Group 8
Jigyasa Bardalaye 240019637
Noor Karaman 1900014984
Teanna Puthucheary 240063221
Theofano Toufexi 240029546

Group 8
Jigyasa Bardalaye 240019637
Noor Karaman 1900014984
Teanna Puthucheary 240063221
Theofano Toufexi 240029546

1.0 Data Analysis Procedure

This analysis aims to improve the efficiency of Silicon's R&D projects for microprocessors by examining knowledge exchange among employees. We identify network characteristics that could optimise project duration, success, and innovation by treating the network as one-node, unweighted, and undirected.

We start by assessing network connectedness to ensure reliable analysis with NetworkX algorithms. We calculate node degrees and log-transform the degree values to understand the spread of knowledge-sharing links among employees; the skewed distribution is expected from a small-world network. The average degree indicates network density, providing insight into knowledge circulation and its potential impact on project outcomes. These network indicators reveal structural aspects that are key to enhancing Silicon's R&D efficiency. Figure 1.0 illustrates a skewed distribution of employee connections, as expected from a small world network, which can be best modelled using a Poisson distribution.

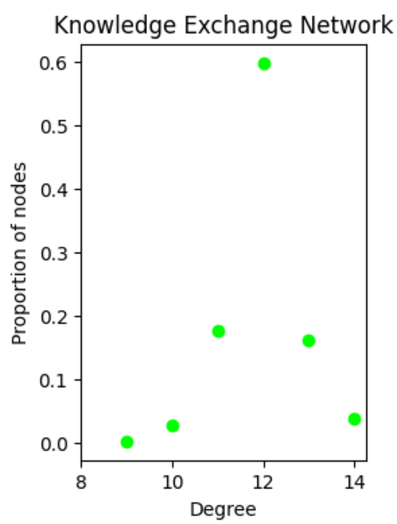

Figure 1.0

1. Identifying Network Analysis Indicators

We identified two key density metrics to assess team connectivity: the clustering coefficient and the average degree. The average degree identifies central employees who facilitate rapid knowledge exchange and serve as hubs within the network.

We use Burt's constraints and node betweenness centrality to measure bridging ties, highlighting employees who connect otherwise isolated clusters.

2. Assessing attributes of the DataFrames

To facilitate analysis at the team level, we converted these dictionaries into consolidated Pandas DataFrame.

3. Merging Network Metrics with Team Data

Before merging the data, we inspect the dataset for missing values and visualise the pre-existing employee affiliation dataset. The figure below illustrates that the employee affiliation network is unconnected; however, it has no isolated employees, which means it consists of multiple smaller components (e.g., teams).

Figure 1.1

Grouping the data from a team perspective allows us to calculate the average values of each analysis indicator, matching the data grouping from the project outcome dataset.

4. Aggregating Metrics at Team Level

Grouping the data from a team perspective allows us to calculate the average values of each analysis indicator, matching the data grouping from the project outcome dataset.

5. Merging Aggregate Data with Project Outcomes

With these DataFrames, we created a new CSV file called ‘outcomes_final’ to store the project outcomes with each team’s average network metrics values of the knowledge exchange variables.

6. Statistical Analysis

To visualise the relationship between the variables in the ‘outcomes_final’ dataset, we load the seaborn package to create univariate boxplots for dummy variable project success and scatterplots for project duration and project novelty. Creating a correlation heatmap (figure 1.2) summarised the correlation between each variable, a more effective method than running individual Pearson's correlation tests.

To carry out statistical analysis, we load the relevant packages before preparing to run an Ordinary Least Squares (OLS) linear regression model. We create a DataFrame including all variables to be used as predictors in the regression, remembering to add a constant (intercept) column. Then, we model different OLS linear regressions, testing each project outcome as the dependent variable.

