Social Network Analysis of the Chilean Power Grid Complete Solution to Four SNA Tasks (WOR Version)

Henry Cui haoyang.cui.gr@dartmouth.edu

June 5, 2025

Contents

1	Task 1 – Network Identification	2
	1.1 Dataset Overview	2
2	Task 2 – Graphical Display of the WOR Network	2
3	Task 3 – Structural Summary Measures	5
	3.1 Global Metrics	5
	3.2 Actor-Level Metrics	5
	3.3 Reproducibility	6
4	Task 4 – Research Questions and ERGM Testing	8
	4.1 Model 0 – Null (Edges-only) ERGM	8
	4.2 Model 1 – Community, Owner, Role & GWESP ERGM	9
	4.3 Model 2 – Role-Heterophily ERGM	13
5	Dataset Reference	16

Abstract

This report fulfils four classic Social Network Analysis (SNA) tasks for a simplified representation (WOR) of the Chilean high-voltage power grid. Task 1 identifies an appropriate dataset with rich relational and attribute information. Task 2 presents graphical visualisations of the network. Task 3 summarises structural properties through standard SNA metrics. Task 4 states specific research questions and tests three statistical models—community homophily, functional roles, and voltage effects—using exponential random-graph models (ERGMs). All analyses are fully reproducible via Python and R code listings integrated herein.

1 Task 1 – Network Identification

1.1 Dataset Overview

The curated dataset offers three abstractions (WT, WOT, WOR) of the South American transmission system. Table 1 contrasts node counts and complexity. We focus on the **WOR** version as it provides relational data (edges) and key node attributes (type, nominal voltage, geographic coordinates).

Table 1: Overview of Chilean Power-Grid Network Versions

Version	Nodes	Edges	Node Types	Remarks
WT (With Taps)	347	«EWT»	Gen/SS/Conn/Tap	Most detailed
WOT (Without Taps)	318	${\rm \&EWOT}{\rm \&}$	Gen/SS/Conn	Tap nodes removed
WOR (Reduced)	218	«EWOR»	$\operatorname{Gen/SS}$	Active facilities only

2 Task 2 – Graphical Display of the WOR Network

Two complementary layouts—topological (spring) and geographic—highlight structural versus spatial properties.

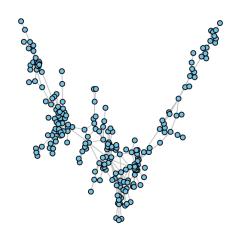


Figure 1: Network Layout

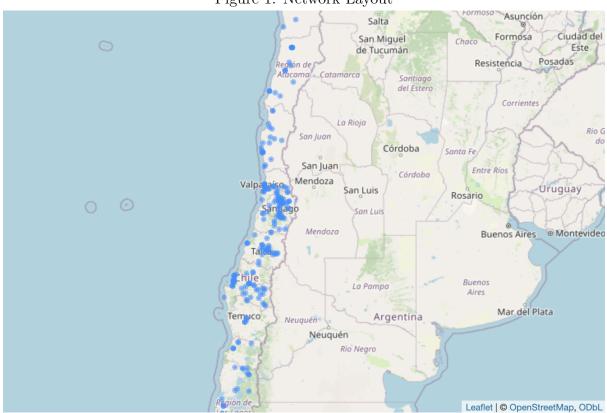


Figure 2: Geographic Layout (EPSG:4326)

