

Agrupación de colonias dentro de la Ciudad de México con mayor prioridad para recibir la vacuna contra COVID-19

Martínez Ostoa N.

Profesores: Martínez Sandoval L., Almendra Hernández V.

Matemáticas Discretas y Algoritmos

I.I.M.A.S. - U.N.A.M.

5 de febrero del 2021

Resumen—Este proyecto tiene un fundamento fuerte en la complejidad económica y bajo este esquema de análisis se plantea analizar las alcaldías y colonias prioritarias dentro de la Ciudad de México para maximizar el impacto económico y de salud al vacunar ciertas colonias.

I. PLANTEAMIENTO DEL PROBLEMA

Por un lado, tenemos que al momento la cantidad de vacunas es limitada tanto en unidad y precio. El objetivo es lograr maximizar el impacto de salud y de reactivación económica utilizando los vacunas disponibles hasta el momento. Para un análisis inicial, consideré las tres alcaldías con un mayor índice de complejidad económica. Este índice, de acuerdo con DATAMÉXICO e HIDALGO, es un indicador sobre las capacidades productivas de una localidad a partir de la presencia de actividades económicas en dicha localidad. En la tabla I se muestran las alcaldías con su índice de complejidad económica (ECI).

Al ordenar la tabla I, podemos observar que las alcaldías con un mayor ECI son:

1. Miguel Hidalgo - 4.15
2. Azcapotzalco - 3.87
3. Cuajimalpa de Morelos - 3.85

Continuando con el planteamiento, ya conocemos el ECI para cada alcaldía, ahora, para identificar la diversidad de cada alcaldía tomé a todas los negocios que solicitaron un préstamo económico al gobierno de la Ciudad de México. Con estos negocios, lo que hice fue filtrarlos por alcaldía y agruparlos por categoría de producto/negocio. Después de hacer el filtrado, obtuvimos 13 categorías principales de productos. Estas categorías se muestran en la tabla II

Alcaldía	ECI
Cuauhtemoc	3.47
Venustiano Carranza	2.50
Azcapotzalco	3.87
Coyoacán	2.80
Cuajimalpa de Morelos	3.85
Gustavo A. Madero	2.68
Iztacalco	2.94
Iztapalapa	2.63
La Magdalena Contreras	1.62
Milpa Alta	0.34
Álvaro Obregón	3.82
Tláhuac	1.63
Tlalpan	2.43
Xochimilco	1.21
Benito Juárez	3.79
Miguel Hidalgo	4.15

Cuadro I
ALCALDÍAS DE LA CIUDAD DE MÉXICO Y SU RESPECTIVO ÍNDICE DE COMPLEJIDAD ECONÓMICA (ECI)

Categoría
Arreglos en general
Industria
Restaurantes
Fotografía
Venta Productos
Comida
Ropa
Papelería
Belleza
Electrónicos
Música
Salud
Mascotas

Cuadro II
CATEGORÍAS/PRODUCTOS ENCONTRADOS EN LOS NEGOCIOS DE LAS COLONIAS DE LAS 3 ALCALDÍAS CON MAYOR ECI

Dentro de los datos de préstamos económicos, únicamente consideré aquellos que ya hayan sido cobrados y después de aplicar el filtro por alcaldía, obtuve que hay 13 colonias dentro de las tres alcaldías principales. Dicha lista de colonias se muestra en la tabla III

Colonias
Jaime Torres Bodet
El Rosario
Tezozomoc
Doctores
Olivar del conde 1a sección
Campamento 2 de octubre
Popotla
Tlaxpana
Escandon
San Martin Xochinahuac
San Juan Tlihuaca
Xaltocán
Santa Maria Nativitas

Cuadro III

COLONIAS DENTRO DE LAS TRES ALCALDÍAS DE LA CIUDAD DE MÉXICO CON MEJOR ECI

Concretamente, el planteamiento del problema lo pensé de la siguiente forma: crear una gráfica bipartita B que represente la relación entre una colonia c y un producto/categoría p . Una vez que tengo a B , la idea es explorar cada una de los nodos de colonias para determinar los siguientes factores:

