



PROPUESTA PARA SEGMENTACIÓN DE VIVIENDAS

Censo Nacional de Hogares,
Población y Viviendas 2020

INDEC

Avances al 05/2019

Índice

| | |
|--|-----------|
| Restricciones identificadas por INDEC | 3 |
| Propuesta para segmentación de viviendas por orden secuencial de manzanas del radio | 5 |
| 1. Orden secuencial de manzanas del radio | 5 |
| Objetivo | 5 |
| Pasos para el orden secuencial de manzanas del radio: | 5 |
| Ejemplos | 7 |
| 2. Segmentación de viviendas por manzanas | 8 |
| Objetivo | 8 |
| Pasos de orden de lados manzana: | 8 |
| Manzanas extremas inicio-fin | 8 |
| Manzanas siguientes | 9 |
| Ejemplos | 11 |
| 3. Resultados esperados | 12 |
| Recursos | 14 |
| Spatial sort / fill | 14 |
| Propuestas matemáticas utilizadas para segmentación | 15 |
| Grids (celdas / superceldas) | 16 |
| Problema del cartero chino. Casos OSMnx y PGRouting, Grafos y CHPP (Chinese Postman Problem) | 17 |

Restricciones identificadas por INDEC

El siguiente es un resumen de las restricciones y exigencias por parte de INDEC a la hora de ordenar y segmentar las viviendas.

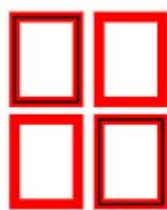
Ordenamiento: De norte a sur, de izquierda a derecha

Tamaño del segmento: 40 viviendas aprox. por segmento

Área a segmentar: radio censal urbano

Adyacencias:

Casos posibles de adyacencia o contigüidad



mzas completas
adyacentes



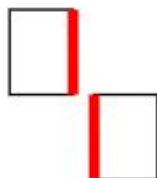
Lados adyacentes
de diferentes mzas



Lados adyacentes
de la misma mza

Casos no aceptados como adyacentes o contiguos

Son los casos en los que la proximidad es a través de un nodo común o se corresponde con lados de la misma manzana pero no respeta la condición de continuidad



mzas unidas por
el vértice



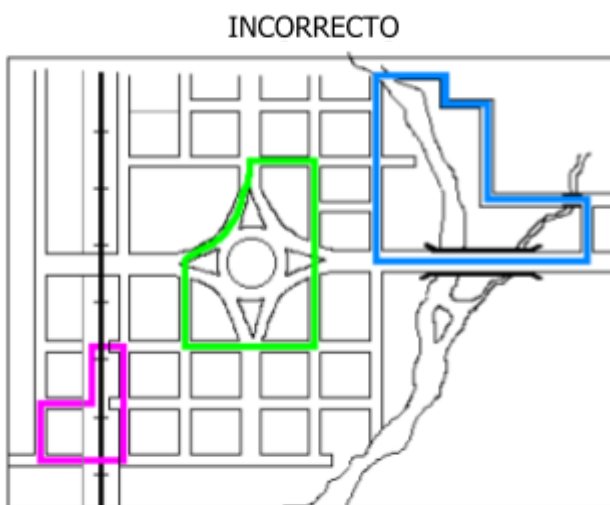
Lados no contiguos de
diferentes mzas



Lados no contiguos de la
misma mza

Cruces: en lo posible de no incluir un cruce de avenidas, rutas, vías de ferrocarril o cursos de agua para evitar una demora por obstáculos en el recorrido.

Evitar los recorridos discontinuos o con cruces en diagonal.



Segmentos completos y entre-segmentos:

Son segmentos completos los que estén:

- dentro de cada radio, por ningún motivo el segmento tomará parte de un radio y parte de otro
- dentro de cada manzana
- dentro de cada lado de la manzana o anexando lados completos
- con un edificio completo o anexando edificios completos
- dentro de un edificio, tomando piso completo o anexando pisos completos y continuos
- entre dos segmentos consecutivos de una manzana, no deberán quedar direcciones que no estén incluidas en alguno de ellos

Propuesta para segmentación de viviendas por orden secuencial de manzanas del radio

La propuesta de segmentación que nos propusimos investigar se compone de:

- Un orden secuencial de manzanas del radio para garantizar el cumplimiento de las restricciones en los recorridos.
- Una segmentación de viviendas siguiendo el orden secuencial de manzanas del radio, deconstruyendo el camino en una secuencia única lineal de viviendas agrupadas en grupos de 40 unidades.

1. Orden secuencial de manzanas del radio

Objetivo

Para la segmentación de viviendas creemos necesario definir un orden secuencial de manzanas del radio.

Esto es importante para garantizar el cumplimiento de las restricciones en los recorridos de los censistas, evitando huecos o viviendas “desconectadas”, es decir viviendas que no puedan ser asignadas, ya que de hacerlo, se estarían violando restricciones necesarias.

Pasos para el orden secuencial de manzanas del radio:

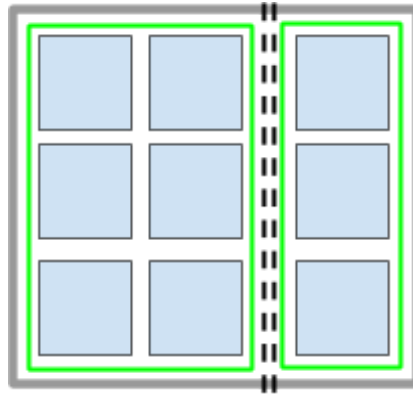
En una primera instancia se realizaron pruebas vectoriales sin grafos (sin transformar las geometrías a una estructura de grafos - nodos y vectores), utilizando la información original brindada por INDEC en formato de geometrías polígono, línea y punto.

La primera forma de encarar el orden secuencial fue con las herramientas PLSQL, SQL, PHP y las funciones espaciales de Postgis v2.3.

Para las representaciones visuales usamos la herramienta QGIS.

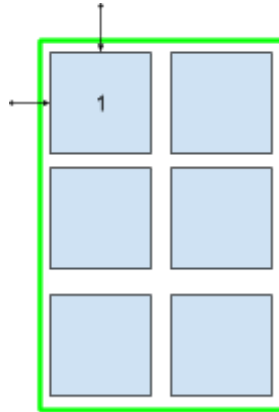
Pasos para el orden secuencial de manzanas del radio:

1. selección de manzanas por radio censal o subradio que no quiebre restricciones espaciales. Es decir:
 - a. Total de manzanas de un radio censal
 - b. Subconjunto de manzanas dentro del radio censal que no rompa restricciones. Por ej, si un radio es cruzado por una vía ferrocarril el radio se subdivide en 2 partes, siendo cada parte el conjunto de manzanas para el orden secuencial:



2. Búsqueda de la manzana inicial:

- a. Tomando el id de manzana(mzatxt) menor
- b. Tomando el min(geom)
- c. Contando las adyacencias por manzana entre manzanas del radio y ordenando por la que posee la menor cantidad de adyacencias.



