Optimal Link Capacity Distribution and Network based Study

Divyayan Dey

June 2020

1 Introduction

The study of capacity/bandwidth allocation in real networks has been of vital importance since the gargantuan growth of data and its concerned flow methodology in complex networks. As traffic keeps growing exponentially with time, efficient data routing and forwarding protocols have emerged necessary for a properly-functioning network. The development of a proper bandwidth allocation strategy is important in the context of congestion and data loss. The design of a network so that it can reduce the inherent losses is also an esteemed area of work in the field of network science. An optimal link capacity has to be obtained for efficient flow of traffic and economic management of network resources. The definition of congestion has been kept very general as it can have several implications depending upon the network under consideration.

Various ways of alleviating network traffic by formulating routing protocols and data processing elements in a network have been shown in previous literature[add citations]. Many of them

Our work is based on the evaluation of probability mass function of packets flowing through each edge. Unlike in Zhao et al.(2005)[2], packet flow here is given by a statistical distribution rather than a numerically defined data rate. Hence, we do not observe any critical point of congestion in the networks. We will employ the idea of shortest paths for the flow of packets as our principal routing protocol. The packet flow model is described in the next section. The method of evaluation of the pmf has been taken from [1] where they discuss an algorithmic approach of finding a probability distribution of packets through links. Through the ongoing discussion, we will present a statistically defined conceptual bandwidth allocation procedure and the influence of network models on the scheme. Capacity of a link is defined as the number of packets sent above which the link becomes congested. If packets more than the assigned capacity are sent through a link, those packets are considered lost and will be a basis of optimization later in this work. Before we get to the evaluation of pmf and placing link capacities, let us define the nature of traffic we have used for our study and simulation. A number of networks are also presented which have necessary real world properties along with random graphs. Our goal will be to arrive at inferences which are as much global as possible, independent of network metrics and data flow nature.

2 Traffic Flow Model definition

The packet flow model which we propose for our work consists of a packet generation model and a packet routing model. Packets will be generated at source nodes following a stochastic distribution and will be sent towards destination nodes. To avoid any formation of queues at intermediate relay nodes, the packet transmission speed is taken much higher than the packet generation speed. When a packet enters a node to be routed, the connected links have to be taken into consideration. According to our protocol, the shortest path between source and destination nodes will be accounted for. If there are multiple shortest paths between the nodes, one of them is chosen at random assuming all of them have same efficiency of transmission.We employ the Open Shortest Path First(OSPF) routing protocol for our routing strategy. To reciprocate with real world communication systems, we use a Poisson distribution of packet flow throughout the model. However, our proposed model is independent of the nature of traffic flow. To avoid congestion in links, there has to be a capacity allocated to each link in the network, which can be either specified by a statistical congestion criteria(such as links remain free for x% of a time frame), or it may be resource specified(given a finite capacity to distribute among links). Hence, we have applied both ideas of finite and infinite bandwidth for edges to satisfy both the aforementioned criteria. All links are assumed homogeneous and processing time period at intermediate nodes is infinitesimally small.For the sake of simplicity, the capacity for processing information is considered equal at all nodes.

Here, a link in the network is said to be congested if the number of packets flowing through it exceeds its assigned capacity. For a Poisson distribution , packets are transferred from one node to another with a mean data rate λ. Higher the value of λ,more will be the packets transferred per unit time. Intuitively, when the data rate is low, capacity assigned to links need not be as high as during rapid transfer of data. This would indicate a higher allocation of bandwidth to maintain the statistical condition of congestion. But to maintain a finite resource specification, the packet transfer rate has little to do with the capacity allocated, as it then becomes dependent on the link's network properties, such as edge betweenness centrality which tells how often does an edge appear in the network's shortest paths.

3 Defining Network Models

We have implemented several network models , **(a)**that follow real world properties such as preferential attachment and growth, the Barabasi-Albert (BA) graph. The BA network is specified by a given number of nodes and num-

