**Supplementary Material**

**I: Derivation for (6).**

, according to its definition in Case II, is

. (A-1)

Furthermore, . Then,

. (A-2)

is defined in (10).

Also,

. (A-3)

Expanding the block matrix multiplication within the trace and simplifying the result, we can get:

. (A-4)

is defined in (9). Inserting (A-4) and (A-2) into (A-1) and re-organizing the terms, (6) can be obtained.

**II: Proof of Theorem 1.**

(5) is a convex optimization, which can be solved by a Block Coordinate Descent (BCD) algorithm. We consider two coordinates in our problem, old domains as a whole and the new domain, respectively. Then, BCD works by alternately optimizing each coordinate. Specifically, at the -the iteration, , BCD solves the following two optimizations:

, (A-5)

. (A-6)

(A-5) is to optimize the new domain, , treating old domains as fixed by using estimates from the previous iteration, . (A-6) then optimizes the old domains, , treating the new domain as fixed by using the estimate from (A-5), .

The objective function in (5), i.e., , consists of a non-differentiable term, . According to the seminal work by Tseng (2001), when a convex objective function includes a non-differentiable term, BCD will converge to the optimal solution if the term is separable according to the coordinates. This is exactly our case, i.e., . Therefore, the BCD in (A-5) and (A-6) will converge to the global optimal solution (i.e., the solution to (5) in Case I). Furthermore, the convergence enjoys a monotone property (Tseng, 2001), i.e.,

**.** (A-7)

Let the initial values, , be the knowledge of old domains in Case II, i.e., . Then, (A-7) gives:

. (A-8)

Next, according to (A-5), is

. The second “=” follows from (6). is dropped in the last equation because it is a constant. Comparing (A-8) and (11), we get . Therefore, (A-8) becomes . When , it means that BCD attains the optimal solution in one coordinate (the old domains). Then, it must attain the optimal solution in the other coordinate (the new domain), i.e., . This completes the proof for Theorem 1.

**III: Proof of Theorem 2.**

Both (12) and (13) can be solved analytically, i.e., and Let and . Then, it can be derived that . Using and , we can show that . Therefore,

**.** (A-9)

In the last equation in (A-9), the cross-product, , is omitted. This is because , as an ordinary least squares estimator, is unbiased, and therefore . Continuing the derivation in (A-9), we can obtain:

**.** (A-10)

Perform an eigen-decomposition for , i.e., . is a diagonal matrix of eigenvalues . consists of corresponding eigenvectors. Then the in (A-10) can be shown to be:

and . (A-11)

Furthermore, let and denote the elements of by . Then, the last term in (A-10) can be shown to be:

. (A-12)

Inserting (A-12) and (A-11) into (A-10),

.

When , . To show that at some , we only need to show that there exists a such that for . To make , a sufficient condition is to make every term in the summation smaller than zero, i.e., , or equivalently, . This proves the existence of and thereby proves Theorem 1.

**IV: Proof of Theorem 3.**

According to (A-10), for a fixed , changes only with respect to . The smaller the , the smaller the . According to Definition 1, is the transfer learning distance . Therefore, the smaller the transfer learning distance, the smaller the . This gives

. (A-13)

Let and . Then, **.** The second inequality follows from (A-13). This completes the proof for Theorem 3.

**V: Proof of Theorem 4**

**Lemma 1**: The optimization in (17) is equivalent to:

(A-14)

*Proof*:To prove Lemma 1 is to prove . Start from the left-hand side. Write , where is a diagonal matrix of the nodes’ degrees, i.e., . is matrix of the edge weights, i.e., . The diagonal elements of are zero. Then,

**.** (A-15)

Plugging in the definition that , (A-15) becomes

. (A-16)

Because the graph is unidirectional, , (A-16) becomes

.

The can be absorbed by .

Next, we prove Theorem 4. Denote the objective function in (A-14) by , i.e.,

.

Because and are solutions to the optimization problem in (A-14) and they are non-zero, they should satisfy: and , i.e.,

, (A-17)

. (A-18)

Focusing on (A-17), the third term on the left-hand side can be written into:

. (A-19)

The last step follows from the given assumption that . Similarly, the third term on the left-hand side of (A-18) can be written into

. (A-20)

Considering (A-19) and (A-20) and taking the difference between (A-17) and (A-18), cancels with because it is known that , and we get:

. (A-21)

. (A-22)

Furthermore, we can get:

. (A-23)

We would like to have an upper bound that does not include . To achieve this, we adopt the following strategy: Because is the optimal solution, should be the smallest. Therefore, , i.e.,

. (A-24)

Then,

, and

. (A-25)

Inserting (A-25) into (A-23), we get

.

**VI: Obtaining (24) by the gradient method.**

Given , the optimization problem in (24) with respect to and is:

Using the gradient method, set the partial derivatives of to be zero:

,

i.e.,

, (A-26)

. (A-27)

From (A-26), we can get:

. (A-28)

Inserting (A-28) into the third term of (A-27) and through some algebra, we can get:

. (A-29)

According to (22), . Using this in (A-29),

**.**  (A-30)

Furthermore, according to (A-28),

**.**

Using (A-30) in the second term, we can get.

Finally, in order to prove that the and are optimal solutions for a minimization problem, we will need to show . “ ” denotes a matrix being positive definite. It can be derived that:

.

Furthermore,

,

where and . So .