- Diversificación de la colonia c : de los 13 productos, cuántos produce c
- Complejidad de la colonia c : valor numérico que representa, de los productos que produce, cuántos otras colonias también lo producen
- ECI: índice de complejidad de la alcaldía a a la que pertenece la colonia c
- Porcentaje de infectados: porcentaje de personas infectadas por COVID-19 en c

Una vez que tengamos los puntos anteriores, la idea fue crear una gráfica completa G ($k = 13$) con pesos en donde el peso entre las colonias c_i y c_j , representa el impacto de salud que tiene vacunar a los negocios en c_i y c_j . Finalmente, para obtener un orden óptimo de vacunación, lo que hice fue encontrar el árbol de peso mínimo en G y eso me asegura que el orden de vacunación dentro de las 13 colonias es óptimo para maximizar el beneficio económico y de salud. Con base en esto, la función

que asigna pesos asignará un menor peso a dos colonias cuyo beneficio de vacunar directamente es mayor a otras dos colonias cuyo beneficio no sea óptimo.

II. JUSTIFICACIÓN DE ELECCIÓN DE LOS DATOS

Para construir el dataset final, cuyas columnas están conformadas por:

- Alcaldía
- ECI por alcaldía
- Código Postal
- Colonia
- Negocio
- Categoría del negocio
- Porcentaje de infectados en la colonia
- Número de empleados

se obtuvieron datos de dos fuentes principales:

1. Datos del gobierno de la Ciudad de México ¹,
2. DataMéxico ², 3. Datos construidos de manera independiente. Concretamente los siguientes datasets:

Datos del gobierno de la Ciudad de México:

- Colonias de la Ciudad de México con número de infectados
- Empresas que solicitaron un préstamo económico a la Ciudad de México

Datos de DataMéxico;

- Índice de complejidad por alcaldía

Datos independientes:

- Lista de alcaldías de la Ciudad de México

Después de emplear todos estos datos, junto con la etapa de filtrado (top 3 colonias de la CDMX con el ECI más grande), obtuvimos un dataset de 353 puntos de información. En la siguiente figura se ve un ejemplo de los primeros 5 elementos dentro del dataset final:

	alcaldia	eci	c_p	colonia	negocio	categoria	porcentaje_infectados	num_empleados
0	azcapotzalco	3.865509	13530	jaime torres bodet	zapatosbelleza	arreglosgeneral	0.1978	2
1	azcapotzalco	3.865509	2100	el rosario	manufacturados	industria	0.1573	0
2	azcapotzalco	3.865509	2459	tezozomoc	antojerias	restaurantes	0.1303	3
3	azcapotzalco	3.865509	6720	doctores	fotografia	fotografia	0.1169	3
4	azcapotzalco	3.865509	2459	tezozomoc	catalogo	ventaproducos	0.1303	2

Figura 1. Ejemplo del dataset final

¹<https://datos.cdmx.gob.mx/>

²<https://datamexico.org/es>

III. MODELACIÓN MATEMÁTICA

Una vez que tenemos las colonias y los productos/categorías, el modelo matemático se reduce a los siguientes puntos:

- Crear una matriz P que represente el espacio de productos dentro de cada uno de los negocios de las respectivas colonias.
 - Las columnas de P representan las 13 categorías
 - Los renglones de P representan las 13 colonias
 - $P_{ij} = 1$ si la colonia i produce al menos un negocio dentro de la categoría j
 - $P_{ij} = 0$ si la colonia i no produce ningún negocio dentro de la categoría j
- Crear una gráfica bipartita B la cual represente esta matriz de transición para visualizar las relaciones directas entre una colonia i y un producto/categoría j
- Generar para cada colonia un vector de agrupación:

$$\langle M_{c0}, M_{c1}, ECI_{ac}, S_c \rangle$$

donde:

- M_{c0} ³: diversificación de productos para la colonia c
- M_{c1} : complejidad de producción para la colonia c
- ECI_{ca} : índice de complejidad (ECI) para la alcaldía a de la colonia c
- S_c : porcentaje de infectados de COVID-19 para la colonia c
- Construir la gráfica G con base en los vectores de agrupación de cada colonia maximizando los pesos entre dos colonias. Para el cálculo de pesos diseñe la siguiente ecuación:

$$W = \frac{1}{\Gamma \cdot ECI_c \cdot (\alpha + \beta)} \quad (1)$$

donde:

- $\Gamma = (S_i + S_j)^2$: donde S_i es el porcentaje de infectados para la colonia i
- $ECI_c = \max\{ECI_{ai}, ECI_{aj}\}$: representa el máximo de los índices de complejidad para las colonias i y j

³Tanto M_{c0} como M_{c1} tienen fundamento en los papers de HIDALGO

- $\alpha = M_{i0} + M_{j0}$
- $\beta = \max\{M_{i1}, M_{j1}\}$

- Encontrar el árbol de peso mínimo dentro de G

IV. PROBLEMA ALGORÍTMICO A RESOLVER

El problema algorítmico se resume a los siguientes subproblemas:

- Hacer un filtrado en tiempo lineal de los conjuntos de datos para obtener el dataset final con las columnas: *alcaldía, ECI, CP, colonia, negocio, categoría, porcentaje infectados, número de empleados*
- Construir la matriz P en tiempo $O(n^2)$ y espacio $O(n^2)$. Dicha matriz se muestra en [X-C](#) junto con su representación gráfica en [X-D](#)
- Construir la matriz bipartita B (ver Anexo, sección [X-E](#))
- Construir la matriz completa $k - 13$ G (ver Anexo, sección [X-F](#))
- Correr el algoritmo de Prim sobre G

V. SOLUCIÓN CON ALGORITMOS COMBINATORIOS

Como se mencionó en las secciones [III](#) y [IV](#), la solución al problema de optimización se encontró mediante un árbol de peso mínimo con el algoritmo de Prim. Concretamente, este problema no se planteó como un problema combinatorio, mas que en la generación del dataset final.

VI. ANÁLISIS DE CORRECTITUD, TIEMPO Y ESPACIO

Análisis de tiempo:

1. Generación del dataset final - $O(n^3)$
2. Generación de P - $O(n^2)$
3. Generación de B - $O(n)$
4. Generación de G - $O(n^2)$
5. Algoritmo de Prim sobre G - $O(n^2)$

Análisis de espacio:

1. Generación del dataset final - $O(n)$:
Concretamente, el dataset final tiene 353 puntos de información.
2. Generación de P - $O(n^2)$
3. Generación de B - $O(n^2)$
4. Generación de G - $O(n^2)$
5. Algoritmo de Prim sobre G - $O(n)$:

La complejidad en espacio de Prim es lineal porque se crean listas auxiliares para ir guardando los estados visitados e ir construyendo el árbol de peso mínimo

VII. ESTRATEGIA DE DISEÑO Y ESTRUCTURAS DE DATOS UTILIZADAS

- Listas: esta estructura de datos se ocupa frecuentemente durante todo el desarrollo del proyecto para diversas funciones
- Conjuntos (sets): esta estructura de datos se emplea
- Diccionarios: se emplean dos principales- uno para almacenar los productos y colonias junto con sus respectivos índices.
- Gráficas (*networkx*): se emplean las clases de gráficas de la librería *networkx*, que en memoria están implementadas como listas de adyacencia

VIII. APLICACIÓN A LOS DATOS

Tomando como referencia el dataset final, en esta sección se presenta una exploración general del dataset que llevaron a construir las gráficas *B* y *G* del anexo (ver secciones *X-E*, *X-C*).

VIII-A. Número de productos producidos en cada colonia

Lo que podemos observar en el histograma 2 es que hay 6 colonias en las cuales se concentra más del 80 % de todo lo producido en las alcaldías.

