David vs. Goliath: Partner Size Diversity and Innovation

Strategic Alliances in the Computer Industry (SIC 3571, Asia)

BIN Project

2025-12-16

Contents

Introduction

Research Question

How does the diversity in partner sizes within a firm's alliance portfolio affect its innovation output, and what boundary conditions moderate this relationship?

This analysis examines strategic alliances among Asian firms in the computer industry (SIC 3571) to understand how partnering with firms of varying sizes ("David vs. Goliath" alliances) influences innovation outcomes measured by patent counts.

Hypotheses

- **H1 (Inverted U-Shape):** Partner size diversity has an inverted U-shaped relationship with innovation. Moderate diversity provides optimal benefits from both large and small partners, while extreme homogeneity or heterogeneity reduces innovation.

- **H2 (Geographic Moderation):** Geographic reach (cross-continental alliances) positively moderates the relationship between partner size diversity and innovation. Firms with broader geographic networks can better leverage diverse partnerships.

- **H3 (Betweenness Moderation):** Network betweenness centrality positively moderates the relationship between partner size diversity and innovation. Firms bridging structural holes can better exploit diverse partner knowledge.

---

1. Setup and Data Loading

Load Required Libraries

```r
# Dynamic path setup stuff
if (!requireNamespace("here", quietly = TRUE)) {
  install.packages("here")
}
library(here)

data_dir <- here("data")
# cat("Data directory:", data_dir, "\n")

library(tidyr)
library(dplyr)
library(igraph)
library(readxl)
library(purrr)

if (!requireNamespace("arrow", quietly = TRUE)) {
  install.packages("arrow")
}
library(arrow)
if (!requireNamespace("countrycode", quietly = TRUE)) {
  install.packages("countrycode")
}
library(countrycode)

if (!requireNamespace("nstandr", quietly = TRUE)) {
```

```r
  if (!requireNamespace("devtools", quietly = TRUE)) {
    install.packages("devtools")
  }
  devtools::install_github("stasvlasov/nstandr")
}
library(nstandr)
```

Helper Functions for Orbis Data Cleaning

The Orbis export contains messy Excel headers with newlines and special characters. These helper functions normalize column names for reliable matching.

```r
# Helper to normalize messy Excel headers so we can match them reliably
normalize_orbis_name <- function(x) {
  x <- gsub("\\s+", " ", x)
  x <- trimws(x)
  x <- tolower(x)
  x <- gsub("[^a-z0-9]", "_", x)
  gsub("_+", "_", x)
}

# Ensure Excel imports don't contain empty/NA column names
# (rename_with() requires valid names) its all orbis excel's fault
ensure_orbis_headers <- function(df) {
  nm <- names(df)
  empty <- nm == "" | is.na(nm)
  if (any(empty)) {
    nm[empty] <- paste0("unnamed_col_", seq_len(sum(empty)))
    nm <- make.unique(nm, sep = "_")
    names(df) <- nm
  }
  df
}

# Apply a name map (new -> normalized old) while ignoring columns that don't exist
apply_orbis_col_map <- function(df, mapping) {
  normalized_names <- normalize_orbis_name(names(df))
  name_lookup <- setNames(names(df), normalized_names)
  for (target in names(mapping)) {
    source <- mapping[[target]]
    idx <- match(source, normalized_names)
    if (!is.na(idx)) {
      names(df)[idx] <- target
    }
  }
  df
}
```