ber of neighbours for each node. This is a well known model that follows the power law distribution $P(k) = k^{-\gamma}$.A node is created in the network on every iteration and is connected to neighbours having higher degree centrality in the network. Such nature has been mimicked from real world communication networks such as the World Wide Web.**(b)**Networks that directly follow the power law degree distribution such as Power-law cluster network ,**(c)** Random graphs such as Erdos-Renyi with given number of nodes and edges(here edges and links convey the same meaning) and the n-m random graph ,having n nodes and m edges have also been studied to observe results in arbitrary random networks. **(d)**Another type of real world network used is the small world network,which bolsters many social networks. One such is the Newman-Watts-Strogatz network, which creates a ring lattice on a number of nodes each having the same number of neighbours. New edges are added to the network with a probability p. Hence, our proposed allocation strategy will be applicable for any network at large. We introduce an additional network parameter q_{xy} ,the probability of link existence between nodes x and y ,which will be used throughout the paper.This value of q has been kept same between all nodes for simplicity. In a later section, we will discuss if different networks have any influence at all, or of what kind on our capacity allocation mechanism.
In the next section, we will discuss capacity allocation in two different ways: one is based on the statistical criteria fulfilment,i.e.,to ensure that for x % of a time frame under consideration, all links remain free; the second is based on a resource management criteria,i.e., a finite network bandwidth to be allocated on the edges so that on a whole, the network remains the least congested(considering all links). The capacity of a link in a network is dependent on the statistical criteria ;intuitively, higher performance requirements need more capacity per link compared to lower performance values. To provide free links for larger period of time, the link bandwidth has to be higher.

The probability mass function here is expressed as the linear convolution of all the random variables that make up the sum. $\Omega_{ij}^{mn}(\lambda_{mn},q_{mn},f_{ij}^{mn},k)$ denotes the probability that k packets sent from node m to node n will pass through the link ij ; λ_{mn} is the mean rate of transfer of packets from node m to node n , q_{mn} is the probability that a link exists between nodes m and n , f_{ij}^{mn} is the probability that links from node m to n have edge ij on them.
To calculate Π_{ij}, we define the set $A_{ij} = \{(x, y) \mid x \in V, y \in V, x \neq y, f_{ij}^{xy} \neq 0\}$
For all $(x, y) \in A_{ij}$, evaluate $\Omega_{ij}^{xy}(\lambda_{xy}, q_{xy}, f_{ij}^{xy}, \sigma)$ for σ being an integer lying in the range $[0, Q-1]$ such that:

$$\Omega_{ij}^{xy}(\lambda_{xy}, q_{xy}, f_{ij}^{xy}, Q-1) << \epsilon\Omega_{ij}^{xy}(\lambda_{xy}, q_{xy}, f_{ij}^{xy}, \lambda_{xy}) \quad ,0 < \epsilon << 1$$

and store the values as defined in our paper in the vector Ω_{ij}^{xy}
Perform the convolution of all the vectors Ω_{ij}^{xy} for all $(x, y) \in A_{ij}$.The resultant distribution will be the probability mass function of packets flowing through edge ij.

4 Allocation based on Statistical Criteria

In this discussion, we will investigate into the variation of the network capacity with varying network topologies and the network parameters taken under consideration in our work. The relation of capacity allocated for each link of the network and also, the total capacity of the network with the prescribed parameters mentioned below,will be studied. Our motive is to arrive at a universal allocation strategy irrespective of how packets are sent or how the network architecture is defined.
For each edge ij in the graph, we will find its probability mass function Π_{ij} which has been evaluated earlier to determine its capacity. To ensure a given performance that links remain free for x % of the time, we need to find the cumulative distribution function so that $\sum_{k=0}^{l} \Pi_{ij} = x/100$, where l is the capacity to be allocated for link ij. The summation of all these capacities $l \, \forall \, ij$,is referred to as our network capacity. The variation of network capacity with different networks helps us decide on a good topology which will require minimum resources among all structures. We will also look at the marginal increase in allocated capacity with increasing performance criteria. That tells us how much additional resource were to be needed if links were to be made more congestion free. For now, we have assumed homogeneous congestion criteria(all links stay free for the same fraction of time). However, these simplifications won't cause any significant deviation in results when applied to a real world network.
Network architectures of different types were taken for simulation. The results are noted down in a later section.

5 Allocation based on Finite Capacity Distribution

In practical traffic networks,where capacity of the network is limited by dataflow hosts, it is an important issue to distribute the network data capacity among the links so that loss of data through the network can be kept minimum. In this discussion,we produce the idea of data capacity distribution in traffic networks and the factors that control the distribution of these capacities among the network links.
We are given a network (V, E) and a network capacity $S \in Z$,which has to be distributed among the N edges of the graph,so that the summation of loss of data flowing over all edges of the networks is minimum. For the sake of simplicity, we assume data is transmitted in packets and those packets are generated from source nodes following a Poisson distribution with mean λ. Another parameter used here is the probability of link existence q discussed in our paper. The loss of packets for an edge is drawn from the probability mass function of the packets flowing through that edge.
Therefore,our problem can be expressed as follows:
$min \ \sum_{ij} \sum_{c_{ij}}^{\infty} \Pi_{ij} \ \ s.t.$
$\sum_{ij} c_{ij} = S, \forall \, ij \in E$ and c_{ij} is the capacity assigned to edge ij

