Journal of Engineering and Development, Vol. 10, No.1, March (2006) ISSN 1813-7822

Network Topology Optimization Using Social Network Analysis

Prof. Dr. Siddeeq Y. Ameen
Computers & Information Technology Eng. Dept.
University of Technology, Baghdad, Iraq

Asst. Lect. Alaa N. Al-Qeicy
Laser and Electronic Optics Eng. Dept.
University of Technology, Baghdad, Iraq

Abstract

The topological design of networks in this work consists of finding a topology (directed and undirected links flow) that improve both the reliability index and increase the available shortest paths in the network without overloading the routing table or network with more links.

This paper proposes a procedure for generating a topology between ring and full-mesh topology which is the partial-mesh. This optimization work was done using the principle of Social Network Analysis (SNA) by focusing on the network betweenness centralization to improve the reliability.

The implementation of this work has been subjected to node/link failures before the optimization to find the important (critical) nodes links and after the design in order to test the validly of the improved network.

الخلاصــــة

إن التصميم الطوبوغرافي للشبكات في هذا البحث يتضمن إيجاد الطبوغرافية (المباشرة وغير المباشرة عبر مسارات المعلومات) التي تحسن كل من عامل الموثوقية وتزيد اتاحية اقصر مسار في الشبكة دون أي زيادة في الحمل لجدول مسارات الشبكة أو استخدام شبكة بمسارات أخرى.

يقترح هذا البحث سياق لتوليد طبوغرافية مابين الطبوغرافية الحلقية و الطبوغرافية المتشعبة والتي هي طبوغرافية التشعب الجزئي . وقد تم الاستناد في تحسين الأداء استخدام مبدأ تحليل الشبكة الاجتماعي وذلك بالتركيز على بينيات مركزية الشبكة لرفع مستوى الموثوقية .

إن التنفيذ تم بموجب التوقفات بين المسار والعقد قبل وبعد التحسين لإيجاد مسارات العقد المهمة والحرجة بعد التصميم لغرض فحص الشبكة المحسنة .

1. Introduction

It is difficult to imagine what modern living would be like without ready access to reliable, economical, and efficient means of computer communication. Although in many instances computers are used to perform their intended role in a stand-alone mode, in others there is need to interwork and exchange data with other computers. Computer network begun to start delivering services to private individuals at home such as to transfer a file of data from one personal computer to another or to access information from a public database using the switched telephone network, in the office for the exchange of electronic mail (e-mail), in schools and colleges to share the use of an expensive peripheral such as a laser printer, in universities and other research establishments to exchange and sharing information [1]. Another application of computer networks is access to information systems like the current World Wide Web (WWW), which contains information about the arts, business, health, sports, history, travel, and many other topics, and so on [1,2]. The Internet (usually mesh topology) is growing faster than ever with traffic across the core of the network quadrupling over the last year with a large geographical areas covering. Furthermore, the services ranging from video conferencing to large multimedia downloads are increased [3]. All these demands and network growth are required to design a network with good reliability.

2. Social Network Analysis (SNA)

Social network analysis is the mapping and measuring of relationships and flows between people, groups, organizations, computers and other information or knowledge processing entities. The nodes in the network are peoples and groups while the links show relationships or flows between the nodes. SNA provides both a visual and a mathematical analysis of complex human systems. One of the methods used to understand networks and their participants is to evaluate the location of actors in the network. Measuring the network location is finding the centrality of a node. All network measures discussed here are based on geodesics, the shortest path between any two nodes [4,5]. **Figure (1)** shows a network example used to distinction between the three most popular centrality measures: activity or degree, closeness, and betweenness.

