An Application of Minimum Spanning Trees to Travel Planning

Lakoa Fitina, John Imbal,
Vanessa Uiari, Nathaline Murki
& Elizabeth Goodyear

Abstract

Travel in Papua New Guinea is quite expensive compared to many of its neighbouring countries. On the other hand Papua New Guinea presents travelers with a tremendous array of sites and scenes that they would not find anywhere else in the world. This is the first of a series papers, resulting from collaboration between the departments of Tourism and Hospitality Management, and Mathematics and Computing Science at Divine Word University, on quantifying the issue of tourism in Papua New Guinea. In this paper we demonstrate the use of Prim's algorithm in graph theory to determine less expensive routes between nodes in a connected graph. We suggest an application to determining cheap transport routes in the country.

Key words: minimum spanning trees, travel planning, Prim's algorithm,

Introduction and definitions

The concept of a network of friends, of businesses, or of computers is pervasive in our society today. Tourism is a networked industry consisting of businesses, and personal relationships between companies and managers in businesses such as national tourism offices, hotels, attractions, transport, tours, travel agents and restaurants. It is this network of relationships that allows the tourism industry to deliver its product. It may be argued therefore that the tourism industry provides the ideal context for study of networks. In a series of papers we will be focusing on different aspects of networking in the tourism industry. In this paper we study networks that arise from flight scheduling and the costing of such flights.

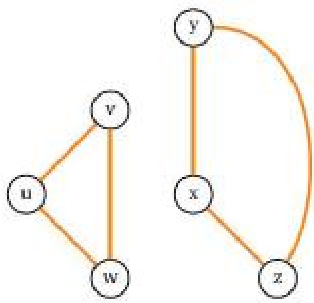
In Sociology a network is defined as a specific type of relation linking defined sets of persons, objects or events (Scott, Baggio, & Cooper, 2009). The sets of persons, objects or events upon which the network is defined are called *actors* or *nodes*. A network therefore is set of nodes and relationships between these nodes.

From a mathematical perspective a network may be represented by a diagram in which the various elements are represented by dots and the connections between them by lines connecting the dots. A connection between two nodes is called an *edge* between those nodes. These diagrams are called *graphs*, and *graph theory* is the branch of mathematics that studies such objects and their features.

For convenience we will often denote a graph by letters G or H; the set of vertices of a graph may be denoted by V or W, and the set of edges E or F. Nodes are denoted by the lowercase letters u, v etc and edges, e, f etc.

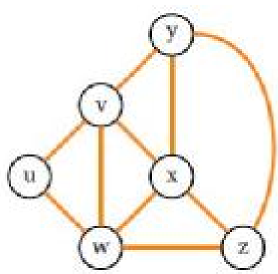
Two vertices u and v are said to be *adjacent* or *connected* if there is an edge between them. If u and v are nodes then u and v are said to be *path-connected* if there is a sequence of nodes $uu_1u_2\ldots u_nv$ such that u and u_1 are connected, u_1 and u_2 are connected … , u_n and v are connected. The sequence $uu_1u_2\ldots u_nv$ is then called a *path* from u to v. A path $uu_1u_2\ldots u_nv$ is called a *cycle* if $u = v$. That is, $uu_1u_2\ldots u_nu$ is a cycle.

Figure 1: In the graph below, *y* and *z* are path-connected, but *y* and *v* are not path-connected.



A graph in which every pair of nodes is path-connected is called a *connected* graph. The above graph is not connected. However, the graph below is connected.

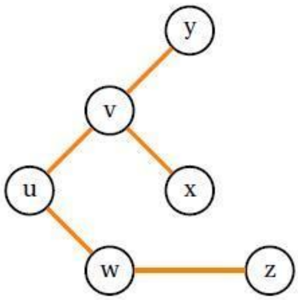
Figure 2: A connected graph



Note that although the above two graphs have the same set of nodes, they are two different graphs.

A graph is said to be ***simple***, if there is at most one edge between any two vertices. That is, any pair of vertices is either not connected with an edge, or there is precisely only one edge between the nodes. The two graphs above, therefore, are simple. A ***simple path*** and ***simple cycle*** are similarly defined. A ***tree*** is a connected graph without any cycles. The graph below is a tree.

