Take-Home\_Ex02

Teo Wee Siang Roy

2025-05-21

# Mini-Case 3

Over the past decade, Oceanus has transitioned from a fishing-based economy to one increasingly reliant on ocean tourism, following crackdowns on illegal fishing. This shift has caused tensions, especially with recent high-profile events like pop star Sailor Shift’s planned music video shoot on the island.

Journalist Clepper Jessen, formerly of FishEye and now with the Hacklee Herald, has been investigating these tensions. His latest focus is the sudden closure of Nemo Reef. Through intercepted communications and investigative analysis, Clepper uncovered a network of expedited approvals and covert arrangements implicating Oceanus officials, Sailor Shift’s team, powerful local families, and the conservation group The Green Guardians—suggesting a potential case of corruption and manipulation.

Task: Create innovative visualizations and visual analytics tools to help Clepper expose the truth behind these events.

# Set-up

## Import library package

pacman::p\_load(tidyverse, jsonlite, lubridate, ggraph, tidygraph, visNetwork)

## Import data and extract nodes and edges

# Load graph structure  
graph\_raw <- fromJSON("TH02/data/MC3\_graph.json", flatten = TRUE)  
nodes <- graph\_raw$nodes  
edges <- graph\_raw$edges

## Parse Timestamp and Extract Day

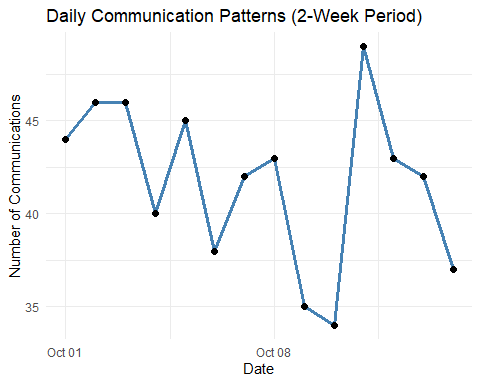
# Convert timestamp to datetime and extract date  
comm\_events <- nodes %>%  
 filter(type == "Event", sub\_type == "Communication") %>%  
 select(id, timestamp) # Make sure 'timestamp' is the actual column name  
  
# Step 2: Parse timestamp into datetime and extract the date  
comm\_events <- comm\_events %>%  
 mutate(  
 start\_time = ymd\_hms(timestamp, quiet = TRUE),  
 day = as.Date(start\_time)  
 )

## Summarise daily Communication Volume

# Count number of communications per day  
daily\_comms <- comm\_events %>%  
 count(day, name = "num\_comms")

## Visualise the Temporal Pattern

# Plot as a time series  
ggplot(daily\_comms, aes(x = day, y = num\_comms)) +  
 geom\_line(color = "steelblue", linewidth = 1.2) +  
 geom\_point(size = 2) +  
 labs(title = "Daily Communication Patterns (2-Week Period)",  
 x = "Date", y = "Number of Communications") +  
 theme\_minimal()



# Question 2

## Filter People + Vessels only

# Filter relevant entity nodes  
entity\_nodes <- nodes %>%  
 filter(type == "Entity", sub\_type %in% c("Person", "Vessel")) %>%  
 select(id, name, sub\_type)

## Assign groups based on name/subtype for better analytics

entity\_nodes <- entity\_nodes %>%  
 mutate(  
 group = case\_when(  
 name %in% c("Green Guardians", "Reef Guardian", "EcoVigil", "Sentinel") ~ "Environmentalism",  
 name %in% c("Sailor Shift", "Sailor Shifts Team") ~ "Sailor Shift",  
 name %in% c("Mariners Dream", "Marlin", "Recreational Fishing Boats") ~ "Fishing/Leisure",  
 name %in% c("Serenity", "Horizon", "Osprey", "Remora", "Neptune", "Mako") ~ "Tourism/Leisure",  
 TRUE ~ sub\_type # fallback: Person or Vessel  
 )  
 )

