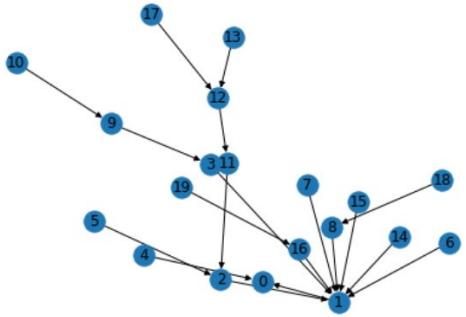
Identification of critical connectors in the directed reaction-centric graphs of microbial metabolic networks



Individual Analysis Project
Miguel Guardado
12/2/2020

Predicting Essential Reactions Using Network Centrality Metrics

Methods

Goal-Create and find a model based off network statistics to predict the essentiality of a node in a metabolic reactions-centric graph.

Data- 5 Network from the paper : Thanks Group 5 for Essential Reaction Data!

E. coli , B. subtilis, G. metallireducens, K. pneumoniae , and S.cerevisiae .

Aim1-Use Logistic Regression, SVM, and Random Forest to try and predict the essentiality of a node based on several centrality/graph metrics.

Aim2 - Perform PCA and determine if Dimension Reduction can give us a better idea into if essentiality can be predicted with our data.

Trying to Model Essential Reactions

```
8 x 1 sparse Matrix of class "dgCMatrix"
                                  1
(Intercept)           -1.4207506
k_total                .
BetweennessCentrality 12.7724577
ClusteringCoeff        .
Bridge                 .
BridgingCentrality     .
CascadeNum             0.2390955
EigenCentrality        .
```

```
Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)           -1.76598    0.05015 -35.214  < 2e-16 ***
BetweennessCentrality 28.85829    4.49544   6.419 1.37e-10 ***
CascadeNum             0.31835    0.02973  10.709  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

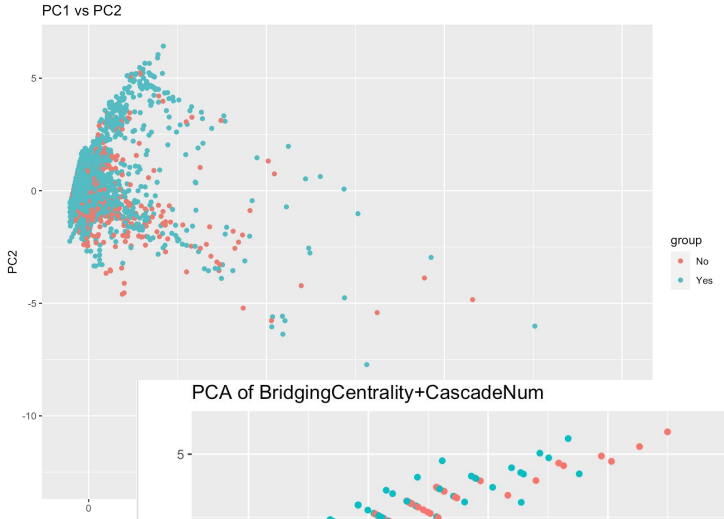
Model Type	Accuracy
LogR: BC+CN	62%
LogR: BC*CN	62%
SVM: BC+CN	57%
RF: BC+CN	64%

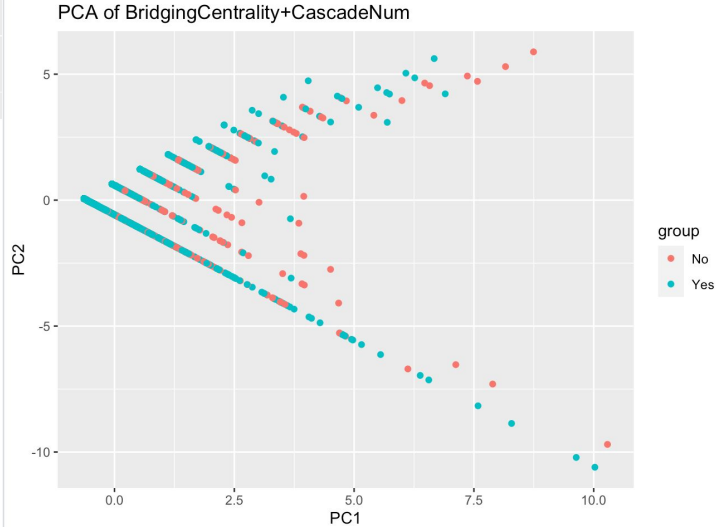
```
Call:
 randomForest(formula = EssentialReaction ~ BetweennessCentrality +      CascadeNum, data = train)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 1

          Mean of squared residuals: 0.1403682
                    % Var explained: 9.04
```

```
      RMSE       Rsquared    MAE
0.4205174  0.04116118  0.2149807
```

PCA plots of Essential Reactions+ Reflection





	BetweennessCentrality	CascadeNum	Essential
229	2.451889e-04	0	Yes
230	2.328977e-03	0	Yes
231	2.451889e-04	0	Yes
232	4.958115e-03	0	No
233	5.366426e-03	0	Yes
234	3.945265e-03	0	Yes
235	8.849893e-04	1	Yes
236	5.680689e-03	0	Yes
237	4.709084e-03	0	Yes
238	3.239921e-03	0	Yes
239	8.785842e-04	1	Yes
240	0.000000e+00	0	No
241	5.097410e-05	0	Yes
242	1.494007e-03	0	No
243	4.116649e-03	0	Yes
244	1.683827e-03	0	Yes
245	1.367693e-02	0	No

Questions

Graph that ended up being FAKE!!!! And giving me false hope