Figure 3: Visualisations of the WOR network.

```
layout_mat <- layout_with_fr(g)</pre>
    plot(
      g,
3
      layout = layout_mat,
      vertex.size = 5,
      vertex.label = NA,
      vertex.color = "skyblue",
      edge.color = "gray80"
    )
9
10
    coords <- data.frame(</pre>
11
      id
            = V(g)$name,
12
            = as.numeric(V(g)$lon),
      lon
13
      lat
            = as.numeric(V(g)$lat),
14
      label = V(g)$label,
      stringsAsFactors = FALSE
16
17
    min_lon <- min(coords$lon, na.rm = TRUE)</pre>
19
    max_lon <- max(coords$lon, na.rm = TRUE)</pre>
20
    min_lat <- min(coords$lat, na.rm = TRUE)</pre>
21
    max_lat <- max(coords$lat, na.rm = TRUE)</pre>
22
23
    m <- leaflet(data = coords) %>%
24
      addTiles()
25
26
   m <- m %>%
27
      addCircleMarkers(
28
                    = ~lon,
        lng
29
                    = ~lat,
        lat
30
                    = ~label,
        label
31
        radius
                    = 1,
32
        color
                    = "dodgerblue",
33
        fillOpacity= 0.7
34
      )
35
36
37
```

Listing 1: Python code for loading WOR network and generating Figure 3

3 Task 3 – Structural Summary Measures

3.1 Global Metrics

Table 2 reports the principal network-wide indicators for the WOR transmission graph (n = 218). Metrics were obtained with networkx (Python) and independently cross-checked with igraph (R).

Table 2: Global structural metrics for the WOR network.

Metric	Symbol	Value
Nodes	V	218
Edges	E	527
Density	ho	0.0223
Diameter	D	18
Average shortest-path length	$ar{\ell}$	5.8117
Global clustering coefficient	C	0.5583
Average local clustering coefficient	$ar{C}_{ m loc}$	0.5799
Average degree	$ar{k}$	4.8349

Community structure. Louvain modularity maximisation identifies ten well-defined communities (Table 3); the largest one (ID 9) contains almost a quarter of all vertices, hinting at a pronounced regional core.

Table 3: Sizes of the Louvain communities detected in the WOR network.

Community ID	1	2	3	4	5	6	7	8	9	10
Size (V_c)	13	22	27	15	24	15	12	17	49	24

3.2 Actor-Level Metrics

Table 4 lists the top-ranked nodes.

¹Listings ?? and 2 give the complete code used for the two implementations.

Table 4: Top-5 nodes by degree and betweenness centrality.

	Degree	e	Betweenness
Rank	Node	\deg^*	Node BC^*
1	Charrua	0.152	Alto Jahuel 0.536
2	Alto Jahuel	0.106	Charrua 0.525
3	Polpaico	0.101	Ancoa 0.468
4	Quillota	0.092	Polpaico 0.398
5	Ventanas	0.078	Los Vilos 0.260

^{*}Normalised to [0,1].