3. Búsqueda de la siguiente manzana candidata.

Para continuar el recorrido es necesario encontrar la siguiente manzana a través de las manzanas candidatas.

Las manzanas candidatas poseen determinadas características, que una vez ponderadas, permiten identificar a la manzana siguiente. A saber:

Partiendo de la primera manzana:

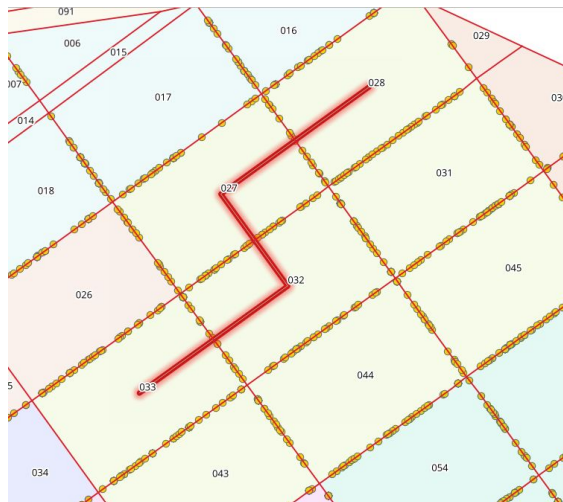
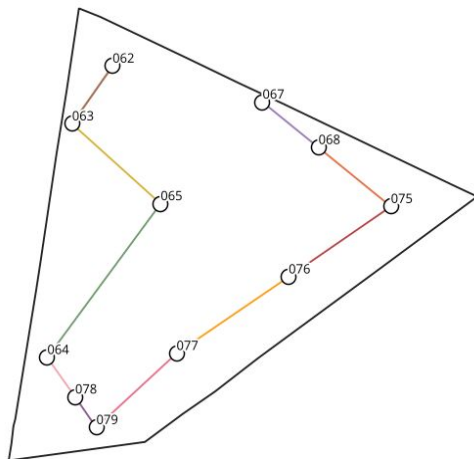
- a. Que no sea una manzana previamente asignada
- b. Adyacencia lindera contra primera manzana
- c. Adyacencia al *envelope* perimetral del subconjunto
- d. Id de manzana contiguo o más cercano (en los casos analizados en las pruebas, dentro del radio censal, los id de manzanas suelen ser contiguos y poseen un criterio de orden)
- e. Distancia entre centroides de la manzana inicial y la candidata
- f. Porcentaje de lado de manzana compartido por adyacencia
- g. Cantidad de lados adyacentes de la manzana candidata con sus candidatas (a mayor cantidad menos periférica es)

Al ponderar bajo una variable estos atributos podemos escoger la manzana siguiente candidata a través de la composición de un grupo de atributos.

De esta forma generamos el primer vector de recorrido entre manzana inicial (manzana 1) y su primera candidata (manzana 2).

Para poder completar el total del recorrido, simplemente es cuestión de iterar esta búsqueda y selección de manzana candidata siguiente hasta completar el total de manzanas del subconjunto.

Ejemplos



Array ([0] => 79 [1] => 77 [2] => 76 [3] => 75 [4] => 68 [5] => 67 [6] => 65 [7] => 62 [8] => 63 [9] => 64 [10] => 78)

RA-FR:0306

Array ([0] => 28 [1] => 27 [2] => 32 [3] => 33)

RA-FR:0108

Array ([0] => 57 [1] => 56 [2] => 55 [3] => 58)

RA-FR:0202

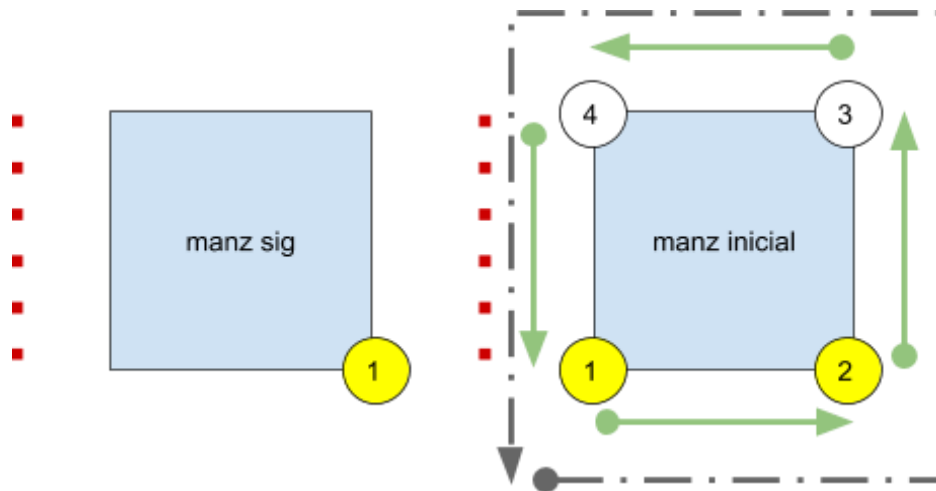
Array ([0] => 18 [1] => 19 [2] => 20 [3] => 29 [4] => 39 [5] => 68 [6] => 69 [7] => 75 [8] => 74 [9] => 70 [10] => 37 [11] => 31) Array ([0] => 31 [1] => 30) Array ([0] => 30 [1] => 38)

Importante

Este proceso puede ser optimizado utilizando grafos con el algoritmo del Chinese Postman Problem o problema del cartero chino.

Para poder ejecutar dicho algoritmo es necesario convertir todas las geometrías a una estructura de grafos (nodos, vértices y distancias).

