AVERAGE-CASE ANALYSES OF VICKREY COSTS

PRASAD CHEBOLU*, ALAN FRIEZE†, PÁLL MELSTED‡, AND GREGORY B. SORKIN◇

ABSTRACT. We explore the average-case "Vickrey" cost of structures in a random setting: the Vickrey cost of a shortest path in a complete graph or digraph with random edge weights; the Vickrey cost of a minimum spanning tree (MST) in a complete graph with random edge weights; and the Vickrey cost of a perfect matching in a complete bipartite graph with random edge weights. In each case, in the large-size limit, the Vickrey cost is precisely 2 times the (non-Vickrey) minimum cost, but this is the result of case-specific calculations, with no general reason found for it to be true.

Separately, we consider the problem of sparsifying a complete graph with random edge weights so that all-pairs shortest paths are preserved approximately. The problem of sparsifying a given graph so that for every pair of vertices, the length of the shortest path in the sparsified graph is within some multiplicative factor and/or additive constant of the original distance has received substantial study in theoretical computer science. For the complete graph K_n with additive edge weights, we show that **whp** $\Theta(n \ln n)$ edges are necessary and sufficient for a spanning subgraph to give good all-pairs shortest paths approximations.

1. INTRODUCTION

The motivation for this work is that a "Vickery-Clarke-Groves" (VCG) auction, while having the benefits of being "truthful" and maximizing "social welfare", may result in arbitrarily large overpayments. We are interested in whether the overpayments are reasonably small in an average-case setting. In this introduction we first recapitulate the VCG auction mechanism and introduce a small amount of notation, then state our average-case results.

1.1. **The VCG auction mechanism.** Suppose that in a graph, each edge is provided by an independent, selfish agent who incurs a cost for supplying it (or for allowing us to drive over it, transmit data over it, or whatever). This "private" cost, the price point at which the agent is neutral between selling the edge or not, is known only to herself. We wish to buy some structure, for example a path between two particular points, or a spanning tree, as cheaply as possible. An obvious "mechanism" to do this is to ask each agent the cost of her edge, find the cheapest structure, and pay each agent accordingly. The problem with this and many other mechanisms is that agents have an incentive to lie: by inflating her claimed cost, an agent will get more money (up to the point where she prices herself out of competition).

A Vickrey-Clarke-Groves (VCG) auction [Vic61, Cla71, Gro73] is a cleverly designed "truthful" mechanism: assuming that the agents act without collusion, in a VCG auction it is in each agent's best interest to name her true cost (if her edge is used, she will get paid at least this, but typically more). Under the same assumption, a VCG auction also maximizes "social welfare": the structure selected is the one that is genuinely cheapest (and so the least possible resource is consumed in road maintenance, data-server support, or whatever).

In a VCG auction, an "auctioneer" first finds a cheapest structure S^*, according to the edge costs $c(e)$ declared by the agents. (This might be a cheapest path, for example; VCG was first explicitly applied to the shortest-path problem in [NR99, NR01].) For each edge $e \in S^*$ in this structure, the auctioneer pays the corresponding agent not the stated cost $c(e)$ of the edge, but a measure of the benefit it provided, namely the difference between what a cheapest structure would have cost if the edge were not present or had infinite cost, call it $c(S_e^\infty)$, and what the cheapest structure would have cost if the edge were free, call it $c(S_e^0)$. It is clear that neither of these terms depends on $c(e)$. An agent whose edge is not used, $e \notin S^*$, is not paid anything.

We can now confirm three important properties of the auction. First, define an "incentive cost" for edge e as

$$(1) \qquad\qquad c^+(e) = c(S_e^\infty) - c(S^*).$$

Then for any edge, used or not, $c^+(e) \geq 0$, and assuming for convenience of discussion that there is a unique cheapest structure S^*, e is used iff $c(S^*) < c(S_e^\infty)$, i.e., iff $c^+(e) > 0$. If edge e is used, the payment for it is

$$c(S_e^\infty) - c(S_e^0) = [c(S_e^\infty) - c(S^*)] + [c(S^*) - c(S_e^0)]$$
$$= c^+(e) + c(e),$$

so an edge is used precisely when the payment would exceed the agent's stated cost.

This also shows that the auction is truthful. Since by manipulating her price $c(e)$ an agent cannot influence what she would be paid, but only whether or not she will, her best strategy is to get paid iff the payment exceeds her true cost, which she can achieve by setting $c(e)$ to be the true cost. Setting $c(e)$ lower may result in her being paid less than cost; setting it higher may cause her to lose out on what would have been a profitable sale.

Finally, assuming that every agent does state her true cost, the structure selected is one that is genuinely cheapest, and social welfare is maximized.

We will later take advantage of an observation based on the fact that the incentive cost (1) is 0 for edges not used, namely that

$$\mathrm{VCG} = \sum_{e \in S^*} [c(e) + (c(S_e^\infty) - c(S^*))]$$
$$= c(S^*) + \sum_{e \in S^*} (c(S_e^\infty) - c(S^*))$$
$$= c(S^*) + \sum_{e \in E} (c(S_e^\infty) - c(S^*))$$
$$(2) \qquad\qquad = \sum_{e \in E} c(S_e^\infty) - (|E| - 1)c(S^*).$$

1.2. **Average-case analysis.** Naturally, the VCG mechanism pays more than the cost of the cheapest structure, and unfortunately the overpayment can be arbitrarily large. For an illustrative example, consider finding a cheapest s-to-t path in a graph consisting of two disjoint paths from s to t, one with all edges costing 0, the other with one edge costing 1 and all others 0. The first path is cheapest with true cost 0, but every edge in it receives a VCG incentive payment of 1 (the cost difference between the two paths), so the total VCG cost is the path length. In [AT02, AT07] it is shown that any truthful mechanism has bad worst-case s–t path overpayment. Additional investigations of shortest paths in a worst-case setting appear in [MPS03, ESS04, CR04, Elk05].

A worst-case analysis of VCG costs may be overly pessimistic. One alternative is through real-world measurements of the VCG overpayment, and such a study appears in [FPSS02]. Another alternative, and the one we adopt here, is to compare the VCG cost with the minimum cost in an average-case setting. This was done for shortest paths in certain graphs in [MPS03, CR04,

KN05, FGS06]. We consider the expected VCG overpayment in three settings, in each of which the expected minimum cost is a classical result in the analysis of random structures.

1.2.1. *Shortest paths.* We first consider shortest paths in the complete graph K_n, or complete digraph $\vec{K}_n$, with i.i.d. exponential(1) edge weights, where exponential(1) denotes the exponential distribution with mean 1. (We use the terms edge weight, cost, or length interchangeably, and a shortest path is a cheapest path.) Janson [Jan99] has shown that **whp** the distance between two vertices, say 1 and n, in this model is $(1 + o(1)) \log n / n$. We prove that the asymptotic expected Vickrey cost is twice as large.

