

DETECCIÓN DE COMUNIDADES Y FILTRADO DE REDES SOBRE DATOS EN SALUD

ANA VIGNOLO ^a, RODRIGO VITABAR ^a, RAMÓN ÁLVAREZ-VAZ ^a

^aInstituto de Estadística, Facultad de Ciencias Económicas y Administración,
Universidad de la República, Montevideo, Uruguay, Gonzalo Ramírez 1926, Piso 1 Of. 20. CP: 11200
e-mail: ramon.alvarez@fcea.edu.uy, ana.vignolo@fcea.edu.uy

El presente trabajo surge de los datos generados para el estudio de las Enfermedades No Transmisibles (ENT) para pacientes que se atendieron entre 2015 y 2016 en la Facultad de Odontología de la Universidad de la República en Montevideo, Uruguay a través de una muestra probabilística de pacientes obtenida mediante un diseño sistemático, para el que parecen primeros resultados en el capítulo 7 del libro Actualidad en Probabilidad y Estadística [1]. Del estudio original surge este nuevo trabajo, que consiste en el uso de análisis de redes sobre los datos, con los cuales se construye una red formada por los pacientes, los cuales se toman como nodos y se considera que dos individuos están conectados si comparten al menos un factor de riesgo o patología, que aparecen codificados como 11 variables binarias. Posteriormente, se considera un umbral más exigente para definir que dos individuos se encuentran conectados por un enlace, mediante la aplicación de una función de filtrado. Con este criterio *ad hoc*, se considera la red formada por pacientes que comparten al menos cinco patologías. Más allá de que esto permite descartar asociaciones "débiles" y facilita la visualización de la red, es interesante en la medida que a través de diferentes algoritmos de detección de comunidades se logran detectar grupos de pacientes con diferentes perfiles epidemiológicos, en base a la prevalencia local (en los clústeres) de las variables que se usan para formar la red y de otros atributos como sexo y edad .

Keywords: Análisis de Redes, Comunidades, Filtrado, ENT (Enfermedades No Transmisibles)

1. Introducción

Las enfermedades no transmisibles ENT, en las que pueden agruparse enfermedades como las cardiovasculares, diabetes, cáncer y enfermedades respiratorias crónicas, son actualmente la causa de mortalidad a nivel mundial más importante. En el ámbito de los estudios epidemiológicos existen situaciones donde se manejan ENT, en particular en salud bucal, donde los tipos de variables que se manejan son binarias y pueden ser intercambiables en cuanto a quien hace el rol de factor de riesgo, [1]. En este tipo de estudios epidemiológico se suele indaga por las ENT y es habitual trabajar con variables binarias que reflejan la presencia de

determinadas enfermedades, las que a su vez se asocian con otro conjunto de enfermedades, que se conocen como comorbilidades, que también muy frecuentemente se medidas también a través de variables binarias, [1].

2. Objetivos

Los objetivos de este trabajo tal como se plantea en el resumen y como extensión del trabajo origina se plantean del siguiente modo:

- Obtener redes que no sean extremadamente densas mediante funciones de filtrados.
- Determinar perfiles epidemiológicos bien diferenciados en base a las variables de salud

y de factores de y crear grupos mediante estrategias de detección de comunidades.

- Estudiar y comparar la topología de la red filtradas, así como la prevalencia de las patologías y de los factores de riesgo en las comunidades detectadas.

3. Metodología

3.1. Métricas usadas en las descripción de las redes. Para medir cuán conectada es una red, existen distintas medidas de resumen globales (métricas para caracterizar la topología de la red), entre las que se destacan:

- **Densidad:** Mide la proporción de enlaces que efectivamente hay en una red, respecto a los que potencialmente podrían existir. Para una red no dirigida, esto puede expresarse como:

$$den(G) = \frac{|E(G)|}{\frac{N(N-1)}{2}} \quad (1)$$

donde $|E(G)|$ es el tamaño del conjunto de enlaces de un grafo G , $E(G)$, y $\frac{N(N-1)}{2}$ es la cantidad enlaces posibles en un grafo con N vértices. Si la red fuera dirigida, el número de potencial de enlaces sería el doble, por lo que el denominador de la fórmula sería $N(N-1)$.