## Filter Edges Between People and Vessels (Communication events)

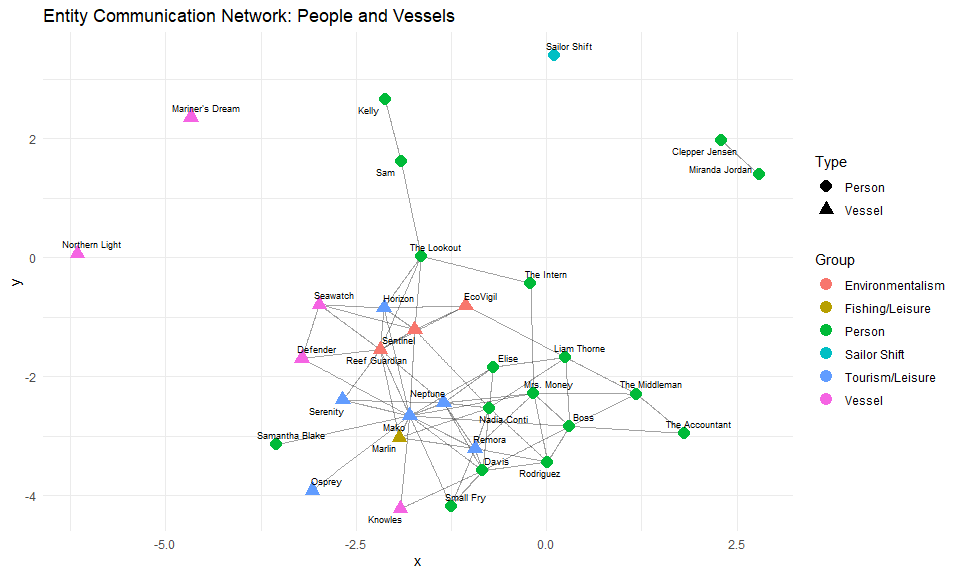
# Identify valid entity IDs  
valid\_ids <- entity\_nodes$id  
  
# Get communication events  
comm\_event\_ids <- nodes %>%  
 filter(type == "Event", sub\_type == "Communication") %>%  
 pull(id)  
  
# Filter 'sent' and 'received' edges for communication events  
sent\_edges <- edges %>%  
 filter(type == "sent", target %in% comm\_event\_ids) %>%  
 select(event\_id = target, entity\_id = source)  
  
received\_edges <- edges %>%  
 filter(type == "received", source %in% comm\_event\_ids) %>%  
 select(event\_id = source, entity\_id = target)  
  
# Combine and create entity pairs  
entity\_event\_edges <- bind\_rows(sent\_edges, received\_edges)  
  
entity\_pairs <- entity\_event\_edges %>%  
 inner\_join(entity\_event\_edges, by = "event\_id") %>%  
 filter(entity\_id.x != entity\_id.y) %>%  
 distinct(entity\_id.x, entity\_id.y) %>%  
 rename(source = entity\_id.x, target = entity\_id.y) %>%  
 filter(source %in% valid\_ids & target %in% valid\_ids)

## Build and Cluster Graph

graph\_tbl <- tbl\_graph(nodes = entity\_nodes, edges = entity\_pairs, directed = FALSE)  
  
# Detect clusters (communities) for color grouping  
graph\_tbl <- graph\_tbl %>%  
 mutate(cluster = as.factor(group\_infomap()))

## Visualise Groups

ggraph(graph\_tbl, layout = "fr") +  
 geom\_edge\_link(alpha = 0.2) +  
 geom\_node\_point(aes(color = group, shape = sub\_type), size = 4) +  
 geom\_node\_text(aes(label = name), repel = TRUE, size = 2.5) +  
 labs(title = "Entity Communication Network: People and Vessels",  
 color = "Group", shape = "Type") +  
 theme\_minimal()



# 2A Interactive graph

# ---- NODES PREPARATION ----  
nodes\_vis <- graph\_tbl %>%  
 as\_tibble() %>%  
 mutate(  
 id = id,  
 label = name,  
 group = group,  
 shape = ifelse(sub\_type == "Person", "dot", "triangle"),  
 title = paste0(  
 "<div style='color:black;'>",  
 "<b>", name, "</b><br>",  
 "Type: ", sub\_type, "<br>",  
 "Group: ", group, "<br>",  
 "Cluster: ", cluster,  
 "</div>"  
 )  
 ) %>%  
 select(id, label, group, shape, title)  
  
