



Islas Canarias
Del 15 al 19 de noviembre de 2021



Instituto de Estadística
y Cartografía de Andalucía
Consejería de Transformación Económica,
Industria, Conocimiento y Universidades

PGROUTING PARA EL CÁLCULO DE INDICADORES DE ACCESIBILIDAD

María Escudero Tena

Instituto de Estadística y Cartografía de Andalucía
maria.escudero.tena@juntadeandalucia.es

Alberto Quintanilla Cabañero

Smart&City Solutions
alberto.quintanilla@smartandcity.com

Iria Enrique Regueira

Instituto de Estadística y Cartografía de Andalucía
iria.enrique@juntadeandalucia.es

Introducción

Esta comunicación es fruto del partenariado entre la Agencia Andaluza de Cooperación Internacional para el Desarrollo (AACID), el Instituto de Estadística y Cartografía de Andalucía (IECA) y el Programa de las Naciones Unidas para el Desarrollo (PNUD) con el objetivo de reforzar el papel institucional de la Estadística Pública en el seguimiento y monitorización de los ODS (Objetivos de Desarrollo Sostenible) en el ámbito local en Andalucía. En este marco, ha surgido un proyecto de creación de información pública y oficial para seguimiento y monitorización de los ODS cuyo objetivo era crear una metodología de cálculo para cuatro indicadores identificados en el Manual Europeo para la realización de Informes Locales Voluntarios de los ODS (VLR) que tomara ventaja en la mayor medida posible de los conjuntos de datos espaciales y poblacionales georreferenciados mantenidos por IECA particularmente adecuados.

Algunos Objetivos de Desarrollo Sostenible, en particular el 11 (Lograr que las ciudades sean más inclusivas, seguras, resilientes y sostenibles) tienen mucho que ver con el acceso de la población a recursos y servicios, y por esto el citado manual explora la dimensión geográfica en muchos indicadores. Sin embargo, la simple proximidad física a recursos como zonas verdes, transporte público, instalaciones deportivas, etc. puede resultar engañosa a la hora de averiguar si existe este acceso. Así, se pueden encontrar algunas barreras arquitectónicas o de infraestructura, la trama urbana puede estar desconectada o configurada de manera ineficiente, provocando la segregación de una parte de la población. Por tanto, el análisis de red es la metodología ideal para determinar el acceso real a los recursos, y pgrouting es la librería de código abierto de referencia para este proceso.

Objetivos

El objetivo de esta comunicación es mostrar la aplicación de los algoritmos y software de análisis de red o de rutas en el cálculo de dos de los indicadores seleccionados en el proyecto. La metodología en ambos casos es muy similar en cuanto al análisis de rutas, por lo que en general nos referiremos a lo largo de la comunicación al primero en varios ejemplos.

Metodología

Datos disponibles

En los indicadores de aplicación se contaba con la oportunidad de utilizar conjuntos de datos georreferenciados, en su mayoría creados y mantenidos por el IECA que permiten cierta sofisticación adicional en las metodologías respecto a las definidas en el manual VLR:

- Estructura y ubicación de la población por edificios (portales).
- Ubicación de las zonas verdes y datos sobre su accesibilidad.
- Red viaria.
- Ubicación de las paradas de transporte público y datos de paso de vehículos (datos ajenos a IECA, procedentes de portales de datos abiertos).

Aun considerando el análisis de red o análisis de rutas como la herramienta más adecuada para el problema de la accesibilidad, no existe un único enfoque o método de cálculo, sino varios diferentes teniendo en cuenta la estructura de los datos sobre los que se ha aplicado (más aún si consideramos las herramientas disponibles). En todos los casos, sin embargo, son necesarias unas tareas previas de limpieza y simplificación de los datos para hacerlos compatibles con los algoritmos de análisis de rutas, ya que, en esencia, debemos reducir el problema a dos conjuntos de puntos que se relacionen por una red de enlaces.

Uno de los principales desafíos del proyecto ha sido gestionar grandes conjuntos de datos, lo que podía generar tiempos de cálculo prolongados, al no haber utilizado un hardware especialmente potente. Nuestro objetivo es ejecutar los algoritmos con los lotes de datos más grandes posibles, para evitar efectos frontera no deseados, pero manteniéndolo en un rango razonable para que estén al alcance de cualquier usuario. Cuando ha sido posible, todas las operaciones se han realizado utilizando capas y tablas que contienen la totalidad de datos de Andalucía. Sin embargo, algunos pasos que requieren mucho tiempo no se pudieron realizar de esta manera y se han dividido en provincias o en bloques. Esto se destacará durante la descripción de la metodología.

