Guided Tour of Machine Learning in Finance

Week 3: Unsupervised Learning

Minimum Spanning Trees, Kruskal algorithm, and equity correlation matrices

Igor Halperin

NYU Tandon School of Engineering, 2017

The Kruskal algorithm

Task:
Given a connected weighted graph with N nodes and connection weights d_{ij}, find a minimum spanning tree (MST) - a graph without loops connecting N nodes with $N-1$ links, such that the total weight of all edges is minimized.

Meaning: from $N(N-1)/2$ links, select $N-1$ shortest links that span all the nodes without forming loops.

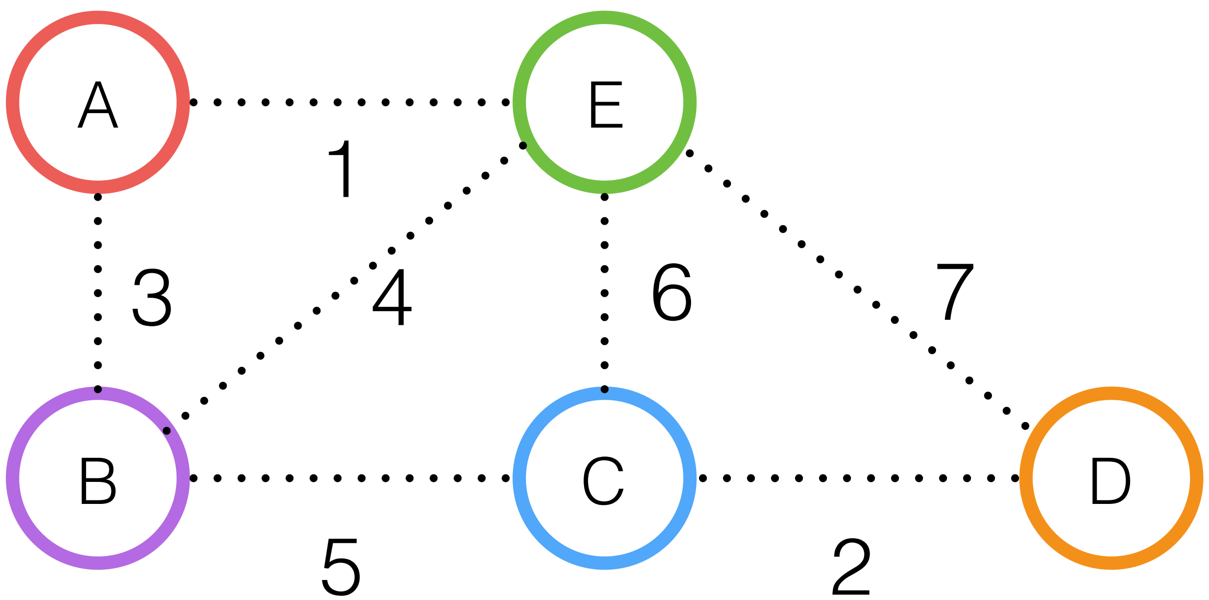
The Kruskal algorithm:

1. Find a minimum weight on a graph. Color it in any color (e.g. red)
2. Find the minimum uncolored edge that does not cross a colored or a red circuit. Mark this edge by a new color.
3. Repeat step 2 until connecting every vertex on the graph. The red edges form a MST.