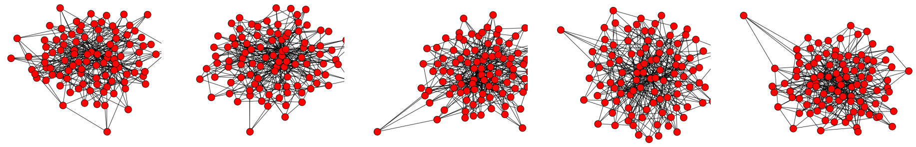

(a) (b) (c) (d) (e)

Figure 1: Barabasi-Albert graphs with 100 nodes and 4 connected neighbours. Number of edges in each graph=384

, provided nodes i and j agree to transfer packets between them.

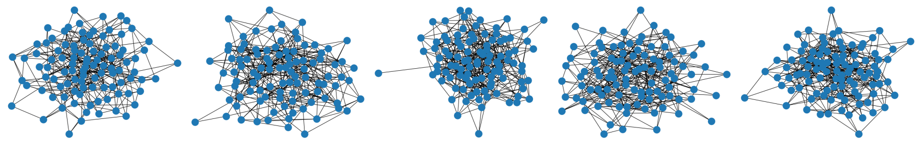
6 Solution Methodology

6.1 Statistical Performance wise allocation

We started with a Barabasi-Albert graph with 30 nodes and 3 connected neighbours. . We then performed our simulations with increasing number of edges for a Barabasi Albert graph by increasing its number of connected neighbours - 3,4,5,6,keeping q and λ fixed. These will show the capacity distribution with different data rates and different number of edges for the same network architecture. We will then move on to different networks - Newman-Watts-Strogatz, Erdos-Renyi keeping nodes and edges same so that a comparative analysis can be done among them as to the amount of resources needed for each of them. The sum of capacity for the network vs the probability of being congestion free were plotted. We will refer to this as **Method 1** in the results.
5 different network architectures as mentioned - (a) Barabasi-Albert, (b) Newman-Watts-Strogatz, (c) Erdos-Renyi, (d) Power-law Cluster graph and (e) n-m random graph which is a arbitrarily generated graph with n nodes and m edges are taken. The probability of being congestion free for each link of the network(i.e., the local performance criteria) was varied from 0.1 to 0.9 in steps of 0.1. Number of nodes in all graphs is 100 and number of edges are close to each other, due to random nature of these networks. The parameters λ ,i.e, the mean rate of packet transfer and q ,i.e, the probability of link existence($0< q <1$) between any two nodes of the network was set to 4 and 0.6 respectively. The graph instances are shown in Figures 1-5.

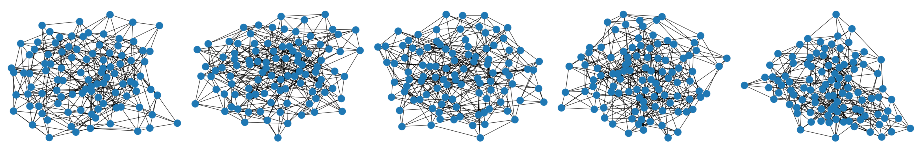
6.2 Finite Capacity Distribution

Our problem is a Mixed-Integer Nonlinear Programming Problem(MINLP). We used a BA graph with 30 nodes and degree=4 for our analysis.The value of λ was varied from 1 to 8 throughout the simulation. Varying the capacity of an edge,the loss for that edge has a variation as shown below(Figure 1). As we observe,the region to the left of the vertical line is non-convex while the region



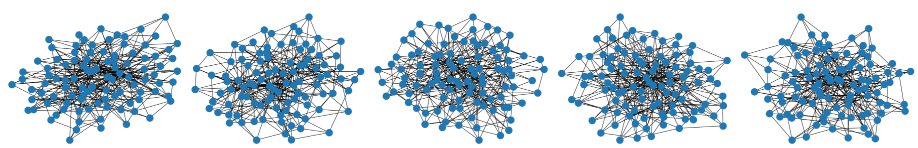
(a) (b) (c) (d) (e)

Figure 2: Erdos-Renyi graphs with (a)100 nodes and 318 edges, (b)100 nodes and 347 edges, (c)100 nodes and 383 edges, (d)100 nodes and 350 edges, (e)100 nodes and 360 edges



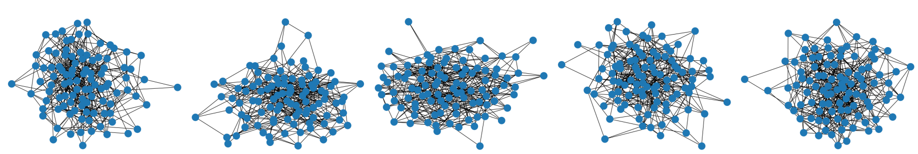
(a) (b) (c) (d) (e)

Figure 3: Newman-Watts-Strogatz graphs with (a)100 nodes and 337 edges, (b)100 nodes and 350 edges, (c)100 nodes and 352 edges, (d)100 nodes and 343 edges, (e)100 nodes and 344 edges



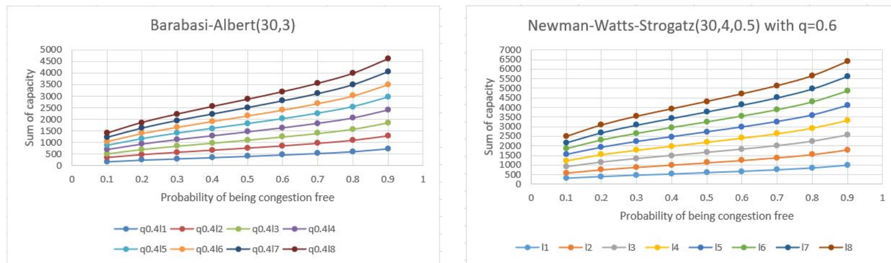
(a) (b) (c) (d) (e)

Figure 4: Power-law cluster networks with (a)100 nodes and 380 edges, (b)100 nodes and 378 edges, (c)100 nodes and 383 edges, (d)100 nodes and 380 edges, (e)100 nodes and 382 edges



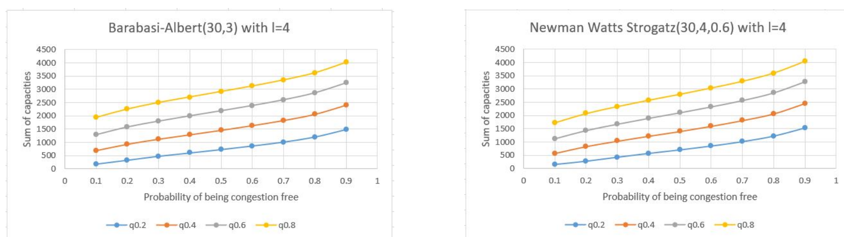
(a) (b) (c) (d) (e)

Figure 5: Random graphs with 100 nodes and 384 edges



(a) (b)

Figure 6: Capacity variation with λ



(a) (b)

Figure 7: Capacity variation with q

to the right is convex.Non-convex functions do not guarantee reaching a global optimum. Hence, we have tried two approaches:
(a) Vary the capacity of all edges from 0 to infinity(computationally it will be upto a fixed length as the pmf is generated upto a finite length).The optimization is run for several times and observed if they show any considerable changes,because the optimization might get stuck each time on a local minima.
(b) Vary the capacity over a portion of the entire range $(0, \infty)$. As the loss versus capacity curve changes from concave to convex,there must exist a point of inflexion at $x = x_0$ where the change occurs.Optimizing the capacities ranging from x_0 to the end will be a convex optimization which guarantees a global minimum.This point x_0 must be the peak of the probability mass function for the packet flow at that edge.This becomes clear from the nature of Poisson distribution as shown in Figure 2.The distribution rises initially,reaches a peak value,then falls again. If the rising edge of the distribution gives a concave nature of loss,then the falling edge must give an opposite nature,i.e., convex.

The comparative analysis of the two approaches is discussed in the section below. This will be referred to as **Method 2**.

7 Simulation Results

7.1 Method 1

As observed, the capacity of the network increases with increasing edges. This is clear from the fact that adding more edges to a network needs additional capacity to be allocated to the new edges. Barabasi-Albert(30,3) denotes 30 nodes and 3 connected node neighbours in the graph. Similarly, Newman-Watts-Strogatz(30,4,0.6) denotes 30 nodes , 4 node neighbours and probability of wiring a new edge = 0.6. The nature of increase is roughly linear in this case. Although we used Poisson distribution of packets between nodes in the network, these properties should hold true for any stochastic distribution of packet transfer.

