

Supplementary material for KDD 2009 submission: Meme-tracking and the Dynamics of the News Cycle

Jure Leskovec, Lars Backstrom, Jon Kleinberg

We now show that DAG Partitioning is computationally intractable to solve optimally.

Proposition 1. *DAG Partitioning is NP-hard.*

Proof. In more detail, given an instance of Multiway Cut, specified by H , T , and W , we create an equivalent instance of DAG Partitioning as follows. First, let n be the total number of nodes in H , and let $T = \{t_1, t_2, \dots, t_k\}$. We define the graph G for DAG Partitioning to have roots r_1, \dots, r_k , as well a node u_i corresponding to each node v_i of H (including the terminals), and a node u_{ij} corresponding to each edge (v_i, v_j) of H . For each node u_{ij} corresponding to the edge (v_i, v_j) , we create edges (u_{ij}, u_i) and (u_{ij}, u_j) . For each node u_i corresponding to a terminal t_j of H , we create the edge (u_i, r_j) ; and for each node u_i not corresponding to a terminal, we create edges (u_i, r_ℓ) for each $\ell = 1, 2, \dots, k$. Finally, if w_{ij} is the weight of edge (v_i, v_j) in H , we put a weight of w_{ij} on each of the edges (u_{ij}, u_i) and (u_{ij}, u_j) . We put a very large weight of X on each edge of the form (u_i, r_j) .

Now we ask: is there a solution to the DAG partitioning with a total edge weight of at most $W + (k - 1)(n - k)X$? We claim such a solution exists if and only if the original Multiway Cut instance had a solution of weight at most W . Indeed, we must delete at least $k - 1$ edges out of each node u_i not corresponding to a terminal, and we won't cut all k due to the large value of X . We can now define a partition of H by placing v_i in the set of nodes associated with terminal t_j if and only if (u_i, r_j) is the undeleted edge out of u_i in G . Now, given a node u_{ij} , if both u_i and u_j choose the same r_j for their undeleted edge, then we do not need to delete either edge out of u_{ij} ; but otherwise, we must delete one of them. Thus, the total amount of edge weight deleted out of nodes of the form u_{ij} is precisely the total edge crossing from one set in the partition of H to another. The other direction of the equivalence is proved similarly, using a partition of H to directly determine which edges out of the nodes u_i and u_{ij} to delete. ■