

Reporte de trabajo: Análisis de Redes con datos del Tamizaje para la detección de riesgos a la Salud Mental por Covid-19

Por Elena Villalobos Nolasco

El presente documento tiene los avances realizados en el proyecto de Análisis de Redes como alternativa de análisis para estudiar psicopatologías y tiene como base principal el artículo: *Estimating psychological networks and their addequacy: A tutorial paper. (2018) Epskamp, S., Borsboom, D. & Fried, E..*

El documento se organiza de la siguiente manera:

1. Conceptos básicos para entender el análisis de redes.
2. Descripción de cómo leer los gráficos principales para estudiar redes.
3. Resultados preliminares

1. Conceptos básicos para entender el análisis de redes

El análisis de redes, basado en la teoría de grafos, tiene como objetivo estudiar las interrelaciones existentes entre entidades. En el análisis de redes los nodos representan entidades (aeropuestos, personas, etc.), y las conexiones, también conocidas como aristas, son observadas, medidas y conocidas (número de vuelos entre aeropuestos, amistades, etc.).

De manera similar pero estructuralmente diferente al análisis de redes, las redes psicológicas consisten en nodos que representan las variables observadas que están interconectados por aristas que representan relaciones estadísticas. Otra manera de decirlo, es que las conexiones entre los nodos son parámetros que se estiman a partir de los datos, por lo que una red mejor estimada será aquella que tiene más datos.

El análisis de redes psicológicas tiene dos pasos involucrados:

1. Estimar el modelo estadístico sobre los datos, es decir, la estimación de la red con sus pesos correspondientes entre las variables observadas.
2. Analizar si las estimaciones de los pesos son adecuadas, a partir de medidas desarrolladas en teoría de grafos.

Ellos simularon una red con un modelo gráfico gaussiano, donde los nodos representan las variables observadas y las aristas representan coeficientes de correlación parcial entre variables, condicionadas sobre las otras variables observadas. Una manera de evaluar la importancia de nodos en esta red es computando las medidas de centralidad de la estructura de la red, estas son:

- Fuerza del nodo: que tan bien el nodo es **directamente** conectado con otros nodos.
- Closeness: cuantificar que tan bien el nodo es indirectamente conectado a otros nodos.
- Betweeness: que tan importante es un nodo en el camino promedio entre dos nodos.

Después, a partir del modelo antes descrito, simularon datos de 500 sujetos y salió que de hecho la red se parece, sin embargo, las medidas de centralidad difieren mucho. Por lo que se debe de tener cuidado en este tipo de evaluaciones.

El artículo propone una metodología para evaluar la precisión de la estructuras de redes psicológicas que va en tres pasos:

1. Estimar los intervalos de confianza sobre los pesos de las aristas.
2. Evaluar la estabilidad de los índices de centralidad observados sobre subconjuntos de datos.
3. Test de diferencias significativas entre los pesos de las aristas y los índices de centralidad.

Un modelo de redes que se usa popularmente para estimar redes psicológicas es un Campo Aleatorio de Markov por Pares (Pairwise Markov Random Field, PMRF). En este modelo, los nodos representan variables, y se conectan a través de aristas no-direccionadas, indicando dependencia condicional entre dos variables; por otro lado si dos variables que no están conectadas, son independientes después de condicionar sobre las otras variables. Cuando los datos son normales multivariados, esta independencia condicional correspondería a una correlación parcial igual a cero.

Cuando los datos son binarios, el modelo PMRF para usar es el modelo Ising. Cuando los datos siguen una densidad normal multivariada, el modelo apropiado de PMRF es el modelo gráfico gaussiano (GGM), en los que las aristas pueden ser directamente interpretadas como coeficientes de correlación parcial. El GGM requiere un estimado de la matriz de covarianza como input, que para correlaciones policóricas (correlación para variables ordinales de variables latentes) pueden ser usadas en caso de que los datos sean ordinales.

