➢ MI-MDTSB Biclustering Algorithm

✓ Consider S stocks observed over T time intervals over N trading days, with returns matrix X .

✓ For each trading day, our MI-MDTSB algorithm first clusters the rows (stocks) using our **average distance gradient change** (ADGC) algorithm (described below) [BC-Step 1].

✓ Within each cluster, our MI-MDTSB algorithm operates on X by
- deleting columns (time points of minimum length d) [BC--Step 2]
- deleting rows (stocks) based on MI based distance, [BC-Step 3]
- adding back rows appropriately [BC-Step 4].

✓ BC-Step 4 outputs the identified bicluster.

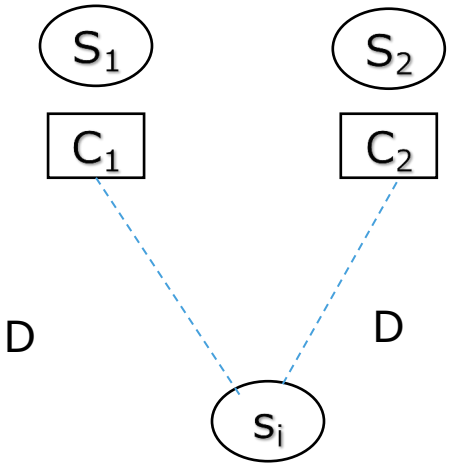
✓ α, β and γ are user defined thresholds.

➢ MI-MDTSB Biclustering Algorithm – BC Step 1: ADGC

- ✓ Cluster rows (stocks) using our average distance gradient change (ADGC) algorithm.
- ✓ ADGC detects a gradient change in a sequence of average distances of cluster seeds, and is an alternative to the well-known average silhouette criterion usually used for selecting number of clusters.
- ✓ The output is K^* clusters.
- ✓ The steps in the ADGC algorithm are illustrated below.

➢ ADGC Algorithm-Step 1. Select seeds for the first two clusters

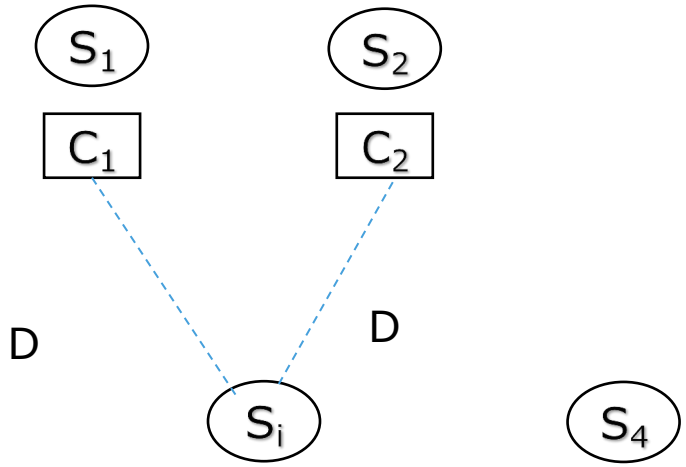
- Compute pairwise D = JMI distance for all pairs of stock returns.
- Find the pair with the largest distance, say S_1 and S_2. These become seeds for the first two clusters, C_1 and C_2.



- To prepare for Step 2, find the MI based distance between all other stocks with the seeds S_1 and S_2.

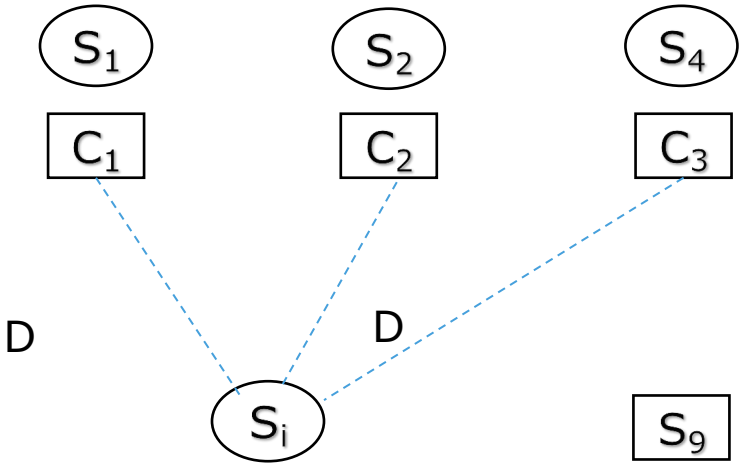
➢ ADGC Algorithm-Step 2. Determine number of clusters and corresponding seeds for successive clusters.

- Find the stock with the largest average distance from all current clusters (S1 and S2). Is S4 the third seed?



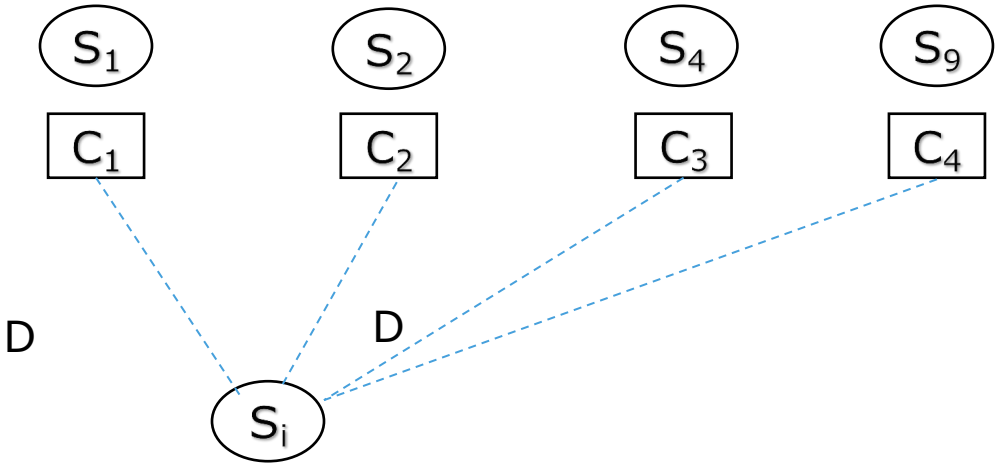
➢ ADGC Algorithm 2-2

S4 becomes a new seed for cluster C3.



