# Modelado Molecular Para la Optimización de Compuestos y Caracterización de Propiedades Mediante Grafos

Robert Daniel Fonseca, Ana Sofia Avila, Jose Nicolas Rios Teoría de grafos Matemáticas Aplicadas y Ciencias de la Computación Escuela de Ingeniería, Ciencia y Tecnología Universidad del Rosario

#### I. Introducción

A química orgánica es una rama de la química que se centra en el estudio de los compuestos basados en carbono, los cuales son fundamentales para la vida en el planeta Tierra. En este contexto, los grafos moleculares han surgido como una herramienta poderosa para comprender y representar la estructura y propiedades de las moléculas. Los grafos moleculares son representaciones gráficas de las moléculas, donde los átomos se representan como vértices y las conexiones entre ellos como aristas [1]. Estos grafos capturan la información topológica y estructural de las moléculas, lo que permite analizar y predecir propiedades químicas, reactividad y comportamiento de las sustancias.

#### II. DESCRIPCION DEL PROBLEMA

Para identificar las propiedades químicas de un compuesto orgánico, es necesario realizar una transformación de su fórmula molecular a un diagrama de estructura de Lewis, gráfico donde se ve la distribución espacial de los átomos de la molécula junto con el número de enlaces que hay entre ellos. En este diagrama es posible llevar a cabo la caracterización y clasificación del tipo de compuesto a través del análisis de la adyacencia de sus átomos, atributos específicos y tipos de enlace. Tanto la creación como el análisis de este diagrama constituyen tareas que requieren una inversión significativa de tiempo por parte del profesional, y estas labores pueden ser agilizadas mediante la utilización de software basado en la teoría de grafos. Esta automatización no solo simplifica la manipulación de estructuras complejas, sino que también permite gestionar eficientemente estructuras de gran tamaño.

#### III. OBJETIVOS

#### A. General

Agilizar el trabajo de investigación e ingeniería de los profesionales que necesitan caracterizar y comparar moléculas orgánicas de forma eficiente, mediante el diseño de un software de modelado que permita visualizar y analizar la estructura dada por medio del uso de grafos.

# B. Especificos

 Graficar y predecir la estructura de una molécula a partir de una fórmula química condensada proporcionada por el usuario.

- Identificar grupos funcionales mediante el análisis de adyacencias e incidencias.
- Identificar las conexiones específicas de un átomo determinado.
- Identificar la masa total de la molécula y la región de masa máxima.
- Hallar los índices topológicos de la molécula dada.
- Analizar los puntos de ebullición aproximados de algunos alcanos por medio de los índices topológicos.

# IV. MARCO TEÓRICO

- A. Conceptos fundamentales de química orgánica
- Formula química condensada: La fórmula química condensada o molecular indica la proporción de átomos al interior de un compuesto.[2]
- Estructura de Lewis: Es una forma de mostrar los electrones de la capa exterior de un átomo. En la estructura de Lewis, se colocan los átomos con sus letras específicas.[3]
- Grupos funcionales:
  - Hidrocarburos: Compuestos orgánicos formados únicamente por átomos de carbono e hidrógeno, enlazados entre sí por uniones covalentes. [4]
    - \* Alcanos: Hidrocarburos que contienen únicamente enlaces simples entre sus átomos.[4]
    - \* **Alquenos:** Hidrocarburos que contienen al menos un enlace doble entre sus átomos, siendo este el enlace de mayor grado.[4]
    - \* **Alquinos:** Hidrocarburos que contienen al menos un enlace triple entre sus átomos.[4]
  - **Hidroxilos:** Grupos funcionales que se forman en la union de oxígeno, hidrógeno y otros átomos.
    - \* Carbonilos: Radical formado por un átomo de carbono con doble enlace a un oxígeno.
    - \* Ácido Carboxilicos: Radical formado por un átomo de carbono con doble enlace a un oxígeno y enlace sencillo a un alcohol (OH).
    - \* Esteres: Radical formado por un átomo de carbono con doble enlace a un oxígeno y enlace sencillo a un oxígeno, conectado con otro átomo diferente, generalmente carbono.
  - Aminas: Moléculas que contienen enlaces carbononitrógeno. El átomo de nitrógeno de una amina tiene

1

un par solitario de electrones y tres enlaces con otros átomos, ya sean de carbono o de hidrógeno.

