A network approach to ResPOnsE COVID-19 dataset

Arturo Bertero

Introduction

This paper summarizes the analyses I performed on ResPOnsE COVID-19 dataset, wave 1. The focus is on results, hence the theoretical part is missing and will be integrated in the future. The broad aim of this paper is to compare the two most important methodologies for network estimation: *mgm* and CAN.

The Mixed Graphical Model (*mgm*) from Haslbeck & Waldorp (2020) allows to insert continuous and categorical variables in the model. CAN model is better able to generate predictive hypotheses regarding network's dynamics, although being limited to dummy variables only. In order to understand the validity of these methods when applied to empirical case studies, this paper examine how network methodologies can help us to understand the network of attitudes regarding compliance toward preventive behaviors against COVID-19. While doing so, a Structural Equation Model will be used as a *benchmark* to inspect whether these methodologies are capable of grasping complex interaction between an heterogeneous set of variables.

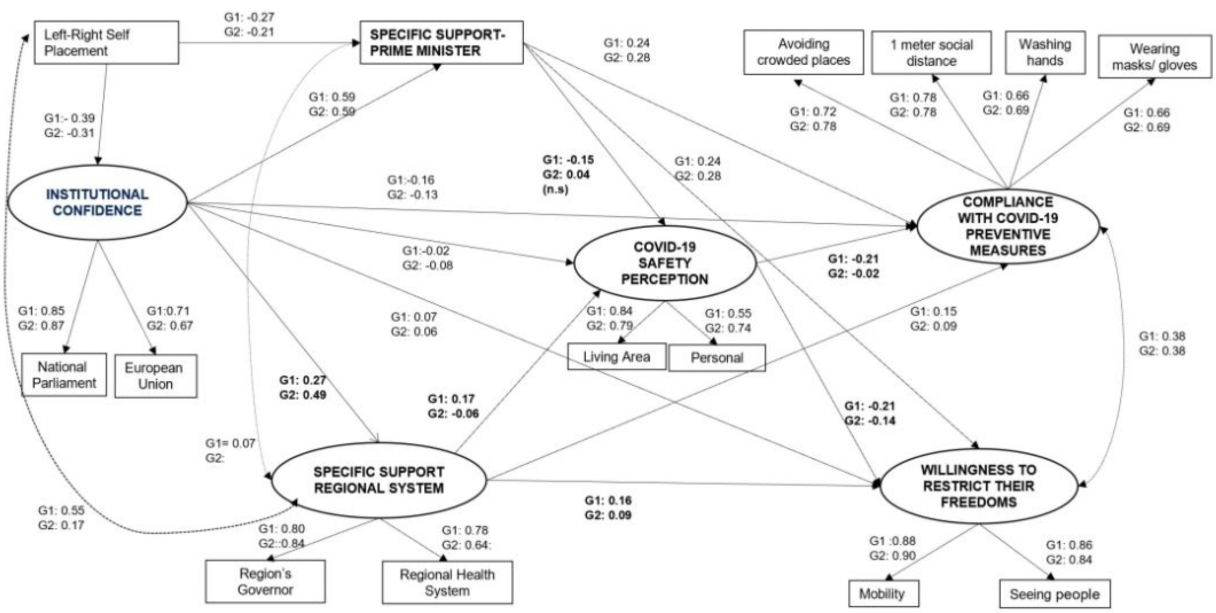

Figure 1: SEM by Guglielmi et al.(2020)

My idea is to use the SEM in Figure 1 above as the the "true model" against which CAN and *mgm* will be compared. Network methodologies and SEM present important differences. SEM is theory driven, and arrows are modeled on the basis of sound theories. In psychological networks ties are undirected, so that causal inferences are always bidirectional. These methods are more exploratory in nature; a tie is drawn when two variables (nodes) are statistically dependent, even after controlling for every other node in the network. Hence, these methods can help us to discover new and unforeseeable patterns among survery's items.

This brief paper is structured in two sections. First, I will apply CAN and *mgm* to the same variables and data composing the SEM in Figure 1. Here I will examine the differences between different network estimation procedures, and those between them and SEM. Second, I will enlarge the scope of the analysis, by widening the list of nodes included in the compliance's networks.

Hypotheses

Some exploratory hypotheses can be formulated. In first place, the two methods should be able to isolate the most important arrows forming Figure 1, meaning that they were able to correctly detect causal inferences predicted by the robust SEM methodology. However, since CAN and *mgm* draws ties only when the associations between nodes hold controlling for each other predictors in the model, it is likely that the smaller associations in Figure 1 will be estimated differently.

- *H1*: *mgm* will isolate the strongest arrows in Figure 1, while neglecting the weakest.
- *H2*: CAN will isolate the strongest arrows in Figure 1, while neglecting the weakest.

Moreover, an additional hypothesis can be formulated regarding the differences between the two estimated networks. Since CAN works with dummy variables, whilst *mgm* with full-scaled ones, CAN network should estimate stronger ties (that is, wider edges) than *mgm*. However, since the logic behind network estimation is similar, no structural differences should appear instead. This means that the statistical test NCT [Network Comparison Test] should shows networks only differing in the strength of their ties, not in the signs of these associations (negative-positive), nor in the number of estimated arrows.

