

A NOTE ON ESTIMATING GLOBAL SUBGRAPH COUNTS BY SAMPLING

SVANTE JANSON AND VALENTAS KURAUŠKAS

ABSTRACT. We give a simple proof of a generalization of an inequality for homomorphism counts by Sidorenko [11]. A special case of our inequality says that if d_v denotes the degree of a vertex v in a graph G and $\text{hom}_\Delta(H, G)$ denotes the number of homomorphisms from a connected graph H on h vertices to G which map a particular vertex of H to a vertex v in G with $d_v \geq \Delta$, then

$$\text{hom}_\Delta(H, G) \leq \sum_{v \in G} d_v^{h-1} \mathbf{1}_{d_v \geq \Delta}.$$

We use this inequality to study the minimum sample size needed to estimate the number of copies of H in G by sampling vertices of G at random.

1. INTRODUCTION

This paper consists of two main parts. In Section 2 we present a simple proof of an inequality that generalizes Sidorenko's inequality on homomorphism counts. Our motivation for this result comes from an application to estimating global subgraph counts by sampling, which is discussed in Section 3.

Acknowledgement. Much of this research was done during the 28th Nordic Congress of Mathematicians at Aalto University, Helsinki, Finland in August 2022, where the authors met. We thank the organizers for making this possible.

2. THE INEQUALITY

Let H and G be graphs. We let $\text{Hom}(H, G)$ denote the set of homomorphisms $H \rightarrow G$, and $\text{hom}(H, G) := |\text{Hom}(H, G)|$ the number of them.

Sidorenko [11] proved the following theorem:

Theorem 1 (Sidorenko, 1994). *For any connected graph H on $h \geq 1$ vertices and any graph G ,*

$$\text{hom}(H, G) \leq \text{hom}(K_{1, h-1}, G). \quad (2.1)$$

In fact, Sidorenko [11] showed this for trees H , but this is immediately equivalent to our formulation, since $\text{hom}(H, G) \leq \text{hom}(T, G)$ for any spanning tree T of H .

If H is a rooted graph, with root o , and $\Delta \geq 0$, we also define

$$\text{hom}_\Delta(H, G) := |\{\varphi \in \text{Hom}(H, G) : d_{\varphi(o)} \geq \Delta\}|, \quad (2.2)$$

where d_v here and below denotes the degree of a vertex v in a graph. (The graph will be clear from the context; in this section it is always G .) We show the following extension of Theorem 1.

Date: 20 October, 2022.

SJ Supported by the Knut and Alice Wallenberg Foundation.

Theorem 2. *For any connected rooted graph H on $h \geq 1$ vertices, any graph G , and any $\Delta \geq 0$,*

$$\text{hom}_\Delta(H, G) \leq \sum_{v \in G} d_v^{h-1} \mathbf{1}_{d_v \geq \Delta}. \quad (2.3)$$

Note that Sidorenko's theorem is the special case $\Delta = 0$ of our theorem.

We will use induction to prove a more general statement. Let $\alpha := (\alpha_w)_{w \in H}$ be a vector of non-negative real numbers α_w indexed by the vertices in H , and let

$$\text{hom}_{\Delta, \alpha}(H, G) := \sum_{\varphi \in \text{Hom}(H, G)} \mathbf{1}_{\{d_{\varphi(o)} \geq \Delta\}} \prod_{w \in H} d_{\varphi(w)}^{\alpha_w}. \quad (2.4)$$

In particular, taking all $\alpha_w = 0$, we have $\text{hom}_{\Delta, \mathbf{0}}(H, G) = \text{hom}_\Delta(H, G)$. Hence, Theorem 2 is a special case of the following result.

Theorem 3. *For any connected rooted graph H on $h \geq 1$ vertices, any graph G , any $\Delta \geq 0$, and any non-negative vector $\alpha = (\alpha_w)_{w \in H}$,*

$$\text{hom}_{\Delta, \alpha}(H, G) \leq \sum_{v \in G} d_v^{h-1+|\alpha|} \mathbf{1}_{d_v \geq \Delta}, \quad (2.5)$$

where

$$|\alpha| := \sum_{w \in H} \alpha_w. \quad (2.6)$$

Proof. To prove (2.5), we use a double induction over the number of vertices h in H and the number of non-root vertices w such that the weight $\alpha_w > 0$. Hence, we may assume that (2.5) holds if we replace the pair (H, α) by a pair (H', α') such that either

- (i) H' has fewer vertices than H , or
- (ii) H' has the same number of vertices as H , but there are fewer non-root vertices $w \in H'$ with $\alpha'_w > 0$ than non-root $w \in H$ with $\alpha_w > 0$.

