Chapter 5

Wide-area network design

5.1 The network design process

The network design process possesses many stages and seemingly independent processes. The various parts of the design process are normally approached individually due to inherent interaction between the factors that influence the design of wide-area WDM optical networks. There are however many principles and functions that optical network design shares with the design of other communication systems.

The basic characteristics of a communication system are information sources, information destinations and the transport of information between these. The source of an information transport is usually geographically displaced from the destination, hence the need for transport networks. The need therefore exists to establish physical infrastructure between the various nodes of a communication network, to provide logical connectivity that can satisfy the communication demand of the source and destination pairs.

In conjunction with physical connectivity, a communication network requires mechanisms for managing the flow of information over the physical infrastructure to ensure security, reliability and quality of service. The way in which these management functions are implemented is greatly influenced by the underlying technologies and protocols used for data transfer. In this section the network design process will be discussed with reference to optimisation parameters, commercial and proprietary design software, and the integrated design methodology.

5.1.1 Optimisation parameters

There always exist certain expectations from a communication network's users, funders and operators of what the network's characteristics should be. The concept of optimisation and the optimisation parameters that can be optimised for, is key to addressing and managing these expectations. A network can be designed in such a way as to provide for the expectations of one party for a certain period, but as user demand and market conditions change a poorly optimised network may quickly lose its ability to satisfy the requirements of all its stakeholders.

In the keynote address of the international IEEE AFRICON 2002 conference in George, South Africa, Dr. Hiromasa Haneda elaborated at length on the differences between the terms *optimal* and *optimum* and how they should be interpreted in optimisation problems. The crux of the matter was that the term *optimum* should be used with extreme care since it implies superiority of a solution with regard to all conceivable criteria. An optimal solution, on the other hand, forms part of a collection of solutions to a problem, each achieved by an optimisation process with a specific optimisation function being considered. The challenge thus lies in the careful selection of an optimisation function that defines the selected optimisation parameter.

The classic optimisation parameters of any optimisation problem are cost and performance. In the case of cost the aim is to achieve as low a cost as possible, whereas performance has the aim of being as high as possible. These parameters unfortunately have the troublesome characteristic of being mutually destructive, leading to a difficult trade-off situation where an increase in the one, for instance higher performance, has a negative impact on the other, greater cost in this case. Due to the inverse logic nature of cost as optimisation parameter, it is important to note that a positive impact on cost is defined as a lowering in cost, where a negative impact on cost is define as an increase in cost.

One could be tempted to define other optimisation parameters in conjunction with the two fundamental parameters of cost and performance. Parameters such as reliability, capacity, and scalability can however be considered as being performance characteristics since they also typically result in a trade-off situation with the cost parameter. It is therefore important to clearly define the exact composition of the performance metric when it is stated as the aim of an optimisation process.

Capacity is the metric usually associated with performance, where data rates and bandwidth dominate as user-level interpretations of a network's value. It is proposed here that reliability, as discussed in section 4.4, falls in the same category as capacity when considering the performance of a network. This can be justified by the user-level perception of bad network performance usually being due to insufficient reliability or the effects of the processes that attempt to restore network functionality in the event of fault or malfunction.

Scalability of a network relates to the ease with which it can allow for the growth of its user-base, capacity or geographical coverage. When optimising a network design, a situation similar to that of training a neural network can be created. Neural networks should be trained with data that represents the statistical distributions of parameters that will exist under normal operating conditions. The design of communication networks, like the training of neural networks, should avoid over fitting that may inhibit the resultant communication network's ability to cope with changes in its composition not anticipated or accommodated for in the original design process.

A good balance between current and possible future requirements should be maintained to avoid a situation where a network is currently utilised at a very low percentage of its capacity because it was designed to suit possible future requirements. Investors in communication network infrastructure generally demand maximum return on investment in the shortest time frame possible and would thus not be satisfied with a network that operates at very low utilisation levels if a cheaper network operating at higher utilisation levels would have resulted in the same revenue generation and user requirement satisfaction potential.

It is true that some level of interaction between reliability, capacity and scalability exist, but these are superficial when compared to that of the cost parameter on these three performance parameters. For many optimisation problems it would be possible to create an optimisation function that optimises for a combination of these three performance parameters without sacrificing too much compared to when they are considered as individually exclusive optimisation problems.

5.1.2 Commercial and proprietary design software

Network design software tools play an important role in assisting network designers in the design process. These software tools vary from very specific algorithms that require the computational speed of a computer, to involved design suites with extensive user interfaces that guide designers through various stages of the design process.

Many of the software tools are available commercially, but usually at extremely high cost due to the limited market for these tools. Most players in the optical networking industry do however also utilise proprietary software tools that are customised for their specific products offerings and equipped with a wealth of knowledge obtained through years of experience. These tools are seldom mentioned in open literature, or where reference is made to them very limited information about their features and functionality are made public. This can be expected from such a highly competitive industry where trade secrets are often concealed in the value of a mysterious constant or unpublished equipment characteristic.

Network design tools exist for various stages in the design process. The RCA function is responsible for developing the virtual topology from the physical and logical topologies of a network. This problem has received a lot of attention from network designers and researchers and is often mistaken for being the only part of the network design process suited for computer-based solving. As a matter of fact, the RCA function, however important, merely brings the design process to a climax where optimisation can be considered and the design eventually concludes.

In addition to purely network design tools, there exists a type of software tool relevant to the network designer, namely network evaluation tools. Two main categories of network evaluation tools exist, namely: simulation tools and optimisation tools. Simulation tools are used to investigate the behaviour of already designed networks, whether theoretical or practical. Optimisation tools, on the other hand, are used where the values of various network parameters are manipulated in order to optimise a predetermined optimisation function, and often form an integral part of the design process.

Network design tools can be categorised based on the specific function that they perform. These functions include: traffic forecasting, trunk engineering, SONET/SDH transport, signalling network design, access network planning, distribution network planning, and backbone network planning. Commercially available software tools available from around $20,000 to more than $500,000 offering various combinations of these design functions include: OPNET by MIL3, COMNET by CACI Products Company, COMPOSIS by AixCom, NetScene by Network Design House, NetMaker by Make Systems, WESTPLAN by Westbay Engineers, AUTONET by NDA Corporation, WinMIND by Network Analysis Center, CANE by ImageNet, NetSuite by NetSuite Development, and NetCracker by NetCracker Technologies.

Figure 5.1 shows the interfacing of the various software modules developed by the CATO project [57]. From this figure it is apparent that the RCA component, here shown as the routing and wavelength assignment function, is one of several network design functions accommodated for in a software-based tool for the design of optical networks. Functions such as restoration and protection, discussed in section 4.4.1, and resource allocation and placement also utilise physical and logical topology inputs and interaction between each other and the RCA function, to produce a virtual network design that can be assessed for optimality.

Another academic tool set for optical network optimisation, modelling and design called NoMAD [58] has been developed as an application of hybrid genetic algorithm and heuristic optimisation techniques to optical network design. Object oriented design methodology makes NoMAD easy to understand, flexible and extensible. The genetic algorithms approach to network design makes use of objective functions and fitness levels to evaluate the mutations of successive generations.

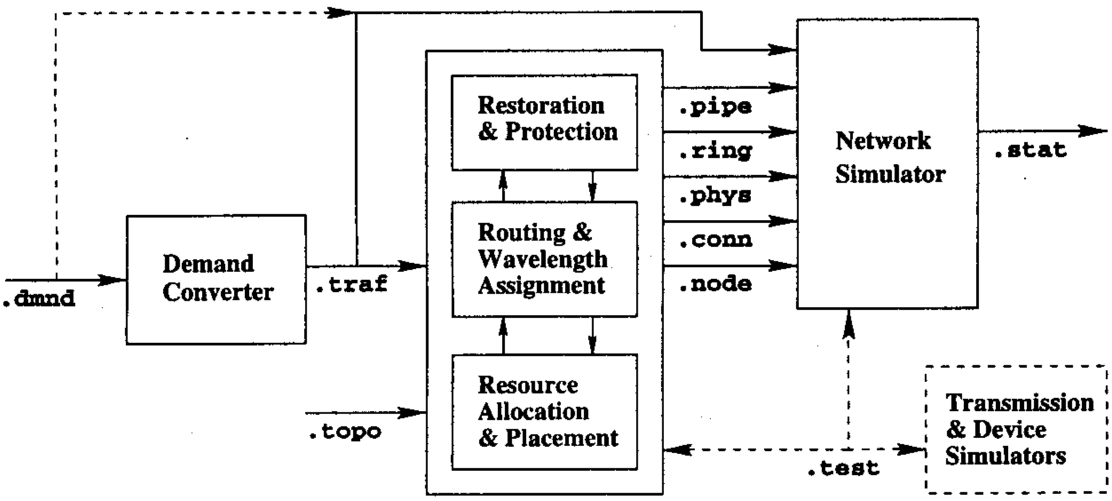

Figure 5.1: High-level diagram of CATO, a computer aided design (CAD) tool for optical networks and interconnects [57].