- **Distancias:** La distancia geodésica mide la cantidad de enlaces que se deben atravesar para ir de un nodo a otro a través del camino más corto. Como medida de resumen global, con frecuencia se utiliza la distancia media, la cual mide la cantidad de pasos que se deben dar, en promedio, para llegar de un vértice a otro [6]. Por otro lado, puede ser de interés el diámetro de la red, el cual mide el camino más largo existente entre dos nodos de la misma.
- **Transitividad:** A nivel local, la transitividad es la proporción de los vecinos de un cierto nodo que se encuentran conectados entre sí. Es decir, que dado el conjunto de vértices con los que se comparte un enlace, se mide el porcentaje de vínculos que efectivamente existen entre ellos, en relación a los que

teóricamente podría haber. A nivel global, la transitividad puede entenderse como la probabilidad de que dos vecinos de un vértice estén conectados entre sí ([6]). Dicha probabilidad puede aproximarse como:

$$\begin{aligned} trans(G) &= P(e_{ik} \in E | e_{ij} \in E; e_{jk} \in E) \\ &\approx \frac{\# \text{ Tríos cerrados}}{\# \text{ Tríos totales}} \end{aligned} \quad (2)$$

Donde $P(e_{ik} \in E | e_{ij} \in E)$ es la probabilidad de que los vecinos del vértice j , i y k , estén conectados entre sí mediante un vértice e_{ik} .

- **Asortatividad:** Se entiende por asortatividad (“*assortativity*”) u homofilia a la conexión selectiva entre vértices en función de ciertas características. Con frecuencia, se considera la asortatividad en función del grado de los nodos. Toma valores entre -1 y 1 y tiene una interpretación análoga al coeficiente de correlación de Pearson: valores cercanos dan la pauta que los nodos tienden a asociarse con otros de grado similar, y valores cercanos a -1 implican que los vértices tienden a “rechazar” dichas conexiones y por lo tanto existe “disasortatividad”. Por su parte, una homofilia cercana a cero denota una ausencia de una “búsqueda preferencial” de los nodos a la hora de asociarse.

Si bien se presentan muy brevemente algunas métricas, otras como las que miden para los vértices como centralidad de intermediación (**Betweenness centrality**) o centralidad de proximidad, (**Closeness centrality**) y que ya se detallaban en [1], solo se usarán para ver el cambio en la topología de la red completa o filtrada. Esta aclaración va en el sentido de que en este tipo de análisis de redes, no es relevante identificar nodos en particular (que son pacientes con nombre), sobre las que habitualmente se evalúan las métricas individuales antes mencionadas, sino tal vez el tipo de paciente que termina teniendo valores muy altos o bajos de algunas de ellas. Lo que si interesa es una vez formada la red con cierta topología es lograr segmentar los vértices (que son los pacientes),

para formar grupos de nodos, tarea que se trata de hacer mediante la detección de comunidades.

3.2. Métodos para detección de comunidades. Existen varias forma de detectar comunidades y siguiendo los planteos de [5] y que también aparecen en [1] en este caso para avances de los resultados que forman parte de este trabajo extendido, presentados para las memorias SIEPXV, se usan los siguientes 3:

- Algoritmo **Fast Greedy** Para poder determinar los clusters en la red se usa una función que tratar de evidenciar grafos densos, que también se llaman comunidades, optimizando el score de modularidad, a través del algoritmo fast greedy, [3]
- Algoritmo **Random Walk** El algoritmo usado en en este caso para detectar comunidades estructurales trata de encontrar subgrafos densos, usando heurísticas, paseos aleatorios (random walk) y donde la idea es que, los paseos aleatorios cortos tienden a estar en la misma comunidad, [7]
- Algoritmo **cluster louvain** Se basa en la medida de modularidad y un enfoque jerárquico. Inicialmente, cada vértice se asigna a una comunidad por sí mismo. En cada paso, los vértices se reasignan a las comunidades de manera local: cada vértice se traslada a la comunidad con la que logra la mayor contribución a la modularidad. Cuando no se pueden reasignar vértices, cada comunidad se considera un vértice por sí mismo y el proceso comienza de nuevo con las comunidades fusionadas. El proceso se detiene cuando solo queda un vértice o cuando la modularidad no se puede aumentar más en un paso [2], [6].