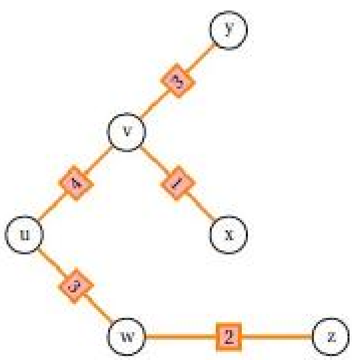
Figure 3: A tree graph



Given a graph G, if we take out some vertices, or some edges from the graph, then we end up with what is called a ***subgraph*** of the graph. Thus, a subgraph is a graph that is contained in a bigger graph. A subgraph that is also a tree, such that every node in the bigger graph is also in the subgraph, is called a ***spanning tree*** of the graph. For example, the graph in Figure 3 is a spanning tree of the graph in Figure 2. The graph in Figure 1 has no spanning tree.

A weighted graph (or tree) is a graph (or tree) in which each edge is assigned a weight or number. The following graph is a weighted tree. A weighted tree with minimum weight is called a ***minimum spanning tree*** (MST). The weight of a graph (or tree) is the sum of the weights of all the edges in the graph (or tree). For example, the weight of the tree in Figure 4 is 13.

Figure 4: A weighted tree graph



Travel in Papua New Guinea

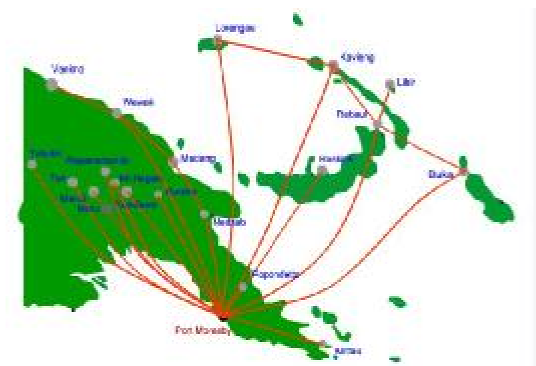
Papua New Guinea is a country in the southwestern Pacific Ocean. The country's geography is diverse and, in places, extremely rugged. A spine of mountains runs the length of the island of New Guinea, forming a populous highlands region. Dense rainforests can be found in the lowland and coastal areas. This terrain has made it difficult for the country to develop transportation infrastructure. In some areas, airplanes are the only mode of transport (Wikipedia, *http://en.wikipedia.org/wiki/Papua_New_Guinea)*

Airlines

Air Niugini is the national airline of Papua New Guinea. It operates flights from the capital city Port Moresby to at least 10 overseas destinations. Within Papua New Guinea it services all the major airports in the country. On any given day it operates between 50 to 70 flights between any two destinations (nodes). Bookings may be made in one of several Air Niugini offices and agencies around the country and overseas. As well, travelers can make bookings over the internet at the address http://www.airniugini.com.pg. On this website there are two interactive maps, a domestic version and an international version, which convey the same type of information.

The domestic interactive map consists of the map of Papua New Guinea with the major cities in towns, presumably where the airports are located, depicted with grey circles. These are the nodes. When a user places the mouse over a node, red lines appear that lead out from that particular node, to other nodes on the map that have flights to and from the chosen node. These red lines are our edges. The figure below shows what happens on the interactive map when the mouse points to the Port Moresby node.

Figure 5: Domestic airline interactive map of Papua New Guinea



Airlines PNG is a private airline that operates flights between Papua New Guinea and Australia and between numerous destinations within Papua New Guinea. It also has a website at http://www.apng.com/ where the intending traveler can make bookings. This site includes a route map which is basically a graph depicting destinations (nodes) and all routes (edges) between the nodes. In addition to the airlines described above, there are several other private and provincial-owned airlines that operate within certain regions in the country.

Sea travel

Passenger ships, freighters, charters, outboard dingies and canoes all offer links between several ports and islands. Perhaps traveling by ship is cheaper, although sometimes can be uncomfortable. For example, boats travel between Lae and Madang and are run by Lutheran Shipping with facilities including passenger cabins, accommodation and meals. They also stop in Finschhafen and Umboi Island.

Several cruise ship companies offer cruises to various ports (nodes) in the country. For example the Aurora offers cruises from Sydney to Rabaul and thence other ports in the country. The Orion Cruise Expeditions offers 5-star luxury trips to the Milne Bay islands, thence to the Solomons and other Pacific destinations.

Travel costs and the weighting of a travel route

Travel cost is generally expensive; travel by boat within the country is generally cheaper than traveling by air. For travelers on a budget, choosing a particular route may be important if one is to keep within one's budget. In the next section we investigate a mathematical method that may be used either manually, or within a computerized system, that can determine less expensive routes between any two nodes that are far apart.