5.1.3 Integrated design methodology

The network planning and design process is complex with many influencing factors, optimisation parameters, tasks, interactions and dependencies. The scope of a network planning and design methodology depends on the employed evaluation criteria. If a methodology is expected to produce the number of required wavelengths for a given scenario, the RCA function will be sufficient. If the economic viability of VoIP needs to be determined, much greater scope would be required. In the context of enterprise network planning and design the following tasks have been identified [49]:

strategic business modelling is the task responsible for analysis of business requirements and revenue generating opportunities. It identifies applications and services that need to be supported by the network infrastructure under design.

industry and technology trends analysis identifies and analyses the influence of technology trends on business. Phases such as introduction, maturity, acceptability, and standardisation are used to assess technology trends.

strengths, weaknesses, opportunities and threats analysis is the task responsible for assessing the *status quo* in terms of usage and capability of communication infrastructure. Strengths and weaknesses of the current infrastructure are determined in order to identify opportunities and threats.

network architecture planning developes functional architecture models defining key functionalities and their interaction on each other.

network planning and design defines the detail of how the architectural planning objectives and requirements can be satisfied. Optimisation techniques are employed to minimise cost for a specified requirements and constraints scenario.

business justification and transition planning is responsible for the development of strategies for closing the gap between the current and desired communication network infrastructure. Economic tools and methods are used to provide alternatives and justification to communication infrastructure investment.

network infrastructure engineering and implementation is the technical task responsible for the deployment and implementation issues of providing appropriate communication network infrastructure.

The network planning and design task is most relevant to this investigation, although some attention has been given to some of the business aspects on the network planning and design problem as discussed in section 4.5. The main activities of the network planning and design task include: requirement specification, network topology design, network dimensioning, and design analysis and verification.

The requirement specification activity involves the identification and estimation of traffic characteristics and resource requirements. The network topology activity determines the number and positions of network nodes and their interconnections, which can be influenced by constraints imposed by existing infrastructure. Performance and reliability modelling are performed by the network dimensioning activity to determine optimum network configurations under various traffic load conditions. The design analysis and verification activity involves iterative sensitivity analysis to evaluate the robustness of a network design under various conditions and for different scenarios.

Figures 5.2 and 5.3 show the integrated design methodology developed through the research conducted in this investigation. The process commences by taken all the influencing factors into consideration when developing the physical and logical topologies. Conventional approaches rely on subjective decisions, as shown in figure 5.2, by network designers that are familiar with the country or region for which the wide-area network is to be designed. Expert knowledge about the network nodes to be inter-connected and the traffic distributions to be expected between them, characterise the subjective nature of the decision making process. The RCA function has been identified as the integrating process responsible for producing a virtual topology. Capacity, reliability and scalability requirements are satisfied through the recursive optimisation process that eventually terminates the design process.

The level of designer interaction in the design process is indicated in figure 5.2 by the manual and automated domain boundaries on the right. Due to the subjective nature of the decision making process responsible for interpretting the influencing factors in developing the physical and logical topologies, manual interaction is required right up to their initial stages. The RCA function is highly automated through the use of several established computer-based algorithms, although the interpretation of results and optimisation processes are again very manual in nature. It is also important to note that the network designer has to take final responsibility when terminating the design process.

As with most processes, there is always an aim to automate as much of the process as possible. In network design processes an increase in design automation can shorten the design cycle and produce more repeatable results. In order to achieve greater automation in the network design process, it is suggested that the interpretation of influencing factors be automated to effectively minimise designer interaction. Figure 5.3 shows how the design methodology presented in figure 5.2 can be modified to allow for greater design process automation. The *subjective decisions* component is replaced with an *objective decisions* component that should result in greater design process automation, since computer-based decision logic structures are inherently objective.

The concept of objectivity in the context of design process automation is defined as the ability to make decisions based on a set of criteria without allowing bias or prejudice to negatively influence the repeatability of the process. A subjective decision, on the other hand, is defined as being influenced by knowledge not explicitly declared as relevant to the criteria framework, thus resulting in low repeatability due to the unpredictability of designer bias and prejudice.

The contribution of the research conducted in this investigation into the field of wide-area optical network design is in the technique suggested for achieving increased objectivity in the network design process. In section 5.3, clustering is presented as technique for improving objectivity in the network design process through its algorithmic nature and resultant load balancing characteristics.

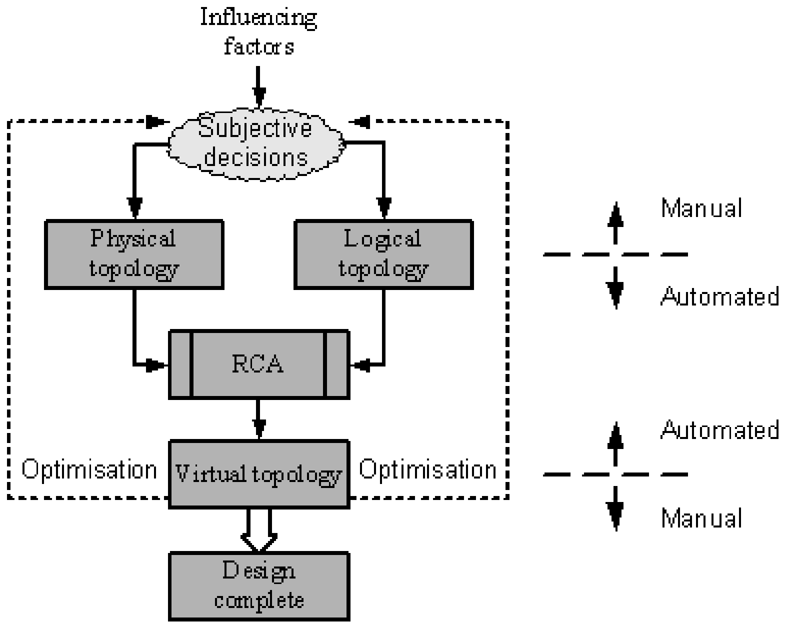

Figure 5.2: Traditional subjective integrated network design methodology.

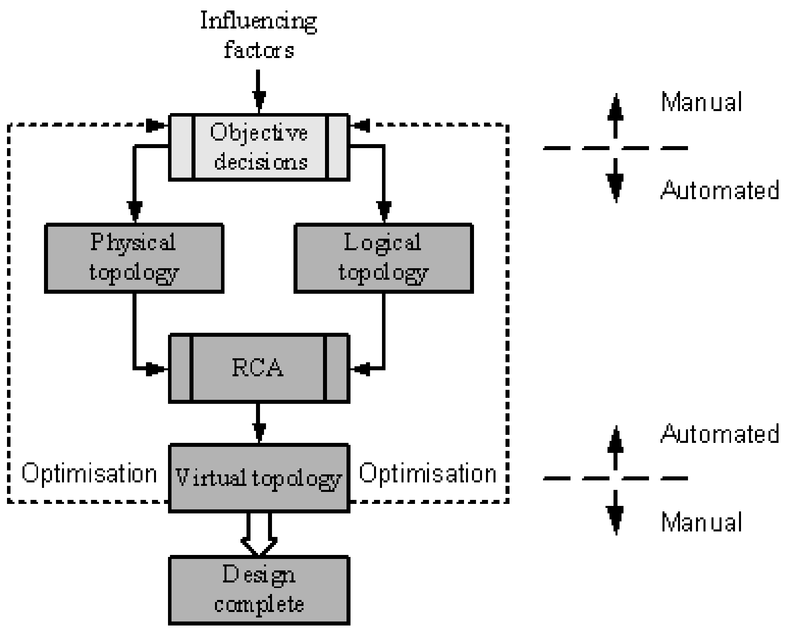

Figure 5.3: Proposed objective integrated network design methodology.

5.2 Methodology for finding hub nodes from economic activity statistics

A methodology was developed for finding hub nodes, clusters and the demand matrix for a network under design by using economic statistics as input to the process. Section 3.1.3 and equation 3.2 in particular show how economic activity can be used as a nodal weight in a modified gravity model. The developed methodology, shown in figure 5.4, applies to any nodal weights and not only economic activity statistics in particular.

In figure 5.4 it can be seen that actual economic activity statistics or a demand matrix generator can be used to generate a full demand matrix. The use of actual economic statistics and geographical coordinates will be demonstrated in chapter 6 and the use of the demand matrix and geographical coordinate generators will be demonstrated in section 5.3.4. The demand matrix reducer will be employed in chapter 6 due to memory constraints, as described by equation 5.1, when implementing the methodology on a computer with finite memory.

The methodology contains an iterative segment where the intra/inter-cluster traffic ratio is used to determine the suitable number of hub nodes. The issue of how many network nodes there should be on the various levels of the multi-level network model, discussed in section 4.1, is non-trivial. One of the attributes of this methodology is its ability to objectively determine the number of nodes required on each of the levels of the multi-level network model.

