

# Aplicaciones de técnicas de ATD con datos del INEGI

José de Jesús Aguirre-Tepole <sup>1</sup>  
Angel Emilio de León-Gutiérrez <sup>2</sup>

## Resumen

En el marco de la Tercera Escuela de Análisis Topológico de Datos y Topología Estocástica que se llevó a cabo del 22 al 28 de Enero de 2017 en las instalaciones de ABACUS en el Estado de México, uno de los cursos que se ofrecieron fue “Network Data Analysis: Metrics and Persistent Homology”. El presente reporte muestra el uso de las técnicas expuestas utilizando información económica de México.

*2010 Mathematics Subject Classification: Clasificación. 55N35, 68U05, 62P20*

*Keywords and phrases: TDA, datos económicos*

## 1 Aspectos teóricos

El objeto de trabajo del curso fueron las redes [CM17a]. Una red es un objeto que muestra las relaciones con y entre los datos y que también puede verse como parte de la teoría de gráficas. Definiendo precisamente, tenemos:

**Definición 1.1.** Sea  $X$  un espacio topológico numerable y sea  $\omega_X$  una función continua de  $X \times X$  (con la topología del producto) a  $\mathbb{R}$ . Una red es el par  $(X; \omega_X)$  y se denotará por  $\mathcal{N}$  la colección de todas las redes.

---

<sup>1</sup>tepole@esfm.ipn.mx.

<sup>2</sup>angel.deleon@banxico.org.mx.

Dada una red  $(X; \omega_X)$ , se refiere a los punto de  $X$  como *nodos* y  $\omega_X$  es la función de peso de  $X$ . Nodos que se encuentren conectados pueden ser referidos como *aristas* o *bordes*. Dado  $A \subset X$ , representamos por  $(A, \omega_X|_{A \times A})$  la subred de  $X$  inducida por  $A$ . En los ejemplos se trabajó con redes finitas, las cuales son representadas como

$$\mathcal{FN} := \{(X, \omega_X) : X \text{ discreto}, \omega_X : X \times X \rightarrow \mathbb{R} \text{ cualquier mapeo}\}.$$

Considerando que las redes en el mundo real son finitas y susceptibles a ser utilizadas en algoritmos computacionales; se considera además, la colección de *redes compactas*, las cuales son redes  $(X; \omega_X)$  con la restricción adicional de que  $X$  sea *compacto*. Explícitamente, se usa la notación  $\mathcal{CN}$  para el conjunto

$$\{(X; \omega_X) : X \text{ numerable y compacto}, \omega_X : X \times X \rightarrow \mathbb{R} \text{ continua}\}.$$

Debido a que el enfoque del curso fue redes y homología persistente, recordamos que un *complejo simplicial*  $K$  con vértices en  $V$  es  $K \subset \mathcal{P}(V)$  tal que

$$\tau \subseteq \sigma \in K \implies \tau \in K.$$

Una *filtración* de un complejo simplicial  $K$  es una secuencia de complejos  $K_i$  tal que

$$\emptyset = K_0 \subset K_1 \subset \dots \subset K_n = K$$

Además, el complejo de *Rips* [CM16] para un espacio métrico  $(X, d_X)$  está definido para cada  $\delta \geq 0$  tal que

$$R(X, \delta) := \{\sigma \in \mathcal{P}(X) : \text{diam}(\sigma) \leq \delta\} \text{ con } \text{diam}(\sigma) := \max_{x, x' \in \sigma} d(x, x')$$

Mientras que el complejo de *Čech* (también conocido como Dowker) para un conjunto de puntos  $S$  está definido para  $r > 0$  como la cubierta de bolas cerradas con centro  $x \in S$  y radio  $r$ , es decir,  $\check{C}(S, r) = \{\sigma \subseteq S \mid \bigcap_{s \in \sigma} \overline{B(s, r)} \neq \emptyset\}$ .

## 2 Aplicación

El conjunto de datos que se utilizo en el curso fue el de 15 sectores económicos de Estados Unidos en el periodo 1997-2015 simbolizados como 19 redes que representan el flujo de productos a través de estos sectores.

Como ejercicio, nosotros decidimos replicar ese ejemplo pero con datos de México. Buscamos en el *INEGI* [Ine17] información y encontramos la matriz insumo producto con año base 2008. En dicha matriz se encuentran los sectores primarios (2) de la economía en millones de pesos. Con estos datos calculamos la homología persistente a partir de la filtración de Rips y Dowker. Para calcular la homología utilizamos las funciones del paquete *PersNet*<sup>3</sup> [CM17b] creado por los instructores. Dicho paquete se ejecuta en el ambiente de *Matlab* y requiere que esté instalado *Javaplex*<sup>4</sup>.