4. Aplicación

Como fue dicho, los datos utilizados constan de 11 variables dummy que toman el valor 1 si el paciente presenta una cierta característica. Dichas características se agrupan en tres grupos como aparecen en y que son:

- factores de riesgo:Fuma a diario, Consumo nocivo de alcohol, Actividad física insuficiente (**bloque 1**)
- enfermedades no transmisibles (ENT):IMC sobrepeso/obesidad, Razón de cintura-cadera, Hipertensión, Diabetes (**bloque 2**)
- patologías bucales:Prevalencia de bolsa, Pérdida dentaria, Prevalencia de caries, Prevalencia de PIP (**bloque 3**)

4.1. Red completa. En particular, las cuatro patologías bucales consideradas afectan a más de la mitad de los pacientes. De ese modo quedan definidas varios tipos de redes una que se denominará *Red completa* y considera a todos los encuestados y otra que se consigna como *Red filtrada*, para las cuales se hace una caracterización de su topología mediante las métricas antes mencionadas y se aplican los algoritmos de detección de comunidades.

4.2. Red filtrada. Se desarrolla una función para poder filtrar y tener una *Red con umbral*. Para disminuir la cantidad de enlaces de la red, se impone un umbral de cinco, de manera que se supone que dos individuos están conectados si comparten al menos cinco patologías y/o factores de riesgo. Dado que el bloque de factores de riesgo, el de enfermedades no transmisibles y el de patologías bucales están compuestos por 3, 4 y 4 variables, respectivamente, un umbral de cinco garantiza que los individuos compartan características en al menos dos de dichos bloques. Al filtrar las relaciones más "débiles", muchos pacientes pasan a estar desconectados del resto de la red, en la medida que no comparten suficientes características con los demás. Se trabaja únicamente con los 381 vértices (de los 601 originales) conectados a la red mediante alrededor de 27.000 enlaces.

4.3. Detección de comunidades. Para clasificar los pacientes en distintos *clusters* desde la óptica del análisis de redes, se utilizaron los algoritmos presentados en la sección 3.2. De esta manera, se aplicaron los métodos *fast greedy*,

Random Walk y *louvain* a la red filtrada sin nodos aislados. Para cada comunidad obtenida mediante cada uno de los algoritmos, se obtuvo la prevalencia de cada una de las 11 patologías.

4.4. Implementación y reproducibilidad. El presente trabajo fue realizado mediante distintas funciones de librerías como *igraph*, *tidyverse*, [9] del *software* estadístico R, [8] y con gráficos de la librería *ggplot*, [10]. Para asegurar la reproducibilidad de los resultados, se dispone el código y los datos utilizados en un repositorio público al que se puede acceder en <https://gitlab.com/GimeseIesta2/analisis-de-redes-sociales-en-salud> y en la plataforma de Open Science Framework <https://osf.io/j6nt4/>.

5. Resultados

5.1. Red completa. Si se considera que dos pacientes se encuentran conectados por un enlace si comparten al menos una característica, se obtiene una red de 601 vértices y más de 170.000 enlaces (ver gráfico 1). Como se advierte a partir de la tabla 1, esto determina que su densidad sea altísima, de alrededor de 0,95. Dado que el grafo está casi completamente interconectado, la distancia media entre nodos resulta ser de alrededor de un paso. Por su parte, el diámetro de la red es de dos pasos.

Medida	Valor
Densidad	0,95
Distancia media	1,05
Diámetro	2
Transitividad	0,96

Table 1. Medidas de resumen para la red completa.

La transitividad de la red resultó ser de 0,96, de manera que la probabilidad de que los vecinos de un cierto vértice estén conectados entre sí es sumamente alta. Nuevamente, esto es razonable dada la cantidad de enlaces de la red.