Limpieza y simplificación de los datos

El procedimiento descrito ha sido diseñado para minimizar la cantidad de datos gestionados. Algunas partes no son esenciales en la metodología y los resultados serían similares si no se aplicaran. El lector que quiera repetir este proceso para una sola ciudad o un área más pequeña probablemente no necesitará estas simplificaciones para lograr tiempos de cómputo aceptables.

De acuerdo con la definición (Poelman, 2018), las líneas del viario que estén más lejos de los 833 m (distancia que puede caminar una persona promedio en 10 minutos) desde cualquier zona verde no se utilizarán en el análisis, ya que sería la distancia teórica máxima que podría cubrirse en caso de que existiera un camino recto partiendo del punto

de servicio. Por lo tanto, estos tramos se pueden eliminar para aligerar las tablas de datos. En el caso del indicador de zonas verdes, utilizando QGIS se ha creado un búfer de 833 m alrededor de todas ellas y se han seleccionado solo las líneas que intersecan con él. Antes de exportar la capa resultante a la base de datos postgis, QGIS también se ha utilizado para convertir a geometría de una sola línea.

Hay dos problemas principales para el uso de las zonas verdes en el análisis de red con pgrouting. En primer lugar, los polígonos (como los que forman la capa de zonas verdes) no se pueden usar como parámetros para funciones basadas en pgrouting. Además, las zonas verdes generalmente no se cruzan con las vías (sus límites son paralelos a las calles en los casos más comunes). Por lo tanto, necesitamos definir cómo transformamos estas áreas en puntos dentro del viario que se puedan utilizar como origen de nuestro análisis de red. Los siguientes pasos se han realizado utilizando QGIS:

1. Los polígonos se han transformado en sus líneas perimetrales.
2. Las líneas obtenidas se han dividido en puntos separados por una distancia regular. Se han aplicado dos distancias de paso puntuales diferentes:
 - a. Cada 10 m para áreas verdes menores de 102 ha. Este límite se ha elegido para garantizar que todas las zonas verdes urbanas se procesen de esta manera, y esta es la superficie más grande que se encuentra en la capa de zonas verdes del Sistema Urbano de DERA¹.
 - b. Cada 50 m para todas las demás zonas verdes.
3. Como complemento al filtro de vías mencionado anteriormente, se eliminan los puntos perimetrales de las zonas verdes situados más lejos de 833 m de cualquier vía.
4. QGIS finalmente se ha utilizado para ajustar los puntos resultantes a la vía más cercana.

