Oded (May 2020, revised July 16, 2020): On the main result of [1].

This memo presents what I consider to be the main result and technique of [1], which is a randomized worst-case to average-case reduction for counting k-cliques in k-partite graphs.[1] Here the average-case problem refers to the uniform distribution over (k-partite n-vertex) graphs and the error rate (of the potentail average-case solver) is noticeable alas vanishing error rate. Specifically, the allowed error rate is $(\log n)^{-\widetilde{O}(k^2)}$, where n is the number of vertices in the graph.

The model. For fixed k, we consider counting k-cliques in k-partite n-vertex graphs, where each part has size $\lfloor n/k \rfloor$. Assume for simplicity that n is a multiple of k, and let $m = \binom{k}{2} \cdot (n/k)^2 < \binom{n}{2}$ denote the number of possible edges in a k-partite n-vertex graph. Let $C_k : \{0,1\}^m \to \mathbb{N}$ denote the function that represents the number of k-cliques in a k-partite n-vertex graph represented by (the non-redundant part of) its adjacency matrix.

Our aim is to randomly reduce computing C_k in the worst-case to computing C_k on the uniform distribution. Below, we present a reduction that makes $(\log n)^{\widetilde{O}(k^2)}$ queries such that each query is uniformly distributed in $\{0,1\}^m$. The establishes the foregoing claim.

The reduction

For a prime $p \in (n^k, 2 \cdot n^k]$, consider the extension of C_k to a polynomial P_k over $\mathcal{F} = \mathrm{GF}(p)$, and observe that P_k is multilinear in $\binom{k}{2}$ sets of variables, where each set corresponds to the bipartite graph that connects two parts of the k-partite graph. Using the hypothesis that $p > n^k$ and the fact that the value of C_k on any binary sequence does not exceed n^k, it follows that the value of P_k on binary strings equals the value of C_k on those strings. Thus, computing C_k reduces (in the worst-cvase sense) to computing P_k.

By the standard self-reduction of polynomials, it follows that evaluating P_k in the worst-case is randomly reducible (using $\binom{k}{2} + 1 < k^2$ queries) to computing $P_k : \mathcal{F}^m \to \mathcal{F}$ on the average with error rate of at most $1/3k^2$. Hence, we focus on reducing the computation of P_k on random inputs (in $\mathcal{F}^m$) to the computation of C_k on random inputs (in $\{0,1\}^m$).

Looking at a generic term of P_k, observe that it has the form $\prod_{\alpha<\beta\in[k]} X^{(\alpha,\beta)}_{v_\alpha,v_\beta}$, where $v_1, ..., v_k \in [n/k]$ and $X^{(\alpha,\beta)}_{v_\alpha,v_\beta}$ corresponds to a vertex-pair (i.e., (v_α, v_β)) with endpoints in parts α and β, respectively. Letting X denote the corresponding sequence of variables, observe that

$$P_k(X) = \sum_{v_1,...,v_k\in[n/k]} \prod_{\alpha<\beta\in[k]} X^{(\alpha,\beta)}_{v_\alpha,v_\beta} \tag{1}$$

where $(v_1, ..., v_k)$ corresponds to a potential k-clique in the k-partite graph. Let $\ell = \log_2(n^{k+3})$ and define the function $F_k : \{0,1\}^{m\cdot\ell} \to \mathbb{N}$ such that

$$F_k(x) \stackrel{\text{def}}{=} \sum_{v_1,...,v_k\in[n/k]} \prod_{\alpha<\beta\in[k]} \sum_{i\in[\ell]} x^{(\alpha,\beta,i)}_{v_\alpha,v_\beta} \cdot 2^{i-1} \tag{2}$$

[1] Let me stress that [1] has many other results, which the authors consider even more interesting.

where $x^{(\alpha,\beta,1)}_{v_\alpha,v_\beta},...,x^{(\alpha,\beta,\ell)}_{v_\alpha,v_\beta}$ represents the (ℓ-bit long) block that corresponds to the variable $X^{(\alpha,\beta)}_{v_\alpha,v_\beta}$ in P_k. Then,

$$\begin{aligned} F_k(x) &= \sum_{v_1,...,v_k\in[n/k]} \sum_{(i_{1,2},...,i_{k-1,k})\in[\ell]^{\binom{k}{2}}} 2^{\sum_{\alpha<\beta\in[k]}(i_{\alpha,\beta}-1)} \cdot \prod_{\alpha<\beta\in[k]} x^{(\alpha,\beta,i_{\alpha,\beta})}_{v_\alpha,v_\beta} \\ &= \sum_{(i_{1,2},...,i_{k-1,k})\in[\ell]^{\binom{k}{2}}} 2^{\sum_{\alpha<\beta\in[k]}(i_{\alpha,\beta}-1)} \cdot \sum_{v_1,...,v_k\in[n/k]} \prod_{\alpha<\beta\in[k]} x^{(\alpha,\beta,i_{\alpha,\beta})}_{v_\alpha,v_\beta}, \end{aligned}$$

