}{\binom{n}{i}} + \sum_{i=n_0+1}^n \frac{C_i^{(k-1)}}{\binom{n}{i}}, \quad k \geq 1, n > n_0.$$

We need the following lemma whose proof we defer until the completion of the proof of the proposition.

Lemma 1. *For each n there exists a constant c_n such that*

$$(8.21) \quad C_n^{(k)} \leq c_n k^n, \quad k \geq 1.$$

From (8.20) and (8.21), it follows that for each n_0 there exists a constant γ_{n_0} such that

$$(8.22) \quad C_n^{(k)} \leq \gamma_{n_0}(k-1)^{n_0} + \sum_{i=n_0+1}^n \frac{C_i^{(k-1)}}{\binom{n}{i}}, \quad k \geq 1, n > n_0.$$

Let

$$(8.23) \quad d_{n_0} \equiv \sup_{n \geq n_0+1} \sum_{i=n_0+1}^n \frac{1}{\binom{n}{i}}.$$

It is easy to see that

$$(8.24) \quad \lim_{n_0 \rightarrow \infty} d_{n_0} = 1.$$

Letting

$$A_{n_0}^{(k)} \equiv \sup_{n > n_0} C_n^{(k)},$$

we have from (8.22) and (8.23) that

$$(8.25) \quad A_{n_0}^{(k)} \leq \gamma_{n_0}(k-1)^{n_0} + d_{n_0} A_{n_0}^{(k-1)}, \quad k \geq 1.$$

It is not hard to show that

$$\sup_{n \geq 0} C_n^{(0)} = \sup_{n \geq 0} \sum_{i=0}^n \frac{1}{\binom{n}{i}} = \frac{8}{3};$$

however, all we need for our purposes is that this quantity is bounded, and this is very easy to see. Thus, we have

$$(8.26) \quad A_{n_0}^{(0)} \leq \frac{8}{3}.$$

It is easy to show that if $\{x_j\}_{j=0}^k$ satisfies the recursive inequalities $x_0 \leq \frac{8}{3}$ and $x_j \leq C + d_{n_0}x_{j-1}$, for $1 \leq j \leq k$, then $x_k \leq C(1 + d_{n_0} + \cdots + d_{n_0}^{k-1} + \frac{8}{3}d_{n_0}^k) = C(\frac{d_{n_0}^k - 1}{d_{n_0} - 1} + \frac{8}{3}d_{n_0}^k)$. Applying this with $C = \gamma_{n_0}(k-1)^{n_0}$, it follows from (8.25) and (8.26) that

$$(8.27) \quad \sup_{n > n_0} C_n^{(k)} = A_{n_0}^{(k)} \leq \gamma_{n_0}(k-1)^{n_0} \left(\frac{d_{n_0}^k - 1}{d_{n_0} - 1} + \frac{8}{3}d_{n_0}^k \right), \quad k \geq 1.$$

By (8.24), for any $\delta > 0$, there exists an n_0 such that $d_{n_0} \leq 1 + \delta$. Using this with (8.27), and using (8.21) with $n \leq n_0$, one concludes that (8.19) holds. This completes the proof of the proposition. \square

We now return to prove Lemma 1.

Proof of Lemma 1. Fix $n \geq 1$. Let B denote the $n \times n$ matrix with entries b_{ij} , $1 \leq i, j \leq n$, given by $b_{ij} = \frac{1}{\binom{i}{j}}$, for $i \geq j$, and $b_{ij} = 0$, for $j > i$. Let v^0 denote the n -vector with entries v_j^0 , $1 \leq j \leq n$, given by $v_j^0 = C_j^{(0)} = \sum_{l=0}^j \frac{1}{\binom{j}{l}}$. Then from the recursive definition of the $\{C_m^{(k)}\}_{m,k=0}^\infty$, it follows that

$$(8.28) \quad C_n^{(k)} = (B^k v^0)_n, \quad k \geq 1,$$

where $(B^k v^0)_n$ denotes the n -th coordinate of the n -vector $B^k v^0$. Since B is lower triangular with all ones on the diagonal, it follows that there exist vectors v^1, \dots, v^{n-1} such that $Bv^0 = v^0 + v^1, Bv^1 = v^1 + v^2, \dots, Bv^{n-2} = v^{n-2} + v^{n-1}$ and $Bv^{n-1} = v^{n-1}$. From this, it follows that

$$(8.29) \quad B^k v^0 = \sum_{l=0}^{k \wedge n} \binom{k}{l} v^l.$$

Thus, from (8.28) and (8.29), we obtain

$$(8.30) \quad C_n^{(k)} = \sum_{l=0}^{k \wedge n} \binom{k}{l} v_n^l.$$

where v_n^l is the n -th coordinate of v^l . The lemma follows immediately from (8.30). \square

REFERENCES

- [1] Baik, J., Deift, P. and Johansson, K. *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. **12** (1999), 1119-1178.
- [2] Bollabás, B., *Modern Graph Theory*, Graduate Texts in Mathematics, 184, Springer-Verlag (1998).
- [3] Logan, B. F. and Shepp, L. A. *A variational problem for random Young tableaux*, Advances in Math. **26** (1977), 206-222.
- [4] Dwass, M., *The number of increases in a random permutation*, J. Combinatorial Theory Ser. A **15** (1973), 192-199.
- [5] Kaplansky, I., *The asymptotic distribution of runs of consecutive elements*, Ann. Math. Statistics **16** (1945), 200-203.

- [6] Komlós, J. and Szemerédi, E., *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. 43 (1983), 55-63.
- [7] Pinsky, R., *Law of large numbers for increasing subsequences of random permutations*, Random Structures Algorithms **29** (2006), 277-295.
- [8] Pinsky, R. G., *When the law of large numbers fails for increasing subsequences of random permutations*, Ann. Probab. **35** (2007), 758-772.
- [9] Vershik, A. M. and Kerov, S. V. *Asymptotic behavior of the maximum and generic dimensions of irreducible representations of the symmetric group*, Functional Anal. Appl. **19** (1985), 21-31.

DEPARTMENT OF MATHEMATICS, TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY,
HAIFA, 32000, ISRAEL

E-mail address: `pinsky@math.technion.ac.il`

URL: <http://www.math.technion.ac.il/~pinsky/>