



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

DQ.0120.2024

Abril 13, 2024

Dr. Román Linares Romero
Presidente del Consejo Divisional
División Ciencias Básicas e Ingeniería
PRESENTE

Le solicito que ponga a consideración del Consejo Divisional de la División de Ciencias Básicas e Ingeniería (DCBI) dos proyectos de investigación de nueva creación:

Proyecto	Responsable	Participantes
Área Académica de Química Cuántica		
Correlaciones de alto orden en el espacio químico representado en un hipergrafo dirigido	Humberto Laguna Galindo	2 posdoctorantes, 1 alumno de maestría
Área Académica de Química Inorgánica		
Estudio teórico experimental de la N-formilación de aminoácidos en ausencia de anhídrido acético	Alejandro Islas Jácome	1 profesor titular, 1 profesor curricular, 1 alumno de maestría, 1 servicio social

La documentación correspondiente se anexa a esta solicitud.

Agradezco su atención a la presente, le envío un cordial saludo y quedo a sus órdenes para cualquier duda o aclaración.

Atentamente
Casa abierta al tiempo



Dr. Jorge Garza Olguín
Jefe del Departamento de Química

UNIDAD IZTAPALAPA

División de Ciencias Básicas e Ingeniería

Departamento de Química

Av. Ferrocarril San Rafael Atlixco 186. Col. Leyes de Reforma 1A Sección. Alcaldía Iztapalapa. 09310, CdMx, México. Edificio R primer piso. Oficina R-118

E-mail: @izt.uam.mx. <http://www.quimica.izt.uam.mx>

1. Nombre del proyecto:

Correlaciones de alto orden en el espacio químico representado en un hipergrafo dirigido

2. Nombre del responsable y tiempo de dedicación al proyecto:

Dr. Humberto Laguna Galindo, Profesor Titular C de Tiempo Completo, Área Académica de Química Cuántica, Departamento de Química

Tiempo de dedicación: 15 horas

3. Nombre de los participantes y tiempo de dedicación al proyecto:

Dr. Ángel Alejandro García Chung, Departamento de Química, realiza estancia posdoctoral.

Tiempo de dedicación: 10 horas

Dr. José Antonio Zárate Colín, Departamento de Química, realiza estancia posdoctoral.

Tiempo de dedicación: 5 horas

Q. Yaffet Zambrano González Departamento de Química, alumno de maestría.

Tiempo de dedicación: 4 horas

4. Área, Departamento y División

Área Académica de Química Cuántica, Departamento de Química, División de Ciencias Básicas e Ingeniería.

5. Objetivos generales y particulares

General

Realizar un estudio exhaustivo de las correlaciones estadísticas de alto orden presentes en hipergrafos orientados completos y en hipergrafos dirigidos completos, que representan el espacio químico para distintos números de sustancias.

Estudiar y caracterizar hipergrafos aleatorios en un esquema similar a los de Erdős-Renyi y de Watts-Strogatz, en particular, nos interesa estudiar las transiciones de fase que éstos puedan presentar como función de la probabilidad, y la capacidad de las medidas de correlación de alto orden de captarlas, evaluando también la capacidad de las medidas de correlación por pares.

Particulares

Establecer las diferencias y similitudes entre las distintas distribuciones de probabilidad que resultan de considerar las estructuras tanto de las sustancias como de las reacciones, a través de las medidas de correlación de alto orden.

Identificar las transiciones de fase en el modelo tipo Erdős-Renyi del hipergrafo orientado aleatorio y del hipergrafo dirigido aleatorio.

Implementar y explorar la estructura de "pequeño mundo" en un modelo de Watts-Strogatz para hipergrafos orientados empleando las medidas de información definidas anteriormente.

Comparar el esquema anterior, diseñado para un hipergrafo completo (orientado o dirigido) con los datos de Reaxys apuntando al estudio de posibles correlaciones que las medidas de información puedan poner de relieve.

Consolidar nuevas líneas de generación y aplicación del conocimiento para la formación de estudiantes de licenciatura y posgrado de la UAM-I en el estudio de la teoría de redes y sus aplicaciones a la química.

6. Antecedentes

Reaxys [1] es una base de datos que está formada por el registro anual de las sustancias y las reacciones químicas que fueron reportadas, tanto en revistas científicas como en patentes desde 1771. Contiene la información recolectada de más de 260 millones de sustancias y 73 millones de reacciones químicas indexadas, y esto la convierte en la base de datos de la química con mayor envergadura.

