

Métodos de Visualización de Datos Georeferenciados: Aplicación

Diego Correa Tristain

email: algoritmia@labormedia.cl

encomendado por :

Guillermo Acuña Arquitectos Asociados

conceptos claves: validación de datos, matriz de adyacencia, lista de adyacencia, clustering de markov

Resumen-- Se propone una forma de aplicación que permita analizar y visualizar la información de red contenida en tablas de información georeferenciadas.

I. Introducción

Para el propósito de analizar y visualizar extensas y complejas tablas de información, es necesario identificar las etapas que una aplicación debe contener para cumplir las diversas tareas enlazadas y de esa manera lograr una visualización en base a una interacción adecuada. Esto, a su vez, producirá contenido apto para ser usado en la conversación entre distintos estamentos, y son la base para la comprensión conjunta de tareas con un alto grado de complejidad, procesos que llevan tiempo no sólo de realización sino también de reflexión.

II. Validación de Información

Un aspecto primario en el análisis de datos es la validación de la información, tanto en el formato de su contenido como en el contenido en sí mismo. En este caso, desearemos que las tablas de información contengan información relevante para construir redes de situaciones, redes de Markov probabilísticas en dos formas de tablas: lista de adyacencia y matriz de adyacencia.

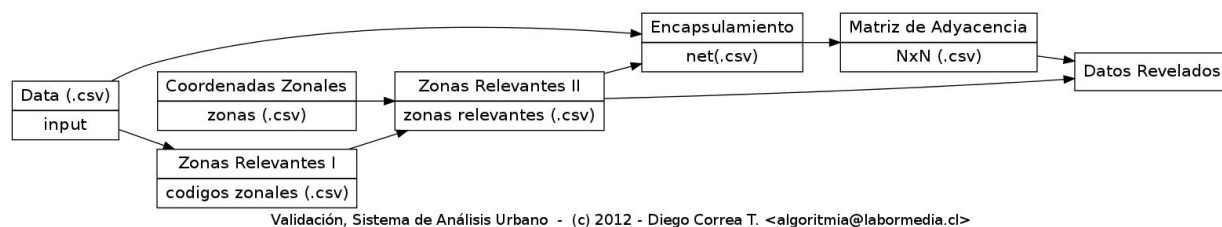
En nuestro ejemplo, la tabla original vendrá en algún formato de datos con separación de tal manera que podrá ser revisado como una planilla para cálculo. Los tipos de datos

genéricos para este tipo de transferencia son datos de caracteres separados por un carácter arbitrario, que puede ser una coma así como un carácter tipo tab.

Nuestro ejemplo considera la información de la encuesta de transporte origen a destino diseñada para el sistema de información de transporte (SINTIA), a cargo de la Secretaría de Planificación de Transporte (SECTRA) del Ministerio de Transporte y Telecomunicaciones (MTT)¹.

La encuesta consiste en preguntas y respuestas realizadas en periodos temporales para zonas de la ciudad de Santiago. De esta manera se obtiene un panorama de qué viajes se realizan, desde donde, hacia donde y está tabulada en base a criterios, como por ejemplo el propósito de viaje.

Las tablas de información disponibles en este sistema son entregadas como lista de adyacencias, vale decir, por cada zona territorial de la encuesta existe una cierta cantidad de datos tanto de origen como de destino y la tabla relaciona el origen con el destino junto con el tiempo y el propósito, como ejemplo.