As the rate of packet transfer λ is increased, the capacity allocated for each link has to be increased ,thereby increasing network capacity, as is evident from Figure 6.The capacity also increases with increasing values of q, based on the reason that higher q indicates more edges created in the network, hence, increase in allocated capacity(Figure 7). This trends of variation with q and λ hold true irrespective of any network used.Although we used Poisson distribution of packets sent between nodes in the network, these properties should hold true for any distribution of packet transfer. The capacity allocated to an edge was found to be strongly dependent on its edge betweenness centrality, as it showed a high correlation coefficient(> 0.99) for all network architectures. The nature of variation is independent of any of the parameters or networks as seen from the plots. It shows an almost linearly increasing nature of variation.

The Newman-Watts-Strogatz network shows the highest allocation of capacity among all and the Barabasi-Albert network requires minimum allocation of capacity.

To investigate more into the nature of capacity allocation , we ran simulations on a number of random graphs - (a)Barabasi-Albert graph,(b)Newman-Watts-Strogatz to observe any unique nature in the distributions(Figure 8(a)). The network intrinsic parameters were set in a way such that the number of edges and nodes for each remain almost same(all of them being random graphs). All architectures showed a similar increasing nature of capacity as the local performance criteria was increased. This is justified because the less congestion in links, more the packet flow, so increase in capacity is implied. The variation of capacities with the link congestion can be assumed to be universal in nature, under the said constraints of our paper. To measure the marginal increase in allocated capacities with congestion, the slope of each of the plots in Figure 4(a) was drawn as in Figure 8(b). As observed, the marginal increase is lower at lower values of congestion, but increases after the local performance criteria crosses 0.5. Simplifying, a lower resource addition needs to be done when links are more congested than that when links become relatively free.

A number of random networks have been studied here and all of them show the same behaviour. Hence, this considerations can be taken to be global keeping under consideration our definition of congestion and nature of packet loss.

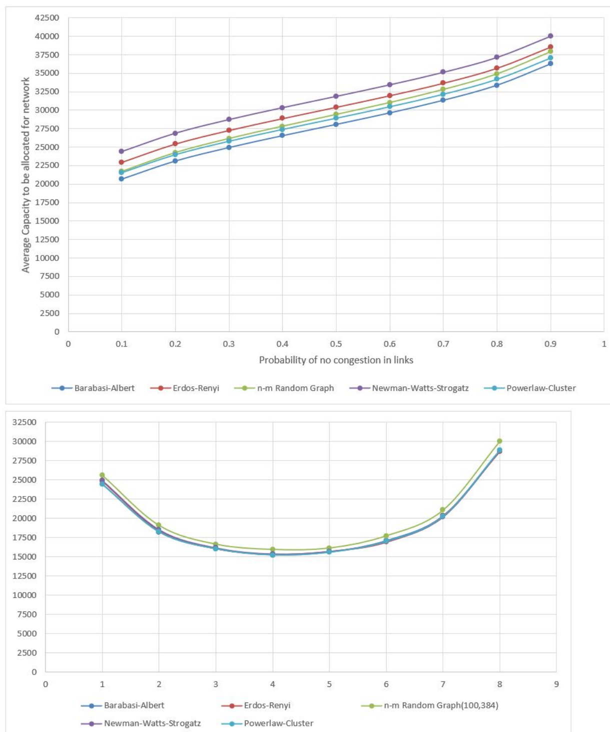


Figure 8: (a)The network capacity allocated versus the probability of no congestion ,i.e, free links, (b)The marginal increase in network capacities at varying probabilities of congestion in links