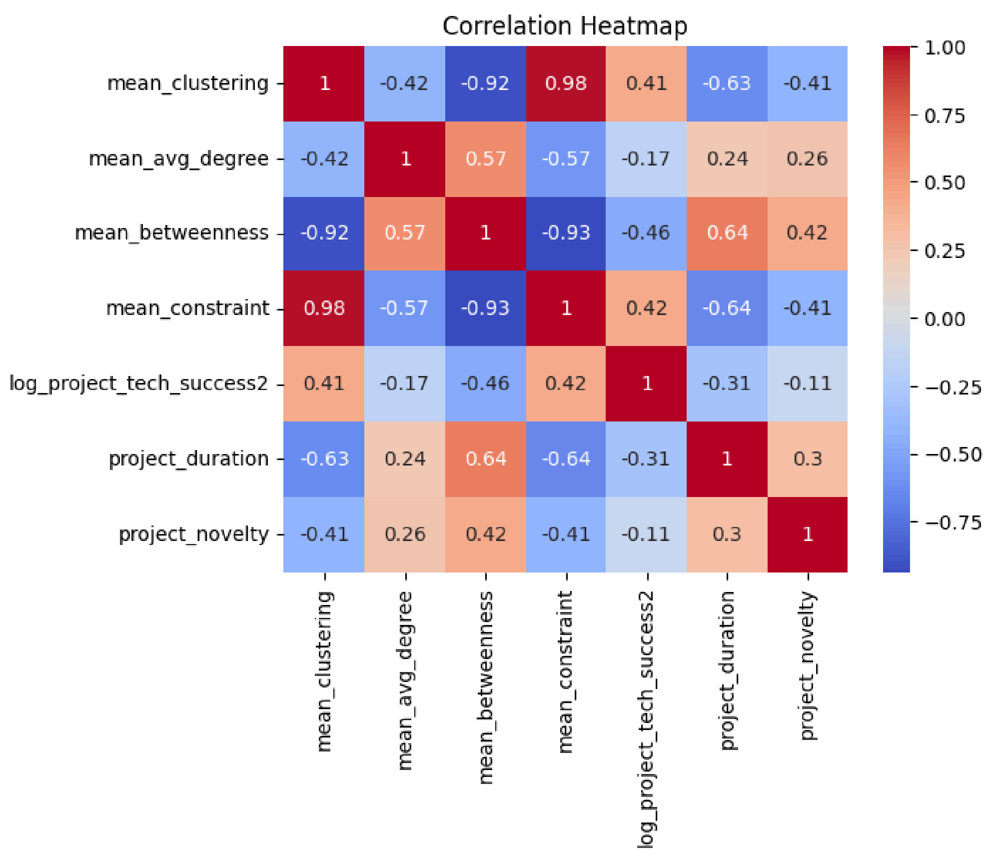

Figure 1.2

2.0 Results

By importing the new data set ‘outcomes_final.csv’ into R studio, we constructed multiple linear regression models on the variables to analyse the network structure’s influence on project efficiency empirically. With project_tech_success being a dummy variable, we choose to log its values using the equation:

```
(log(project_tech_success)+ 1)
```

Figure 2.0

It is important to note that the Akaike Information Criterion (StepAIC) may be a more accurate depiction of the complexity of real-world settings (Bozdogan, 2000), as it illustrates how variables work in conjunction. However, we focus on analysing the results based on significant p-values in the regression model, as we are trying to determine a causal relationship between network variables and project outcomes to formulate specific and actionable business recommendations to enhance efficiency.

```
                                OLS Regression Results
====================================================================================
Dep. Variable:     log_project_tech_success2   R-squared:                      0.254
Model:                                   OLS   Adj. R-squared:                 0.207
Method:                        Least Squares   F-statistic:                    5.339
Date:                       Sat, 09 Nov 2024   Prob (F-statistic):          8.86e-05
Time:                               12:48:54   Log-Likelihood:               -3.7145
No. Observations:                        101   AIC:                            21.43
Df Residuals:                             94   BIC:                            39.73
Df Model:                                  6
Covariance Type:                   nonrobust
===================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               -4.7727      3.651     -1.307      0.194     -12.022       2.477
mean_avg_degree      0.2883      0.142      2.024      0.046       0.005       0.571
mean_clustering     -7.9352      4.890     -1.623      0.108     -17.644       1.773
mean_betweenness   -82.3147     30.535     -2.696      0.008    -142.943     -21.686
mean_constraint     33.2377     22.473      1.479      0.142     -11.383      77.858
project_duration     0.0025      0.008      0.297      0.767      -0.014       0.019
project_novelty      0.0335      0.044      0.757      0.451      -0.054       0.121
==============================================================================
Omnibus:                       11.754   Durbin-Watson:                   1.830
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               12.704
Skew:                          -0.863   Prob(JB):                      0.00174
Kurtosis:                       3.196   Cond. No.                     1.04e+05
==============================================================================
```

Figure 2.1

Initially, there appears to be no significant relationship among the project outcome variables; therefore, each variable will be analysed individually.