According to the Independent Consumer and Competition Commission (review of the PNG air transport industry, 2006), the transport cost and the lack of transport infrastructure have been identified as major contributors to the poor performance of the tourism industry as a sector that is particularly reliant on air transportation. Other factors contributing to choice of air travel/transportation other than fares are, seat comfort, meals provided onboard, staff presentation, onboard information and entertainment, reliability in performance, journey time, reputation, passenger benefit programs, disability facilities, free baggage allowance, coordination, etc. can form part of the weighting on a graph (or tree).

We see therefore that from the mathematical point of view the weight of an edge is a factor of fare, comfort, reliability, time traveled, distance traveled, free baggage allowance, etc. To simplify things we will assume that

$$w = w(k, d, b, r)$$

where w, k, d, b and r denote weight, unit fare, distance, free baggage allowance, and reputation of the transport company respectively.

Suppose that u and v are any two nodes that represent two travel destinations that are connected, or that have an edge between them. Then the weight w of the edge *uv* is a real number that is directly proportional to the fare k per unit of distance traveled, the distance d of travel, and is inversely proportional to the reputation r (the *reliability index of*) the transport provider, free baggage allowance b.

We will define $w = c\dfrac{kd}{br}$, where c is a constant of proportionality. For simplicity, we use $c = 1$. That is, r is a value between 0 (unreliable) to 1 (completely reliable). We note that the actual fare f between the two nodes is $f = kd$, so we may rewrite the formula as

$$w = \frac{f}{br}.$$

We point out that since this is (as far as the authors can tell), is the first paper on the number w, that this is only an approximation at this stage of the research. It may well be that after more research is done this formula may be extended or improved upon, but for this paper it will be sufficient to illustrate the point about the weighting of an edge.

Generally the cost of travel is fixed between any two nodes. In some rare cases one may negotiate the fare. This is especially true if the transport provider is a small operator running a boat between islands, or a small taxi business still trying to expand its customer base.

In this study we assume that the cost f is fixed. The number r is a measure of how reliable an operator is. For specific operators some time may be needed to determine the limits of this number. In this paper we will assume r to be a function

$$r : \Omega \to [0,1]$$

from the set Ω of transport providers to the closed interval from 0 to 1.Therefore r is a fuzzy indicator. We will investigate this function more carefully in a later paper.

The number b is the number of kilos of baggage one is allowed free, after which the passenger is then asked to pay ‘excess baggage’ fare. For most transport operators this is a fixed integer.

In the case of travel via an Air Niugini flight, the number b is 16 (kg). The current fare between Port Moresby and Madang is 566.70 (K). For argument’s sake suppose we assume that r = 0.8. Then the weight along the Madang Port Moresby route is

$$w = \frac{566.70}{16 \times 0.8} = 44.3$$

We will study the number w in more detail in a subsequent paper. For now we will demonstrate the use of Prim's algorithm to determining the route of least costing between any two nodes in within a travel plan. In this case $c = b = r = 1$, so w is equal to the cost, f.

Prim's minimum spanning tree algorithm

Given a graph, one is often interested in determining a minimum spanning tree from the theoretical point of view. In this paper our interest in minimum spanning trees arises from our interest in applying minimum spanning tree algorithms to determining low cost and perhaps most efficient and therefore satisfying route between all the nodes one hopes to visit. In particular, we will apply the Prim's algorithm.

Prim's algorithm is an algorithm in graph theory that determines a minimum spanning tree for a connected weighted graph. This algorithm determines a sub-graph of the weighted graph, which is a weighted tree that contains all of the nodes from the original graph.

Prim's algorithm has the property that the edges in the set A always form a single tree. We begin with some vertex v in a given graph G with sets of vertices V and set of edges E, defining the initial set of vertices A. Then, in each iteration, we choose a minimum-weight edge (u, v), connecting a vertex v in the set A to the vertex u outside of set A. Then vertex u is brought in to A. This process is repeated until a spanning tree is formed. We always choose the smallest-weight edge joining a vertex inside set A to the one outside the set A. The implication of this fact is that it adds only edges that are "safe" for A; that is, when the algorithm terminates, the edges in set A form a MST.