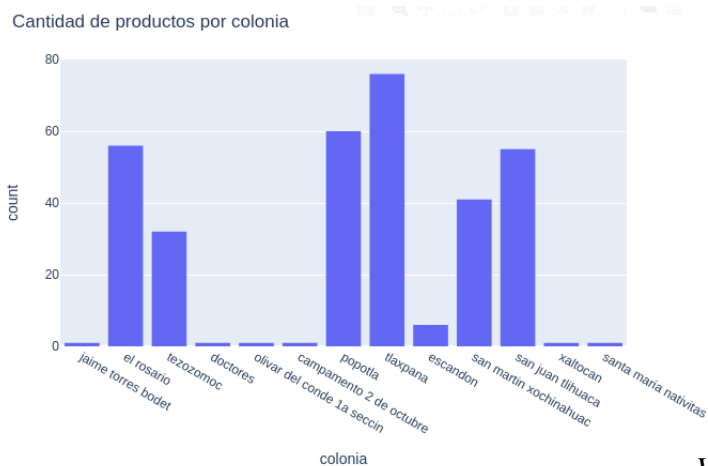


Figura 2. Número de productos producidos en cada una de las 13 colonias

VIII-B. Cantidad de productos por alcaldía

De nuevo, en el histograma 3 podemos observar que en las colonias de Azcapotzalco y Miguel Hidalgo es donde se concentran más del 80 % de todo lo producido en las 3 alcaldías más complejas.

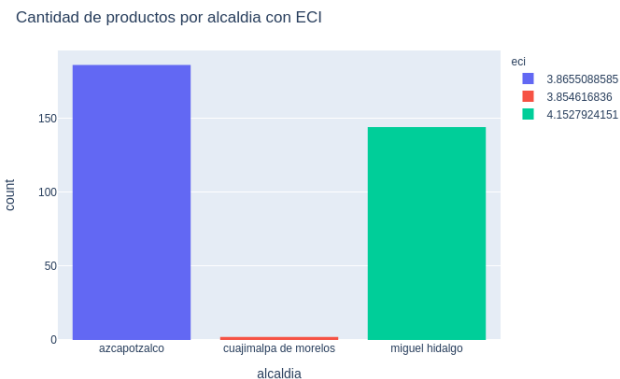


Figura 3. Cantidad de productos producidos en las tres alcaldías más complejas de la CDMX

VIII-C. Distribución de los productos en las colonias

En el histograma 4 podemos observar que las categorías más comunes son los restaurantes y la industria. En la industria, se encuentran conglomerados de negocios que se encargan de productos mecánicos, vulcanizadoras hasta empresas de muebles.

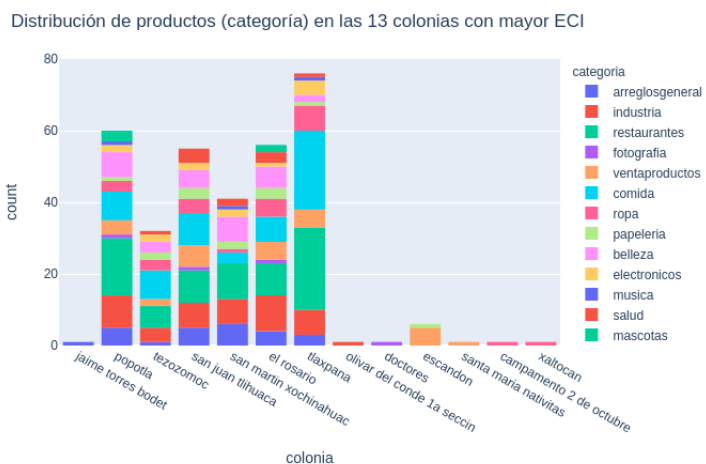


Figura 4. Distribución del tipo de productos en las alcaldías

VIII-D. Distribución de productos para las 3 colonias con mayor cantidad de productos

Finalmente, en los histogramas de esta sección vemos la distribución de los productos más comunes

en las 3 colonias más complejas.