Project Success

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     log_project_tech_success2   R-squared:                       0.249
Model:                                   OLS   Adj. R-squared:                  0.217
Method:                        Least Squares   F-statistic:                     7.943
Date:                       Sat, 09 Nov 2024   Prob (F-statistic):           1.42e-05
Time:                               12:48:54   Log-Likelihood:                -4.0864
No. Observations:                        101   AIC:                             18.17
Df Residuals:                             96   BIC:                             31.25
Df Model:                                  4
Covariance Type:                   nonrobust
==============================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              -4.4115      3.192     -1.382      0.170     -10.748       1.925
mean_avg_degree     0.2884      0.134      2.147      0.034       0.022       0.555
mean_clustering    -8.1522      4.708     -1.732      0.087     -17.497       1.193
mean_betweenness  -78.1450     29.324     -2.665      0.009    -136.352     -19.938
mean_constraint    33.5440     21.464      1.563      0.121      -9.062      76.150
==============================================================================
Omnibus:                       11.416   Durbin-Watson:                   1.881
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               12.311
Skew:                          -0.851   Prob(JB):                      0.00212
Kurtosis:                       3.163   Cond. No.                     1.38e+04
==============================================================================
```

Figure 2.2

The model's fit is indicated by an R-squared value, showing that the model explains approximately 24.9% of the variation in project success. This is supported by a low p-value for the F-statistic, confirming that at least one network variable is statistically significant.

By fitting a multiple linear regression model to understand the relationship between project success and different network metrics, it is clear that average degree (p = 0.034) and betweenness centrality (p =0.009) are significant metrics that impact the likelihood of teams completing successful projects. This suggests that teams that regularly share information, forming a tightly-knit team on average, are more likely to be part of a successful team, supporting the bonding social capital theory (Moore et al., 2018). The average degree metric, representing the average number of direct connections a team has, has a positive relationship with project success. For every unit increase of mean average degree between teams, the likelihood of project success increases by a factor of 1.33, calculated by taking the exponential of the beta coefficient as the dependent variable was log-transformed. These observations align with the patterns shown in the boxplots in Figure 5.0.

Holding a brokerage position with a high value of betweenness centrality is significant in affecting project success due to early access to information and novel ideas; the impact of increasing one's brokerage position had a marginal effect (1.28×10^{-34}). It can be argued that teams with a high mean average degree are likely to have stronger connections, representing a distal cause (Borgatti & Ofem, 2010), and could indirectly amplify the impact of team connections on project success.

Group 8
Jigyasa Bardalaye 240019637
Noor Karaman 1900014984
Teanna Puthucheary 240063221
Theofano Toufexi 240029546

Project Duration

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       project_duration   R-squared:                       0.475
Model:                            OLS   Adj. R-squared:                  0.453
Method:                 Least Squares   F-statistic:                     21.69
Date:                Sat, 09 Nov 2024   Prob (F-statistic):           9.06e-13
Time:                        12:48:54   Log-Likelihood:                -256.97
No. Observations:                 101   AIC:                             523.9
Df Residuals:                      96   BIC:                             537.0
Df Model:                           4
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const              202.7665     39.038      5.194      0.000     125.276     280.257
mean_avg_degree     -4.8304      1.642     -2.941      0.004      -8.090      -1.571
mean_clustering    113.6022     57.570      1.973      0.051      -0.673     227.878
mean_betweenness   894.4745    358.587      2.494      0.014     182.684    1606.265
mean_constraint   -646.2810    262.476     -2.462      0.016   -1167.291    -125.271
==============================================================================
Omnibus:                        2.246   Durbin-Watson:                   1.876
Prob(Omnibus):                  0.325   Jarque-Bera (JB):                1.725
Skew:                          -0.145   Prob(JB):                        0.422
Kurtosis:                       2.430   Cond. No.                     1.38e+04
==============================================================================
```

Figure 2.3

This regression analysis investigates the effect of various network metrics on project duration, which yields statistically significant results between variables' average degree, betweenness centrality, and Burt's constraints. An adjusted R-squared of 0.453 and an F-statistic p-value of 9.153e-13 indicate that a network's structure has a significant impact on project duration.

According to Figure 2.3, the average degree (p=0.004) has a significant impact on project duration. Increasing the average number of connections between employees in a team by one unit could decrease the project duration by 4.8 days. Referencing the correlation heatmap (Figure 1.2) and the mean constraint coefficient from Figure 2.3, it can be inferred that project duration has a large negative relationship with Burt's constraints. The negative beta coefficient implies a reduction in project duration as average constraints increase, as significant as 645 days for every 1 unit increase of average team constraints, supporting the bonding theory and reaffirming the understanding that bridging networks lack coordination (Long et al., 2013).

Interestingly, the other bridging ties metric, betweenness centrality, has the opposite effect on project duration at higher severity, explaining that employees who act as bridges have lower constraints limiting their cross-team collaboration; this is supported by the relationship observed in Figure 5.1. An increase in 0.1 unit of team average betweenness centrality increases project duration by 89.4 days. When a team holds a position of high betweenness centrality, it creates a dynamic where many other teams would become reliant on them to relay crucial information, and this can create a bottleneck, hence prolonging the time taken to complete projects.