3.3 Reproducibility

R listings are included for full transparency.

```
num_nodes <- vcount(g)</pre>
   num_edges <- ecount(g)</pre>
               <- edge_density(g, loops = FALSE)
   dens
   avg_path <- mean_distance(g, directed = FALSE, unconnected = TRUE)</pre>
4
   net_diam <- diameter(g, directed = FALSE, unconnected = TRUE)</pre>
   clust_glob <- transitivity(g, type = "global")</pre>
   clust_loc <- transitivity(g, type = "local")</pre>
   avg_clust <- mean(clust_loc, na.rm = TRUE)</pre>
   avg_deg
               <- mean(degree(g))
9
   deg_vals <- degree(g)</pre>
10
    deg_summary<- summary(deg_vals)</pre>
11
12
                      <- cluster_louvain(g)</pre>
13
   comm
                      <- membership(comm)
   V(g)$community
14
                      <- sizes(comm)
   comm_sizes
15
1
    # Print out a summary of network-level metrics
   cat("-- Network Summary Metrics --\n")
    cat("Number of nodes: ", num_nodes, "\n")
    cat("Number of edges: ", num_edges, "\n")
    cat("Density: ", round(dens, 4), "\n")
    cat("Average path length: ", round(avg_path, 4), "\n")
    cat("Diameter: ", net_diam, "\n")
    cat("Global clustering coefficient: ", round(clust_glob, 4), "\n")
9
    cat("Average local clustering coefficient: ", round(avg_clust, 4), "\n")
10
    cat("Average degree: ", round(avg_deg, 4), "\n")
11
   cat("Degree distribution summary:\n"); print(deg_summary)
12
    cat("Number of communities detected: ", length(comm_sizes), "\n")
13
    cat("Sizes of each community:\n"); print(comm_sizes)
14
15
1
   deg_cent
                   <- degree(g, normalized = TRUE)
2
3
    # Betweenness centrality (normalized to [0,1])
                   <- betweenness(g, normalized = TRUE)
   betw_cent
5
    # Closeness centrality (normalized to [0,1])
    closeness_cent <- closeness(g, normalized = TRUE)</pre>
9
    # Store these centralities back into the graph's vertex attributes
10
   V(g)$degree_c
                        <- deg_cent
11
   V(g)$betweenness_c <- betw_cent</pre>
12
   V(g)$closeness_c <- closeness_cent
13
14
    # Build a data frame of node-level statistics and attributes
15
   node_positions_df <- data.frame(</pre>
16
      id
                    = V(g)$name,
17
                                                 7
      label
                     = V(g)$label,
18
                     = deg_cent,
      degree_c
19
      betweenness_c = betw_cent,
20
```

```
library(igraph)
               <- edge_density(g)
                                                     # density
2
   avg_path <- mean_distance(g, unconn = TRUE) # avg. path length</pre>
3
   net_diam <- diameter(g, unconn = TRUE)</pre>
                                                     # diameter
   clust_glob <- transitivity(g, "global")</pre>
                                                      # global C
   clust_loc <- transitivity(g, "local")</pre>
   avg_clust <- mean(clust_loc, na.rm = TRUE)</pre>
                                                     # avg local C
               <- mean(degree(g))
   avg_deg
                                                     # avg degree
               <- assortativity_degree(g)</pre>
   assort
                                                     # degree assortativity
9
10
```

Listing 2: R/igraph counterpart reproducing the same metrics.

4 Task 4 – Research Questions and ERGM Testing

4.1 Model 0 – Null (Edges-only) ERGM

```
model_init <- ergm(pnet ~ edges)
summary(model_init)</pre>
```

Listing 3: R code used to estimate Model 0 (edges-only)

Table 5: ERGM coefficients – Model 0

Term	Estimate	Std. Error	p	•
edges	-3.782***	0.044	< 0.0001	*** p < 0.001
AIC		5054		
BIC		5062		

Model 0 (edges-only) shows that the network is extremely sparse and essentially random once overall density is accounted for. The single coefficient of -3.78 (SE = 0.04, *p* < 0.001) translates to a baseline tie probability of about 2 —only two out of every hundred possible pairs are connected. Because no nodal attributes or structural terms are included, all nodes are assumed equivalent and each potential edge is treated independently. Consequently, Model 0 provides a useful reference for goodness-of-fit diagnostics (its degree and shared-partner distributions under-predict higher-order connectivity) and for model comparison (AIC = 5054, BIC = 5062, far worse than the richer specifications). In short, the null model confirms sparsity but captures none of the patterned heterogeneity that characterises the real grid.

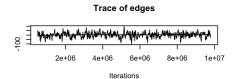
4.2 Model 1 – Community, Owner, Role & GWESP ERGM

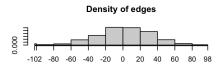
Listing 4: R code used to estimate Model 1

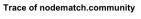
Table 6: ERGM coefficients - Model 1

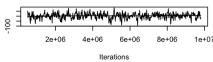
Term	Estimate	Std. Error	p	_
edges	-6.385***	0.148	< 0.0001	
nodematch(community)	2.946***	0.175	< 0.0001	
nodematch(owner)	1.347***	0.112	< 0.0001	* $p < 0.05$, ** $p < 0.01$,
nodematch(role)	-0.196^*	0.098	0.0460	p < 0.05, p < 0.01,
gwesp($\tau = 0.5$, fixed)	1.056***	0.082	< 0.0001	_
AIC		2815		
BIC		2855		