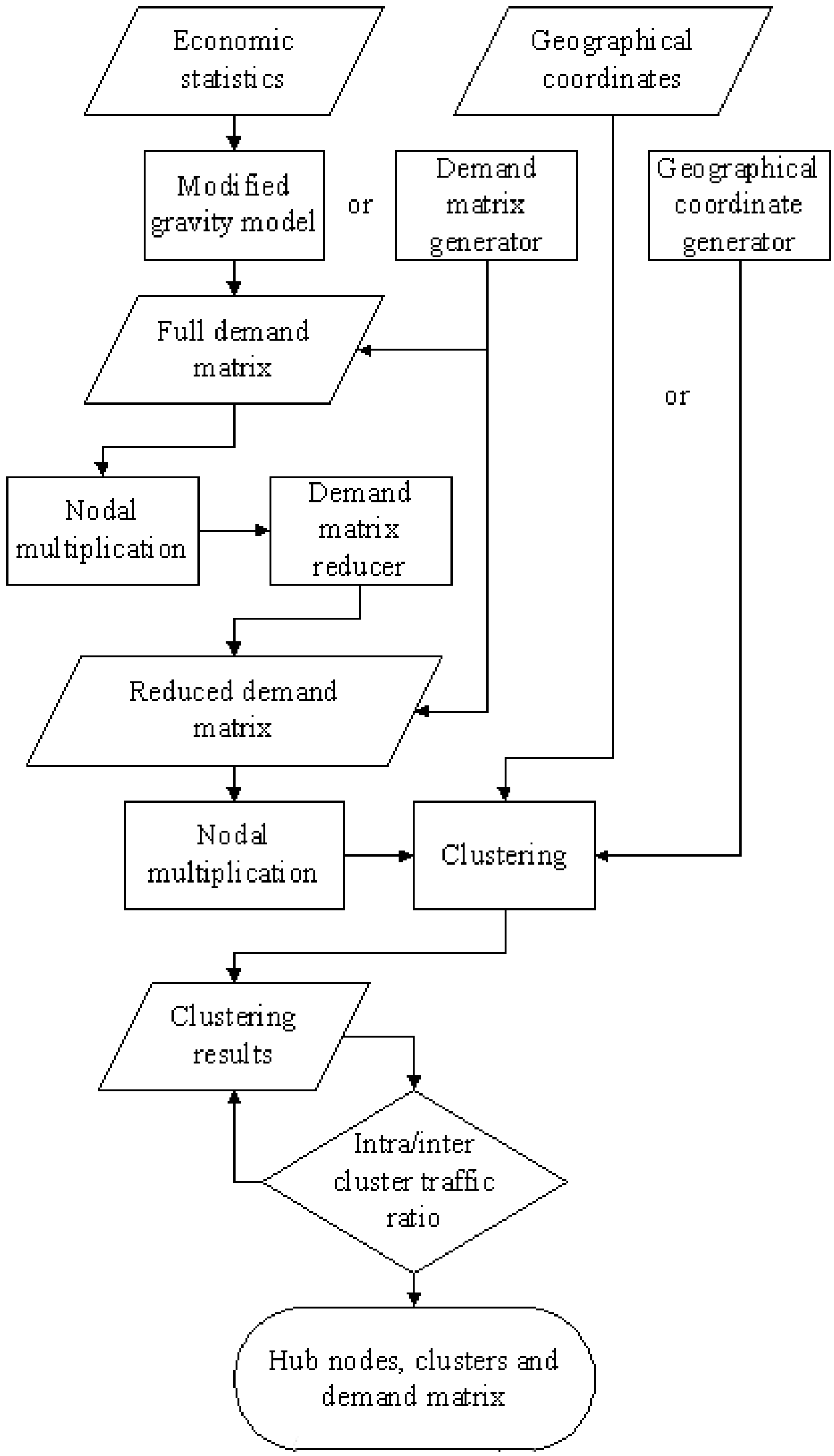

Figure 5.4: Flow diagram of methodology for finding hub nodes from economic statistics.

5.3 Clustering of network nodes

A conventional approach to selecting hub nodes for wide-area optical networks is by means of exhaustive searching. This approach guarantees optimal solutions, but requires vast amounts of time, which make it impractical for network design problems exceeding moderate complexity. The clustering approach to selecting wide-area optical network hub nodes is a statistical pattern recognition technique whereby the number of clusters is reduced one by one through the merging of neighbouring clusters judged most similar based on a selected similarity metric. Although consuming considerably less time than exhaustive searching, this technique has memory requirements that increase non-linearly with the number of network nodes of the network:

$$M \propto \frac{n(n-1)}{2}, \tag{5.1}$$

where M is the required memory and n is the number of physical network nodes being clustered.

There are various well-known similarity metrics, including: shortest distance, largest distance, average distance, and centroid distance. In this implementation of the clustering approach to hub node selection, the employed similarity metric is known as the Ward linkage [59], which can be described as an incremental sum of squares metric. This metric is very well suited to the creation of clusters around hub nodes, due to the way in which it rewards low interference on hub node location when selecting clusters to merge. The two clusters judged to be most similar would be the clusters that minimise the Ward linkage as follows:

$$
\begin{aligned}
s_{Ward}(H) &= min_H \left(\frac{N_i N_j \| \frac{1}{N_i} \sum_{\bar{x} \in H_i} \bar{x} - \frac{1}{N_j} \sum_{\bar{x'} \in H_j} \bar{x'} \|^2}{N_i + N_j} \right) && (5.2) \\
&= min_H \left(\frac{N_i N_j \| \mu_{H_i} - \mu_{H_j} \|^2}{N_i + N_j} \right), && (5.3)
\end{aligned}
$$

where min_H is interpreted as the minimum over all possible combinations of clusters, H is the collection of clusters, x and x' are the nodes of arbitrary clusters in the collection, i and j are cluster indices, N is the number of network nodes in a cluster, and μ_{H_i} is the centroid of cluster i.

During each iteration of the clustering process, two clusters are merged to form one cluster containing all their respective network nodes. Iteration of the clustering process thus reduces the total number of clusters by one cluster per iteration. It is intuitive that clusters would be more similar early in the clustering process than late in the clustering process. Figure 5.5 shows how the value of the similarity metric changes with the number of remaining clusters. In this clustering problem, 1801 clustering iterations were required to end up with one cluster containing all the virtual network nodes that were created from 60 actual network nodes through a process of node multiplication. For the clustering process to consider node weighting, multiple instances of actual network node are created according to its estimated add/drop traffic, resulting in a higher number of virtual network nodes than actual network nodes.

5.3.1 Background to similarity metrics

The purpose of data clustering is to identify clusters that appear naturally in a given data set. When starting a clustering process all data points are regarded as individual

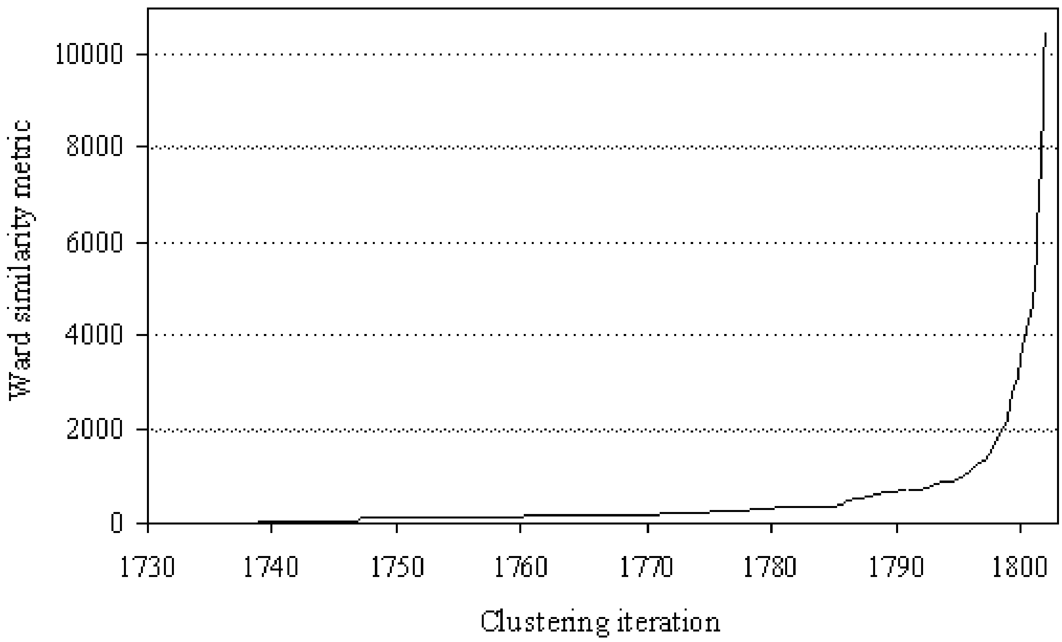

Figure 5.5: The non-zero values of the similarity metric are shown as a function of the number of elapsed clustering iterations.

clusters. Two clusters that are judged to be most similar, based on a chosen metric, are combined to form one new cluster. This process is repeated until the desired number of clusters is reached.

Clustering by implementing the minimum distance s_{min} similarity metric means that two clusters are combined to form one new cluster when the smallest Euclidean distance between data points in the two clusters is smaller than the smallest Euclidean distance between any other two data points from any two different clusters. The s_{min} similarity metric can be expressed as:

$$s_{min}(H) = min_H \left(min_{\bar{x} \in H_i, \bar{x'} \in H_j} \|\bar{x} - \bar{x'}\| \right), \tag{5.4}$$

where min_H is interpreted as the minimum over-all possible combinations of clusters.

Clustering by implementing the maximum distance s_{max} similarity metric means that two clusters are combined to form one new cluster when the largest Euclidean distance between data points in the two clusters is smaller than the largest Euclidean distance between any other two data points from any two different clusters. The s_{max} similarity metric can be expressed as:

$$s_{max}(H) = min_H \left(max_{\bar{x} \in H_i, \bar{x'} \in H_j} \|\bar{x} - \bar{x'}\| \right), \tag{5.5}$$

where min_H is interpreted as the minimum over-all possible combinations of clusters.

Clustering by implementing the average distance s_{avg} similarity metric means that two clusters are combined to form one new cluster when the average Euclidean distance between data points in the two clusters is smaller than the average Euclidean distance between the data points of any other two clusters. The s_{avg} similarity metric can be expressed as:

$$s_{avg}(H) = min_H \left(\frac{1}{N_i N_j} \sum_{\bar{x} \in H_i} \sum_{\bar{x'} \in H_j} \|\bar{x} - \bar{x'}\| \right), \tag{5.6}$$

where min_H is interpreted as the minimum over-all possible combinations of clusters.