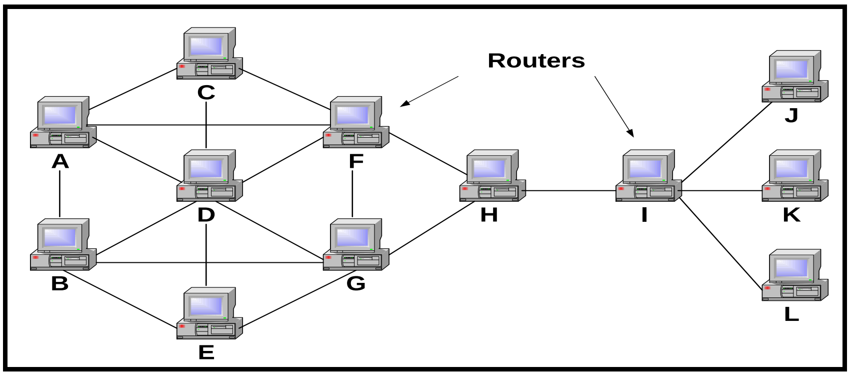

Figure (1) Social Network Analysis Model

Degree or activity is the number of direct connections a node has. In this human network, router D has the most direct connections in the network, making it the most active node (central) in the network with the highest degree count. Common wisdom in personal networks is “the more connections, the better” The important matters is where those connections lead to and how they connect the otherwise unconnected.

Closeness is the measure of shortest path from one node to all others. Router F and G have fewer connections than D, yet the pattern of their ties allow them to access all the nodes in the network more quickly than anyone else. While routers (J, K, and L) have greatest distance to others, it considers having less closeness because it passes through more intermediate nodes. Maximizing the closeness of only one or a few routers leads to counterproductive results.

Betweenness centrality is the measures of best powerful node in the network that plays as a “broker” or single point of failures. While router D has many direct ties, node H has few direct connections fewer than the average in the network. Yet, in many ways, it has one of the best locations in the network. It is a boundary spanner and plays the role of broker. It is between two important constituencies, in a role similar to that of a border router. The good news is that it play a powerful role in the network, the bad news is that it is a single point of failure. Without H, router I, J, K, and L would be cut off from information and knowledge in router D’s cluster.

2-1 Average Distance in Network

The shorter the path, the fewer hops it takes to go from one node to another. In human networks, short paths imply quicker communication with less distortion. In computer networks, the signal degradation and delay is usually not an issue. Nonetheless, a network with many short paths connecting all nodes will be more efficient in passing data through and to rearranging when any failures or topology changes happened. If there is more than one optimum path (geodesics) then the algorithm uses the shortest optimum path. Average distance length is strongly correlated with closeness throughout the network. As the closeness of all nodes to each other improves, the average distance also improves [4, 6].

2-2 Network Centralization

Individual network centralities provide insight into the individual’s location in the network. The relationship between the centralities of all nodes can reveal much about the overall network structure. A very centralized network is dominated by one or a few very central nodes. If these nodes are removed or damaged, the network quickly fragments into unconnected subnetworks. Highly central nodes can become critical points of failure. A network with a low centralization score is not dominated by one or a few nodes such a network has no single points of failure. It is resilient in the face of many local failures. Many nodes or links can fail while allowing the remaining nodes to still reach each other over new paths. One of the most important uses of social network analysis is identification of "most central" units in a network [4, 7]. In this work, we focus on network betweenness centralization to improve the networks.

3. Network Topology Design and Optimization

Social network analysis was used to design and optimize network topology, two different cases were enhanced, and the complete analysis was given.

3-1 Undirected Network Design

In this case the SNA was applied to National Science Foundation NETwork (NSFNET) backbone network shown in **Fig.(2)**. NSFNET backbone is used to connect the supercomputing centers in the USA in 1989. NSFNET was created by NSF (the U.S. National Science Foundation), The Complete NSFNET consists of backbone network connecting supercomputer and some (eventually about 20) regional networks that connected to the backbone to allow users at thousands of universities, research labs, libraries, and museums to access any of the supercomputers and to communicate with one another [2,4].

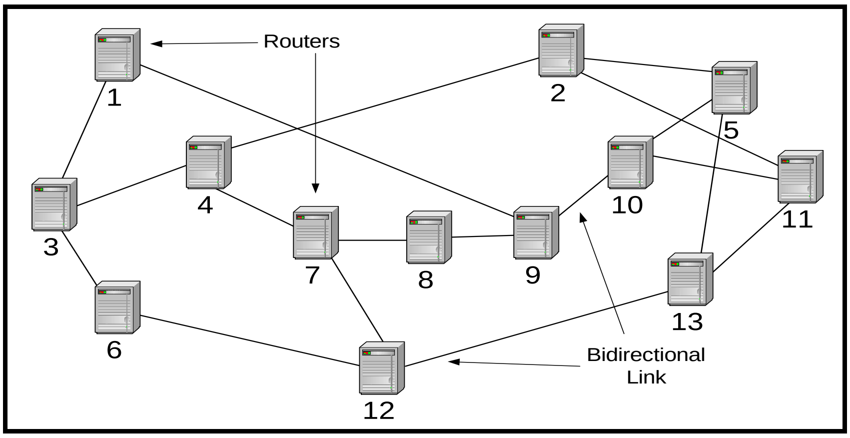

Figure (2) The NSFNET Undirected Backbone

To achieve good network design, three competing goals are balanced for it [4]:

i. Reducing hop count.