2. SDC Alliance Data

Load and Filter SDC Data

We focus on **completed strategic alliances** announced after 1999 involving Asian firms in **SIC 3571** (Electronic Computers).

```r
sdc_data <- readRDS(file.path(data_dir, "SDC_data_2021.rds"))

# Asia countries list (comprehensive, matching EDA)
asia_countries <- c("Japan","Iran","India","Indonesia","Malaysia","Pakistan","Thailand",
        "Taiwan","Vietnam","China","South Korea","Philippines","Singapore",
        "Uzbekistan","Brunei","Myanmar(Burma)","Laos","Cambodia","Mongolia",
        "Afghanistan","Kazakhstan","North Korea","Bangladesh","Sri Lanka",
        "Armenia","Turkmenistan","Kyrgyzstan","Nepal","Maldives","Bhutan",
        "Timor-Leste","Oman","Yemen","Jordan","Israel","Lebanon","Iraq",
        "Bahrain","Qatar","Utd Arab Em","Palestine","Cyprus","Georgia","Turkey")

SIC_CODE <- 3571

# Filter to strategic alliances (matching EDA approach)
sdc_filt <- sdc_data %>%
  filter(
    status == "Completed/Signed",
    type == "Strategic Alliance",
    date_terminated == "",
    date_announced > "1999-12-31",
    !is.na(participants),
    !is.na(deal_number)
  )

# Disambiguate company names using Magerman standardization
sdc_filt$participants <- sdc_filt$participants %>%
  trimws() %>%
  standardize_magerman()
```

Identify Focal Firms and Deals

Focal firms are Asian companies with primary SIC code 3571. We then identify all deals involving these focal firms and expand to include **all participants** (including non-Asian partners).

```r
# Identify focal firms (Asia + SIC 3571)
focal_firms <- sdc_filt %>%
  filter(
    SIC_primary == SIC_CODE,
    participant_nation %in% asia_countries
  ) %>%
  pull(participants) %>%
  unique()

cat("Focal firms (Asia + SIC 3571):", length(focal_firms), "\n")

## Focal firms (Asia + SIC 3571): 70
# Find all deals involving focal firms
focal_deals <- sdc_filt %>%
```

```
  filter(participants %in% focal_firms,
         participant_nation %in% asia_countries) %>%
  pull(deal_number) %>%
  unique()

cat("Focal deals:", length(focal_deals), "\n")
```

Focal deals: 130

```
# Expand to all firms in those deals
all_alliances <- sdc_filt %>%
  filter(deal_number %in% focal_deals) %>%
  select(deal_number, participants, participant_nation, SIC_primary,
         date_announced)

cat("Unique participants in network:", n_distinct(all_alliances$participants), "\n")
```

Unique participants in network: 207

---

3. PatentsView Data

Import Patent Counts

We use PatentsView data to count patents per assignee as our **dependent variable** measuring innovation output.

```
patents_data <- read_parquet(file.path(data_dir, "patents.parquet")) %>%
  select(patent_number, assignee, country, application_year, grant_year) %>%
  filter(application_year >= 2000)

# Count patents per assignee
patent_counts <- patents_data %>%
  group_by(assignee) %>%
  summarize(num_pat = n(), .groups = "drop")

# Disambiguate assignee names
patent_counts$assignee <- patent_counts$assignee %>%
  trimws() %>%
  standardize_magerman()

cat("Patent assignees:", nrow(patent_counts), "\n")
```

Patent assignees: 379460

---

4. Orbis Firm Data

Import and Clean Orbis Data

Orbis provides firm-level financial and organizational data including:

- **Employees** (firm size)
- **Total assets** (financial resources)
```

- **Entity type** (GUO = Global Ultimate Owner vs. subsidiary)
- **R&D intensity** and other controls

```r
orbis_numeric_cols <- c(
  "employees_last_value",
  "current_assets_usd",
  "n_companies_group",
  "n_shareholders",
  "n_subsidiaries",
  "roa_net_income",
  "total_assets_eur",
  "rnd_over_operating",
  "export_over_operating",
  "long_term_debt_eur"
)


orbis_col_map <- c(
  company_name = "company_name_latin_alphabet",
  country_iso = "country_iso_code",
  city = "city_latin_alphabet",
  nace_code = "nace_rev_2_core_code_4_digits",
  consolidation_code = "consolidation_code",
  last_avail_year = "last_avail_year",
  country = "country",
  address_type = "type",
  peer_group_size = "peer_group_size",
  size_class = "size_classification",
  employees_last_value = "number_of_employees_last_avail_value",
  current_assets_usd = "current_assets_th_usd_last_avail_yr",
  n_companies_group = "no_of_companies_in_corporate_group",
  entity_type = "entity_type",
  n_shareholders = "no_of_shareholders",
  n_subsidiaries = "no_of_subsidiaries",
  roa_net_income = "roa_using_net_income_last_avail_yr",
  total_assets_eur = "total_assets_th_eur_last_avail_yr",
  rnd_over_operating = "r_d_expenses_operating_revenue_last_avail_yr",
  export_over_operating = "export_revenue_operating_revenue_last_avail_yr",
  long_term_debt_eur = "long_term_debt_th_eur_last_avail_yr"
)