- Preferentemente el punto de inicio debe ser el más cercano al boundary del radio.
- Preferentemente el punto de inicio debe tener como manzanas adyacentes a las más lejanas en el orden secuencial de manzanas del radio (mayor periferia)



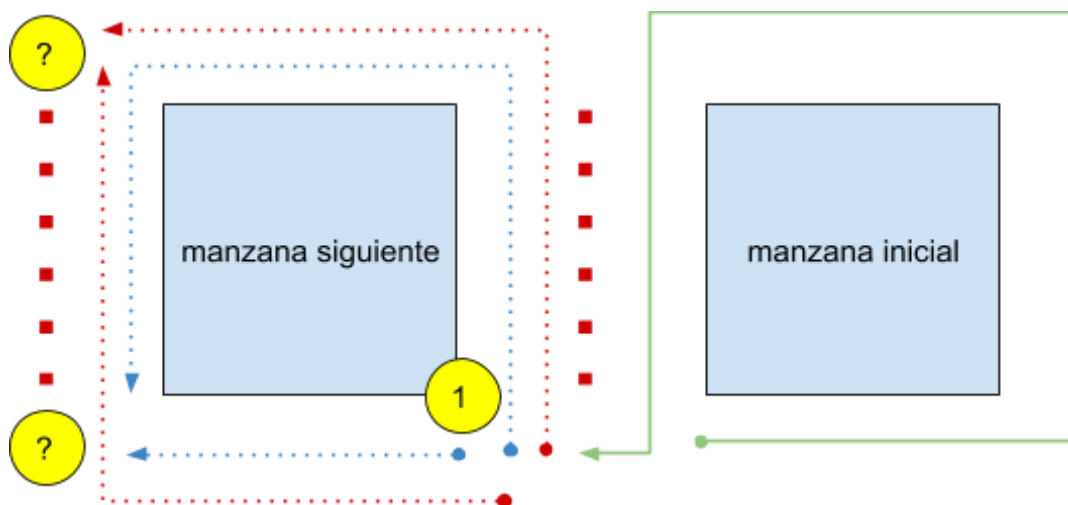
De esta forma obtenemos:

- el recorrido ordenado segmentado de la primera manzana.
- el punto de inicio para la siguiente manzana.

Manzanas siguientes

La manzana inicial posee un mismo punto de inicio-fin, pero para las manzanas siguientes los puntos de inicio fin deben ser diferentes.

El punto de inicio ya lo tenemos, porque va a ser el punto de fin de la manzana inicial. Pero para obtener el punto fin de la manzana siguiente, debemos resolver cuál de los puntos de la línea adyacente entre la manzana presente y la posterior (línea adyacente entre manzana2 y manzana3) va a ser el mejor candidato del punto final (Diag3).



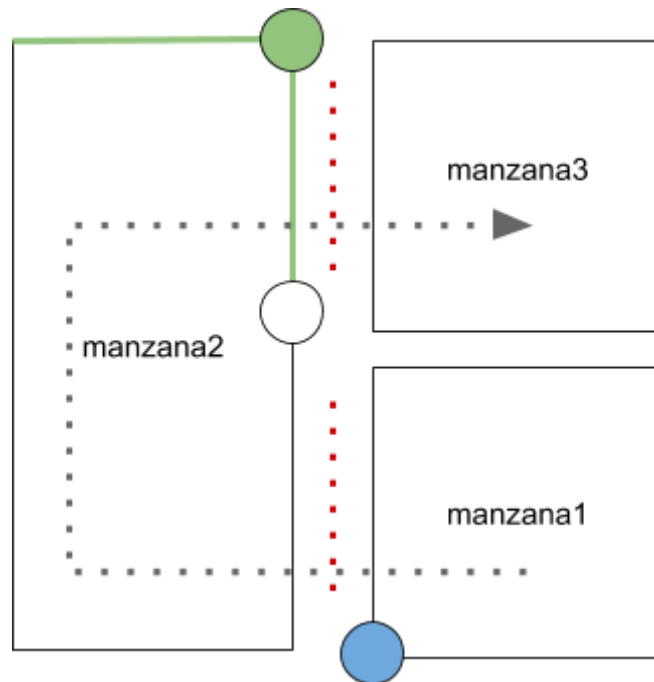
Para esto se pone en valor reglas que nos permiten elegir el punto fin candidato de la manzana siguiente. A saber:

- Preferentemente tener una distancia menor entre punto inicio y punto fin candidato.

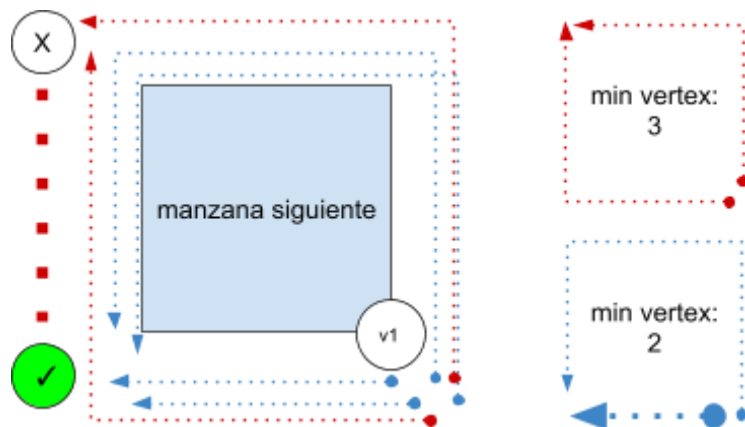
- Preferentemente tener la menor cantidad de vértices entre punto inicio y punto fin candidato.
- Tener como manzanas cercanas a las más lejanas siguiendo el orden secuencial de manzanas del radio (mayor periferia).
- Preferentemente no intersectar con otras manzanas, más que la anterior y la siguiente.
- Preferentemente que el punto final intersecte o esté más cercano contra el boundary del radio censal o el subgrupo de manzanas.
- Preferentemente ser esquina

Ejemplos

Punto final candidato sobre esquina boundary donde intersecta únicamente manzana2 y manzana3 (no manzana1)



Por mínima cantidad de vértices:



3. Resultados esperados

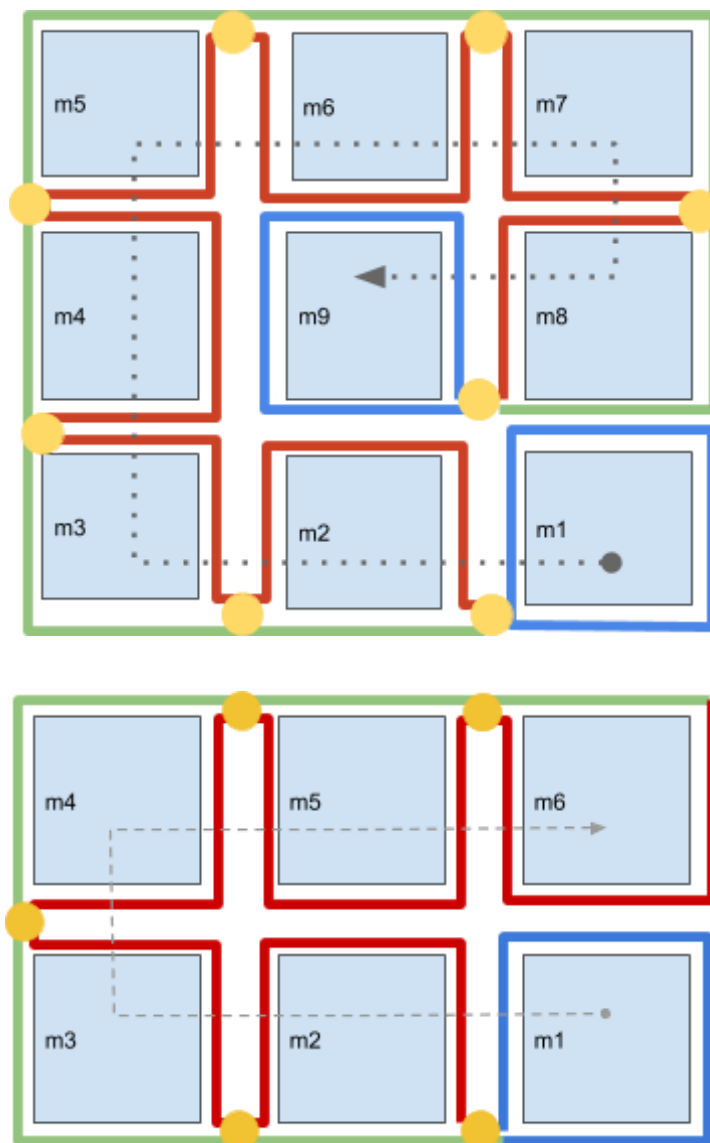
Una vez resuelto el punto final candidato generamos 2 paths.

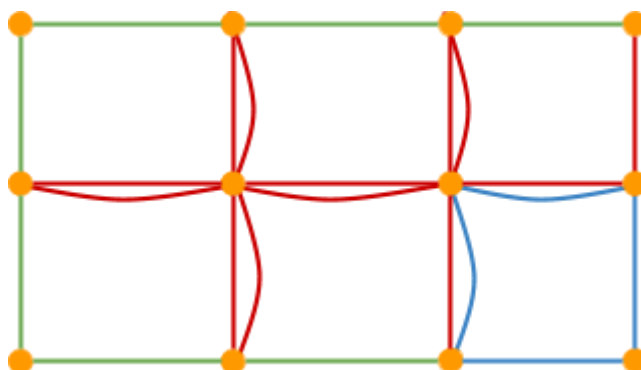
path 1: el de menor medida (length) y cantidad de vértices entre lado manzana inicio y lado manzana fin.

path 2: el de mayor medida (length) y cantidad de vértices entre lado manzana inicio y lado manzana fin.

De esta forma generamos:

- 2 manzanas extremos que necesitan auto-completarse a sí mismas al tener un mismo punto de inicio-fin (azul).
- 2 paths (verde y rojo) los cuales llegan al mismo punto fin de manzanas inicio-fin, donde un path acompaña la cara externa y el otro la cara interna y adyacencias.

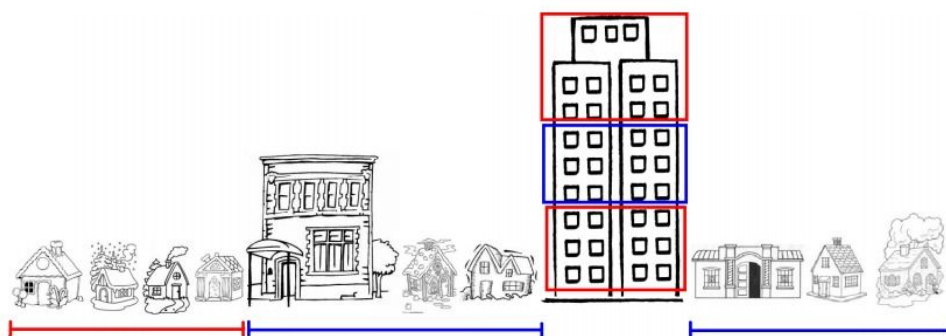




Generando así el recorrido lineal de líneas de manzana:



Esto permite agrupar las viviendas de forma secuencial hasta completar el cupo necesario:



Para la asignación por listado de viviendas es necesario aplicar una georreferenciación entre el listado de viviendas y la cartografía disponible. Esto permite identificar orden ascendente o descendente por segmento línea de manzana y ordenar adecuadamente el listado de viviendas.

De esta forma el listado se acomoda según las alturas más cercanas a cada vértice de los segmentos de líneas listando por menor o mayor altura siguiendo el orden secuencial.

Ej listado de viviendas FR-RA:0108 (google spreadsheet)

https://docs.google.com/spreadsheets/d/1JKORG6cHRuAyzNBSEZMX1wq6rRb0-TjdzQ_WXncyEI/edit?usp=sharing

Tener en cuenta:

HN es el número de puerta

H4 es el tipo de vivienda (tipos)

HP es el número de piso en un edificio

HD es el número de departamento en un edificio

Los buckets de censistas pueden ser acomodados por script.

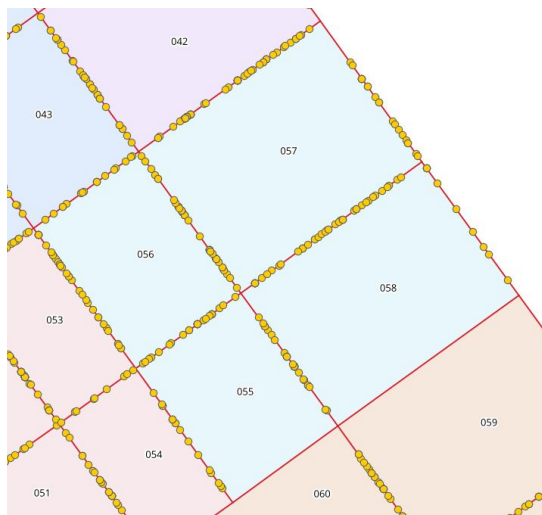
3. Resultados obtenidos

Ejemplo de segmentación FRACC01 RADIO08:

Listado de viviendas FR-RA:0108 (google spreadsheet)