The base case $h = 1$ is trivial, since in this case (2.5) is an identity. To prove the induction step, we consider three cases separately.

Case 1: H has a leaf $w \neq o$ with $\alpha_w = 0$. Let v be the unique neighbour of w in H . Define $H' := H \setminus \{w\}$, and let $\alpha'_v := \alpha_v + 1$, and $\alpha'_u := \alpha_u$ for all other $u \in H'$. Then $\text{hom}_{\Delta, \alpha}(H, G) = \text{hom}_{\Delta, \alpha'}(H', G)$, and thus (2.5) follows by the induction hypothesis, since H' has one vertex less than H .

Case 2: H has (at least) two (distinct) non-roots v and w with $\alpha_v, \alpha_w > 0$. Here we use Hölder's inequality, in a way that is essentially the same as in Sidorenko [11] (although he does it in a more general way).

By decomposing the sum in (2.4) according to the values of $\varphi(v)$ and $\varphi(w)$, we obtain

$$\text{hom}_{\Delta, \alpha}(H, G) = \sum_{x, y \in G} \mu_{x, y} d_x^{\alpha_v} d_y^{\alpha_w}, \quad (2.7)$$

for some numbers $\mu_{x, y} \geq 0$ that do not depend on α_v and α_w . We regard the numbers $\mu_{x, y}$ as a measure μ on the finite set $V(G) \times V(G)$, and rewrite (2.7) as

$$\text{hom}_{\Delta, \alpha}(H, G) = \iint_{V(G) \times V(G)} d_x^{\alpha_v} d_y^{\alpha_w} d\mu(x, y). \quad (2.8)$$

Hölder's inequality now yields

$$\begin{aligned} & \text{hom}_{\Delta, \alpha}(H, G) \\ & \leq \left(\iint_{V(G) \times V(G)} d_x^{\alpha_v + \alpha_w} d\mu(x, y) \right)^{\frac{\alpha_v}{\alpha_v + \alpha_w}} \left(\iint_{V(G) \times V(G)} d_y^{\alpha_v + \alpha_w} d\mu(x, y) \right)^{\frac{\alpha_w}{\alpha_v + \alpha_w}} \\ & = \text{hom}_{\Delta, \alpha'}(H, G)^{\frac{\alpha_v}{\alpha_v + \alpha_w}} \text{hom}_{\Delta, \alpha''}(H, G)^{\frac{\alpha_w}{\alpha_v + \alpha_w}}, \end{aligned} \quad (2.9)$$

where

$$\alpha'_v := \alpha_v + \alpha_w, \quad \alpha''_v := 0, \quad (2.10)$$

$$\alpha'_w := 0, \quad \alpha''_w := \alpha_v + \alpha_w, \quad (2.11)$$

$$\alpha'_u := \alpha''_u := \alpha_u \quad \text{for } u \notin \{v, w\}. \quad (2.12)$$

Hence (2.5) follows from the induction hypothesis, since both α' and α'' have one less non-root vertex with positive weight than α .

Case 3: The remaining case. If none of the cases above applies, then every non-root leaf has positive weight, and there is at most one non-root vertex with positive weight. In particular, there is at most one non-root leaf. If also $|V(H)| \geq 2$, then H must have exactly one non-root leaf, say v , and thus H is a path with end vertices o and v . Furthermore, only v and (possibly) the root o can have positive weight. Thus

$$\text{hom}_{\Delta, \alpha}(H, G) = \sum_{\varphi \in \text{Hom}(H, G)} \mathbf{1}_{\{d_{\varphi(o)} \geq \Delta\}} d_{\varphi(o)}^{\alpha_o} d_{\varphi(v)}^{\alpha_v}. \quad (2.13)$$

In this case, we use a variant of an argument that has been used to show other inequalities (see, e.g., [7, Theorems 43 and 236] and [6, Theorem 2.4]). We write, for $x \in G$,

$$f(x) := d_x^{\alpha_o} \mathbf{1}_{d_x \geq \Delta}, \quad g(x) := d_x^{\alpha_v}. \quad (2.14)$$