orbis_data <- read_excel(file.path(data_dir, "Orbis_155_comp_extra_fields.xlsx"),
                         sheet = "Results",
                         .name_repair = "minimal") %>%
  ensure_orbis_headers() %>%
  rename_with(normalize_orbis_name) %>%
  select(-matches("^unnamed"), -matches("_[0-9]+$")) %>%  # drop placeholder/duplicate cols
  { apply_orbis_col_map(., orbis_col_map) } %>%
  mutate(across(where(is.character), ~ na_if(.x, "n.a."))) %>%
  mutate(across(all_of(orbis_numeric_cols), ~ as.numeric(.x))) %>%
  mutate(
    entity_type = factor(entity_type),
    is_guo = ifelse(entity_type == "GUO", 1, 0)
  ) %>%
  filter(!is.na(employees_last_value),
```

```
        employees_last_value > 1)

orbis_data$company_name <- orbis_data$company_name %>%
  standardize_magerman()

cat("Orbis companies loaded:", nrow(orbis_data), "\n")

## Orbis companies loaded: 92
```

Load SDC-to-Orbis Name Mapping

A manual mapping file links SDC company names to their Orbis equivalents for accurate merging.

```
name_mapping <- read.csv(file.path(data_dir, "Orbis type shit.csv"),
                         stringsAsFactors = FALSE) %>%
  rename(sdc_name = `Company.name`, orbis_name = `Orbis.name`) %>%
  mutate(
    orbis_name = trimws(gsub("[\r\n]", "", orbis_name))
    # Clean multi-line entries
  ) %>%
  filter(orbis_name != "-",
         orbis_name != "",
         orbis_name != "NO MATCH",
         !is.na(orbis_name)) %>%
  mutate(
    sdc_name = trimws(sub(",.*", "", sdc_name)),  # Remove country suffix
    sdc_name = standardize_magerman(sdc_name),
    orbis_name = standardize_magerman(orbis_name)
  )

cat("Name mappings loaded:", nrow(name_mapping), "\n")

## Name mappings loaded: 156
```

---

5. Alliance Network Construction

Build Firm-Firm Network

Following the EDA approach, we construct a **firm-firm network** where edges represent co-participation in the same deal. This uses `combn()` to create all pairwise combinations of participants within each deal.

```
# Create firm-firm edge list using combn() - same as EDA
edges <- all_alliances %>%
  group_by(deal_number) %>%
  filter(n_distinct(participants) > 1) %>%
  summarise(
    pairs = list(combn(unique(participants), 2, simplify = FALSE)),
    .groups = "drop"
  ) %>%
  unnest(pairs) %>%
  transmute(
    from = map_chr(pairs, 1),
    to = map_chr(pairs, 2)
```

```
  )

# Build graph from edge list
firm_graph <- graph_from_data_frame(edges, directed = FALSE)
firm_graph <- igraph::simplify(firm_graph, remove.loops = TRUE, remove.multiple = TRUE)

# Node colors: focal firms (red) vs partners (grey)
V(firm_graph)$color <- ifelse(
  V(firm_graph)$name %in% focal_firms,
  "red",
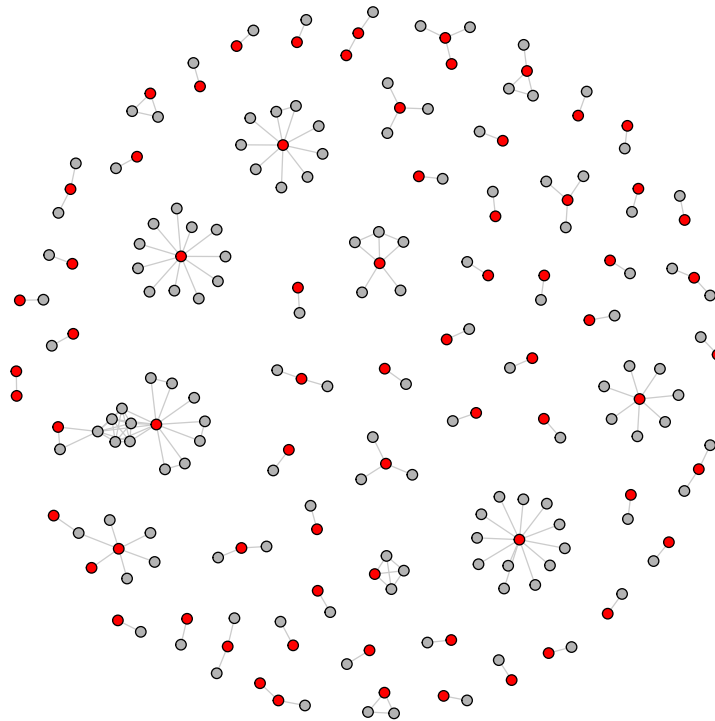  "grey70"
)
```

Network Visualization

