Part I

1. $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} \right\}$

   a. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = (1)(0) + (0)(1) + (0)(2) = 0$

   $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} = (1)(0) + (0)(-2) + (0)(1) = 0$

   $\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} = (0)(0) + (1)(-2) + (2)(1) = 2 - 2 = 0$

   b. Orthonormal basis: $\left\{ \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} \right\| \right\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \right\}$

   c. Decomposition

   $a_1 = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = -2$

   $a_2 = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \frac{7}{\sqrt{5}}$

   $a_3 = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}}$

   $\rightarrow -2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{7}{\sqrt{5}} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} + \frac{1}{\sqrt{5}} \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix}$

2. $A = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

   a. Eigenvectors: $\left\{ \begin{pmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{pmatrix}, \begin{pmatrix} 0.3913 \\ -0.2475 \\ -0.8863 \end{pmatrix}, \begin{pmatrix} 0.8247 \\ 0.5216 \\ 0.2185 \end{pmatrix} \right\}$
   Eigenvalues: $\{0, 0.8377, 7.1623\}$

   b. $0 \begin{pmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{pmatrix} - 2.7627 \begin{pmatrix} 0.3913 \\ -0.2475 \\ -0.8863 \end{pmatrix} + 2.5234 \begin{pmatrix} 0.8247 \\ 0.5216 \\ 0.2185 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

   c. $Ax = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + a_3 \lambda_3 v_3 = \begin{pmatrix} 14 \\ 10 \\ 6 \end{pmatrix}$

3. Centrality

a.

$$\begin{pmatrix} Protein & PI3K & RAS & PKC & PLC & PIP3 & PIP2 \\ PI3K & 0 & 1 & 0 & 0 & 1 & 1 \\ RAS & 1 & 0 & 1 & 0 & 0 & 0 \\ PKC & 0 & 1 & 0 & 1 & 0 & 1 \\ PLC & 0 & 0 & 1 & 0 & 1 & 1 \\ PIP3 & 1 & 0 & 0 & 1 & 0 & 1 \\ PIP2 & 1 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

b. Leading eigenvector is: $\begin{pmatrix} -0.3851 \\ -0.2473 \\ -0.3851 \\ -0.4294 \\ -0.4294 \\ -0.5230 \end{pmatrix}, \lambda = 3.1149$

c. Most central is PIP2, least central is RAS, and this matches up directly to the number of connections each protein has. PIP2 has the most (4) and RAS has the least (2). The protein with the largest number of direct connections is <u>not</u> always the most central node. Centrality is also determined by the centrality of the nodes it is connected to.

Part II

```
1. load('HW5_data.mat');
2. T1 = table2array(training_lines)';
   I_DC = [1 1 1 1 1 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
   -1 -1]';
   % 1 : IDC
   % -1 : DC
   M = fitcsvm(T1,I_DC)
```

```
M =

  ClassificationSVM
             ResponseName: 'Y'
    CategoricalPredictors: []
               ClassNames: [-1 1]
           ScoreTransform: 'none'
          NumObservations: 28
                    Alpha: [26×1 double]
                     Bias: -0.0392
         KernelParameters: [1×1 struct]
           BoxConstraints: [28×1 double]
          ConvergenceInfo: [1×1 struct]
           IsSupportVector: [28×1 logical]
                   Solver: 'SMO'
```

```
3. xval4 = crossval(M,'Kfold',4)
   kfoldLoss(xval4)
   xval28 = crossval(M,'Kfold',28)
   kfoldLoss(xval28)
```

```
    Properties, Methods
xval4 =
  ClassificationPartitionedModel
    CrossValidatedModel: 'SVM'
           PredictorNames: {1×8750 cell}
             ResponseName: 'Y'
          NumObservations: 28
                    KFold: 4
                Partition: [1×1 cvpartition]
               ClassNames: [-1 1]
           ScoreTransform: 'none'


    Properties, Methods
ans = 0.2500
xval28 =
  ClassificationPartitionedModel
    CrossValidatedModel: 'SVM'
           PredictorNames: {1×8750 cell}
             ResponseName: 'Y'
          NumObservations: 28
                    KFold: 28
                Partition: [1×1 cvpartition]
               ClassNames: [-1 1]
           ScoreTransform: 'none'


    Properties, Methods
ans = 0.2143
```

ans = 0.2500	ans = 0.2143
ans = 0.1786	ans = 0.2143
ans = 0.2143	ans = 0.2143
ans = 0.2143	ans = 0.2143
ans = 0.2500	ans = 0.2143

4. 4-fold repeats: Leave-one-out repeats:
    a. Accuracy changes depending on the method because when the number of groups, k, is smaller than the number of datapoints, there is stochasticity in the validation.
5. P = table2array(patient_samples)';
6. TestResults = predict(M,P)

Patients 1 and 2 are predicted to have regular DC, and Patient 3 is predicted to have IDC.

```
TestResults = 3×1
        -1
        -1
         1
```