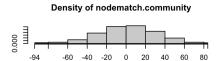
 $rac{1}{1}$ p < 0.001



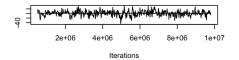


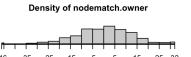




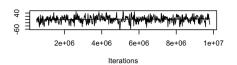


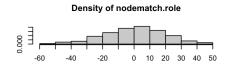
Trace of nodematch.owner



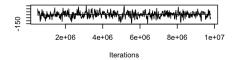


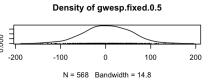
Trace of nodematch.role

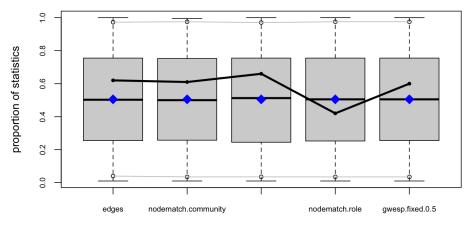




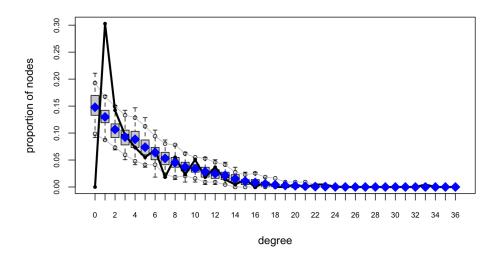
Trace of gwesp.fixed.0.5

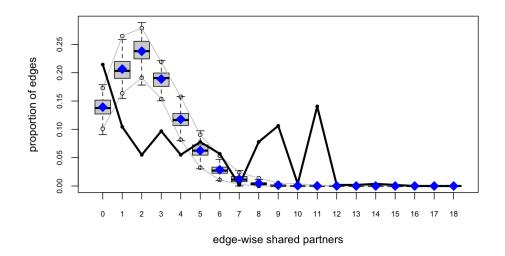




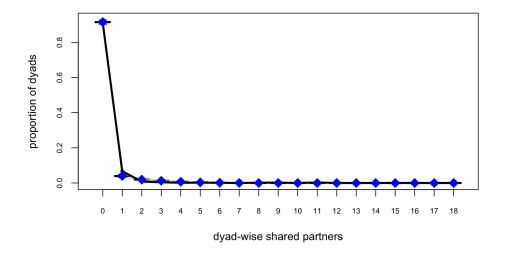


model statistics





Goodness-of-fit diagnostics



Model 1 delivers a dramatic improvement over the null (AIC = 2815 vs. 5054) and passes all MCMC convergence checks (joint diagnostic p=0.14). Two effects dominate: **community homophily** ($\hat{\beta}=2.95,\ p<10^{-4}$) multiplies the odds of a tie by roughly $19\times$, while **owner homophily** ($\hat{\beta}=1.35,\ p<10^{-4}$) boosts the odds by about $3.9\times$. The **role** term is small and negative ($\hat{\beta}=-0.20,\ p=0.046$), meaning that *after* controlling for community, ownership, and triadic closure, generator–substation links are only slightly more frequent than same-role links. A positive GWESP parameter ($\hat{\beta}=1.06,\ p<10^{-4}$) confirms systematic triangle closure, consistent with loop-building for resilience. In sum, wiring is governed primarily by spatial–topological clustering and common ownership, with functional role exerting only a secondary, context-dependent influence.