ii. Reducing available paths.

iii. Increasing the number of failures the network can withstand.

Taking the three competing goals internetwork design in consideration, failures were applied to the links and nodes in the network and show what happens to the metrics given, and the three goals. Two strategies were adopted in this work to improve the network topology. One possible way to improve the network topology is by moving the existing links to connect different pairs of routers or computers. Many cases were taken to rearrange links, but the case that we move the lowest edge (less importance) betweenness link (11-2) and (2-5) as shown in **Fig.(2)** to connect another pair give little improvement to some metrics.

The second strategy to improve the network topology is to increase the number of physical links. Better topology and robust network can be found by increasing the number of links in the network, and also the number of geodesics path and longest path (hops) was reduced. The following points were taken in consideration when the number of links in the network was increased:

1. The procedure of adding links begun with the nodes that have less direct connection to others (i.e. less activity or degree) [4].

2. The addition of links must be in places that decrease the importance of most centralized nodes and edge betwenness links in order to enhance the reliability of the network.

3. In order to reduce the distance/steps between nodes (intermediate node between source and destination). The adding of new links is preferable to be in the right place in the network.

4. The addition of new links must be suitable and not to overload the routing tables. The geodesics no. increased with links addition but it must be adequate.

5. Construct a topology similar to partial topology (between ring and full-mesh topology). Yet it quickly converges on a good solution even large networks improve quickly with just few added links such as the ring topology could reach the partial mesh topology with only four links with (network centralization = 0). **Table (1)** shows the options taken to improve the network.

Table (1) Strategy -2- Possible Network Improvement

Scenario	Network Centralization C_D	C_C	C_B	No. of Geodesics Path	Longest Path (hops)	Average Distance
Original net. -Fig.(2)-	0.023	0.102	0.062	200	4	2.205
1) add link: (1-6)	0.007	0.073	0.074	198	3	2.141
2) add link: (1-8)	0.007	0.084	0.050	202	4	2.142
3) add links: (1-6)(1-8)	0.090	0.047	0.046	190	3	2.090
4) add links: (1-6) (1-8) (7-10)	0.076	0.228	0.089	198	3	2.013
5) add links: (1-6) (1-8) (7-10) (2-6)	0.060	0.182	0.079	214	3	1.936
6) add links: (1-6) (1-8) (7-10) (2-6) (8-11)	0.045	0.142	0.051	246	3	1.872

where, C_D, C_C, and C_B represent network centralization for degree, closeness, and betweenness measures respectively. Options 1, 3, and 4 give the best improvement to network than others in the following metrics:

1. The longest shortest path was reduced to three hops.
2. The average distance was reduced.
3. Network centralization did not increase that much.
4. The number of paths for router to remember was reduced.

The design achieves efficient topology but the network is still weak to node, and link failure. Therefore, starting from these options links was added and the network was remeasured again. The iterative process was continued until no farther improvements happen (preferred to be partial-mesh topology) or when the network reached a good score centralization. **Figure (3)** shows the relationship of increasing links versus the average number of hops (average distance). Improvements of option (1) in **Table (1)** were actually implemented in NSFNET (July 1988-November 1992) [4,2]. Link (7-10) and (11-8) was also implemented (July 1988-July 1989) [8]. In computer network topology, the best and optimal topology that can stand to any nodes or links failure is the full mesh topology. However, to reach this topology, routers must have direct links to all routers in the network. The improved network now is more robust, even with the increasing in geodesics paths. **Figure (4)** shows the improved network.