En consonancia con lo anterior, la alta densidad de la red también implica que los nodos presentan muchos enlaces incidentes. Como se

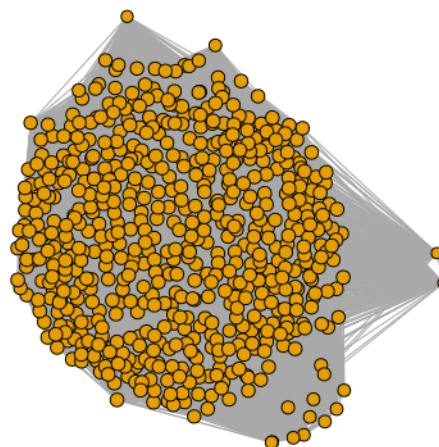


Fig. 1. Red completa con 601 vértices y 170.599 enlaces.

aprecia en la Figura 2, esto implica que la distribución del grado de los vértices presente una fuerte asimetría hacia la derecha, con una media de 567,7. Algunos pacientes comparten al menos una característica con todos los demás, por lo que su grado es de 600.

En la Figura 3, se advierte que la distribución de la intermediación de los nodos de la red se encuentra bastante dispersa con una cierta asimetría a la derecha. La correspondiente media es de 16,1. Es decir, que en promedio, alrededor de 16 caminos más cortos pasan por cada nodo.

A partir del análisis anterior, se concluye que la alta densidad de la red dificulta visualizar la estructura de las relaciones entre los pacientes en lo que refiere a sus factores de riesgo y sus enfermedades. En consecuencia, puede ser útil establecer un criterio más exigente para poder afirmar que dos pacientes comparten un enlace.

5.2. Red filtrada. Como se muestra en la Figura 4, al filtrar las relaciones más "débiles" muchos pacientes pasan a estar desconectados del resto de la red, en la medida que no comparten suficientes características con los demás. En lo

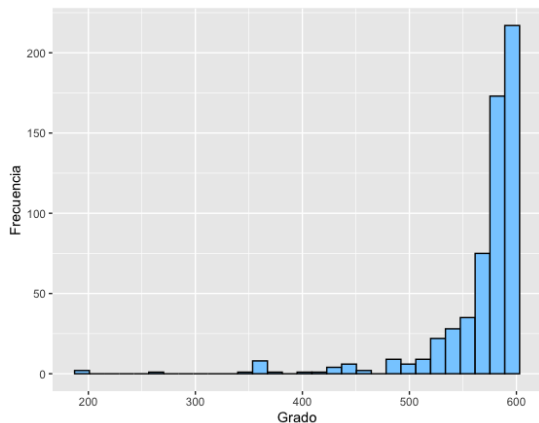


Fig. 2. Histograma de distribución de grados

que sigue, se trabaja únicamente con los 381 vértices conectados a la red mediante alrededor de 27.000 enlaces. En el correspondiente gráfico, se observa que la mayoría de los nodos se encuentran fuertemente conectados entre sí. En línea con esto, el *clique* más grande de la red se compone de 127 vértices (ver tabla 2).

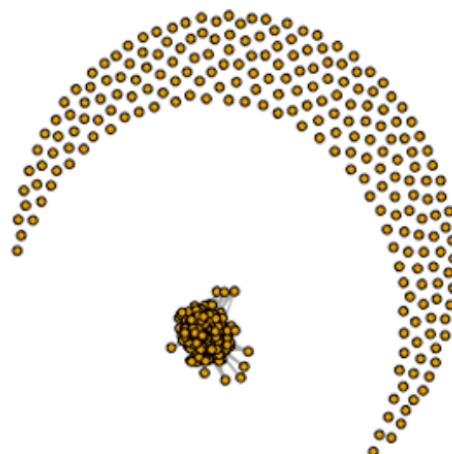


Fig. 4. Red con 601 vértices y 27.382 enlaces

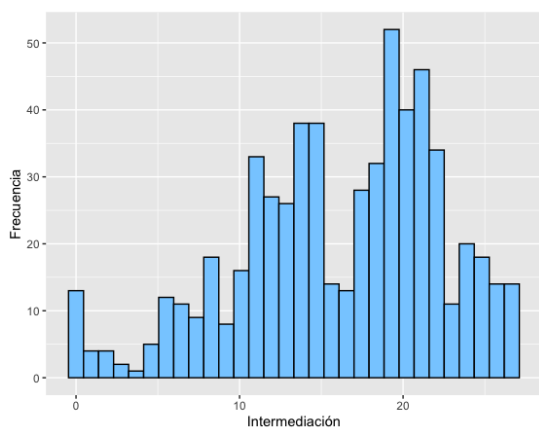


Fig. 3. Histograma de métrica *intermediación*

Medida	Valor
Densidad	0,38
Distancia media	1,62
Diámetro	3
Transitividad	0,66

Table 2. Medidas de resumen para la red filtrada sin nodos aislados.