The Kruskal algorithm

Wikipedia's example $(N = 5)$

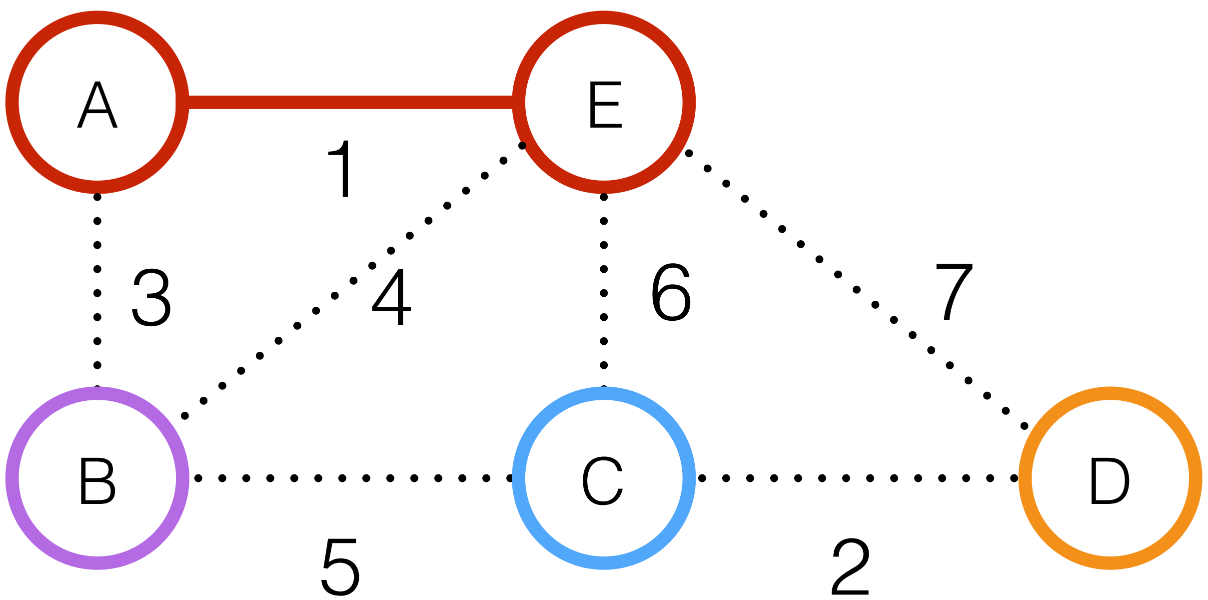
Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7