Clustering by implementing the centroid distance s_{mean} similarity metric means that two clusters are combined to form one new cluster when the Euclidean distance between the means of the data points in the two clusters is smaller than the Euclidean distance between the means of the data points of any other two clusters. The s_{mean} similarity metric can be expressed as:

$$s_{avg}(H) = min_H \left\| \frac{1}{N_i} \sum_{\bar{x} \in H_i} \bar{x} - \frac{1}{N_j} \sum_{\bar{x'} \in H_j} \bar{x'} \right\|, \tag{5.7}$$

where min_H is interpreted as the minimum over-all possible combinations of clusters.

5.3.2 Clustering of weighted network nodes

The way in which node weighting, as discussed in section 3.1.3, translates into node multiplication ensures that clustering decisions are based on node weights as well as absolute and relative node locations. If, for instance, the nodes that have heavier weights are multiplied by a greater factor than the nodes that have lighter weights, a situation will be created where the clustering process is biased towards heavily weighted network nodes. The opposite might seem to be a better situation, where the gap between heavily and lightly weighted nodes can be reduced through a greater multiplication factor for nodes with lighter weights. Theoretically it would however be preferable to not introduce any bias at the node multiplication stage, since the node weighting stage can be used for this. In practice it might not be possible to avoid bias at the node multiplication stage, since only integer multiples can be created and the nodal weights might not be integer multiples of a common denominator.

The demand matrix developed in section 3.1.3 is used to produce nodal add/drop traffic values. This is done by determining the sum of the values in the row and column of each network node in the demand matrix as follows:

$$w_i = \sum_{j=1}^{N} D_{i,j} + D_{j,i}. \tag{5.8}$$

where w_i is the add/drop traffic of physical network node i, $D_{i,j}$ is the demand between nodes i and j as defined in equation 3.2, and N is the number of network nodes. For a symmetrical demand matrix this would simplify to:

$$w_i = 2\sum_{j=1}^{N} D_{i,j} \propto \sum_{j=1}^{N} D_{i,j}. \tag{5.9}$$

The clustering process is initialised by regarding all network nodes, of which most are multiples resulting from nodal add/drop traffic weighting, as individual clusters. The number of initial clusters is given by:

$$n_{weighted} = K_2 \sum_{i=1}^{N} w_i, \tag{5.10}$$

where w_i is the add/drop traffic of physical network node i and N is the number of physical network nodes. K_2 is a normalisation factor chosen in such a way as to ensure that each physical network node translates into at least one virtual network node. Due to memory considerations when implementing the algorithm this is achieved through the rounding up of normalised individual nodal add/drop traffic and not by normalisation alone as would be theoretically preferable. Equation 5.10 can be interpreted as the number of physical network nodes multiplied by the normalised average nodal add/drop traffic.

5.3.3 Intra/inter-cluster traffic ratio

One of the most important problems to be solved when designing an optical network, and any other communication network for that matter, is that of how many network nodes there should be on each of the levels of the multi-level network model. In the context of wide-area network design the number of backbone nodes is of importance. The developed methodology utilises a metric referred to as the intra/inter-cluster traffic ratio to assist in solving the problem of how many network nodes there should be on each of the levels of the multi-level network model.

The intra/inter-cluster traffic ratio for a specific cluster is defined as follows:

$$R = \frac{\text{traffic with source and destination in cluster}}{\text{traffic with source or destination in other cluster.}} \tag{5.11}$$

For high numbers of clusters this ratio is typically low, since a lot of inter-cluster traffic would exist. For low numbers of clusters this ratio is typically high, since strong communities of interest exist within the clusters resulting in high intra-cluster traffic. Finding the number of clusters for which the intra/inter-cluster traffic ratios are acceptable is thus a justified approach to selecting the number of hub nodes for a specified network level.

Figure 5.6 presents a special case of the multi-level network model, where only three levels exist and no network node can appear on more than two of these levels. Guidelines for selecting suitable intra/inter-cluster traffic ratios when solving the problem of how many network nodes should appear on the various levels of the multi-level network model are given through the 10%, 30% and 60% geographical traffic distribution estimates. Two target intra/inter-cluster traffic ratios can be derived from these guidelines. The first ratio is valid at the interface of the backbone level and regional levels, where the ratio is as follows:

$$R = \frac{30\% + 60\%}{10\%} = 9. \tag{5.12}$$

The second ratio is valid at the interface of the regional level and local levels, where the ratio is as follows:

$$R = \frac{60\%}{30\% + 10\%} = 1.5. \tag{5.13}$$

These target intra/inter-cluster traffic ratios should however only be seen as a guidelines, since the required configuration of multi-level network model levels for a network under design might not be conform to the three level network model presented here.

The means and standard deviations of the different numbers of intra/inter-cluster traffic ratios are used in order to enable comparison between the ratios of different numbers

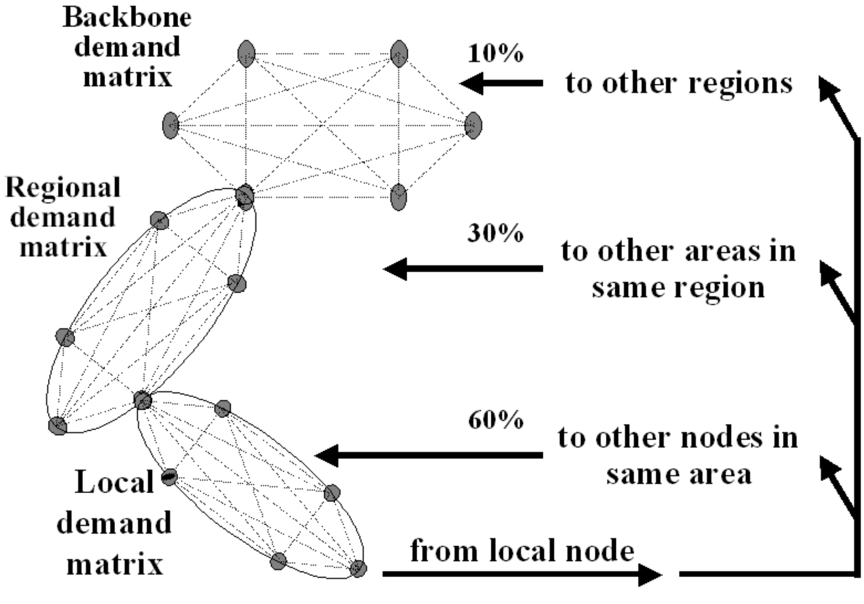

Figure 5.6: Intra/inter-cluster traffic ratios for various levels of a three-level network model.

of clusters. The mean intra/inter-cluster traffic ratio is favourable when it is close to a target value deemed appropriate for the specific network level. Proximity to this target value identifies potentially optimal numbers of clusters. When attempting to find the optimal number of high-level hub nodes, the first incidence of mean intra/inter-cluster traffic ratios above or on and subsequently on or below the target value, are regarded as suitable candidates. The candidate with the lowest standard deviation of the intra/inter-cluster traffic ratios is deemed optimal, since a lower standard deviation would indicate that the clusters have intra/inter-cluster traffic ratios closer to the mean intra/inter-cluster traffic ratio.

5.3.4 Simulation experiment

The purpose of this simulation experiment is to determine how the integrated design methodology presented in section 5.1.3 can be used to evaluate the means and standard deviations of the intra/inter cluster traffic ratios, as described in section 5.3.3, for the following theoretical scenarios:

1. Geographical coordinates generated from a beta distribution with $a = b = 1$ and a demand matrix generated from a beta distribution with $a = b = 0.1$.
2. Geographical coordinates generated from a beta distribution with $a = b = 1$ and a demand matrix generated from a beta distribution with $a = b = 1$.
3. Geographical coordinates generated from a beta distribution with $a = b = 5$ and a demand matrix generated from a beta distribution with $a = b = 0.1$.
4. Geographical coordinates generated from a beta distribution with $a = b = 5$ and a demand matrix generated from a beta distribution with $a = b = 1$.

The beta distribution was employed due to the ease with which an iterative process can be performed on its parameters, resulting in the creation of a potentially wide range of probability distribution functions should the need arise. The statistical nature of the four scenarios defined above are described in the following sections.

Simulation context definition

The simulation was conducted in Matlab using the Statistics Toolbox version 3.0. Definition of the context of the two types of input parameters to the simulation, geographical coordinates and a demand matrix, is important for it determines the context of the simulation.

A square plane spanning 10° by 10°, approximately 1.2×10^6 km^2, was used for the geographical coordinate sets to be generated. This corresponds roughly to the surface-area of a medium-sized country like South Africa, which has a surface area of $1,219,090\ km^2$. The square plane under consideration is purely theoretical and should not be misconstrued as being representative of an irregular shaped country, like most countries tend to be.

For the purpose of this simulation an aggregate network capacity of 1 Tbps was assumed. This value relates to all the traffic between the nodes of the network and not to the traffic that exists at lower levels of the multi-level network model that are beyond the scope of the network design under consideration. In a practical scenario this kind of traffic would for example include a circuit-switched telephone connection established between two telephones connected to the same exchange.

The number of network nodes taken into consideration in this simulation has been set at 60. This value is small enough to allow for the memory-hungry clustering of multiplied virtual network nodes, as described in section 5.3.2, and large enough to ensure meaningful clustering results.

Statistics of the input parameters

The input parameters to the simulation are generated geographical coordinates and a demand matrix. These were generated according to a beta distribution by a random number generator in Matlab. The beta distribution has two parameters of its own, as shown in equation 5.14, and these were chosen to be equal at values of 0.1 and 1 for generating two different demand matrices, and 1 and 5 for generating geographical coordinates on the square plane defined in section 5.3.4.