- \* Aminas primarias: Aquellas cuya conexión carbono nitrógeno tiene 1 solo enlace. El nitrógeno tiene enlaces a 2 hidrógenos.
- \* Aminas secundarias: Aquellas cuya conexión carbono nitrógeno tiene 2 enlaces. El nitrógeno tiene enlace a 1 hidrógeno.
- Amidas: Unión de un grupo carbonilo y una amina.
   Es decir, un carbono con doble enlace a oxígeno y un enlace sencillo a un nitrógeno.
- Regla de Octeto: La regla del octeto establece que un átomo es considerado estable cuando completa su orbital de valencia con 8 electrones, ya sea a través de electrones de valencia propios o mediante enlaces químicos. Esta regla fomenta la tendencia de los átomos a unirse en moléculas, las cuales son estables cuando todos sus átomos componentes cumplen con esta regla.[7]
- Masa atómica: La masa atómica, también conocida como masa molar, es una propiedad física que se refiere a la masa promedio de un átomo de un elemento químico en una unidad de masa atómica (u). La cual está estandarizada para los elementos en la tabla periódica sin variación en la cantidad de neutrones.[7]
- Masa molecular: La masa molecular, también conocida como masa total de una molécula, es la suma de las masas atómicas de todos los átomos que la conforman.[7]
- Grafo molecular: Grafo en donde los vértices equivalen a los átomos y las aristas a los enlaces de covalencia.
- **Indice topológico:** Es un invariante del grafo molecular. Relaciona propiedades estructurales de una molécula con propiedades físicas de la misma.

#### B. Conceptos de teoría de grafos

- **Distancia:** Si el grafo G tiene un u, v-camino, la distancia de u a v, notada d(u, v), es la longitud mínima de un u, v-camino.
- **Vecindad:** Es el conjunto de vértices adyacentes a un vértice. Se nota como N(v).
- Índice de Randić: El índice de Randić es una medida utilizada en química computacional para evaluar la complejidad de un grafo molecular no cíclico al comparar los grados de los vértices. Este índice relativo se emplea para comparar la ramificación de dos moléculas, donde un valor más alto indica una mayor ramificación [6]. Se define como:

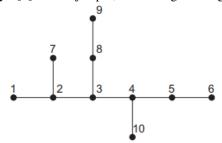
$$R_i = \sum_{j=1}^n \frac{1}{\sqrt{d_i d_j}}$$

• Índice de Weiner: Es la suma de las distancias entre todos los pares de vértices en un grafo. Se define como:

$$\sum_{u,v \in V(G)} d(u,v)$$

• Número de caminos de Wiener: Son los números  $^iP$  donde  $^iP$  es el número de pares de vértices en el grafo separados por i aristas. Por ejemplo, en un  $C_3$  tendriamos

- que  ${}^{1}P$  es 3, y el resto de ellos serían 0. Es un índice topológico del grafo molecular.[8]
- Metil: Se define como el número de vértices de grado uno que son adyacentes a un vértice de grado tres o mayor[8]. Por ejemplo, en el siguiente grafo:



Tenemos que mth = 3 (observe los vértices 1, 7, 10). Es un índice topológico del grafo molecular.

#### V. MODELAMIENTO DEL PROBLEMA

Con el propósito de visualizar y analizar las moléculas objetivo, estas se representan mediante grafos. En este esquema, cada vértice representa un átomo de la estructura, mientras que las aristas representan la existencia de enlaces entre átomos adyacentes. La cantidad de arista indica la cantidad de electrones que participan en el enlace tomando solo un átomo, lo que, a su vez, implica el tipo de enlace presente en la molécula, es decir, 2 átomos con 2 aristas paralelos entre ellos indica que cada átomo está aportando 2 electrones a ese enlace y tomando otros 2. Se presenta de forma análoga para los enlaces triples y 3 arístas paralelas.

Los átomos que seran tomados en cuenta para la implementación son aquellos que componen los compuestos orgánicos más comunes, es decir, carbono, hidrógeno, oxígeno y nitrógeno. Para cada uno de ellos se implementara en la creación del grafo atributos que almacenen el tipo al que pertenece cada vértice (atributo "name") junto con sus características de electronegatividad (atributo "electronegatividad") y masa atómica (atributo "masa").

En la figura 1 se muestra el modelamiento de la molécula CO2 mediante el grafo con las características establecidas.

# **HCONH2**

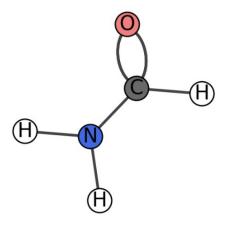


Figura 1. Modelo preliminar

#### VI. SOLUCIÓN PROPUESTA

Para la realización de cada objetivo específico, se plantean diferentes soluciones. Se explicarán en detalle a continuación.