The Kruskal algorithm

Wikipedia's example $(N = 5)$

Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7

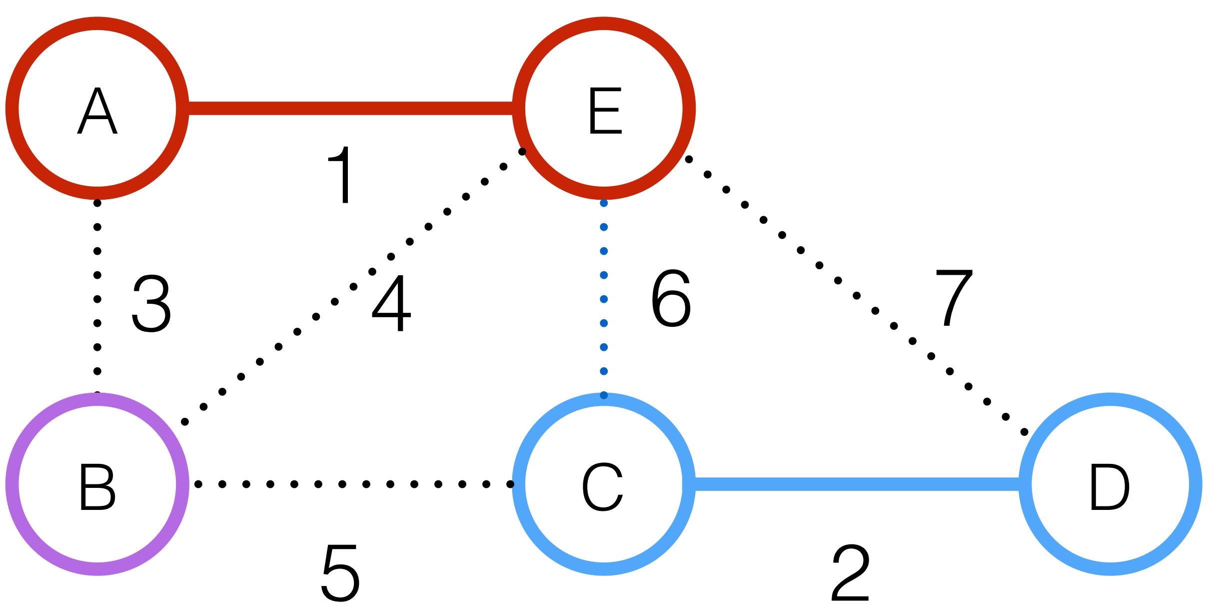


Select the smallest weight

The Kruskal algorithm

Wikipedia's example $(N = 5)$

Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7

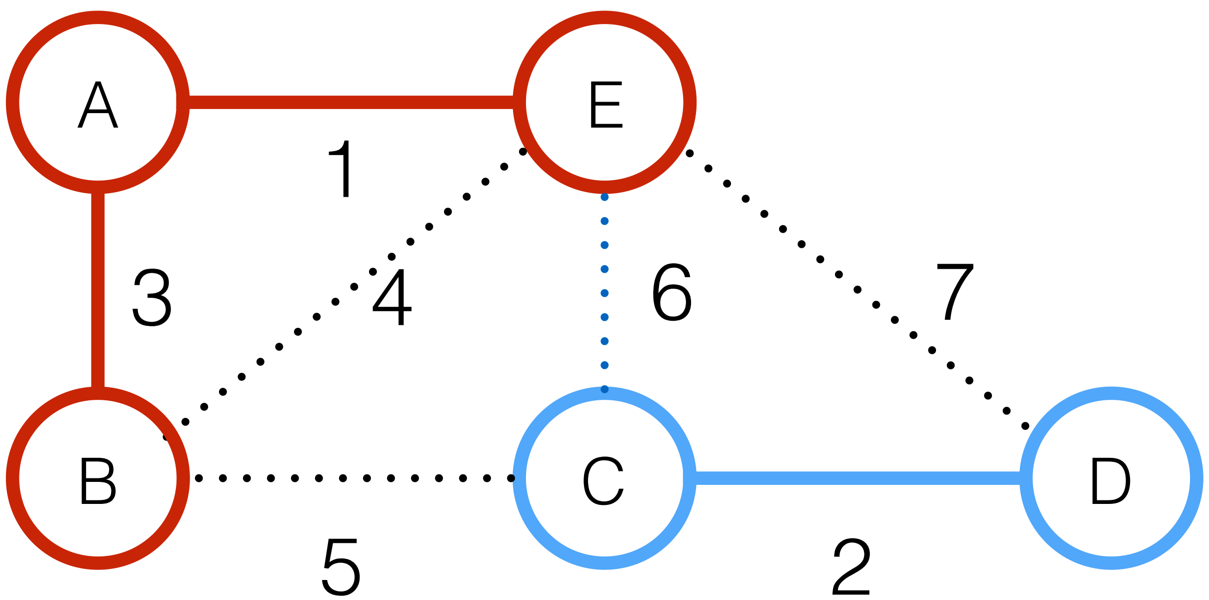


Select another smallest weight