Figure 5.7 shows the plots of the beta pdf for $a = b = 0.1$, $a = b = 1$, and $a = b = 5$ respectively. Note how $a = b = 1$ results in a uniform distribution and $a = b = 1$ has a similar shape to the normal distribution although it is confined to the range $(0,1)$. By manipulating the two input parameters of the beta distribution it is possible to create virtually any pdf in the range $(0,1)$, which makes it well-suited to simulations such as this where various probability distributions are iteratively required. The well-known beta probability distribution function (pdf) is given by:

$$y = f(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} I_{(0,1)}(x), \tag{5.14}$$

where a and b are input parameters determining the shape of the beta pdf, x is the value for which probability is evaluated, the indicator function $I_{(0,1)}$ ensures nonzero probability for values of x in the range $(0,1)$, and $B(\cdot)$ is the Beta function:

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \tag{5.15}$$

where $\Gamma(a)$ is the gamma function defined by the integral:

$$\Gamma(a) = \int_0^{\infty} e^{-t} t^{a-1} dt. \tag{5.16}$$

In order to allow for the comparison of results obtained from the different beta distributions, two random number generator seeds were used, one for the generation of the two different demand matrices and the other for the generation of the two different geographical coordinate sets.

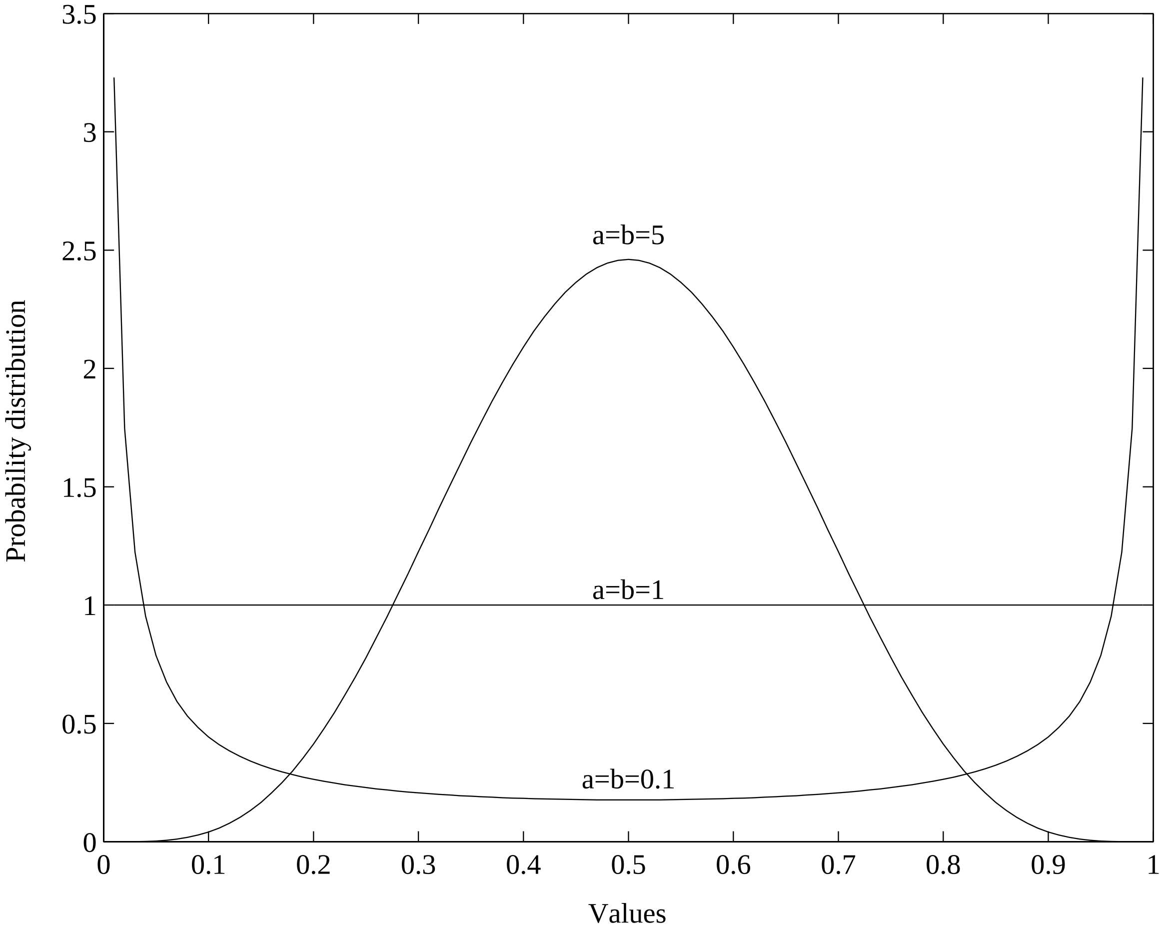

Figure 5.7: Beta probability distribution functions for a=b=0.1, a=b=1, and a=b=5.

Geographical position as input parameter

Two sets of geographical positions for the network nodes to be clustered were generated. The one set was generated from a beta distribution with $a = b = 1$ and the other from a beta distribution with $a = b = 5$. The first set is shown in figure 5.8 where the uniform nature of the coordinate distribution over both latitude and longitude can be observed. The second set is shown in figure 5.9 where the concentration of networks nodes towards the middle of the plane can be observed.

Demand matrix as input parameter

Two demand matrices were generated with nodal add/drop traffic values from beta distributions with $a = b = 0.1$ and $a = b = 1$ respectively. The individual values for the demands between nodes were determined as follows:

$$D_{i,j} = \frac{D_i \times D_j}{C}, \tag{5.17}$$

where D_i is the add/drop traffic of node i, D_j is the add/drop traffic of node j, and C is the chosen aggregate network capacity requirement of the network under design, 1 Tbps in this case. i and j are allowed to be equal, since the 60 nodes considered here are assumed to be above the lowest level of the multi-level network model as defined in section 4.1.

Figures 5.10 and 5.11 show how 8-bin histograms of the nodal add/drop traffic from the generated demand matrices follow the beta distributions from which they were generated. Eight bins were used in the histogram because 8 is the closest integer to

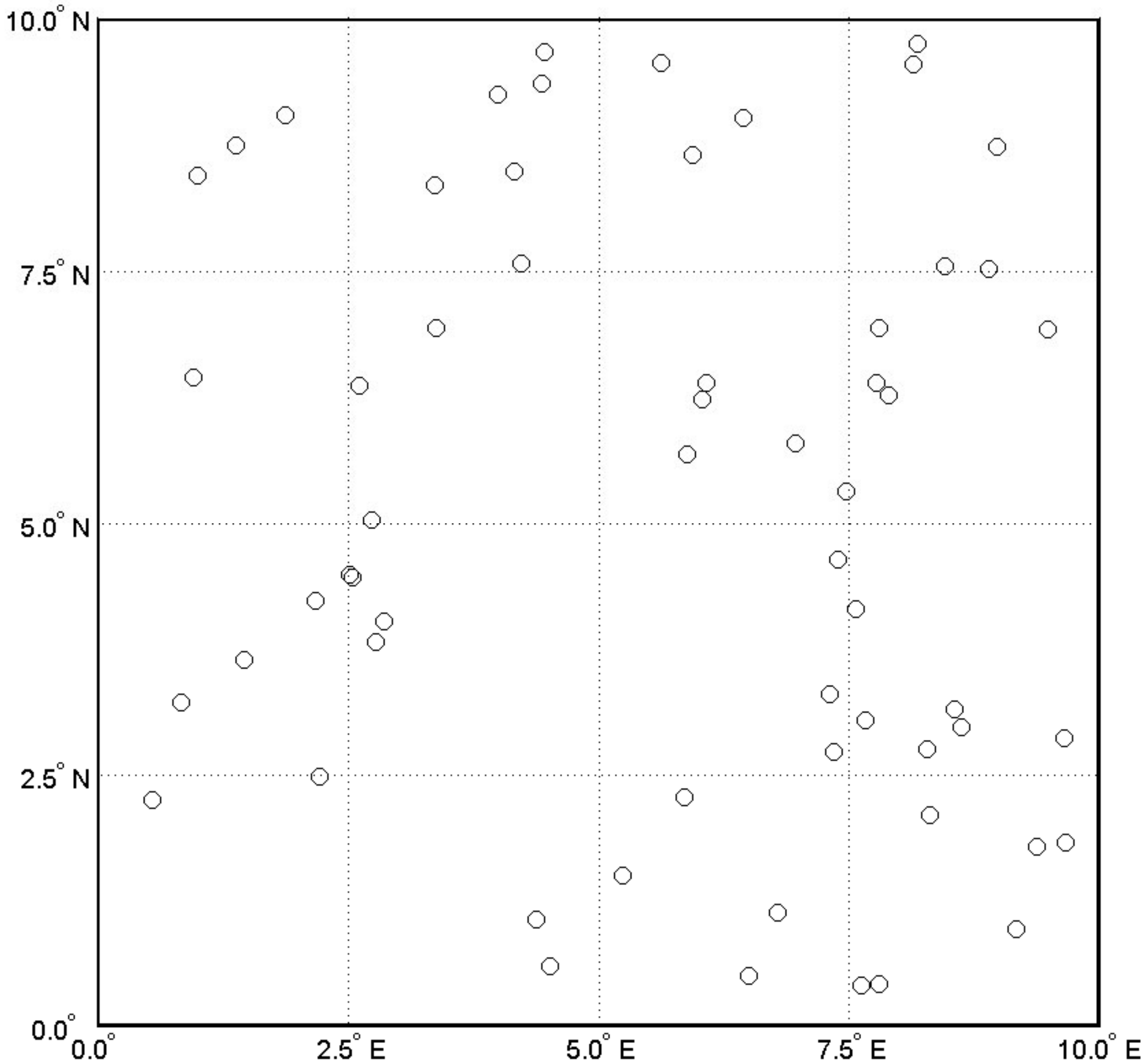

Figure 5.8: Geographical coordinates when generated from beta distribution with $a = b = 1$.