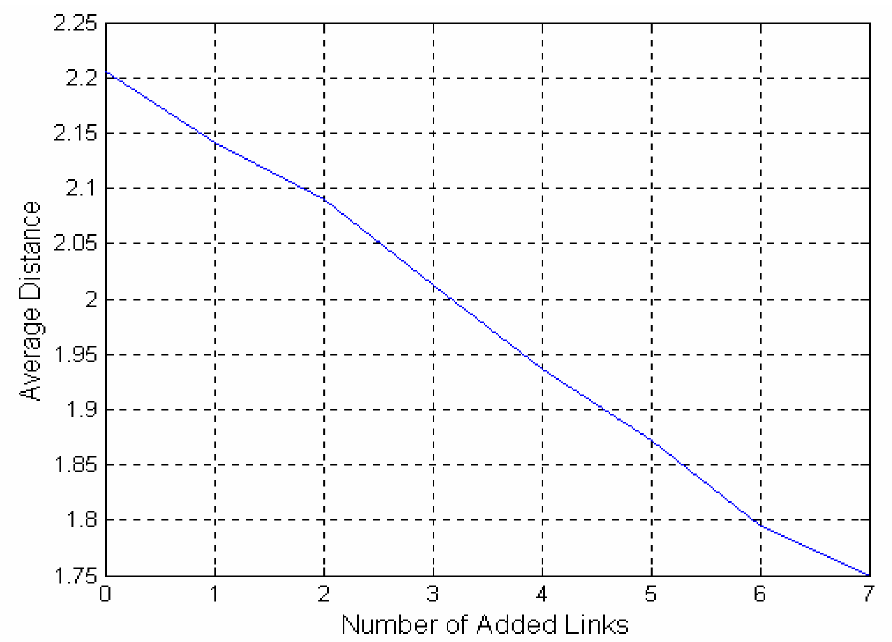

Figure (3) Relationship between the Numbers of added Links versus Average Distance

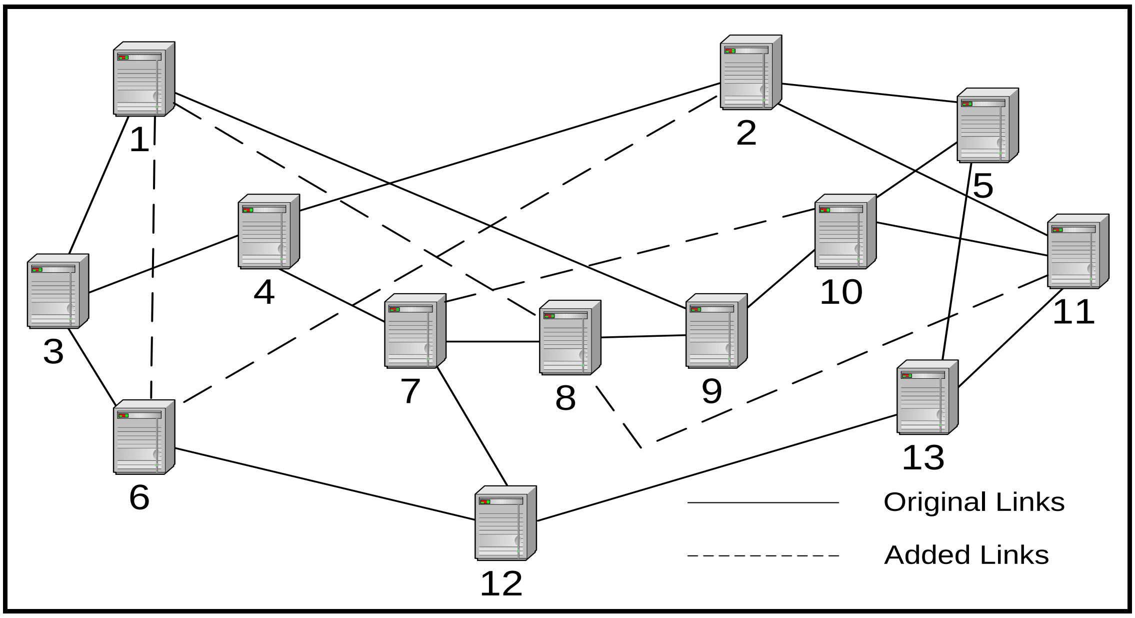

Figure (4) The Improved Undirected Network (NSFNET)

Failures were applied again to the most important nodes and links in the network. The results show that the designed topology was efficient and can still work in the face of failures. **Figure (5)** shows the relationship between the number of links failure (8-9, 5-13, 1-3, 2-4, 7-12, 6-12 respectively) versus the average distance and network centralization.