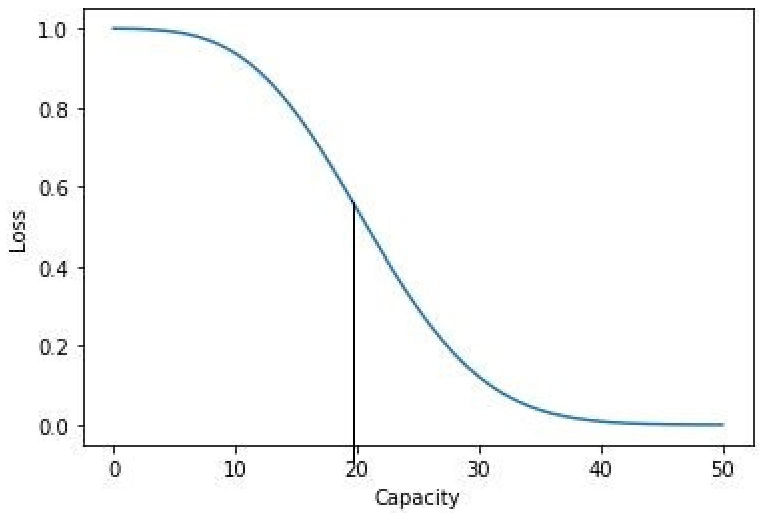


Figure 9: The plot of probabilistic loss of packets versus capacity allocated for an edge

7.2 Method 2

We used a Barabasi-Albert graph with 30 nodes and degree=4 for our analysis. Varying the capacity of an edge,the loss for that edge has a variation as shown (Figure 9). As we observe,the region to the left of the vertical line is non-convex while the region to the right is convex.Non-convex functions do not guarantee reaching a global optimum. Hence, we have tried two approaches:
(a) Vary the capacity of all edges from 0 to infinity(computationally it will be upto a fixed length as the pmf is generated upto a finite length).The optimization is run for several times and observed if they show any considerable changes,because the optimization might get stuck each time on a local minima.
(b) Vary the capacity over a portion of the entire range $(0, \infty)$. As the loss versus capacity curve changes from concave to convex,there must exist a point of inflexion at $x = x_0$ where the change occurs.Optimizing the capacities ranging from x_0 to the end will be a convex optimization which guarantees a global minimum.This point x_0 must be the peak of the probability mass function for the packet flow at that edge.This becomes clear from the nature of Poisson distribution as shown in Figure 10.The distribution rises initially,reaches a peak value,then falls again. If the rising edge of the distribution gives a concave nature of loss,then the falling edge must give an opposite nature,i.e., convex.

The Pyomo[3,4] *IPOPT* solver was used to allocate capacities to edges. The optimization was run on Erdos-Renyi graph with probability of edge creation(p) varying from 0.1 to 0.9. A capacity of 3000 was taken. To avoid discrepancies, we set the variable bounds to be within the convex region of the problem,i.e., from the pmf maxima to the pmf length for each edge. It should be taken care

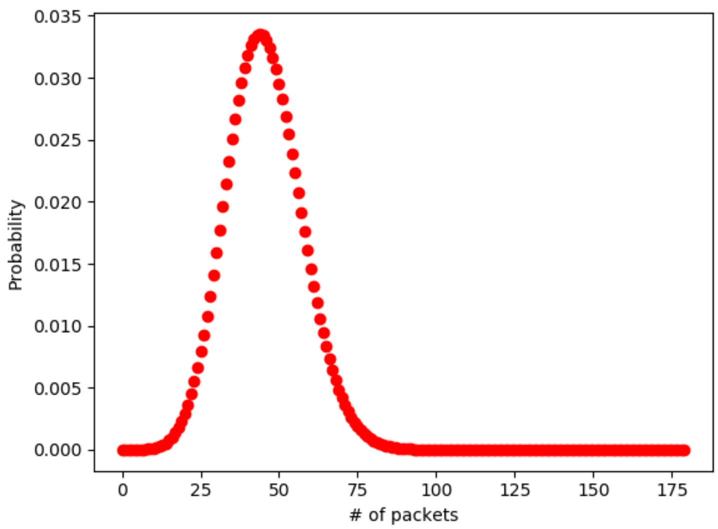
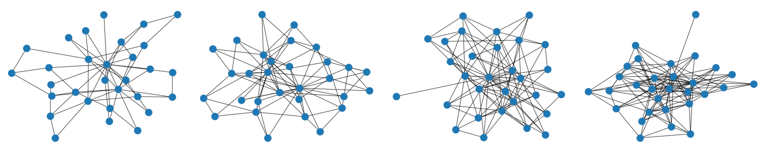