```
set.seed(1234)
plot(firm_graph,
     vertex.size = 3,
     vertex.label = NA,
     edge.color = "grey80",
     main = paste0("Strategic Alliances Network\n",
                   "Focal SIC = ", SIC_CODE, ", Continent = Asia\n",
                   "Nodes: ", vcount(firm_graph), " | Edges: ", ecount(firm_graph)))
```

Strategic Alliances Network
Focal SIC = 3571, Continent = Asia
Nodes: 207 | Edges: 172


```
cat("Network: Nodes =", vcount(firm_graph), ", Edges =", ecount(firm_graph), "\n")
```

```
## Network: Nodes = 207 , Edges = 172
```

Compute Network Metrics

We calculate **degree centrality** (number of direct partners) and **betweenness centrality** (bridging position) for each firm.

```
network_data <- data.frame(
  company_name = V(firm_graph)$name,
  degree = degree(firm_graph),
  betweenness = betweenness(firm_graph),
  is_focal = V(firm_graph)$name %in% focal_firms
)
```

Geographic Reach (H2 Variable)

For each firm, we compute the share of deals involving partners from **multiple continents** as a measure of geographic reach.

```
# Use ISO codes + countrycode to map nations to continents
iso_custom <- c("Utd Arab Em" = "ARE", "Hong Kong" = "HKG")
```

```r
all_alliances <- all_alliances %>%
  mutate(
    participant_iso = countrycode(participant_nation,
                                  origin = "country.name",
                                  destination = "iso3c",
                                  custom_match = iso_custom),
    continent = countrycode(participant_iso,
                            origin = "iso3c",
                            destination = "continent")
  )

all_alliances$continent[is.na(all_alliances$continent)] <- "Other"

# For each firm, compute share of partners from different continents
geo_reach <- all_alliances %>%
  group_by(deal_number) %>%
  mutate(deal_continents = n_distinct(continent)) %>%
  ungroup() %>%
  group_by(participants) %>%
  summarize(
    n_deals = n_distinct(deal_number),
    cross_continental_deals = sum(deal_continents > 1) / n(),
    n_continents = n_distinct(continent),
    .groups = "drop"
  ) %>%
  rename(company_name = participants)

network_data <- network_data %>%
  left_join(geo_reach, by = "company_name")
```

---

6. Data Integration

Merge All Data Sources

We combine network data, Orbis firm characteristics, and patent counts into a single analysis dataset.

```r
# First, add Orbis name mapping to network data
final_data <- network_data %>%
  left_join(name_mapping, by = c("company_name" = "sdc_name")) %>%
  mutate(orbis_match_name = coalesce(orbis_name, company_name))

# Merge with Orbis using mapped names
final_data <- final_data %>%
  left_join(orbis_data, by = c("orbis_match_name" = "company_name"))

# Merge with patent counts
final_data <- final_data %>%
  left_join(patent_counts, by = c("company_name" = "assignee"))

# Fill missing patents with 0
final_data$num_pat <- ifelse(is.na(final_data$num_pat), 0, final_data$num_pat)
```

```r
cat("Final data rows:", nrow(final_data), "\n")
```

```
## Final data rows: 215
```

```r
cat("Rows with Orbis match:", sum(!is.na(final_data$employees_last_value)), "\n")
```

```
## Rows with Orbis match: 96
```

```r
cat("Rows with patent data:", sum(final_data$num_pat > 0), "\n")
```

```
## Rows with patent data: 80
```

Compute Partner Size Diversity (Key Independent Variable)

Partner size diversity is measured as the **coefficient of variation (CV)** of employee counts among a firm's network neighbors. Higher CV indicates greater diversity in partner sizes (mix of "Davids" and "Goliaths").