Al reducirse la cantidad de enlaces, la densidad de la red se reduce a 0,38, y su diámetro aumenta a tres. En consecuencia, la distancia media entre nodos aumenta a 1,6. Si bien la transitividad disminuye sensiblemente, la misma continúa siendo bastante alta (0,66).

Evidentemente, la disminución en la cantidad de enlaces redonda en una caída en el grado de los vértices. Como se aprecia en la figura 6, la distribución del grado de cada nodo se desplaza a la izquierda, con un máximo de 372

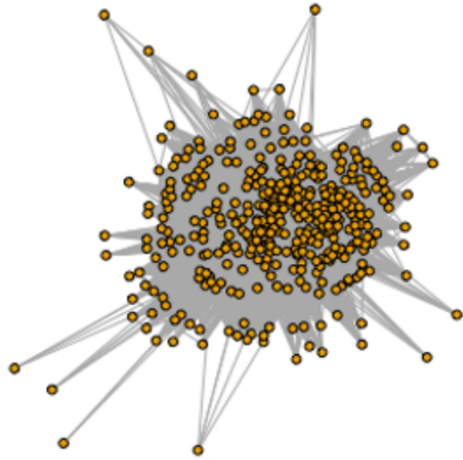


Fig. 5. Red sin nodos aislados (381 vértices y 27.382 enlaces)

y una media de 143,7.

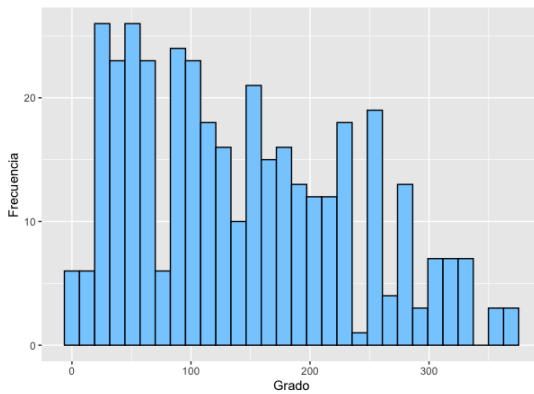


Fig. 6. Distribución del grado de los vértices de la red filtrada.

Al disminuir la densidad de la red, algunos nodos incrementan su importancia para la conexión de la red. De esta forma, se observan algunos valores inusualmente altos en la distribución del coeficiente de intermediación de la red (ver figura 7). De todos modos, dicha distribución se encuentra sumamente concentrada en valores bajos del recorrido, con una media de

118,6.

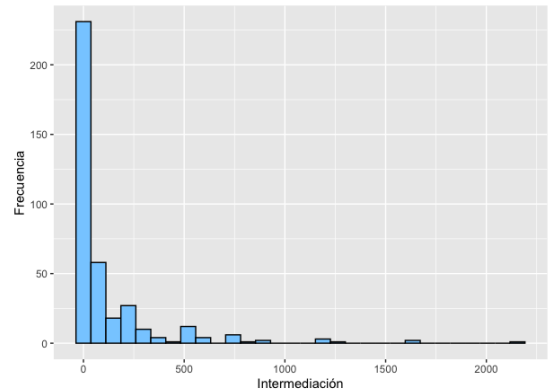


Fig. 7. Distribución de la intermediación de los vértices de la red filtrada.

A partir de la figura 9, se advierte un cierto aumento en el indicador de cercanía de cada nodo a los demás, de forma que la media de la distribución se incrementa a 0,002. Al existir menos enlaces, es razonable que aumente la distancia geodésica de cada nodo a los demás.