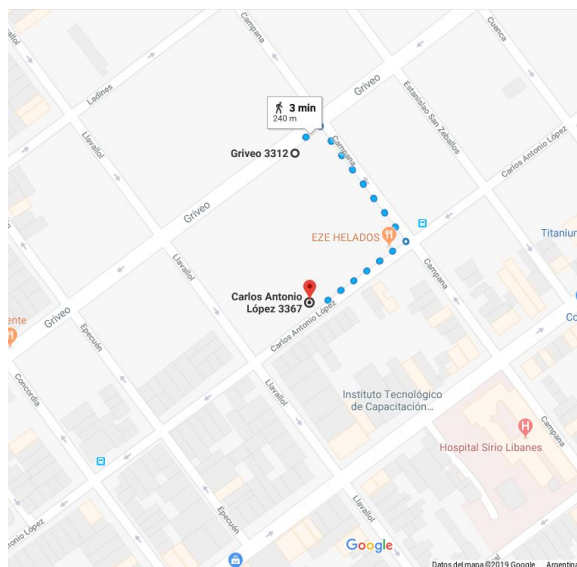
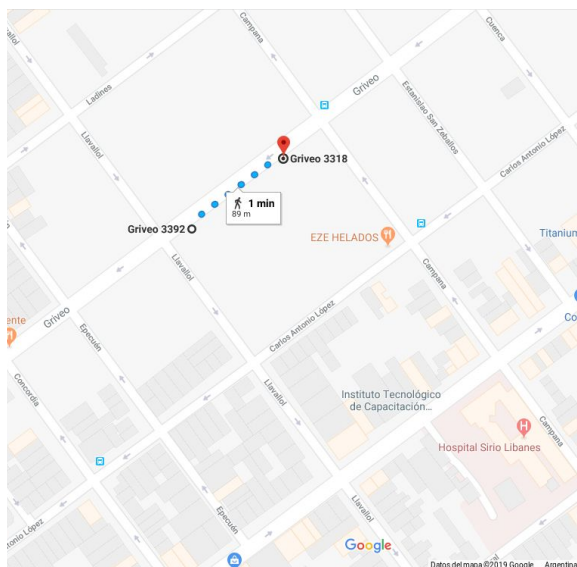
https://docs.google.com/spreadsheets/d/1JKORGC6cHRuAyzNBSEZMX1wq6rRb0-TjdzQ_WXncyEI/edit?usp=sharing

Manzanas FRACC01 RADIO08

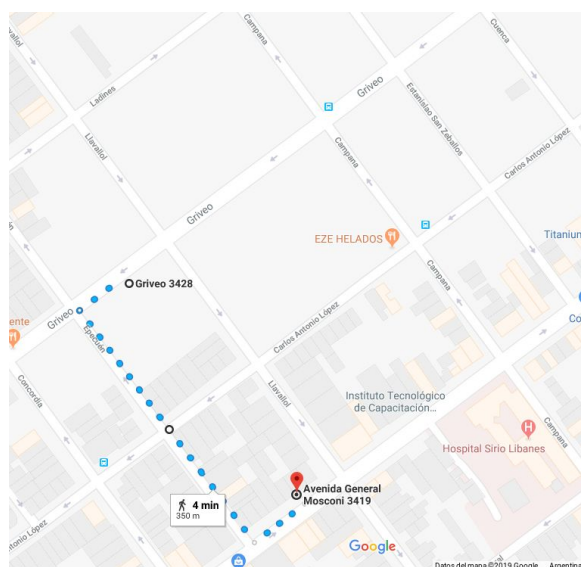
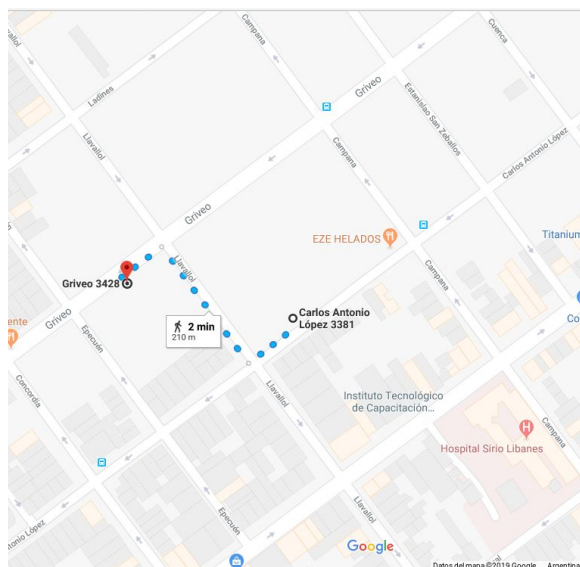


Recorrido por censista ordenado por recorrido:

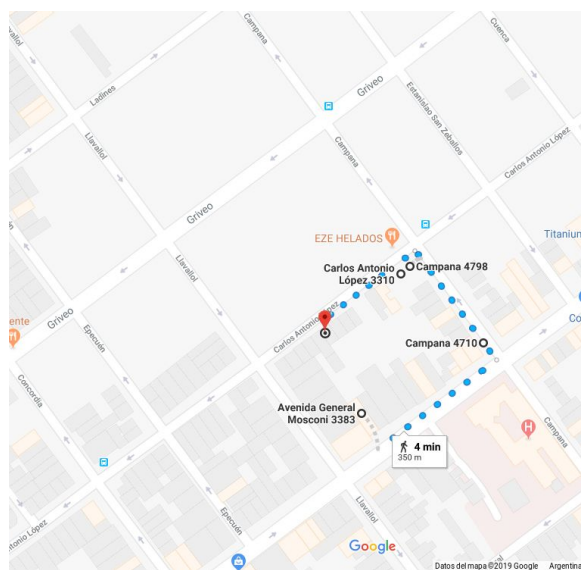
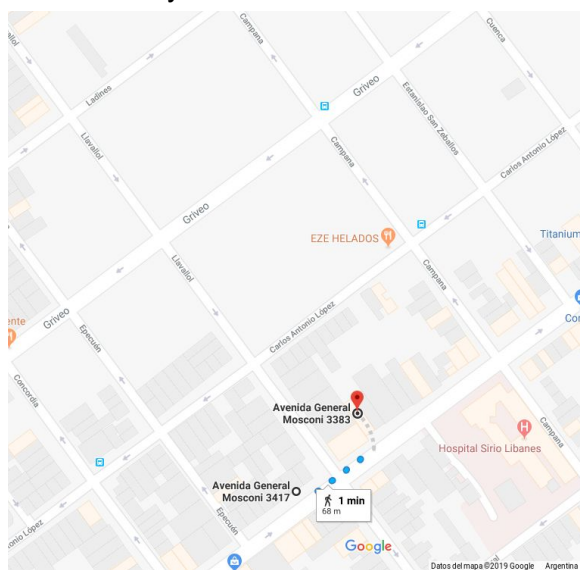
Censistas 1 y 2