- *H3*: CAN's ties will be stronger than *mgm*'s ones
- *H4*: no other structural differences will emerge between the networks.

Finally, the fifth hypotheses regards the ecological validity of the SEM study. A clear advantage of network methodologies is that they allow us to enlarge the picture of the interactions between many variables composing a survey. Therefore, by creating a bigger network we can investigate if relationships that are statistically significant in the restricted list of variables remain valid also in a broader context.

- *H5*: a broader network will eliminate some of the arrows present in Figure 1

Method

Data used in this paper are from Wave 1 of ResPOnsE data. Three different dataset are created from it. CAN dataset includes every variable selected by Guglielmi et al. (2020). These items were binarized to accommodate CAN's requirements. The second dataset consists of the very same variables, but here they retain their full scale of measurement. Finally, the third dataset is required to answer to *H5*, and it contains more items that are potentially relevant to investigate compliance to preventive behaviors. Also here variables are in their full-scale.

In network methodologies variables are nodes of the networks. However, since many items belonged to standard batteries, factor analysis was performed to reduce the number of the nodes in the graph. This is consistent with the methodology adopted by the only study of compliance to preventive behaviors performed with network methodology (Chambon et al. 2022). Data reduction ensures better readability of the final graph as well better estimation of the ties, since the most interrelated variables are simplified into a unique node. Factor analysis was conducted with PCA method, and results are reported in the additional material section at the end of the paper. Descriptive statistics of each dataset are reported in in the final section as well.

Results: *mgm* vs CAN

Figure 2 shows the comparison between CAN and *mgm* methods. Blue (red) ties represents positive (negative) relationships. The geographical position of nodes is not meaningful, since cartography is subjected to a drawing algorithm that only aims at placing highly interacting nodes closer in space. In the graphs, node belonging to the same communities share the same color. The community stability algorithm isolated two communities in both network. The community consisting of compliance and willingness to restrict is clearly the "behavioral" one, whilst the other involves several attitudes, which are mainly political. The fact that willingness to restrict clusters with compliance is already a precious information, since it means that that node is likely to be a strong predictor of compliance.

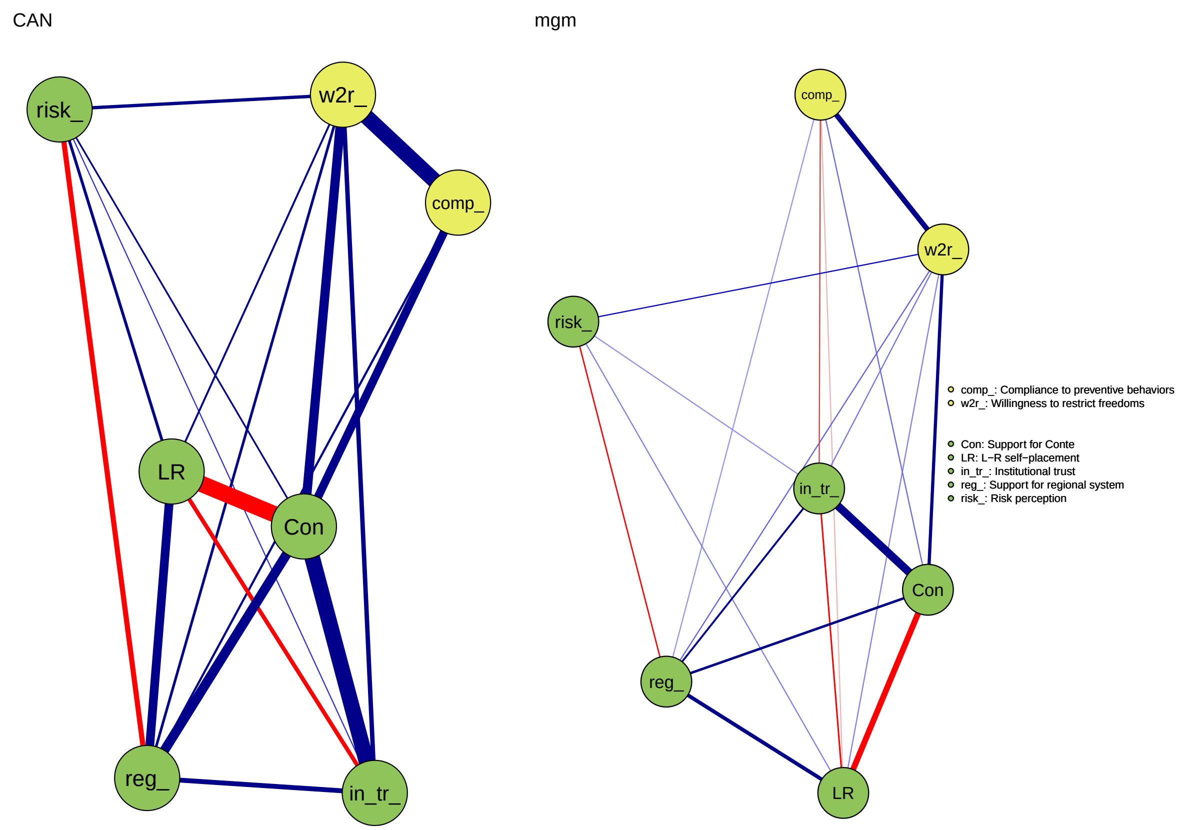

Figure 2: Centrality table

Figure 3 shows the centralty of each node in the two estimated network. This table gives indirect support to *H1* and *H2* by showing that the different procedure rendered similar networks, at least in terms of centrality metrics. The most important node is support for Premier Conte, follwed by "Willingness to restrict". Both of them are charecterized by a strategic position in the network, brokering between clusters.