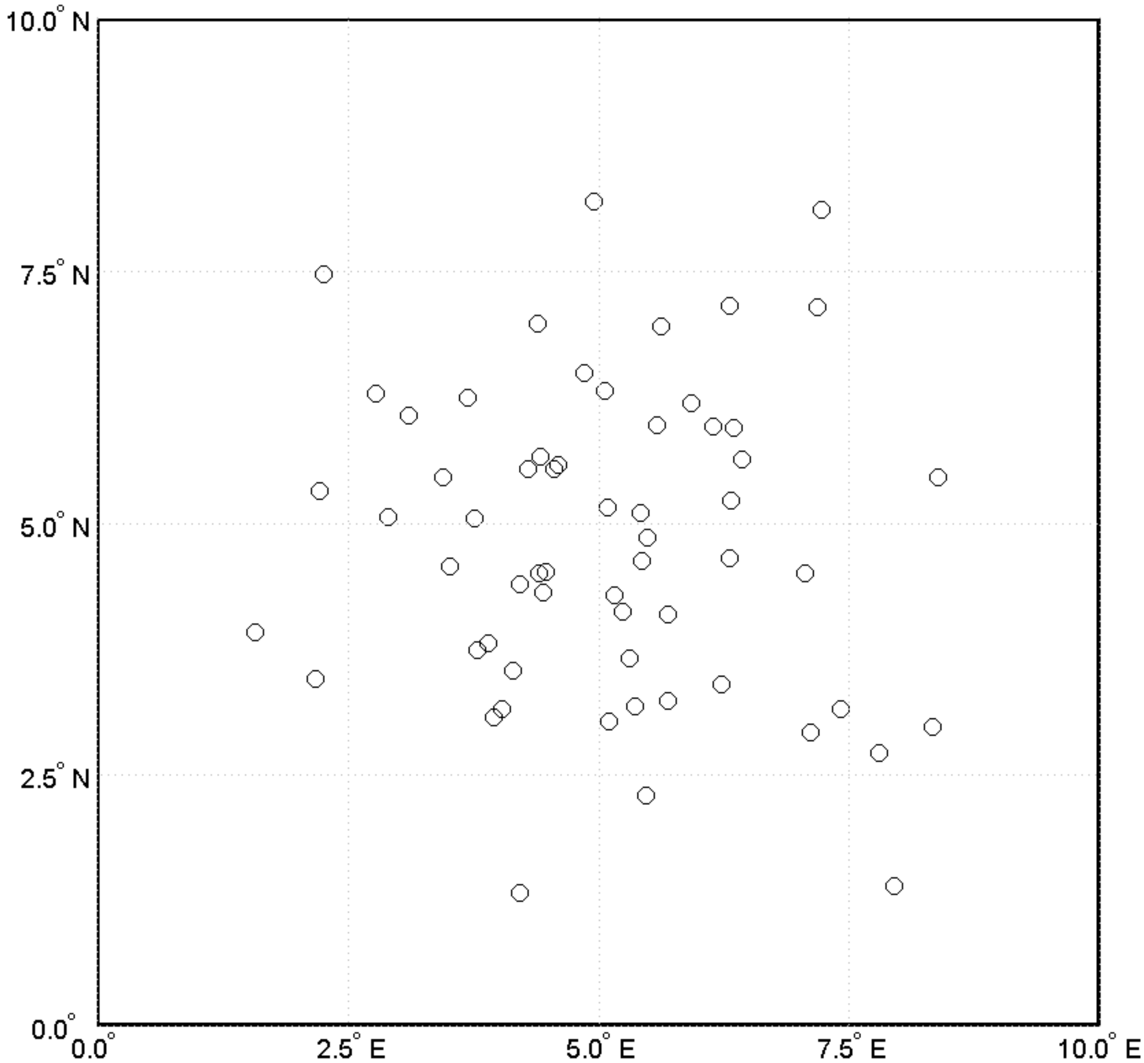

Figure 5.9: Geographical coordinates when generated from beta distribution with $a = b = 5$.

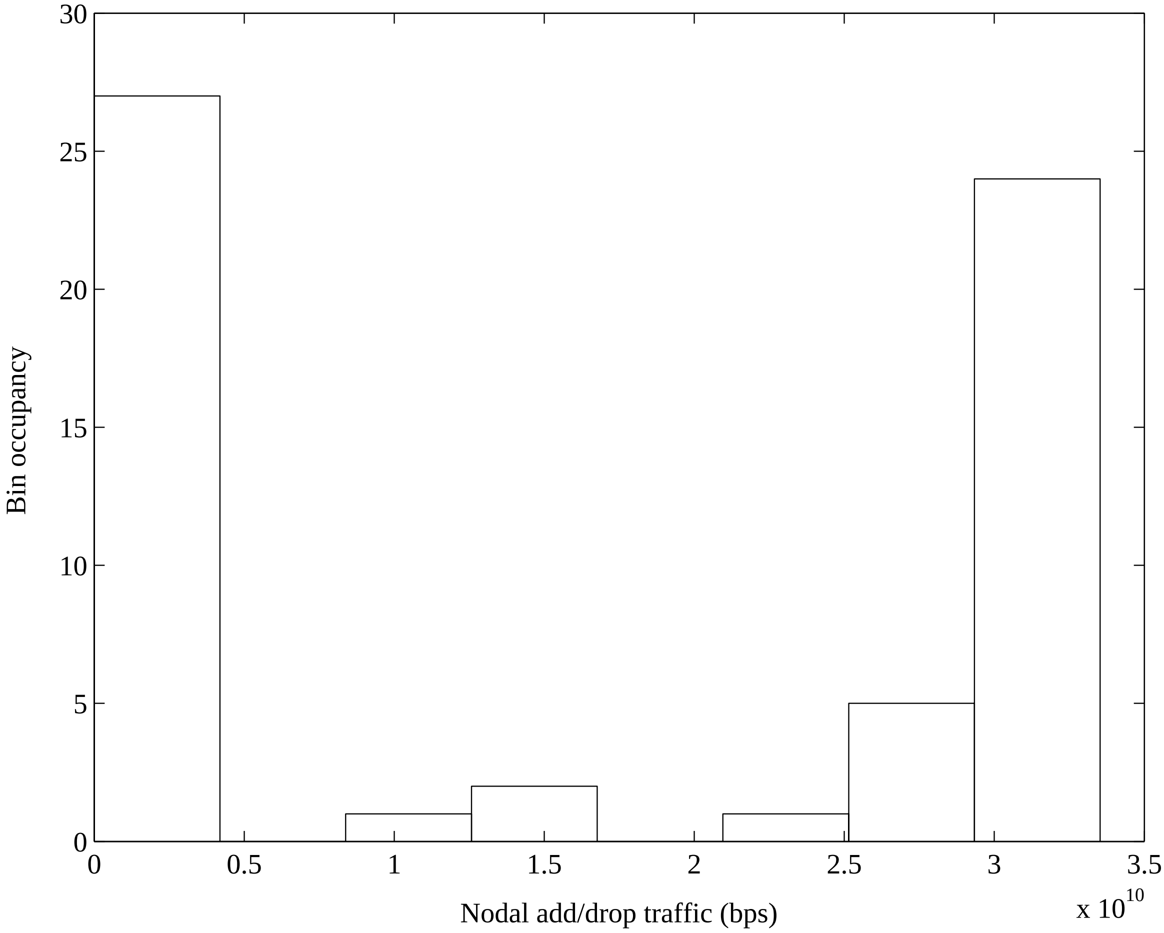

Figure 5.10: Histogram with 8 bins of nodal add/drop traffic when generated from beta distribution with $a = b = 0.1$.

the square root of 60, which is the number of network nodes. The demand matrices generated in this simulation have the constraint of being symmetrical, due to the method by which beta distributed nodal add/drop traffic values are used to generate the demand matrices.