Then both $f(x)$ and $g(x)$ are (weakly) increasing functions of d_x , and thus, for all $x, y \in G$,

$$(f(x) - f(y))(g(x) - g(y)) \geq 0. \quad (2.15)$$

Consequently, using also the symmetry of H interchanging o and v ,

$$\begin{aligned} 0 & \leq \sum_{\varphi \in \text{Hom}(H, G)} (f(\varphi(o)) - f(\varphi(v)))(g(\varphi(o)) - g(\varphi(v))) \\ & = 2 \sum_{\varphi \in \text{Hom}(H, G)} f(\varphi(o))g(\varphi(o)) - 2 \sum_{\varphi \in \text{Hom}(H, G)} f(\varphi(o))g(\varphi(v)) \\ & = 2 \text{hom}_{\Delta, \alpha'}(H, G) - 2 \text{hom}_{\Delta, \alpha}(H, G), \end{aligned} \quad (2.16)$$

where

$$\alpha'_o := \alpha_o + \alpha_v, \quad (2.17)$$

$$\alpha'_w := 0 \quad \text{for } w \neq o. \quad (2.18)$$

Thus $\text{hom}_{\Delta, \alpha}(H, G) \leq \text{hom}_{\Delta, \alpha'}(H, G)$, and thus (2.5) follows by induction, since α' has one less non-root vertex with positive weight than α . \square

Proof of Theorem 2. As mentioned above, this is the special case $\alpha = 0$ of Theorem 3. \square

3. APPLICATIONS

Let now H be a fixed connected graph on h vertices and G an arbitrary (large) graph on n vertices. Let $\text{Emb}(H, G)$ denote the set of embeddings (injective homomorphisms) $H \rightarrow G$; we will be interested in estimating the number $\text{emb}(H, G) := |\text{Emb}(H, G)|$ by sampling a rather small number of vertices of G and exploring small neighbourhoods of them.

A similar problem for sequences of graphs with a weak local limit has been studied in [8]; there a uniform integrability condition on the $(h-1)$ th power of the random vertex degree was used. Uniform integrability of graph degrees or their powers is natural for sequences of graphs, and has been used both in theoretical work and in applications, see, e.g., [1; 2]. In our setting, we use instead the related (3.3) below. The general problem of estimating small subgraph counts in a given graph has been considered by many authors for a variety of applications, see, e.g., the survey paper [10].

To estimate $\text{emb}(H, G)$, we may assume that H is a rooted graph with a root o (o can be chosen arbitrarily). For a vertex $v \in G$, let $X(v) = X(H, G, v)$ denote the number of embeddings $\sigma \in \text{Emb}(H, G)$ such that $\sigma(o) = v$. We may then estimate $\text{emb}(H, G)$ from the numbers $X(v_i^*)$ for some randomly sampled vertices v_i^* in G . However, since vertices of high degree in G may give outliers with exceptionally high numbers of such embeddings, we use truncation in order to obtain our error bounds.

Choose an arbitrary rooted spanning tree T of H with the same root o . Say that a vertex $u \in T$ is *internal* if $d_u^T > 1$ or $u = o$ (where d_u^T denotes the degree of u in T). Denote by i_T the number of internal vertices in T . Choose also a positive integer Δ . For a vertex $v \in G$, let $\bar{X}(v) = \bar{X}(H, G, v, T, \Delta)$ denote the number of embeddings $\sigma \in \text{Emb}(H, G)$ such that $\sigma(o) = v$ and $d_{\sigma(u)} < \Delta$ for all internal vertices $u \in T$.

Let $N \geq 1$ and let v_1^*, \dots, v_N^* be drawn from $V(G)$ independently and uniformly at random. Consider the following estimate for $n^{-1} \text{emb}(H, G)$:

$$\hat{X}_N := \frac{1}{N} (\bar{X}(v_1^*) + \dots + \bar{X}(v_N^*)). \quad (3.1)$$

Note that the random variable $\bar{X}(v_1^*)$ is bounded, since for every $v \in G$,

$$0 \leq \bar{X}(v) \leq (\Delta - 1)^{h-1}. \quad (3.2)$$

Its mean can be estimated by \hat{X}_N , with an error that can be bounded using, for example, Hoeffding's bound, which we do to prove the following result.

