

# sidoNet: Síndrome de Desgaste Ocupacional Docente modelado con grafos

Ana María Ruiz-Ruano García, Javier Rodríguez Fragoso, Francisco José Moya-Faz, Enrique Javier Garcés de los Fayos Ruiz, y Jorge López Puga

## Introducción

Este proyecto contiene los archivos de datos, el código y las figuras generadas para el artículo titulado xxxxxxxxxxxx xxxxxxxx xxxxxx y que ha sido remitido a la revista XXXXXX XXXXX para su revisión por pares. Contiene la misma información que la que está contenida en el archivo README.pdf.

Este proyecto se encuentra alojado en la *Open Science Framework* (OSF) en la web <https://osf.io/4cbsz/> y en GitHub en la web <https://github.com/dr46/sidoNet>. **Advertencia:** nótese que los hipervínculos que contiene este documento y el archivo README.pdf están referidos a la versión del proyecto disponible en GitHub. Por tanto, para navegar por el conjunto de archivos contenidos en este proyecto se recomienda utilizar la versión web del proyecto disponible en GitHub. **Se hace notar** también que para generar los archivos .csv que contienen tablas se ha usado tanto el punto y coma (habitual en algunas hojas de cálculo configuradas para la lengua española con el objetivo de que la separación de columnas no se confunda con el delimitador decimal) como la coma (más común en entornos en los que se usa el punto como delimitador decimal). Por ello, es posible que algunas tablas se vean en formato de texto plano en GitHub mientras que deberían verse en formato tabulado en la OSF. También se hace notar que tildes del español pueden no verse correctamente en la web de la OSF. Sin embargo, las tildes son legibles perfectamente en las páginas de GitHub.

## Carpetas o directorios del proyecto

Este proyecto contiene un conjunto de carpetas en las que se organizan los archivos. Más concretamente, las carpetas del proyecto y la descripción de la información que contienen es esta:

- dat: esta carpeta contiene archivos de datos.
- stx: esta carpeta contiene archivos de código.
- fig: esta carpeta contiene las figuras generadas por el proyecto.
- tab: esta carpeta contiene tablas de resultados.
- fun: esta carpeta contiene funciones de código R que se invocan en diferentes archivos.
- par: esta carpeta contiene archivos que contienen algunos de los parámetros fijos que se utilizarán para estimar los grafos.

## Archivos

### Archivos de datos

- cuerpo.csv: esta tabla contiene una única columna en la que se indica a qué cuerpo laboral pertenecía

la población ( $N = 20766$ ) de estudio en el momento en que se estimó el tamaño muestral necesario para llevar a cabo la investigación.

- demo.csv: esta tabla contiene información sociodemográfica de la muestra de participantes. Las variables que incluye, así como los posibles valores que pueden asumir, son las siguientes:
  - id: código de identificación del participante.
  - sex: indica si el participante es mujer (0) u hombre (1).
  - ec: indica el estado civil del participante. Sus posibles categorías son:
    - \* Casado/casada.
    - \* Convivencia en pareja.
    - \* Divorciado/divorciada.
    - \* Pareja estable.
    - \* Soltero/soltera.
    - \* Viudo/viuda.
  - nh: indica el número de hijos del participante.
  - ne: indica el nivel de estudios del participante. Los valores posibles que asume esta variable son los siguientes.
    - \* Diplomatura.
    - \* Doctorado.
    - \* Formación profesional.
    - \* Grado.
    - \* Licenciatura.
  - esp: especialidad del nivel de estudios. Esta variable asume los siguientes valores con sus correspondientes significados:
    - \* 1: Magisterio.
    - \* 2: Filosofía y Letras.
    - \* 3: Psicología y logopedia.
    - \* 4: Filología Hispánica.
    - \* 5: Psicología.
    - \* 6: Geografía e Historia.
    - \* 7: Educación Infantil.
    - \* 8: Inglés.
    - \* 9: Matemáticas y ciencias.
    - \* 10: Traducción e interpretación.
    - \* 11: Física.
    - \* 12: Educación Física.
    - \* 13: Química.
    - \* 14: Bellas Artes.
    - \* 15: Francés.
    - \* 16: Administración y finanzas.
    - \* 17: Ingeniería.
    - \* 18: Biología.
    - \* 19: Pedagogía.
    - \* 20: Música.
  - mp: indica la respuesta a la pregunta *¿Ha cursado el título de Máster del Profesorado?* cuyas posibles respuestas son:
    - \* 0: No.
    - \* 1: Sí.
  - rol: indica el rol que desempeña el participante en su centro de trabajo y asume los siguientes posibles valores.
    - \* Dirección.
    - \* Educación Infantil.