Jigyasa Bardalaye 240019637
Noor Karaman 1900014984
Teanna Puthucheary 240063221
Theofano Toufexi 240029546

The strong negative correlation (-0.92 and -0.93) in Figure 1.2 involving betweenness centrality and clustering coefficients as well as betweenness centrality and Burt's constraint, implies that employees who facilitate communication between teams are not deeply embedded within a single close-knit team. The results of this analysis indicate that traditional network metrics–such as density and bridging ties- significantly impact the duration of Silicon's R&D project.

Project Novelty

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        project_novelty   R-squared:                       0.195
Model:                            OLS   Adj. R-squared:                  0.161
Method:                 Least Squares   F-statistic:                     5.811
Date:                Sat, 09 Nov 2024   Prob (F-statistic):           0.000313
Time:                        12:48:54   Log-Likelihood:                -89.306
No. Observations:                 101   AIC:                             188.6
Df Residuals:                      96   BIC:                             201.7
Df Model:                           4
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const               -4.3549      7.423     -0.587      0.559     -19.089      10.379
mean_avg_degree      0.3632      0.312      1.163      0.248      -0.257       0.983
mean_clustering    -14.9457     10.946     -1.365      0.175     -36.674       6.782
mean_betweenness    57.6278     68.180      0.845      0.400     -77.709     192.965
mean_constraint     57.3521     49.906      1.149      0.253     -41.710     156.415
==============================================================================
Omnibus:                        0.362   Durbin-Watson:                   1.696
Prob(Omnibus):                  0.835   Jarque-Bera (JB):                0.079
Skew:                          -0.021   Prob(JB):                        0.961
Kurtosis:                       3.130   Cond. No.                     1.38e+04
==============================================================================
```

Figure 2.4

According to this model, network structure does not have a significant impact on innovative solutions which contradicts research from Burt (2005, p. 92). Burt identified a nonlinear negative association between network constraints and the generation of innovative ideas. The modest negative correlation observed in Figure 1.2 (-0.41) aligns partially with this concept. However, given that we are applying linear regression to approximate a nonlinear relationship, some distortion in the results is expected.