Estudiar esta cantidad de datos supone un reto mayor. De las varias perspectivas posibles para explorar los datos, en este proyecto nos enfocamos en una ruta muy concreta: el modelado de los datos, primero como un hipergrafo orientado y en segundo lugar como un hipergrafo dirigido, para el estudio de su estructura y la estructura de sus relaciones.

Proponemos desarrollar un modelo para estudiar correlaciones estadísticas de pares de variables aleatorias en los hipergrafos utilizando la información mutua, definida en el contexto de la teoría de la información, pues una caracterización de la base de datos empleando medidas informacionales, permitirá explicar y comprender las distintas regiones del enorme conjunto de datos, su grado de entrelazamiento así como su geometría y topología. Esto supone un resultado notable si se tiene en cuenta que se dispone de la información de más de 220 años de datos de sustancias y reacciones químicas, con lo cual nuestros análisis temporales de las medidas de información, aportarán un entendimiento robusto, basado en datos, sobre el comportamiento y evolución de la química como ciencia. Hecho que, además, permitirá realizar extrapolaciones para inferir tendencias para los siguientes años. En este proyecto pretendemos dar un paso más en esta caracterización calculando las correlaciones estadísticas de alto orden [2], es decir más allá de las correlaciones por pares.

Hasta la fecha, se ha utilizado un enfoque estadístico para encontrar patrones en los datos que permiten distinguir etapas de la ciencia (y hablar propiamente de etapas protorgánica, orgánica y organometálica) [3], así como los patrones que permiten ubicar características de la comunidad que trabaja con sustancias y reacciones químicas [4]. El conocimiento de los datos que se refleja en estos trabajos previos y en otros desarrollos actuales que estamos realizando, nos dan la pauta para responder las preguntas de esta investigación y también nos permiten diseñar los pasos para la investigación propuesta en este proyecto.

7. Descripción

El enfoque que proponemos adoptar para el estudio es el modelado de los datos para detectar y explicar patrones. Ello reviste un interés intrínseco para estudiar y describir la dinámica de la comunidad científica. Para el modelado, en este proyecto nos planteamos la relación entre el espacio de sustancias químicas (un espacio que tiene 260 millones de elementos hasta la fecha) y el espacio de las reacciones químicas (que tiene 73 millones de elementos hasta la fecha) a través de un hipergrafo orientado y de un hipergrafo dirigido, que se construye de la siguiente manera:

Se consideran las sustancias químicas de interés y las reacciones químicas entre ellas. Cada conjunto de reactivos constituye un hipernodo, y cada conjunto de productos otro hipernodo, y la arista entre ellos un hipervértice. Es importante señalar que los hipernodos

pueden compartir sustancias, pues cada una de ellas puede aparecer en diferentes reacciones químicas, como producto o como reactivo. Esto produce patrones diversos en el hipergrafo cuyas propiedades matemáticas son de interés.

Los hipergrafos orientados y los dirigidos en sí mismos revisten interés teórico que nos planteamos estudiar en este proyecto. Se trata de objetos matemáticos de relativamente reciente aparición en la literatura científica [6, 7, 8] cuyas propiedades matemáticas están bajo estudio [9, 10, 11, 12]. El interés en los hipergrafos orientados o dirigidos proviene de formular extensiones a la teoría de redes [13, 14], hoy plenamente consolidada y con crecientes aplicaciones, que permitan estudiar aspectos que no aparecen de forma explícita en los modelos de redes más conocidos, es decir, que los hipergrafos (orientados y/o dirigidos) son estructuras de orden mayor y por lo tanto, contienen más información que la contenida en los grafos. Nos planteamos por tanto, realizar un estudio de sus características topológicas utilizando teoría de la información [15], para detectar correlaciones estadísticas de pares y correlaciones estadísticas de alto orden [2, 16] entre hipernodos que nos permitan caracterizar al hipergrafo orientado y/o dirigido.

Proponemos utilizar modelos de pocas sustancias antes de abordar el estudio de la red real construida con Reaxys.