- \* Educación Primaria.
- \* Educación Secundaria.
- \* Jefatura de Estudios.
- \* Orientación.
- \* Otro.
- \* Secretaría.
- inf: indica los ingresos netos mensuales en el núcleo familiar.
- cue: cuerpo docente al que pertenece el participante. Las categorías de respuesta presentes en esta variable son las siguientes:
  - \* Cuerpo de Catedráticos de Enseñanza Secundaria.
  - \* Cuerpo de Maestros.
  - \* Cuerpo de Profesores de Enseñanza Secundaria.
  - \* Cuerpo de Profesores Técnicos de Formación.
  - \* Otro.
  - \* Profesores de Religión (Primaria).
  - \* Profesores de Religión (Secundaria).
  - \* Profesores Especialistas.
- d21.csv: esta tabla contiene las respuestas a la escala abreviada de depresión, ansiedad y estrés (DASS-21) de Lovibond y Lovibond (1995). Este archivo tiene cuatro conjuntos de variables. El significado de las variables es el siguiente:

- id: es el identificador de cada uno de los casos del estudio.
- de e1 a e7: se refieren a los ítems de estrés del DASS-21.
- de a1 a a7: se refieren a los ítems de ansiedad del DASS-21.
- de d1 a d7: se refieren a los ítems de depresión del DASS-21.

El significado de cada una de las respuestas a estos ítems es el siguiente:

- 0: no aplicable a mí.
- 1: un poco aplicable a mí o durante parte del tiempo.
- 2: bastante aplicable a mí, o durante una buena parte del tiempo.
- 3: muy aplicable a mí, o la mayor parte del tiempo.
- cesqt.csv: esta tabla contiene las respuestas a la escala para la evaluación del síndrome de quemarse por el trabajo (CESQT-PE) de Gil-Monte et al. (2009). Esta matriz de datos contiene 21 columnas que se agrupan en cinco bloques. El significado de cada una de las columnas es el siguiente:
  - id: es el identificador de cada uno de los casos del estudio.
  - de il1 a il5: se refieren a los ítems de la dimensión de ilusión por el trabajo del CESQT-PE.
  - de in1 a in6: se refieren a los ítems de la dimensión de indolencia del CESQT-PE.
  - de cu1 a cu5: se refieren a los ítems de la dimensión de culpa del CESQT-PE.
  - de de1 a de4: se refieren a los ítems de desgaste psíquico del CESQT-PE.

Las opciones de respuesta para cada una de estas variables oscilan entre 0 y 4 y su significado es el siguiente:

- 0: nunca.
- 1: raramente: algunas veces al año.
- 2: a veces: algunas veces al mes.
- 3: frecuentemente: algunas veces por semana.
- 4: muy frecuentemente: todos los días.
- co.csv: esta tabla contiene la pertenencia a los conglomerados para cada participante del estudio. Contiene dos columnas cuyos significados son:

- id: es el código de identificación asignado a cada participante del estudio.
  - co: indica el conglomerado a que pertenece cada uno de los participantes.
- sido.csv: esta tabla es una versión actualizada de la base de datos demo.csv pero incluyendo la pertenencia a los conglomerados encontrados en el estudio, la puntuación total de la escala DASS-21, la puntuación total de la escala CESQT-PE así como las puntuaciones en las subescalas de susodichos tests. El significado de las columnas añadidas en esta base de datos respecto a la tabla demo.csv es el siguiente.
    - co: indica el conglomerado al que pertenece el participante.
    - ilu: puntuación de la subescala de ilusión por el trabajo de la escala CESQT-PE.
    - ind: puntuación de la subescala de indolencia de la escala CESQT-PE.
    - cul: puntuación de la subescala de culpa de la escala CESQT-PE.
    - des: puntuación de la subescala de desgaste psíquico de la escala CESQT-PE.
    - do: puntuación total de desgaste ocupacional al corregir la escala CESQT-PE.
    - est: puntuación de la subescala de estrés de la escala DASS-21.
    - ans: puntuación de la subescala de ansiedad de la escala DASS-21.
    - dep: puntuación de la subescala de depresión de la escala DASS-21.
    - dass: puntuación total de la escala DASS-21.

## Archivos de código

- sidoNet.R: es el archivo de código maestro del proyecto. Desde este archivo se van ejecutando el resto de los archivos de código del proyecto.
- 01\_muestra.R: este archivo realiza estimaciones del tamaño muestral teniendo en cuenta los posibles conglomerados de docentes latentes y teniendo en cuenta el tamaño de los estratos poblacionales estudiados. Genera las tablas muestra.conglomerados.csv y muestra.estratificada.csv.
- 02\_demo.R: este archivo ejecuta el análisis de las características sociodemográficas de la muestra del estudio. Este archivo genera las tablas t.frec.csv y edin.csv que son simultáneamente proyectadas en la consola de R. Este archivo también produce las figuras edad.pdf y imf.pdf.
- 03\_fiabil.R: este archivo estima los coeficientes de fiabilidad ( $\omega$  de McDonald y  $\alpha$  de Cronbach) y sus intervalos de confianza utilizando técnicas de remuestreo aleatorio (1000 muestras en todos los casos) de los datos muestrales. Este archivo genera la tabla fiabil.csv que se almacena en la carpeta *tab*. **Atención:** ejecutar el archivo de código 03\_fiabil.R puede consumir una gran cantidad de recursos y de tiempo. El tiempo empleado para ejecutar las computaciones dependerá de los recursos del ordenador en que se ejecute el comando correspondiente. Es por ello por lo que la línea de código que desencadena los análisis de fiabilidad ha sido comentada en el archivo maestro del proyecto (sidoNet.R). En su lugar, se ha optado por introducir la línea de código que permite recuperar los resultados (fiabil.csv) que produce dicho archivo y que están almacenados en la carpeta *tab*. **Se hace notar**, también, que para ejecutar este archivo es necesario cargar el paquete “MBESS” (Kelley, 2021). Si el paquete no está instalado, se producirán errores y la computación se interrumpirá.
- 04\_cluster.R: este archivo estima el número de conglomerados que hay en los datos y el vector que asocia a cada participante con cada uno de los conglomerados. Este archivo de código usa el paquete “NbClust” (Charrad et al., 2014). Si el paquete no está instalado, se producirán errores y la computación se interrumpirá. Este archivo también produce los gráficos contenidos en el archivo clust.pdf e imprime el resumen del análisis (en inglés) así como el tamaño de los conglomerados encontrados en la consola de R. Por último, este archivo de código genera la tabla de datos co.csv que se almacena en la carpeta *dat* y que contiene la pertenencia a cada conglomerado para cada participante.
- 05\_escon.R: este archivo estudia las diferencias que se observan en las variables registradas en el estudio en función de la pertenencia a los clusters observados en los datos. Para que funcione correctamente

deben estar instalados los paquetes “effectsize” (Ben-Shachar et al., 2020) y “rcompanion” (Mangiafico, 2024) porque son requeridos al ejecutar el archivo de código. Este archivo de código genera la base de datos intermedia *sido.csv* que se almacena en la carpeta *dat*. También produce las tablas *comp.csv* y *tj.csv* que se almacenan en la carpeta *tab*. Y, por último, este archivo de código produce el archivo *puntos.pdf* que contiene gráficos para comparar cómo se comportan los clusters identificados en los datos en las escalas y subescalas utilizadas en el estudio.