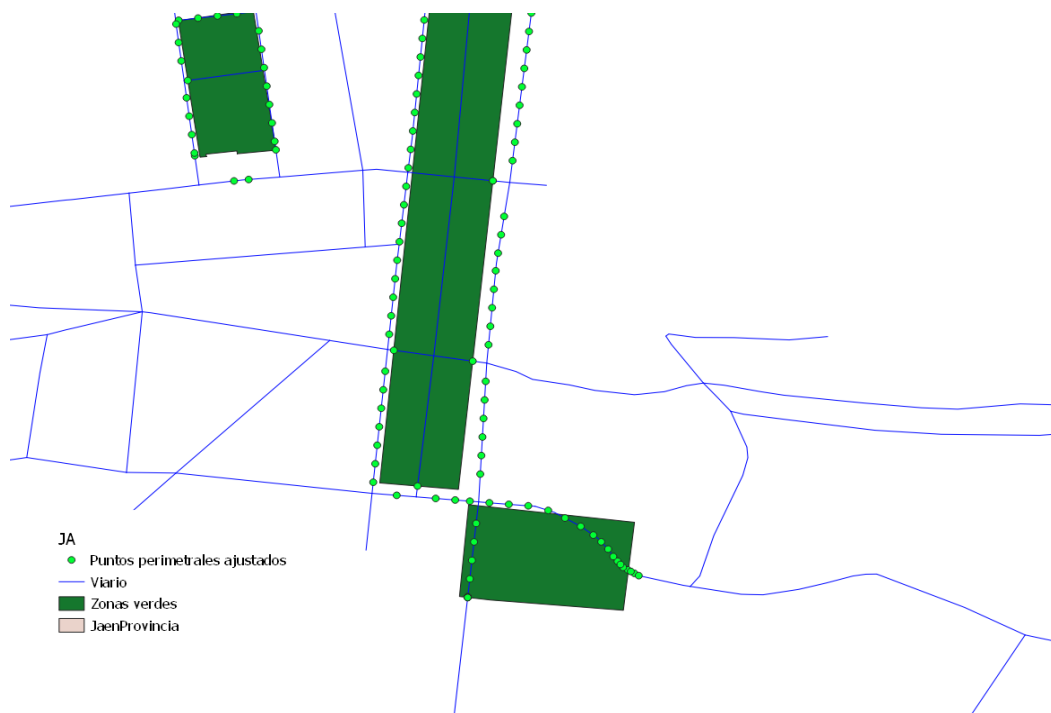


Figura 1 - Zonas verdes modeladas como puntos perimetrales ajustados al viario

¹ Datos Espaciales de Referencia de Andalucía

Después de crear la capa de puntos perimetrales del área verde, una inspección rápida muestra muchos puntos aislados derivados de los micropolígonos que quedaron en la creación de la capa de datos de la zona verde². En algunas provincias estos puntos representan casi el 50% del total, aumentando artificialmente el tiempo de cálculo.

La capa de edificios está formada por puntos (idealmente es el punto de acceso al edificio, aunque existen excepciones). Al igual que con los puntos perimetrales de las zonas verdes, deben estar al lado de una vía, y se han movido a la más cercana utilizando la herramienta “Ajustar a Geometría” de QGIS. Existe una alternativa con código postgis para esto, pero después de algunas pruebas ha mostrado un peor rendimiento.

Método de cálculo

Como comentábamos al principio, existen varios enfoques posibles, y en este estudio hemos barajado dos:

1. Hallar un área de servicio alrededor de los recursos: esta es un área, común a todos los puntos o superficies que dan el servicio, dentro de la cual dicho servicio es accesible.
2. Encontrar, para cada punto de población, el camino óptimo a un punto o superficie prestadora del servicio si lo hay.

Se ha adoptado el primer enfoque, ya que intuitivamente parece más eficiente realizar un cálculo único que realizar iterativamente el análisis de red para una gran cantidad de puntos. Sin embargo, este método también ha presentado algunos inconvenientes (en la práctica, el cómputo del área de servicio debe realizarse también un número similar de veces al comenzar desde zonas verdes, debido a la conversión de los polígonos en puntos).

Las herramientas de enrutamiento trabajan con grafos, no con geometrías (Cormen, y otros, 2001). Es decir, a partir de las geometrías originales de la red viaria se construye una representación lógica (topología de red), formada por nodos y enlaces. Dependiendo del problema tratado podemos asignar diferentes pesos necesarios para atravesar los enlaces de la topología, poner restricciones en los giros alrededor de los bordes o asignar diferentes clases a los enlaces (residencial, autopista, puente de peaje).

El proceso de extraer la información necesaria de las tablas espaciales se denomina "construcción de la topología de red". Pgrouting proporciona la función `pgr_createTopology` (`pgr21`) para construir esta topología basándose en la información de geometría. Requiere los siguientes parámetros:

- `Edge_table`: tabla de enlaces. En nuestro caso, es la tabla que contiene la información de la red viaria.
- `Tolerance`: separación entre bordes desconectados. Se utiliza para decidir si varios nodos cercanos dentro de la tolerancia se consideran el mismo nodo.
- `the_geom`: nombre de la columna de geometría de la tabla de enlaces.
- `Id`: nombre de la columna de clave principal de la tabla de enlaces.
- `Source`: nombre de la columna de fuente u origen de la tabla de enlaces.
- `Destino`: nombre de la columna de destino de la tabla de enlaces. Estas dos columnas se construyen durante la ejecución.

² La creación de esta capa se describe en la comunicación “Estadística Pública en el Seguimiento y Monitorización de los ODS en el ámbito local en Andalucía”

Como resultado, la topología se almacena en dos tablas:

- `Edge_table`: modificada por la función llenando las columnas de origen y destino con los dos vértices que definen cada enlace. Los enlaces conectan los nodos de la cuadrícula.
- `Vertices_table`: almacena los nodos de la topología.

Pgrouting también proporciona funciones para analizar y reparar (si es necesario) la topología (`pgr21`).

