

CSS Analysis

George G. Vega Yon

October 22, 2018

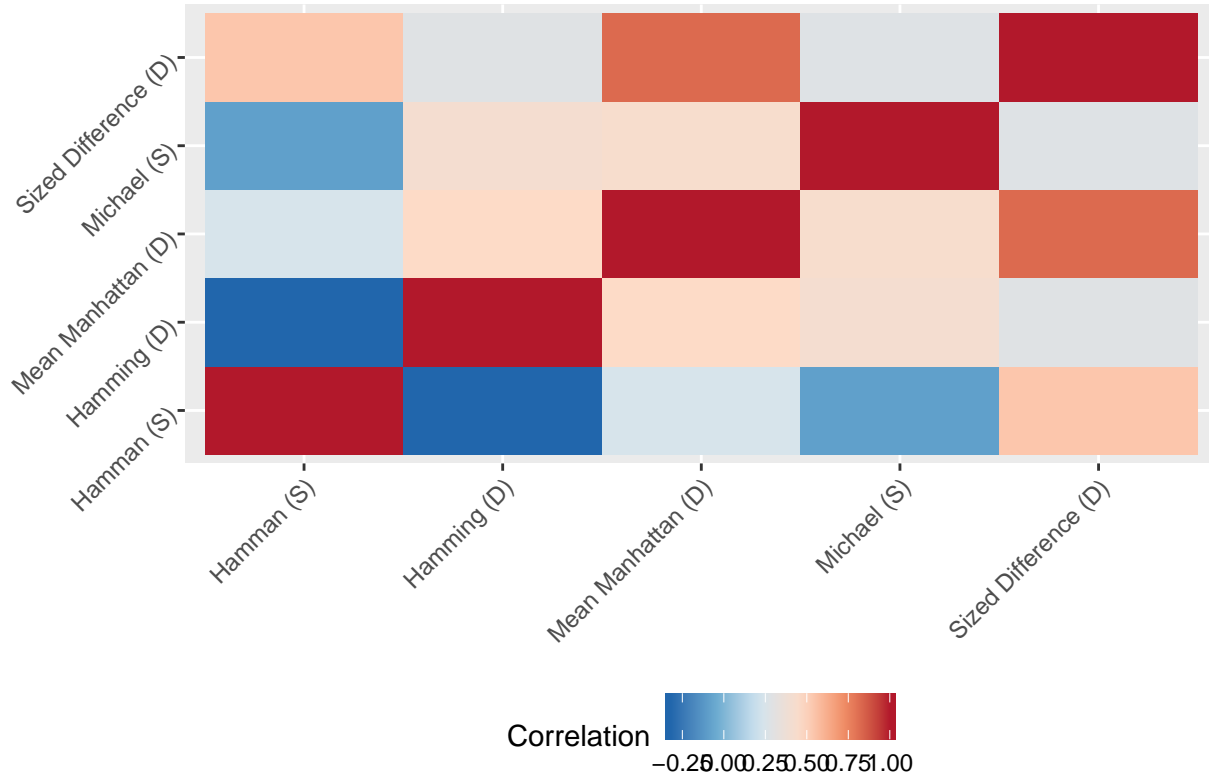
Locally Aggregated Structures (LAS)

- For each group $g \in G$ we generated what the literature calls Locally Aggregated Structure networks (LAS networks).
- A tie (i, j) in the LAS exists if and only if $(i, j) \in CSS_i$ and $(i, j) \in CSS_j$, i.e. if both i and j report the existence of such tie.

Levels of agreement

- For each $i \in N_g$ (individual i in group G), we calculated the following metrics

Correlation between similarity statistics



Observation: The hamming normalized seems to be the same as the Mean manhattan (s)