The Kruskal algorithm

Wikipedia's example $(N = 5)$

Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7

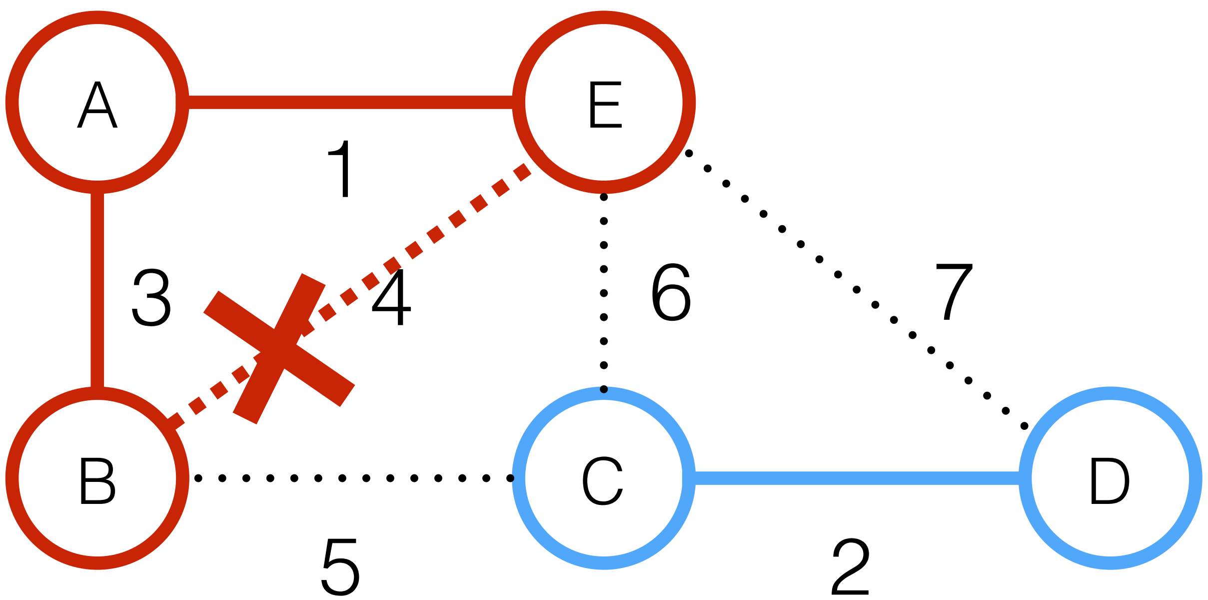


Select another smallest weight

The Kruskal algorithm

Wikipedia's example $(N = 5)$

Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7

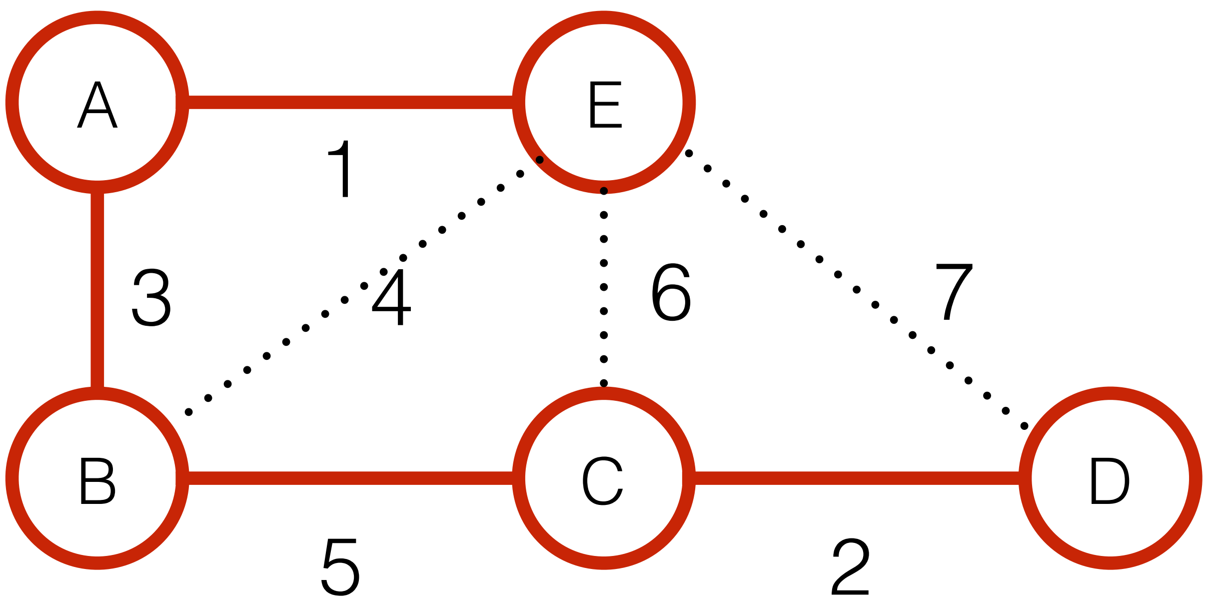


Selecting BE creates a loop!

