The sparse learning of probabilistic graphical models to identify the significant compounds

Ziyi Yang

7 de mayo de 2019

1. Results

We use different sparse learning methods of probabilistic graphical models to identify the significant compounds and record the results.

- **Neighborhood selection method** [1].
  Number of edge = 1916;
  Degree of node > 20: 15;
  Name of compounds:
  Table 1 shows the degree and name of the selected compounds by neighborhood selection method.
  Figure 1 shows the network estimating using neighborhood selection method.

- **Graphical lasso** [2].
  Number of edge = 1687;
  Degree of node > 20: 32;
  Name of compounds:
  Table 2 shows the degree and name of the selected compounds by graphical lasso.
  Figure 2 shows the network estimating using graphical lasso.

- **Scale free networks by reweighed L_1 regularization (glasso-SF)** [3].
  Number of edge = 1735;
  Degree of node > 20: 76;
  Name of compounds:
  Table 3 shows the degree and name of the selected compounds by glasso-SF.
  Figure 3 shows the network estimating using glasso-SF.

- **Sparse PArtial Correlation Estimation (space)** [4].
  Number of edge = 1261;
  Degree of node > 20: 7;
  Name of compounds:

Table 4 shows the degree and name of the selected compounds by sparse partial correlation estimation.
Figure 4 shows the network estimating using sparse partial correlation estimation.

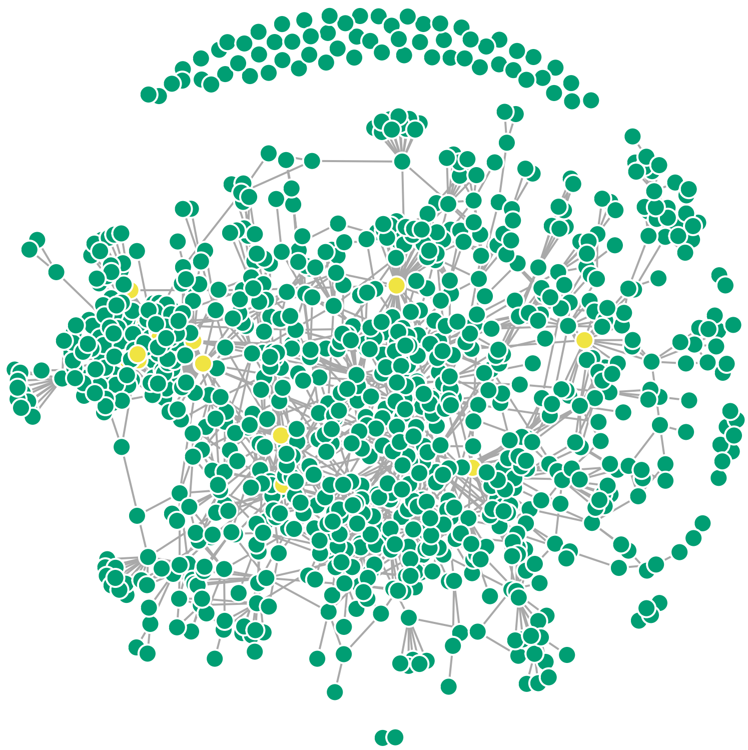

Figura 1: The network estimating using neighborhood selection method. Each node represents a compound and compounds with a degree of greater than 20 are plotted in yellow.

Cuadro 1: The significant compound selected by neighborhood selection method.

Degree	Compound
101	Com_370
56	Com_1057
51	Com_323
40	Com_1548883
36	Com_5484202
34	Com_BS02
33	Com_136419
33	Com_31405
30	Com_637542
26	Com_4723
25	Com_31404
24	Com_445789
24	Com_W7
21	Com_GC71
21	Com_73659

Cuadro 2: The significant compound selected by graphical lasso.