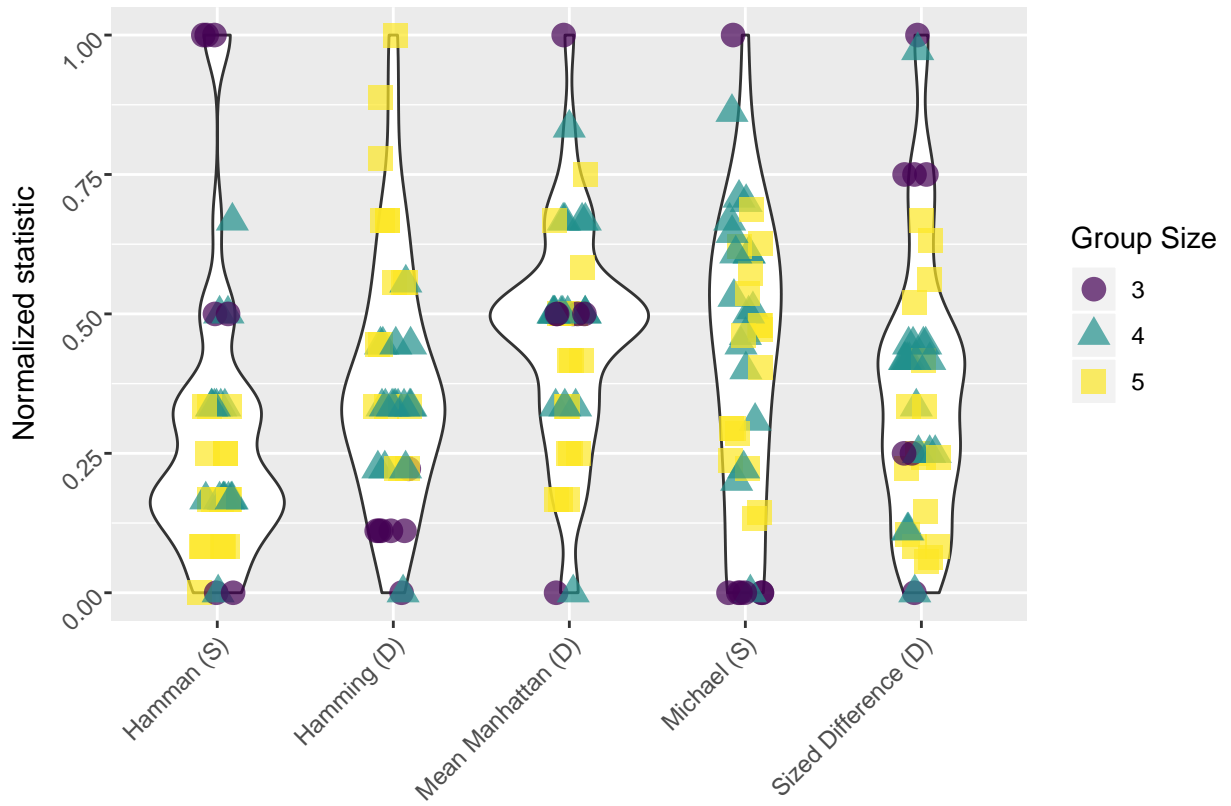


Figure 1: Distribution of Within Group Ranges of Similarity (S) and Distance (D) Statistics with respect to the LAS. Values are normalized to range between 0 and 1.