H1 and H2

The SEM in Figure 1 modeled the following "causal" relationships:

- institutional confidence -> support prime minister [positive]
- institutional confidence -> support regional system [positive]
- institutional confidence -> compliance [negative, almost null]
- institutional confidence -> willingness to restrict [positive, almost null]

- support prime minister -> compliance [positive]
- support prime minister -> risk perception [negative]

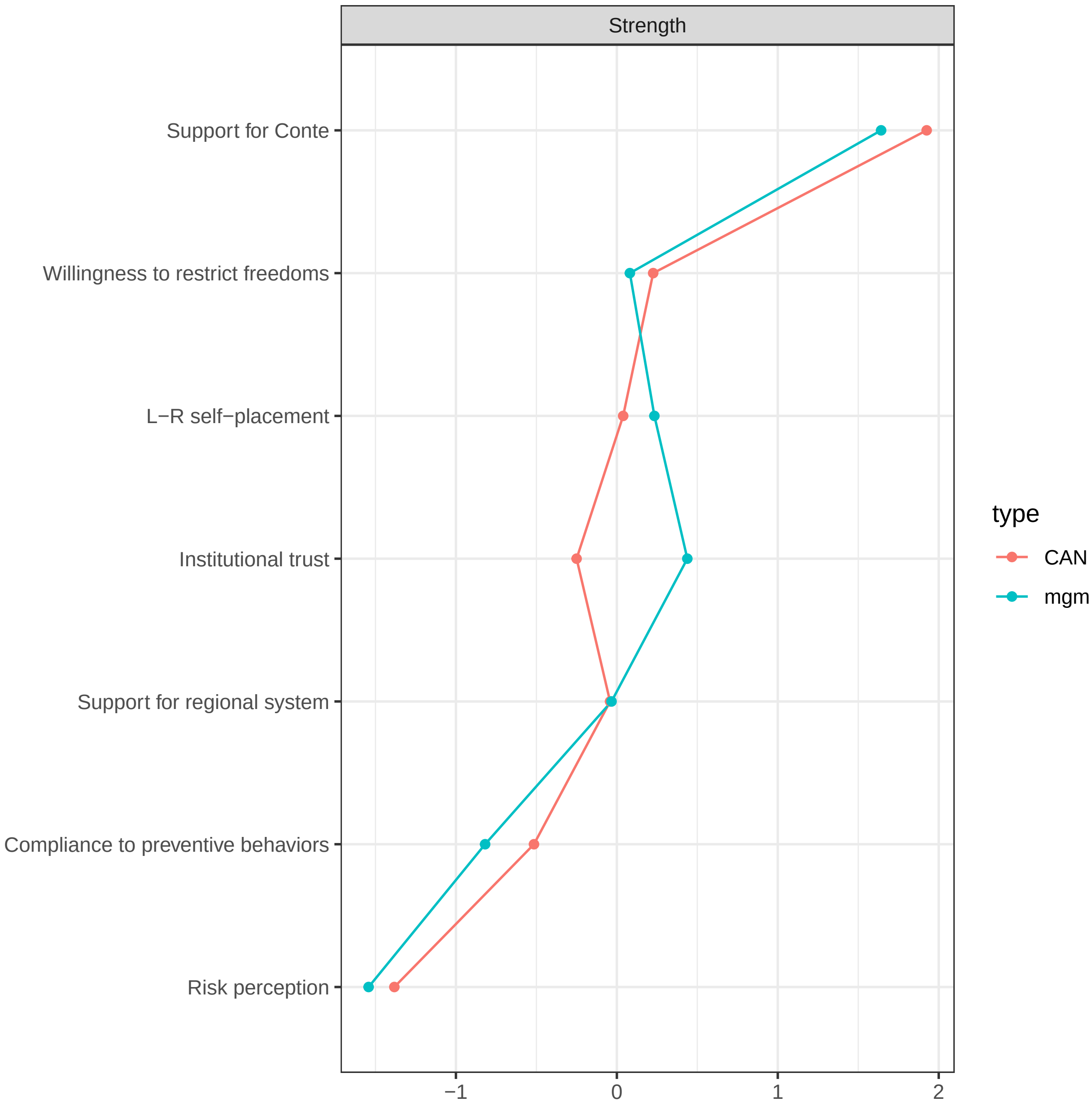

Figure 3: Network comparison

- support prime minister -> willingness to restrict [positive]