Theorem 1. *Suppose that the edges of the complete graph K_n (respectively, digraph $\vec{K}_n$) have i.i.d. exponential mean-1 edge weights. Let $\mathbf{E}(\mathrm{SP})$ be the expected cost of a shortest path from 1 to n. Then*

$$\mathbf{E}(\mathrm{VCG}) \sim 2\mathbf{E}(\mathrm{SP}).$$

In a small digression, we will also consider the problem of sparsifying the random edge-weighted digraph so that **whp** shortest path distances are (approximately) preserved. Janson [Jan99] also showed that the weighted diameter in this model is $(3 + o(1)) \log n / n$. It follows that **whp** the subgraph consisting of the $4n \log n$ cheapest edges contains the shortest path between each pair of vertices. If we only keep Dn edges (so D is the average in+out degree), with $D = D(n) = O(\log n)$, how good an approximation can we find in the all pairs shortest path problem?

Theorem 2. *In a complete digraph $\vec{K}_n$ with i.i.d. exponential(1) edge weights, suppose that we keep only Dn edges to form a digraph H. Then **whp** there exists a pair of vertices s, t such that $d_H(s,t)/d_{\vec{K}_n}(s,t) \geq \frac{\log n}{4D}$, where $d_H(s,t)$ and $d_{\vec{K}_n}(s,t)$ denote shortest distance in H and $\vec{K}_n$ respectively.*

1.2.2. *Minimum Spanning Tree.* We next consider a minimum spanning tree of K_n with uniform $[0, 1]$ edge weights. It was shown by Frieze [Fri85] that the expected cost $\mathbf{E}(\mathrm{MST})$ of a minimum spanning tree on K_n satisfies $\lim_{n\to\infty} \mathbf{E}(\mathrm{MST}) = \zeta(3)$. Even though there is no nice expression for the exact expectation for finite n, we prove that the expected VCG cost is *exactly* (not just asymptotically) twice as large.

Theorem 3. *Suppose that the edges of the complete graph K_n have i.i.d. uniform $[0, 1]$ edge weights. Let $\mathbf{E}(\mathrm{MST})$ be the expected cost of a minimum spanning tree. Then*

$$\mathbf{E}(\mathrm{VCG}) = 2\mathbf{E}(\mathrm{MST}).$$

1.2.3. *Assignment.* Finally, we consider the VCG cost of a perfect matching in a complete bipartite graph with random edge weights, known as the "random assignment problem". When the edge weights are i.i.d. exponential(1) random variables, Mézard and Parisi [MP85, MP86, MP87] gave a sophisticated mathematical physics argument, using the "replica method" (related to the "cavity method"), that the minimum cost AP satisfies $\lim_{n\to\infty} \mathbf{E}(\mathrm{AP}) = \zeta(2) = \pi^2/6$. Aldous [Ald92, Ald01] made this mathematically rigorous through reasoning about a "Poisson weighted infinite tree". For finite values of n, Parisi [Par98] conjectured the expected cost to be $\sum_{i=1}^{n} i^{-2}$, Coppersmith and Sorkin [CS99] extended the conjecture to cheapest cardinality-k assignments in $K_{m,n}$, and these results were proved simultaneously, by different methods, by Linusson and Wästlund [LW04] and Nair, Prabhakar and Sharma [NPS05]. A beautiful short proof was later found by Wästlund [Wäs].

As in the previous cases, we find that the expected VCG cost is twice the minimum cost asymptotically (but not for finite n as it is for MST).

Theorem 4. *Suppose that the edges of the complete bipartite graph $K_{n,n}$ have i.i.d. exponential mean-1 edge weights. Let $\mathbf{E}(\mathrm{AP})$ be the expected cost of a minimum weight perfect matching. Then*

$$
\begin{aligned}
\mathbf{E}(\mathrm{VCG}) &= \mathbf{E}(\mathrm{AP}) + n\left(\frac{1}{n-1} + \sum_{l=1}^{n-1}\frac{1}{l}\frac{n-l}{n} - \sum_{l=2}^{n-1}\frac{1}{l(l-1)}\sum_{i=0}^{l-1}\frac{n-i}{n}\prod_{j=i+1}^{l}\frac{(n-j)j}{(n-j+1)j-1}\right)\\
&\sim \ 2\mathbf{E}(\mathrm{AP}).
\end{aligned}
$$

2. Shortest Paths

The cost model for this section will be that each edge e of the complete graph K_n or digraph $\vec{K}_n$ is given an independent cost X_e where X_e is exponential with mean 1, i.e., $X_e \sim \mathrm{exponential}(1)$ are i.i.d. random variables. We will compute the expected Vickrey cost of a shortest path from vertex 1 to vertex n. We start by computing the expected cost $\mathbf{E}(\mathrm{SP})$ of this path. We follow the analysis of Dijkstra's algorithm due to Janson [Jan99]. Janson actually considered the symmetric (undirected) case, but there is no essential difference in the analysis of the two. We begin with the directed (asymmetric) case in Section 2.1, treating the undirected case in Section 2.2.

2.1. **Asymmetric (directed) model.** In the complete digraph $\vec{K}_n$, if D_k is the distance to the kth closest vertex from vertex 1 then $D_1 = 0$ and for $k > 1$ we have

$$
(3) \qquad\qquad\qquad\qquad \mathbf{E}(D_{k+1}) = \mathbf{E}(D_k) + \frac{1}{k(n-k)}.
$$

Explanation: Suppose that we are about to add the $(k+1)$th closest vertex to our shortest path tree T. Let v be the jth vertex added to T and suppose that $w \notin T_k$ the state of T after k rounds. The conditional distribution of the edge length $c(v,w)$ is exponential mean-1, conditional on being at least $D_k - D_j$, i.e., $D_{j-1} + c(v,w) = D_k + X_{v,w}$ where the $X_{v,w}$ are independent exponentials with mean 1. This condition ensures that w has not already been added to the tree. There are $k(n-k)$ edges and the expected value of the minimum of m mean-1 exponentials is $1/m$.

It follows from (3) that

$$
(4) \qquad\qquad\qquad\qquad \mathbf{E}(D_{k+1}) = \sum_{i=1}^{k}\frac{1}{i(n-i)}.
$$

Vertex n is added to T at a random stage of this process, so

$$
(5) \qquad\qquad \mathbf{E}(\mathrm{SP}) = \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{i=1}^{k}\frac{1}{i(n-i)} = \frac{1}{n-1}\sum_{i=1}^{n-1}\sum_{k=i}^{n-1}\frac{1}{i(n-i)} = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{i}.
$$

Now let SP_e denote the shortest path length when we are not allowed to use edge $e = (r,s)$. Letting $D_k^{(r,s)}$ denote the distance to the kth closest vertex v_k from vertex 1 when (r,s) is excluded, in place of (3) we have

$$
(6) \qquad\qquad \mathbf{E}(D_{k+1}^{(r,s)}) = \mathbf{E}(D_k^{(r,s)}) + \mathbf{E}\left(\frac{1}{k(n-k) - \theta(r,s,k)}\right)
$$

where $\theta(r,s,k)$ is the indicator for $r \in T_k$, $s \notin T_k$. We will need to estimate the following under the assumption that (r,s) is a random edge:

$$
\begin{aligned}
\mathbf{E}(\theta(r,s,i)\mathbf{1}(v_k = n)) &= \mathbf{P}(\mathbf{1}(v_k = n) \mid \theta(r,s,i) = 1)\mathbf{P}(\theta(r,s,i) = 1)\\
&= \mathbf{P}(\mathbf{1}(v_k = n) \mid \theta(r,s,i) = 1)\frac{i(n-i)}{n(n-1)}.
\end{aligned}
$$