The Kruskal algorithm

Wikipedia's example $(N = 5)$

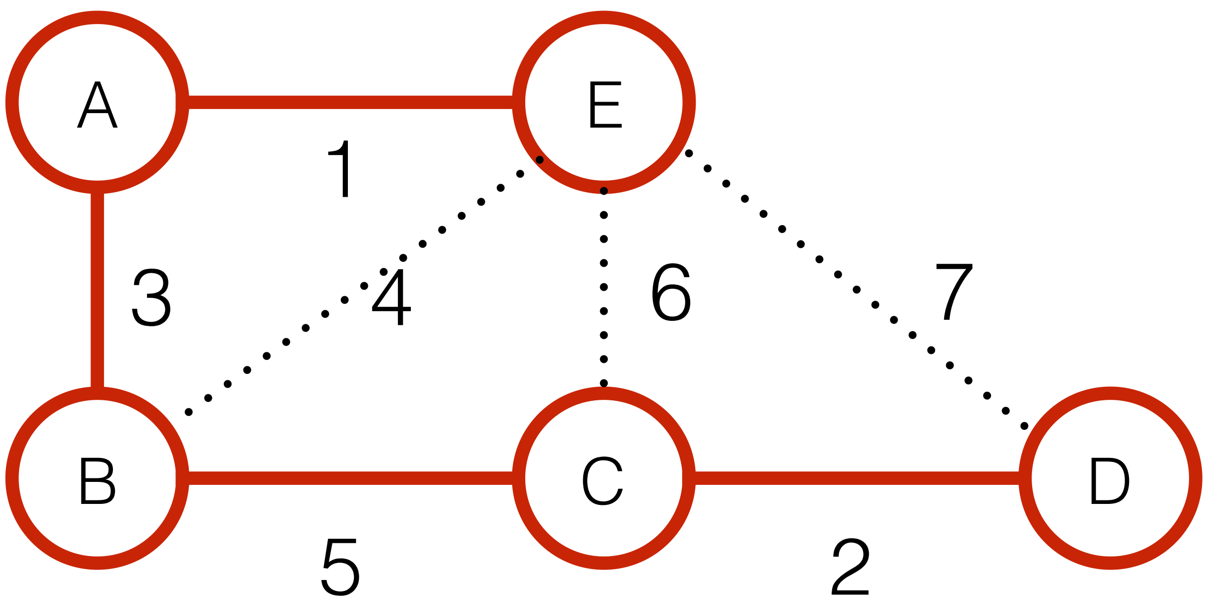
Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7



The Kruskal algorithm

Wikipedia's example $(N=5)$

Edge	AE	CD	AB	BE	BC	EC	ED
Weight	1	2	3	4	5	6	7



The minimum spanned tree is completed!

The MST compresses the data representation from $N(N-1)/2$ parameters to $N-1$ parameters

Clustering-based filtering

Use the MTS to de-noise the empirical distance matrix $D(\Delta t) = \left\{ d_{ij}(\Delta t) \right\}_{i,j=1}^{N}$ (and the correlation matrix $\mathrm{C}(\Delta \mathrm{t})$!) matrix by replacing