Por otro lado, hipergrafos orientados o dirigidos pueden construirse a partir de la base de datos Reaxys, seleccionando un conjunto de sustancias y estableciendo las reacciones químicas entre ellas que han sido registradas en la base de datos. Estos hipergrafos se distinguirán de las versiones de hipergrafos orientados o dirigidos completos en que no aparecerán todos los hipervértices, pues las leyes de la química, la física, e incluso las contingencias sociales (guerras, disponibilidad de reactivos, costo de los experimentos, por mencionar algunos factores) determinan las reacciones químicas que efectivamente se han llevado a cabo, y por lo tanto los hipervértices que sí aparecerán. Se trata de la química real, tal como ha sido en estos últimos 220 años.

La comparación entre los hipergrafos de la química real y las químicas idealizadas reviste interés teórico para el estudio de los hipergrafos como objetos matemáticos, pero también las herramientas de la teoría de la información que se habrán desarrollado en el caso de los hipergrafos orientados o dirigidos completos, serán de utilidad para estudiar el espacio químico real.

Con este proyecto, esperamos obtener una descripción exhaustiva de las correlaciones de alto orden en la estructura de los hipergrafos completos (orientados y dirigidos) utilizando teoría de la información. También el entendimiento del rol de la estructura de las sustancias y de las reacciones en las distribuciones de probabilidad que resulten de las variables aleatorias construidas.

Nos enfocaremos también en el estudio y determinación de cotas para las entropías de Shannon en el contexto de los hipergrafos orientados y dirigidos. Este tema es de importancia para la comunidad de teoría de la información y forma parte de la caracterización de la estructura de los hipergrafos orientados y dirigidos en la medida en que las medidas informacionales aportan entendimiento y comprensión sobre la misma. Un aspecto a enfatizar en este punto, es la sospecha del rol que pueden tener las medidas de

la información en el análisis del surgimiento de estructuras ya que los hipergrafos son estructuras de orden mayor que los grafos.

Además estudiaremos la detección de patrones (ciclos de sustancias relacionadas entre sí, como un ejemplo notable de lo que estaremos buscando, pero no limitados a eso) en el hipergrafo orientado completo y el hipergrafo dirigido completo, su entendimiento para la posterior detección de patrones en el hipergrafo construido a partir de los datos de Reaxys.

Adicionalmente, estudiaremos para su caracterización, las transiciones de fase que ocurren en hipergrafos orientados y dirigidos aleatorios empleando medidas de correlación de alto orden. Este enfoque es novedoso y promete un camino muy interesante para el estudio de los fenómenos críticos en redes complejas utilizando medidas de correlación más allá de las de pares.

Por último, esperamos obtener un entendimiento del tamaño y complejidad de las subregiones de la química existentes en Reaxys, así como su cercanía a ser un “pequeño mundo”. Esperamos descubrir las zonas menos exploradas y aquellas que están en marcado auge.

Referencias:

- [1] Reaxys, www.reaxys.com
- [2] Salazar SJC, Laguna HG, Sagar RP, 2022, Pairwise and higher-order statistical correlations in excited states of quantum oscillator systems, *Eur. Phys. J. Plus*, 137, 19 (doi: 10.1140/epjp/s13360-021-02215-z)
- [3] Llanos EJ, Leal W, Luu DH, Jost J, Stadler PF, Restrepo G, 2019, Exploration of the chemical space and its three historical regimes, *Proc. Nat. Acad. Sci.*, 116, 12660.
- [4] Leal W, Llanos EJ, Bernal A, Stadler PF, Jost J, Restrepo G, 2022, The expansion of chemical space in 1826 and in the 1840s prompted the convergence to the periodic system, *Proc. Nat. Acad. Sci.*, 119, e2119083119.
- [5] Garcia-Chung A, Bermúdez-Montaña M, Stadler PF, Jost J, Restrepo G, 2023, Chemically inspired Erdős-Rényi oriented hypergraphs, arXiv:2309.06351v1 [cs.DM].
- [6] Mulas R, Horak D, Jost J, 2022 *Graphs, Simplicial Complexes and Hypergraphs: Spectral Theory and Topology*. Cham: Springer International Publishing. (doi: 10.1007/978-3-030-91374-8\ 1).
- [7] Jost J, Mulas R, 2019 Hypergraph Laplace operators for chemical reaction networks. *Advances in Mathematics* 351, 870–896.
- [8] Ausiello G, Laura L, 2017 Directed hypergraphs: Introduction and fundamental algorithms—A survey. *Theoretical Computer Science* 658, 293–306. (doi:<https://doi.org/10.1016/j.tcs.2016.03.016>).
- [9] Eidi M, Jost J, 2019 Ollivier Ricci curvature of directed hypergraphs. *Scientific Reports* 10, 1, 12466.
- [10] Mulas R, Kuehn C, Jost J, 2020 Coupled dynamics on hypergraphs: Master stability of steady states and synchronization. *Physical Review E* 101, 062313. (doi:10.1103/PhysRevE.101.062313).
- [11] Sun H, Bianconi G, 2021 Higher-order percolation processes on multiplex hypergraphs. *Phys. Rev. E* 104, 034306. (doi:10.1103/PhysRevE.104.034306).
- [12] Thakur M, Tripathi R, 2009 Linear connectivity problems in directed hypergraphs.