- support regional system -> willingness to restrict [positive]
- support regional system -> risk perception [positive]
- support regional system -> compliance [positive]

- risk perception -> compliance [negative]
- risk perception -> willingness to restrict [negative]

- L/R self-placement -> institutional confidence [negative]
- L/R self-placement -> support prime minister [negative]

According to *H1* and *H2*, Figure 2 should isolate these relationships, also preserving their sign (i.e.: negative or positive). The left side of Figure 2 plots the results of CAN network estimation. The following relationships can be observed, with bold content in squared parentheses to highlight different relationships:

- institutional confidence -> support prime minister [positive]
- institutional confidence -> support regional system [positive]
- institutional confidence -> compliance [**null**]
- institutional confidence -> willingness to restrict [positive]

- support prime minister -> compliance [positive]
- support prime minister -> risk perception [**positive, almost null**]
- support prime minister -> willingness to restrict [positive]

- support regional system -> willingness to restrict [positive]
- support regional system -> risk perception [**negative**]
- support regional system -> compliance [positive]

- risk perception -> compliance [**null**]
- risk perception -> willingness to restrict [**positive**]

- L/R self-placement -> institutional confidence [negative]
- L/R self-placement -> support prime minister [negative]

Hence, only 5 out of 14 associations differed in the CAN model. It is interest to point out that these five relationships are characterized by very small SEM coefficients. CAN may have deleted these associations since its estimation procedure iterativelly regress each variable on each other, drawing a tie only when a certain association has passed this form of statistical control. Hence, rather than “missed” or “uncorrect” relationships, the absence of these ties may be indicative of a finest approach to multivariate associations adopted by CAN.

In addition to these relationships, CAN was able to detect other associations. Since the most important variable in the graph is compliance, the behavioral one, it is appropriate to focus on the ties in which it is involved. In Figure 1 compliance was predicted by (decreasing order):

- support prime minister [positive]
- institutional confidence [negative]
- support regional system [positive]
- risk perception [negative]

CAN reports the following associations instead (decreasing order):

- Willingness to restrict [positive]
- Support prime minister [positive]
- support regional system [positive]

Therefore there are two main differences. In first place, the relationship between risk perception and compliance is not validated by CAN. This is interesting, since risk perception emerged as a marginal node in the network considering the centrality metrics (Figure 3 above). The absence of this tie means that this statistical association vanished when controlling for the other node of the network. Second, the strongest tie in which compliance is involved is the one with willingness to restrict. This relationship was indicated as a covariance in the SEM (Figure 1), whilst it is here a bidirectional ties, indicating bidirectional causal inference.

The right part of Figure 2 reports the *mgm* network instead. Here we observe the following relationships:

- institutional confidence -> support prime minister [positive]
- institutional confidence -> support regional system [positive]
- institutional confidence -> compliance [negative]
- institutional confidence -> willingness to restrict [positive]

- support prime minister -> compliance [positive]
- support prime minister -> risk perception [**null**]
- support prime minister -> willingness to restrict [positive]

- support regional system -> willingness to restrict [positive]
- support regional system -> risk perception [positive]
- support regional system -> compliance [**negative**]

- risk perception -> compliance [**null**]
- risk perception -> willingness to restrict [**positive**]

- L/R self-placement -> institutional confidence [**positive**]
- L/R self-placement -> support prime minister [negative]

Again, five relationships differs in the estimated network. Once again these relationships involve SEM's arrows characterized by small coefficients. The behavioral node was connected to (decreasing order):

- Willingness to restrict [positive]
- Support prime minister [positive]
- Support regional system [positive]
- institutional trust [negative]
- L-R [negative]

Hence, with respect to SEM, *mgm* suppressed the weakest association (risk-compliance), while adding Willingness and L-R as predictors of compliance. In sum, *H1* and *H2* are at least partially confirmed. The most important relationship in Figure 1 are indeed present in Figure 2, in both CAN and *mgm* network. Five relationships are different in these network; as predicted, these are the weakest associations characterizing the SEM of Guglielmi et al. (2020).

H3 and H4

According to these hypotheses, networks should differ only in the strength of the estimated ties. We already observed that CAN and *mgm* produced comparable but slightly different networks. For example, Figure 2 showed that compliance is involved in more ties in the *mgm* (vs CAN) network. However, the statistical test NCT can check for both H3 and H4, by investigating whether these differences are statistically significant.

Figure 4 shows the result of NCT. NCT graphs are empty when comparing structural equivalent networks. When networks are different in their ties, edges are displayed instead. Here, blue ties are edges that are stronger in CAN (vs *mgm*) network, in a statistically significant way. Red ties would indicate the reverse, and their absence confirm *H3*. Figure 4 also shows that *H4* is falsified. Structural differences emerge in the networks. In CAN's network, "Support for premier" is associated with risk perception, unlike in the *mgm* one. In the opposite manner, the tie compliance-risk perception is present in the *mgm* network and absent in the CAN one.