$$d_{ij} \rightarrow d_{ij}^{<}$$

Maximum distance d_{kl} among single steps in a path from i to j in a MST

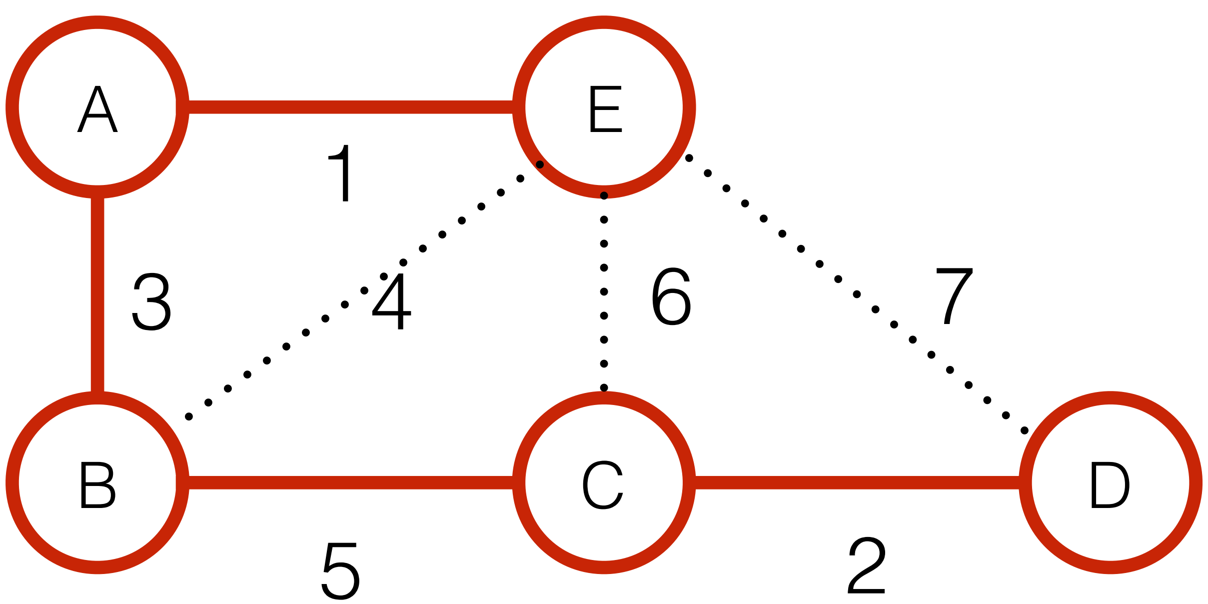
Clustering-based filtering

Use the MTS to de-noise the empirical distance matrix $D(\Delta t) = \left\{ d_{ij}(\Delta t) \right\}_{i,j=1}^{N}$ (and the correlation matrix $\mathrm{C}(\Delta \mathrm{t})$!) matrix by replacing