➢ ADGC Algorithm 2-3

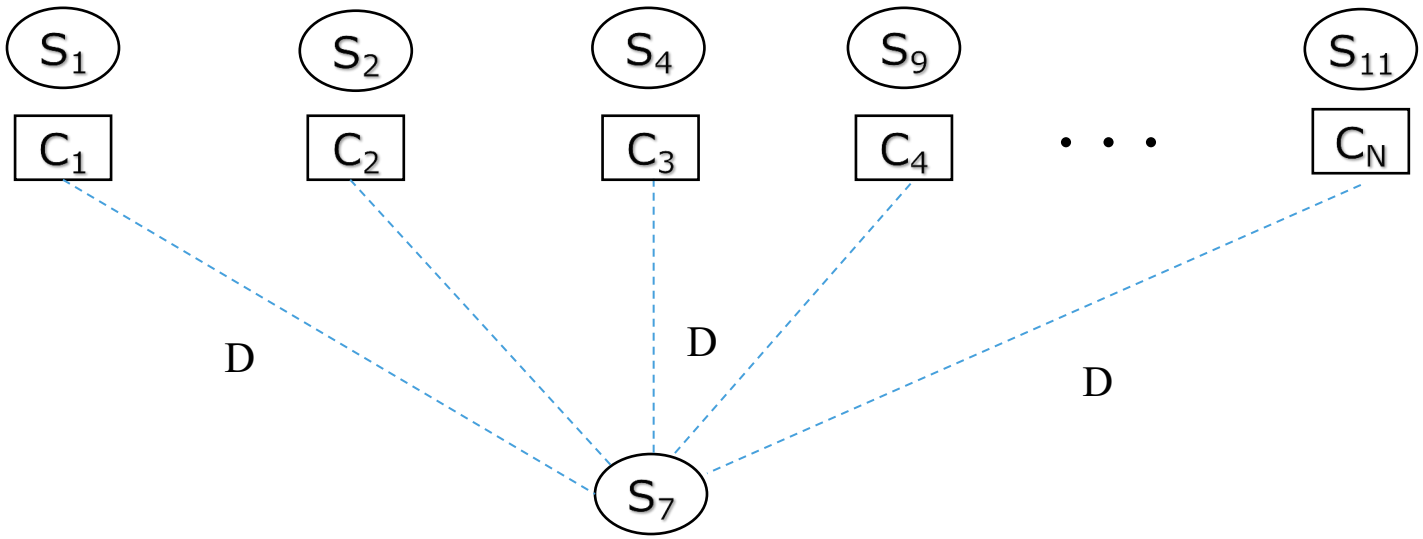
S9 becomes a new seed for cluster C4.



Stop when $\bar{D}(i-1) < \bar{D}_i$, optimal number of clusters K* is determined

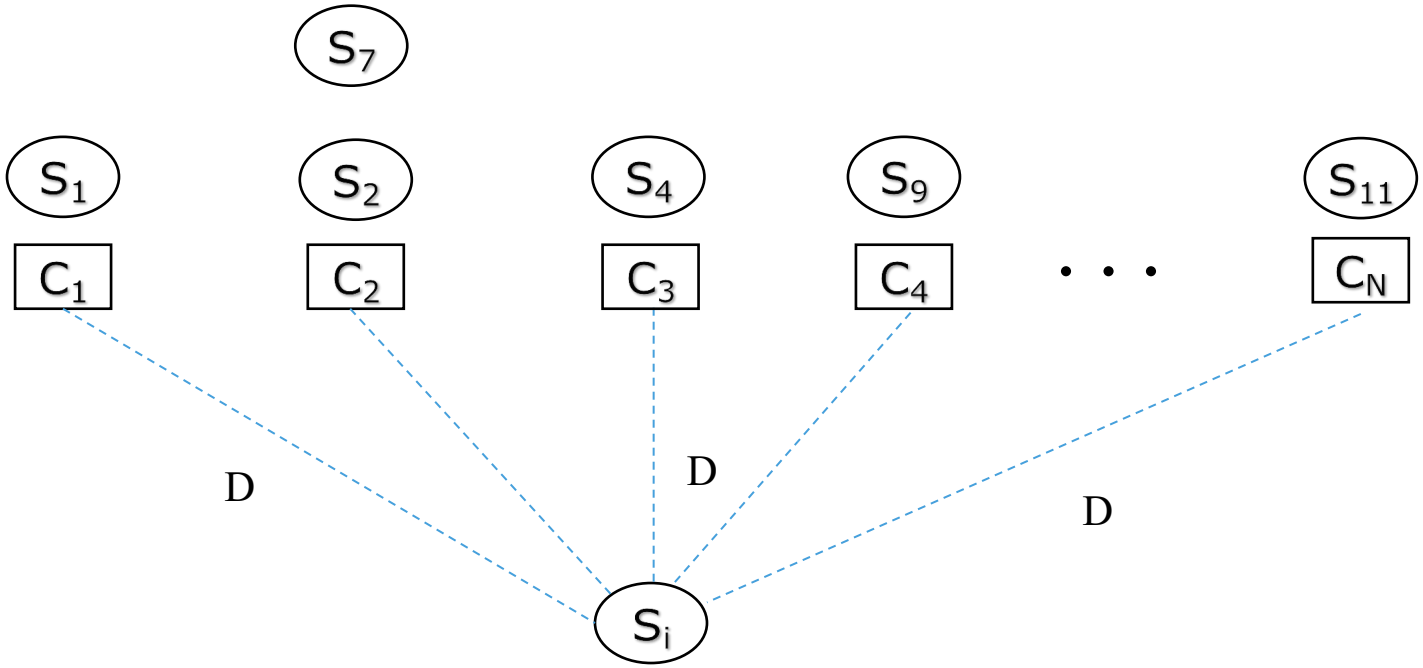
➢ ADGC algorithm Step 3: Allocate the stocks to K* clusters

- ✓ Assign a stock to the cluster from which it has the smallest distance.
- ✓ When a cluster has more than 1 unit, use the maximum of all distances.