Theorem 4. *Let H be a connected graph on $h \geq 1$ vertices with a rooted spanning tree T , let G be a graph on $n \geq 1$ vertices, and let D be the degree of a uniformly random vertex in G . Suppose a positive integer Δ and a non-negative λ satisfy*

$$\mathbb{E} [D^{h-1} \mathbf{1}_{D \geq \Delta}] \leq \lambda. \quad (3.3)$$

Let $s > 0$ and $p \in (0, 1]$. If

$$N \geq \frac{(\Delta - 1)^{2(h-1)}}{2s^2} \ln \frac{2}{p}, \quad (3.4)$$

then

$$\mathbb{P}(\hat{X}_N - s \leq n^{-1} \text{emb}(H, G) \leq \hat{X}_N + s + i_T \lambda) \geq 1 - p. \quad (3.5)$$

Proof. We have

$$\text{emb}(H, G) = \sum_{v \in V(G)} \bar{X}(v) + \left| \bigcup_u \{ \sigma \in \text{Emb}(H, G) : d_{\sigma(u)} \geq \Delta \} \right|, \quad (3.6)$$

where the union is over the internal vertices of T .

The first term on the right is equal to $n \mathbb{E} \bar{X}(v_1^*)$. By the union bound and Theorem 2 applied to each tree that can be obtained from T by rerooting at an internal vertex, the second term is at most $i_T n \mathbb{E} [D^{h-1} \mathbf{1}_{D \geq \Delta}] \leq i_T \lambda n$. Therefore

$$\mathbb{E} \bar{X}(v_1^*) \leq n^{-1} \text{emb}(H, G) \leq \mathbb{E} \bar{X}(v_1^*) + i_T \lambda. \quad (3.7)$$

Hoeffding's classical inequality says that for a sum of independent random variables X_1, \dots, X_N with values in $[0, 1]$ and $\mu = N^{-1} \mathbb{E}(X_1 + \dots + X_N)$ we have

$$\mathbb{P}(|N^{-1}(X_1 + \dots + X_N) - \mu| \geq t) \leq 2 \exp(-2Nt^2). \quad (3.8)$$

The claim follows by applying this to the random variables $X_i := \bar{X}(v_i^*)/(\Delta - 1)^{h-1}$ and using (3.7) (or using $\bar{X}(v_i^*) = 0$ if $\Delta = 1$). \square

In particular, we obtain by choosing $s = \epsilon \mathbb{E} D^{h-1}$ in Theorem 4 the following corollary. (Choosing s in this way makes sense when $n^{-1} \text{emb}(H, G)$ is of the same order as its upper bound $n^{-1} \text{hom}(K_{1,h-1}, G) = \mathbb{E} D^{h-1}$, which often may be reasonable to expect in practice.)

Corollary 5. *With notation as above, assume that (3.3) holds. If $\epsilon > 0$, $p \in (0, 1]$, and*

$$N \geq \frac{(\Delta - 1)^{2(h-1)}}{2\epsilon^2 (\mathbb{E} D^{h-1})^2} \ln \frac{2}{p}, \quad (3.9)$$

then

$$\mathbb{P}(\widehat{X}_N - \epsilon \mathbb{E} D^{h-1} \leq n^{-1} \text{emb}(H, G) \leq \widehat{X}_N + \epsilon \mathbb{E} D^{h-1} + i_T \lambda) \geq 1 - p. \quad (3.10)$$

If we are able to draw edges K_2 , wedges P_3 or other small subgraphs in G uniformly at random, we can estimate certain small subgraph densities using a much smaller sample size using the following generalization of Theorem 4. Other authors have used other methods to obtain some practical results in estimating densities of graphs H on $h \leq 5$ vertices using algorithms which sample random paths on up to 5 vertices as their first step, see, e.g., [10, Section 4.3].

To state the generalization, let now O be a non-empty subgraph of H . Let, as above, T be a spanning tree of H . We declare that a vertex $v \in T$ is O -internal if either $v \in T \setminus V(O)$ and $d_v^T \geq 2$, or $v \in O$ and there is an edge $uv \in T$ with $u \notin O$. We let i_T^O denote the number of O -internal vertices in T . Note that for $h \geq 2$ if O consists of a single vertex o , this agrees with the previous definition.

Assume $\text{emb}(O, G) \geq 1$ and let $\nu \in \text{Emb}(O, G)$. Let $\bar{X}^O(\nu)$ denote the number of embeddings $\sigma \in \text{Emb}(H, G)$ such that ν is the restriction of σ to O , and $d_{\sigma(u)} < \Delta$ for all O -internal vertices u . Let ν_1^*, \dots, ν_N^* be independent, uniformly random elements of $\text{Emb}(O, G)$. Consider the estimate

$$\widehat{X}_N^O := \frac{1}{N} (\bar{X}^O(\nu_1^*) + \dots + \bar{X}^O(\nu_N^*)). \quad (3.11)$$

Similarly as above we have:

