

Name: Aman Jiddewar.

Data Preparation:

I have corrected the diversification to handle the case of smaller number of investments which skew the diversification factor. This is similar to shrinkage estimator used in IMDB ratings.

$$\text{Corrected Diversification} = \frac{v}{(v + m)} * D + \frac{m}{(v + m)} * C$$

Where:

R = Diversification of firm A to B

m = Minimum number of investments needed to be considered as significant

v = Number of investments

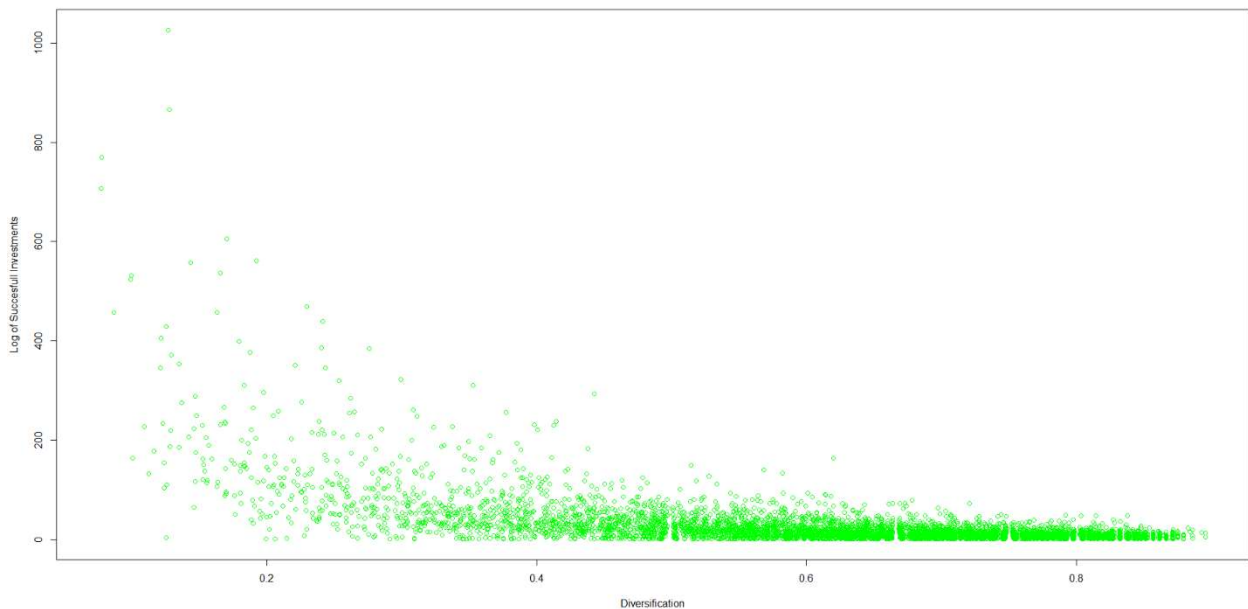
C = The mean diversification across the whole set

The value of m is set to 10 in HW.

Question 1.

Part A:

The diversification plot is on corrected diversification which is calculated using method mentioned in data correction part.



Part B:

```
call:
glm(formula = successful_investments ~ corrected_diversification,
     family = "poisson", data = investment_by_diversification)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-23.620  -1.155  -1.155  -0.141   32.938

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    6.564859   0.005424   1210  <2e-16 ***
corrected_diversification -6.856696   0.009536   -719  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 531833  on 25512  degrees of freedom
Residual deviance: 139767  on 25511  degrees of freedom
AIC: 210809

Number of Fisher Scoring iterations: 5
```