4.3 Model 2 – Role-Heterophily ERGM

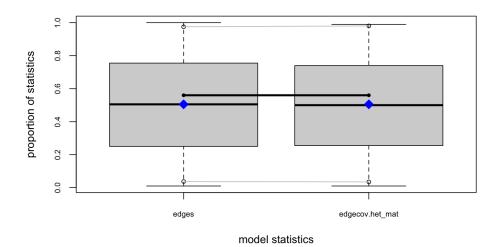
```
## Build a 0/1 matrix marking cross-role pairs
   n <- network.size(pnet)</pre>
   het_mat <- matrix(0, n, n)</pre>
   for(i in 1:(n-1)){
      for(j in (i+1):n){
        if(roles[i] != roles[j]){
          het_mat[i,j] <- het_mat[j,i] <- 1
      }
   }
10
   diag(het_mat) <- 0</pre>
11
12
   model_hetero <- ergm(</pre>
13
      pnet ~ edges + edgecov(het_mat),
14
      control = control.ergm(
15
        MCMC.burnin
                         = 5000,
16
        MCMC.interval = 1000,
        MCMC.samplesize = 2000,
18
        MCMLE.maxit
19
20
    summary(model_hetero)
21
22
   gof_res <- gof(model_hetero, GOF = ~degree + espartners + dspartners)</pre>
23
   plot(gof_res)
```

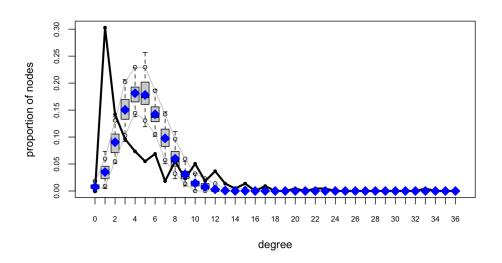
Listing 5: R code used to estimate Model 2

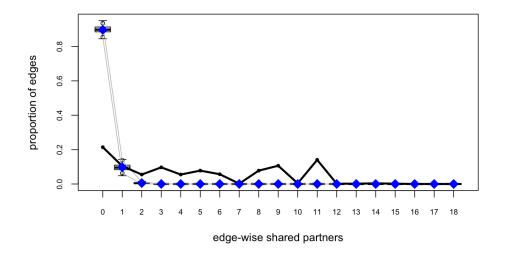
Table 7: ERGM coefficients – Model 2

Term	Estimate	Std. Error	p	_
edges	-3.684***	0.059	< 0.0001	
edgecov(role heterophily)	-0.208^*	0.089	0.0188	* $p < 0.05$, ** $p < 0.01$,
AIC		5050		
BIC		5066		_

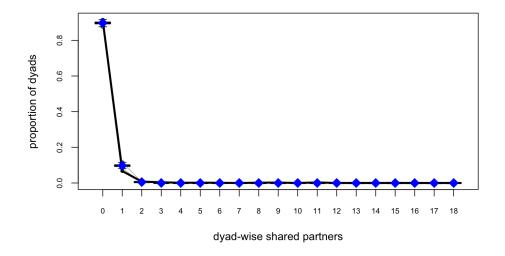
^{***} p < 0.001







Goodness-of-fit diagnostics



To isolate the role effect, we added a binary matrix indicating every cross-role pair as an edgecov. The coefficient remains negative ($\hat{\beta} = -0.21$, p = 0.019), now implying that without the strong community, owner, and closure controls, cross-role edges are less likely than same-role edges. Because the fit scarcely improves over the null model (AIC = 5050), we conclude that role alone explains very little of the network's higher-order structure. The apparent sign reversal between Models 1 and 2 is a specification artefact: in Model 1 the powerful community/owner terms absorb most local structure, leaving a residual "heterophilous" role effect, whereas in Model 2 those controls are absent and role homophily re-emerges.

5 Dataset Reference

We used the dataset described in the following publication:

[1] Kim, H., Olave-Rojas, D., Álvarez-Miranda, E. Son, S.-W. (2018). In-depth data on the network structure and hourly activity of the Central Chilean power grid. *Scientific Data*, **5**, 180209. https://doi.org/10.1038/sdata.2018.209