H5

Figure 5 below plots the full network. The algorithm of community detection isolated 4 clusters. The red cluster contains two political attitudes: support for Conte and approval of governmental measures to fight COVID-19. The yellow cluster features two demographics variables (age, income of the family) two indicator

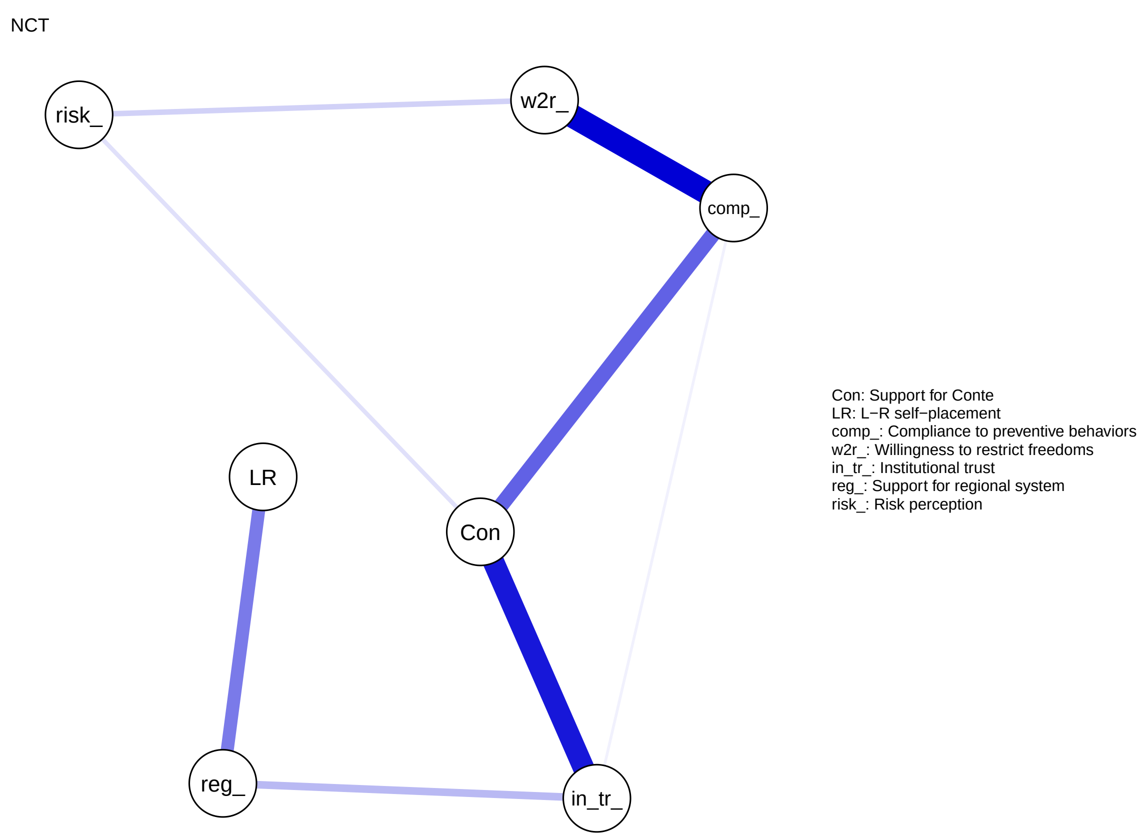

Figure 4: Network comparison

of well-being (health, happyness) and risk perception. Anxiety and depression clustered together (blue). The remaining variables formed a huge green community instead. This community involves political attitudes (L-R, political interest, regional support, economy over public health, collectivism) as well as generic attitdue toward science and medicine. It is important to note that the behavioral node (compliance), and the attitude with which this node was strongly connected with in the previous networks (willingness to restrict) are within this cluster.