We then write

$$
(7) \qquad \frac{k-1}{k(n-k)} \leq \mathbf{P}(\mathbf{1}(v_k = n) \mid r, s \text{ and } v_j \neq n, 1 \leq j < k) \leq \frac{k}{k(n-k)-1}.
$$

This is because the choice of v_k is always derived from the random choice of one of $k(n-k)-\theta(r,s,k)$ edges and $k-1$ or k will lead to $v_k = n$. Thus,

$$
\begin{aligned}
\mathbf{P}(\mathbf{1}(v_k = n) \mid \theta(r,s,i) = 1) \ &\leq \ \frac{k}{k(n-k)-1} \prod_{j=1}^{k-1}\left(1 - \frac{j-1}{j(n-j)}\right) \\
&= \ \frac{1}{n-k}\left(1 + \frac{1}{k(n-k)-1}\right)\prod_{j=1}^{k-1}\frac{n-j-1}{n-j}\left(1 + \frac{1}{(j-1)(n-j)}\right) \\
&\leq \ \frac{1}{n-1}\exp\left\{\sum_{j=1}^{k}\frac{1}{j(n-j)-1}\right\} \\
&= \ \frac{1}{n-1}\left(1 + O\left(\frac{\log n}{n}\right)\right).
\end{aligned}
$$

It follows that

$$
(8) \qquad \mathbf{E}(\theta(r,s,i)\mathbf{1}(v_k = n)) \leq \frac{i(n-i)}{n(n-1)^2}\left(1 + O\left(\frac{\log n}{n}\right)\right).
$$

With inequality (8) in hand, we have

$$
\begin{aligned}
\sum_e \mathbf{E}(\mathrm{SP}_e) \ &= \ \sum_{r,s}\sum_{k=1}^{n-1}\sum_{i=1}^{k}\mathbf{E}\left(\frac{\mathbf{1}(v_k = n)}{i(n-i)-\theta(r,s,i)}\right) \\
&= \ \sum_{k=1}^{n-1}\sum_{i=1}^{k}\sum_{r,s}\mathbf{E}\left(\frac{\mathbf{1}(v_k = n)}{i(n-i)}\right) + \sum_{k=1}^{n-1}\sum_{i=1}^{k}\sum_{r,s}\mathbf{E}\left(\frac{\mathbf{1}(v_k = n)\,\theta(r,s,i)}{i(n-i)(i(n-i)-\theta(r,s,i))}\right) \\
(9) \qquad &= \ n(n-1)\mathbf{E}(\mathrm{SP}) + n(n-1)\sum_{k=1}^{n-1}\sum_{i=1}^{k}\mathbf{E}\left(\frac{\mathbf{1}(v_k = n)\,\theta(r,s,i)}{i(n-i)(i(n-i)-\theta(r,s,i))}\right) \\
&\leq \ n(n-1)\mathbf{E}(\mathrm{SP}) + \frac{1}{n-1}\sum_{i=1}^{n-1}\sum_{k=i}^{n-1}\frac{1}{i(n-i)-1}\left(1 + O\left(\frac{\log n}{n}\right)\right) \\
&= \ n(n-1)\mathbf{E}(\mathrm{SP}) + \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{i}\left(1 + O\left(\frac{\log n}{n}\right)\right)
\end{aligned}
$$

where from (9) onwards we treat (r,s) as a random edge.

Going back to (2) and (5) we see that

$$
(10) \qquad \mathbf{E}(\mathrm{VCG}) \leq 2\mathbf{E}(\mathrm{SP}) + O\left(\frac{\log^2 n}{n^2}\right).
$$

For a lower bound we go back to (7) and write

$$
\begin{aligned}
\mathbf{P}(\mathbf{1}(v_k = n) \mid \theta(r,s,i) = 1) \;&\geq\; \frac{k-1}{k(n-k)} \prod_{j=1}^{k-1} \left(1 - \frac{j}{j(n-j)-1}\right) \\
&=\; \left(1 - \frac{1}{k}\right) \frac{1}{n-k} \prod_{j=1}^{k-1} \frac{n-j-1}{n-j} \left(1 - \frac{1}{j(n-j)-1}\right) \\
&=\; \left(1 - \frac{1}{k}\right) \frac{1}{n-1} \left(1 - O\left(\frac{\log n}{n}\right)\right).
\end{aligned}
$$

It follows that

$$
(11) \qquad \mathbf{E}(\theta(r,s,i)\mathbf{1}(v_k = n)) \geq \left(1 - \frac{1}{k}\right) \frac{i(n-i)}{n(n-1)^2} \left(1 + O\left(\frac{\log n}{n}\right)\right).
$$

Going back to (9)

$$
\begin{aligned}
\sum_e \mathbf{E}(\mathrm{SP}_e) \;&\geq\; n(n-1)\mathbf{E}(\mathrm{SP}) + \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{1}{i(n-i)} \sum_{k=i}^{n-1} \left(1 - \frac{1}{k}\right)\left(1 - O\left(\frac{\log n}{n}\right)\right) \\
&\geq\; n(n-1)\mathbf{E}(\mathrm{SP}) + \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{n-i-\log n}{i(n-i)} \left(1 - O\left(\frac{\log n}{n}\right)\right)
\end{aligned}
$$

and it follows that

$$
\mathbf{E}(\mathrm{VCG}) \geq 2\mathbf{E}(\mathrm{SP}) - O\left(\frac{\log^2 n}{n^2}\right).
$$

With the upper bound from (10), this lower bound proves the asymmetric case of Theorem 1.

2.2. **Symmetric (undirected) model.** Our analysis extends easily to the case where we have a complete graph K_n, as opposed to a complete digraph, still using i.i.d. exponential(1) edge weights. Firstly, Janson's result (originally proved in this model, in fact) is unchanged, i.e., (5) still holds. We then observe that $\mathbf{P}(\theta(r,s,i) = 1)$ is now $\frac{2i(n-i)}{n(n-1)}$, leading to a doubling of the expression (8). On the other hand, the factor $n(n-1)$ in (9) becomes $\binom{n}{2}$. The factor two loss in the first sum halves $|E|$ as it should, and the factor two loss in the second sum compensates for the doubling described in (8). We end up with the same expression as before.

3. All pairs shortest paths

We give an outline proof of Theorem 2, which is somewhat peripheral to the main theme of the paper. Let an edge $e = (s,t)$ be *bad* if (i) it is "short", with cost $c(e) \leq 2D/n$, and (ii) $d_{\vec{K}_n \setminus e}(s,t) \leq \frac{\log n}{2n}$. For a random edge e, $\mathbf{P}(e\ is\ bad) = \mathbf{P}(\mathrm{i})\mathbf{P}(\mathrm{ii})$ because these events are independent, depending respectively on the length of e and the lengths of the other edges. The theorem's conclusion is null unless $D = O(\log n)$, thus $D = o(n)$, in which case $\mathbf{P}(\mathrm{i}) = 1 - \exp(-2D/n) \sim \frac{2D}{n}$. For (ii) we use that $d_H(s,t) \geq d_{\vec{K}_n}(s,t)$; we have seen in (5) that $\mathbf{E}(d_{\vec{K}_n}(s,t)) \sim \frac{\log n}{n}$ and Janson [Jan99] has shown that $\mathbf{Var}(d_{\vec{K}_n}(s,t)) = O(n^{-2})$. (These results are shown for undirected graphs but follow similarly for directed graphs.) By Cheychev's inequality, then, $\mathbf{P}(\mathrm{ii}) = O(1/\log^2 n)$. Multiplying $\mathbf{P}(\mathrm{i})$ and $\mathbf{P}(\mathrm{ii})$, the number of bad edges Z satisfies $\mathbf{E}(Z) = O(Dn/\log^2 n)$, and by Markov's inequality, $Z \leq Dn/\log n$ **whp**. From $\mathbf{P}(\mathrm{i}) \sim 2D/n$ it is immediate that the (binomially distributed) number of such short edges is at least $\frac{3}{2}Dn$ **whp**. Assuming that all the above high-probability events hold, for any way of selecting Dn edges, there will be at least $Dn/2$ short edges not selected, and since $Z \leq Dn/\log n$, at least $Dn/3$ of these are not bad. These edges are short yet are not bad, thus