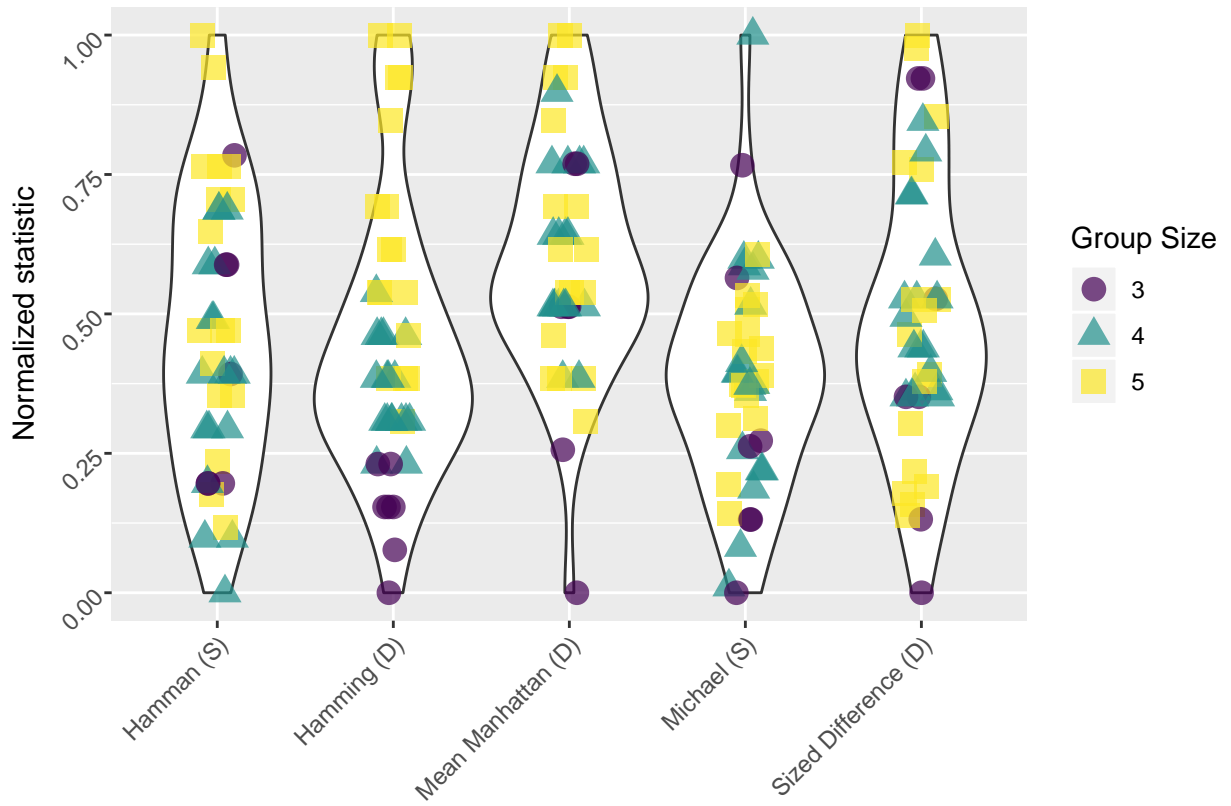


Figure 2: Distribution of Within Group Ranges of Similarity (S) and Distance (D) Statistics within the groups. Values are normalized to range between 0 and 1.

Within groups variability

TCI analysis

LAS

Descriptive statistics

Statistic	GrpSize	N	Min	Max	Variance
Hamman (S)	3	7	0.0000000	2.0000000	0.8095238
Hamman (S)	4	18	0.0000000	1.3333333	0.1045752
Hamman (S)	5	17	0.0000000	0.6666667	0.0412582
Hamming (D)	3	7	0.0000000	2.0000000	0.3333333
Hamming (D)	4	18	0.0000000	5.0000000	1.1764706
Hamming (D)	5	17	2.0000000	9.0000000	4.4705882
Mean Manhattan (D)	3	7	0.0000000	1.0000000	0.0833333
Mean Manhattan (D)	4	18	0.0000000	0.8333333	0.0326797
Mean Manhattan (D)	5	17	0.1666667	0.7500000	0.0310458
Michael (S)	3	7	0.0000000	2.0000000	0.5714286
Michael (S)	4	18	0.0000000	1.7230769	0.1804503
Michael (S)	5	17	0.2666667	1.3730769	0.1201329
Sized Difference (D)	3	7	0.0000000	1.0000000	0.1339286
Sized Difference (D)	4	18	0.0000000	0.9722222	0.0419289
Sized Difference (D)	5	17	0.0555556	0.6666667	0.0415099