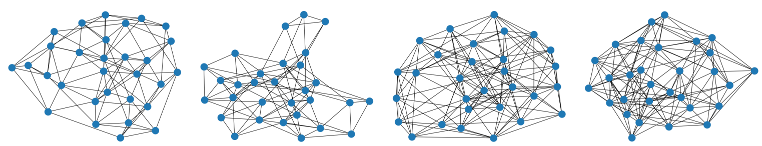
Figure 10: Probabilistic distribution of packets flowing through an edge

of that the network capacity should be above the sum of all pmf maximas, so that our problem remains a convex optimization. The results are shown below:
Erdos-Renyi Graph -
p=0.2, number of edges = 90, correlation coefficient=0.982
p=0.25, number of edges = 102, correlation coefficient=0.986
p=0.3, number of edges = 131, correlation coefficient=0.952
p=0.35, number of edges = 145, correlation coefficient=0.948
p=0.4, number of edges = 173, correlation coefficient=0.951
p=0.45, number of edges = 195, correlation coefficient=0.912
p=0.5, number of edges = 218, correlation coefficient=0.900
p=0.55, number of edges = 222, correlation coefficient=0.545
p=0.6, number of edges = 258, correlation coefficient=0.675
p=0.65, number of edges = 275, correlation coefficient=0.1613
For $p > 0.65$, correlation values were below 0.15. The capacities were obtained as floating pint values, although they should be positive integers , most likely the $IPOPT$ solver has this intrinsic behaviour. Here , we observe that the correlation coefficient drops drastically when the link existence probability exceeds 0.55. The primary reason behind this is that, at high values of p, the link centralities become highly homogeneous(variance in the order of 10^{-4}), leaving no significance for the centrality of an edge and the Pearson's correlation coefficient becomes dependent on the total network capacity instead of a link's individual existence in the graph.
We observed the capacity vs centrality plot for the network's edges at different



(a) (b) (c) (d)

Figure 11: (a)Barabasi-Albert graph with 30 nodes and 2 neighbours. Edges=56 (b)Barabasi-Albert graph with 30 nodes and 3 neighbours.Edges=81 (c)Barabasi-Albert graph with 30 nodes and 4 neighbours.Edges=104



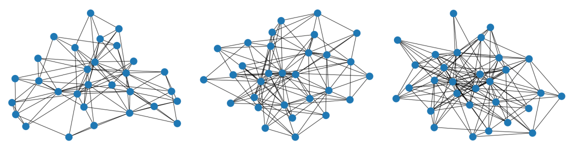
(a) (b) (c) (d)

Figure 12: (a)Newman Watts Strogatz graph with 30 nodes,4 neighbours and probability of link existence=0.6.Edges=91 (b)Newman Watts Strogatz graph with 30 nodes,5 neighbours and probability of link existence=0.6.Edges=95 (c)Newman Watts Strogatz graph with 30 nodes,6 neighbours and probability of link existence=0.6.Edges=141 (d)Newman Watts Strogatz graph with 30 nodes,7 neighbours and probability of link existence=0.6.Edges=145

values of p. The link capacity shows a linear dependence at low values of p , but at high values such as p=0.8, the relation is obfuscated.
Apart from a Erdos-Renyi Graph, other network architectures whose edges are not equitably distributed throughout, the edge capacities show a clean coherence commensurate with the edge's betweenness centrality.

8 Appendix

[1]. '*Link Capacity Distributions and Optimal Capacities for Competent Network Performance*',2020 - Pal S., Bakshi D., Chaterjee A., Mukherjee A.
[2]. '*Onset of traffic congestion in complex networks*',2005 - Zhao L., Lai Y., Park K., Ye N.
[4]. '*A review and prospect on the network congestion control*',2010 - Yan-tan T., Ping-ping X.
[3]. "*Pyomo: modeling and solving mathematical programs in Python.*" Mathematical Programming Computation 3, no. 3 (2011): 219-260 - Hart, William E., Jean-Paul Watson, and David L. Woodruff.
[4].'*Pyomo – Optimization Modeling in Python*'. Springer, 2017 - Hart, William



(a) (b) (c)

Figure 13: (a)Power-law cluster network with 30 nodes,3 neighbours and probability of adding a triangle p=0.6. Edges=102 (b)Power-law cluster network with 30 nodes,5 neighbours and p=0.6. Edges=123(c)Power-law cluster network with 30 nodes, 6 neighbours and p=0.6. Edges=140