- `pgr_analyzeGraph`: Analiza la topología de la red y termina de llenar la información en la tabla de nodos (`Vertices_table`) que hará más rápidos los siguientes pasos.
- `pgr_nodeNetwork`: por lo general, los datos GIS no suelen estar "enlazados" correctamente. Esto crea topologías no válidas, lo que dará como resultado rutas incorrectas al intentar usarlas con `pgrouting`. Estos enlaces significan que en cada intersección de la red de calles, todos los bordes se dividirán en segmentos de vía separados. Hay casos como una intersección de paso elevado y paso inferior donde no se puede atravesar desde el paso elevado hasta el paso inferior, pero esta función no tiene la capacidad de detectar y adaptarse a esas situaciones. Esta función lee la tabla de enlaces (`Edge_table`) que hemos creado en el primer paso, que tiene un ID de columna de clave primaria y una columna de geometría llamada `the_geom`, y cruza todos los segmentos en ella con todos los otros segmentos y con ello crea una tabla `edge_table_noded`.

Después de seleccionar la tabla corregida (`edge_table_noded`), la topología se genera nuevamente para crear todos los vértices nuevos.

La mayoría de las funciones de `pgrouting` están orientadas a trabajar con nodos en la topología. Sin embargo, esta limitación no se ajusta al problema que estamos tratando ya que la distancia a los vértices tanto de los edificios como de los puntos perimetrales de las áreas verdes es relevante para el problema (de forma muy inconveniente para este estudio, los portales no están oportunamente situados en las intersecciones de las calles). La familia de funciones `pgrouting "with_points"` nos permite proporcionar también puntos ubicados en los enlaces, pero en la medida en que el análisis de red no funciona con geometrías, deben transformarse en algo que las funciones puedan comprender. Necesitamos crear duplas compuestas por:

- Enlace: donde se encuentra el punto
- Fracción: Valor en el rango $<0, 1>$ que indica la posición relativa entre los extremos del enlace. Si es 0, el punto se ubicará en el punto de origen del enlace. Si es 1, estará en el punto destino.

La creación de estas duplas es un proceso lento con `postgis`, por tanto, es uno de los problemas que abordamos por partes, utilizando la función `ST_LineLocatePoint` en lotes de 10,000 puntos.

La cantidad de puntos procesados puede reducirse aún más una vez calculadas todas las fracciones agrupar las duplas que coinciden (que corresponden a puntos perimetrales cuyo punto más cercano en la vía es el mismo).

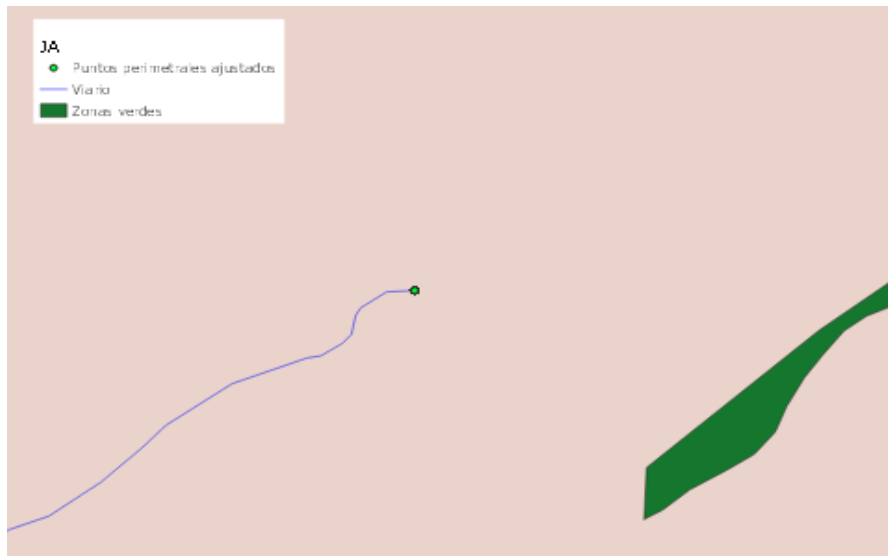


Figura 2 - Varios puntos perimetrales de una zona verde coinciden en la misma dupla Enlace/Fracción

La función `pgr_withPointsDD` permite el uso de estos puntos y, utilizando el algoritmo de Dijkstra, extrae todos los nodos y puntos que requieren para alcanzarlos un coste menor o igual al valor de distancia especificado desde un punto de partida. Los enlaces extraídos en la ejecución construyen el árbol de expansión correspondiente.

