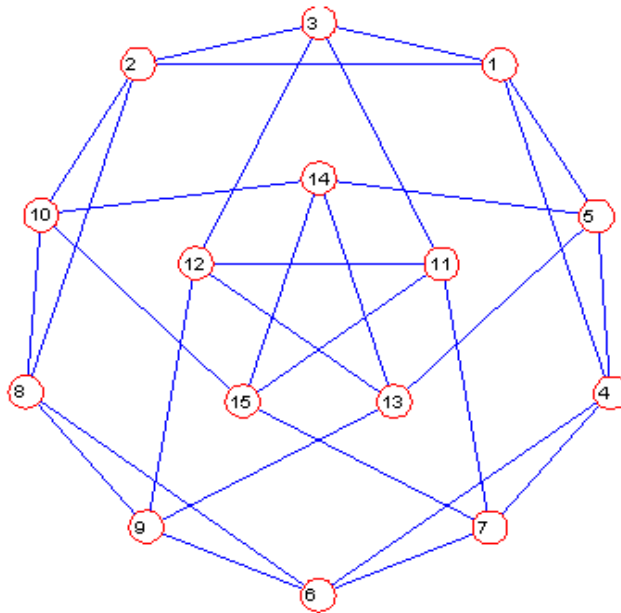


Practice Questions:

- 1) Name one thing that makes genome assembly difficult?
- 2) Calculate the sequencing coverage of a genome that is 2Mb and has been sequenced with 500,000 reads each with a length of 200bp?
- 3) Is the following Eulerian? If yes, provide a path. If no, state why it isn't.



- 4) What is genome assembly and why do we need it?
- 5) Given the following reads use the de Bruijn method ($n\text{-mer}=3$) to determine the genome sequence (show the graph and the final genome sequence).
TTGT
GCGT
ATGC
GTTT
- 6) What affect would decreasing the minimum ORF length cut-off have on sensitivity of a gene prediction program?
- 7) Write out the exact contents of my_file.txt after running the following program.

```
out=open("my_file.txt","w")
a="ATGCTGA\n"
x=a.find("TGA")
out.write(">id1234\n")
out.write(a)
out.write(x)
out.close()
```

- 8) Explain how GeneMark and Glimmer improve gene prediction beyond simply identifying ORFs.