

Elaboración de Perfiles Epidemiológicos en Estudios Sanitarios mediante Análisis de Redes

Anónimo

Abstract El presente trabajo surge de los datos generados para el Estudio de las ENT para pacientes que se atienden entre 2015 y 2016 en la Facultad de Odontología de la Universidad de la República (Álvarez-Vaz and Massa 2020), a través de una muestra probabilística obtenida mediante un diseño sistemático de pacientes. Del estudio madre surge el siguiente trabajo que consiste en un análisis de redes sobre los datos, donde se construye una red formada por los pacientes que se toman como nodos. Se considera que dos individuos están conectados si comparten al menos un factor de riesgo o patología, que aparecen consignados como 11 variables binarias.

Posteriormente, se considera un umbral más exigente para definir que dos individuos se encuentran conectados por un eje, mediante la aplicación de una función de filtrado. De manera arbitraria, 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 recoge información sobre potenciales relaciones de comorbilidad. Finalmente a través de diferentes algoritmos de detección de comunidades se elaboran los perfiles epidemiológicos.

Palabras clave: Redes - Comunidades - ENT (Enfermedades NO Transmisibles)

Introducción

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, (Álvarez-Vaz and Massa 2020).

Objetivos

El objetivo de este trabajo es obtener perfiles epidemiológicos bien diferenciados en base a los atributos binarios partiendo de un conjunto de variables, sin discriminar cuales son variables de respuesta, proponiendo la creación de grupos mediante la estrategias de detección de comunidades.

Metodología

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 (Álvarez-Vaz and Massa 2020) 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**)

Teniendo en cuenta que varios de estos indicadores son sumamente prevalentes en la población, se decide trabajar filtrando la red, imponiendo un umbral para determinar si dos individuos están conectados o no, para evitar tener una red extremadamente densa. O trabajar con subredes al considerar los 3 tipos de variables por separado.

Para cada tipo de red a su vez se evalúan diferentes tipos de métricas como son:

- Centralidad de intermediación (Betweenness centrality), (Brandes 2001)
- Centralidad de proximidad (Closeness centrality), (Freeman 1979)
- Cohesión de la red (cliques)
- Modularidad
- Enlace selectivo (Asortatividad)

Finalmente para las redes completas o filtradas se usan diferentes algoritmos de detección de comunidades como :

- Algoritmo **Random Walk** El algoritmo usado 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, (Pons and Latapy 2015)
- Algoritmo **Fast Greedy** Para poder determinar los clusters en la red se usa una función que trata de evidenciar grafos densos, que también se llaman comunidades, optimizando el score de modularidad, a través del algoritmo fast greedy, (Clauset, Newman, and Moore 2004)
- 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 (Blondel et al. 2008).

Para asegurar la reproducibilidad de los resultados del análisis realizado, se dispuso el código y los datos utilizados en un repositorio público al que se puede acceder a través de este link <https://gitlab.com/Gimesefesta2/analisis-de-redes-sociales-en-salud>

Aplicación

La descripción de las variables binarias se hace mediante **tidyverse** (Wickham et al. 2019) y para el análisis de las redes se usa la librería **igraph** (Csárdi et al. 2023).

En particular, las cuatro patologías bucales consideradas afectan a más de la mitad de los pacientes. De ese modo quedan definidas 3 tipos de redes **Red completa**, para la cual se caracteriza su topología mediante las métricas mencionadas y se aplican los algoritmos de detección de comunidades.

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.

Por último se elaboran **Redes por tipo de variable**

- Red de factores de riesgo: 407 nodos y 53000 enlaces
- Red de enfermedades no transmisibles : 452 nodos y alrededor de 87.000 enlaces
- Red de patologías bucales: 481 y 153000 enlaces

Dado que cada red por tipo de variable tiene diferentes nodos, la detección de comunidades se hace para la red completa filtrada. Con los diferentes algoritmos la estructura de comunidades queda con 3 para el **Random Walk**, 4 para el **Fast Greedy** y 4 para **cluster_louvain**

Conclusiones

Con los resultados encontrados que muestran una tipología de perfiles bien diferenciados en las diferentes comunidades detectadas (en la presencia y ausencia de las variables), y una distribución de las métricas también diferencial por comunidad, se propone continuar con esta estrategia, al trabajar con las Redes por tipo de variable y ver si cambia la topología de cada red y por ende las comunidades, cuando se consideraban en forma simultánea, así como la proyección en las redes, de otros atributos de los pacientes como sexo, edad.

10 Álvarez-Vaz, Ramón, and Fernando Massa. 2020. “Aplicación de Análisis de Redes Para La Elaboración de Perfiles Epidemiológicos En Estudios Sanitarios.” *Actualidad En Probabilidad y Estadística*.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. “Fast Unfolding of Communities in Large Networks.” *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10): P10008. <https://doi.org/10.1088/1742-5468/2008/10/p10008>.

Brandes, Ulrik. 2001. “A Faster Algorithm for Betweenness Centrality.” *The Journal of Mathematical Sociology* 25 (2): 163–77. <https://doi.org/10.1080/0022250X.2001.9990249>.

- Clauset, Aaron, M. E. J. Newman, and Cristopher Moore. 2004. "Finding Community Structure in Very Large Networks." *Physical Review E* 70 (6). <https://doi.org/10.1103/physreve.70.066111>.
- Csárdi, Gábor, Tamás Nepusz, Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, and Kirill Müller. 2023. *igraph: Network Analysis and Visualization in r*. <https://doi.org/10.5281/zenodo.7682609>.
- Freeman, Linton C. 1979. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1 (3): 215. <http://proxy.timbo.org.uy:443/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edo&AN=17189509&lang=es&site=eds-live>.
- Pons, Pascal, and Matthieu Latapy. 2015. "Computing Communities in Large Networks Using Random Walks." <http://arXiv.org/abs/physics/0512106v1>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.