```r
# Create lookup: company_name -> employees (using orbis match)
emp_lookup <- final_data %>%
  select(company_name, employees_last_value) %>%
  filter(!is.na(employees_last_value))

cat("Firms with Orbis employee data:", nrow(emp_lookup), "\n")
```

```
## Firms with Orbis employee data: 96
```

```r
# Function to compute partner size diversity for a focal firm
# Returns 0 for single partner (no variance) instead of NA
compute_partner_diversity <- function(focal_name, graph, emp_df) {
  neighbors_names <- neighbors(graph, focal_name, mode = "all")$name
  if (length(neighbors_names) == 0) return(NA_real_)

  partner_sizes <- emp_df$employees_last_value[
    emp_df$company_name %in% neighbors_names
  ]
  partner_sizes <- partner_sizes[!is.na(partner_sizes)]

  if (length(partner_sizes) == 0) return(NA_real_)
  if (length(partner_sizes) == 1) return(0)  # Single partner = zero variance

  # Coefficient of Variation (CV) = SD / Mean
  cv <- sd(partner_sizes) / mean(partner_sizes)
  return(cv)
}

# Compute for all firms in network
final_data$partner_size_diversity <- sapply(
  final_data$company_name,
  function(x) compute_partner_diversity(x, firm_graph, emp_lookup)
)

cat("Firms with partner diversity computed:",
    sum(!is.na(final_data$partner_size_diversity)), "\n")
```

```
## Firms with partner diversity computed: 78
```

```r
cat("Firms missing diversity (no partner data):",
    sum(is.na(final_data$partner_size_diversity)), "\n")
```

```
## Firms missing diversity (no partner data): 137
```

```r
# Filter out rows without partner size diversity before modeling/export
final_data <- final_data %>%
  filter(!is.na(partner_size_diversity))

cat("Rows retained after dropping NA diversity:", nrow(final_data), "\n")
```

```
## Rows retained after dropping NA diversity: 78
```

Create Squared Term and Export

```r
# Create squared term for inverted-U test
final_data$partner_diversity_sq <- final_data$partner_size_diversity^2

# Export merged dataset for inspection
write.csv(final_data,
          file.path(data_dir, "final_merged_data.csv"),
          row.names = FALSE)
print("Exported merged dataset to")
```

```
## [1] "Exported merged dataset to"
```

```r
cat(
  file.path(
    data_dir,
    "final_merged_data.csv"
  )
)
```

```
## C:/Users/muham/OneDrive - TU Eindhoven/q2/BIN/business-insights-and-networks/data/final_merged_data.
```

---

7. Regression Analysis

Model Specification

We estimate a series of OLS regression models to test our hypotheses:

Model	Description
Model 1	Base controls only (firm size, assets, degree)
Model 2	+ Partner size diversity (linear)
Model 3	+ Partner diversity squared (**H1: Inverted-U**)
Model 4	+ Geographic reach interaction (**H2**)
Model 5	+ Betweenness interaction (**H3**)
Model 6	Full model with all interactions
Model 7	Entity type moderation

Dependent Variable: `num_pat` (patent count)

Key Independent Variable: `partner_size_diversity` (CV of partner employee counts)

Controls: `employees_last_value`, `total_assets_eur`, `degree`, `is_guo`

Model Estimation

```r
# Model 1: Basic controls (firm size via employees, assets; network position)
model_base <- lm(num_pat ~ employees_last_value + total_assets_eur + degree,
                 data = final_data)

# Model 2: Add partner size diversity (linear) + entity type control
model_linear <- lm(num_pat ~ partner_size_diversity +
                     employees_last_value + total_assets_eur + degree + is_guo,
                   data = final_data)