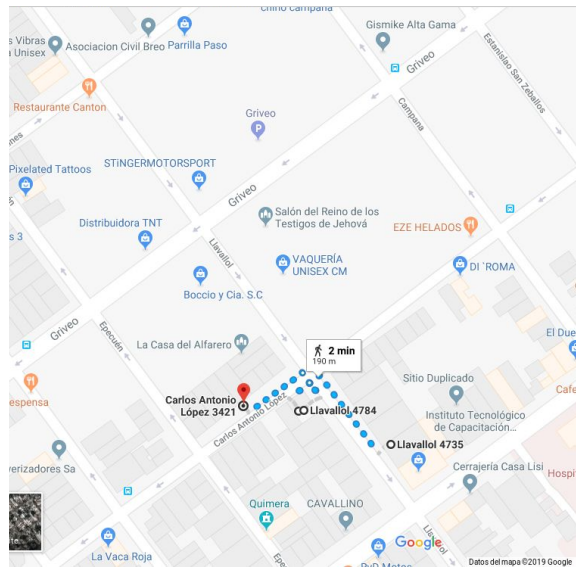
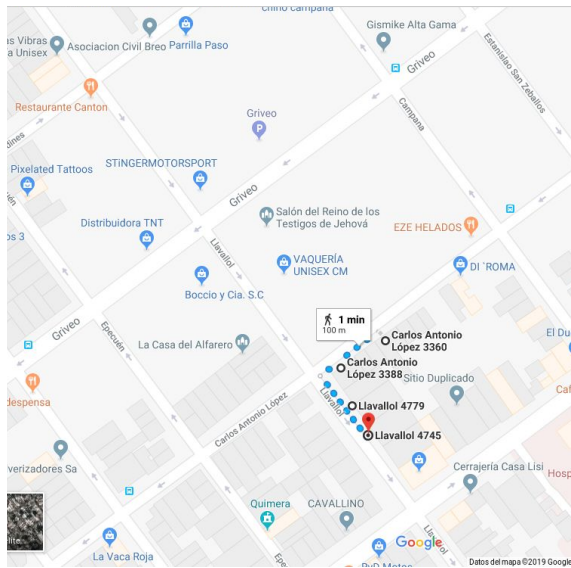
Censistas 3 y 4



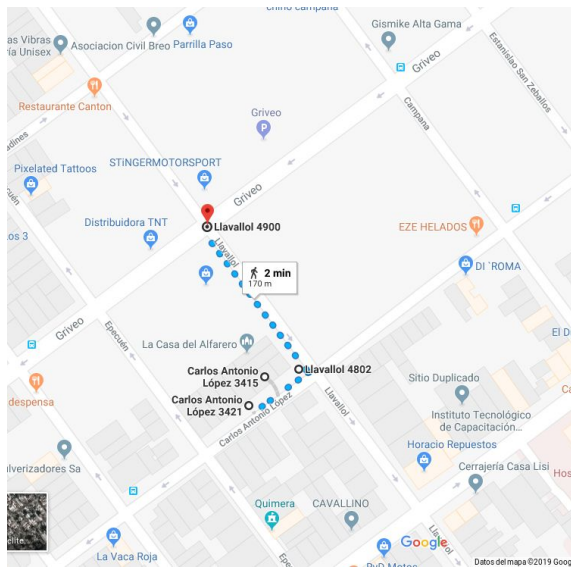
Censistas 5 y 6



Censistas 7 y 8



Censista 9



Actualización Junio-Julio 2019

Tras encontrarnos problemas con la funcionalidad exclusivamente geométrica o exclusivamente topológica, se decidió romper el problema en problemas más pequeños, y encontrar así un punto intermedio utilizando ambas funcionalidades dependiendo del problema.

Problema 1: recorrido de manzana.

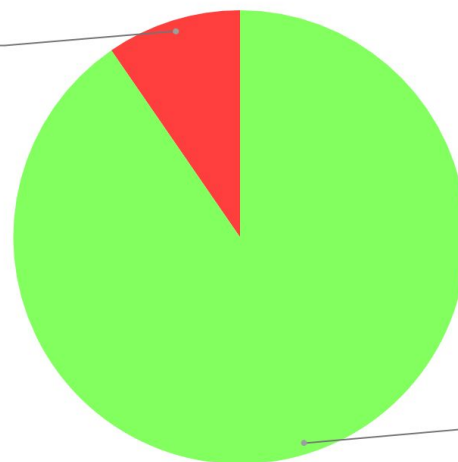
Para resolver este problema de recorrido de manzana se utilizan funciones geométricas.

To-do: radios con manzanas inner no boundaries:

Radios impactados

Radio con

9,6%



Radio solo

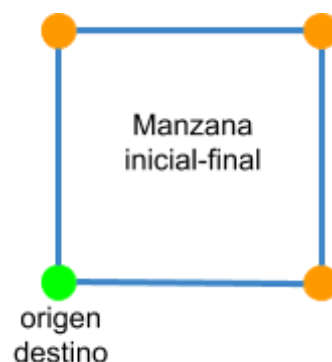
90,4%

Problema 2: Recorrido de manzanas de mismo punto inicial - final.

Se identifica el nodos de inicio y fin de recorrido.

Para este tipo de manzanas inicio-fin se utiliza la función de topología pgr_TSP, pgr_dijkstraCostMatrix, es decir un aproximamiento TravelSalesmanProblem para estas manzanas.

El resultado nos devuelve un único camino:

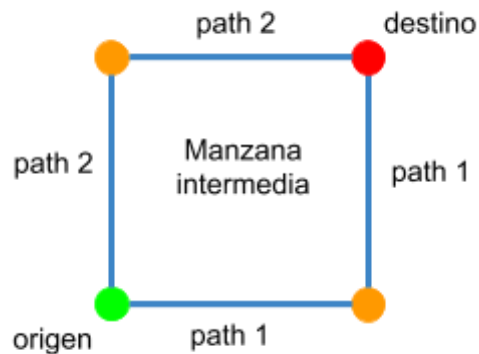


Problema 3: Recorrido de manzanas medias con punto inicial - final diferentes.

Se identifican los nodos de inicio y fin del recorrido.

Para este tipo de manzanas se utiliza la función de topología pgr_ksp es decir K shortest.

El resultado de este problema devuelve 2 caminos:



Problema 4: Orden del recorrido.