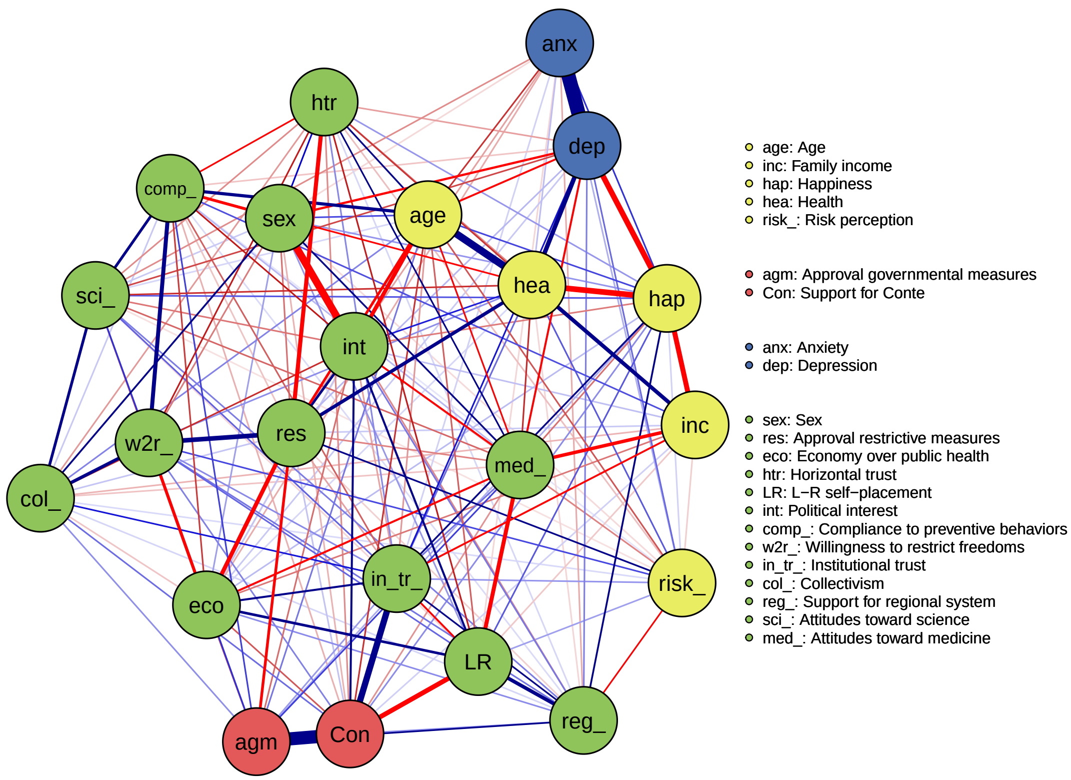

Figure 5: Extended mgm network

Figure 5 can be simplified in Figure 6, where smaller relationships are omitted. These links are still present in the model, and simply excluded from the graph to ease readability.

Figure 7 below plot the centrality of each node. The most important nodes in the network are anxiety and depression. Their high strength is at least partly due to their strong tie. Moreover, depression is also strongly connected with the yellow cluster by means of the happiness node. Support for Conte is instead the most important political attitude, since it is strongly connected to the green cluster forming two important ties with L-R self-placement and trust in institution. An important observation regards risk perception. This node is very marginal in the network, and therefore at the bottom of Figure 7. Risk perception plays a relevant role in the SEM of Figure 1. However, this can be due to choices made while building the model, rather than being due to its true importance in predicting compliance. Moreover low centrality may be indicative of low validity of this survey items. If risk is truly important for compliance, as stressed in the literature, how can it be so marginal in a compliance network?

To answer *H5* we can compare edges of Figure 6 with arrows of the SEM in Figure 1. A comparison with

