INFSCI 1070/TELCOM 2000 Homework 3

1. Modern data analytics projects involve the analysis of very large data sets (hence the term “big data analytics”). Suppose you collected data at various geographically distant sites (A-G)
   1. Suppose further that you could perform your analytics at one site only. If the sites were interconnected as below, which site would that be?
      1. Create an adjacency matrix (in Gephi, if you wish)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G |
| A | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| B | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| C | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| D | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| E | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| F | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

* + 1. Choose one of the centrality measures (e.g., closeness, betweenness, degree, etc.) and make the case why that is the right one

Closeness centrality; through Gephi, we see that nodes C and E both have high closeness centrality or shorter average distance to other nodes. This centrality measure should be selected for its intended purpose of finding the direct access to other nodes, which is important for collecting data from opposite distant sites. Degree centrality would not be adequate as it focuses on the sheer number of contacts to the node, and betweenness centrality focuses on the fraction of shortest paths that go through that particular node.

* + 1. Choose the best node based on this.
  1. Assume that all nodes have “store and forward” capability.
     1. If each dataset was 1 terabyte (1012 bytes) how long would it take for all of the data to arrive at the processing site if each link was 1 Gigabit/second?

Transmission delay = file size/bit rate

File size: 1 terabyte = 1000000000000bytes/8 = 1.25x10^11 bits

Bit rate: 1 gigabit/sec = 10^9bits/sec

1.25x10^11 bits/10^9bits/sec = **125 sec**

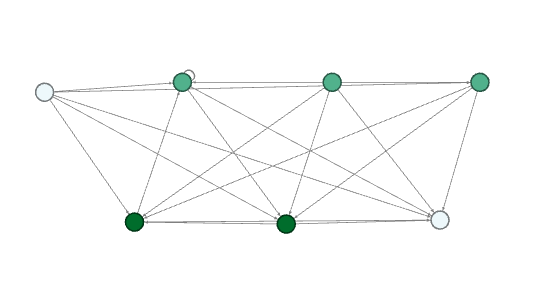
* + 1. How long it would take for 10 Gigabits/second?

**12.5 sec**

* + 1. What kinds of delay can you calculate based on the data you have?

Propagation delay cannot be calculated due to the lack of data on length, processing delay cannot be calculated due to lack of packet information. Total delay cannot be calculated to lack of information about other types of delay.

So only transmission delay can be calculated.

* 1. Suppose you decided to instead to send 1 TB flash cards via overnight mail (5pm drop off, delivered by 10am the next day) from each location directly to the destination?
     1. How does this change in the communication strategy change the connectivity graph (draw a picture)?
     2. What is the bit rate of each of these channels?

Bit rate = M/Tt

M: 1 TB = 10^12 bytes = 1.25x10^11 bits

T: 17 hours x 60 seconds = 1020 sec

1.25x10^11 bits/1020 sec = 122549019.6 bits/sec = **1.22x10^8 bits/sec**

1. (INFSCI 1070) Consider the following system:

Link A

Link B

The attributes of this system are as follows:

**Link A** Bit rate = 2.0 megabits/sec

Length = 240 km

Propagation velocity = 0.8 \* speed of light (3\*105 km/sec)

**Link B** Bit rate = 1.5 megabits/sec

Length = 135 km

Propagation velocity = 0.9 \* speed of light

Node 1 generates messages that are 1 kilobit long.

Node 2 is a relay that does not alter the packets, but only buffers them as needed for retransmission to Node 3.

**Next page for answers/work, two versions of part 2 due to confusion about the question (asked professor and came up with the first solution, worked with separate professor and derived second solution)**

1. What is the maximum packet transmission rate that Node 1 can have before the system begins to fail?

The maximum packet transmission rate is limited by the slower link, therefore the maximum packet transmission rate is **1.5 megabits/sec** from Link B.