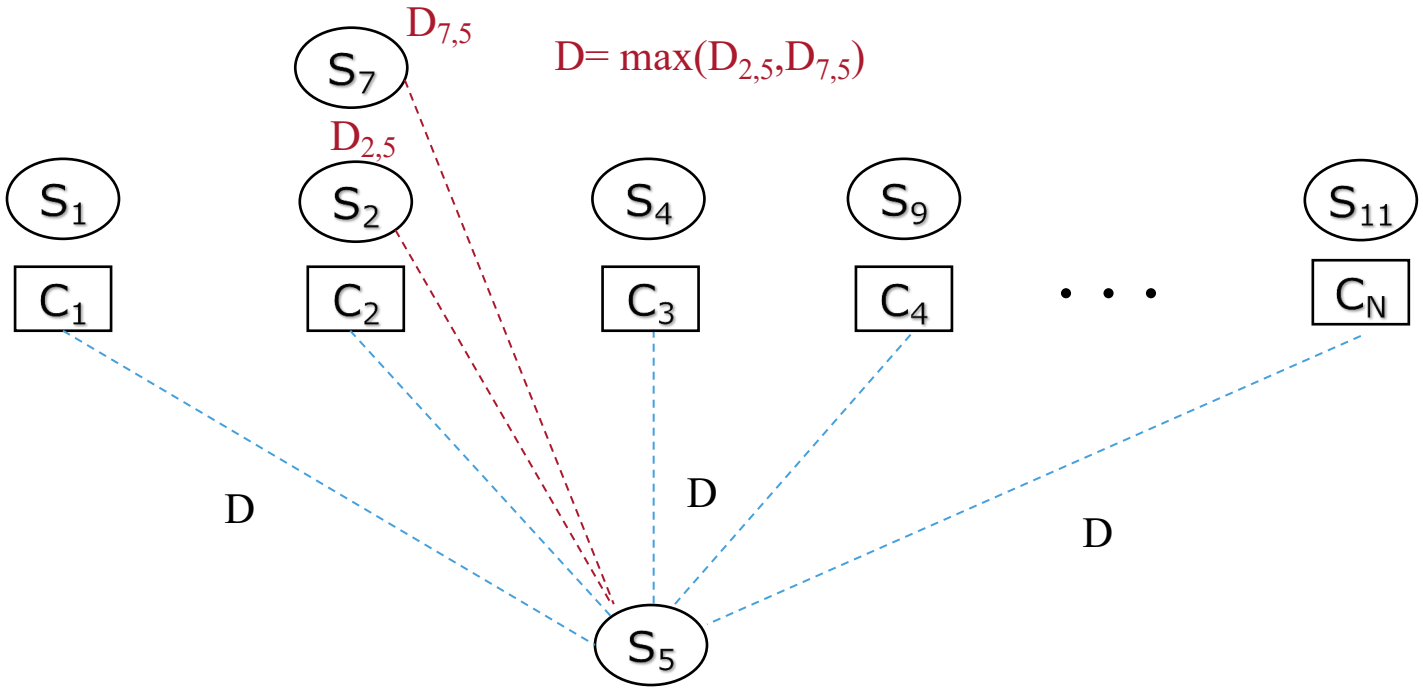
➢ ADGC Algorithm 3-2

- S7 is allocated to C2.



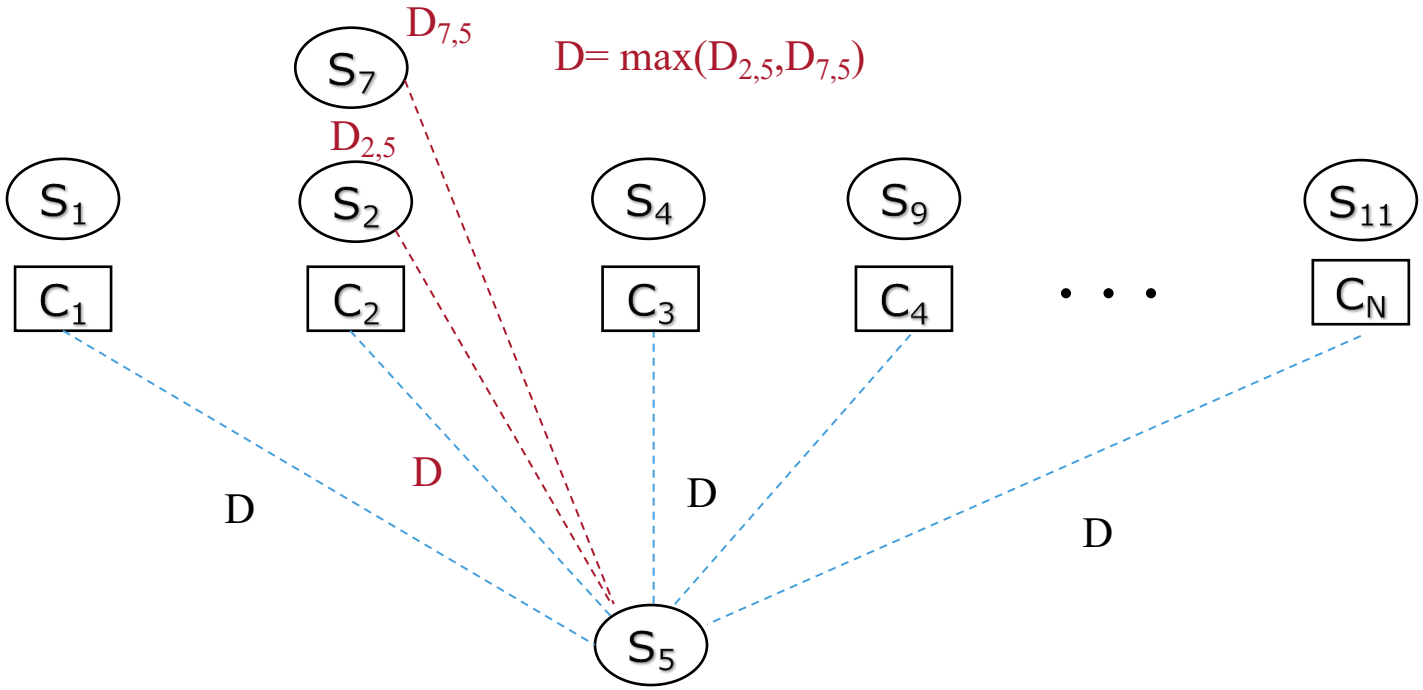
➢ ADGC Algorithm 3-3

- Calculate distance between a stock and a cluster.