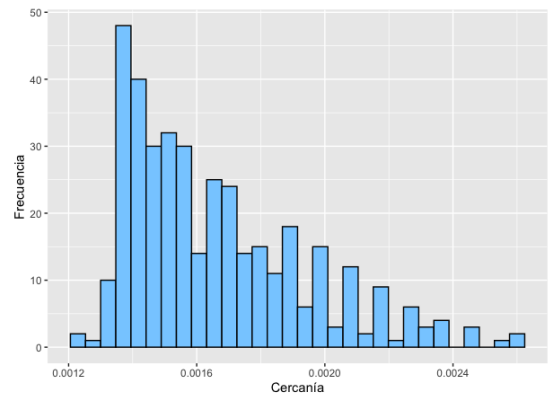


Fig. 8. Histograma de métrica de cercanía

Fig. 9. Distribución de la cercanía de los vértices de la red filtrada.

5.3. Detección de comunidades. A continuación, se analizan las comunidades detectadas mediante cada uno de los tres

algoritmos considerados para la red filtrada sin nodos aislados, compuesta por 381 pacientes.

5.3.1. Algoritmo *fast greedy*. Como se muestra en la figura 10, a partir del algoritmo *fast greedy* se detectaron tres comunidades. Mientras que las comunidades 1 y 2 se componen de 171 y 191 individuos respectivamente, la comunidad 3 cuenta con sólo 19 pacientes. Como se aprecia a partir de la tabla 3, cada grupo presenta diferentes características.

En primer lugar, la comunidad 1 es un grupo con relativamente baja presencia de factores de riesgo comportamentales tales como tabaquismo y alto consumo de alcohol. En contraposición, una alta proporción de los pacientes pertenecientes a este grupo poseen factores de riesgo generales como sobrepeso y obesidad, alta razón de cintura-cadera, hipertensión y diabetes. En términos de salud bucal, para la mayoría de las patologías la comunidad 1 exhibe prevalencias similares a las del total de la red filtrada. La única excepción corresponde a la variable de pérdida dentaria: mientras que casi un 58 % de la población general ha perdido piezas dentales, para la comunidad 1 esta proporción fue de sólo 35 %.

En segundo lugar, la comunidad 2 se destaca principalmente por la alta prevalencia de pérdida dentaria y caries, en ambos casos superior al 80 %. Para las demás patologías vinculadas a la salud bucal, las proporciones fueron similares a las de toda la red. Por otra parte, dentro de esta comunidad, la prevalencia de patologías vinculadas a la salud general resultó ser menor al total general. Finalmente, en lo que respecta a los factores de riesgo comportamentales, la comunidad 2 se caracteriza por la relativamente alta prevalencia de pacientes que fuman a diario.

Como fue señalado, la comunidad 3 está conformada por únicamente 19 individuos, con lo cual las prevalencias obtenidas deben ser interpretadas con precaución. Este grupo presenta una alta prevalencia de factores de riesgo tanto comportamentales como vinculados a la salud general (a excepción de la diabetes, donde la proporción fue similar a la de toda la red). A su vez, se advierten prevalencias relativamente bajas para la mayoría de patologías bucales.

Únicamente la prevalencia de PIP resultó ser superior a la total.

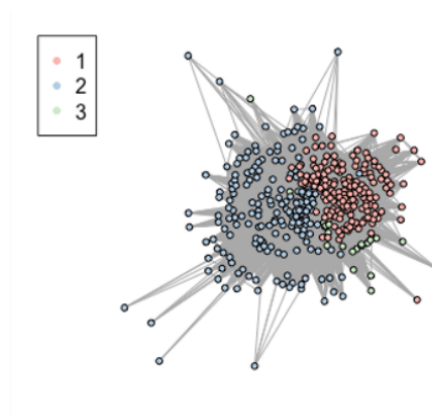


Fig. 10. Detección de comunidades mediante el algoritmo *fast greedy* para la red filtrada sin nodos aislados.