III. Análisis.

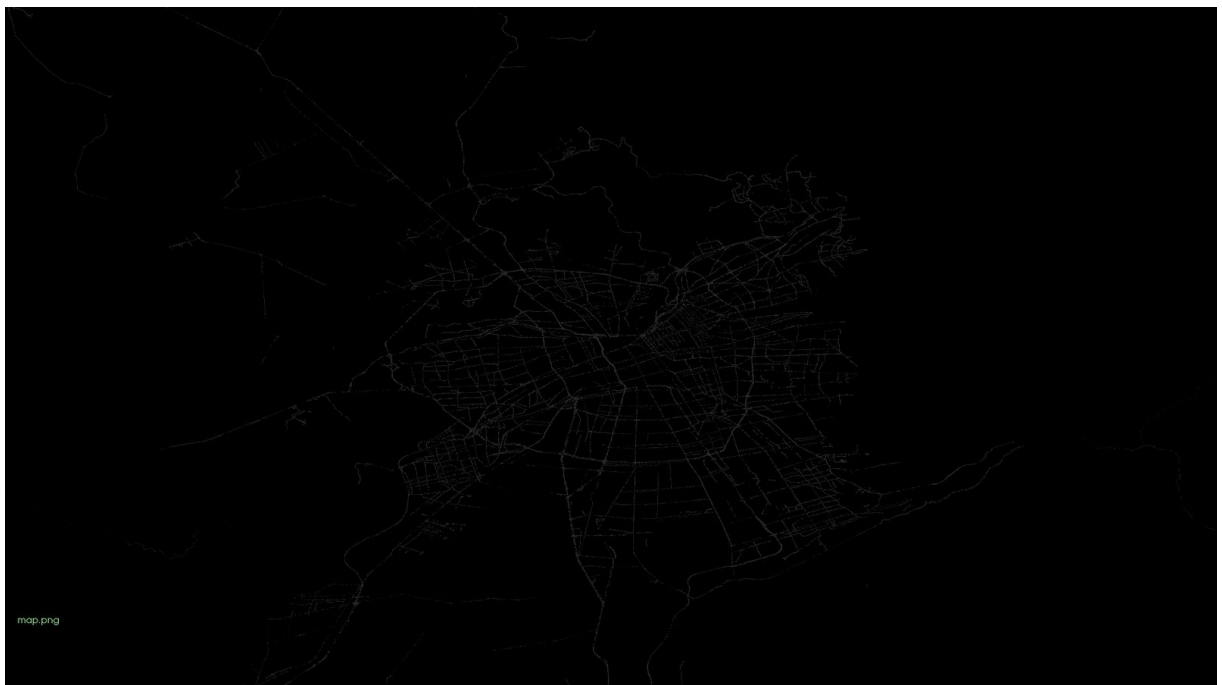
Para el análisis de la información en base al método de clustering propuesto es necesario transformar la información contenida en las tablas de adyacencia entregadas por el sistema de información de transporte a matrices de adyacencia. Estas consisten en matrices cuyas filas y columnas coinciden en su información zonal, tanto en el orden como en las zonas en si mismas. De esta manera obtenemos una matriz cuadrada NxN, donde N es la cantidad de zonas contenidas en la lista de adyacencia original.

El método de análisis de Markov identifica ciclos en la red y los ordena en base a su intensidad de interconexión. Esto implica que la red este autoreferenciada a si misma en cada uno de sus nodos, de tal forma que todos los nodos de ella pertenezcan inicialmente a un ciclo dentro de la red. Esto se traduce en la eliminación de las filas nulas de la matriz de adyacencia, i.e. los nodos cuyas filas no contienen información de destino.

En base este procedimiento de validación es entonces coherente iniciar el método de análisis en base a tres parámetros: cantidad de secuencias, cantidad de iteraciones e inflación. La cantidad de secuencias son las veces en que se iterará para cada resultado, la cantidad de iteraciones son la cantidad de veces en que el método iterará por cada secuencia y la inflación es una medida de suavidad en la cual el sistema es decrecido en base a una curva cuadrática calculada como el cuadrado de la probabilidad. Esto nos da diversos resultados para diversos parámetros, los cuales incidirán en la cantidad de clusters identificados mediante el análisis. La heurística de este proceso dependerá del propósito del análisis en base a los criterios antes mencionados. Por ejemplo, si se requiere un resultado de mayor interconexión de la red completa, entonces la medida de inflación tendrá un mayor valor.

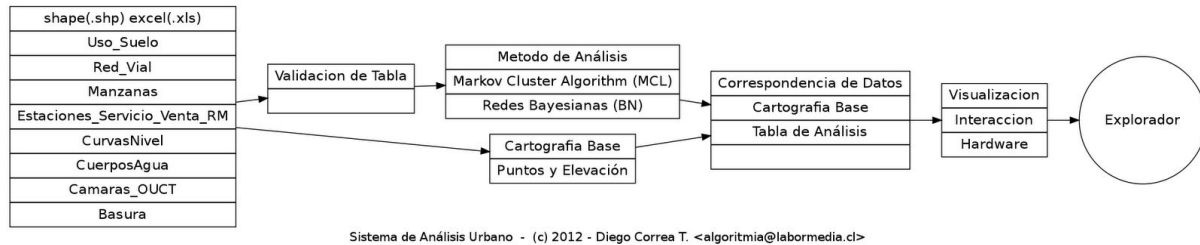
IV. Cartografía

Para la representación gráfica de la información, es pertinente señalar mediante puntos de referencia la ubicación de la información tanto contenidas en las tablas iniciales como las que constituyen parte del análisis. Para este propósito, para cada zona de origen y destino se asocia un centroide de esa zona georeferenciada y se sobrepone las líneas paramétricas de las tablas mediante los centroides de las zonas, sobre una carta básica que sirva como referencia dentro del espacio cartesiano en el cual esta información se representa.



Cartografía base para la referencia de los datos representados de la tabla.

V. Visualización e Interacción



La síntesis del diseño general, entonces, se compondrá de los siguientes aspectos:

Validación de tabla

Método de Análisis

Cartografía Base

Correspondencia de datos: Análisis + Cartografía.

Visualización: Interacción en base al hardware a utilizar.

¹ <http://sintia.sectra.gob.cl/>