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Prop. of Males	0.10 (0.48)	-0.08 (0.45)	-0.08 (0.45)	-0.09 (0.45)	0.05 (0.47)	-0.09 (0.52)	-0.19 (0.48)	-0.06 (0.48)	-0.24 (0.50)	-0.08 (0.50)
Prop. of Non-white	-0.10 (0.33)	-0.02 (0.30)	-0.05 (0.30)	-0.21 (0.30)	-0.11 (0.31)	-0.42 (0.36)	-0.31 (0.33)	-0.31 (0.33)	-0.57 (0.35)	-0.39 (0.34)
Age Range	0.04 (0.03)	0.02 (0.02)	0.03 (0.02)	0.04 (0.02)	0.04 (0.03)	0.04 (0.03)	0.02 (0.03)	0.02 (0.03)	0.03 (0.03)	0.02 (0.03)
Avg GPA	-0.06 (0.16)	-0.10 (0.15)	-0.11 (0.15)	-0.09 (0.15)	-0.07 (0.16)	0.11 (0.18)	0.05 (0.17)	0.06 (0.17)	0.06 (0.17)	0.10 (0.17)
Size = 4	0.26 (0.24)	0.47* (0.23)	0.23 (0.20)	0.48* (0.23)	0.16 (0.22)					
Size = 5	0.82** (0.26)	1.20*** (0.28)	0.68** (0.21)	0.97*** (0.22)	0.68** (0.23)	0.56** (0.17)	0.75*** (0.17)	0.45** (0.16)	0.52** (0.16)	0.53** (0.16)
(Intercept)	-0.22 (1.27)	0.35 (1.19)	0.68 (1.22)	0.28 (1.19)	0.21 (1.26)	-0.79 (1.32)	-0.09 (1.26)	-0.03 (1.26)	0.02 (1.32)	-0.53 (1.28)
Distance/Similarity										
Hamman (S)	0.05 (0.18)					-0.15 (0.31)				
Hamming (D)		-0.11* (0.05)					-0.11* (0.05)			
Mean Manhattan (D)			-0.85* (0.38)					-1.01* (0.45)		
Michael (S)				-0.36* (0.16)					-0.40 (0.21)	
Sized Difference (D)					-0.40 (0.33)					-0.63 (0.40)
R ²	0.37	0.45	0.45	0.44	0.39	0.34	0.43	0.44	0.41	0.39
Adj. R ²	0.24	0.33	0.34	0.33	0.27	0.20	0.31	0.32	0.28	0.26
Num. obs.	42	42	42	42	42	35	35	35	35	35
RMSE	0.48	0.45	0.45	0.45	0.47	0.47	0.44	0.44	0.45	0.45

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2: Regression using different distance/similarity measurements

Within

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Prop. of Males	0.06 (0.48)	0.07 (0.51)	0.06 (0.50)	0.03 (0.46)	0.06 (0.50)	-0.17 (0.52)	-0.18 (0.56)	-0.20 (0.56)	-0.15 (0.51)	-0.35 (0.56)
Prop. of Non-white	-0.11 (0.32)	-0.11 (0.32)	-0.11 (0.32)	-0.11 (0.31)	-0.11 (0.32)	-0.39 (0.35)	-0.39 (0.36)	-0.39 (0.35)	-0.40 (0.34)	-0.39 (0.35)
Age Range	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.03 (0.03)	0.04 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
Avg GPA	-0.07 (0.16)	-0.06 (0.16)	-0.06 (0.17)	-0.05 (0.16)	-0.06 (0.16)	0.08 (0.18)	0.10 (0.18)	0.09 (0.18)	0.10 (0.18)	0.12 (0.18)
Size = 4	0.22 (0.22)	0.25 (0.25)	0.24 (0.22)	0.30 (0.21)	0.24 (0.22)					
Size = 5	0.80*** (0.22)	0.82* (0.35)	0.80** (0.23)	0.85*** (0.21)	0.79** (0.22)	0.63*** (0.17)	0.64** (0.22)	0.59** (0.16)	0.58** (0.16)	0.58** (0.16)
(Intercept)	0.00 (1.28)	-0.13 (1.26)	-0.04 (1.30)	0.18 (1.21)	-0.07 (1.28)	-0.57 (1.33)	-0.68 (1.34)	-0.58 (1.37)	-0.43 (1.31)	-0.59 (1.30)
Distance/Similarity										
Hamman (S)	-0.10 (0.20)					-0.17 (0.21)				
Hamming (D)		-0.01 (0.04)					-0.01 (0.04)			
Mean Manhattan (D)			-0.18 (0.64)					-0.35 (0.76)		
Michael (S)				-0.43 (0.25)					-0.39 (0.30)	
Sized Difference (D)					-0.15 (0.52)					-0.70 (0.63)
R ²	0.37	0.37	0.37	0.42	0.37	0.35	0.34	0.34	0.38	0.37
Adj. R ²	0.24	0.24	0.24	0.30	0.24	0.21	0.20	0.20	0.24	0.23
Num. obs.	42	42	42	42	42	35	35	35	35	35
RMSE	0.48	0.48	0.48	0.46	0.48	0.47	0.47	0.47	0.46	0.46

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3: Regression using different distance/similarity measurements