- 1) Para graficar la molécula, se hace uso de la librería igraph que permite el modelamiento de grafos. Así como la librería matplotlib para la parte gráfica de los grafos. Por otro lado, se hace uso de un algoritmo diseñado por los propios estudiantes que desarrolla la configuración de la molécula a partir de la fórmula molecular (cadena de texto), esto mediante la revisión de la regla del octeto en los átomos de Carbono, Hidrógeno, Nitrógeno y Oxígeno.
- 2) El proceso de identificación del grupo funcional se realiza mediante la matriz de adyacencia y una lista de etiquetas que contiene a que tipo de átomo corresponde cada fila y columna. En dicha matriz según el numero de la entrada y sus étiquetas correspondientes se identifica que tipo de enlace hay entre átomos específicos, de forma que si se encuentra una adyacencia correspondiente a un grupo funcional este lo idéntifica. Por ejemplo en el átomo de CO2 de la figura 1 se identifica al menos un átomo de oxígeno con enlacce doble a un carbono y se descartan otras adyacencias correspondientes a otros grupos Hidroxilo.
- 3) La masa total se calcula recorriendo el grafo con busqueda a lo ancho, ya que las moléculas no suelen tener una profundidad muy alta, y sumando la masa de cada átomo que compone el grafo desde el atributo de cada vértice. Con el mismo tipo de recorrido la región de masa máxima será la vecindad que posea la mayor suma de masas.
- 4) La electronegatividad total se calcula recorriendo el grafo con busqueda a lo ancho y sumando la electronegatividad de cada átomo que compone el grafo desde el atributo de cada vértice. Con el mismo tipo de recorrido la región de electronegatividad máxima será la vecindad que posea la mayor suma de electronegatividades.
- 5) En cada molécula se calculan 4 índices topológicos, el índice de Randić, indice de Wiener y metil que se usa para comparar las ramificaciones, y el índice de Wiener, para evaluar tanto las ramificaciones como el tamaño de la molécula, así como los caminos de Wiener.
- 6) Parar calcular el punto de ebullición aproximado de un alcano se usa el modelo planteado en el texto [8], el cuál es:

$$\begin{array}{llll} f(^1P,\dots,^8P,\mathrm{Mth}) & = & -167.49997 & + \\ 84.28344(^1P)^{0.46072} & + & 15.94534(^2P)^{0.00348} & + \\ 17.42198(^3P)^{0.53517} & + & 11.16515(^4P)^{0.00164} & + \\ 4.74582(^5P)^{0.00089} & + & 5.23270(^6P)^{0.00143} & + \\ 6.67018(^7P)^{0.14687} & + & 5.27829(^8P)^{0.96677} & - \\ 5.53723\mathrm{Mth}^{0.00072} & & & \\ \end{array}$$

#### VII. IMPLEMENTACIÓN DE LA SOLUCIÓN

Para la implementación de la solución, se desarrolló un algoritmo de construcción de matrices de adyacencia, el cual se inicializa con una matriz de ceros, y una lista de las moleculas ordenadas por columna en la matriz. El algoritmo

toma en cuenta los valores dados por la regla del octeto para cada átomo, y se encarga de probar todas las posibles construcciones de una molécula hasta hallar la que es estable. Una molecula estable dentro del algoritmo implica que todos sus átomos están completos y que su matriz de adyacencia corresponde a un grafo conexo.

Como parte de la construcción del algoritmo, se implementaron 7 funciones, una que inicializa la matriz en ceros, 4 de las cuales verifican en qué estado se encuentra un átomo (incompleto, completo, con sobrecarga de enlaces) y 2 que verifican que la matriz esté completa y el grafo resultante sea conexo.

Además del algoritmo y las funciones que le asisten, se generó la visualización de la molécula a partir del grafo, dandole el color de los átomos que la componen, así como mostrando la cantidad de enlaces entre cada una y el nombre en la parte superior. Esto se desarrolla con la ayuda de las librerias igraph y cairocffi, utilizando un grafo generado por la librería igraph a partir de la matriz y y la función "plot()".

Por último, se generaron 16 funciones que trabajan sobre la matriz de adyacencia de la molécula para caracterizarla, esto incluye colocar su masa, electronegatividad, masa máxima, electronegatividad máxima y en que átomo se concentran, así como la identificación de los grupos funcionales detallados parrafos atrás. Además de esto, se tienen cuatro funciones, aquella que cuenta el número de caminos de Wiener que posee la molecula, una que encuentra su punto de ebullición calculado a partir de los metiles que posee, otra que calcula cuántos metiles posee, y la última que calcula el índice de Wiener de la molécula.