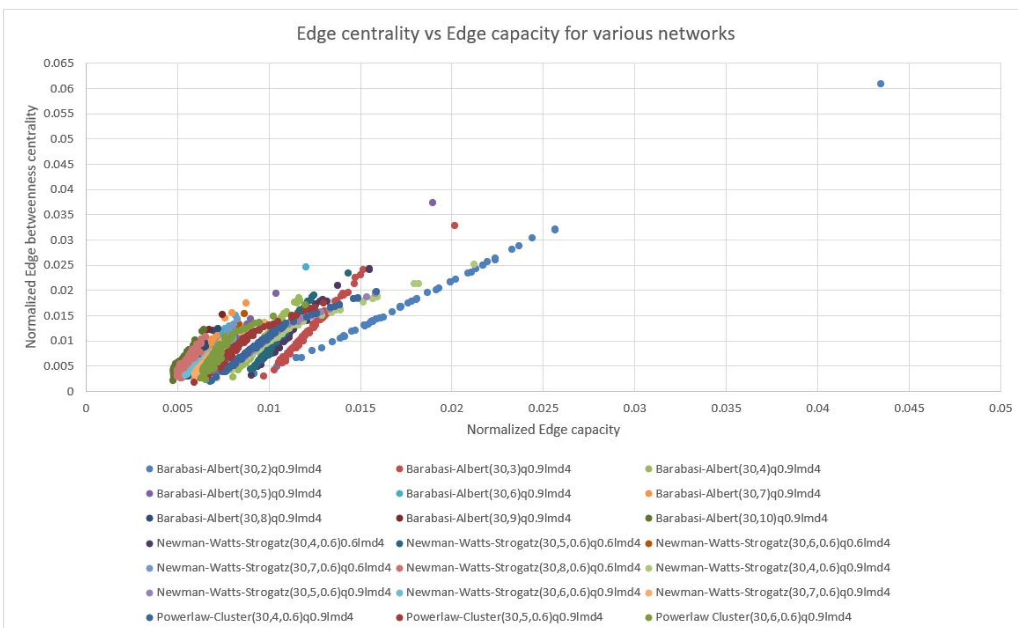


Figure 14: The plot describing how the betweenness centrality of edges in a network directly relates to its allocated capacity in the network. Plots for three major networks - the Barabasi-Albert, Newman-Watts-Strogatz and the Power-law Cluster graph are shown. The notation for each network is {Network_Name}q{value of q}lmd{value of λ}

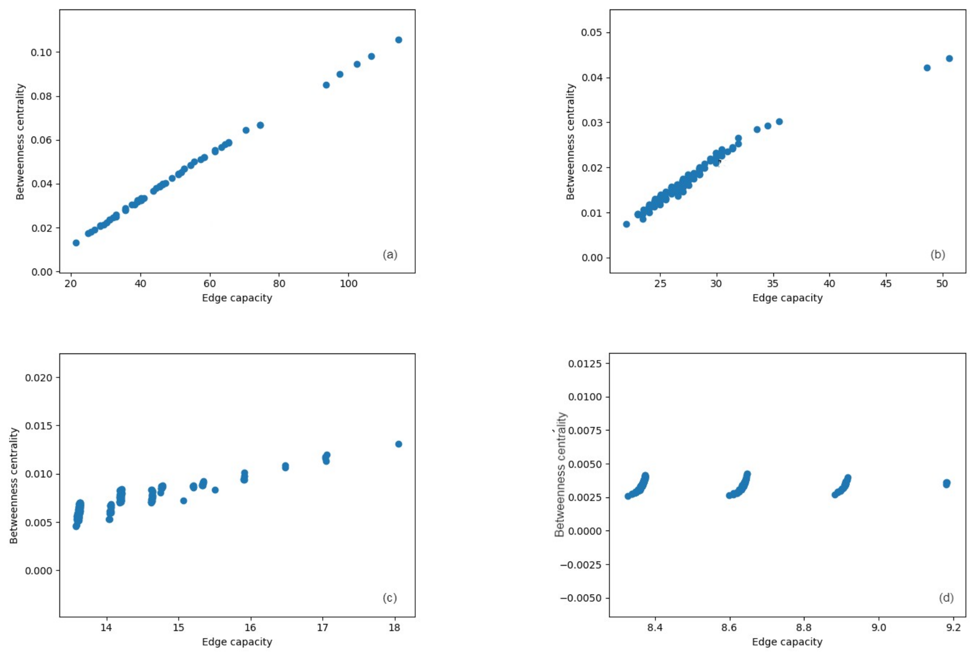


Figure 15: Scatter plots for betweenness centrality vs capacity of an edge for an Erdos-Renyi graph with 30 nodes at (a)p=0.15,(b)p=0.25,(c)p=0.50,(d)p=0.80

E., Carl Laird, Jean-Paul Watson, David L. Woodruff, Gabriel A. Hackebeil, Bethany L. Nicholson, and John D. Siirola.