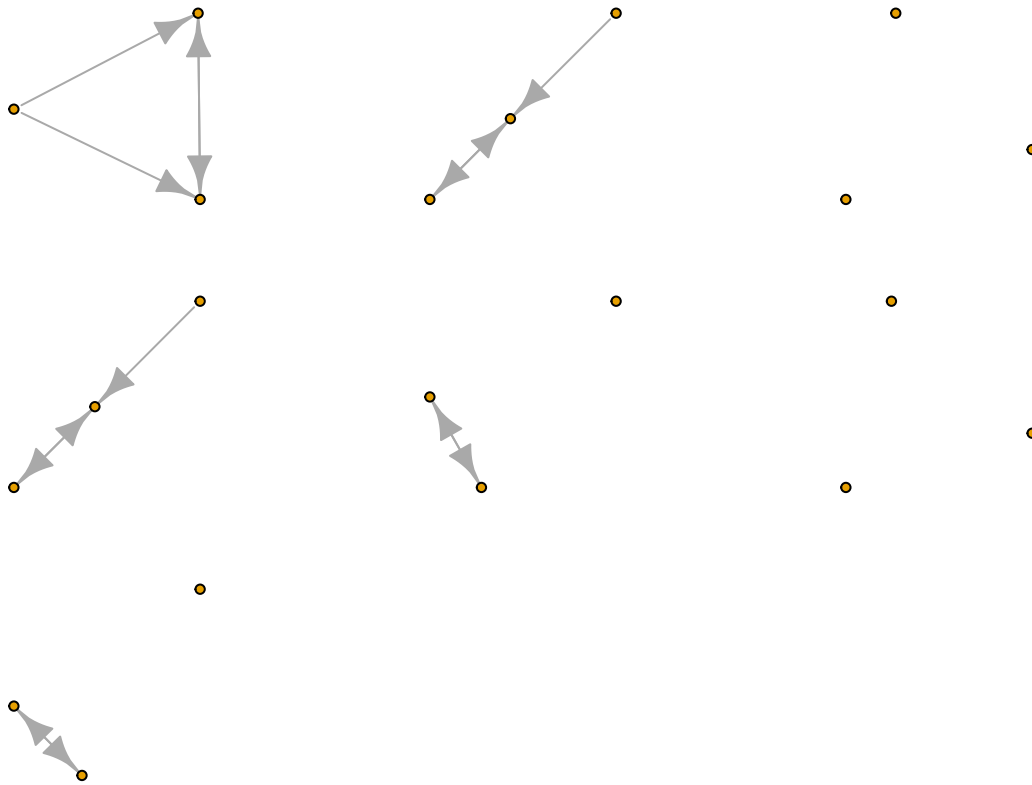


Figure 3: LAS Networks of size 3

Network plots

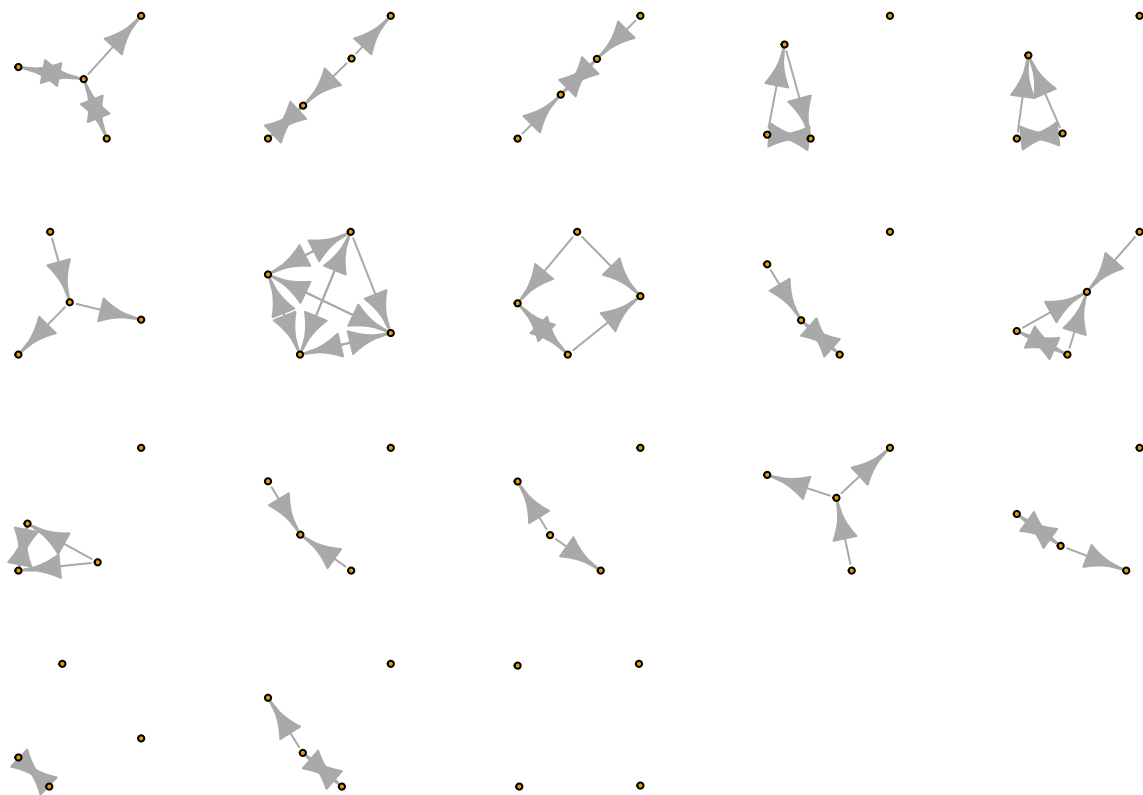