#### VIII. RESULTADOS

Para evidenciar el funcionamiento al final del programa se realiza un bloque de código que condensa la formula molecular ingresada, el grafo molecular creado algorítmicamente junto con la masa total, zona de masa máxima, electronegatividad total, zona de electronegatividad máxima, grupo funcional identificado o ausencia de él y punto de ebullición si se detecta que la molecula es un alcano. En el anexo se presenta el *output* de una molécula para cada grupo funcional trabajado.

#### IX. CONCLUSIONES

A partir del trabajo desarrollado se derivan las siguientes conclusiones:

- Las moléculas orgánicas, si bien son capaces de escribirse de forma compacta con una fórmula molecular, tienen varios isómeros que deberían ser tomados en cuenta si se desarrolla la idea más a fondo.
- 2) Se desarrolló un algoritmo capaz de encontrar una versión estable de cualquier molécula orgánica que se le ingrese, si bien toma bastante tiempo con moléculas grandes debido a la complejidad computacional que esto conlleva. Sin embargo, el usuario es capaz de mejorar el rendimiento proporcionando la fórmula semi desarrollada de la molécula
- Los conceptos de teoría de grafos, en especial los índices de Randic y Wiener, tienen aplicaciones importantes en la química para el análisis de moléculas

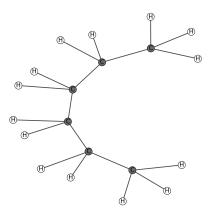
# BIBLIOGRAFÍA

- [1] M. Aguirre Elorza, CLASIFICACIÓN DE GRAFOS MOLECULARES CON APRENDIZAJE AUTOMÁTICO, Trabajo Fin de Grado, Universidad Rey Juan Carlos, Madrid, 2022. [En línea]. Disponible: https://burjcdigital.urjc.es/handle/10115/22592
- [2] Formulas estructurales condensadas ALCANOS. Pagina Principal, Materiales de Aprendizaje.org. Accedido el 23 de octubre de 2023. [En línea]. Disponible: https://shorturl.at/avPSW
- [3] A. Z. Fernandes. Qué es la estructura de Lewis y cómo se representa (con ejemplos). Toda Materia. Accedido el 23 de octubre de 2023. [En línea]. Disponible: https://www.todamateria.com/estructura-de-lewis/
- [4] HIDROCARBUROS ALIFÁTICOS. Didáctica Multimedia. Accedido el 23 de octubre de 2023. [En línea]. Disponible: https://shorturl.at/pqCGU
- [5] Cso.go.cr. [En línea]. Disponible en: https://shorturl.at/tDNR5
- [6] I. Gutman, B. Furtula y V. Katanić, Randić index and information, AKCE Int. J. Graphs Combinatorics, vol. 15, n.º 3, pp. 307–312, diciembre de 2018. Accedido el 23 de octubre de 2023. [En línea]. Disponible: https://doi.org/10.1016/j.akcej.2017.09.006
- [7] J. Clayden, N. Greeves, S. Warren, P. Wothers, Organic Chemistry", 1st ed. Oxford University Press. 2001.
- [8] K. J. Burch, Chemical applications of graph theory, en Mathematical Physics in Theoretical Chemistry, S. M. Blinder y J. E. House, Eds. Elsevier, 2019, pp. 261-294.
- [9] S. Medina Carrasco, Influencia de los grupos hidroxilos como grupo funcional primario o secundario en la adsorción de moléculas orgánicas sobre grafito, 2012.
- [10] P. Flowers, K. Theopold, R. Langley, y W. R. Robinson, 20.4 Aminas y amidas, Química 2ed, 02-jun-2022. [En línea]. Disponible en: https://shorturl.at/gGUVW

#### X. ANEXOS

#### A. Hidrocarburos

• Alcanos:

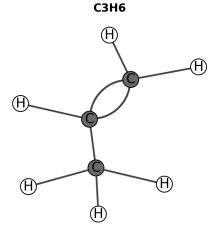


```
Masa total: 86.17576 U
La masa máxima se concentra en: C
Electronegatividad total: 46.100000000000001 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randic es: 9.656854249492381
Su índice de wiener es: 35
Sus caminos de wiener son: p1: 5, p2: 4, p3: 3, p4: 2, p5: 1, p6: 0, p7: 0, p8: 0
Su número de metiles es: 0
Su grupo funcional principal es: alcano
Su punto de ebullición aproximado es: 72.72853389624701
```

Figura 1. Molécula Alcano

#### • Alquenos

☐ Ingrese la fórmula molecular: C3H6 Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3



```
Masa total: 42.080040000000004 U
La masa máxima se concentra en: C
Electronegatividad total: 20.8499999999998 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randic es: 1.6927053408400363
Su índice de wiener es: 4
Sus caminos de wiener son: p1: 2, p2: 1, p3: 0, p4: 0, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 1
Su grupo funcional principal es: alqueno
```