Variable	Comunidad 1	Comunidad 2	Comunidad 3	Total
Fuma a diario	18,1 %	46,6 %	42,1 %	33,6 %
Consumo nocivo de alcohol	7,6 %	14,7 %	31,6 %	12,3 %
Actividad física insuficiente	45,0 %	49,7 %	78,9 %	49,1 %
IMC sobrepeso/obesidad	81,3 %	51,3 %	94,7 %	66,9 %
Razón de cintura-cadera	81,3 %	46,6 %	89,5 %	64,3 %
Hipertensión	70,2 %	34,0 %	100,0 %	48,6 %
Diabetes	34,5 %	15,2 %	26,3 %	24,4 %
Presencia de bolsa	64,3 %	68,6 %	57,9 %	66,1 %
Pérdida dentaria	35,1 %	82,7 %	10,5 %	57,7 %
Prevalencia de caries	72,5 %	82,7 %	47,4 %	76,4 %
Prevalencia de PIP	74,9 %	71,7 %	89,5 %	74,0 %
Cantidad de nodos	171	191	19	381

Table 3. Prevalencias de patologías totales y por comunidad según el algoritmo *fast greedy* para la red filtrada y sin nodos aislados.

5.3.2. Algoritmo *random walk*. El algoritmo *random walk* arrojó cuatro comunidades de pacientes. Las primeras tres constan de 32, 139 y 209 individuos respectivamente. Por su parte, el cuarto grupo se compone de un único paciente con sobrepeso u obesidad y razón de cintura-cadera alta, así como presencia de bolsa y de PIP. En consecuencia, no tiene sentido la caracterización de estas comunidades, ya que implicaría excluir del análisis ese nodo que queda siendo parte de una sola comunidad. Dado que la idea es poder comparar la misma cantidad de nodos, se opta por presentar los resultados al usar el Algoritmo *louvain*, que como ya se presentó en la sección 4.3 tiene una lógica de construcción mediante métodos jerárquicos en fases.

5.3.3. Algoritmo *louvain*. A partir del algoritmo de *louvain*, se obtuvieron cuatro grupos de 66, 126, 66 y 123 individuos, respectivamente. Como se muestra en la tabla 4, la comunidad 1 se caracteriza por la alta proporción de miembros que realizan insuficiente actividad física y tienen IMC y razón de cintura-cadera excesivamente altos. En lo que refiere a la salud bucal, se trata de un grupo con baja pérdida de piezas dentales y alta prevalencia de PIP.

La comunidad 2 exhibe una alta proporción de pacientes que fuman a diario. Aunque presenta prevalencias relativamente bajas de factores de riesgo vinculados a la salud general, se aprecia una muy alta proporción de pacientes que padecen patologías bucales.

Al igual que la comunidad 1, la comunidad 3 se caracteriza por la alta proporción de individuos que realizan actividad física y con problemas de sobrepeso y obesidad. Sin embargo, este grupo presenta relativamente bajos niveles de tabaquismo y alcoholismo si se lo compara con el total de la red filtrada sin nodos aislados. Además, se trata de un grupo con bajas prevalencias de bolsa y PIP y altas prevalencias de pérdida dentaria y caries.

Finalmente, la comunidad 4 presenta bajos niveles de factores de riesgo comportamentales y alta prevalencia de patologías vinculadas a la salud general. En particular, este grupo posee tasas particularmente altas de hipertensión y diabetes. En lo que refiere a enfermedades odontológicas, es particularmente llamativa la alta prevalencia de PIP.

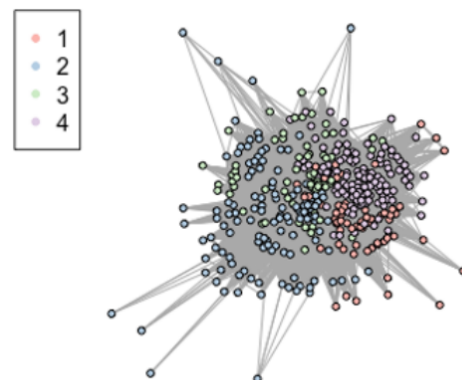


Fig. 11. Detección de comunidades mediante el algoritmo *louvain* para la red filtrada sin nodos aislados.