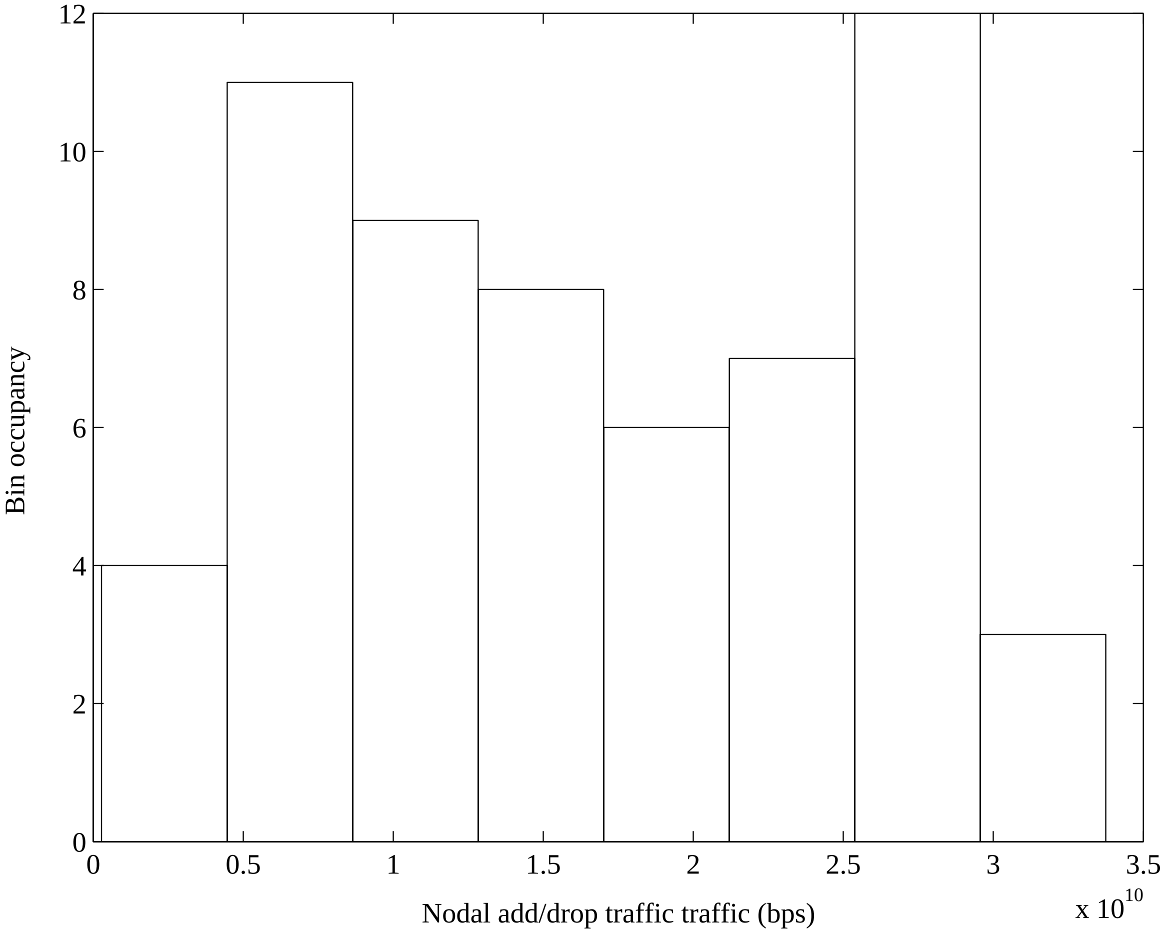

Figure 5.11: Histogram with 8 bins of nodal add/drop traffic when generated from beta distribution with $a = b = 1$.

5.3.5 Results and discussion

Evaluation of intra/inter-cluster traffic ratio

The results obtained for scenarios 1-4, as defined in section 5.3.4, are presented in figures 5.12 to 5.15. The mean intra/inter-cluster traffic ratio value for each number of hub nodes was calculated as follows:

$$\mu_N = \frac{1}{N}\sum_{l=1}^{N} R_m, \tag{5.18}$$

where N is the number of hub nodes for which a mean of the intra/inter-cluster traffic ratios is being calculated, and R_m is the intra/inter-cluster traffic ratio, as defined in section 5.3.3, of cluster m.

The standard deviation value for the intra/inter-cluster traffic ratio for each number of hub nodes was calculated as follows:

$$\sigma_N = \left(\frac{1}{N-1}\sum_{l=1}^{N} (R_m - \mu_N)^2\right)^{\frac{1}{2}}, \tag{5.19}$$

where N is the number of hub nodes for which a standard deviation of the intra/inter-cluster traffic ratios is being calculated, R_m is the intra/inter-cluster traffic ratio, as defined in section 5.3.3, of cluster m, and μ_N is the mean intra/inter-cluster traffic ratio for N hub nodes as defined in equation 5.18.

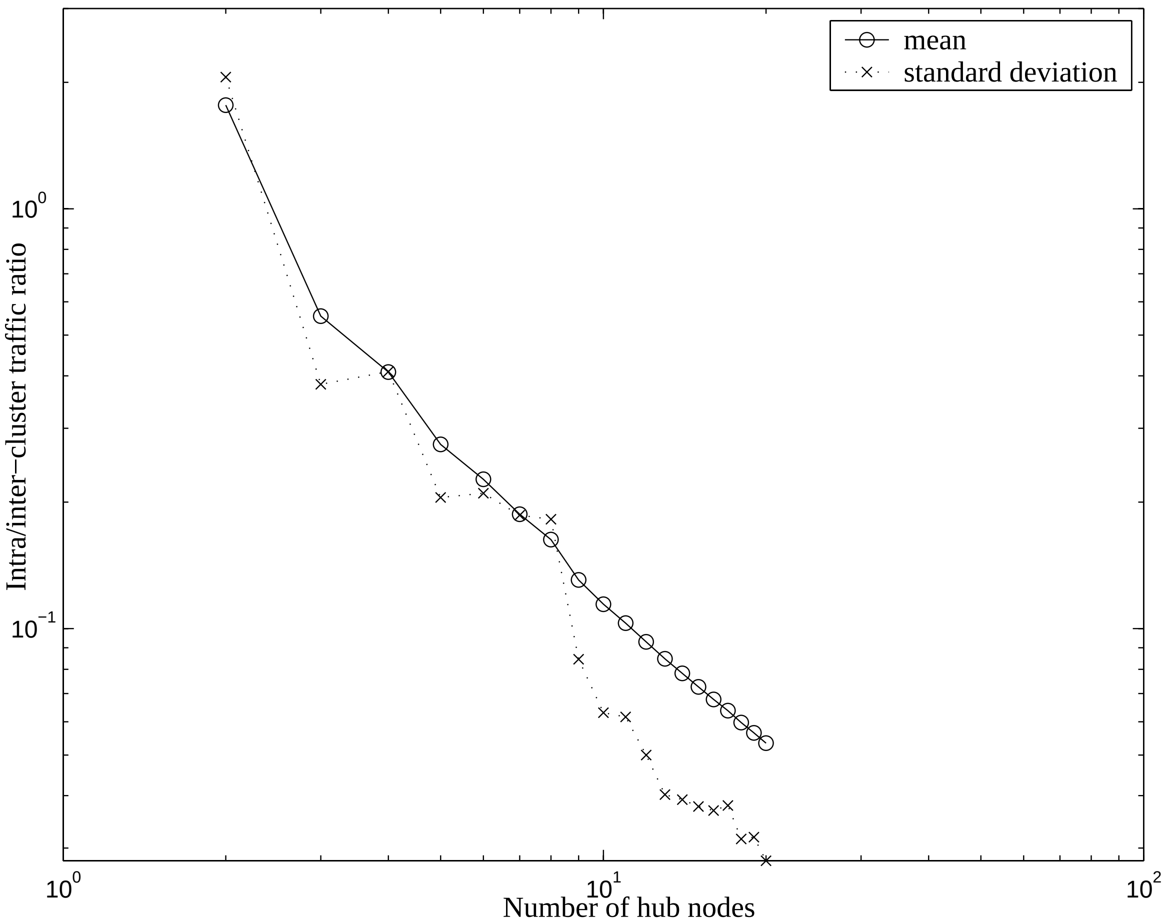

Figure 5.12: Means and standard deviations of the intra/inter-cluster traffic ratio as a function of the number of hub nodes, for scenario 1 as defined in section 5.3.4.

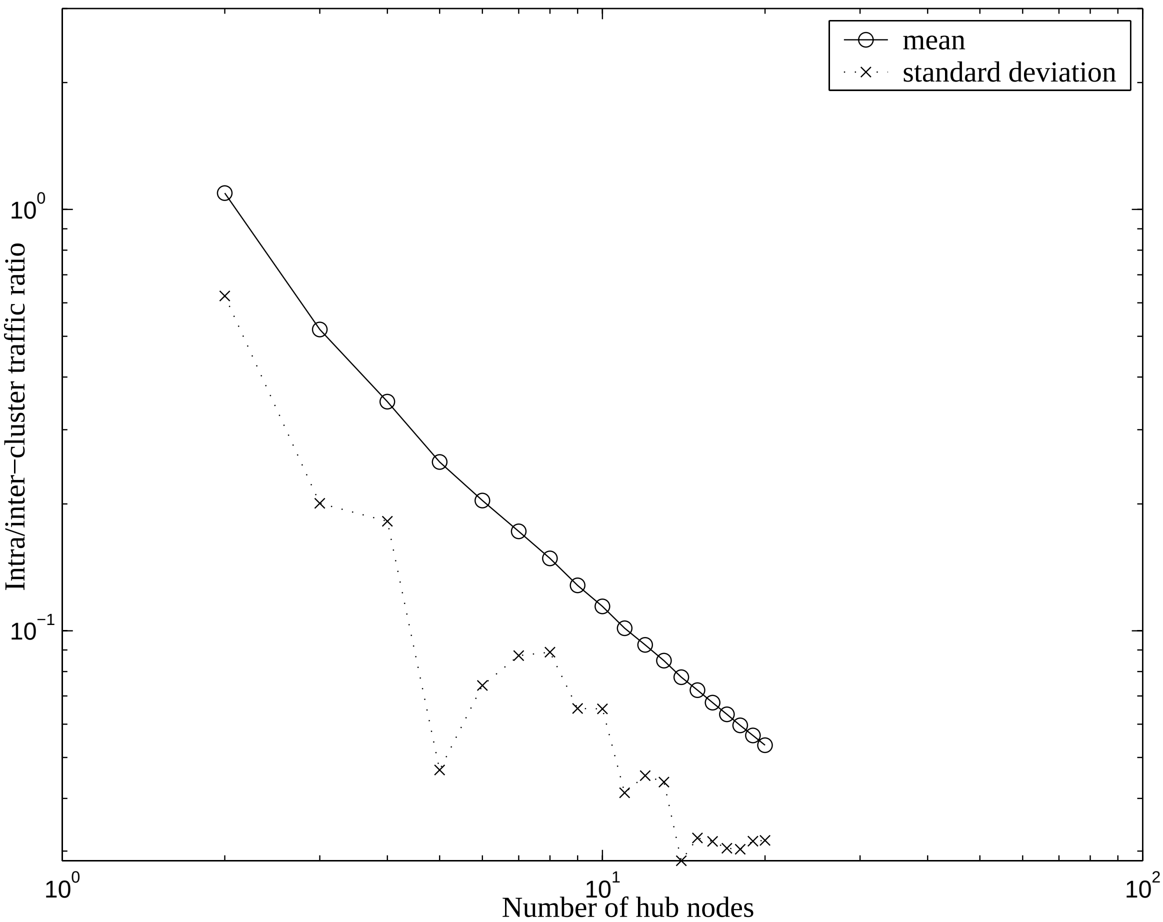

Figure 5.13: Means and standard deviations of the intra/inter-cluster traffic ratio as a function of the number of hub nodes, for scenario 2 as defined in section 5.3.4.