must violate (ii), so that $d_{\vec{K}_n}(s,t) \le c(e) \le \frac{2D}{n}$ while $d_H(s,t) \ge d_{\vec{K}_n \setminus e}(s,t) > \frac{\log n}{2n}$. This completes the proof of Theorem 2, since for these edges $d_H(s,t)/d_{\vec{K}_n}(s,t) \ge \frac{\frac{1}{2}\log n/n}{2D/n} = \frac{\log n}{4D}$.

In fact it shows more: for a large number of edges ($\Theta(n)$ of them), not only is the approximation ratio poor, but the additive gap is large as well: of order $\Omega(\log n/n)$, the same order as a typical distance.

4. Minimum Spanning Tree

The cost model for this section will be that each edge of the complete graph K_n is given an independent cost X_e where X_e is uniform $[0,1]$ for all $e \in E(K_n)$. We use the integral formula of Avram and Bertsimas [AB92]: For a connected graph $G = (V, E)$ with uniform $[0, 1]$ edge weights $X_e, e \in E$, let $\mathrm{MST}_G = \mathrm{MST}(G, X)$ denote the length of the minimum spanning tree with these edge weights. Then

$$(12) \qquad \mathbf{E}(\mathrm{MST}_G) = \int_{p=0}^1 \mathbf{E}(\kappa(G_p))dp - 1$$

where G_p is the random subgraph of G obtained by including each edge independently with probability p and $\kappa(G_p)$ is the number of connected components of G_p.

For $1 \le s \le n$ let $C_{s,m}$ denote the number of connected graphs of order s and size m. Using

$$\kappa(G) = \sum_{s=1}^n \binom{n}{s} \mathbf{1}(\text{the } s \text{ vertices induce a connected subgraph of } G)$$

$$(13) \qquad = \sum_{s=1}^n \binom{n}{s} \sum_{m=s-1}^{\binom{s}{2}} \sum_{i=1}^{C_{s,m}} \mathbf{1}(\text{the } i\text{th } s, m \text{ graph is a component of } G),$$

$$\mathbf{E}(\mathrm{MST}_{K_n}) = \sum_{s=1}^n \sum_{m=s-1}^{\binom{s}{2}} \binom{n}{s} C_{s,m} \int_{p=0}^1 p^m (1-p)^{\binom{s}{2}-m+s(n-s)} dp - 1$$

$$(14) \qquad = \sum_{s=1}^n \sum_{m=s-1}^{\binom{s}{2}} C'_{s,m} - 1$$

where, by $\int_{p=0}^1 p^m (1-p)^k dp = \frac{m!k!}{(m+k+1)!}$,

$$(15) \qquad C'_{s,m} = \binom{n}{s} C_{s,m} \frac{m! \left(\binom{s}{2} - m + s(n-s)\right)!}{\left(\binom{s}{2} + s(n-s) + 1\right)!}.$$

Going back to (12) and reasoning similarly as for (13),

$$\sum_{e \in K_n} \mathbf{E}(\mathrm{MST}_{K_n \setminus e}) = \sum_{e \in K_n} \left[\sum_{s=1}^n \sum_{m=s-1}^{\binom{s}{2}-\mathbf{1}(s=n)} \int_0^1 \mathbf{E}\left(\#(s\text{-vertex } m\text{-edge components of } (K_n \setminus e)_p)\right) dp - 1 \right]$$

$$(16) \qquad = \sum_{s=1}^n \sum_{m=s-1}^{\binom{s}{2}-\mathbf{1}(s=n)} (A_{s,m}^0 + A_{s,m}^2 + A_{s,m}^1) - \binom{n}{2}$$

where $A_{s,m}^0$ is the integral over p of the sum over e of the expected number of s-vertex m-edge components of a random probability-p subgraph of $K_n \setminus e$ containing *neither* endpoint of e, $A_{s,m}^2$ is the like sum for components containing *both* endpoints, and $A_{s,m}^1$ that for exactly *one* endpoint.

For an edge e of K_s, let $\widehat{C}_{s,m}$ denote the number of m-edge spanning connected subgraphs of K_s containing e. We have

$$
(17) \qquad \binom{s}{2}\widehat{C}_{s,m} = mC_{s,m}
$$

since both sides of this equation count the number of pairs (f, H) where H is an m-edge spanning subgraph of $[s]$ and f is an edge of H. (Throughout, we take $\binom{n}{k} = 0$ if $n < k$. Above, if $s = 1$ the left side is 0, and so is the right, since $s = 1$ implies $m = 0$.) To calculate $A^0_{s,m}$ we select s vertices, whereupon edge e must have both endpoints outside these s, while on these s vertices any subgraph $C_{s,m}$ is acceptable, and we integrate the probability that the chosen subgraph is induced and isolated in $G = K_n \setminus e$:

$$
\begin{aligned}
A^0_{s,m} &= \binom{n}{s}\binom{n-s}{2}C_{s,m}\int_{p=0}^{1} p^m(1-p)^{\binom{s}{2}-m+s(n-s)}dp \\
&= \binom{n-s}{2}\binom{n}{s}C_{s,m}\frac{m!\left(\binom{s}{2}-m+s(n-s)\right)!}{\left(\binom{s}{2}+s(n-s)+1\right)!} \\
&= \binom{n-s}{2}C'_{s,m},
\end{aligned}
$$

where the last line uses (15). The formula correctly evaluates to 0 for $s > n - 2$, and thus can safely be applied for all pairs s, m in the sum (16).

In calculating $A^2_{s,m}$, both endpoints of e are within the s vertices selected, but then the graph under consideration is $G = K_n \setminus e$ and a subgraph of G cannot use e, so that by (15) the number of valid subgraphs is $C_{s,m} - \widehat{C}_{s,m}$, and the probability that a subgraph is induced is adjusted to reflect that the size of G is $\binom{s}{2} - 1$:

$$
\begin{aligned}
A^2_{s,m} &= \binom{n}{s}\binom{s}{2}(C_{s,m} - \widehat{C}_{s,m})\int_{p=0}^{1} p^m(1-p)^{\binom{s}{2}-1-m+s(n-s)}dp \\
&= \binom{n}{s}\left(\binom{s}{2}C_{s,m} - mC_{s,m}\right)\frac{m!\left(\binom{s}{2}-m-1+s(n-s)\right)!}{\left(\binom{s}{2}+s(n-s)\right)!} \qquad \text{(from (17))} \\
&= \left(\binom{s}{2}-m\right)\frac{\binom{s}{2}+s(n-s)+1}{\binom{s}{2}-m+s(n-s)}C'_{s,m}.
\end{aligned}
$$

The formula properly evaluates to 0 for $s = 1$ (where $m = 0$) and for $m = \binom{s}{2}$. It cannot be applied when $s = n$, $m = \binom{n}{2}$, where it is $0/0$, but anyway this pair is excluded from the sum (16).

Calculating $A^1_{s,m}$ is similar but now the missing edge is among the $s(n-s)$ cross edges, any of the $C_{s,m}$ subgraphs on the s vertices is acceptable, and the probability that a subgraph is isolated is adjusted to reflect that the number of cross edges is $s(n-s) - 1$:

$$
\begin{aligned}
A^1_{s,m} &= \binom{n}{s}s(n-s)C_{s,m}\int_{p=0}^{1} p^m(1-p)^{\binom{s}{2}-m+s(n-s)-1}dp \\
&= s(n-s)\binom{n}{s}C_{s,m}\frac{m!\left(\binom{s}{2}-m-1+s(n-s)\right)!}{\left(\binom{s}{2}+s(n-s)\right)!} \\
&= s(n-s)\frac{\binom{s}{2}+s(n-s)+1}{\binom{s}{2}-m+s(n-s)}C'_{s,m}.
\end{aligned}
$$

This formula properly evaluates to 0 for $s = n$, except that like the formula for A^2 it is invalid for the pair $s = n$, $m = \binom{n}{2}$ excluded from (16).

Now observe that

$$A_{s,m}^2 + A_{s,m}^1 = \left(\binom{s}{2} - m + s(n-s)\right) \frac{\binom{s}{2} + s(n-s) + 1}{\binom{s}{2} - m + s(n-s)} C'_{s,m}$$

$$= \left(\binom{s}{2} + s(n-s) + 1\right) C'_{s,m}.$$

The terms being canceled are 0 only in the case $m = \binom{s}{2}$ and $s = n$ excluded from (16). Thus

$$A_{s,m}^0 + A_{s,m}^2 + A_{s,m}^1 = \left(\binom{n-s}{2} + \binom{s}{2} + s(n-s) + 1\right) C'_{s,m} = \left(\binom{n}{2} + 1\right) C'_{s,m}.$$

Substituting this into (16), and writing $N = \binom{n}{2}$ and $E = \mathbf{E}(\mathrm{MST}_{K_n})$, we have

$$\sum_{e \in K_n} \mathbf{E}(\mathrm{MST}_{K_n \setminus e}) = \sum_{s=1}^{n} \sum_{m=s-1}^{\binom{s}{2} - \mathbf{1}(s=n)} (N+1) C'_{s,m} - N.$$

For $s = n$ we extend the sum to include $\binom{n}{2} = N$, subtracting out $(N+1)C'_{n,N}$ to correct:

$$= (N+1) \left(\sum_{s=1}^{n} \sum_{m=s-1}^{\binom{s}{2}} C'_{s,m} - C'_{n,N} \right) - N.$$

From (14) the double sum is $E + 1$, while $C_{n,N} = 1$ and thus $C'_{n,N} = 1/(N+1)$, so this is

$$= (N+1) \left(E + 1 - \frac{1}{N+1} \right) - N$$

$$= (N+1)E.$$

Going back to (2) we see that

$$\mathbf{E}(\mathrm{VCG}) = (N+1)E - (N-1)E = 2E$$

and this completes the proof of Theorem 3.

5. Assignment Problem

Let the weight of edge (i,j) be denoted $x_{i,j}$. Let X be the $n \times n$ matrix with entries $x_{i,j}$. Let Y_j, $j = 1, \ldots, n$, be the $(n-1) \times (n-1)$ matrices obtained from X by deleting row n and column j. Let T_k, $k = 0, 1, \ldots, n-1$, denote the minimum assignment costs of Y_j, $j = 1, \ldots, n$, **sorted into increasing order**. Nair, Prabhakar and Sharma [NPS05] proved that, for a matrix of i.i.d. exponential(1) variables, the increments $T_k - T_{k-1}$ are independent, and

(18) $$T_k - T_{k-1} \sim \exp(k(n-k)).$$

A minimum assignment of a (random) $n \times n$ matrix is given by a value in the "missing" n'th row and the minimum assignment in the complementary submatrix. If $x_{n,\pi(n)}$ belongs to the minimum assignment, its Vickrey bonus is the cost difference between this assignment and the smallest assignment using a different element in row n. We are thus interested in gap between the smallest and second-smallest values of

(19) $$x_{n,\sigma(j)} + T_j,$$

where

$$x_{n,j} \sim \text{exponential}(1)$$

are independent random variables and $\sigma(j)$ is a random permutation. This is the incentive cost of the nth row; the total Vickrey incentive cost is the sum of similar values for all rows, and thus the

total expected Vickrey incentive cost is n times the expectation of the difference of the minimum and second-minimum values of (19). The computation of this value is done next.

We explicitly compute the expected Vickrey incentive cost. Going back to (18) and (19) we now fix the values of $T_i = t_i$ for $i = 0, \ldots, n-1$ and define $t_n = \infty$. Let $Y_j = x_{n,\sigma(j)} + t_j$ and $Y_{(1)} < Y_{(2)}$ be the two smallest values of the Y_j. We want to evaluate the following integral

$$(20) \qquad \int_0^\infty x \mathbf{P}(Y_{(2)} = x) dx.$$

Assume $Y_{(1)} = Y_i$ and $Y_{(2)} = Y_j = x$. Then we must have $Y_i < x$, $Y_j = x$ and $Y_k > x$ for $k \in \{0, \ldots, n-1\} \setminus \{i,j\}$. Since $x > Y_i \geq t_i$ and $x = Y_j \geq t_j$ we must have $x \geq t_{\max\{i,j\}}$. We break up the integral (20) into integrals over $[t_l, t_{l+1}]$ where $\max\{i,j\} \leq l \leq n-1$. The integral (20) is then obtained by summing over all possible pairs i, j where $i \neq j$ giving

$$(20) = \sum_{i \neq j} \sum_{l=\max\{i,j\}}^n \int_{t_l}^{t_{l+1}} x \mathbf{P}(Y_i < x, Y_j = x, Y_k \geq x, k \in \{0, \ldots, n-1\} \setminus \{i,j\}) dx$$

$$= \sum_{i \neq j} \sum_{l=\max\{i,j\}}^n \int_{t_l}^{t_{l+1}} x(1 - e^{-(x-t_i)}) e^{-(x-t_j)} \prod_{k \in \{0,\ldots,n-1\}\setminus\{i,j\}} \mathbf{P}(Y_k \geq x) dx$$

$$= \sum_{i \neq j} \sum_{l=\max\{i,j\}}^n \int_{t_l}^{t_{l+1}} x(1 - e^{-(x-t_i)}) e^{-(x-t_j)} \prod_{k \in 0,\ldots,l \setminus\{i,j\}} e^{-(x-t_k)} dx$$

$$(21) \qquad = \sum_{i \neq j} \sum_{l=\max\{i,j\}}^n \int_{t_l}^{t_{l+1}} x \left(e^{-(lx-t_0-\ldots-t_{i-1}-t_{i+1}-\ldots t_l)} - e^{-((l+1)x-t_0-\ldots t_l)} \right) dx.$$