Figura 2: Molécula Alqueno

# • Alquinos

☐ Ingrese la fórmula molecular: C4H6
Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3

# 

**C4H6** 

Masa total: 54.09103999999999 U
La masa máxima se concentra en: C
Electronegatividad total: 23.4 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randic es: 3.25
Su índice de wiener es: 10
Sus caminos de wiener son: p1: 3, p2: 2, p3: 1, p4: 0, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 2
Su grupo funcional principal es: alguino

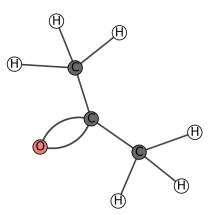
Figura 3. Molécula Alquino

# B. Hidroxilos

• Carbonilo

☐ Ingrese la fórmula molecular: CH3COCH3
Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3

#### СНЗСОСНЗ



Masa total: 58.079040000000000 U
La masa máxima se concentra en: C
Electronegatividad total: 24.2899999999999 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randic es: 3.7677669529663684
Su índice de wiener es: 9
Sus caminos de wiener son: p1: 3, p2: 3, p3: 0, p4: 0, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 2
Su grupo funcional principal es: carbonilo

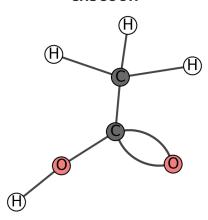
Figura 4. Molécula carbonilo

# · Ácidos carboxílico

☐ Ingrese la fórmula molecular: CH3COOH

Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3

# СНЗСООН



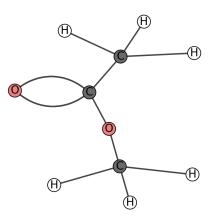
Masa total: 60.05136 U
La masa máxima se concentra en: C
Electronegatividad total: 20.7799999999999 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randíc es: 3.7677669529663684
Su índice de wiener es: 9
Sus caminos de wiener son: p1: 3, p2: 3, p3: 0, p4: 0, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 2
Su grupo funcional principal es:

ácido carboxilico

# Figura 5. Molécula ácido carboxílico

#### • Ester

Ingrese la fórmula molecular: CH3COOCH3 Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3 CH3COOCH3



Masa total: 74.07804 U
La masa máxima se concentra en: C
Electronegatividad total: 27.72999999999997 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randic es: 6.035533905932738
Su índice de wiener es: 18
Sus caminos de wiener son: p1: 4, p2: 4, p3: 2, p4: 0, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 1
Su grupo funcional principal es:

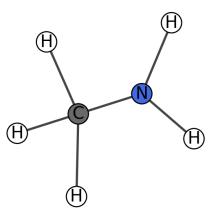
Figura 6. Molécula ester

#### C. Aminas

# • Primaria

Ingrese la fórmula molecular: CH3NH2
Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3

# CH3NH2



Masa total: 31.0569000000000002 U
La masa máxima se concentra en: C
Electronegatividad total: 16.59 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randic es: 1.0
Su índice de wiener es: 1
Sus caminos de wiener son: p1: 1, p2: 0, p3: 0, p4: 0, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 0

Su grupo funcional principal es: amina primaria

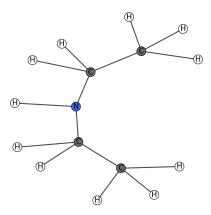
Amina secundaria

Figura 7. Molécula amina primaria

# • Secundaria

Ingrese la fórmula molecular: CH3CH2NHCH2CH3
Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3

CH3CH2NHCH2CH3



Masa total: 73.13694000000000 U
La masa máxima se concentra en: C
Electronegatividad total: 37.4400000000000 kJ/mol
La electronegatividad máxima se concentra en: C
Su índice de randíc es: 6.742640687119286
Su índice de wiener es: 20
Sus caminos de wiener son: p1: 4, p2: 3, p3: 2, p4: 1, p5: 0, p6: 0, p7: 0, p8: 0
Su número de metiles es: 0
Su grupo funcional principal es:

Figura 8. Molécula amina secundaria

# D. Amida

☐ Ingrese la fórmula molecular: CH3CONH2

Ingrese el número de enlaces máximo deseado (simple, doble, triple) (Si no tiene esta información, ingrese 3): 3

# CH3CONH2

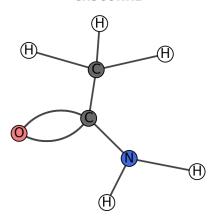


Figura 9. Molécula amida