Una vez obtenido los recorridos manzanas inicial - medias - final, se debe ordenar los mismos para que se complete como un recorrido lineal único de censistas.

1. Manzana inicial
2. Manzanas medias, camino 1 (más corto)
3. Manzana final
4. camino inverso de Manzanas medias, camino 2

Problema 5: Segmentación por censista cada 30 viviendas

Al resultado final, se le aplicó la función NTILE utilizando como valor, la máxima cantidad de registros (viviendas) / 30, rompiendo así el total de registros en grupos de 30 viviendas.

Observaciones:

Para los segmentos de línea que no se encontraron domicilios georreferenciados, el registro figura como un único registro que contempla el total de alturas de dicho segmento de línea, y es obligatorio para el censista recorrer dicho segmento de línea por más que no existan domicilios georreferenciados.

Recursos

Spatial sort / fill

Spatial Sort: Morton Key:

http://www.spatialdbadvisor.com/oracle_spatial_tips_tricks/138/spatial-sorting-of-data-via-morton-key

PG Morton:

https://en.wikipedia.org/wiki/Z-order_curve,

https://en.wikipedia.org/wiki/Matrix_representation, <https://github.com/Oslandia/pgmorton>

Gridding / fishnets:

https://spatialdbadvisor.com/postgis_tips_tricks/258/gridding-a-geo-object-postgis,

https://spatialdbadvisor.com/postgis_tips_tricks/300/generating-a-grid-fishnet-of-points-or-polygons-for-postgis

Space filling curve:

https://en.wikipedia.org/wiki/Space-filling_curve

Hilbert curve:

https://en.wikipedia.org/wiki/Hilbert_curve

<https://books.google.com.ar/books?id=WtZTDwAAQBAJ&pg=PA540>

PGRouting TSP:

<https://docs.pgrouting.org/2.0/es/src/tsp/doc/index.html>

OSMNx:

<https://github.com/gboeing/osmnx>

Python for Street Networks:

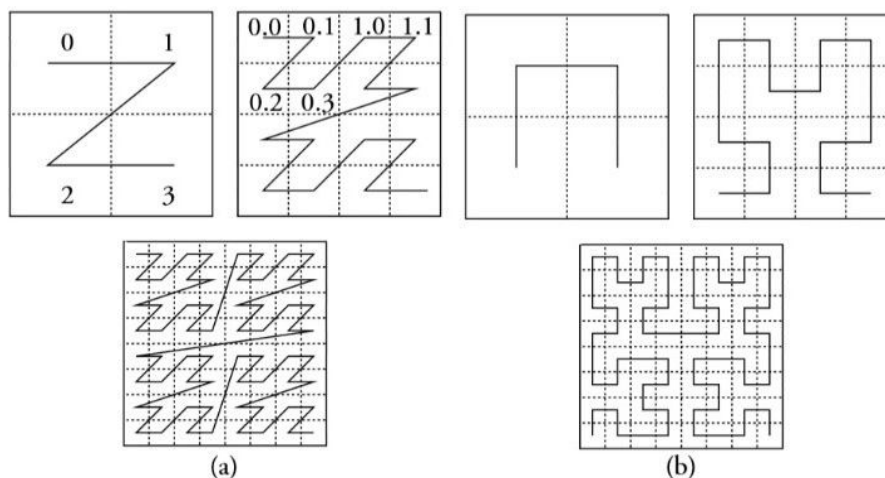
<https://geoffboeing.com/2016/11/osmnx-python-street-networks/>

Solving the Chinese Postman Problem:

<http://brooksandrew.github.io/simpleblog/articles/intro-to-graph-optimization-solving-cpp/>

graph optimization:

<http://brooksandrew.github.io/simpleblog/articles/intro-to-graph-optimization-solving-cpp/>



Propuestas matemáticas encontradas

Segmentación matemática utilizada para censo 2010

http://www.dm.uba.ar/materias/investigacion_operativa/2010/2/slides.pdf

<http://www.dii.uchile.cl/~ris/RISXXV/censo.pdf>

<https://www.improntait.com/en/novedades-en/gis-para-censo-nacional-de-poblacion-hogares-y-viviendas>

<https://usuarios.fceia.unr.edu.ar/~nasini/Censo.Interfaces.pdf>

Tesis de Florencia Fernández Slezak analizando caso de estudio: "Programación Matemática para distribución eficiente de censistas en censos nacionales"

http://mate.dm.uba.ar/~gduran/docs/tesis/tesis_Florencia.pdf

Optimización de recorridos en ciudades. Una aplicación al sistema de recolección de residuos sólidos urbanos en el Municipio de Concordia

<https://dcc.fceia.unr.edu.ar/sites/default/files/tesinas/77.pdf>

Sistema de Información Geográfica como apoyo a las actividades del Censo Nacional de Población, Hogares y Viviendas 2010 de la provincia de Buenos Aires

<http://40jaiio.sadio.org.ar/sites/default/files/T2011/JUL/565.pdf>

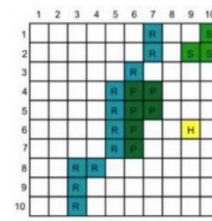
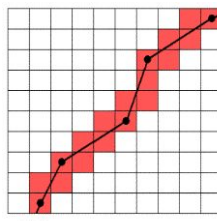
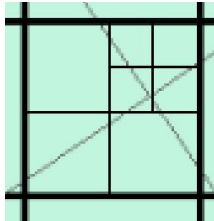
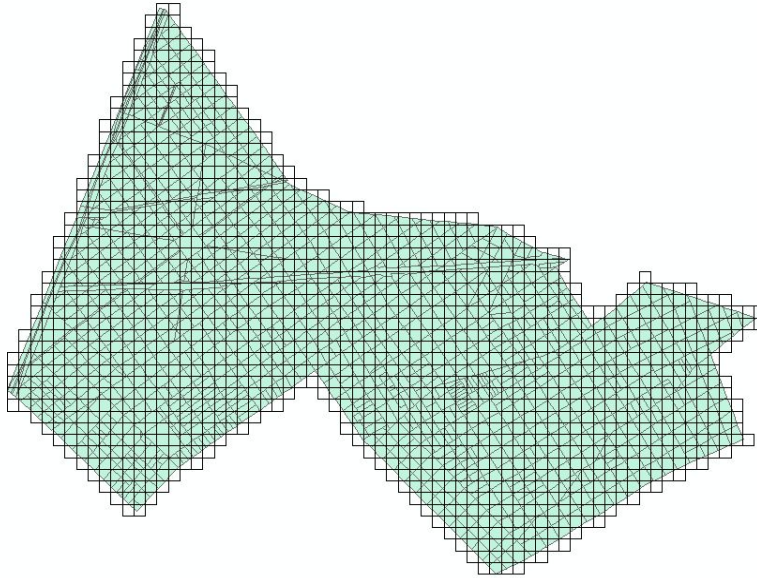
Estructuras geográficas y mecanismos de ordenamiento espacial para la definición de recorridos (Aguas Argentinas)