El modelo PMRF tiene el problema de que el número de parámetros a estimar crece mucho con el tamaño de la red y en psicología no se tienen tantos datos que compensen dicha estimación. Para tratar dicho problema se utiliza la forma de regularización de LASSO (Least Absolute Shrinkage and Selection Operator). Dicha técnica penaliza el uso de muchos parámetros, limitando la suma de valores paramétricos absolutos. Por lo que LASSO regresa un modelo de red más conservador, es decir sólo un pequeño número de aristas se utilizan para explicar la covariación de la estructura en los datos. El paquete de qgraph utiliza glasso en combinación con la selección de modelos EBIC para estimar un GGM regularizado.

Después de hacer la estimación de la red se sugiere seguir estos métodos para la evaluación de la precisión de la red:

- a. Estimación de la precisión de los pesos de las aristas, utilizando Intervalos de Confianza con bootstrap.
- b. Investigar la estabilidad de los índices de centralidad, con subconjuntos de datos.
- c. Realizar tests de diferencias con bootstrapp entre los pesos de las aristas y los índices de centralidad para evaluar si estas diferencias son significativas entre ellas.

Precisión de los pesos de las aristas.

Para evaluar la variabilidad de los pesos de las aristas se pueden estimar Intervalos de Confianza. Para construirlos necesitamos saber la distribución muestral del estadístico de interés, pero saberlo para estadísticos como medidas de centralidad puede ser difícil. Por lo que utilizaremos bootstrap, técnica que implica estimar repetidamente un modelo con datos muestreados o simulados y obtener la estadística de interés. Se puede hacer bootstrap de dos maneras, paramétrico y no paramétrico. En el no paramétrico, las observaciones de los datos se remuestran con reemplazo para crear un nuevo conjunto de datos, mientras que en el paramétrico muestrea nuevas observaciones del modelo paramétrico que fue estimado de los datos originales, lo que crea una serie de valores que pueden ser utilizados para estimar la distribución muestral. Con esto se aconseja utilizar bootstrap no paramétrico para datos ordinales. También importante mencionar que los resultados del bootstrap no deberían ser utilizados para hacer test de significancia diferente de cero, pues LASSO ayuda a quitar los pesos que no son importantes en la red, por lo que los pesos que ya aparecen dentro de la red, pueden ser significativamente diferentes de cero, es decir, son importantes.

Por lo que la IC con bootstrap de los pesos de las aristas no deberían ser interpretados como diferentes de cero, si no que sólo muestra la precisión de los pesos y comparar las aristas entre ellas. Cuando los ICb, son anchos, es difícil interpretar la fuerza de una aristas. Así mismo LASSO quita los pesos que son positivos y negativos, por lo que, toma el signo de la red como algo *real*.

Como las medidas de centralidad están en función directa de los pesos, los intervalos largos de los pesos, probablemente serán resultado de poca precisión en los índices de centralidad.

Estabilidad de centralidad

Para saber la estabilidad de las medidas de centralidad no se puede utilizar bootstrap, porque genera distribuciones muestrales sesgadas, por lo que proponen estudiar dichas medidas con subconjuntos de los datos. Se dice que los índices de centralidad son estables cuando el orden de los índices es el mismo, en diferentes submuestras del mismo conjunto de datos, o con diferentes nodos (no tan usado). Aplicar esta técnica de remuestreo (el bootstrap regular) para diversas proporciones (ya sean participantes o variables), puede ser utilizado para evaluar la correlación entre los índices de centralidad originales y aquellos obtenidos en las submuestras. Si la correlación cambia completamente después de quitar, digamos 10% de la muestra, entonces las interpretaciones de centralidad pueden ser erróneas. A esto se le llama bootstrap de subconjuntos con case-dropping. Para cuantificar dicha estabilidad se propone la medida de coeficiente de estabilidad de correlación (CS-coefficient) que se recomienda sea mayor a 0.7.

Test para diferencias significativas.

Este test también se hace con bootstrap para saber si una arista es significativamente distinta de otra o si las medidas de centralidad son significativamente distintas de otras. Pero como en todo test de significancia, debes de tener cuidado en la interpretación porque no rechazar la hipótesis nula no es evidencia de que esta es verdadera.