Variable	Comunidad 1	Comunidad 2	Comunidad 3	Comunidad 4	Total
Fuma a diario	33,3 %	58,7 %	24,2 %	13,0 %	33,6 %
Consumo nocivo de alcohol	10,6 %	18,3 %	3,0 %	12,2 %	12,3 %
Actividad física insuficiente	62,1 %	42,9 %	71,2 %	36,6 %	49,1 %
IMC sobrepeso/obesidad	80,3 %	39,7 %	81,8 %	79,7 %	66,9 %
Razón de cintura-cadera	77,3 %	35,7 %	77,3 %	79,7 %	64,3 %
Hipertensión	16,7 %	31,7 %	53,0 %	80,5 %	48,6 %
Diabetes	22,7 %	14,3 %	25,8 %	35,0 %	24,4 %
Presencia de bolsa	68,2 %	76,2 %	39,4 %	69,1 %	66,1 %
Pérdida dentaria	25,8 %	80,2 %	87,9 %	35,8 %	57,7 %
Prevalencia de caries	77,3 %	83,3 %	83,3 %	65,0 %	76,4 %
Prevalencia de PIP	80,3 %	81,7 %	33,3 %	84,6 %	74,0 %
Cantidad de nodos	66	126	66	123	381

Table 4. Prevalencias de patologías totales y por comunidad según el algoritmo *louvain* para la red filtrada y sin nodos aislados.

6. Conclusiones

En el presente trabajo, se realizó un primer análisis exploratorio de una red integrada por una muestra de 601 pacientes que se atendieron en la Facultad de Odontología en 2015 y 2016, para los que se contaba con 11 variables indicatrices de distintos factores de riesgo y enfermedades. Se consideraron distintos criterios para afirmar que dos individuos se encuentran conectados mediante un enlace.

En primer lugar, se definió que dos personas se encuentran conectados si comparten al menos un factor de riesgo, enfermedad no transmisible y/o patología bucal. En este caso, la red resultante es demasiado densa para reconocer la estructura de relaciones entre pacientes. Por este motivo, en segundo lugar, se fijó un umbral de cinco características compartidas para que dos individuos se encuentren conectados mediante un enlace. Al imponer esta restricción, muchos pacientes pasan a estar aislados de la red. En este trabajo, se optó por excluirlos del análisis. De esta manera, se trabajó con 381 nodos conectados.

A partir de la red filtrada sin vértices aislados, se buscó crear perfiles epidemiológicos mediante distintos algoritmos de detección de comunidades. Los resultados variaron mucho dependiendo del método implementado, tanto en términos de cantidad de comunidades detectadas como de su tamaño y prevalencia de patologías. No se advierten patrones muy claros en los grupos de nodos tanto para enfermedades ENT o patologías bucales y factores de riesgo, que pueda

ayudar a decidir clara,memnte cual es la mejor agrupación, por lo que queda profundizar sobre este punto en trabajos futuros.

Referencias

- [1] Ramón Álvarez-Vaz y Fernando Massa (2020) Aplicación de análisis de redes para la elaboración de perfiles epidemiológicos en estudios sanitarios. Actualidad en Probabilidad y Estadística (pp. 81–102). BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA Dirección General de Fomento Editorial. <https://doi.org/10.5281/zenodo.7149166>.
- [2] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. doi:10.1088/1742-5468/2008/10/p10008
- [3] Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6). doi:10.1103/physreve.70.066111
- [4] Csardi G, Nepusz T (2006). “The igraph software package for complex network research.” *InterJournal, Complex Systems*, 1695. <https://igraph.org>

- [5] Kolaczyk, E. and Csárdi, G. (2014). *Statistical analysis of network data with R*. Springer, New York.
- [6] Pablo Mones y Ramón Álvarez-Vaz (2021) Análisis de la red de investigadores del IESTA, parte 1 (Serie Documentos de Trabajo; 7/21). Montevideo: Universidad de la República. Facultad de Ciencias Económicas y de Administración, Instituto de Estadística. <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/10518>.
- [7] Pons, P., and Latapy, M. (2015). Computing communities in large networks using random walks, <http://arXiv.org/abs/physics/0512106v1>
- [8] R Core Team (2023). R: A Language and Environment for Statistical Computing, [url = https://www.R-project.org/](https://www.R-project.org/)
- [9] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686
- [10] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.