MA 544: Programming Assignment - 2

In [ ]:
```
# Import required packages
```

Question 1: Visualization of Sparse Matrices

Create the sparse COO representation of the adjacency matrix of the following graph. One need not find the 2-dimensional representation for doing so. Find its CSR representation using appropriate Python function. Visualize this matrix by using spy() function. Refer to the notebook on sparse Matrices and check the SciPy documentation.
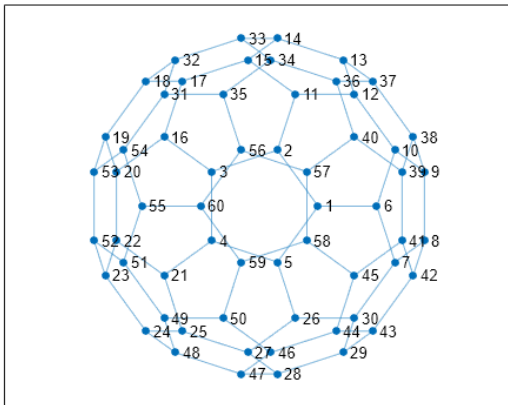

Image source: Mathworks

In [ ]:
```
# Your code starts here
```

Question 2: Page Ranking

Visit the SNAP networks database to find the Berkley-Stanford Webgraph from 2002. This dataset gives the hyperlink structure of 685230 webpages having a total of 7600595 hyperlinks.

A. Download the datafile and read the data to create a link matrix Q in the DoK format.

B. Use the optimized page-ranking algorithm with damping factor values $\alpha = 0.85$ and $\alpha = 0.95$ to find the page-ranking vector $\mathbf{r}$ while maintaining the same tolerance level for convergence. Report on the time it takes to perform these calculations.

C. Find the indices of the 10 highest ranking web-pages in decreasing order for the two runs.

D. Now use a personalization vector that gives equal non-zero probability to pages with prime indices otherwise zero. Find the indices of the 10 highest ranking web-pages. Use only $\alpha = 0.85$ in this case.

In [ ]:
```
# Your code starts here
```

Question 3: Clustering

Pick up a dataset of your choice for clustering (UCI dataset). Apply various clustering algorithms (including at least k-means and spectral clustering) to this dataset and explain your observations by using insightful plots or other some other metrics. Compare these algorithms based on some metric such as silhouette score, adjusted rand index etc.

- Read this blog for some help.
- You can also refer to the codes and discusion for this Kaggle dataset.

In [1]:
```
# Your code starts here
```