Now set $s_l = t_0 + \ldots + t_l$ and then the innermost integral in (21) can be evaluated as

$$\int_{t_l}^{t_{l+1}} x \left(e^{-(lx-s_l+t_i)} - e^{-((l+1)x-s_l)} \right) dx$$

$$= \left[-\frac{x}{l} e^{-(lx-s_l+t_i)} - \frac{1}{l^2} e^{-(lx-s_l+t_i)} + \frac{x}{l+1} e^{-((l+1)x-s_l)} + \frac{1}{(l+1)^2} e^{-((l+1)x-s_l)} \right]_{t_l}^{t_{l+1}}.$$

Notice that our double summation in (21) can be split into four parts each of the form

$$\sum_{i \neq j} \sum_{l=\max\{i,j\}}^n b_l a_i = \sum_{l=1}^{n-1} b_l \left(\sum_{i \neq j, i, j \leq l} a_i \right) = \sum_{l=1}^{n-1} b_l \left(\sum_{i=0}^l l a_i \right)$$

$$(22) \qquad = \sum_{l=1}^{n-1} l b_l \left(\sum_{i=0}^l a_i \right)$$

where $a_i = e^{-t_i}$ for the first two terms of (21) and $a_i = 1$ for the last two. We now evaluate each part of (21) separately, call them $I_1, \ldots, I_4$.

$$I_1 = \sum_{i \neq j} \sum_{l=\max\{i,j\}}^n \left[-\frac{x}{l} e^{-(lx-s_l+t_i)} \right]_{t_l}^{t_{l+1}}$$

$$= \sum_{l=1}^{n-1} \left[t_l e^{-(lt_l-s_l)} - t_{l+1} e^{-(lt_{l+1}-s_l)} \right] \left(\sum_{i=0}^l e^{-t_i} \right)$$

$$= \sum_{l=1}^{n-1} \left[t_l e^{-((l-1)t_l-s_{l-1})} - t_{l+1} e^{-(lt_{l+1}-s_l)} \right] \left(\sum_{i=0}^l e^{-t_i} \right)$$

$$= t_1 + \sum_{l=1}^{n-1} t_l e^{-(lt_l - s_{l-1})}$$

where by abuse of notation we have used the fact that $t_n e^{-((n-1)t_n - s_{n-1})} = 0$. We have also used the identity, under the assumption $u_n = 0$,

$$\sum_{l=1}^{n-1} (u_l - u_{l+1}) \sum_{i=0}^{l} v_i = u_1 v_0 + \sum_{l=1}^{n-1} u_l v_l.$$

The second part of (21) is

$$I_2 = \sum_{i \neq j} \sum_{l=\max\{i,j\}}^{n-1} \left[-\frac{1}{l^2} e^{-(lx - s_l + t_i)} \right]_{t_l}^{t_{l+1}}$$

$$= \sum_{l=1}^{n-1} \left[\frac{1}{l} e^{-(lt_l - s_l)} - \frac{1}{l} e^{-(lt_{l+1} - s_l)} \right] \left(\sum_{i=0}^{l} e^{-t_i} \right)$$

$$= \sum_{l=1}^{n-1} \left[\frac{1}{l} e^{-((l-1)t_l - s_{l-1})} - \frac{1}{l} e^{-(lt_{l+1} - s_l)} \right] \left(\sum_{i=0}^{l} e^{-t_i} \right)$$

$$= 1 + \sum_{l=1}^{n-1} \frac{1}{l} e^{-(lt_l - s_{l-1})} - \sum_{l=1}^{n-1} \frac{1}{l(l+1)} e^{-(lt_{l+1} - s_l)} \left(\sum_{i=0}^{l} e^{-t_i} \right)$$

$$= 1 + \sum_{l=1}^{n-1} \frac{1}{l} e^{-(lt_l - s_{l-1})} - \sum_{l=2}^{n-1} \frac{1}{l(l-1)} e^{-(lt_l - s_l)} \left(\sum_{i=0}^{l-1} e^{t_l - t_i} \right).$$

Note that in the final double summation we have replaced the index l by $l - 1$.
The third part of (21) is

$$I_3 = \sum_{i \neq j} \sum_{l=\max\{i,j\}}^{n-1} \left[\frac{x}{l+1} e^{-((l+1)x - s_l)} \right]_{t_l}^{t_{l+1}}$$

$$= \sum_{l=1}^{n-1} l \left[t_{l+1} e^{-((l+1)t_{l+1} - s_l)} - t_l e^{-((l+1)t_l - s_l)} \right]$$

$$= \sum_{l=1}^{n-1} l \left[t_{l+1} e^{-((l+1)t_{l+1} - s_l)} - t_l e^{-(lt_l - s_{l-1})} \right]$$

$$= -\sum_{l=1}^{n-1} t_l e^{-(lt_l - s_{l-1})}.$$

The final part of (21) is

$$I_4 = \sum_{i \neq j} \sum_{l=\max\{i,j\}}^{n-1} \left[\frac{1}{(l+1)^2} e^{-((l+1)x - s_l)} \right]_{t_l}^{t_{l+1}}$$

$$= \sum_{l=1}^{n-1} \frac{l}{l+1} \left[e^{-((l+1)t_{l+1} - s_l)} - e^{-((l+1)t_l - s_l)} \right]$$

$$= \sum_{l=1}^{n-1} \frac{l}{l+1} \left[e^{-((l+1)t_{l+1} - s_l)} - e^{-lt_l - s_{l-1})} \right]$$

$$= -\sum_{l=1}^{n-1} \frac{1}{l(l+1)} e^{-(lt_l - s_{l-1})}.$$

Notice that

$$I_1 + I_3 = t_1$$

$$I_2 + I_4 = 1 + \sum_{l=1}^{n-1} \frac{1}{l+1} e^{-(lt_l - s_{l-1})} - \sum_{l=2}^{n-1} \frac{1}{l(l-1)} e^{-(lt_l - s_l)} \left(\sum_{i=0}^{l-1} e^{t_l - t_i} \right).$$

In [NPS05] it is shown that the minimum Y_j has expected value

$$t_0 + 1 - \sum_{l=1}^{n-1} \frac{1}{l(l+1)} e^{-(lt_l - s_{l-1})}.$$

So the expected difference $D_n = I_1 + I_3 + I_2 + I_4 - \min Y_j$ is given by

$$(23) \qquad D_n = (t_1 - t_0) + \sum_{l=1}^{n-1} \frac{1}{l} e^{-(lt_l - s_{l-1})} - \sum_{l=2}^{n-1} \frac{1}{l(l-1)} e^{-(lt_l - s_l)} \left(\sum_{i=0}^{l-1} e^{t_l - t_i} \right).$$