➢ ADGC Algorithm 3-4

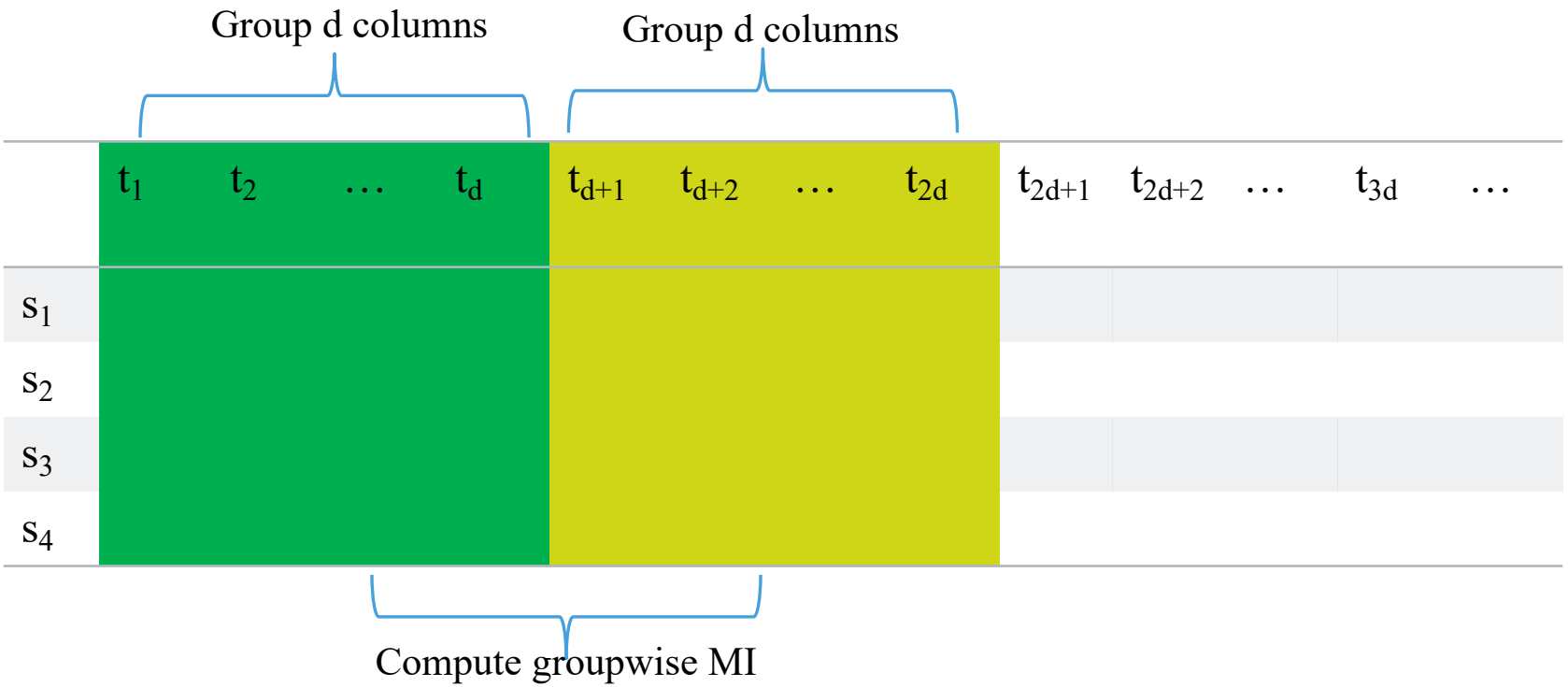
- Where should S5 go?



- The S stocks are now in K* clusters.