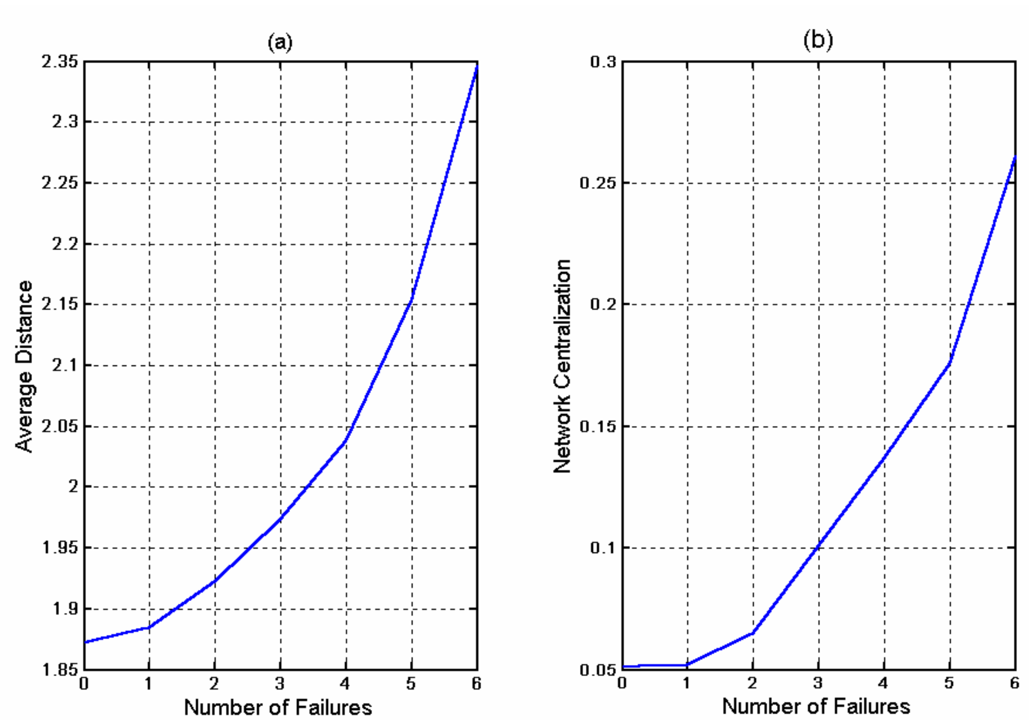

Figure (5) Relationship between Numbers of Links Failure versus (a) Average Distance (b) Network Betweenness Centralization

As shown in **Fig.(5)** above, influence of increasing links failure was appear to both average distance and network centralization but not that much as before the improvement, the improved network can still work even with a number of links failure because another choices with accessible distance are available now.

The SNA measures were applied to the cases and the results were obtained, similar result was obtained using UCINET software for windows. It's a program for the analysis of social networks. The data was inputted using NetDraw program. Also it can read UCINET system files, text files, and others [9].

3-2 Directed Network Design

The principle of SNA and steps mentioned before was applied also to directed network. **Figure (6)** shows the directed network which differ from the previous case in that the direction of information flow has been considered in the network topology design [10].