Taking the expectation over the T_i's we get from [NPS05] and (18) that

$$\mathbf{E}[e^{-(lT_l - S_{l-1})}] = \mathbf{E}[e^{-\sum_{j=1}^{l} j(T_j - T_{j-1})}] = \prod_{j=1}^{l} \mathbf{E}[e^{-j(T_j - T_{j-1})}]$$

$$= \prod_{j=1}^{l} \frac{n-j}{n-j+1} = \frac{n-l}{n}.$$

Similarly we get for $i = 0, \ldots, l-1$

$$\mathbf{E}[e^{(T_l - T_i)} e^{-(lT_l - S_{l-1})}] = \mathbf{E}[e^{((T_l - T_{l-1}) + \ldots + (T_{i+1} - T_i)) - \sum_{j=1}^{l} j(T_j - T_{j-1})}]$$

$$= \mathbf{E}[e^{-\sum_{j=1}^{i} j(T_j - T_{j-1}) - \sum_{j=i+1}^{l}(j-1)(T_j - T_{j-1})}]$$

$$= \prod_{j=1}^{i} \mathbf{E}[e^{-j(T_j - T_{j-1})}] \prod_{j=i+1}^{l} \mathbf{E}[e^{-(j-1)(T_j - T_{j-1})}]$$

$$= \frac{n-i}{n} \prod_{j=i+1}^{l} \frac{(n-j)j}{(n-j+1)j - 1}.$$

Plugging this into (23) we get

$$D_n = \frac{1}{n-1} + \sum_{l=1}^{n-1} \frac{1}{l} \frac{n-l}{n} - \sum_{l=2}^{n-1} \frac{1}{l(l-1)} \sum_{i=0}^{l-1} \frac{n-i}{n} \prod_{j=i+1}^{l} \frac{(n-j)j}{(n-j+1)j - 1}.$$

We now show that $D_n \sim \frac{\pi^2}{6n}$. First note that we can write

$$\frac{n-i}{n} \prod_{j=i+1}^{l} \frac{(n-j)j}{(n-j+1)j-1} = \frac{n-i}{n} \prod_{j=i+1}^{l} \frac{(n-j)j}{(n-j+1)j} \cdot \frac{1}{1 - \frac{1}{(n-j+1)j}}$$

$$= \frac{n-l}{n} \prod_{j=i+1}^{l} \frac{1}{1 - \frac{1}{(n-j+1)j}}$$

$$= \frac{n-l}{n} \exp\left\{ \left(\sum_{j=i+1}^{l} \frac{1}{(n-j+1)j} + O\left(\frac{1}{j^2(n-j+1)^2}\right) \right) \right\}$$

$$= (1 + O(n^{-2}))\frac{n-l}{n} \exp\left\{ \sum_{j=i+1}^{l} \frac{1}{(n-j+1)j} \right\}$$

$$= \left(1 + O\left(\frac{\log^2 n}{n^2}\right)\right) \frac{n-l}{n} \left(1 + \sum_{j=i+1}^{l} \frac{1}{(n-j+1)j}\right)$$

$$= \frac{n-l}{n} \left(1 + \sum_{j=i+1}^{l} \frac{1}{(n-j+1)j} + O\left(\frac{\log^2 n}{n^2}\right)\right).$$

Using this approximation we can write the expected difference D_n as

$$\frac{1}{n-1} + \sum_{l=1}^{n-1} \frac{1}{l}\frac{n-l}{n} - \sum_{l=2}^{n-1} \frac{1}{l(l-1)} \sum_{i=0}^{l-1} \frac{n-l}{n} \left(1 + \sum_{j=i+1}^{l} \frac{1}{(n-j+1)j} + O\left(\frac{\log^2 n}{n^2}\right)\right)$$

$$= O\left(\frac{\log^3 n}{n^2}\right) + \frac{1}{n-1} + \sum_{l=2}^{n-1} \frac{1}{(l-1)n} - \sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n(n+1)} \sum_{i=0}^{l-1} \sum_{j=i+1}^{l} \left(\frac{1}{j} + \frac{1}{n-j+1}\right).$$

Now reversing the order of summation we can re-write the last summation as

$$\sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n} \sum_{j=1}^{l} \sum_{i=0}^{j-1} \frac{1}{n+1} \left(\frac{1}{n-j+1} + \frac{1}{j}\right)$$

$$= \sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n} \sum_{j=1}^{l} \frac{1}{n+1} \left(\frac{j}{n-j+1} + 1\right)$$

$$= \sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n} \sum_{j=1}^{l} \frac{1}{n-j+1}$$

$$= \sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n^2} + \sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n} \sum_{j=2}^{l} \frac{1}{n-j+1}.$$

The second summation can be re-written

$$\sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n} \sum_{j=2}^{l} \frac{1}{n-j+1}$$

$$= \sum_{j=2}^{n-1} \frac{1}{n-j+1} \sum_{l=j}^{n-1} \frac{n-l}{l(l-1)n}$$

$$= \sum_{j=2}^{n-1} \frac{1}{n-j+1} \sum_{l=j}^{n-1} \left(\left(\frac{1}{l-1} - \frac{1}{l} \right) - \frac{1}{(l-1)n} \right)$$

$$= \sum_{j=2}^{n-1} \frac{1}{n-j+1} \left(\frac{1}{j-1} - \frac{1}{n-1} - \sum_{l=j}^{n-1} \frac{1}{(l-1)n} \right)$$

$$= \sum_{j=2}^{n-1} \frac{1}{n} \left(\frac{1}{n-j+1} + \frac{1}{j-1} \right) - \sum_{j=2}^{n-1} \frac{1}{(n-j+1)(n-1)} - \sum_{j=2}^{n-1} \frac{1}{n-j+1} \sum_{l=j}^{n-1} \frac{1}{(l-1)n}$$

$$= \left(\frac{1}{n} - \frac{1}{n-1} \right) \sum_{j=2}^{n-1} \frac{1}{n-j+1} + \frac{1}{n} \sum_{j=2}^{n-1} \frac{1}{j-1} - \sum_{j=2}^{n-1} \frac{1}{n-j+1} \sum_{l=j}^{n-1} \frac{1}{(l-1)n}$$

$$= O\left(\frac{\log^3 n}{n^2} \right) + \sum_{j=2}^{n-1} \frac{1}{(j-1)n} - \sum_{j=2}^{n-1} \frac{1}{n-j+1} \sum_{l=j}^{n-1} \frac{1}{(l-1)n}.$$

Finally this gives the following expression for D_n:

$$O\left(\frac{\log^3 n}{n^2} \right) + \frac{1}{n-1} + \sum_{l=2}^{n-1} \frac{1}{(l-1)n} - \sum_{l=2}^{n-1} \frac{n-l}{l(l-1)n^2}$$

$$- \sum_{j=2}^{n-1} \frac{1}{(j-1)n} + \sum_{j=2}^{n-1} \frac{1}{n-j+1} \sum_{l=j}^{n-1} \frac{1}{(l-1)n}$$

$$= O\left(\frac{\log^3 n}{n^2} \right) + \frac{1}{n} \sum_{l=2}^{n-1} \frac{H_{n-1} - H_{n-l}}{l-1}.$$

The total expected incentive cost nD_n is then given by

$$(24) \qquad \sum_{l=2}^{n-1} \frac{H_{n-1} - H_{n-l}}{l-1} + O\left(\frac{\log^3 n}{n} \right).$$

Using $H_k = \gamma + \log k + O(\frac{1}{k})$ we can estimate the sum as

$$\sum_{l=2}^{n-1} \frac{\log\left(\frac{n-1}{(n-1)-(l-1)} \right)}{l-1} + O\left(\frac{\log^3 n}{n} \right)$$

$$= \sum_{l=2}^{n-1} \frac{1}{n-1} \frac{-\log(1 - \frac{l-1}{n-1})}{\frac{l-1}{n-1}}.$$