Figure 4: LAS Networks of size 4.

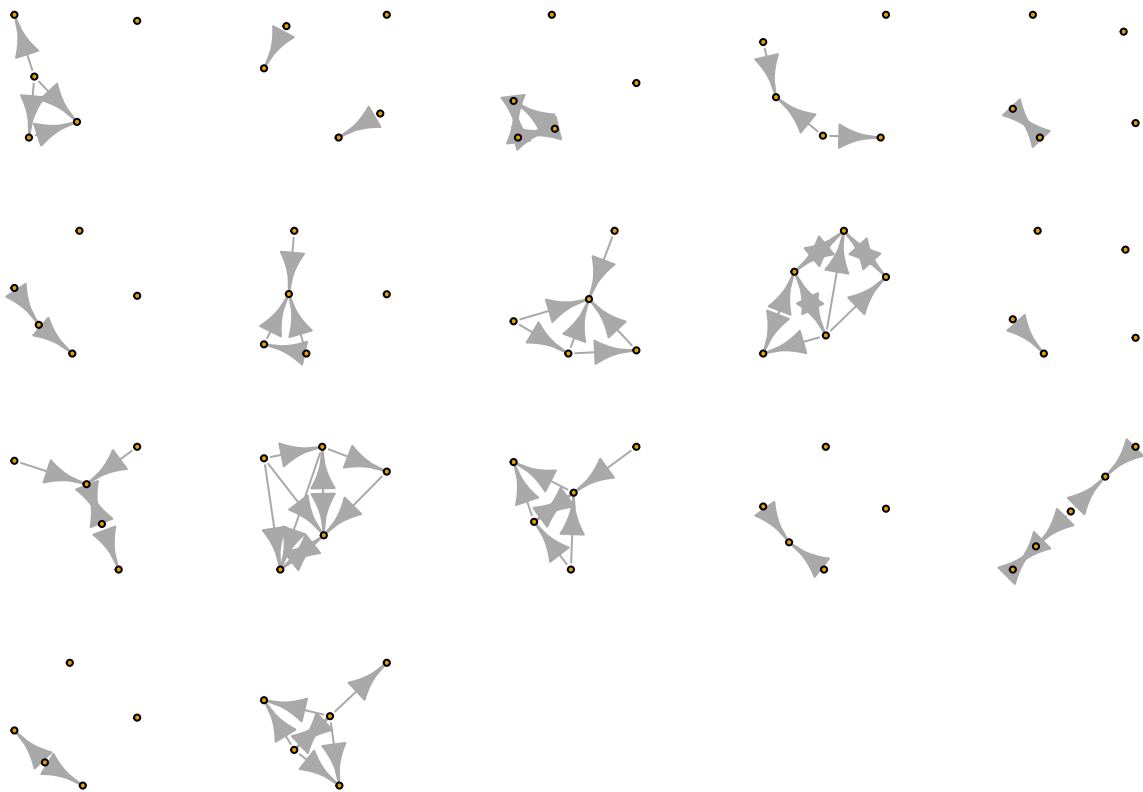


Figure 5: LAS Networks of size 5.