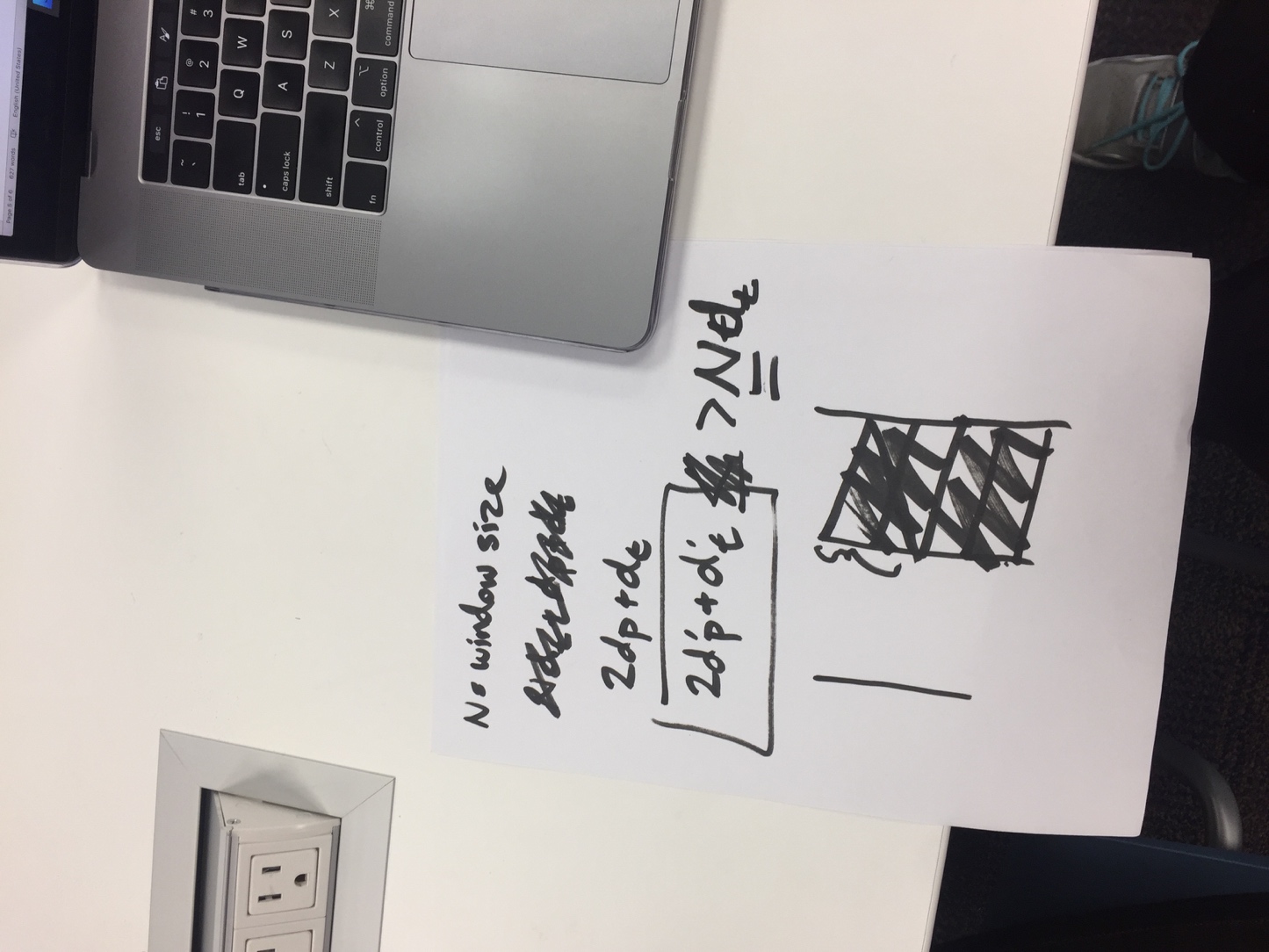
1. If Node 1 uses a sliding window flow control on Link A, what would the maximum window size be for the system not to fail?

Sliding window protocol allows sender to send multiple frames, receiver gives permission to transmit data until window is full. The restricting factor would be Link B.

2x(0.0005)+0.00067 = 0.00167

0.00167/0.00067 = 2.4925

Window size: **2.49**



Alternatively:

