An exact estimate of an algorithm for finding a maximum flow, applied to the problem "on representatives"[1]

Alexander V. Karzanov

It is known that a solution of the problem "on representatives" is reduced to finding a maximum flow in a network Q of a special structure, namely: there are nodes A_i, $i = 1, \dots, n$, of 1st level, which correspond to the rows of the (input) matrix and are connected with a source S by entering arcs, and there are nodes B_j, $j = 1, \dots, m$, of 2nd level, corresponding to the columns of the matrix and connected with a sink T by leaving arcs. There is an arc going from a node of 1st level to a node of 2nd level if and only if the corresponding entry of the matrix is one. All arcs have unit capacities.

In case of the generalized problem "on representatives" the arcs $\overrightarrow{SA_i}$, $i = 1, \dots, n$, have integer capacities a_i, and the arcs $\overrightarrow{B_jT}$, $j = 1, \dots, m$, have integer capacities b_j.

The algorithm of finding a maximum flow either gives a solution to the problem or declares the non-existence of a solution and outputs a maximum size selection (*incomplete solution*) bounded from above by the numbers a_i and b_j.

In the present paper we will show that the algorithm of finding a maximum flow due to E. Dinitz [1] solves the problem on representatives with time bound $Cp\sqrt{\min(n, m)}$, where p is the number of 1's in the matrix and C is a constant. The previously known bound is of the form $Cp \min(n, m)$.

For the generalized problem on representatives, one obtains time bound $Cp\sqrt{N}$, where $N := \sum_1^n a_i = \sum_1^m b_j$. The previous bound is $Cp(n + m)$ (see the presentation of E. Dinitz in the present issue). In case $N \simeq (n + m)^2$ the bounds coincide.

Dinitz' maximum flow algorithm iteratively constructs *shortest augmenting paths subgraphs*, called *manuals*, whose lengths (the distances from S to T) are monotone increasing. We will consider manuals C_k formed by the shortest augmenting paths from S (rather than double-sided) and define their lengths to be the distances from S to T as well.

In each manual, one finds a flow such that deleting the (unit-capacity) arcs in it makes the sink unreachable from the source. The time bound of a standard method of constructing such a flow in a manual, as well as in an arbitrary "combinatorial" network, is Cp (see the above-mentioned presentation of E. Dinitz), and this bound cannot be improved.

Our first goal is to estimate the number of manuals in the case of ordinary problem on representatives.

The first manual C_0 is just Q itself. Suppose that to the moment of constructing a manual C_k we have already obtained a flow of value M_{k-1}. According to the algorithm, the manual C_k is extracted from the network Q_{k-1} which in turn is obtained from Q

[1] Author's translation (preserving the original style and notation as much as possible) from: А.В. Карзанов, Точная оценка алгоритма нахождения максимального потока, примененного к задаче о "представителях", В кн.: *Вопросы кибернетики. Труды семинара по комбинаторной математике* (Москва, 1971), Советское Радио, Москва, 1973, с. 66–70. (A.V. Karzanov, Tochnaya otsenka algoritma nakhozhdeniya maksimal'nogo potoka, primenennogo k zadache "o predstavitelyakh", In: *Voprosy Kibernetiki. Trudy Seminara po Kombinatornoĭ Matematike* (Moscow, 1971), Sovetskoe Radio, Moscow, 1973, pp. 66–70, in Russian.)

by replacing the arcs occurring in the flow by the reverse arcs (and the arcs of the form $\overrightarrow{SA_i}$ or $\overrightarrow{B_jT}$ in the flow are deleted at all). Let us denote the number of such reverse arcs by q_k. Then $q_k = M_{k-1}$, since each reverse arc arising from $\overrightarrow{A_iB_j}$ corresponds to the path SA_iB_jT in the flow (and such paths meet only at S and T).

We refer to the set of nodes in C_k lying at distance r from S as the *layer* O_k^r (then $O_k^0 = \{S\}$).

Let the maximum flow value in the network Q is equal to M (obviously, $M \leq \min(n, m)$).

Consider an arc ℓ going from a layer O_k^{2t} to the next layer O_k^{2t+1}, $t = 1, \ldots, d$ (where the length of C_k is equal to $2d + 3$). By parity reasonings, one can see that ℓ is just a reverse arc in the network Q_{k-1}. Denote the number of arcs going from a layer O_k^r to the layer O_k^{r+1} by γ_k^r. We have $\sum_{t=1}^{d} \gamma_k^{2t} \leq q_k = M_{k-1}$.

On the other hand, since, by the construction of manuals, any arc of Q_{k-1} not contained in C_k but having both ends in C_k cannot go from a layer with a smaller number to a layer with a bigger one, the arcs from O_k^{2t} to O_k^{2t+1} constitute a cut in the network Q_{k-1}. Using the fact that the maximum flow value in Q_{k-1} is equal to $M - M_{k-1}$ and that the flow value does not exceed the minimum cut capacity [2], we obtain $\gamma_k^{2t} \geq M - M_{k-1}$, $t = 1, \ldots, k$ (in view of $k \leq d$).

This implies $M_{k-1} \geq \sum_{t=1}^{k} \gamma_k^{2t} \geq (M - M_{k-1}) \cdot k$, and finally, $k(M - M_{k-1}) \leq M$.

Obviously, $(M - M_{k-1})$ is a strictly decreasing function. While $k \leq \sqrt{M}$, the number of manuals is at most $\sqrt{M}$. And when k becomes greater than $\sqrt{M}$, we have $(M - M_{k-1}) < \sqrt{M}$, so the number of remaining manuals is less than $\sqrt{M}$ as well. Therefore, the total number of manuals is at most $\sqrt{M} \leq \sqrt{\min(n, m)}$, whence the required time bound $Cp\sqrt{\min(n, m)}$ follows.

In case of the generalized problem on representatives, if $\overline{N}$ is the maximum flow value in Q (so $\overline{N} \leq N$) and if N_{k-1} is the value of the flow obtained to the moment of constructing a manual C_k, then we obtain a similar inequality $N_{k-1} \geq \sum_{t=1}^{k} \gamma_k^{2t} \geq (\overline{N} - N_{k-1}) \cdot k$, whence the total number of manuals is at most $2\sqrt{\overline{N}}$ and the time bound is $Cp\sqrt{N}$.

Next we construct a network instance to the ordinary problem on representatives for which the algorithm achieves the time bound $Cn^{2.5}$ (for the generalized problem a construction is similar).

We start with constructing $C_2 = Q_1$. Let $O_1^0 := \{S\}$, $O_1^{\lfloor\sqrt{n}\rfloor+1} := \{T\}$, the layers $O_1^2, O_1^3, O_1^4, O_1^5$ contain n nodes each, and the layers O_1^1 and $O_1^6, \ldots, O_1^{\lfloor\sqrt{n}\rfloor}$ contain $\lfloor\sqrt{n}\rfloor$ nodes each. The nodes of layers O_1^2 and O_1^4 are connected to the nodes of layers O_1^3 and O_1^5, respectively, by arcs forming 1–1 correspondences (i.e. by matchings).