- *06\_rnd.R*: este archivo de código genera y analiza las redes de síntomas para cada una de las dimensiones de la escala DASS-21 de Lovibond y Lovibond (1995). Para ejecutar esta línea de código es necesario tener instalados los paquetes “qgraph” (Epskamp et al, 2012) y “igraph” (Kolaczyk y Csárdi, 2014). Ese archivo estima las redes que se representan en las figuras del archivo *rnd.pdf*. También genera un conjunto de tablas que contienen los estadísticos locales (relativos a cada uno de los nodos de la red) para cada una de las redes estimadas (*al.1*, *al.2*, *el.1*, *el.2*, *dl.1*, y *dl.2*) así como la tabla *est.g.csv* que contiene estadísticos globales para cada una de las redes.

## Tablas

- *exclusion.xlsx*: esta tabla contiene cinco columnas destinadas a clarificar el proceso de toma de decisiones para incluir (o no) los perfiles docentes correspondientes en el estudio. El significado de las columnas de la tabla es el siguiente:
  - *ex*: indica si el perfil docente se excluye del estudio.
  - *mex*: indica el motivo de exclusión del perfil docente.
  - *tpe*: indica el tamaño poblacional del estrato (véase el archivo *cuerpo.csv* y su descripción correspondiente).
  - *cue*: cuerpo docente.
  - *meap*: muestra estimada para el estrato considerando un muestreo estratificado con afijación proporcional al cuerpo docente (León y Montero, 2003).
- *muestra.conglomerados.csv*: esta tabla contiene información sobre la estimación del tamaño muestral basada en el análisis previo de conglomerados latentes posibles. En todos los casos se considera una potencia  $(1 - \beta)$  del 80%, una variabilidad entre grupos reducida (1) y una variabilidad entre-grupos elevada (6) para obtener estimaciones de tamaños muestrales lo más conservadoras posible ( $\alpha = 0.05$ , ver archivo *01\_muestra.R*). Contiene tres columnas y el significado de cada una de ellas es el siguiente:
  - *conglomerados.latentes*: indica el número posible de conglomerados ( $k$ ) latentes de personal docente en la muestra.
  - *tamanyo.conglomerado*: indica el tamaño muestral ( $n_k$ ) estimado por conglomerado latente.
  - *muestra.total.estimada*: indica el tamaño muestral estimado para el estudio considerando el tamaño muestral del conglomerado latente y el número de conglomerados considerados ( $n_k \times k$ ).
- *muestra.estratificada.csv*: esta tabla contiene cinco columnas que informan sobre el tamaño muestral estimado teniendo en cuenta un muestreo estratificado con afijación proporcional para los cuerpos docentes considerados en el estudio. El significado de cada una de las columnas es el siguiente:
  - *cuer*: cuerpo docente considerado.
  - *Freq*: número de personas en la población que pertenecen al cuerpo docente considerado.
  - *as*: estimación del tamaño muestral para el cuerpo docente considerado considerando el criterio de afijación proporcional simple.
  - *pp*: indica la proporción que suponen el número de personas incluida en el estrato considerado.
  - *me*: muestra estimada para el estrato considerado teniendo en cuenta un muestreo con afijación proporcional por estratos.