Theoretical Computer Science 410, 27, 2592–2618.

(doi:<https://doi.org/10.1016/j.tcs.2009.02.038>).

[13] Boccaletti S, De Lellis P, del Genio C, Alfaro-Bittner K, Criado R, Jalan S, Romance M, 2023 The structure and dynamics of networks with higher order interactions.

Physics Reports 1018, 1–64. (doi:<https://doi.org/10.1016/j.physrep.2023.04.002>).

[14] Chodrow PS, 2020 Configuration models of random hypergraphs. Journal of Complex Networks 8, 3, cnaa018. (doi:10.1093/comnet/cnaa018).

[15] Cover TM, Thomas JA, Elements of Information Theory, (Wiley, New York, 1991)

[16] Salazar SJC, Laguna HG, Sagar RP, 2020, Higher-Order Information Measures from Cumulative Densities in Continuous Variable Quantum Systems, Quantum Rep. 2020, 2, 560–578 (doi:10.3390/quantum2040039).

8. Recursos disponibles para el desarrollo del proyecto.

El proyecto requiere de equipo de cómputo personal y acceso a tiempo de súper cómputo para realizar algunos de los cálculos. Los equipos de cómputo personal están disponibles en el Área Académica de Química Cuántica

9. Infraestructura actual en la Universidad disponible para el proyecto.

La Universidad tiene disponible equipo de súper cómputo para la realización de los cálculos necesarios para el desarrollo del proyecto.

10. Fuentes de financiamiento.

El proyecto fue apoyado por dos años en el marco de la “Convocatoria para Postulación de Proyectos de Investigación por Personal Académico de Ingreso Reciente” de la Dirección de Apoyo a la Investigación (DAI) de la UAM.

11. Indicadores de desempeño

Investigación:

- a. 2 artículos de investigación en revistas internacionales de arbitraje estricto.
- b. 3 participaciones en congresos nacionales.

Docencia

- c. 1 asesoría de proyecto terminal
- d. 1 asesoría de tesis de maestría

Preservación y divulgación de la cultura

- e. 1 artículo de divulgación.
- f. 1 seminario especializado para presentar los resultados en el Departamento de Química de la UAM o en el Área Académica de Química Cuántica.

12. Fecha de inicio, duración y planeación a dos años.

Primer año. Julio de 2024 a junio de 2025

Actividades:

- Construcción y caracterización del modelo de hipergrafo completo para distintas triadas de variables aleatorias. Desarrollo de las simulaciones correspondientes empleando matlab.
- Construcción y caracterización del modelo de hipergrafo a la Erdős-Renyi y a la Watts-Strogatz. Desarrollo de las simulaciones correspondientes empleando Matlab.

Entregables:

- Participación en congreso

- Publicación de un artículo científico con los resultados obtenidos.

Segundo año. Julio de 2025 a junio de 2026

Actividades:

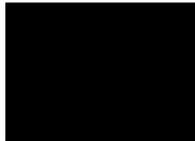
- Elaboración de los hipergrafos reales utilizando la base de datos Reaxys.
- Comparación de los modelos estudiados en el primer año con los hipergrafos reales construidos con Reaxys.
- Aplicación de las distintas medidas de correlación de alto orden a los datos de Reaxys y las distintas regiones de la química.

Entregables:

- Participaciones en congreso
- Redacción y envío para su publicación de un artículo científico con los resultados obtenidos.
- Seminario especializado.
- Redacción y envío para su publicación del artículo de divulgación.

13. Palabras clave

Hipergrafos, entropía de Shannon, Erdős-Renyi, Watts-Strogatz, Reaxys.



Dr. Humberto Laguna Galindo

Responsable Técnico. Profesor Titular C, T.C.

Departamento de Química

Universidad Autónoma Metropolitana-Iztapalapa (UAM-I)