We expect that teams with high values for both bridging ties metrics will significantly and positively impact project novelty. However, Burt's (2005, p. 147) argument that “innovative solutions are expected from teams with brokerage relations” specifies that employees must hold a brokerage position across different divisions. Information from the knowledge exchange network represents employees within the R&D department, hence limiting their applicability as information exchange is isolated within the department. This can explain the non-significant relationship between broker teams and the novelty of project outcomes.

3.0 Business Recommendations

These business recommendations are tailored to help Silicon increase its efficiency within its R&D department. According to Coelli et al. (2005), efficiency can be subcategorised into two groups: technical efficiency and allocative efficiency. In this context, we will be focusing on improving Silicon's allocative efficiency, involving a selection of inputs that produce outputs that add value to the organisation. As there was no significant relationship between project success and project novelty, we can concur that innovative solutions are essential for projects to be classified as successful. We chose not to focus on project novelty, as it doesn't have a direct impact on efficiency but instead could offer a competitive advantage of its product offerings.

Recommendation 1: Increase the average team degree by organising in-person networking opportunities.

As seen in Figure 2.2, the beta coefficient of average team degree suggests that for a unit increase of average team degree, the likelihood of a project being categorised as successful increases by a factor of 1.33. An increased average degree means team members have more connections within teams, which can lead to easier access to information. In a densely connected team, members may interact more frequently, generating diverse ideas and enhancing innovation.

Additionally, increasing the average team degree significantly reduces project duration by 4.82 days for every unit increase in average team degree. The average team degree increase is associated with higher project success rates and shorter project durations; therefore, this recommendation should be prioritised when evaluating the organisational structure. This can be achieved by organising in-person team activities. Pi and Cai (2017) emphasise that while virtual platforms enable quick information access and broader sharing, in-person activities enhance relational trust, a key component that supports the network closure argument.
Organising regular internal workshops, brainstorming sessions, or team-building activities. This would increase the average number of connections within teams, thereby improving the overall project success rate.

It is important to note that this analysis is based on an unweighted graph, which doesn't consider the strength or frequency of information exchange but merely identifies the presence of a relationship. Analysing the knowledge exchange network as a weighted network with assigned values to connections may provide a more nuanced view of the network, improving predictive analysis of Silicon's R&D projects.