- t.frec.csv: esta tabla contiene las frecuencias absolutas y relativas (como porcentajes) para las variables relativas al sexo, al estado civil, al número de hijos al nivel de estudios y al rol que asumen los participantes en su centro educativo. La tabla tiene tres columnas cuyos significados son:
  - fa: frecuencia absoluta.
  - fr: porcentaje.
- edin.csv: esta tabla contiene estadísticos descriptivos para la edad y para los ingresos mensuales familiares (imf) del participante. Las filas de la tabla contienen la siguiente información:
  - 0%: valor mínimo de la variable.
  - 25%: percentil 25 de la variable.
  - 50%: percentil 50 de la variable.
  - 75%: percentil 75 de la variable.
  - 100%: valor máximo de la variable.
  - M: media de la variable.
  - DT: desviación típica de la variable.
- fiabil.csv: esta tabla contiene las estimaciones de fiabilidad para las escalas DASS-21 y CESQT así como para sus correspondientes subescalas. El significado de las columnas que contiene esta tabla es el siguiente:
  - om: estimación puntual del coeficiente de consistencia interna  $\omega$  de McDonald.
  - et.om: estimación del error típico asociado al coeficiente de consistencia interna  $\omega$  de McDonald.
  - li.om: estimación del límite inferior del intervalo de confianza al 95% para el coeficiente de consistencia interna  $\omega$  de McDonald.
  - ls.om: estimación del límite superior del intervalo de confianza al 95% para el coeficiente de consistencia interna  $\omega$  de McDonald.
  - al: estimación puntual del coeficiente de consistencia interna  $\alpha$  de Cronbach.
  - et.al: estimación del error típico asociado al coeficiente de consistencia interna  $\alpha$  de Cronbach.
  - li.al: estimación del límite inferior del intervalo de confianza al 95% para el coeficiente de consistencia interna  $\alpha$  de Cronbach.
  - ls.al: estimación del límite superior del intervalo de confianza al 95% para el coeficiente de consistencia interna  $\alpha$  de Cronbach.

Por su parte, el significado de las filas es el siguiente:

- dass.21: escala DASS-21.
  - dass.21.e: subescala de estrés de la escala DASS-21.
  - dass.21.a: subescala de ansiedad de la escala DASS-21.
  - dass.21.d: subescala de depresión de la escala DASS-21.
  - cesqt: escala CESQT-PA.
  - cesqt.il: subescala de ilusión por el trabajo de la escala CESQT-PA.
  - cesqt.in: subescala de indolencia de la escala CESQT-PA.
  - cesqt.cu: subescala de culpa de la escala CESQT-PA.
  - cesqt.de: subescala de desgaste psicológico de la escala CESQT-PA.
- comp.csv: esta tabla compara los conglomerados encontrados en las variables cuantitativas del estudio. El significado de sus columnas es el siguiente:
  - m.1: media del cluster número 1.
  - m.2: media del cluster número 2.
  - dt.1: desviación típica del cluster número 1.
  - dt.2: desviación típica del cluster número 2.
  - t: valor de la  $t$  de Student para la comparación de medias.
  - df: grados de libertad para la  $t$  de Student.

- p:  $p$ -valor de la  $t$  de Student.
- d:  $d$  de Cohen para la estimación del tamaño del efecto.
- r: estimación del tamaño del efecto basado en  $r$ .

Por su parte, el significado de las filas es el siguiente:

- CESQT-PA: puntuación total de la escala CESQT-PA.
  - Ilu. Traba.: puntuación de la subescala de ilusión por el trabajo de la escala CESQT-PA.
  - Indolencia: puntuación de la subescala de indolencia de la escala CESQT-PA.
  - Culpa: puntuación de la subescala de culpa de la escala CESQT-PA.
  - Desgas. Psi.: puntuación de la subescala de desgaste psíquico de la escala CESQT-PA.
  - DASS-21: puntuación total de la escala DASS-21.
  - Estr.: puntuación de la subescala de estrés de la escala DASS-21.
  - Ansiedad: puntuación de la subescala de ansiedad de la escala DASS-21.
  - Depre.: puntuación de la subescala de depresión de la escala DASS-21.
  - Edad: edad.
  - Hijos: número de hijos.
  - Ingresos: ingresos mensuales del núcleo familiar.
- tj.csv: esta tabla contiene el análisis de relación que se establece entre las variables cualitativas y los conglomerados observados en los datos. El significado de las columnas es el siguiente:
    - ji.2: es el estadístico  $\chi^2$  de Pearson.
    - p: es el  $p$ -valor del contraste de hipótesis para la  $\chi^2$  de Pearson.
    - V: es la  $V$  de Cramer.
  - al.1, al.2, el.1, el.2, dl.1, y dl.2: estas tablas contienen, respectivamente, los estadísticos locales para cada uno de los nodos contenidos en las redes de ansiedad (conglomerados 1 y 2), de estrés (conglomerados 1 y 2), y de depresión (conglomerados 1 y 2). El significado de las filas es el siguiente:
    - GP: grado ponderado del nodo.
    - G: grado no ponderado del nodo.
    - C: índice de cercanía del nodo.
    - I: índice de intermediación del nodo.
    - IG: influencia esperada del nodo.
    - WS: coeficiente de conglomeración de Watts y Strogatz (1998).