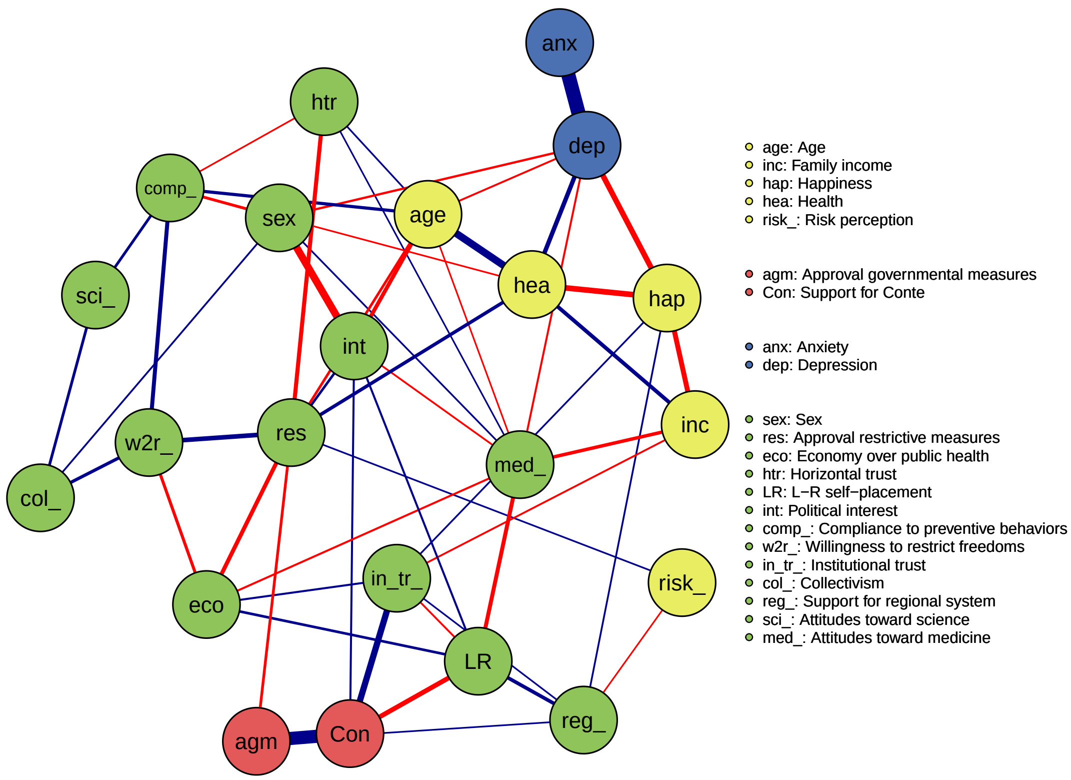

Figure 6: Extended mgm network

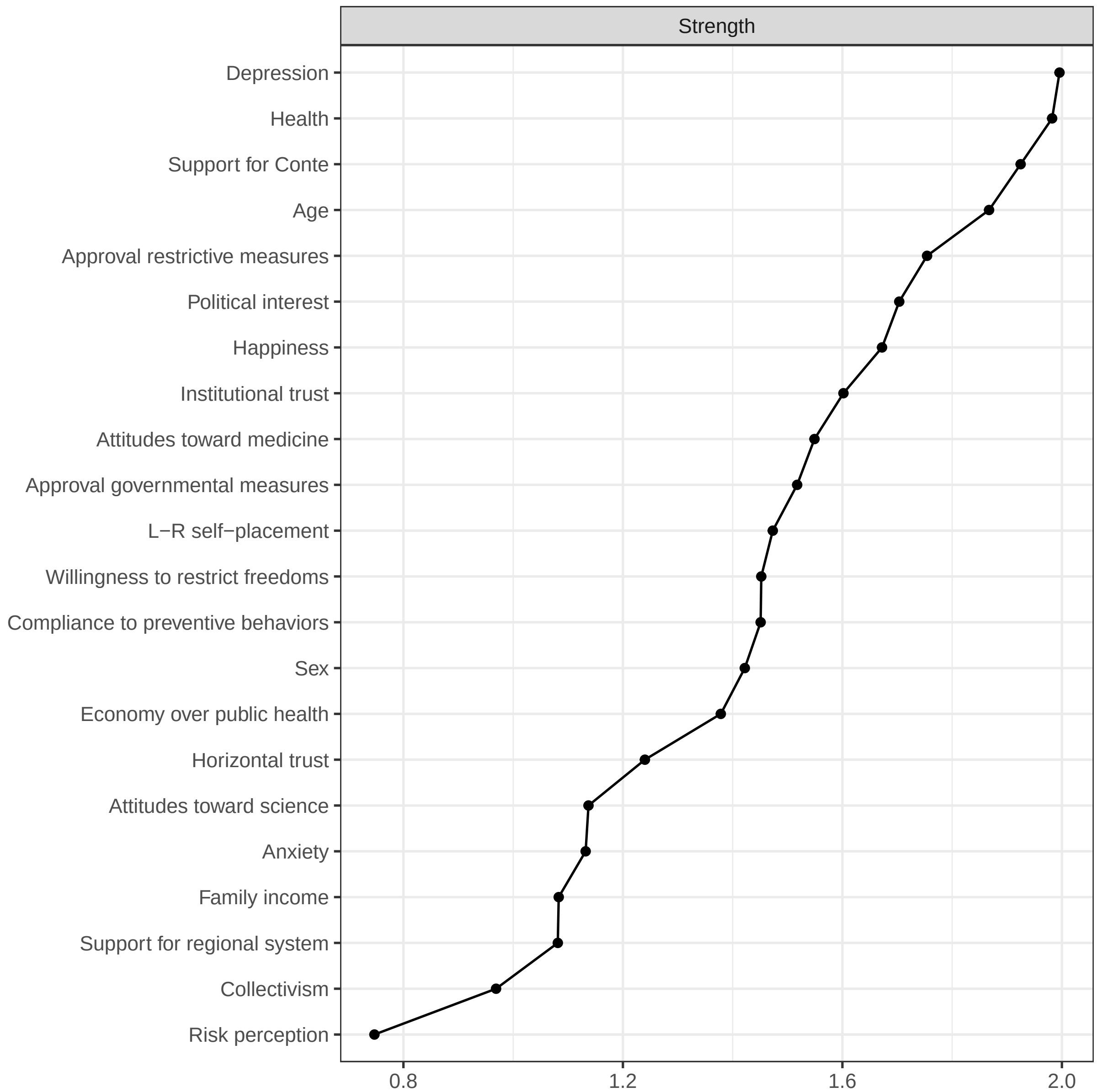

Figure 7: Centrality Table full mgm network

Figure 5 would be painful and pointless. The clear advantage of Figure 6 is that it visually gives an overview of the main bidirectional causal relationships taking place in the extended network.

- institutional confidence -> support prime minister [positive]
- institutional confidence -> support regional system [positive]
- institutional confidence -> compliance [**null**]
- institutional confidence -> willingness to restrict [**null**]

- support prime minister -> compliance [**null**]
- support prime minister -> risk perception [**null**]
- support prime minister -> willingness to restrict [**null**]

- support regional system -> willingness to restrict [**null**]
- support regional system -> risk perception [**negative**]
- support regional system -> compliance [**null**]

- risk perception -> compliance [**null**]
- risk perception -> willingness to restrict [**null**]

- L/R self-placement -> institutional confidence [negative]
- L/R self-placement -> support prime minister [negative]

Finally, predictor of compliance in the completed network are (decreasing order):

- Willingness to restrict [positive]
- Age [positive]
- Attitude toward science [positive]
- Horizontal trust [negative]
- Sex [0 = women]

Hence, *H5* is confirmed. An enlarged picture of ResPOnsE COVID-19 give us a different account of the variables influencing compliance to preventive behaviors. First, some of the variable chosen in the SEM of Figure 1 resulted marginal and peripheral in the full *mgm* network. Second, many of the estimated relationship of that model do not appear in the network representation of ResPOnsE. This means that their statistical significance vanish after controlling for the influence of other variables. Third, the full network gave us a deeper insight at the list of the valid and important predictor of compliance. Within this list, willingness to restrict personal freedom emerged as the most important attitude influencing the behavioral node. Also the demographic variables sex and age strongly influence compliance, and their impact could be calculated only by choosing *mgm* as the estimation procedure. Finally, attitudes toward science are another addition to this list, since they were not present in the model of Figure 1, but indicated as relevant in the network model.