Recommendation 2: Increase team constraints by organising an office seating plan
Our analysis shows that higher Burt constraints, reducing the team's brokerage position, are associated with reducing the average duration of a project by 64.5 days for every tenth unit increase. This could be vital for time-sensitive projects; research from West and Iansiti (2003) highlighted the reliance of R&D projects within semiconductor companies on the race to be early adopters of new technological advancements. This further illustrates the importance of reducing the duration of Silicon's projects.

To increase team constraints, Silicon could create a strategic seating plan, seating all team members together, which can encourage internal knowledge sharing while limiting excessive external collaboration. Based on the proximity principle (Uzzi & Dunlap, 2005), which expects employees in close proximity to form knowledge-sharing connections, enhancing collaboration by making interactions more frequent and seamless. By grouping individuals with similar expertise, the seating arrangement allows team members to easily exchange insights on specialised tasks and technical challenges, fostering strong, cohesive team connections.

Implementing this recommendation is expected to reduce project duration, increase the quarterly project capacity for the R&D department, and allow Silicon's management to allocate capital more effectively. Additionally, it develops a more specialised subgroup that increases productivity between well-defined tasks but may require careful management to avoid becoming information silos (Bento et al., 2020).

Recommendation 3: Decrease team betweenness centrality by decentralising the department structure

The beta coefficient in Figure 2.3 for mean betweenness centrality of 894 signifies the negative impact and an additional unit of team betweenness centrality can have on project duration in days. Prolonging a project unnecessarily can be costly, as capital is allocated inefficiently.

While employees with high betweenness centrality can facilitate important cross-team communication, their central role can also create delays, as many teams rely on them for information exchange. To mitigate this, Silicon should evaluate its reporting structures and shift from a traditional hierarchical structure to a flatter structure. This will allow for better delegation of control across the team, encourage more project ownership across teams, and reduce the reliance on a central figure, thus preventing bottlenecks that could delay projects. Phelps et al. (2012) found that excessive reliance on specific colleagues for knowledge can create bottlenecks and reduce knowledge-sharing efficiency. Therefore, fostering equal contribution and shared ownership throughout a project becomes crucial—achievable by minimising betweenness centrality. It is worth noting that betweenness centrality also has a significant relationship with impacting project success; however, since the impact is negligible, mitigating the impact of project duration outweighs the possible externalities on project success.