Antes de realizar el cálculo de la homología, se normalizaron los datos, tomando en cuenta las recomendaciones dadas en la sesión. Se exportaron los datos a *Matlab* generando una matriz y un vector de etiquetas. Después de esto se ejecutaron las funciones correspondientes para que la matriz sea vista como un conjunto de redes y dependiendo de la filtración seleccionada, obtener los códigos de barras correspondientes. En los instrucciones 1 y 2 se muestran las instrucciones correspondientes dependiendo de la filtración, y en las tablas 3 y 4 se observan los resultados para las componentes conectadas ( $H_0$ ) y los ciclos ( $H_1$ ) de cada una de las filtraciones.

En particular podemos observar que en el caso de  $H_0$ , las únicas componentes en las que coinciden los cálculos con ambas filtraciones son las mostradas en la Tabla 1, lo cual consideramos que se puede interpretar como una confirmación de la estrecha relación entre dichos sectores. Los ciclos obtenidos por ambas filtraciones son distintos. En el caso de la filtración de Rips obtenemos el ciclo en la Figura 1 (que puede verse como un cuadrado). Mientras que con la filtración de Dowker obtenemos el ciclo en la Figura 2, en el que participan cuatro sectores (puede verse como dos triángulos que comparten un lado).



Figura 1: Resultado con la filtración de Rips

<sup>3</sup><https://research.math.osu.edu/networks/Datasets.html>

<sup>4</sup><http://appliedtopology.github.io/javaplex/>

Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	Servicios científicos, técnicos y profesionales
Minería	aprovechamiento forestal, pesca y caza, crianza y explotación de animales, agricultura
Información en medios masivos	Servicios financieros y de seguros
Transportes, correos y almacenamiento	Industrias Manufactureras

Tabla 1: Componentes comunes a las dos filtraciones

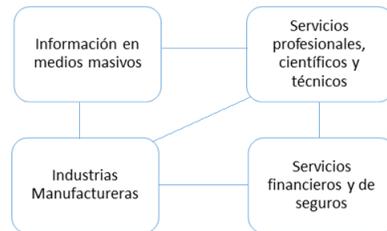


Figura 2: Resultado con la filtración de Dowker

Llama la atención los sectores involucrados en ambos ciclos, y parte de un análisis posterior es verificar si dichos sectores se relacionaron como productor y/o consumidor en el año de los datos, debido a que esas relaciones de primera instancia no son habituales y que justamente esto hace resaltar la importancia de estas técnicas que posiblemente permitan encontrar información que por los métodos tradicionales no surge de forma habitual.

### 3 Apéndice

Listing 1: Persistencia con filtración Rips

```

>> % Instrucciones para obtener la                                1
>> % homología '{i}a persistente                                 2
>> % en Matlab con Persnet y JavaPlex                             3
>> load('inegi.mat')                                           4
>> diam=max(max(inegi))                                         5
>> [sk0,sk1,sk2]=rips(inegi)                                     6
>> computePers(sk0,sk1,sk2,[],diam)                             7
  
```

Listing 2: Persistencia con filtración Dowker

```

>> % Instrucciones para obtener la                               1
>> % homolog\ '{i}a persistente                                 2
>> % en Matlab con Persnet y JavaPlex                           3
>> load('inegi.mat')                                           4
>> diam=max(max(inegi))                                         5
>> [sk0,sk1,sk2]=dowker(inegi)                                   6
>> computePers(sk0,sk1,sk2,[],diam)                             7

```

Sectores económicos
<p>Aprovechamiento forestal, pesca y caza, agricultura, cría y explotación de animales,  Minería.  Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final.  Construcción.  Industrias Manufactureras.  Comercio.  Transportes, correos y almacenamiento.  Información en medios masivos.  Servicios financieros y de seguros.  Servicios inmobiliarios y de alquiler de bienes muebles e intangibles.  Servicios profesionales, científicos y técnicos.  Corporativos.  Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación.  Servicios educativos.  Servicios de salud y de asistencia social.  Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos.  Servicios de alojamiento temporal y de preparación de alimentos y bebidas.  Otros servicios excepto actividades gubernamentales.  Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales.</p>