Limitation

Here i mention the two main limitations of this report. First, compliance is more a constant than a variable in W1 of ResPOnsE dataset (see below, additional material, descriptives). This is a serious shortcoming that would deficit any predictive analysis on this data. Second, community detection algorithm was probably uncapable of detecting the true number of communities. This function requires 1000 iterations to be computed with stable results. However, due to the large sample size, the huge amount of nodes, and the limited computational power I have at disposal I had to run 100 iterations only. I had to stop the full cycle after 30 hours of processing. I have to find a way to compute this function renting an R server.

Additional material

This report and all the analyeses are fully reproducible. This is the link to my GitHub repositorty:

https://github.com/ArtBert96/response

The repository is organized with the IPO protocol. By running on Rstudio all the syntax file in the "Processing" folder, the machine uses the data in "Input" to produce all graphs and images in "Output". In the section below I only add descriptives and summary information about the factor analysis.

Factor analysis

Eigen values of Principal Component Analysis performed on items of the same battery.

Index	Eigenvalue comp.1	Eigenvalue comp.2
Compliance to preventive behaviors	1.6	0.7
Willingness to restrict freedoms	1.3	0.5
Institutional trust	1.3	0.6
Risk perception	1.3	0.7
Collectivism	1.3	0.6
Support for regional system	1.3	0.7
Attitudes toward science	1.4	0.8
Attitudes toward medicine	1.4	0.8

Cronbach's Alpha

Index	Cronbach's Alpha
Compliance to preventive behaviors	0.8
Willingness to restrict freedoms	0.9
Institutional trust	0.8
Risk perception	0.7
Collectivism	0.8
Support for regional system	0.7
Attitudes toward science	0.7
Attitudes toward medicine	0.8

Descriptives

Statistic	N	Mean	St. Dev.	Min	Max
Support for Conte	11,677	0.7	0.5	0	1
L - R selfplacement	11,677	0.4	0.5	0	1
Compliance to preventive behaviors	11,677	0.9	0.2	0	1
Willingness to restrict freedoms	11,677	0.7	0.4	0	1
Institutional trust	11,677	0.5	0.5	0	1
Support for regional system	11,677	0.7	0.5	0	1
Risk perception	11,677	0.2	0.4	0	1

Figure 8: CAN database, descriptives

Statistic	N	Mean	St. Dev.	Min	Max
Support for Conte	11,677	6.2	2.6	0	10
L - R selfplacement	11,677	4.6	3.0	0	10
Compliance to preventive behaviors	11,677	8.5	1.7	0.0	10.0
Willingness to restrict freedoms	11,677	6.6	2.7	0.0	10.0
Institutional trust	11,677	4.6	2.4	0.0	10.0
Support for regional system	11,677	6.1	2.2	0.0	10.0
Risk perception	11,677	1.6	0.9	0.0	4.0

Figure 9: *mgm* database, descriptives

Statistic	N	Mean	St. Dev.	Min	Max
Sex	10,070	0.6	0.5	0	1
Age	10,070	49.1	15.6	18	90
Family income	10,070	0.5	0.5	0	1
Approval governmental measures	10,070	6.6	2.4	0	10
Approval restrictive measures	10,070	0.2	0.4	0	1
Economy over public health	10,070	4.6	2.2	0	10
Horizontal trust	10,070	3.5	2.3	0	10
Support for Conte	10,070	6.3	2.6	0	10
Happiness	10,070	6.1	1.9	0	10
Health	10,070	0.4	0.5	0	1
Anxiety	10,070	0.7	0.5	0	1
Depression	10,070	0.5	0.5	0	1
L-R self-placement	10,070	4.6	3.0	0	10
Political interest	10,070	0.3	0.5	0	1
Compliance to preventive behaviors*	10,070	8.6	1.7	0.0	10.0
Willingness to restrict freedoms*	10,070	6.6	2.6	0.0	10.0
Institutional trust*	10,070	4.6	2.5	0.0	10.0
Risk perception*	10,070	1.6	0.9	0.0	4.0
Collectivism*	10,070	6.0	2.3	0.0	10.0
Support for regional system*	10,070	6.1	2.2	0.0	10.0
Attitudes toward science*	10,070	7.0	1.8	0.0	10.0
Attitudes toward medicine*	10,070	5.5	2.5	0.0	10.0

Figure 10: Extendend *mgm* database, descriptives