Kruskal's Algorithm

In 1938 Burrows published the results of an extensive archaeological analysis of Polynesian islands. By examining cultural traits and complexes including artifacts (tools, canoe types, bark cloth, etc.), aspects of social organization (languages, kinship practices), and religious ideas, he identified 4 subgroups in Polynesia.

- Western Polynesia (Samoa, Tonga)
- Central Polynesia (Society Islands, Tuamotus, Southern Cooks, Australs, Rapa, Hawaii)
- Marginal Polynesia (New Zealand, Easter Island, Marqueses, Mangareva)
- Intermediate Polynesia (Northern Cooks)

In 1992 Irwin proposed that the level of similarity between Polynesian cultures is proportional to the ease of travel between them. In other words, the more remote two islands are, the less similar their cultures. The following matrix gives the mutual accessibility (as computed by Irwin) between the islands in the study.

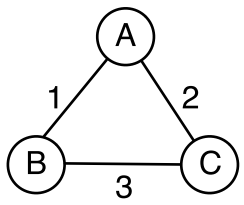
	TON	SAM	SCK	NCK	SOC	MRQ	TUA	MGR	AUS	RAP	HAW	EAS	NZ
TON		.44	.32	.18	.22	.18	.24	.09	.16	.10	.11	.03	.23
SAM	.44		.27	.23	.17	.15	.17	.10	.19	.11	.19	.05	.24
SCK	.32	.27		.23	.54	.30	.46	.14	.38	.18	.15	.08	.24
NCK	.18	.23	.23		.23	.15	.21	.11	.18	.14	.11	.05	.10
SOC	.22	.17	.54	.23		.41	.98	.22	.71	.21	.20	.09	.20
MRQ	.18	.15	.30	.15	.41		.54	.20	.34	.15	.13	.10	.12
TUA	.24	.17	.46	.21	.98	.54		.34	.63	.21	.21	.11	.24
MGR	.09	.10	.14	.11	.22	.20	.34		.23	.21	.08	.14	.08
AUS	.16	.19	.38	.18	.71	.34	.63	.23		.33	.19	.06	.21
RAP	.10	.11	.18	.14	.21	.15	.21	.21	.33		.05	.10	.09
HAW	.11	.19	.15	.11	.20	.13	.21	.08	.19	.05		.03	.09
EAS	.03	.05	.08	.05	.09	.10	.11	.14	.06	.10	.03		.00
NZ	.23	.24	.24	.10	.20	.12	.24	.08	.21	.09	.09	.00	

Assignment

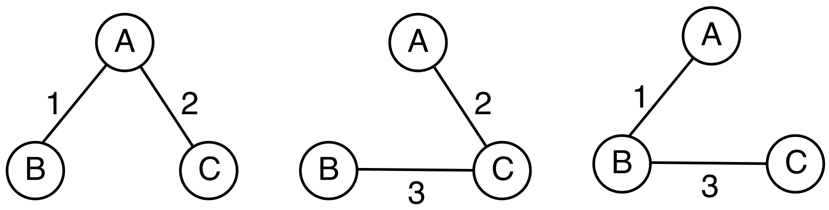
The assignment is to test Irwin's hypothesis by computing a maximum weight spanning tree for the accessibility of these 13 islands and compare the implied clusters with those given in Burrows' study. Additionally, produce a plot of the maximum weight spanning tree on a map of the Western Hemisphere suitable for publication.

Some Graph Theory

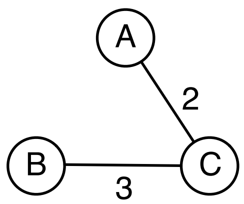
A graph is a set of nodes (or vertices) and edges. A network is a graph where the edges have been assigned weights. The following picture shows a network $\mathbb{G}$ with nodes A, B, and C, and edges AB (weight 1), AC (weight 2), and BC (weight 3).



A spanning tree is a connected subgraph containing no cycles. Here are the three possible spanning trees for $\mathbb{G}$.



The maximum weight spanning tree of $\mathbb{G}$ has the property that no other spanning tree has a larger sum of weights on its edges. Obviously the maximum weight spanning tree of $\mathbb{G}$ is the one where the edges sum to 5.



Kruskal's Algorithm

One method for computing the maximum weight spanning tree of a network $\mathbb{G}$ – due to Kruskal – can be summarized as follows.

1. Sort the edges of $\mathbb{G}$ into decreasing order by weight. Let T be the set of edges comprising the maximum weight spanning tree. Set $T = \emptyset$.
2. Add the first edge to T.
3. Add the next edge to T if and only if it does not form a cycle in T. If there are no remaining edges exit and report $\mathbb{G}$ to be disconnected.
4. If T has $n-1$ edges (where n is the number of vertices in $\mathbb{G}$) stop and output T. Otherwise go to step 3.

Suggested Procedure

1. Create a data frame with three columns: "Island 1", "Island 2", and "Accessibility". Each row should contain the accessibility between a pair of islands in the study. There should be 78 rows.
2. Create a vector named "Component" with one element for each island containing unique integer values (for instance `1:13`).
3. Use the `order` function to sort the rows into decreasing order (note: you will also have to use the subset function or the `[` operator).
4. Examine the edges (this should be done in a loop).
   (a) If "Island 1" and "Island 2" belong to the same component move on to the next edge.
   (b) Otherwise add the edge to the maximum weight spanning tree. Let `ic1` be the component of "Island 1" and `ic2` the component of "Island 2". For every island with component `max(ic1, ic2)`, set the component to `min(ic1, ic2)`.
   (c) Stop after adding 12 edges to the maximum weight spanning tree.
5. Use the R packages *maps* and *mapdata* to draw a map of the relevant region.
6. Draw the maximum weight spanning tree on top of your map. Use dashed lines for the k edges with the smallest weights (where k is a number you pick) to emphasize the clusters.
7. Comment on what you find.

	Latitude	Longitude
Tonga (TON)	−21.13	−175.20
Samoa (SAM)	−13.83	−171.83
South Cook (SCK)	−21.33	−160.27
North Cook (NCK)	−10.88	−165.82
Society (SOC)	−17.67	−149.50
Marquesas (MRQ)	−8.89	−140.13
Tuamotu (TUA)	−15.08	−145.87
Mangareva (MGR)	−23.12	−134.97
Australs (AUS)	−23.38	−149.45
Rapa (RAP)	−27.58	−144.33
Hawaii (HAW)	19.52	−155.51
Easter (EAS)	−27.12	−109.37
New Zealand (NZ)	−41.35	175.02