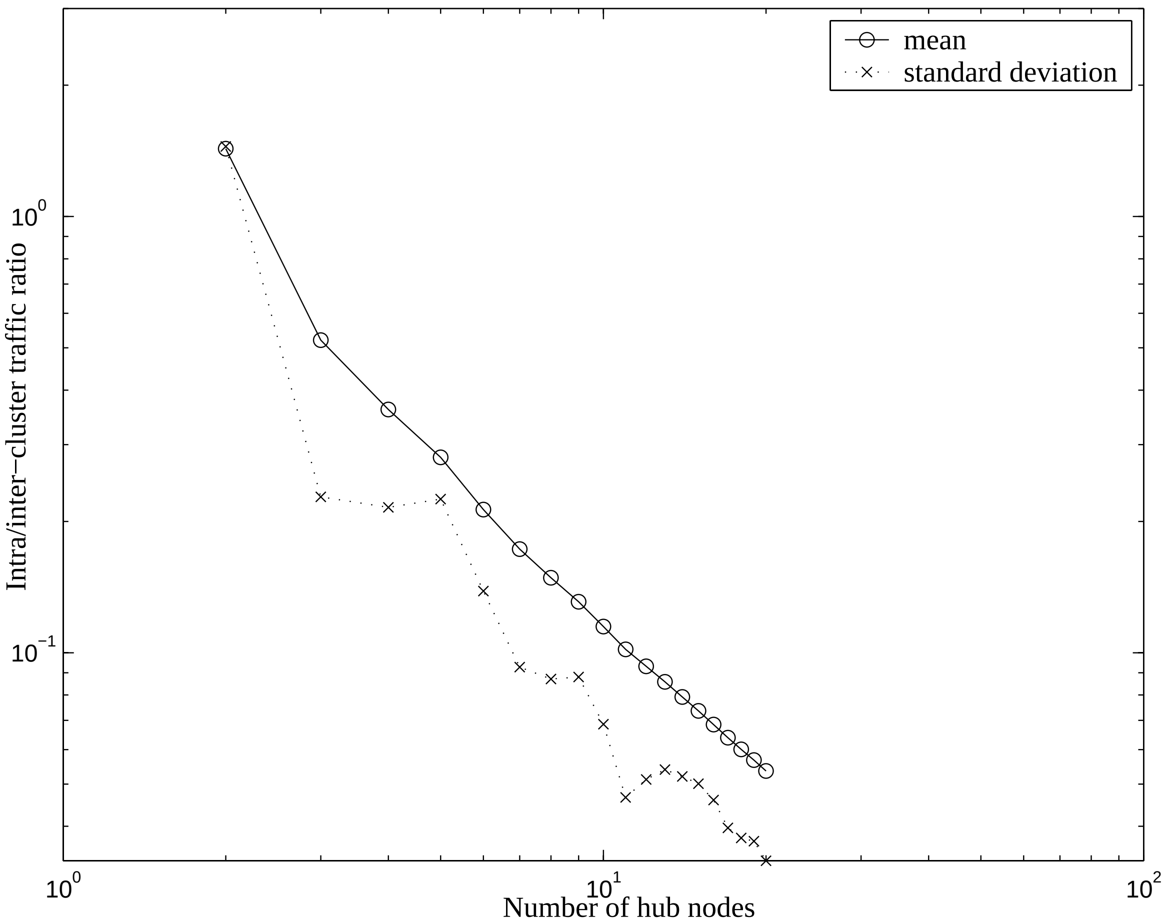

Figure 5.14: Means and standard deviations of the intra/inter-cluster traffic ratio as a function of the number of hub nodes, for scenario 3 as defined in section 5.3.4.

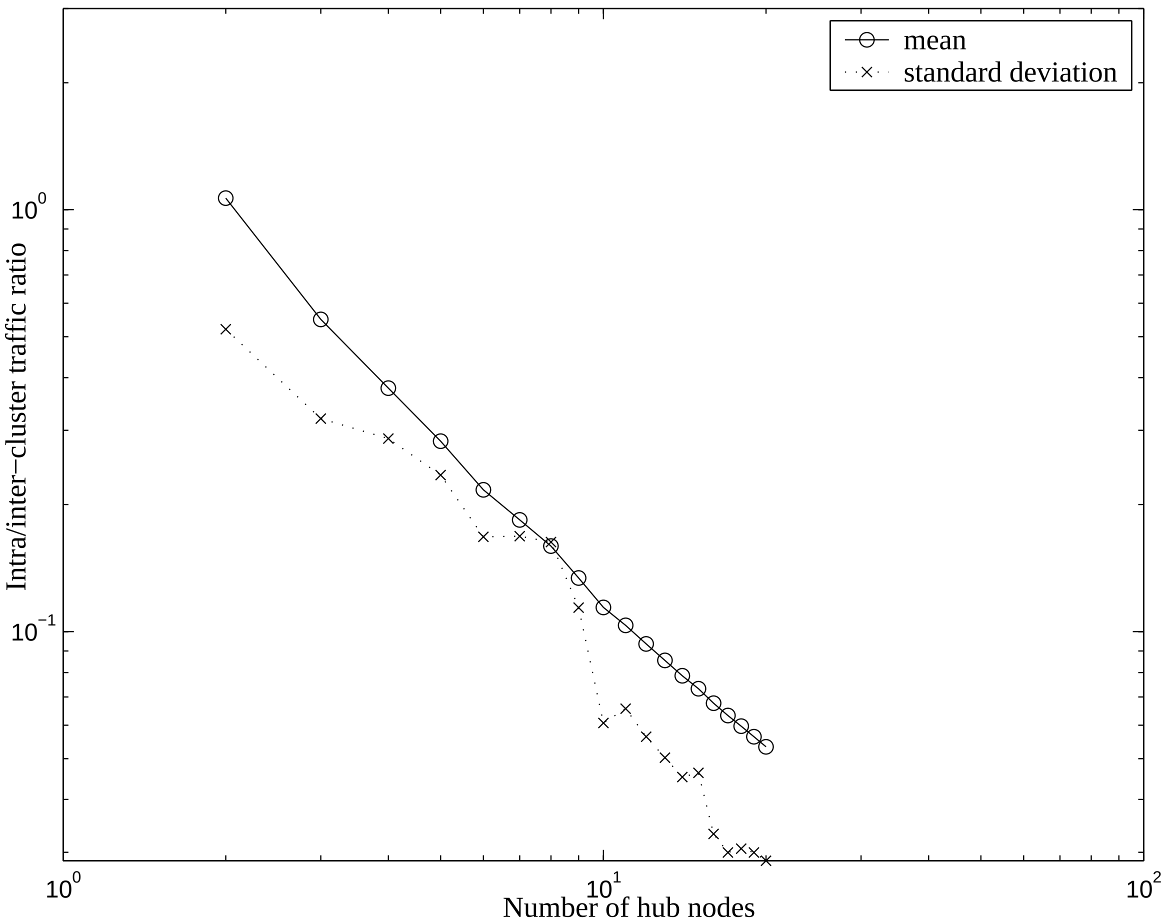

Figure 5.15: Means and standard deviations of the intra/inter-cluster traffic ratio as a function of the number of hub nodes, for scenario 4 as defined in section 5.3.4.

Discussion

In order to determine the number of hub nodes required on the backbone level of the multi-level network model for each of the four scenarios, as defined in section 5.3.4, the results obtained in section 5.3.5 have to be evaluated against a target or preferable intra/inter-cluster traffic ratio. The guidelines provided in the discussion of section 5.3.3 were considered in selecting the target intra/inter-cluster traffic ratio to be 0.5. This target value is lower than the value of 1.5 recommended in section 5.3.3 for the selection of hub nodes on the second level of the multi-level network model.

Table 5.1 shows the number of hub nodes required on the backbone level of the multi-level network model for each of the four scenarios outlined in section 5.3.4, for various target intra/inter-cluster traffic ratios, including the target of 0.5.

The scenarios are quite evenly matched when it comes to the number of backbone hub nodes required at different intra/inter-cluster traffic ratio targets. The results obtained in this experiment can thus not be used to identify any trends about the sensitivity of the number of hub nodes to the distributions of the nodal add/drop traffic or geographical coordinates. It can however be seen that the required number of hub nodes decreases as the target intra/inter-cluster traffic ratio increases.

Scenario	$R_t = 0.2$	$R_t = 0.5$	$R_t = 0.8$
1	7	3	3
2	7	3	3
3	6	4	3
4	7	4	3

Table 5.1: The number of backbone level hub nodes for each of the four scenarios, defined in section 5.3.4, according to various target intra/inter-cluster traffic ratios indicated by R_t.