Social Network Analysis - Assignment 1

Group 7

Floris Vermeulen   Martijn van Iterson   Niek Fleerakkers   Patryk Grodek
Samir Sabitli

2025-09-01

Table of contents

1 Introduction

Hypothesis	ERGM Term	Motivation
H1: Students who I consider friends also consider me their friend	`mutual`	The `mutual` term captures the tendency for outgoing ties to co-occur with incoming ties between actors, capturing the dynamic of reciprocation.

2 Methodology

2.1 Dataset

The study utilises publicly-available data from Bondera

Figure 1: Plot of the Bondera Network with and without Community Detection
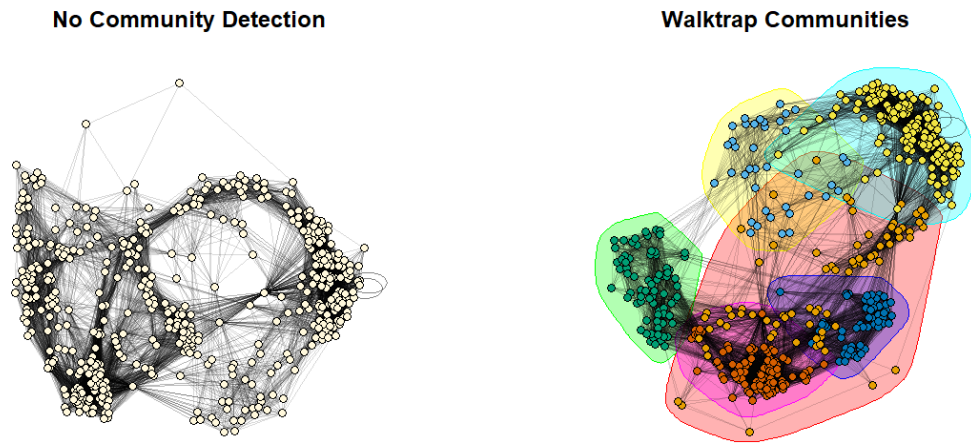
References