So as $n \to \infty$ the sum will approach the integral

$$(25) \qquad \int_0^1 \frac{-\log(1-x)}{x} dx.$$

Integrating the series expansion of $\frac{-\log(1-x)}{x}$ gives

$$\int_0^1 \frac{-\log(1-x)}{x} dx = \int_0^1 \sum_{k=1}^\infty \frac{x^{k-1}}{k} dx = \sum_{k=1}^\infty \frac{1}{k} \left[\frac{x^k}{k} \right]_0^1$$

$$= \sum_{k=1}^\infty \frac{1}{k^2} = \frac{\pi^2}{6}.$$

The switching of the limit and the integral can be justified using the monotone convergence theorem. Let $f_n(x) = \sum_{k=1}^{n} \frac{x^{k-1}}{k}$, then for $x \in [0,1)$, $\lim_{n\to\infty} f_n(x) = \frac{-\ln(1-x)}{x}$, and $f_1, f_2, \ldots$ form a non-decreasing sequence and thus $\int_0^1 \lim_{n\to\infty} f_n(x)dx = \lim_{n\to\infty} \int_0^1 f_n(x)dx$.

This completes the proof of Theorem 4.

6. Open questions

The same question can be raised for any combinatorial optimization problem with random weights. Natural candidates for consideration include a minimum spanning arborescence (rooted tree with all arcs oriented away from the root) in the complete digraph $\vec{K}_n$ (the directed analogue of MST result), a minimum-weight perfect matching in K_n (the non-bipartite analogue of Random Assignment), and the symmetric or asymmetric Traveling Salesman Problem (in K_n or $\vec{K}_n$ respectively). The natural edge weight distributions are i.i.d. uniform $[0,1]$ or i.i.d. exponential(1); asymptotically these will be equivalent, but one or other may be more convenient, and with luck one might give an exact (non-asymptotic) result like the one given here for MST. We conjecture that in these cases too the expected Vickery cost is twice the expected minimum cost, asymptotically. Assuming this pattern holds, it would be most interesting to understand why it is so.

References

[AB92]   Florin Avram and Dimitris Bertsimas, *The minimum spanning tree constant in geometrical probability and under the independent model: a unified approach*, Annals of Applied Probability **2** (1992), 113–130.

[Ald92]   David Aldous, *Asymptotics in the random assignment problem*, Pr. Th. Related Fields **93** (1992), 507–534.

[Ald01]   David J. Aldous, *The $\zeta(2)$ limit in the random assignment problem*, Random Structures Algorithms **18** (2001), no. 4, 381–418.

[AT02]   Aaron Archer and Éva Tardos, *Frugal path mechanisms*, Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, California), January 06-08 2002, pp. 991–999.

[AT07]   Aaron Archer and Éva Tardos, *Frugal path mechanisms*, ACM Trans. Algorithms **3** (2007), no. 1, Art. 3, 22. MR MR2301829 (2008b:68013)

[Cla71]   E.H. Clarke, *Multipart pricing of public goods*, Public Choice **8** (1971), 17–33.

[CR04]   Artur Czumaj and Amir Ronen, *On the expected payment of mechanisms for task allocation (extended abstract)*, Proceedings of the Fifth ACM Conference on Electronic Commerce, 2004, pp. 252–253.

[CS99]   Don Coppersmith and Gregory B. Sorkin, *Constructive bounds and exact expectations for the random assignment problem*, Random Structures Algorithms **15** (1999), no. 2, 113–144. MR 2001j:05096

[Elk05]   Edith Elkind, *True costs of cheap labor are hard to measure: edge deletion and VCG payments in graphs*, Proceedings of the Sixth ACM Conference on Electronic Commerce, 2005, pp. 108–117.

[ESS04]   Edith Elkind, Amit Sahai, and Ken Steiglitz, *Frugality in path auctions*, Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2004, pp. 694–702.

[FGS06]   Abraham Flaxman, David Gamarnik, and Gregory B. Sorkin, *First-passage percolation on a width-2 strip and the path cost in a VCG auction*, Proceedings of the Second International Workshop on Internet and Network Economics, WINE 2006, Patras, Greece, December 15-17, 2006 (Paul G. Spirakis, Marios Mavronicolas, and Spyros C. Kontogiannis, eds.), LNCS, vol. 4286, Springer, 2006, pp. 99–111.

[FPSS02] Joan Feigenbaum, Christos Papadimitriou, Rahul Sami, and Scott Shenker, *A BGP-based mechanism for lowest-cost routing*, PODC '02: Proceedings of the Twenty-First Annual Symposium on Principles of Distributed Computing (New York, NY, USA), ACM Press, 2002, pp. 173–182.

[Fri85]   Alan Frieze, *On the value of a random minimum spanning tree problem*, Discrete Applied Mathematics **10** (1985), 47–56.

[Gro73]   Theodore Groves, *Incentives in teams*, Econometrica **41** (1973), no. 4, 617–631.

[Jan99]   Svante Janson, *One, two, three $\log n/n$ for paths in a complete graph with random weights*, Combinatorics, Probability and Computing **8** (1999), 347–361.

[KN05]   David Karger and Evdokia Nikolova, *Brief announcement: on the expected overpayment of VCG mechanisms in large networks*, PODC '05: Proceedings of the 24th Annual ACM Symposium on Principles of Distributed Computing (New York, NY, USA), ACM Press, 2005, pp. 126–126.

[LW04]   Svante Linusson and Johan Wästlund, *A proof of Parisi's conjecture on the random assignment problem*, Probability Theory and Related Fields **128** (2004), 419–440.

[MP85]   M. Mézard and G. Parisi, *Replicas and optimization*, J. Physique Lettres **46** (1985), 771–778.

[MP86]   _____ , *Mean-field equations for the matching and the travelling salesman problems*, Europhys. Lett. **2** (1986), 913–918.

[MP87]   _____ , *On the solution of the random link matching problems*, J. Physique Lettres **48** (1987), 1451–1459.

[MPS03]   Milena Mihail, Christos Papadimitriou, and Amin Saberi, *On certain connectivity properties of the Internet topology*, FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (Washington, DC, USA), IEEE Computer Society, 2003, p. 28.

[NPS05]   Chandra Nair, Balaji Prabhakar, and Mayank Sharma, *Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures*, Random Structures Algorithms **27** (2005), no. 4, 413–444. MR MR2178256 (2006e:90050)

[NR99]   Noam Nisan and Amir Ronen, *Algorithmic mechanism design (extended abstract)*, Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing (Atlanta, Georgia, United States), 1999, pp. 129–140.

[NR01]   _____ , *Algorithmic mechanism design*, Games Econom. Behav. **35** (2001), no. 1-2, 166–196, Economics and artificial intelligence. MR MR1822468 (2002a:68146)

[Par98]   Giorgio Parisi, *A conjecture on random bipartite matching*, Physics e-Print archive, http://xxx.lanl.gov/ps/cond-mat/9801176, January 1998.

[Vic61]   William Vickrey, *Counterspeculation, auctions and competitive sealed tenders*, Journal of Finance **16** (1961), 8–37.

[Wäs]   Johan Wästlund, *An easy proof of the zeta(2) limit in the random assignment problem*, Electronic Journal of Probability, to appear.