Theorem 6. *With notation and assumptions as above and in Theorem 4, including (3.3), assume also $\text{emb}(O, G) \geq 1$ and, instead of (3.4),*

$$N \geq \frac{(\Delta - 1)^{2(h-|V(O)|)}}{2s^2} \ln \frac{2}{p}. \quad (3.12)$$

Then

$$\mathbb{P}\left(\widehat{X}_N^O - s \leq \frac{\text{emb}(H, G)}{\text{emb}(O, G)} \leq \widehat{X}_N^O + s + \frac{i_T^O \lambda n}{\text{emb}(O, G)}\right) \geq 1 - p. \quad (3.13)$$

Proof. Using the same argument as for (3.7) we get by Theorem 2:

$$\text{emb}(O, G) \mathbb{E} \bar{X}^O(\nu_1^*) \leq \text{emb}(H, G) \leq \text{emb}(O, G) \mathbb{E} \bar{X}^O(\nu_1^*) + i_T^O \lambda n. \quad (3.14)$$

Note that $0 \leq \bar{X}^O(\nu) \leq (\Delta - 1)^{h-|V(O)|}$. To finish the proof, we again apply Hoeffding's inequality. \square

For a simple example, let $H = K_h$ and $T = O = K_{1, h-1}$; then $|V(O)| = h$ and $i_T^O = 0$. Thus the values of Δ and λ are irrelevant; we take $\Delta = n$ and $\lambda = 0$ so that (3.3) holds. Theorem 6 shows that $N \geq 2^{-1} s^{-2} \ln \frac{2}{p}$ suffices. (In this simple case with $V(O) = V(H)$, this follows without invoking Theorem 2 in the proof.)

When we are only able to draw uniform embeddings of O with $|V(O)| < h$, if (3.3) holds with non-trivial values of Δ and λ , we can apply Theorem 6, for example, with $s = \epsilon \mathbb{E} D^{h-1}$ as in Corollary 5 not only to reduce N but also to bound the steps needed to compute $\bar{X}^O(\nu)$.

Finally, we note that Theorem 2 allows us to generalize (the difficult part of) Theorem 2.1 of [8], with a simpler proof that does not require the local weak convergence assumption of [8].

Theorem 7. *Let H be a fixed connected graph on h vertices, and pick an arbitrary vertex $o \in H$ as its root. Let $(G_n, n \in \{1, 2, \dots\})$ be a sequence of graphs. Let V_n be a uniformly random vertex from $V(G_n)$ and let $D_n := d_{V_n}$ be its degree in G_n .*

If $X(H, G_n, V_n)$ converges in distribution to a random variable X^ as $n \rightarrow \infty$, and D_n^{h-1} is uniformly integrable, then*

$$|V(G_n)|^{-1} \text{emb}(H, G_n) \rightarrow \mathbb{E} X^*. \quad (3.15)$$

Proof. Write $X_n := X(H, G_n, V_n)$. Clearly

$$\mathbb{E} X_n = |V(G_n)|^{-1} \text{emb}(H, G_n). \quad (3.16)$$

Hence we need to prove that $\mathbb{E} X_n \rightarrow \mathbb{E} X^*$. Since $X_n \xrightarrow{d} X^*$, we have $\mathbb{E} X^* \leq \liminf_{n \rightarrow \infty} \mathbb{E} X_n$ by Fatou's lemma.

To show the opposite inequality, fix a rooted spanning tree T of H and a positive integer Δ , and write $\bar{X}_n := \bar{X}(H, G_n, V_n, T, \Delta)$. By (3.16) and (3.6) (with G_n instead of G), Theorem 2 implies, as above for (3.7),

$$\mathbb{E} X_n \leq \mathbb{E} \bar{X}_n + i_T \mathbb{E} [D_n^{h-1} \mathbf{1}_{D_n \geq \Delta}]. \quad (3.17)$$

Furthermore, by (3.2), writing $x \wedge y := \min(x, y)$,

$$\bar{X}_n \leq X_n \wedge (\Delta - 1)^{h-1}. \quad (3.18)$$

Since X_n converges in distribution to X^* , we have

$$\lim_{n \rightarrow \infty} \mathbb{E} [X_n \wedge (\Delta - 1)^{h-1}] = \mathbb{E} [X^* \wedge (\Delta - 1)^{h-1}] \leq \mathbb{E} X^*. \quad (3.19)$$