La función se llama con los siguientes parámetros:

- `edge_sql`: consulta (como texto) que proporciona la topología completa utilizada. Esto coincide con nuestra tabla de topología con nodos, incluyendo una columna de coste (en este caso, es simplemente la longitud del enlace).

```
SELECT id, source, target, st_length (geom) AS cost, -st_length (geom)
AS reverse_cost FROM viario_simple_noded
```

- `points_sql`: consulta (como texto) que proporciona los vértices de la topología y los puntos que se deben considerar. En nuestro caso, solo proporcionamos puntos

```
SELECT id AS pid, id_via AS edge_id, fraction FROM fractions
```

- `start_vid`: una matriz de los puntos y/o vértices que se tomarán como puntos de partida. Como convención de signos, si son positivos se entienden como vértices y si son negativos, como puntos. Por lo tanto, en nuestro caso solo proporcionamos identificadores negativos.
- `distance`: distancia máxima que se puede recorrer.
- `directed`: si es VERDADERO, los bordes se pueden recorrer en ambas direcciones.
- `details`: si es VERDADERO, la salida de la función incluye los nodos intermedios.

El tiempo de ejecución de las funciones `pgrouting` aumenta exponencialmente a medida que aumenta la densidad de la topología (número de vértices y enlaces) y

proporcionalmente con la cantidad de puntos de partida (Nestar, 2015). Por tanto, hemos vuelto a dividir el proceso en lotes de 10.000 fracciones.

La salida de la ejecución de `pgr_withpointsDD` es solo parcial y no se puede usar directamente para crear un área de servicio e intersecarla con los puntos de construcción:

- El cálculo solo incluye los vértices dentro de la distancia y el coste consumido (distancia) para alcanzarlos. Puede haber puntos más allá de esos vértices que aún se pueden alcanzar.
- No se utilizan todos los enlaces posibles para llegar a estos vértices.

De hecho, no se requiere que todos los enlaces lleguen a todos los vértices intermedios, y `pgrouting` determina sólo los que tienen el menor coste. Por lo tanto, algunas calles no se incluyen en la tabla de resultados, aunque estén incluidos en el área de servicio. Para completar esta tabla, entonces, necesitamos encontrar estos enlaces no utilizados que aún se pueden recorrer dentro de la distancia especificada:

- Encontramos los enlaces que no se han incluido en el resultado, pero cuyos dos nodos finales sí, junto con su longitud.
- Hallamos la distancia mínima necesaria para llegar a esos nodos.
- Finalmente, añadimos todos los enlaces cuya longitud sea menor que esa distancia mínima restante tras llegar a cualquiera de los dos nodos.

Los enlaces encontrados hasta este momento son los incluidos completamente dentro del área de servicio, por lo que ya es posible intersecar los puntos de construcción con la geometría resultante.