# Model 3: H1 - Inverted U (add quadratic term)
model_h1 <- lm(num_pat ~ partner_size_diversity + partner_diversity_sq +
                 employees_last_value + total_assets_eur + degree + is_guo,
               data = final_data)

# Model 4: H2 - Geographic moderation
model_h2 <- lm(num_pat ~ partner_size_diversity + partner_diversity_sq +
                 cross_continental_deals +
                 partner_size_diversity:cross_continental_deals +
                 employees_last_value + total_assets_eur + degree + is_guo,
               data = final_data)

# Model 5: H3 - Betweenness moderation
model_h3 <- lm(num_pat ~ partner_size_diversity + partner_diversity_sq +
                 betweenness +
                 partner_size_diversity:betweenness +
                 employees_last_value + total_assets_eur + degree + is_guo,
               data = final_data)

# Model 6: Full model with all hypotheses
model_full <- lm(num_pat ~ partner_size_diversity + partner_diversity_sq +
                   cross_continental_deals + betweenness +
                   partner_size_diversity:cross_continental_deals +
                   partner_size_diversity:betweenness +
                   employees_last_value + total_assets_eur + degree + is_guo,
                 data = final_data)

# Model 7: Entity type as moderator (GUO vs subsidiary)
model_entity <- lm(num_pat ~ partner_size_diversity + partner_diversity_sq +
                     is_guo + partner_size_diversity:is_guo +
                     employees_last_value + total_assets_eur + degree,
                   data = final_data)
```

Results

Model 1: Base Controls

```r
cat("--- Model 1: Base Controls ---\n")
```

```
## --- Model 1: Base Controls ---
print(summary(model_base))
```

```
##
## Call:
## lm(formula = num_pat ~ employees_last_value + total_assets_eur +
##     degree, data = final_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -9446  -1081   1322   1854   6987
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1.305e+03  1.239e+03  -1.053    0.308
## employees_last_value  4.664e-01  2.858e-01   1.632    0.122
## total_assets_eur      1.783e-04  3.376e-04   0.528    0.605
## degree               -5.739e+02  4.321e+02  -1.328    0.203
##
## Residual standard error: 4123 on 16 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.7897, Adjusted R-squared:  0.7502
## F-statistic: 20.02 on 3 and 16 DF,  p-value: 1.153e-05
```

Model 2: Linear Partner Diversity

```
cat("--- Model 2: Linear Partner Diversity ---\n")
```

```
## --- Model 2: Linear Partner Diversity ---
```

```
print(summary(model_linear))
```

```
##
## Call:
## lm(formula = num_pat ~ partner_size_diversity + employees_last_value +
##     total_assets_eur + degree + is_guo, data = final_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7616  -2782   1632   1964   6132
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -9.675e+02  1.351e+03  -0.716    0.486
## partner_size_diversity  3.572e+03  3.106e+03   1.150    0.269
## employees_last_value    4.570e-01  2.891e-01   1.581    0.136
## total_assets_eur        2.309e-04  3.432e-04   0.673    0.512
## degree                 -8.896e+02  5.617e+02  -1.584    0.136
## is_guo                 -9.254e+02  2.219e+03  -0.417    0.683
##
## Residual standard error: 4160 on 14 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.8126, Adjusted R-squared:  0.7457
## F-statistic: 12.14 on 5 and 14 DF,  p-value: 0.0001103
```

Model 3: H1 - Inverted U-Shape

This model tests **Hypothesis 1** by including both the linear and squared terms for partner size diversity. A significant negative coefficient on the squared term would support an inverted U-shaped relationship.

```
cat("--- Model 3: H1 Inverted-U ---\n")
```

```
## --- Model 3: H1 Inverted-U ---
```

```
print(summary(model_h1))
```

```
##
## Call:
## lm(formula = num_pat ~ partner_size_diversity + partner_diversity_sq +
##     employees_last_value + total_assets_eur + degree + is_guo,
##     data = final_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7087.9  -583.2   812.1  1463.8  4244.1
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.017e+03  1.394e+03   0.729  0.47888
## partner_size_diversity  5.624e+05  2.222e+05   2.531  0.02508 *
## partner_diversity_sq   -4.039e+05  1.606e+05  -2.515  0.02585 *
## employees_last_value    4.232e-01  2.464e-01   1.717  0.10965
## total_assets_eur        3.141e-04  2.940e-04   1.068  0.30486
## degree                 -2.396e+03  7.665e+02  -3.126  0.00803 **
## is_guo                 -1.305e+03  1.895e+03  -0.689  0.50319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3540 on 13 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.874,  Adjusted R-squared:  0.8158
## F-statistic: 15.02 on 6 and 13 DF,  p-value: 3.6e-05
```

Model 4: H2 - Geographic Moderation