Question 2

```
# Question 2 -----
detailed_deals_q2 <- detailed_deals[!is.na(InvestorId), .(InvestorId, DealId, Lead_Investor)]
edge_list__investors_deals <- merge(detailed_deals_q2, detailed_deals_q2, by = "DealId",
                                   allow.cartesian = TRUE)[InvestorId.x != InvestorId.y,]
edge_list__investors_deals[, Lead_Investor.y := NULL]
setnames(edge_list__investors_deals, old = c("InvestorId.x", "InvestorId.y", "Lead_Investor.x"),
         new = c("Investor_from", "Investor_to", "lead_investor"))

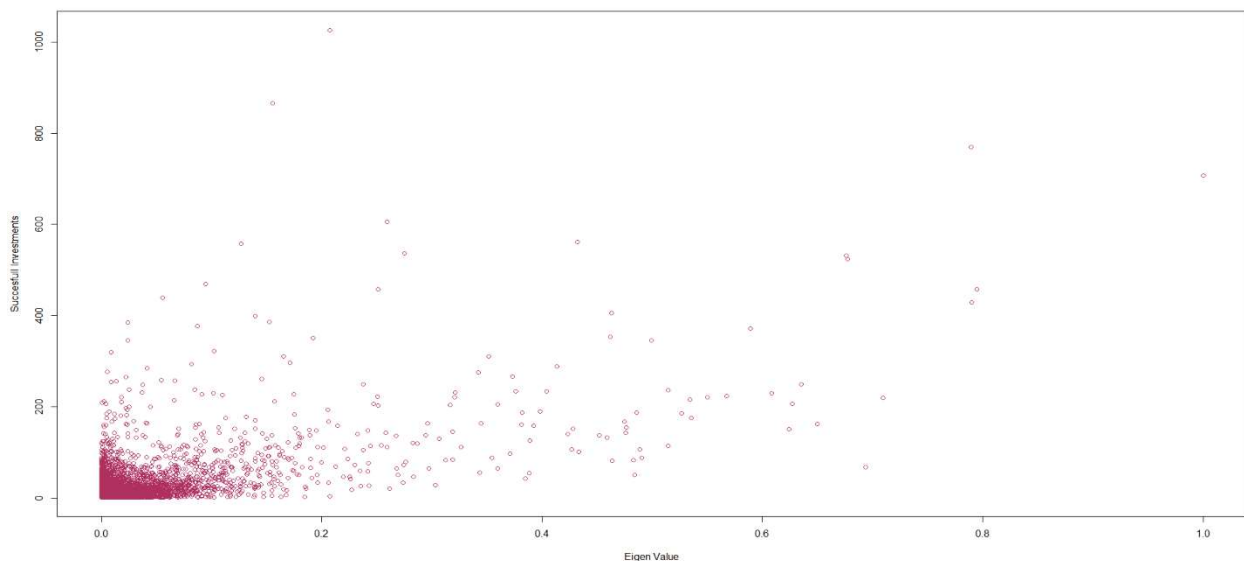
max(edge_list__investors_deals[, .N, by = .(Investor_from, Investor_to)][, N])

directed_edge_list_investors <- edge_list__investors_deals[, .(lead_sum = sum(lead_investor), total_investments = .N),
                  by = .(Investor_from, Investor_to)]

directed_edge_list_investors[, status := (lead_sum/(total_investments))]
investors_graph <- graph_from_data_frame(directed_edge_list_investors, directed = TRUE)
E(investors_graph)$weight <- directed_edge_list_investors[, status]

eigen_centrality <- eigen_centrality(investors_graph)
investors_eigen_centrality <- data.table(InvestorId = names(eigen_centrality(investors_graph)$vector),
                                       eigen = eigen_centrality(investors_graph)$vector)
```

Part A:



Part B:

```
call:
glm(formula = successful_investments ~ (eigen), family = "poisson",
    data = investors_eigen_centrality_sucess_plt)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|--------|--------|
| -62.542 | -3.217 | -2.625 | -0.724 | 72.659 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|---------|------------|
| (Intercept) | 2.114932 | 0.002574 | 821.7 | <2e-16 *** |
| eigen | 6.144372 | 0.011007 | 558.2 | <2e-16 *** |

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 459498 on 17618 degrees of freedom
Residual deviance: 341080 on 17617 degrees of freedom
AIC: 394400

Number of Fisher Scoring iterations: 6

Question 3

```
call:
glm(formula = successful_investments.x ~ eigen * corrected_diversification,
     family = "poisson", data = interaction_terms_model)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|--------|
| -21.487 | -1.419 | -1.340 | 0.125 | 35.341 |

Coefficients:

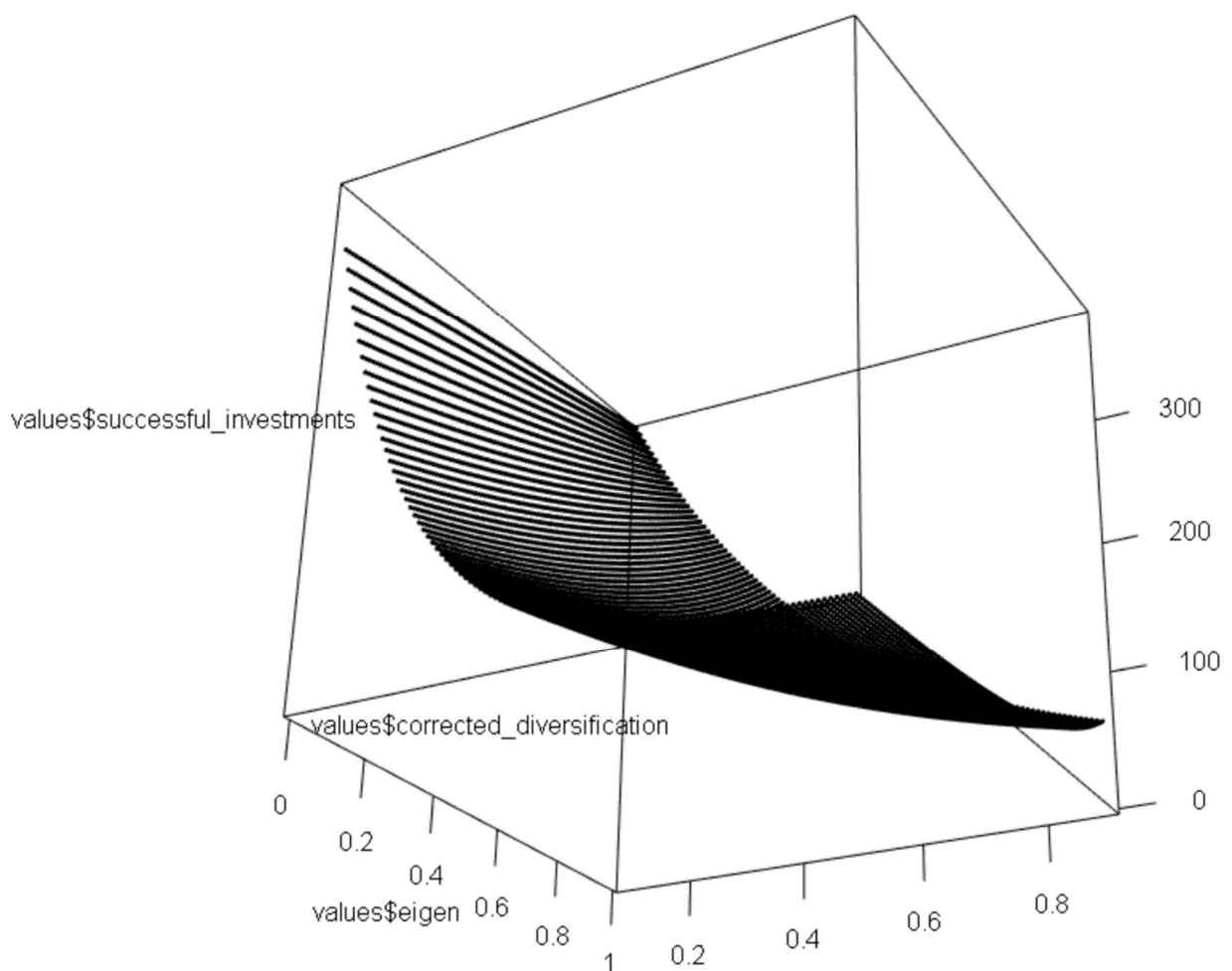
| | Estimate | Std. Error | z value | Pr(> z) |
|---------------------------------|-----------|------------|----------|--------------|
| (Intercept) | 6.319381 | 0.008077 | 782.356 | < 2e-16 *** |
| eigen | -0.187639 | 0.032686 | -5.741 | 9.43e-09 *** |
| corrected_diversification | -6.335254 | 0.012904 | -490.952 | < 2e-16 *** |
| eigen:corrected_diversification | 4.101229 | 0.182168 | 22.513 | < 2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 459498 on 17618 degrees of freedom
Residual deviance: 123994 on 17615 degrees of freedom
AIC: 177318

Number of Fisher Scoring iterations: 5



Question 4:

```
> summary(model)
Call:
multinom(formula = startup_state ~ corrected_diversification +
  eigen + corrected_diversification * eigen, data = merged_status)

Coefficients:
              (Intercept) corrected_diversification      eigen corrected_diversification:eigen
Generating revenue  2.53528429          -0.9283060 -0.7046891             0.5963856
Profitable         2.54107888          -1.8110883 -0.3472399            -18.3444560
Ramp-up            -0.02111366          -0.3683002 -0.8402780             6.9138002

Std. Errors:
              (Intercept) corrected_diversification      eigen corrected_diversification:eigen
Generating revenue  0.02458640          0.03965488 0.07048402             0.4254494
Profitable         0.02568060          0.04177186 0.07981826             0.4936366
Ramp-up            0.03371966          0.05486205 0.09666864             0.5602216

Residual Deviance: 697230.4
AIC: 697254.4
```

As shown below the values are below 0.05, except for status generating revenue the interaction term is not significant. Over all looking at the model above, diversification has drawback for the start up because it has negative correlation with the state of the startup. This sounds counter-intuitive, because more diversity means more exchange of different ideas.

```
> z <- summary(model)$coefficients/summary(model)$standard.errors
> (1 - pnorm(abs(z), 0, 1)) * 2
              (Intercept) corrected_diversification      eigen corrected_diversification:eigen
Generating revenue  0.0000000          0.000000e+00 0.000000e+00             0.1609816
Profitable         0.0000000          0.000000e+00 1.359008e-05             0.0000000
Ramp-up            0.5312148          1.903944e-11 0.000000e+00             0.0000000
```