$d_{ij} \rightarrow d_{ij}^{<}$

Maximum distance d_{kl} among single steps in a path from i to j in a MST

Edge	AE	CD	AB	BE	BC	EC	ED
d_{ij}	1	2	3	4	5	6	7



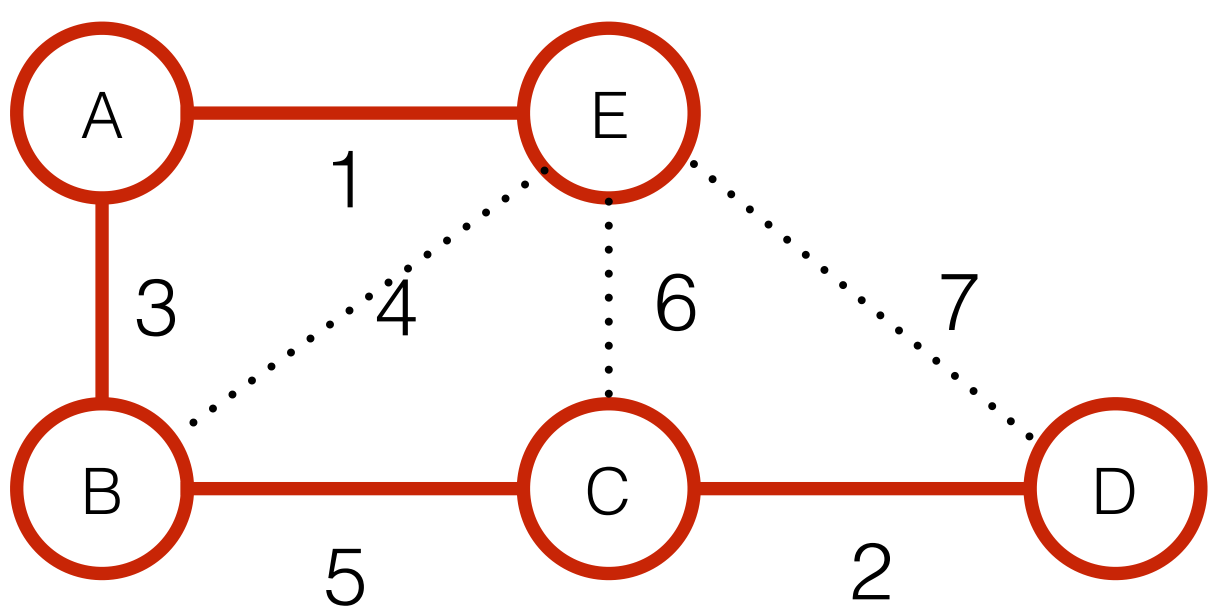
Clustering-based filtering

Use the MTS to de-noise the empirical distance matrix $D(\Delta t) = \left\{ d_{ij}(\Delta t) \right\}_{i,j=1}^{N}$ (and the correlation matrix $\mathrm{C}(\Delta \mathrm{t})$!) matrix by replacing

$d_{ij} \to d_{ij}^{<}$

Maximum distance d_{kl} among single steps in a path from i to j in a MST

Edge	AE	CD	AB	BE	BC	EC	ED
d_{ij}	1	2	3	4	5	6	7
$d_{ij}^{<}$	1	2	3	**3**	5	**5**	**5**