Write $\epsilon_\Delta := \limsup_{n \rightarrow \infty} \mathbb{E}[D_n^{h-1} \mathbf{1}_{D_n \geq \Delta}]$. Then, by (3.17)–(3.19) and $i_T \leq h$,

$$\limsup_{n \rightarrow \infty} \mathbb{E} X_n \leq \limsup_{n \rightarrow \infty} \mathbb{E} \bar{X}_n + h\epsilon_\Delta \leq \mathbb{E} X^* + h\epsilon_\Delta. \quad (3.20)$$

This holds for every $\Delta > 0$, and $\lim_{\Delta \rightarrow \infty} \epsilon_\Delta = 0$ by the uniform integrability assumption; thus $\limsup_{n \rightarrow \infty} \mathbb{E} X_n \leq \mathbb{E} X^*$, which completes the proof. \square

4. CONCLUDING REMARKS

We extended Sidorenko’s inequality and used it to derive bounds on the sample size needed to estimate the number of small subgraphs in a large graph using only the weak assumption (3.3).

Like Hoeffding’s bound, our estimate works for worst case graphs; therefore the lower bound from Theorem 4 may be too pessimistic for specific real-world graphs. Nevertheless it would be interesting to better understand if our results and assumptions of type (3.3) can be useful in practice.

It would also be interesting to find an interpretation or applications of the general case of our inequality (2.5).

REFERENCES

- [1] H. Andersson, Limit theorems for a random graph epidemic model, *Annals of Applied Probability* **8** (1998), 1331–1349.
- [2] I. Benjamini, R. Lyons and O. Schramm, Unimodular random trees, *Ergodic Theory and Dynamical Systems* **35** (2015), 359–373.
- [3] A. D. Broido and A. Clauset, Scale-free networks are rare, *Nature Communications* **10** (2019), 1–10.
- [4] P. Dagum, R. Karp, M. Luby and S. Ross, An optimal algorithm for Monte Carlo estimation, *SIAM Journal on Computing* **29** (2000), 1484–1496.
- [5] D. M. Feehan and A. S. Mahmud, Quantifying population contact patterns in the United States during the COVID-19 pandemic, *Nature Communications* **12** (2021), 1–9.
- [6] G. Grimmett, *Percolation*, 2nd ed., Springer, Berlin, 1999.
- [7] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, 1952.
- [8] V. Kurauskas, On local weak limit and subgraph counts for sparse random graphs, *Journal of Applied Probability* **59** (2022), 755–776.
- [9] V. Mnih, Cs. Szepesvári and J.-Y. Audibert, Empirical Bernstein stopping, In *Proceedings of the 25th International Conference on Machine Learning (ICML ’08)* (2008), Association for Computing Machinery, New York, NY, USA, 672–679.
- [10] P. Ribeiro, P. Paredes, M. E. P. Silva, D. Aparicio and F. Silva, A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets, *ACM Computing Surveys* **54** (2021), 1–36.
- [11] A. Sidorenko, A partially ordered set of functionals corresponding to graphs, *Discrete Mathematics* **131** (1994), 263–277.

APPENDIX A. A NOTE ON REAL-WORLD EXPERIMENTS

During the preparation of this paper we tested Corollary 5 with some real-world degree distributions. For our first experiment, we considered survey data on self-reported human contact count distributions during a single day in the USA in 4 COVID-19 pandemic waves [5]. For our second experiment, we considered degree distributions of more than 500 empirical networks of various types and sizes made available as part of the supporting code of [3].

Although the lower bounds in our first experiment seem to be interesting for further exploration, we believe that the survey sizes in [5] (several thousand respondents in each of the COVID-19 waves) were too small to determine if there exists a practically useful choice of λ and Δ in (3.3), even for $h = 3$. In the second case for 75% of degree distributions we got a lower bound for N exceeding the underlying network size.

We believe that in the first case establishing a better understanding on the degree tails, simply collecting more data or applying the adaptive mean estimation methods [4; 9] might help. In the second case, since the full network data is available, the methods mentioned in [10] and Theorem 6 seem to be more suitable.

The code and the results of our experiments are available at <https://github.com/valentas-kurauskas/subgraph-counts-hoeffding>.

DEPARTMENT OF MATHEMATICS, UPPSALA UNIVERSITY, PO Box 480, SE-751 06 UPPSALA, SWEDEN

Email address: `svante.janson@math.uu.se`

URL: `http://www2.math.uu.se/~svante/`

FACULTY OF MATHEMATICS AND INFORMATICS, VILNIUS UNIVERSITY, AKADEMIJOS 4, 08412, VILNIUS, LITHUANIA

Email address: `valentas@gmail.com`