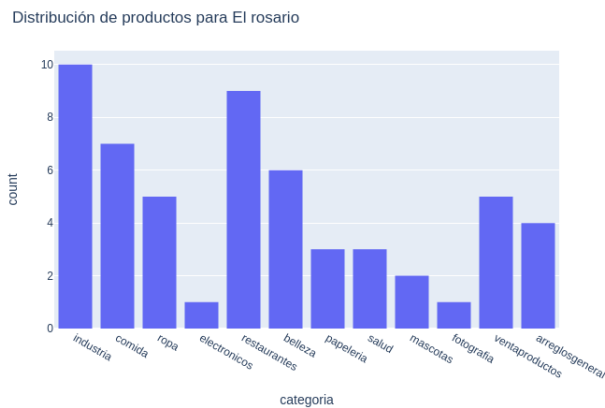


Figura 5. Colonia El Rosario

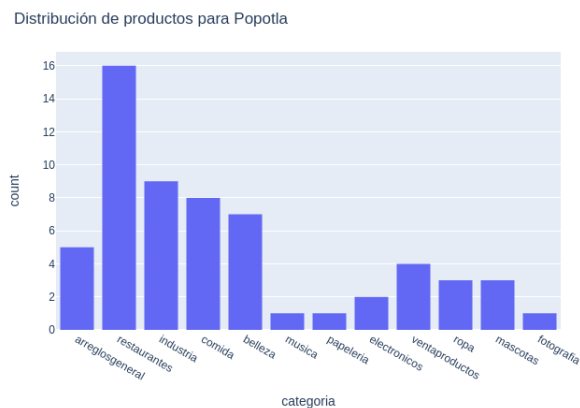


Figura 6. Colonia Popotla

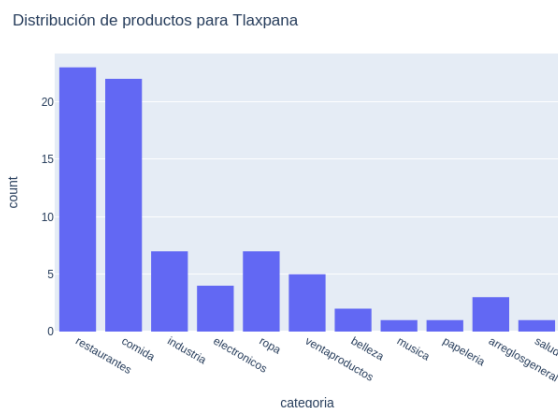


Figura 7. Colonia Tlaxpana

IX. CONCLUSIONES Y TRABAJO A FUTURO

Después de hacer el análisis final y correr los algoritmos sobre las gráficas, obtuve que la prioridad

para vacunar a las colonias es la siguiente:

1. El Rosario
2. Xaltocan
3. San Juan Tlihuaca
4. Tlaxpana
5. San Martín Xochinauhac
6. Popotla
7. Santa María Nativitas
8. Tezozomoc
9. Jaime Torres Bodet
10. Olivar del Conde 1era Sección
11. Escandón
12. Campamento 2 de octubre
13. Doctores

Como trabajo a futuro queda expandir este mismo análisis para todas las colonias con préstamo cobrado y aprobado dentro de las 16 alcaldías de la Ciudad de México. Una de las mayores limitantes de este proyecto fue la limpieza de datos. Gran parte del análisis se restringió a las top 3 alcaldías porque eso arrojaba una clasificación manual de cerca de 400 negocios. En total, de las 16 alcaldías, los negocios obtenidos eran cerca de 11,000.

En la figura 11⁴ se muestra la gráfica completa sobre la cual corrí el algoritmo de Prim:

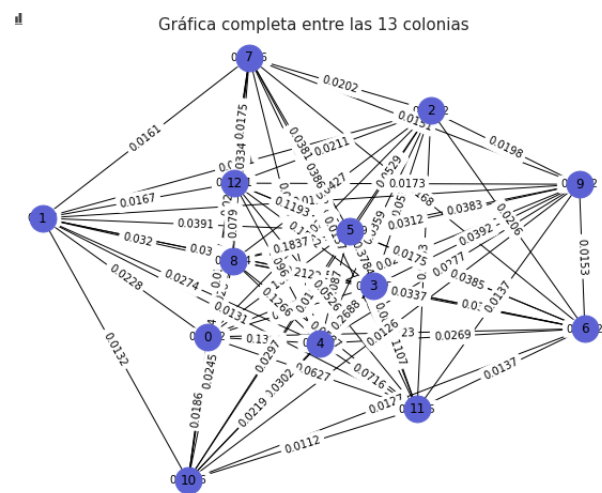


Figura 8. Gráfica G con las colonias y los pesos entre cada una de ellas

REFERENCIAS

VU, TRUNG V. *Economic Complexity and health outcomes: A global perspective* Munich Personal RePEc Archive. 2020