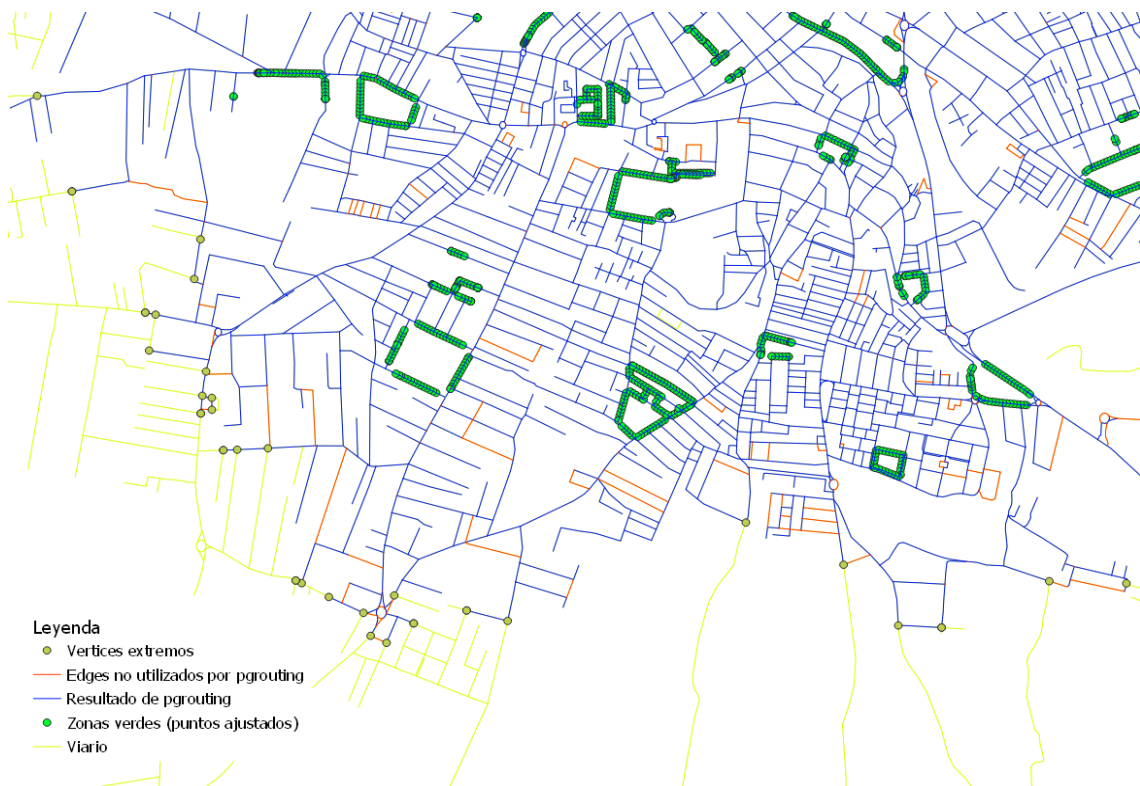


Figura 3 - Área de servicio hallada por `pgrouting`, enlaces no usados y últimos vértices alcanzados

Sin embargo, todavía hay puntos dentro de la distancia a partir de los últimos vértices alcanzados. El último paso consiste en crear un búfer alrededor de esos vértices e

intersecarlo con la capa de edificios y con los enlaces que comienzan en esos vértices. El radio de este búfer es diferente para cada vértice (resultado de restar a 833 m el coste necesario para llegar a ese vértice específico). Este proceso se realiza con la asistencia de Python.

Añadiendo los portales resultantes de este último paso tenemos el conjunto completo de edificaciones con acceso a zonas verdes. Esta tabla se enlaza posteriormente, en el entorno protegido del IECA, con los datos poblacionales sobre las personas que viven en cada portal segregadas por sexo y grupos de edad. También se incorpora a la tabla la variable municipio, donde se ubica el edificio. Esto permite agregar al ámbito municipal la población sin acceso a zonas verdes, desagregada por sexo y edad.

Cálculo iterativo de ruta óptima por edificio

Como se mencionó anteriormente, el enfoque opuesto (comenzar a definir las áreas de servicio desde los puntos de acceso al edificio) podría ser, en principio, no tan diferente en términos de la cantidad de puntos analizados y, por lo tanto, en tiempo de cálculo. De hecho, en Andalucía hay aproximadamente 2 millones de edificios en la capa analizada, y el total de puntos perimetrales de zonas verdes finalmente utilizados para el cálculo está cerca de esa cifra. Sin embargo, la estrategia es completamente diferente. No es el objetivo de este informe documentar este algoritmo alternativo, pero intentaremos esbozar cómo podría implementarse, ya que el diseño algorítmico es un poco más complejo.

1. Primero, no tiene sentido analizar sin pensar todos los portales secuencialmente. Aunque las funciones de pgrouting admiten llamadas 1-n (un punto de partida, muchos destinos), deberían estar asociadas jerárquicamente de alguna manera con las zonas verdes, para limitar la cantidad de puntos objetivo. La forma más fácil es recorrer secuencialmente todas las zonas verdes (al dividir las en puntos, se puede guardar la identificación original de la zona verde) y analizar su área de influencia.
2. Por definición, esta área de influencia podría implementarse como un búfer de 833 m alrededor de la zona verde. Para simplificar, esta zona de influencia podría crearse alrededor del centroide del polígono con

$$\text{radio} = 833 + \text{distancia entre el centroide y el punto más alejado.}$$

3. Una vez que se han localizado todos los edificios dentro de este búfer, el análisis de la red comienza desde el más lejano al más cercano, con todos los puntos de la zona verde como destino en la función `pgr_withpointsdd`.
4. Si la función tiene éxito (se llega hasta un punto perimetral), el portal se etiqueta como accesible y todos los edificios que se cruzan con la ruta seguida también se etiquetan de la misma forma. Esta última es la optimización clave de este algoritmo.
5. Las pruebas continúan analizando solo los portales que no están etiquetados, hasta que no haya más portales sin etiquetar en la zona de influencia. Por lo tanto, en cada iteración hay menos portales para probar.
6. El proceso continúa con la siguiente zona verde, probando solo portales sin etiquetar (ya que las áreas de influencia de distintas zonas verdes se pueden solapar).