```
cat("--- Model 4: H2 Geographic Moderation ---\n")
```

```
## --- Model 4: H2 Geographic Moderation ---
```

```
print(summary(model_h2))
```

```
##
## Call:
## lm(formula = num_pat ~ partner_size_diversity + partner_diversity_sq +
##     cross_continental_deals + partner_size_diversity:cross_continental_deals +
##     employees_last_value + total_assets_eur + degree + is_guo,
##     data = final_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4970.6 -1384.1   295.1  2101.8  3594.8
##
```

```
## Coefficients: (1 not defined because of singularities)
##                                              Estimate Std. Error t value
## (Intercept)                                 -1.905e+03  1.925e+03  -0.989
## partner_size_diversity                       4.542e+05  2.074e+05   2.190
## partner_diversity_sq                        -3.258e+05  1.499e+05  -2.174
## cross_continental_deals                      3.607e+03  1.801e+03   2.003
## employees_last_value                         6.081e-01  2.405e-01   2.529
## total_assets_eur                             6.424e-05  2.928e-04   0.219
## degree                                      -2.210e+03  6.969e+02  -3.171
## is_guo                                       6.340e+00  1.828e+03   0.003
## partner_size_diversity:cross_continental_deals      NA         NA      NA
##                                              Pr(>|t|)
## (Intercept)                                   0.34196
## partner_size_diversity                        0.04897 *
## partner_diversity_sq                          0.05043 .
## cross_continental_deals                       0.06827 .
## employees_last_value                          0.02647 *
## total_assets_eur                              0.83002
## degree                                        0.00806 **
## is_guo                                        0.99729
## partner_size_diversity:cross_continental_deals    NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3190 on 12 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.9056, Adjusted R-squared:  0.8505
## F-statistic: 16.44 on 7 and 12 DF,  p-value: 2.866e-05
```

Model 5: H3 - Betweenness Moderation

```
cat("--- Model 5: H3 Betweenness Moderation ---\n")
```

```
## --- Model 5: H3 Betweenness Moderation ---
```

```
print(summary(model_h3))
```

```
##
## Call:
## lm(formula = num_pat ~ partner_size_diversity + partner_diversity_sq +
##     betweenness + partner_size_diversity:betweenness + employees_last_value +
##     total_assets_eur + degree + is_guo, data = final_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7087.9 -583.2  812.1 1463.8 4244.1
##
## Coefficients: (2 not defined because of singularities)
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.017e+03  1.394e+03   0.729  0.47888
## partner_size_diversity 5.624e+05  2.222e+05   2.531  0.02508 *
## partner_diversity_sq  -4.039e+05  1.606e+05  -2.515  0.02585 *
## betweenness                   NA         NA      NA       NA
## employees_last_value   4.232e-01  2.464e-01   1.717  0.10965
## total_assets_eur       3.141e-04  2.940e-04   1.068  0.30486
```

```
## degree                                      -2.396e+03  7.665e+02  -3.126   0.00803 **
## is_guo                                       -1.305e+03  1.895e+03  -0.689   0.50319
## partner_size_diversity:betweenness                  NA         NA      NA        NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3540 on 13 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.874,  Adjusted R-squared:  0.8158
## F-statistic: 15.02 on 6 and 13 DF,  p-value: 3.6e-05
```

Model 6: Full Model