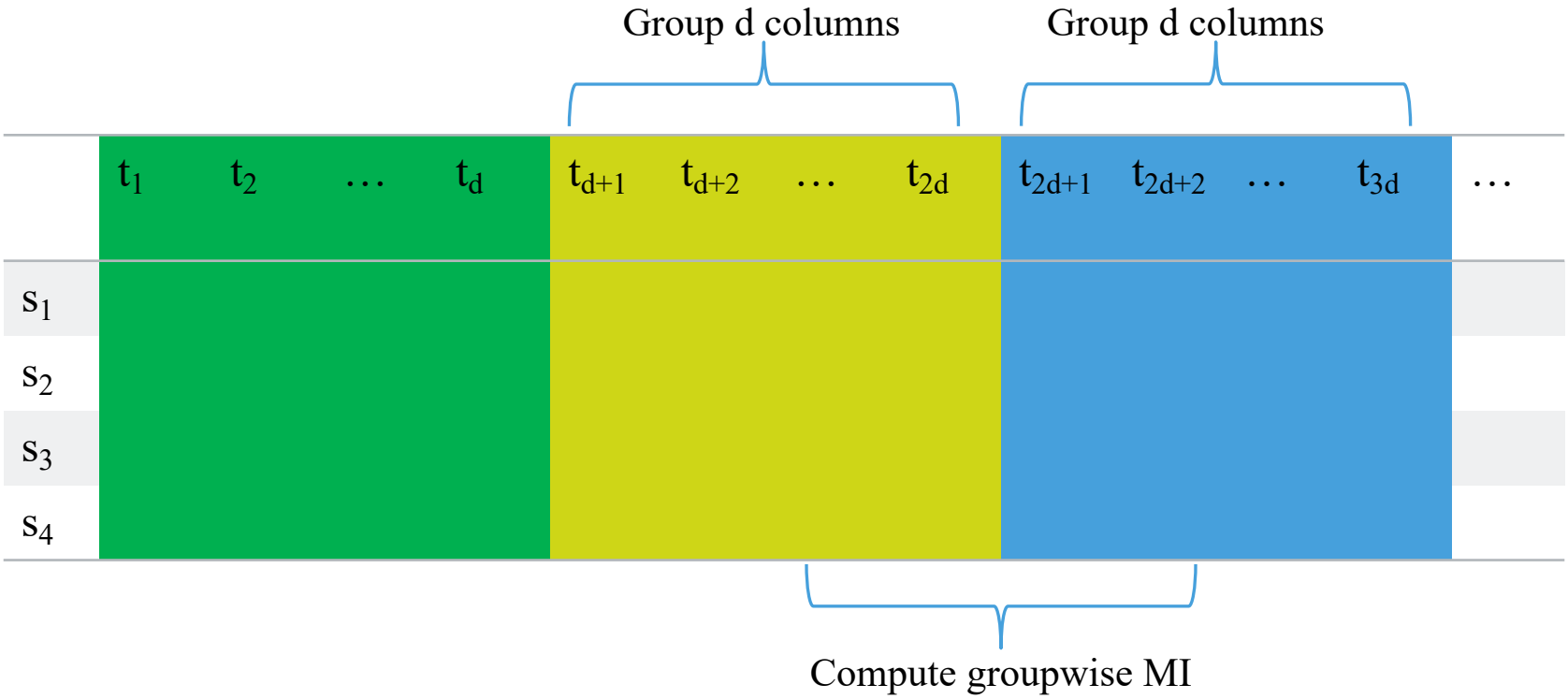
➢ BC-Step 2: Cluster columns/times within C1

- ✓ To preserve time contiguity, divide columns into groups of length d (say 10 mins)
- ✓ Measure the JMI distance between a group and the next group.
- ✓ Keep the groups separate if their JMI distance > alpha.



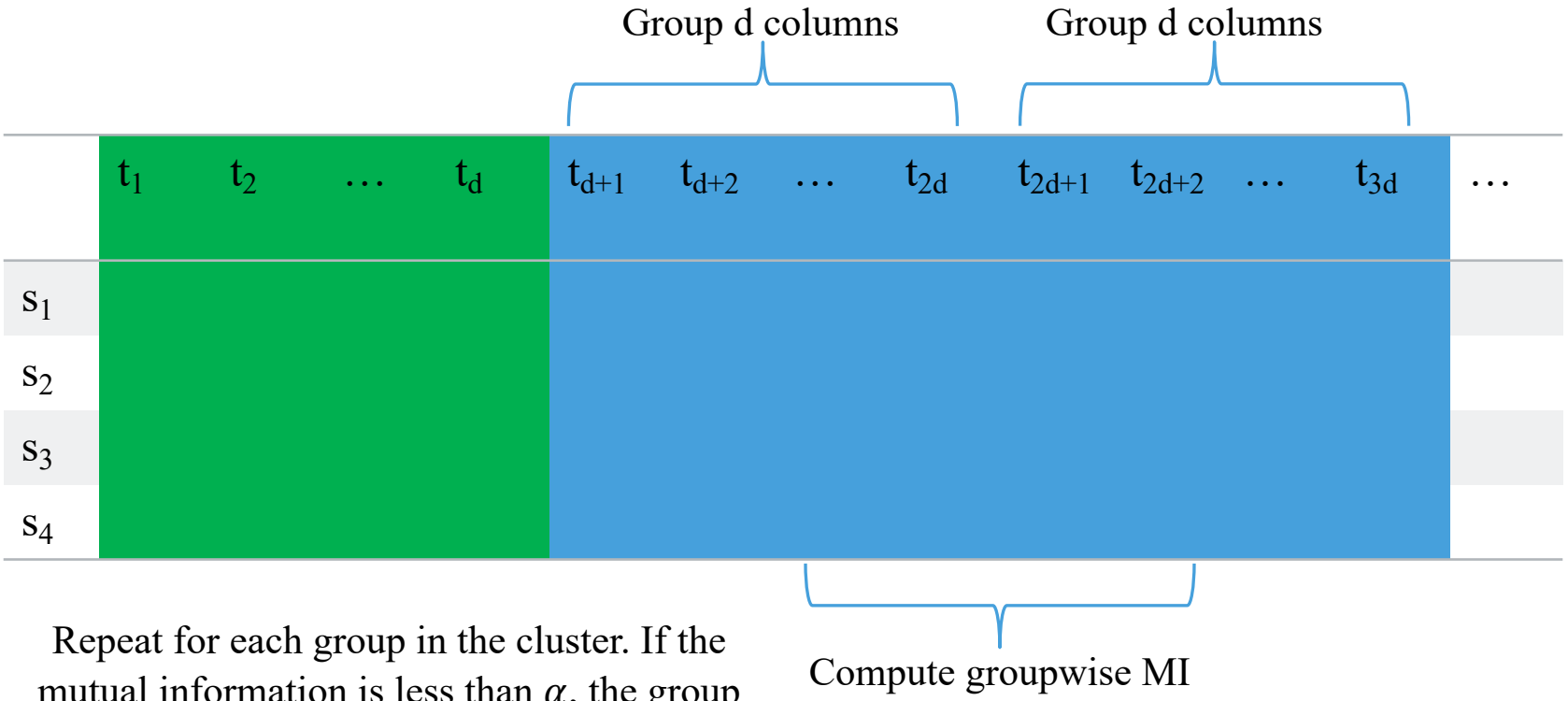
➢ Cluster columns/times BC 2-2

Repeat this for the next groups.



➢ Cluster columns/times: BC 2-3

Since the JMI distance <α, the two groups below are merged into a single group.