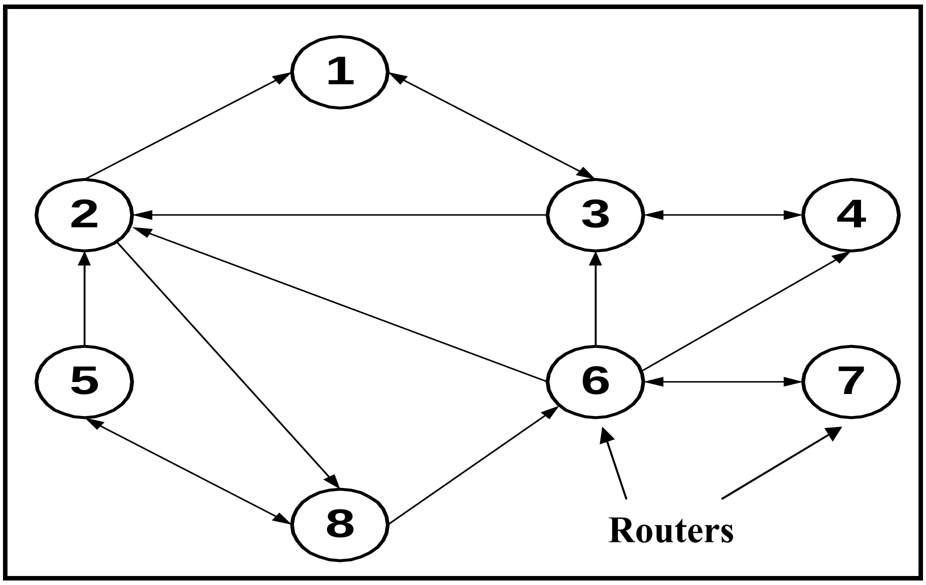

Figure (6) Routers in Directed Network

As shown in **Fig.(6)**. Removing link (2-8) which the most active line in the topology will isolate router (5, 6, 7, and 8) to be reached from all other nodes in the network, also it convert the topology into unconnected one like the bus topology. While removing router (2 and 6) will damage the network. The network becomes so critical when node/link failures happen, and it becomes unconnected (converted to a topology similar to bus) if failures happen in the centralized nodes in the network. The procedure of adding links to enhance the original network topology has been used to stand to node and link failures. The procedure began with the node which has the lowest degree (node 7). Many options were taken as shown in **Table (2)** and the best options were taken to build an efficient topology as shown in **Fig.(7)**.

Table (2) Strategy-2- Possible Network Improvement

Scenario	Network Centralization C_{Din}	C_{Dout}	C_{Cin}	C_{Cout}	C_B	Geodesics No.	Longest Path	Average Distance
Original net.Fig.(6)	0.163	0.326	0.419	0.435	0.280	63	5	2.268
1) add link: (7-1)	0.122	0.288	0.316	0.331	0.179	68	4	2.036
2) add link: Flow (4-7)	0.143	0.306	0.352	0.360	0.178	67	4	2.071
3) add link: (7-5)	0.143	0.306	0.410	0.412	0.282	71	5	2.232
4) add link:(7-1) (4-7)	0.102	0.265	0.268	0.301	0.174	70	4	1.964
5) add link: Option (4) + (4-5)	0.061	0.224	0.147	0.299	0.090	71	3	1.750
6) add link: Option (5) + (7-8)	0.183	0.183	0.214	0.214	0.068	73	3	1.643
7) add link: Option (6) + (1-5)	0.142	0.142	0.165	0.165	0.042	83	3	1.589
8) add link: Option (7) + Flow (2-6)	0.122	0.122	0.118	0.118	0.070	81	2	1.536

where, flow in **Table (2)** represents added link only with one direction flow. Option (7) provides good centralization (betweenness). It maximizes the closeness which leads to improve the average distance.

Adding option (8) will decrease the maximum number hops to two hops only but adding more links will be cost effective and also our procedure stop after five options. The geodesies in the network are increased but not that much value that overload the routing table. A network with many short paths connecting all nodes will be more efficient in passing data and reconfiguring after topology change. **Figure (7)** shows the improved network topology using option (7) that proposed to be evaluated in the implementation of routing algorithms. The options presented in **Table (2)** is not unique solution, links can be added to optimize the network topology to reach good reliability index. However, adding more links to the network can causes:

1. Overload the router table with more than enough shortest paths.
2. Delay during routing table updating when cost change or links failure.
3. Drawbacks because any damage for the network will be expensive.