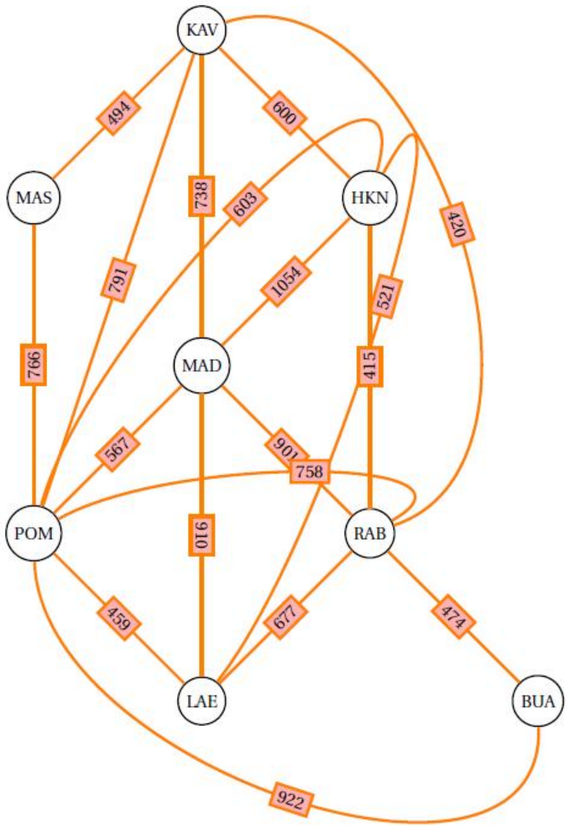
Figure 6: Prim's Algorithm

We describe this algorithm in pseudo-code form as follows:

```
let A be a single vertex u
while (A has fewer than n vertices)
{
  find the smallest edge connecting A to G − A
  add it to A
}
```

To illustrate this algorithm, consider the different ways of traveling from Rabaul (RAB) to Manus (MAS) in a route that includes stops at each of the other locations. The following graph contains routes to and from Rabaul and Manus. For simplicity we exclude the routes to and from Lihir.

Figure 7: Graph contains routes to and from Rabaul and Manus



Here, the nodes are different towns or airports: In particular RAB stands for the town Rabaul, and MAS for Manus, etc. The weights of the edges in the graph are the costs of the fares between the different nodes. To start, we pick the node RAB. This node forms the start of our spanning tree *A*. We look at all the edges going out from RAB, and choose an edge of minimum weight that joins RAB to a node outside of the tree *A*. In this case, we choose the edge with weight 415, that connects RAB to HKN. The tree *A* now contains the nodes RAB, HKN and the edge with weight 415 between these two nodes. We now look outside of *A* for a node that is joined minimally to a node in *A*. Such a node is KAV, which is joined by an edge with weight 420, to KAV. If we continue, we get this sequence of trees:

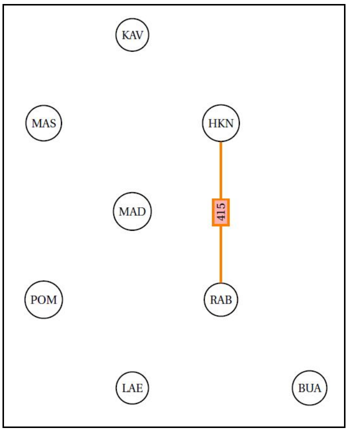


Fig 8-a: *A* with two nodes

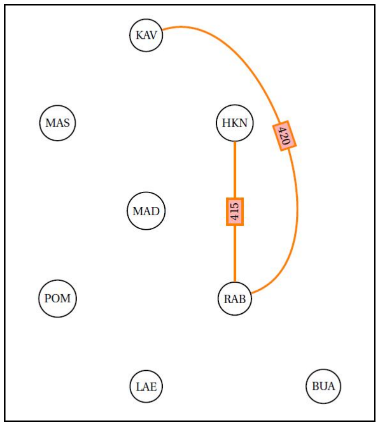


Fig 8-b: *A* with three nodes

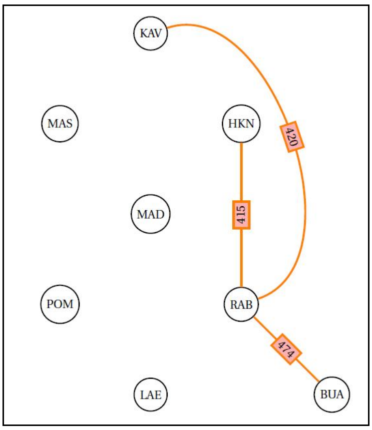


Fig 8-c: *A* with four nodes

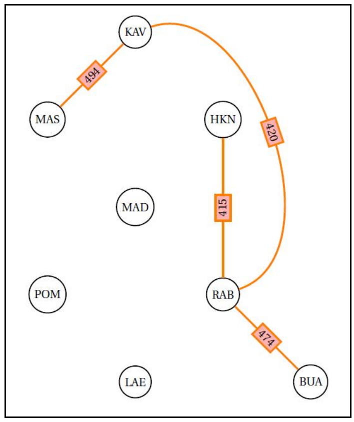


Fig 8-d: *A* with five nodes

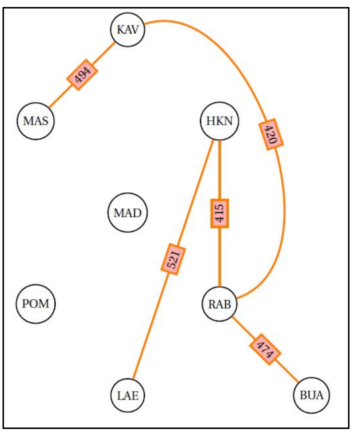


Fig 8-e: *A* with six nodes

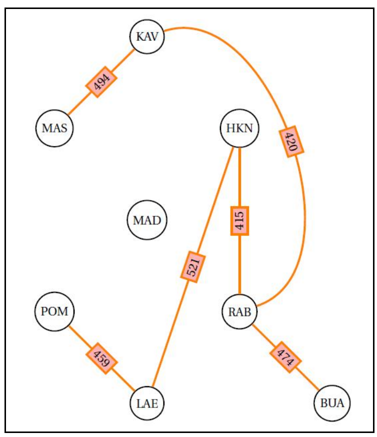


Fig 8-f: *A* with 7 nodes

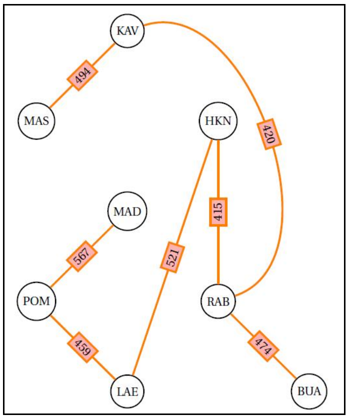


Fig 8-g: *A* with 8 nodes

Fig 8-g represents the minimum spanning tree of the connected graph in Figure 7. This is the subtree with minimum weight. What this means is that if one wants to visit all the nodes in Figure 7, this is the route that one should take in order to minimize cost. In this case, the total cost is K3350.

This is not as simple as it looks. In order to take full advantage of this tree, one has to visit at least one of the nodes more than once. For example a possible route is MAD, POM, LAE, HKN, RAB, BUA, RAB, KAV, MAS. Here, a person starts at MAD, then POM, etc until he/she reaches RAB. From RAB he/she takes the trip to BUA, then back again to RAB and thence to HKN, KAV and MAS. Thus RAB is visited twice.

However, even with the above scenario, one still spends less than any other route in the graph, in which every node is visited at least once (as one can easily verify).

We have deliberately simplified the problem of choosing a route by1) taking a small number of nodes, 2) by taking only edges serviced by airplanes, 3) ignoring reliability, seat comfort, etc 4) by excluding one or two routes. The situation will drastically increase in complexity when boat trips, buses, etc are factored in, including the fares for these transport modes, even for a small number of nodes.

Conclusion

In this paper we demonstrated the use of Prim's algorithm to obtain a minimum spanning tree from a given connected graph. When the nodes of the graph represent transportation nodes then a minimum spanning tree can suggest a route of lowest cost or close to lowest cost through all the nodes in the graph. This may be extremely useful when designing a tour consisting of different destinations and sites in the country.

The algorithm can reduce the amount of time needed to determine a route, and as well, increase the traveller's confidence in the suggested route. The situation can become extremely complex when different types of transportation are included, as well as the different variables discussed in the previous section.

But that will be the focus of a subsequent paper.

References

Airlines PNG website, *http://www.apng.com/*

AirNiugini website , *http://www.airniugini.com.pg*

Independent Consumer and Competition Commission (Review of the PNG Air Transport Industry, Final Report, 20th September 2006),

Scott, Noel; Baggio, Rodolfo & Cooper, Chris 2009, *Network Analysis and Tourism: From Theory to Practice*, Channel View Publications, NY.

Traveler's Point, http://www.travellerspoint.com/guide/Papua_New_Guinea/

Wikipedia, *http://en.wikipedia.org/wiki/Papua_New_Guinea*

Authors

Dr Lakoa Fitina is the Head of the Mathematics and Computing Department and Mr John Imbal Fitina is the Head of the Tourism and Hospitality Management Department at Divine Word University. Ms Vanessa Uiari, Ms Nathaline Murki and Ms Elizabeth Goodyear are lecturers within the Tourism and Hospitality Management Department. The departments are part of the Faculty of Business and Informatics at Divine Word University.