<https://cumincad.architexturez.net/system/files/pdf/8afc.content.pdf>

Grids (celdas / superceldas)

Realizamos pruebas con barridos por grids. Los mismos:

- Permiten un ordenamiento norte surte, izquierda derecha, cubriendo así el total parcelario por radio censal. (order by grid, col , row, geom desc)
- Permite identificar y asignar atributos a los grids: por ejemplo para contemplar restricciones contenidas en dicho gris (cursos de agua, vías de ferrocarril, rutas). Esto podría generar un corte en la asignación de censistas de dicho grid.
- Permite generar los subgrids dentro de un grid, necesarios hasta identificar las unidades manzana y seguir con un barrido.
- Se puede contemplar la cantidad de viviendas por grid / subgrids hasta obtener el valor necesario e ir sumando manzanas adyacentes, de forma ordenada y secuencial, evitando huecos.



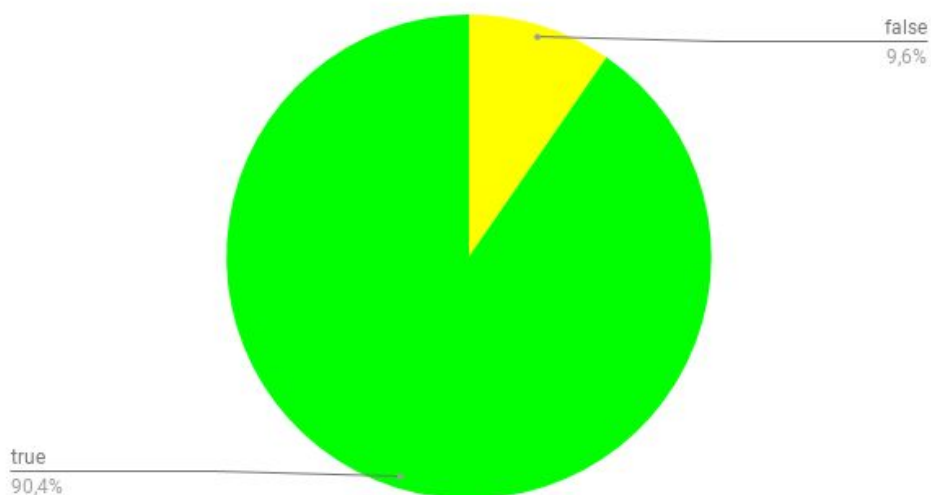
Actualización Mayo 2019

- Resultados positivos para recorridos de manzanas en radios únicamente con manzanas boundaries sin manzanas internas.
- Resultados zonagris para recorridos de manzanas en radios que incluyen manzanas internas sin tocar boundaries.

% Radios únicamente manzanas boundaries

Prov02, Depto11, Codloc10
230 radios , 1245 manzanas

Recuento de radio_manzanas_todas_borde_bool



Alternativa para el recorrido: topología con PGRouting

Casos:

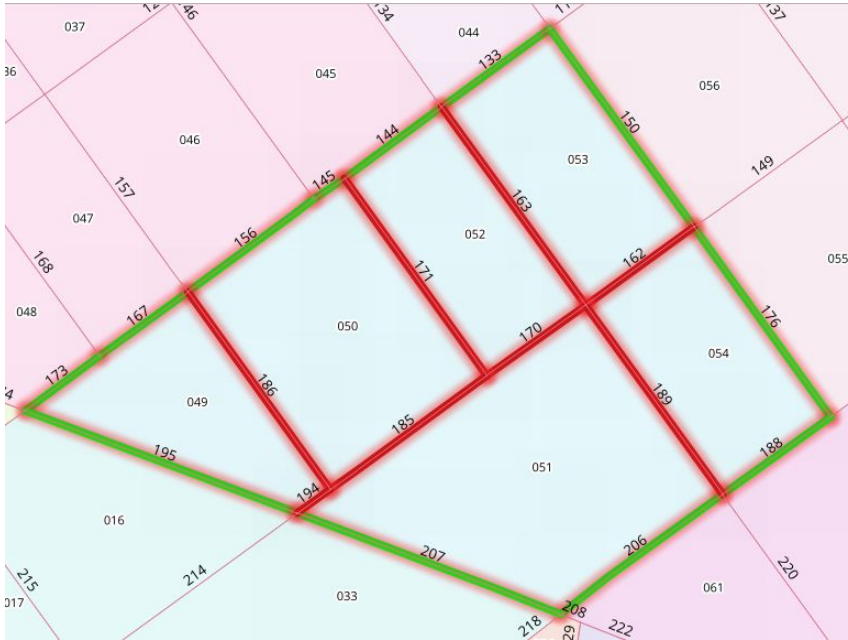
1. Por líneas de manzana y nodos de interseccion entre manzanas:
 - a. camino largo (vuelta a las líneas de manzana)
 - b. camino corto



2. Por borde externo y conjunto de líneas internas

a. Borde externo (pgDijkstra)

b. líneas internas (ChPP)



Porqué no usar ChPP en todos los radios?

<https://docs.pgrouting.org/dev/en/chinesePostmanProblem-family.html>

- There is no path when the graph is not connected.
- We will return **no path found** error if there is no path.

Docs.

Problema del cartero chino. Casos OSMnx y PGRouting, Grafos y CHPP (Chinese Postman Problem)

https://es.m.wikipedia.org/wiki/Problema_del_cartero_chino

<https://arxiv.org/pdf/1611.01890.pdf>

<https://docs.pgrouting.org/dev/en/chinesePostmanProblem-family.html>

https://docs.pgrouting.org/dev/en/pgr_directedChPP.html

Problema del cartero rural a la recolección de residuos reciclables

<http://www.dii.uchile.cl/~ris/RIS2015/residuos.pdf>

Se realizaron pruebas a través del función ChPP disponible en la última versión experimental dev de la extensión PGRouting 3.0.0-dev:

<https://github.com/pgRouting/pgrouting/wiki/GSoC-2018-Minimum-cost-flow-and-ChPP>