➢ BC-Step 3: Row deletion within C1 (which has stocks S1,S2,S3,S4)

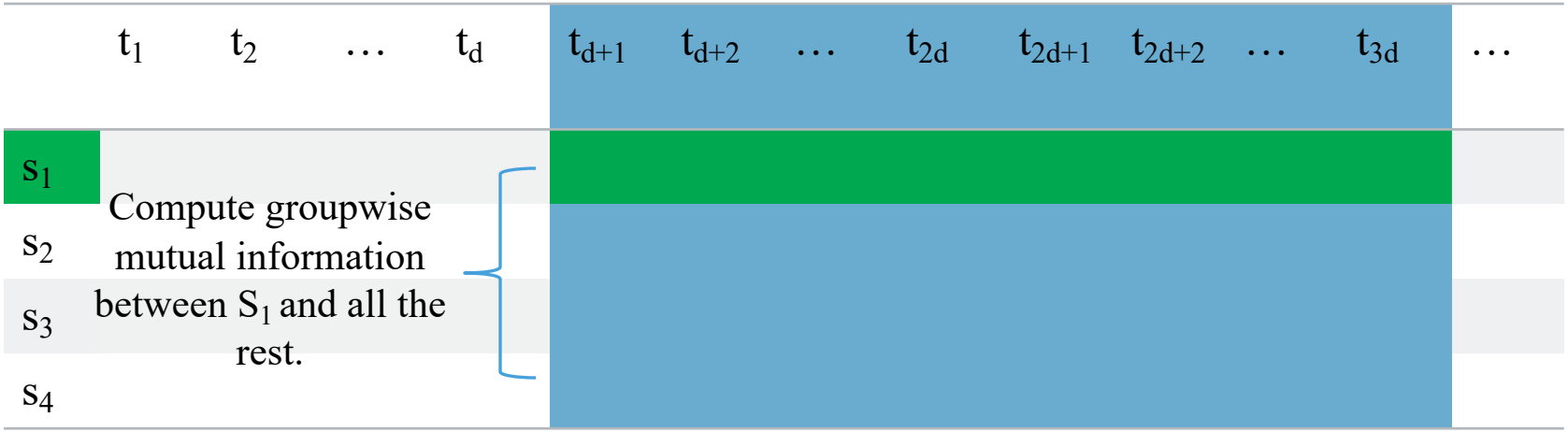
- ✓ Let $\mathbf{X}_{\kappa}^{1}$ be the outcome of BC-Step 2 (blue region).
- ✓ Only focus on time block $\mathbf{X}_{\kappa}^{1}$.
- ✓ For each row (stock) in $\mathbf{X}_{\kappa}^{1}$ measure the groupwise MI distance between this row and all other rows in $\mathbf{X}_{\kappa}^{1}$
- ✓ Keep the rows (stocks) if the row is close to all the other rows.

Keep columns from t_{d+1} to t_{3d} after the column clustering

	t_1	t_2	…	t_d	t_{d+1}	t_{d+2}	…	t_{2d}	t_{2d+1}	t_{2d+2}	…	t_{3d}	…
s_1													
s_2													
s_3													
s_4													