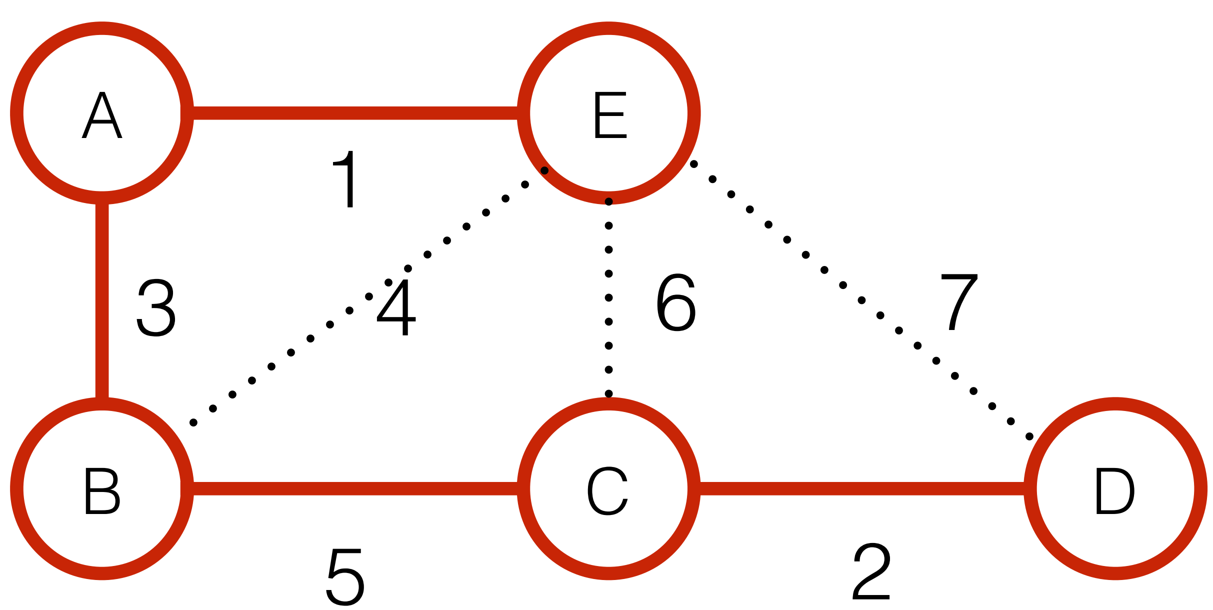
Clustering-based filtering

Use the MTS to de-noise the empirical distance matrix $D(\Delta t) = \left\{ d_{ij}(\Delta t) \right\}_{i,j=1}^{N}$ (and the correlation matrix $\mathrm{C}(\Delta \mathrm{t})$!) matrix by replacing

$$d_{ij} \rightarrow d_{ij}^{<}$$

Maximum distance d_{kl} among single steps in a path from i to j in a MST (the Subdominant Ultrametric Distance)

Edge	AE	CD	AB	BE	BC	EC	ED
d_{ij}	1	2	3	4	5	6	7
$d_{ij}^{<}$	1	2	3	**3**	5	**5**	**5**



Ultrametric distance: violates the triangle inequality of an Euclidean metric:

$$d_{ij} \leq \max_{k} \left\{ d_{ik}, d_{kj} \right\}$$

Subdominant Ultrametric - the largest ultrametric among those that are less or equal to d_{ij}

Application for portfolio optimization

Application for portfolio optimization (E. Pantaleo et al, “When Do Improved Covariance Matrix Estimators Enhance Portfolio Optimization? An empirical study of nine estimators”, 2010):

- Applied different filtering techniques to a sample covariance matrix of US stocks
- Compared realized risk for Markowitz-optimal portfolios obtained with different filtered covariance matrices

Application for portfolio optimization

Application for portfolio optimization (E. Pantaleo et al, “When Do Improved Covariance Matrix Estimators Enhance Portfolio Optimization? An empirical study of nine estimators”, 2010):

- Applied different filtering techniques to a sample covariance matrix of US stocks
- Compared realized risk for Markowitz-optimal portfolios obtained with different filtered covariance matrices

- Alternatives used:
    (i) Sample covariance matrix
    (ii) Random Matrix Theory (RMT) based filtering
    (iii) MST and other graph-based filtering

Application for portfolio optimization

Application for portfolio optimization (E. Pantaleo et al, “When Do Improved Covariance Matrix Estimators Enhance Portfolio Optimization? An empirical study of nine estimators”, 2010):

- Applied different filtering techniques to a sample covariance matrix of US stocks
- Compared realized risk for Markowitz-optimal portfolios obtained with different filtered covariance matrices

- Alternatives used:
  (i) Sample covariance matrix
  (ii) Random Matrix Theory (RMT) based filtering
  (iii) MST and other graph-based filtering

Results:

- Substantially lower-risk optimal portfolios for the regime $T / N > 1$ when short sales are allowed
- No significant improvements over the sample covariance matrix when short sales are not allowed, and/or $T / N < 1$

Control question

Select all correct answers

1. A Minimum Spanning Tree (MTS) constructs a graph with N nodes without loops using $N-1$ links, such that the total weight of all links is **maximized**.
2. A Minimum Spanning Tree (MTS) constructs a graph with N nodes without loops using $N-1$ links, such that the total weight of all links is **minimized**.
3. The MST compresses the parametrization of a correlation matrix from $N(N-1)/2$ parameters to $N-1$ parameters
4. Graph-based filtering of correlation matrices produce lower-risk optimal portfolios for the regime $T/N<1$ when short sales are not allowed
5. Graph-based filtering of correlation matrices produce lower-risk optimal portfolios for the regime $T/N>1$ when short sales are allowed

Correct answers: 2, 3, 5