To maximise project success and efficiency, the R&D department should adopt a decentralised team structure that balances collaboration with autonomy, similar to agile startups. This aligns with Burt’s (2005) bonding social capital theory, emphasising coordination over innovation to enhance Silicon’s project efficiency. To assess the impact of bridging ties on project novelty, Silicon should evaluate the knowledge exchange network across departments. Implementing these recommendations can significantly boost efficiency, aligning with Silicon’s strategic goals.

4.0 Bibliography

Bento, F., Tagliabue, M. and Lorenzo, F., 2020. Organizational silos: A scoping review informed by a behavioral perspective on systems and networks. *Societies*, *10*(3), p.56

Borgatti, S.P. and Ofem, B., 2010. Social network theory and analysis. *Social network theory and educational change*,*17*, p.29.

Bozdogan, H., 2000. Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, *44*(1), pp.62-91.

Burt, R.S., 2005. *Brokerage and closure: An introduction to social capital*. Oxford University Press.

Coelli, T.J., Rao, D.S.P., O'donnell, C.J. and Battese, G.E., 2005. *An introduction to efficiency and productivity analysis*. Springer science & business media.

Cross, R., Borgatti, S.P. and Parker, A., 2002. Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, *44*(2), pp.25-46.

Long, J.C., Cunningham, F.C. and Braithwaite, J., 2013. Bridges, brokers and boundary spanners in collaborative networks: a systematic review. *BMC health services research*, *13*, pp.1-13.

Moore, C.B., Payne, G.T., Autry, C.W. and Griffis, S.E., 2018. Project complexity and bonding social capital in network organizations. *Group & Organization Management*, *43*(6), pp.936-970.

Phelps, C., Heidl, R. and Wadhwa, A., 2012. Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management*, *38*(4), pp.1115-1166.

Pi, S. and Cai, W., 2017. Individual knowledge sharing behavior in dynamic virtual communities: the perspectives of network effects and status competition. *Frontiers of Business Research in China*, *11*, pp.1-17.

Stuart, T.E., 1999. Technological prestige and the accumulation of alliance capital. In Corporate social capital and liability (pp. 376-389). Boston, MA: Springer US.

Uzzi, B. and Dunlap, S., 2005. How to build your network. *Harvard Business Review*, *83*(12), p.53.

West, J. and Iansiti, M., 2003. Experience, experimentation, and the accumulation of knowledge: the evolution of R&D in the semiconductor industry. *Research Policy*, *32*(5), pp.809-825.

5.0 Appendix

<u>Boxplots illustrating the relationship between Project Success and network variables</u>

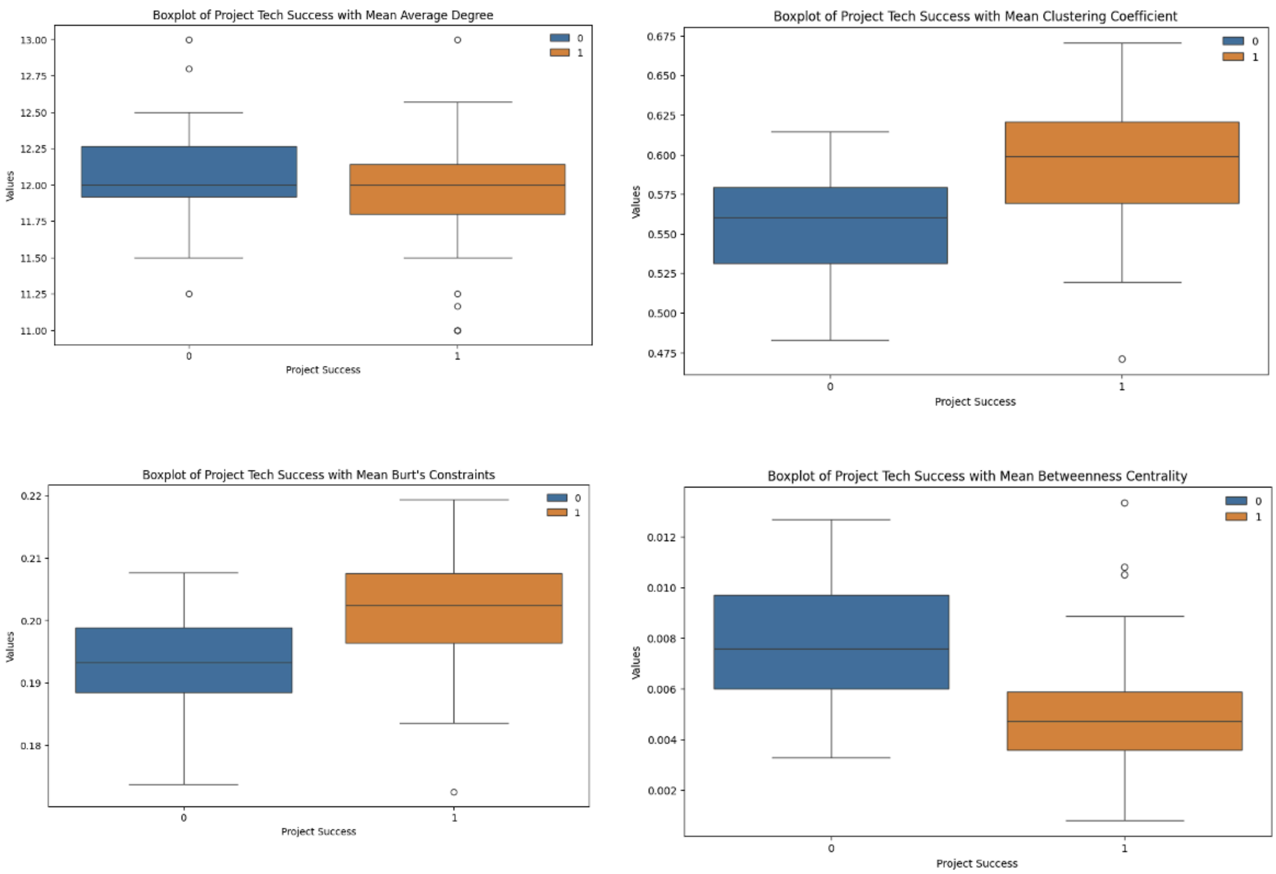

Figure 5.0