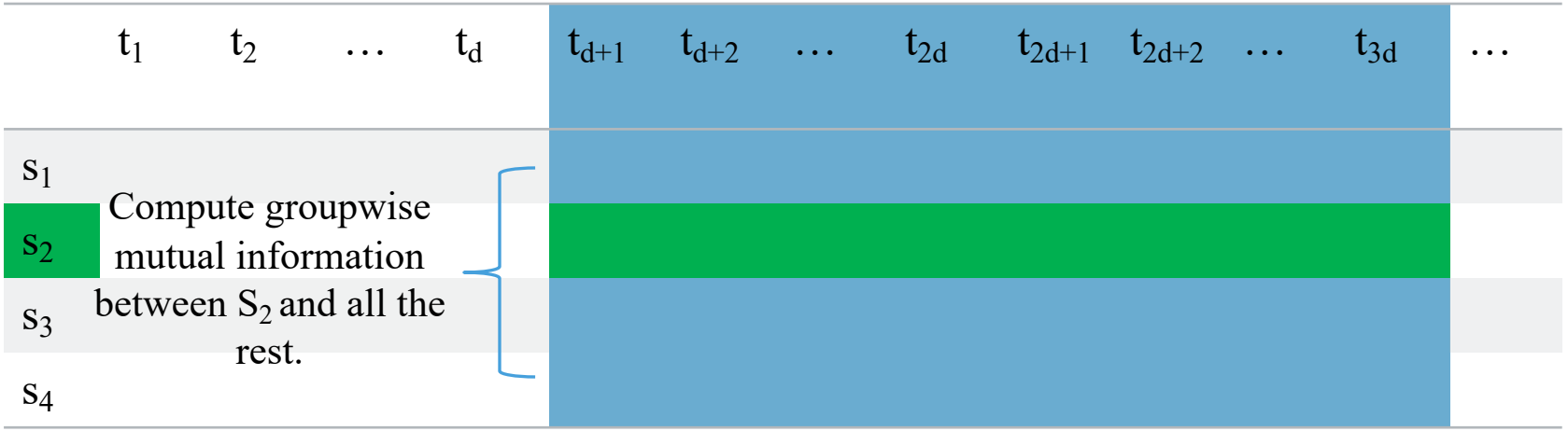
➢ Row deletion: BC 3-2

Keep columns from t_{d+1} to t_{3d} after the column clustering



➢ Row deletion: BC 3-3

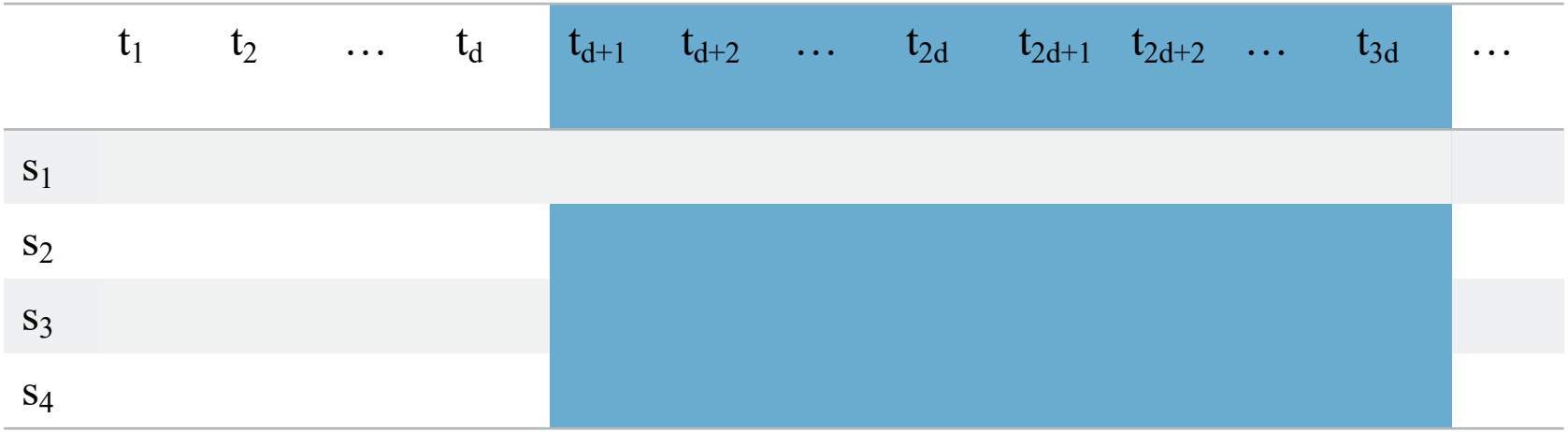
Keep columns from t_{d+1} to t_{3d} after the column clustering



Repeat for each stock in the cluster. If the mutual information is less than β, the stock is removed.

➢ Row deletion: BC 3-4

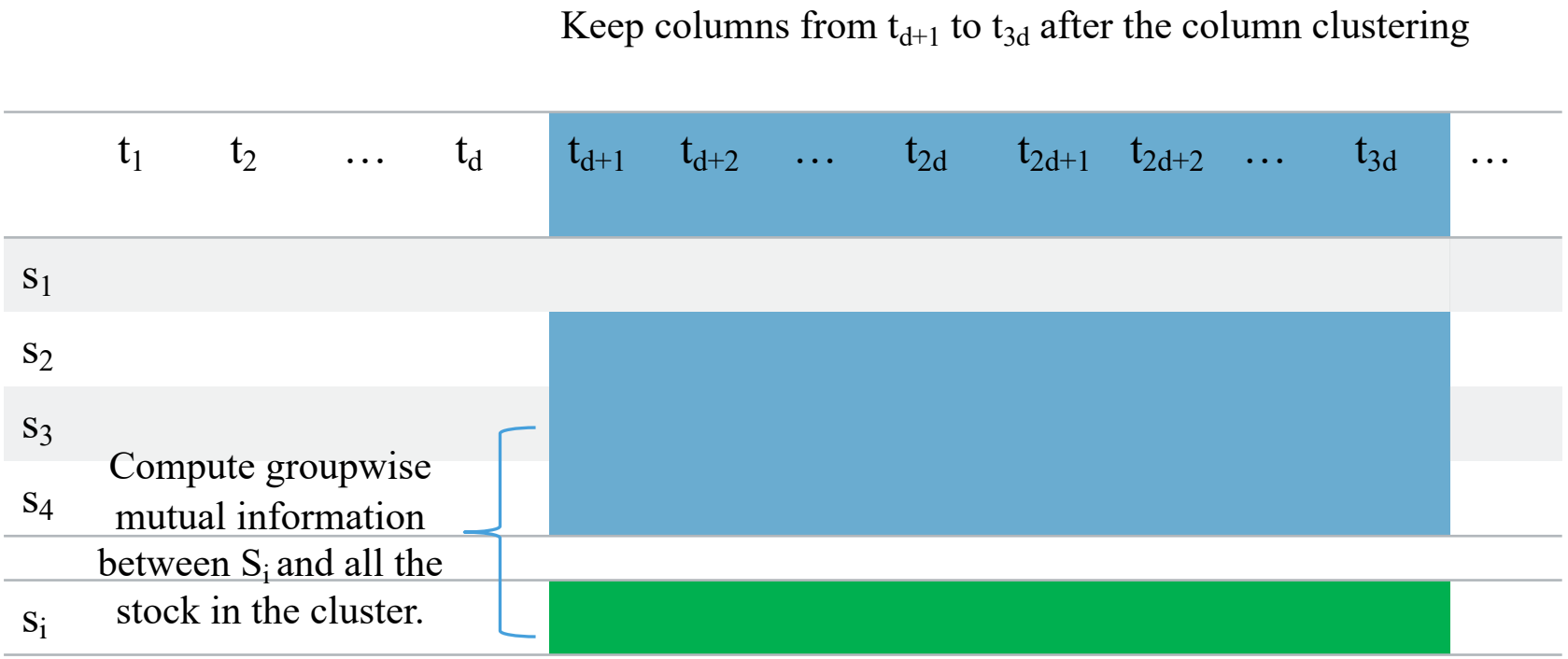
Keep columns from t_{d+1} to t_{3d} after the column clustering



Suppose stock S_1 is removed.