Degree	Compound	Degree	Compound
83	Com_370	24	Com_445789
57	Com_323	24	Com_GC72
47	Com_BS02	24	Com_73299
32	Com_9064	24	Com_7967
32	Com_73659	24	Com_BS22
31	Com_GC71	23	Com_443639
31	Com_1548883	23	Com_442154
30	Com_11504083	23	Com_12305517
29	Com_246728	22	Com_GC56
29	Com_51346141	21	Com_1549111
29	Com_338	21	Com_5481663
28	Com_GC74	21	Com_157139
28	Com_BS03		
28	Com_1057		
27	Com_31404		
27	Com_31405		
26	Com_8051		
26	Com_5280804		
25	Com_72		
25	Com_13572		

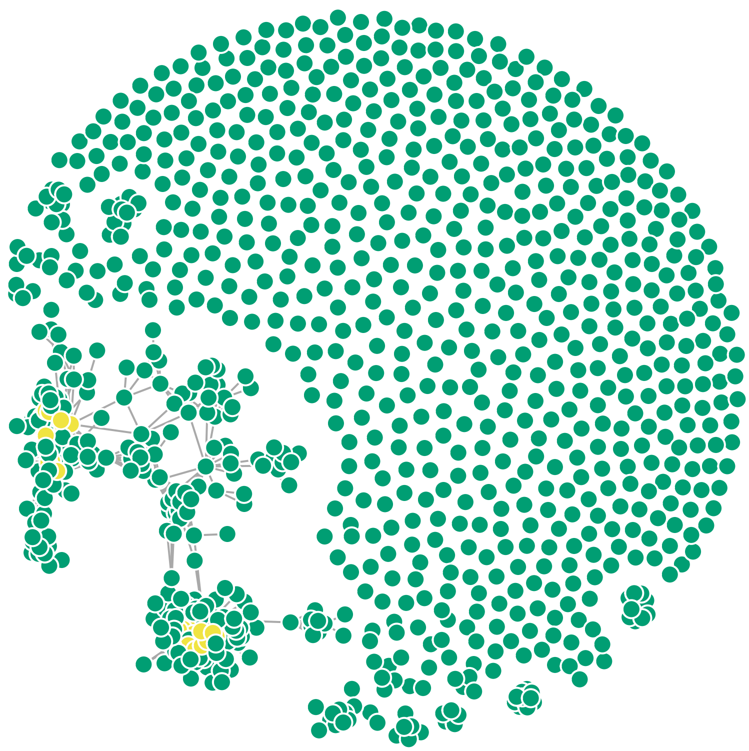

Figura 2: The network estimating using graphical lasso method. Each node represents a compound and compounds with a degree of greater than 20 are plotted in yellow.

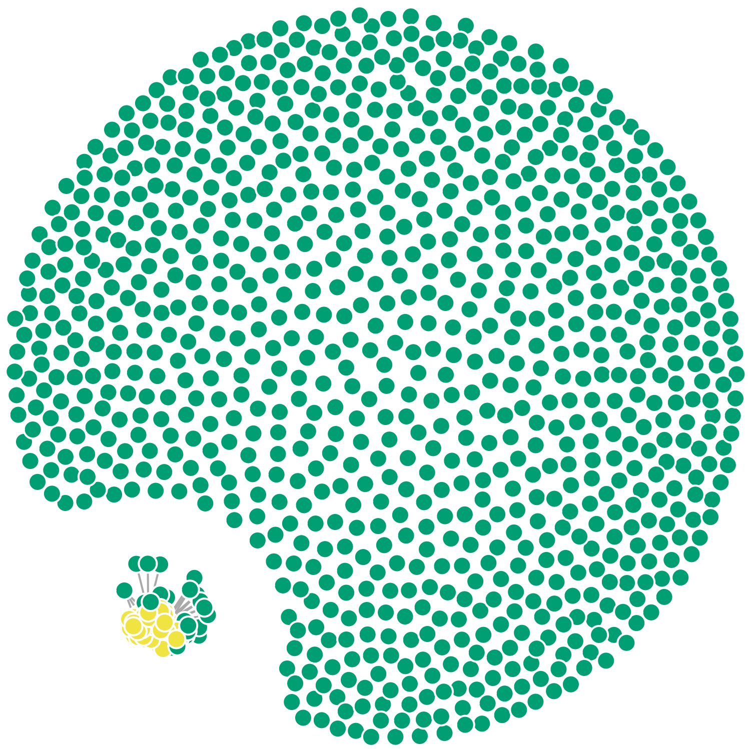

Figura 3: The network estimating using scale free networks by reweighed L_1 regularization method. Each node represents a compound and compounds with a degree of greater than 20 are plotted in yellow.

Cuadro 3: The significant compound selected by glasso-SF.