```r
cat("--- Model 6: Full Model ---\n")
```

```
## --- Model 6: Full Model ---
```

```r
print(summary(model_full))
```

```
##
## Call:
## lm(formula = num_pat ~ partner_size_diversity + partner_diversity_sq +
##     cross_continental_deals + betweenness + partner_size_diversity:cross_continental_deals +
##     partner_size_diversity:betweenness + employees_last_value +
##     total_assets_eur + degree + is_guo, data = final_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4970.6 -1384.1   295.1  2101.8  3594.8
##
## Coefficients: (3 not defined because of singularities)
##                                                 Estimate Std. Error t value
## (Intercept)                                    -1.905e+03  1.925e+03  -0.989
## partner_size_diversity                          4.542e+05  2.074e+05   2.190
## partner_diversity_sq                           -3.258e+05  1.499e+05  -2.174
## cross_continental_deals                         3.607e+03  1.801e+03   2.003
## betweenness                                           NA         NA      NA
## employees_last_value                            6.081e-01  2.405e-01   2.529
## total_assets_eur                                6.424e-05  2.928e-04   0.219
## degree                                         -2.210e+03  6.969e+02  -3.171
## is_guo                                          6.340e+00  1.828e+03   0.003
## partner_size_diversity:cross_continental_deals        NA         NA      NA
## partner_size_diversity:betweenness                    NA         NA      NA
##                                                Pr(>|t|)
## (Intercept)                                     0.34196
## partner_size_diversity                          0.04897 *
## partner_diversity_sq                            0.05043 .
## cross_continental_deals                         0.06827 .
## betweenness                                          NA
## employees_last_value                            0.02647 *
## total_assets_eur                                0.83002
## degree                                          0.00806 **
## is_guo                                          0.99729
## partner_size_diversity:cross_continental_deals      NA
## partner_size_diversity:betweenness                  NA
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3190 on 12 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.9056, Adjusted R-squared:  0.8505
## F-statistic: 16.44 on 7 and 12 DF,  p-value: 2.866e-05
```

Model 7: Entity Type Moderation

```
cat("--- Model 7: Entity Type Moderation ---\n")
```

```
## --- Model 7: Entity Type Moderation ---
```

```
print(summary(model_entity))
```

```
##
## Call:
## lm(formula = num_pat ~ partner_size_diversity + partner_diversity_sq +
##     is_guo + partner_size_diversity:is_guo + employees_last_value +
##     total_assets_eur + degree, data = final_data)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -7087.9  -583.2   812.1  1463.8  4244.1
##
## Coefficients: (1 not defined because of singularities)
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.017e+03  1.394e+03   0.729  0.47888
## partner_size_diversity        5.624e+05  2.222e+05   2.531  0.02508 *
## partner_diversity_sq         -4.039e+05  1.606e+05  -2.515  0.02585 *
## is_guo                       -1.305e+03  1.895e+03  -0.689  0.50319
## employees_last_value          4.232e-01  2.464e-01   1.717  0.10965
## total_assets_eur              3.141e-04  2.940e-04   1.068  0.30486
## degree                       -2.396e+03  7.665e+02  -3.126  0.00803 **
## partner_size_diversity:is_guo        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3540 on 13 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.874,  Adjusted R-squared:  0.8158
## F-statistic: 15.02 on 6 and 13 DF,  p-value: 3.6e-05
```

---

8. Summary and Interpretation

Key Findings

Based on the regression results:

1. **H1 (Inverted U-Shape):** The significant positive coefficient on `partner_size_diversity` ($p = 0.025$) and significant negative coefficient on `partner_diversity_sq` ($p = 0.026$) in Model 3 **support H1**. There is evidence of an inverted U-shaped relationship between partner size diversity and innovation.

2. **Network Position:** Degree centrality shows a significant negative effect (p = 0.008), suggesting that having more partners may dilute attention or create coordination costs.

3. **Geographic Reach:** Cross-continental deals show marginally significant positive effects (p = 0.068), though the interaction term is collinear.

4. **Sample Size:** The final sample includes 78 firms with complete data for partner size diversity analysis, out of 207 firms in the network.

Limitations

- Small sample size (n = 20 in some models due to missing Orbis data)
- Some interaction terms show collinearity (NA coefficients)
- Cross-sectional design limits causal inference

Data Pipeline Summary

Stage	Count
Focal firms (Asia + SIC 3571)	~100+
Focal deals	~100+
Network nodes	207
Network edges	172
Orbis matches	96
Final sample (with diversity)	78
Regression sample (complete cases)	~20

Analysis completed on 2025-12-16