# ---- EDGES PREPARATION ----  
edges\_tbl <- graph\_tbl %>%  
 activate(edges) %>%  
 as\_tibble() %>%  
 distinct(from, to, .keep\_all = TRUE)  
  
nodes\_tbl <- graph\_tbl %>%  
 activate(nodes) %>%  
 as\_tibble() %>%  
 mutate(node\_index = row\_number())  
  
edges\_vis <- edges\_tbl %>%  
 left\_join(nodes\_tbl %>% select(node\_index, id), by = c("from" = "node\_index")) %>%  
 rename(from\_id = id) %>%  
 left\_join(nodes\_tbl %>% select(node\_index, id), by = c("to" = "node\_index")) %>%  
 rename(to\_id = id) %>%  
 select(from = from\_id, to = to\_id)  
  
# ---- INTERACTIVE VISUALIZATION ----  
visNetwork(nodes\_vis, edges\_vis, width = "800px", height = "800px") %>%  
 visIgraphLayout(layout = "layout\_with\_fr") %>%  
 visNodes(  
 font = list(size = 24, color = "black", face = "bold"),  
 shadow = TRUE  
 ) %>%  
 visEdges(  
 arrows = "to",  
 color = list(color = "gray", highlight = "red"),  
 smooth = TRUE  
 ) %>%  
 visOptions(  
 highlightNearest = list(enabled = TRUE, degree = 1),  
 nodesIdSelection = TRUE # This enables the dropdown  
 ) %>%  
 visLegend(  
 position = "left",  
 main = "Group Legend",  
 useGroups = TRUE,  
 ncol = 1,  
 stepX = 100,  
 stepY = 50  
 ) %>%  
 visGroups(groupname = "Environmentalism", color = "#4DAF4A") %>%  
 visGroups(groupname = "Sailor Shift", color = "#E41A1C") %>%  
 visGroups(groupname = "Fishing/Leisure", color = "#377EB8") %>%  
 visGroups(groupname = "Tourism/Leisure", color = "#FFD700") %>%  
 visGroups(groupname = "Officials", color = "#FF7F00") %>%  
 visGroups(groupname = "Industry", color = "#999999")

Node Encoding:

* Shape: Distinguishes people (dots) from vessels (triangles)
* Color: Groups entities by affiliation/role
* Labels: Direct identification of entities

Interactive Features: - Dropdown selection: Allows focused exploration of specific entities

* Hover tooltips: Provide detailed node information
* Highlighting: Shows connected neighbors when selecting nodes
* Legend: Clearly identifies group classifications

# 2B Insights

**How This Helps Clepper**

* **Identifying Key Players:**
* The diagram visually highlights which people and vessels are most interconnected, helping to focus attention on the core of the network.
* **Spotting Isolated Entities:**
* Outliers or isolated nodes (like “Sailor Shift” and “Mariner’s Dream”) may warrant further investigation to understand why they are not connected to the main group.
* **Understanding Group Structure:**
* The use of color and shape makes it easy to distinguish between different types of entities and their likely group affiliations, revealing the structure of the broader community and its subgroups.

**Central Cluster:** The central area of the diagram contains a dense cluster of people and vessels, suggesting frequent or multiple interactions among these entities. This could indicate a core group involved in related activities.

**Group Separation:** Environmentalism vessels (green triangles) are grouped together and connected to several people, indicating that these vessels are involved with or communicate with multiple individuals.

Tourism/Leisure vessels (yellow triangles) are also clustered, with some connections to people in the central group.

Sailor Shift (red circle) is isolated, with no visible direct connections in this snapshot, suggesting either a lack of direct communication in this subset or that this node is involved in different parts of the network not shown here.

**Peripheral People and Vessels:** Some individuals and vessels are not connected to the main cluster, indicating limited or no recent communication events with the central group.

**In summary:** This network diagram provides a clear visual summary of the relationships and communication patterns between people and vessels in the Oceanus community. It shows which entities are most central, which groups are most tightly connected, and which individuals or vessels are on the periphery or isolated from the main activity. This can guide further investigation into group dynamics and potential areas of interest in the knowledge graph