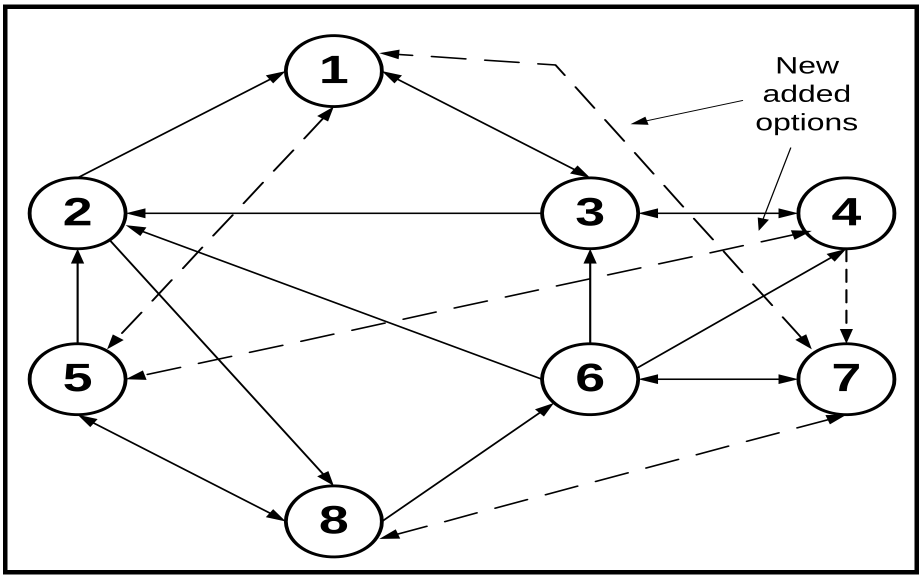

Figure (7) Improved Directed Network

4. Conclusions

The main conclusions obtained from this work are:

1. Social Network Analysis (SNA) was successfully applied to the networks to select topologies for directed and undirected link flow. The SNA measures and assessment of the basic topologies show that the optimum topology is preferable to be between ring and full-mesh topology which is the partial-mesh topology.
2. It is clear that the proposed strategy of increasing the number physical links was leads to better topology design, and good reliability. The network was tested with a number of node/link failures. The results show that the improved topology can stand and work even with these failures but a little change happened in average distance and network centralization value.
3. The results of optimization process were based on the network betweenness centralization than other (degree and closeness) centrality measures because this measure provides the location importance of node to other in the network (i.e. betweenness of node in connecting subnets). Its failure causes problems and damaging the network if it is highly betweenness centralized.
4. It is obvious that the increasing of links leads to maximize the closeness. Maximizing closeness between all routers improves both updating and minimizes hop counts. Furthermore, the results show that every router will have more choices (enough redundancy) in passing data and reconfiguring easily after a topology change. The latter reduces the overhead imposed for router during routing table updating.

Finally the followings can be suggested for future investigation: It is proved that the SNA used here improved the network topology from many sides. By using SNA, an algorithm can be designed using software programming to investigate the best position of the added links in the topology. This can be based on the network design requirements and the important node/link in the network. This provides best optimization for the topology and the options provided can be kept in spare if failures happened or one of the links is not available at any time. Also the successful construction of the improved network topology with a few added links can be used to optimize large networks such as the terrorist 11 September network which is highly architected and can be used to model new networks.

5. References

1. F., Halsall, ***"Data Communications, Computer Networks and Open Systems"***, Fourth Edition, Addison-Wesley Publishing Company, 1996.

2. A. S., Tanenbaum, ***"Computer Networks"***, Third Edition, Prentice-Hall Inc., 2003.

3. James E. Goldman, and Phillip T. Rawles, ***"Local Area Networks"***, Second Edition, John Wiley & Sons Inc., 2000.

4. Valids Krebs, ***"The Social Life of Routers"***, The Internet Protocol Journal, Vol. 3, No. 4, 2000, pp. 14-25.

5. Andrew Parker, ***"An Introduction to Social Network Analysis"***, IBM Institute for Knowledge-Based Organizations, 2001, pp. 4-23.

6. Borgatti S. P., Everett, and L. C., Freeman, ***UCINET, Version 6.12: A Software Package for SNA. Harvard Analytic Technologies***, 2002.

7. Robert A. Hanneman, ***"Introduction to Social Network Methods"***, Department of Society, University of California, 1998, pp. 60-68.

8. ***Measurement and Operations Analysis Team, "The National Science Foundation Network"***, (http://www.moat.nlanr.net/INFRA/NSFNET), 1995, pp. 1-6.

9. Borgatti S. P., Everett, and L. C. Freeman, ***UCINET, Version 6.12: A Software Package for SNA, Harvard Analytic Technologies***, 2002.

10. Dr. Rao, ***"WAN and Long Distance Connectivity"***, (http://www.math.uaa.alaska.edu/~afkjm/cs442/handouts/wans