(Here we capitalize on the fact that in the first expression the sum is over k-long sequences rather than k-subsets; this is due to the fact that C_k and P_k refer to k-cliques in k-partite graphs.) Using the foregoing correspondence (between X and x), it follows that $P_k(X)$ *is congruent to* $F_k(x)$ *modulo* p. This holds not only when each block in x encodes the corresponding field element in X, but also when it encodes a value that is congruent to this field element modulo p.

The latter observation is important because it allows us to encode a uniformly distributed element of $\mathcal{F}^m$ by an almost uniformly distributed element of $\{0,1\}^{m\cdot\ell}$. Specifically, we encode $v\in\mathcal{F}$ by a uniformly distributed sequence $r=(r^{(j)})_{j\in[\ell]}\in\{0,1\}^\ell$ such that $\sum_{j\in[\ell]}2^{j-1}\cdot r^{(j)}\equiv v \pmod p$. Hence, when v is uniformly distributed in $\mathcal{F}$, the resulting r is $p\cdot 2^{-\ell}$-close to being uniformly distributed in $\{0,1\}^\ell$. Recall that $p\cdot 2^{-\ell}<2n^k\cdot n^{-(k+3)}<n^{-1}/\binom{n}{2}$.

The key observation is that, for every $(i_{1,2},...,i_{k-1,k})\in[\ell]^{\binom{k}{2}}$, it holds that

$$\sum_{v_1,...,v_k\in[n/k]} \prod_{\alpha<\beta\in[k]} x^{(\alpha,\beta,i_{\alpha,\beta})}_{v_\alpha,v_\beta} = C_k(y), \tag{3}$$

where $y^{(\alpha,\beta)}_{v_\alpha,v_\beta}$ equals $x^{(\alpha,\beta,i_{\alpha,\beta})}_{v_\alpha,v_\beta}$. Hence, $F_k(x)$ is computed by evaluating C_k at $\ell^{\binom{k}{2}}$ points, and if x uniformly distributed (in $\{0,1\}^{m\cdot\ell}$), then each query to C_k is n^{-1}-close to being uniformly distributed (in $\{0,1\}^m$).

Conclusion. The foregoing worst-case to average-case reduction of C_k makes $q=\ell^{\binom{k}{2}}\cdot(\binom{k}{2}+1)$ queries, and yields a correct answer (with probability at least 2/3) provided that the error rate (of the average-case solver) is at most $1/3q$. Recalling that $\ell=(k+3)\cdot\log_2 n$, this yields an error rate of $(\log n)^{-\widetilde{O}(k^2)}$.

Digest. The key observation is captured by Eq. (3), which implies that $F_k(x)$ can be decomposed to $\ell^{\binom{k}{2}}$ terms such that each term corresponds to a sequence $(i_{1,2},...,i_{k-1,k})\in[\ell]^{\binom{k}{2}}$ and represent the contribution of individual bits in each bipartite graph. Specifically, for $\alpha<\beta\in[k]$, only the contribution of the $i^{\text{th}}_{\alpha,\beta}$ bit of the elements associated with the bipartite graph between the α^{th} and β^{th} parts is taken. This decomposition is possible since we are dealing with k-partite graphs, given that we already expressed $F_k(x)$ in term of the contribution of bits (see Eq. (2)); the latter expression was already used in [2].

Comparison to [3]

The reduction presented above yields the correct answer whenever all queries are answered correctly. In contrast, the reduction in [3] yields the correct answer even if only a noticeable fraction of the

queries are answered correctly. Hence, the current reduction yields a worst-case to average-case reduction when average-case is understood as having noticeable alas vanishing error rate, whereas the result in [3] applies to average-case in a much more relaxed sense (i.e., having vanishing but noticeable success rate). On the other hand, here average-case refers to the uniform distribution over all k-(equi)partite graphs, whereas [3] refers to uniform distribution over a more structured set (which is easily recognizable).

References

[1] Enric Boix-Adsera, Matthew Brennan, and Guy Bresler. The Average-Case Complexity of Counting Cliques in Erdos-Renyi Hypergraphs. In *60th FOCS*, 2019.

[2] Oded Goldreich and Guy Rothblum. Worst-case to Average-case reductions for subclasses of P. *ECCC*, TR17-130, 2017.

[3] Oded Goldreich and Guy Rothblum. Counting t-Cliques: Worst-Case to Average-Case Reductions and Direct Interactive Proof Systems. In *59th FOCS*, 2018.