Interpretación del gráfico:

- a. Cada línea horizontal representa una arista de la red, ordenadas desde la arista con el peso más alto, hasta la arista con el peso más bajo. Es decir, de la red que sacaste, las conexiones más fuertes son entre el 17 y el 16, el 3 y el 4, el 11 y el 5, y estas conexiones son las que están hasta arriba del gráfico de intervalos de confianza de los pesos de las aristas. De hecho, son de manera confiable los tres bordes más fuertes, ya que sus Intervalos de Confianza, no se superponen con los CI de ningún otro borde.
- b. Entre menos caiga la línea, implica más estabilidad a través de las diferentes cantidades de muestra.

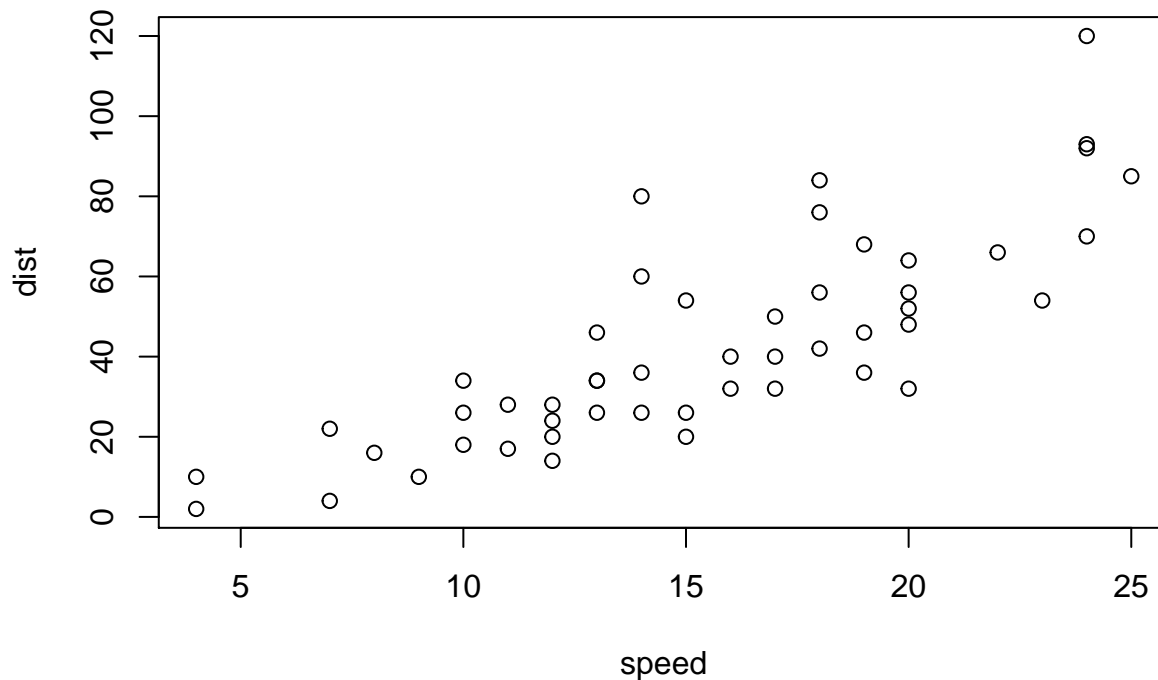
Precisión de la red (Network accuracy)

1. Estimar la precisión de los pesos de las aristas con los IC creados con Bootstrap (bootstrapped CI). Esto se hace con bootstrap no paramétrico.
2. Investigar estabilidad de la centralidad, donde se indica si el orde de los índices centrales permanece igual después de re-estimar la red con menos casos o nodos. Esto se hace un subsets de tu base de datos. (case-dropping subset bootstrap) para la que se tiene la medida de CS-coefficient = correlation stability coefficient.
3. Evaluar diferencias significativas (bootstrapped difference test).

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
plot(cars)
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.