A   Source Code - Data Processing

```r
# —————————————————————————————————————————————————————————— #
# Reset WD
#setwd("../..")

# Import the Bondora P2P Dataset
bondora_raw <- read.csv("dataset/LoanData_Bondora.csv",
                        header = TRUE)
cols <- colnames(bondora_raw)

# —————————————————————————————————————————————————————————— #

# Subset only Columns we need
keep_cols <- c("LoanId", "UserName","NewCreditCustomer","LanguageCode",
               "Age", "Gender", "Country", "Amount", "Interest",
               "LoanDuration","UseOfLoan", "Education", "MaritalStatus",
               "NrOfDependants", "Rating", "Restructured",
               "NoOfPreviousLoansBeforeLoan", "MonthlyPayment")
bondora <- bondora_raw[keep_cols]

# Remove Rows with NAs -> Complete Dataset Preferred
bondora_complete <- na.omit(bondora)
sum(is.na(bondora_complete))

# Observe Class of Each Attribute
sapply(bondora_complete, class)

# Make Binary Indicators Binary
new_customer_mapping <- c("True" = 1, "False" = 0)
bondora_complete$NewCreditCustomer <- new_customer_mapping[
  bondora_complete$NewCreditCustomer]

# Replace User inputs of Blank Dependants with Zero
bondora_complete$NrOfDependants[bondora_complete$NrOfDependants == ""] <- NA
bondora_complete$NrOfDependants[is.na(bondora_complete$NrOfDependants)] <- 0

# Make the Column Numeric
bondora_complete$NrOfDependants <- as.numeric(bondora_complete$NrOfDependants)
bondora_complete$NrOfDependants[is.na(bondora_complete$NrOfDependants)] <- 0

# Make Restructured Binary
bondora_complete$Restructured <- new_customer_mapping[
  bondora_complete$Restructured]

# Randomly Remove Observations until Desired Size is Reached
set.seed(42)
sample_size <- 500
sample_indices <- sample(1:nrow(bondora_complete), sample_size)
bondora_sample <- bondora_complete[sample_indices, ]

# —————————————————————————————————————————————————————————— #

# Choose Feature Subset for Similarity Metric
similarities <- c("LoanDuration","Amount","MonthlyPayment", "NewCreditCustomer",
                  "NoOfPreviousLoansBeforeLoan","LanguageCode")
bondora_similar <- bondora_sample[similarities]

# Standardise Numeric Features in the Similarity Set
bondora_similar_scaled <- scale(bondora_similar)

# Compute Cosine Similarity
cosine_sim <- function(X) {
  # numerator: dot product
  sim <- X %*% t(X)
```

```r
65
66    # denominator: product of norms
67    norms <- sqrt(rowSums(X^2))
68    sim <- sim / (norms %*% t(norms))
69
70    return(sim)
71  }
72
73  similarity_matrix <- cosine_sim(bondora_similar_scaled)
74
75  # Get the Usernames for the Random Lenders
76  vertex_names <- as.character(bondora_sample$UserName)
77
78  # Get upper triangle indices
79  ut <- which(upper.tri(similarity_matrix), arr.ind = TRUE)
80
81  # Filter by threshold
82  ut <- ut[similarity_matrix[ut] >= threshold, ]
83
84  # Create edge list
85  p2p_bondera <- data.frame(
86      from = vertex_names[ut[,1]],
87      to   = vertex_names[ut[,2]],
88      weight = similarity_matrix[ut],
89      stringsAsFactors = FALSE
90  )
91
92  # Difference between attributes present and not present
93  att_diffs <- setdiff(keep_cols, similarities)
94
95  # Merge Data Frames to Ensure other Attributes Appear in Edge List
96  p2p_bondera <- merge(p2p_bondera, bondora_sample[att_diffs],
97                    by.x = "from", by.y = "UserName", all.x = TRUE)
98  colnames(p2p_bondera)[4:14] <- paste0("from_", colnames(p2p_bondera)[4:14])
99
100 p2p_bondera <- merge(p2p_bondera, bondora_sample[att_diffs],
101                   by.x = "to", by.y = "UserName", all.x = TRUE)
102 colnames(p2p_bondera)[15:25] <- paste0("to_", colnames(p2p_bondera)[15:25])
103
104 # ———————————————————————————————————————————————————————————— #
105
106 bondera_sample_atts$name <- bondera_sample_atts$UserName
107
108 p2p_bondera_network <- igraph::graph_from_data_frame(
109     d = p2p_bondera[c('from','to','weight')], directed = FALSE)
110
111 walktrap_comm <- snafun::extract_comm_walktrap(p2p_bondera_network)
112 snafun::g_summary(p2p_bondera_network)
113
114 par(mfrow = c(1, 2))
115
116 plot(p2p_bondera_network,
117     main = "No Community Detection",
118     edge.arrow.size = 0.3,
119     edge.color = rgb(0,0,0, alpha = 0.15),
120     vertex.frame.color = "black",
121     vertex.label = NA,
122     vertex.frame.size = 3,
123     vertex.size = 5,
124     vertex.shape = "circle",
125     vertex.color = "cornsilk",
126     edge.curved = FALSE,
127     layout = igraph::layout.fruchterman.reingold)
128
129 plot(walktrap_comm, p2p_bondera_network,
130     main = "Walktrap Communities",
```

```
131        edge.arrow.size = 0.3,
132        edge.color = rgb(0,0,0, alpha = 0.15),
133        vertex.frame.color = "black",
134        vertex.label = NA,
135        vertex.frame.size = 3,
136        vertex.size = 5,
137        vertex.shape = "circle",
138        vertex.color = "cornsilk",
139        edge.curved = FALSE,
140        layout = igraph::layout.fruchterman.reingold)
141
142 # Add Vertex Attributes
143 igraph::V(p2p_bondera_network)$Age <- bondera_sample_atts$Age[
144    match(igraph::V(p2p_bondera_network)$name, bondera_sample_atts$name)]
```

data_processing_bondera.R

B   Technology statement student(s)

During the preparation of this work, I/We used [**NAME TOOL / SERVICE / VER-SION OF AI TOOL**] in order to [**REASON**]. The following parts of the assignment were affected/generated by AI tool usage: [**INTRODUCTION / METHODS / xxx, DISCUSSION**]. After using this tool/service, [**NAME STUDENT(S)**] evaluated the validity of the tool's outputs, including the sources that generative AI tools have used, and edited the content as needed. As a consequence, [**NAME STUDENT(S)**] take(s) full responsibility for the content of their work.