La principal ventaja de este algoritmo es que el análisis de la red se realiza para pequeños lotes de puntos y un área limitada (y así tenemos la capacidad de limitar la red de calles utilizada), precisamente los dos ítems que aumentaban los tiempos de ejecución de pgrouting. Además, hay una optimización significativa al eliminar puntos durante la ejecución.

Los riesgos potenciales, en comparación con el enfoque centrado en áreas de servicio, se derivan del número de operaciones que deben repetirse muchas veces (del orden de 10,000 - 100,000 para Andalucía): ordenar puntos por distancia, ejecutar individualmente funciones de pgrouting (comparado con las aproximadamente 100 veces requeridas en el otro algoritmo), etc., lo que puede compensar el ahorro de tiempo.

Resultados

En el último paso del cálculo del indicador, se calcula la población que vive en los portales para cada municipio y se vincula a los datos obtenidos. Con esta información se han calculado 3 indicadores:

- Proporción total de población sin acceso a zonas verdes en su vecindario: $(\text{población total} - \text{población con acceso}) / \text{población total}$
- Proporción total de población mayor de 65 años sin acceso a áreas verdes en su vecindario: $(\text{total más de 65} - \text{más de 65 con acceso}) / \text{total más de 65}$
- Diferencia de género en el acceso: $[(\text{total muj.} - \text{muj. acceso}) / \text{total muj.}] - [(\text{total hom.} - \text{hom. acceso}) / \text{total hom.}]$

Los indicadores se han publicado en el portal del Sistema de Indicadores de Desarrollo Sostenible de Andalucía, en el apartado de Estadística Experimental:

<https://ods-municipios-andalucia.github.io/ods-municipios-andalucia/>

Conclusiones

La comparabilidad que se alcanza afrontando el problema para toda la Comunidad Autónoma desde un método unificado es muy valiosa, pero es costosa en cuanto a la capacidad y tiempo de cálculo requeridos. Por ello, y pensando en niveles superiores de comparabilidad, se hace necesario explorar otros algoritmos o herramientas más optimizados, como el método iterativo propuesto, u otros.

El resultado final se ha limitado a un indicador que describe el municipio, pero el material intermedio puede ser de gran utilidad para los procesos de decisión locales, al recopilar datos a nivel micro (calle o manzana). En este sentido, es necesario explorar formas de comunicación que permitan hacer llegar estos datos a los usuarios adecuados que faciliten su uso y comprensión.

Principales referencias bibliográficas

Copernicus Urban Atlas. [En línea] European Environment Agency .
<https://land.copernicus.eu/local/urban-atlas>.

Cormen, Thomas H., y otros. 2001. Section 24.3: Dijkstra's algorithm. *Introduction to Algorithms (Second ed.)*. s.l. : MIT Press and McGraw–Hill, 2001.

Nestar, Noelia. 2015 Evaluación de soluciones para el cálculo de rutas óptimas con bases de datos. Tesis de Fin de Grado. s.l. : Universidad de Valladolid.

pgrouting Manual v3.2.1. [En línea] <https://docs.pgrouting.org/latest/en/pgRouting-introduction.html>.

Poelman, Hugo. 2018. *A Walk to the Park? Assessing Access to Green Areas in Europe's Cities. Update Using Completed Copernicus Urban Atlas Data*. s.l. : European Commission Regional and Urban Policy, 2018.

Poelman, Hugo y Dijkstra, Lewis. 2015. *Measuring Access to Public Transport in European Cities*. s.l. : European Commission Regional and Urban Policy, 2015.

Referencia Transporte Público Estático (GTFS). [En línea] Google.
<https://developers.google.com/transit/gtfs/reference>.

Siragusa, Alice, y otros. 2020. *European Handbook for SDG Voluntary Local Reviews*. Luxembourg : Publications Office of the European Union, 2020.