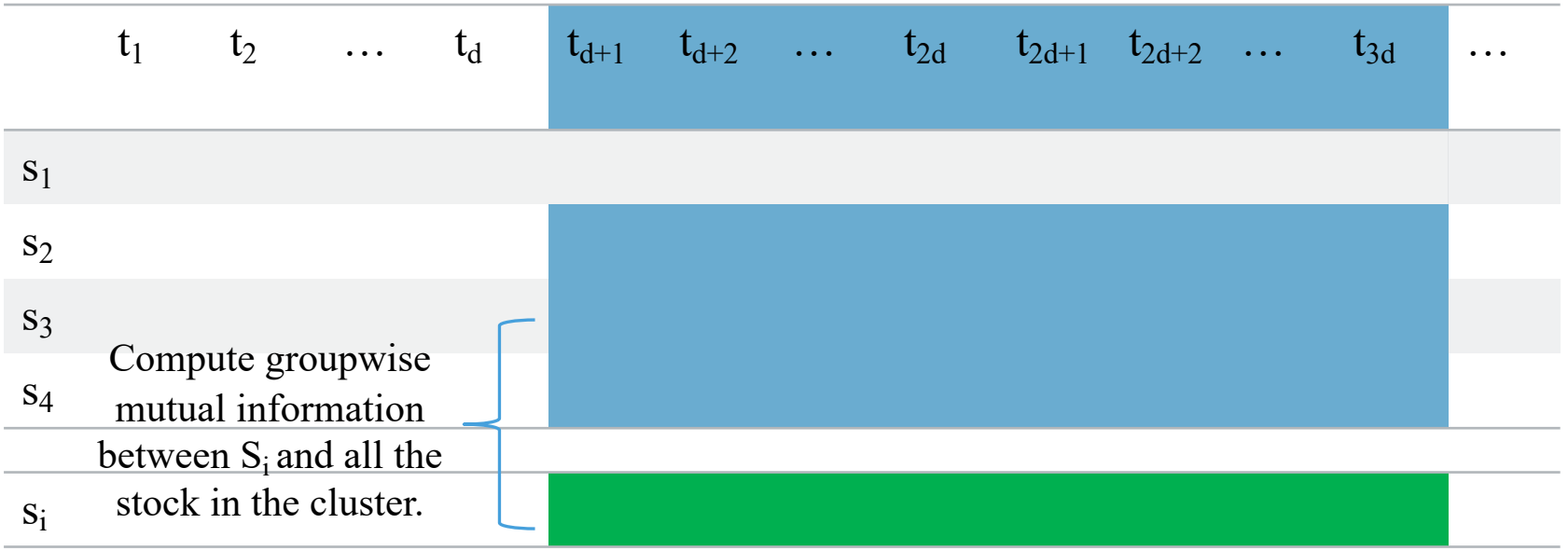
➢ BC-Step 4: Row insertion

- ✓ Let $\mathbf{X}_{\kappa}^{2}$ be the output of BC-Step 3.
- ✓ For each stock not in $\mathbf{X}_{\kappa}^{2}$, measure its distance to $\mathbf{X}_{\kappa}^{2}$.
- ✓ If a stock is close to $\mathbf{X}_{\kappa}^{2}$ we insert the stock to the rows of $\mathbf{X}_{\kappa}^{2}$



➢ Row insertion: BC 4-2

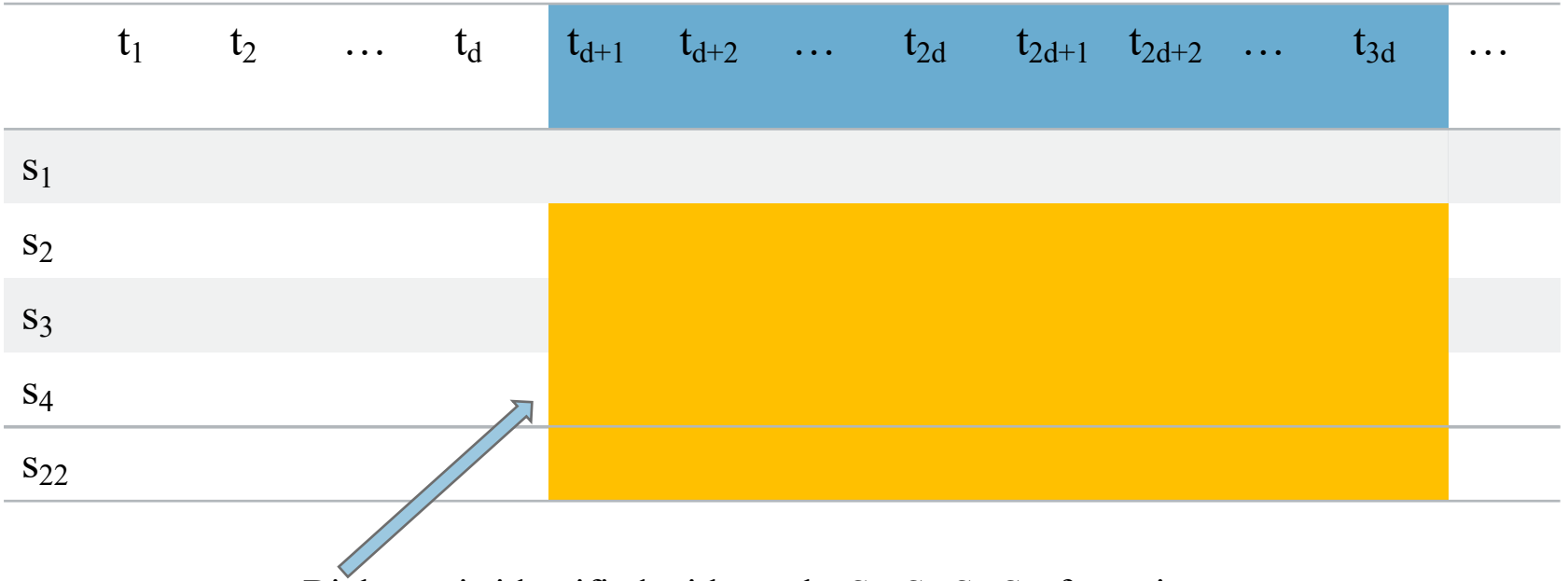
Keep columns from t_{d+1} to t_{3d} after the column clustering



Repeat for each stock S_i not in the cluster. If the mutual information is greater than γ, then S_i is inserted.

➢ Bicluster is identified

Keep columns from t_{d+1} to t_{3d} after the column clustering



Bicluster is identified with stocks S_2, S_3, S_4, S_{22} from time t_{d+1} to t_{3d}