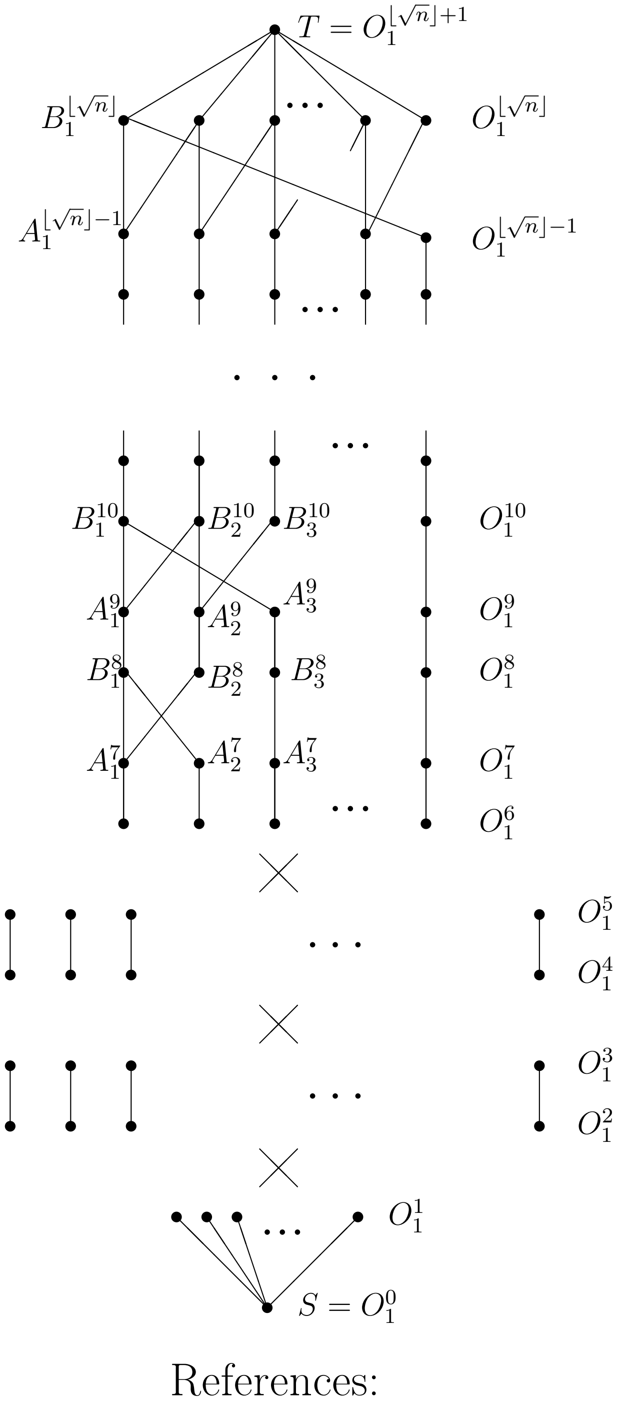
The layers $O_1^0, O_1^1, O_1^3, O_1^5, O_1^{\lfloor\sqrt{n}\rfloor}$ are connected to their next layers by all possible arcs. The remaining layers are connected to their next ones by the arcs as indicated in the picture.

One can check that there does exist an instance of the problem on representatives for which the network Q_1 (obtained after the first iteration) is exactly as described above. This network has Cn nodes. The maximum flow value in Q_1 is equal to $\lfloor\sqrt{n}\rfloor$.

Choose the path $\xi_1 = S \ldots A_1^7 B_1^8 A_1^9 B_1^{10} \ldots A_1^{\lfloor\sqrt{n}\rfloor-1} B_1^{\lfloor\sqrt{n}\rfloor} T$ (on the iteration handling C_2). Deleting the arcs of this path makes T unreachable from S, and we come to the next network Q_2 and manual C_3 having the length $\lfloor\sqrt{n}\rfloor + 3$. Choose in C_3 the path $\xi_2 = S \ldots A_2^7 B_1^8 A_1^7 B_2^8 A_2^9 B_2^{10} \ldots A_2^{\lfloor\sqrt{n}\rfloor-1} B_2^{\lfloor\sqrt{n}\rfloor} T$; deleting its arcs makes T unreachable from S.

Continuing the process, we obtain a sequence of manuals $C_2, C_3, \ldots, C_{\lfloor\sqrt{n}\rfloor} - 1$ such that each contains a path having the above-mentioned property.

This gives the desired lower bound $Cn^2 \cdot \sqrt{n} = Cn^{2.5}$.



References:

[1] E.A. Dinic, An algorithm for solution of a problem of maximum flow in a network with power estimation, *Doklady Akademii Nauk SSSR*, vol. 194, No. 4, 1970, in Russian.

[2] L. Ford and D. Fulkerson, *Flows in Networks*, Mir, Moscow, 1966 (Russian transl.).