Tabla 2: Sectores económicos México 2008

Componentes conectadas ( $H_0$ )	
Minería	Industrias Manufactureras
Comercio	Industrias Manufactureras
Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	Servicios científicos, técnicos y profesionales
Minería	Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza
Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	Servicios profesionales, científicos y técnicos
Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	Servicios financieros y de seguros
Minería	Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final
Información en medios masivos	Servicios financieros y de seguros
Información en medios masivos	Industrias Manufactureras
Transportes, correos y almacenamiento	Industrias Manufactureras
Información en medios masivos	Corporativos
Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	Servicios de alojamiento temporal y de preparación de alimentos y bebidas
Otros servicios excepto actividades gubernamentales	Servicios profesionales, científicos y técnicos
Construcción	Minería
Información en medios masivos	Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos
Servicios educativos	Servicios profesionales, científicos y técnicos
Servicios de salud y de asistencia social	Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación
Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales	Comercio
Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza	
Ciclos ( $H_1$ )	
Construcción	Servicios inmobiliarios y de alquiler de bienes muebles e intangibles
Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza	Servicios de salud y de asistencia social

Tabla 3: Componentes conectadas y ciclos con la filtración de Rips

Componetes conectadas ( $H_0$ )	
Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final	Industrias Manufactureras
Transportes, correos y almacenamiento	Industrias Manufactureras
Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final	Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza
Construcción	Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final
Minería	Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza
Corporativos	Servicios profesionales, científicos y técnicos
Industrias Manufactureras	Servicios de alojamiento temporal y de preparación de alimentos y bebidas
Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	Servicios profesionales, científicos y técnicos
Servicios de salud y de asistencia social	Industrias Manufactureras
Información en medios masivos	Servicios financieros y de seguros
Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	Industrias Manufactureras
Otros servicios excepto actividades gubernamentales	Industrias Manufactureras
Comercio	Servicios profesionales, científicos y técnicos
Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales	Industrias Manufactureras
Comercio	Minería
Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos	Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación
Servicios educativos	Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final
Información en medios masivos	Comercio
Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza	
Ciclos ( $H_1$ )	
Industrias Manufactureras Información en medios masivos	Servicios profesionales, científicos y técnicos
Industrias Manufactureras Servicios financieros y de seguros	Servicios profesionales, científicos y técnicos

Tabla 4: Componentes conectadas y ciclos con la filtración de Dowker

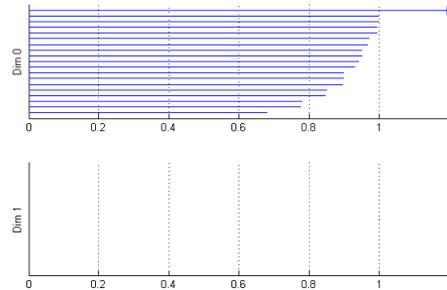


Figura 3: Código de barras de la filtración Rips

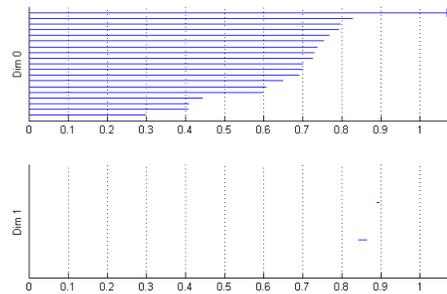


Figura 4: Código de barras de la filtración Dowker

### Agradecimientos

Agradecemos al Dr. Facundo Mémoli y a su alumno Samir Chowdhury por los conceptos y técnicas presentadas en el curso. También a los organizadores por el total apoyo durante el evento.

José de Jesús Aguirre Tepole  
*Departamento de Matemáticas,*  
 Escuela Superior de Física y Matemáticas,  
 Instituto Politécnico Nacional,  
 Av. IPN Edificio 9,  
 U. P. Adolfo López Mateos,  
 Ciudad de México, México,  
 tepole@esfm.ipn.mx

Angel Emilio de León Gutiérrez  
*Oficina de Desarrollo de Sistemas de*  
*Provisión de Liquidez,*  
 D.G. de Sistemas de Pagos y  
 Servicios Corporativos - Banxico,  
 5 de Mayo 6 - Condesa,  
 Ciudad de México, México,  
 angel.deleon@banxico.org.mx

### Referencias

- [CM16] Samir Chowdhury and Facundo Mémoli. Persistent Homology of Asymmetric Networks: An Approach based on Dowker

Filtrations. *ArXiv e-prints*, August 2016.

- [CM17a] Samir Chowdhury and Facundo Mémoli. The metric space of the networks. 2017.
- [CM17b] Samir Chowdhury and Facundo Mémoli. Persnet tutorial, metrics and persistence on networks. 2017.
- [Ine17] Instituto nacional de estadística y geografía, 2017.