⁴Revisar Anexo X-A para ver la numeración de cada colonia

HIDALGO C *Economic complexity theory and applications* Nature Reviews Physics. 2021

SKIENA S *The Algorithm Design Manual* Springer. 2020

HIDALGO C., HAUSMANN R. *The network structure of economic output* Springer Science + Business Media. 2011

HIDALGO C., KLINGER B., BARABÁSI A., Hausmann R. *The Product Space Conditions the Development of Nations* Science. 2007.

HIDALGO C., HAUSMANN R. *The Building Blocks of Economic Complexity* Center for International Development at Harvard University. 2009

PRITCHETT L., SUMMERS L. *Wealthier is Healthier* World Development Report. 1993

X. ANEXO

X-A. Diccionario de colonias

Índice	Colonia
1	jaime torres bodet
2	el rosario
3	tezozomoc
4	doctores
5	olivar del conde 1a sección
6	campamento 2 de octubre
7	popotla
8	tlaxpana
9	escandon
10	san martin xochinahuac
11	san juan tlihuaca
12	xaltocan
13	santa maria nativitas

Cuadro IV
DICCIONARIO DE COLONIAS

X-B. Diccionario de productos/categorías

Índice	Producto/Categoría
51	arreglosgeneral
52	industria
53	restaurantes
54	fotografia
55	ventaproducos
56	comida
57	ropa
58	papelaria
59	belleza
60	electronicos
61	musica
62	salud
63	mascotas

Cuadro V
DICCIONARIO DE PRODUCTOS/CATEGORÍAS

X-C. Matriz P

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & & & & & & & \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & & & & & & & \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & & & & & & & \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & & & & & & & \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & & & & & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & & & & & & & \end{bmatrix} \tag{2}$$

X-D. Representación gráfica de P

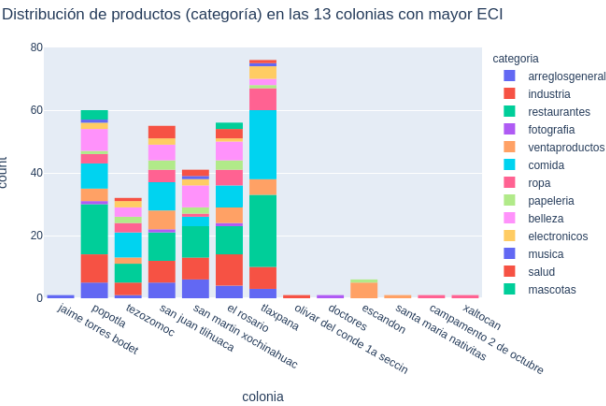


Figura 9. Diversificación de cada colonia en categorías

X-E. Gráfica bipartita B

X-F. Gráfica completa G ($k = 13$)



Figura 10. Gráfica bipartita B entre colonia y producto

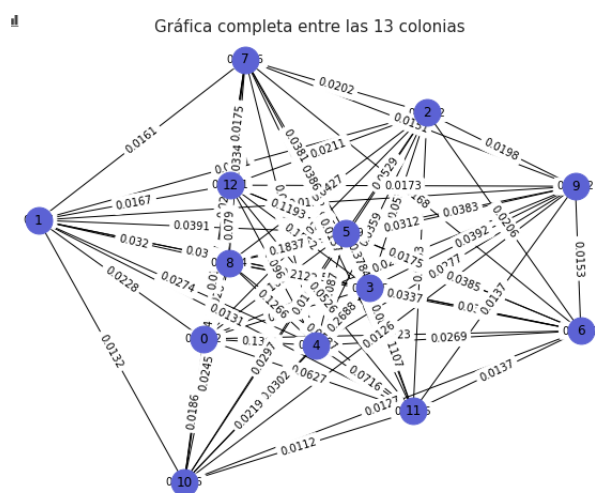


Figura 11. Gráfica completa G