

# Social Network Analysis - Assignment 1

Group 7

Floris Vermeulen      Martijn van Iterson      Niek Fleerakkers      Patryk Grodek  
Samir Sabitli

2025-09-12

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Methodology</b>	<b>2</b>
2.1	Dataset . . . . .	2
	<b>References</b>	<b>3</b>
<b>A</b>	<b>Source Code - Data Processing</b>	<b>4</b>
<b>B</b>	<b>Technology Statement</b>	<b>7</b>

# 1 Introduction

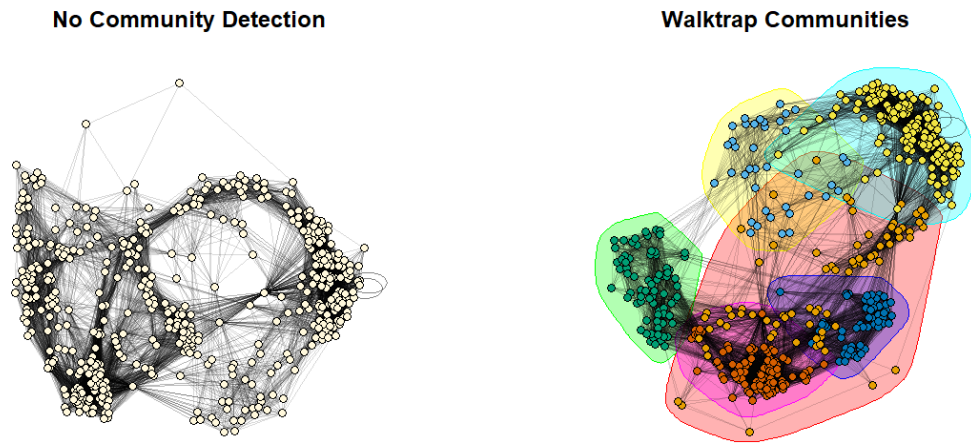
Hypothesis	ERGM Term	Motivation
H1: Students who I consider friends also consider me their friend	<b>mutual</b>	The <b>mutual</b> term captures the tendency for outgoing ties to co-occur with incoming ties between actors, capturing the dynamic of reciprocation.

## 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

```
1
2 # ----- #
3 # Reset WD
4 #setwd("../..")
5
6 # Import the Bondora P2P Dataset
7 bondora_raw <- read.csv("dataset/LoanData_Bondora.csv",
8                       header = TRUE)
9 cols <- colnames(bondora_raw)
10
11 # ----- #
12
13 # Subset only Columns we need
14 keep_cols <- c("LoanId", "UserName", "NewCreditCustomer", "LanguageCode",
15              "Age", "Gender", "Country", "Amount", "Interest",
16              "LoanDuration", "UseOfLoan", "Education", "MaritalStatus",
17              "NrOfDependants", "Rating", "Restructured",
18              "NoOfPreviousLoansBeforeLoan", "MonthlyPayment")
19 bondora <- bondora_raw[keep_cols]
20
21 # Remove Rows with NAs -> Complete Dataset Preferred
22 bondora_complete <- na.omit(bondora)
23 sum(is.na(bondora_complete))
24
25 # Observe Class of Each Attribute
26 sapply(bondora_complete, class)
27
28 # Make Binary Indicators Binary
29 new_customer_mapping <- c("True" = 1, "False" = 0)
30 bondora_complete$NewCreditCustomer <- new_customer_mapping[
31   bondora_complete$NewCreditCustomer]
32
33 # Replace User inputs of Blank Dependants with Zero
34 bondora_complete$NrOfDependants[bondora_complete$NrOfDependants == ""] <- NA
35 bondora_complete$NrOfDependants[is.na(bondora_complete$NrOfDependants)] <- 0
36
37 # Make the Column Numeric
38 bondora_complete$NrOfDependants <- as.numeric(bondora_complete$NrOfDependants)
39 bondora_complete$NrOfDependants[is.na(bondora_complete$NrOfDependants)] <- 0
40
41 # Make Restructured Binary
42 bondora_complete$Restructured <- new_customer_mapping[
43   bondora_complete$Restructured]
44
45 # Randomly Remove Observations until Desired Size is Reached
46 set.seed(42)
47 sample_size <- 500
48 sample_indices <- sample(1:nrow(bondora_complete), sample_size)
49 bondora_sample <- bondora_complete[sample_indices, ]
50
51 # ----- #
52
53 # Choose Feature Subset for Similarity Metric
54 similarities <- c("LoanDuration", "Amount", "MonthlyPayment", "NewCreditCustomer",
55                 "NoOfPreviousLoansBeforeLoan", "LanguageCode")
56 bondora_similar <- bondora_sample[similarities]
57
58 # Standardise Numeric Features in the Similarity Set
59 bondora_similar_scaled <- scale(bondora_similar)
60
61 # Compute Cosine Similarity
62 cosine_sim <- function(X) {
63   # numerator: dot product
64   sim <- X %*% t(X)
```

```

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$UserName)
77 |
78 | # Get upper triangle indices
79 | ut <- which(upper.tri(similarity_matrix), arr.ind = TRUE)
80 |
81 | # Filter by threshold
82 | ut <- ut[similarity_matrix[ut] >= threshold, ]
83 |
84 | # Create edge list
85 | p2p_bondera <- data.frame(
86 |   from = vertex_names[ut[,1]],
87 |   to   = vertex_names[ut[,2]],
88 |   weight = similarity_matrix[ut],
89 |   stringsAsFactors = FALSE
90 | )
91 |
92 | # Difference between attributes present and not present
93 | att_diffs <- setdiff(keep_cols, similarities)
94 |
95 | # Merge Data Frames to Ensure other Attributes Appear in Edge List
96 | p2p_bondera <- merge(p2p_bondera, bondora_sample[att_diffs],
97 |   by.x = "from", by.y = "UserName", all.x = TRUE)
98 | colnames(p2p_bondera)[4:14] <- paste0("from_", colnames(p2p_bondera)[4:14])
99 |
100 | p2p_bondera <- merge(p2p_bondera, bondora_sample[att_diffs],
101 |   by.x = "to", by.y = "UserName", all.x = TRUE)
102 | colnames(p2p_bondera)[15:25] <- paste0("to_", colnames(p2p_bondera)[15:25])
103 |
104 | # ----- #
105 |
106 | bondora_sample_atts$name <- bondora_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

During the preparation of this work, I/We used [NAME TOOL / SERVICE / VERSION 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.