Using a univariate boxplot, we can evaluate the relationship between the dummy variable (project success) and network metrics. It is an effective comparative tool that offers a quick observation of the median, spread, and presence of outliers belonging to each category. This step is important before conducting statistical tests, as it offers a preliminary assessment of whether basic assumptions regarding normality and homogeneity of variance are met.

Figure 5.0 shows that more successful projects, on average, have a higher mean clustering coefficient and Burt's constraints than non-successful projects, which suggests that teams within a denser network are more likely to achieve project success. Contrastingly, a lower average of betweenness centrality for successful projects implies that employees who serve as a connector between otherwise isolated groups are less efficient in delivering on projects.

While this is an initial method of assessing the data, evaluating the p-values will identify whether these relationships are significant with respect to the dependent variable.

Scatterplot Illustrating the Relationship Between Project Duration and Network Variables

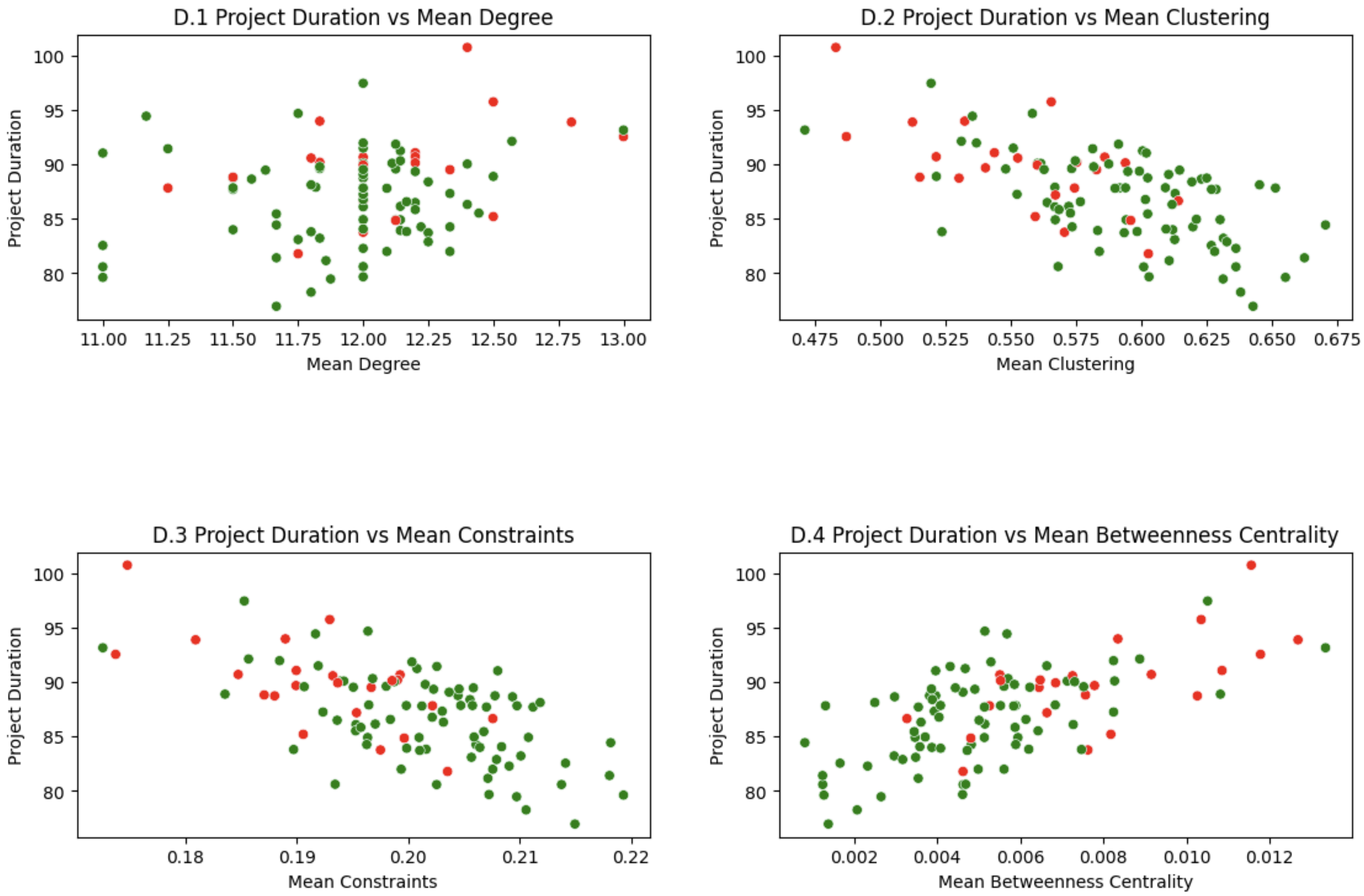

Figure 5.1

A scatterplot allows us to confirm the assumption of the data required for accurate linear regression modelling.

By visually inspecting the scatterplots, we observe a negative linear relationship between project duration and mean clustering (D.2) as well as project duration and mean constraints (D.3). This suggests that employees who selectively share knowledge while building stronger relationships with their direct connections are likely to complete projects more quickly. At first glance, it is challenging to determine the relationship between mean degree and project duration. Further analysis, using additional visualisations or statistical models, may provide a more transparent representation. The positive relationship between project duration and mean betweenness centrality (D.4), reflects the effect of betweenness centrality on project success as seen in Figure 5.0. This proposes that employees who act as brokers take longer to complete tasks.

Scatterplot Illustrating the Relationship Between Project Novelty and Network Variables

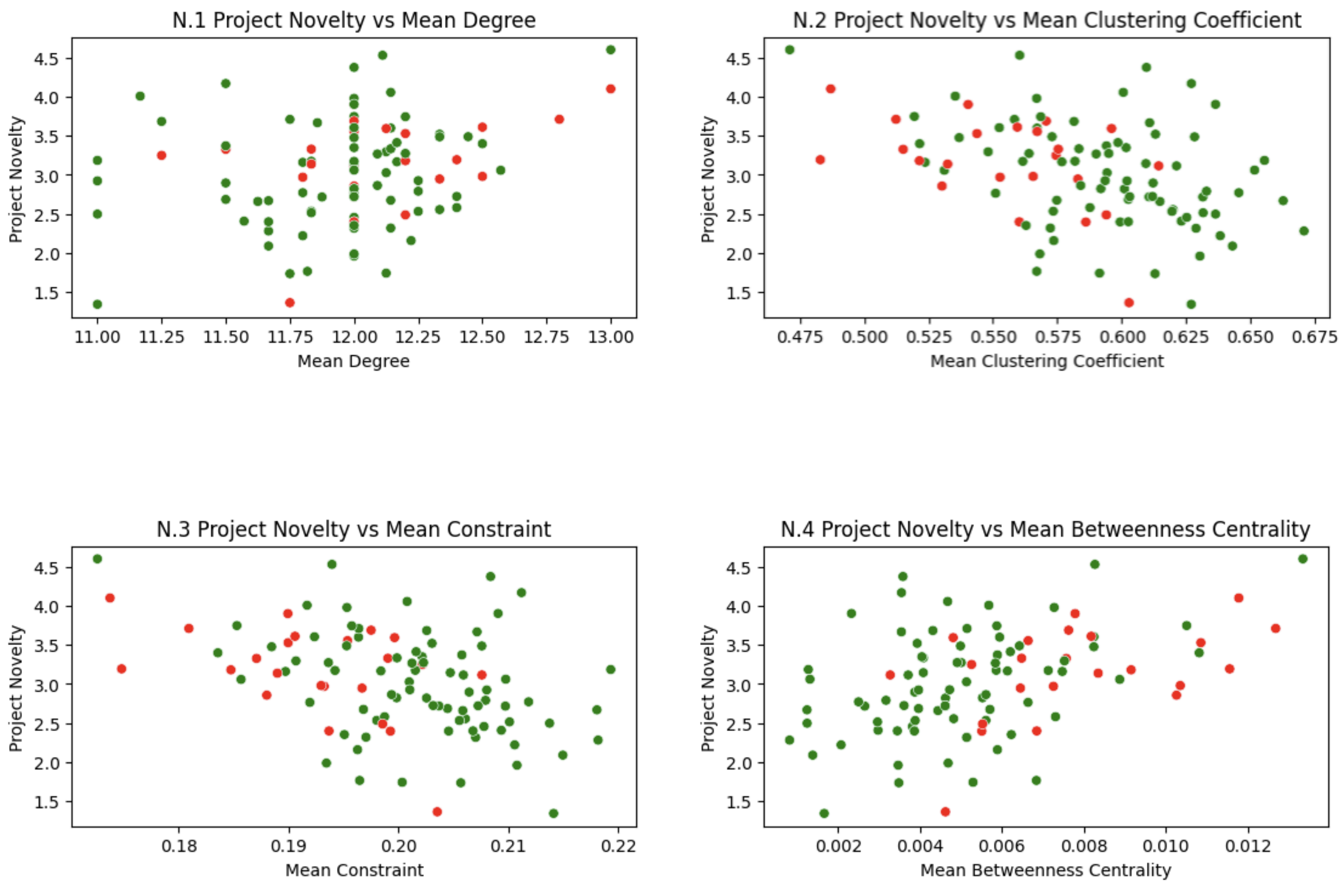

Figure 5.2

Figure 5.2 reveals relationships between network variables and project outcomes, similar to those with Project Duration. A positive relationship between project novelty and betweenness centrality (N.4) suggests that employees who connect across different teams are more likely to generate innovative solutions. Conversely, mean constraints (N.3) and the mean clustering coefficient (N.2) negatively impact novelty, indicating that while these factors may improve project efficiency, they may also limit innovation within the R&D department.

These data visualisations consistently demonstrate similar relationships across all three project outcome variables.