Degree	Compound	Degree	Compound	Degree	Compound	Degree	Compound
79	Com_370	52	Com_GC74	44	Com_GC72	32	Com_7311
68	Com_1549111	52	Com_72	43	Com_736186	32	Com_64971
63	Com_323	52	Com_5282102	42	Com_GC23	29	Com_7826
61	Com_HQ50	52	Com_51346141	41	Com_GC68	29	Com_5280460
60	Com_HQ07	52	Com_HQ38	41	Com_19814101	29	Com_101341142
58	Com_443639	52	Com_338	40	Com_21672692	29	Com_101341025
58	Com_442154	52	Com_31405	39	Com_NBS03	28	Com_15226483
57	Com_14704550	51	Com_9064	39	Com_BS46	28	Com_GC08
57	Com_HQ37	50	Com_BS03	39	Com_64945	28	Com_157139
56	Com_5280804	50	Com_BS22	38	Com_12310283	26	Com_7967
55	Com_1549106	49	Com_11504083	36	Com_GC25	25	Com_HQ40
55	Com_637542	49	Com_73659	36	Com_73398	25	Com_8468
55	Com_BS02	49	Com_5318767	36	Com_GC07	23	Com_246728
55	Com_1548883	49	Com_5280805	36	Com_5318188	23	Com_100633
55	Com_HQ08	48	Com_10494	36	Com_GC44	23	Com_HQ51
54	Com_GC71	48	Com_BS04	36	Com_10114	21	Com_136419
53	Com_GC56	47	Com_BS33	34	Com_BS47		
53	Com_5481663	45	Com_GC76	33	Com_HQ52		
53	Com_12305517	45	Com_73299	33	Com_79197		
52	Com_31404	44	Com_1057	32	Com_3286789		

Cuadro 4: The significant compound selected by sparse partial correlation estimation (space).

Degree	Compound
506	Com_10114
288	Com_GC25
87	Com_HQ08
56	Com_GC44
51	Com_31404
28	Com_BS04
25	Com_GC68

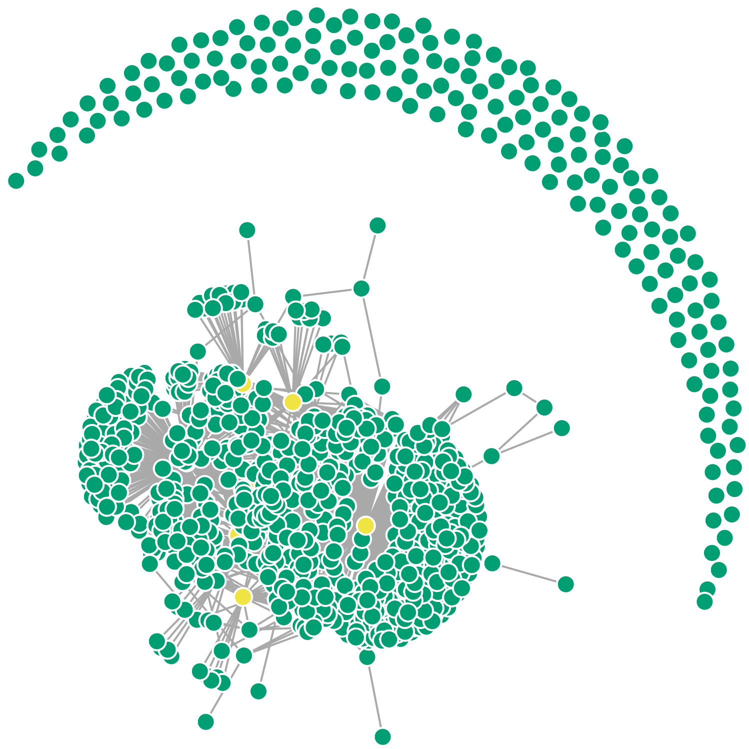

Figura 4: The network estimating using sparse partial correlation estimation. Each node represents a compound and compounds with a degree of greater than 20 are plotted in yellow.

Reference

1. Meinshausen N, Bühlmann P. Variable selection and high-dimensional graphs with the lasso[J]. Annals of Statistics, 2006, 34: 1436-1462.

2. Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model[J]. Biometrika, 2007, 94(1): 19-35.

3. Liu Q, Ihler A. Learning scale free networks by reweighted l1 regularization[C]. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 2011: 40-48.

4. Peng J, Wang P, Zhou N, et al. Partial correlation estimation by joint sparse regression models[J]. Journal of the American Statistical Association, 2009, 104(486): 735-746.