## Figuras

- edad.pdf: es el histograma de la edad.
- imf.pdf: es el histograma de la los ingresos medios familiares de los participantes.
- clust.pdf: este archivo contiene las figuras utilizadas para determinar el número de clusters en la muestra usando el método de Hubert (primera página del documento) y el índice D (segunda página del documento).
- puntos.pdf: este archivo contiene gráficos de puntos para los clusters encontrados y para cada una de las escalas y subescalas utilizadas en este estudio.
- rnd.pdf: este archivo contiene los grafos no dirigidos para cada una de las dimensiones del DASS-21 en función de cada uno de los conglomerados extraídos de los datos. El grosor de los enlaces es proporcional a la fuerza de asociación estimada para cada par de variables.

## Funciones

- 01\_fval.R: esta función sirve para extraer estadísticos de los análisis de fiabilidad.
- 02\_compi.R: esta función sirve para realizar comparaciones descriptivas e inferenciales entre los clusters encontrados en los datos.
- 03\_gc2.R: esta función sirve para generar gráficos de puntos para las puntuaciones de las escalas utilizadas en el estudio (CESQT-PA y DASS-21) y sus subescalas en función del cluster al que pertenece el participante.
- 04\_rn.R: esta función extrae estadísticos del test de  $\chi^2$  de Pearson y de la estimación de la  $V$  de Cramer.

## Archivos de parámetros

- par.rnd.R: este archivo contiene los parámetros fijos de estimación para las redes no dirigidas.

## Detección de casos atípicos y su tratamiento

Se detectaron casos con valores atípicos que fueron tratados consecuentemente en la variable *imf* (ingresos familiares mensuales, ver tabla 02\_demo.R). En concreto, estas fueron las modificaciones que se llevaron a cabo:

- El valor de la variable *imf* era 300 para el caso con id 102 pero se trató como caso perdido por presentar un valor considerablemente bajo y por considerarse un error tipográfico.
- El valor de la variable *imf* era 2 para el caso con id 117 pero se trató como caso perdido por presentar un valor considerablemente bajo y por considerarse un error tipográfico.
- El valor de la variable *imf* era 0 para el caso con id 205 pero se trató como caso perdido por presentar un valor considerablemente bajo y por considerarse un error tipográfico.

## Referencias

- Ben-Shachar, M. S., Lüdtke, D., y Makowski, D. (2020). Effectsize: estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56). Artículo 2815. <https://doi.org/10.21105/joss.02815>
- Charrad, M., Ghazzali, N., Boiteau, V., y Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6), 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., y Borsboom, D. (2012). qgraph: network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4), 1–18. <https://doi.org/10.18637/jss.v048.i04>
- Gil-Monte, P. R., Unda Rojas, S., y Sandoval Ocaña, J. I. (2009). Validez factorial del “Cuestionario para la Evaluación del Síndrome de Quemarse por el Trabajo” (CESQT) en una muestra de maestros mexicanos. *Salud Mental*, 32(3), 205–214.
- Kolaczyk, E. D., y Csárdi, G. (2014). *Statistical analysis of network data with R*. Springer. <https://doi.org/10.1007/978-3-030-44129-6>
- León, O. G., y Montero, I. (2003). *Métodos de investigación en psicología y educación*. McGraw Hill.



- Lovibond, P. F., y Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3), 335-343. [https://doi.org/10.1016/0005-7967\(94\)00075-U](https://doi.org/10.1016/0005-7967(94)00075-U)
- Watts, D., y Strogatz, S. (1998, 4 de Junio). Collective dynamics of ‘small-world’ networks. *Nature* 393(6684), 440–442. <https://doi.org/10.1038/30918>