



PROYECTO DE INVESTIGACIÓN INTERNOS SIN FINANCIAMIENTO O
AUTOGESTIONADOS

ANEXO 1 - DATOS INFORMATIVOS

Fecha de presentación (dd/mm/aa): 04/03/2020

Título del proyecto: Desarrollo de un algoritmo computacional para contar ciclos primarios en redes complejas.

TIPO DE INVESTIGACIÓN

Investigación básica

Investigación aplicada

DEPARTAMENTO(S) Y/O INSTITUTO(S):

1. Departamento de Física
- 2.

LÍNEA(S) DE INVESTIGACIÓN (verificable en el SAEW):

1. Física Aplicada
- 2.

RESUMEN DE INFORMACIÓN DEL DIRECTOR Y COLABORADORES

Director

| Apellidos y nombres | No. de Cédula | HSS | Departamento | Título de mayor nivel y mención. |
|---------------------|---------------|-----|------------------------|----------------------------------|
| Xulvi-Brunet, Ramon | 1757068646 | 5 | Departamento de Física | PhD |

Colaborador(es)

| Apellidos y nombres | No. de Cédula | HSS | Departamento | Título de mayor nivel y mención. |
|---------------------|---------------|-----|--------------|----------------------------------|
| | | | | |
| | | | | |
| | | | | |

Colaboradores Externos

| Apellidos y nombres | No. de identificación | HSS | Institución | Título de mayor nivel y mención. |
|---------------------|-----------------------|-----|-------------|----------------------------------|
| | | | | |
| | | | | |
| | | | | |

* HSS = Horas Semana Semestre

HOJA DE VIDA DEL DIRECTOR DEL PROYECTO



PROYECTO DE INVESTIGACIÓN INTERNOS SIN FINANCIAMIENTO O
AUTOGESTIONADOS

ANEXO 2 – DETALLES DE LA PROPUESTA

| | |
|--|--|
| Investigación Básica | Investigación Aplicada <input checked="" type="checkbox"/> |
| DEPARTAMENTO(S) Y/O INSTITUTO(S): | |
| 1. Departamento de Física | |
| 2. | |
| LINEA(S) DE INVESTIGACIÓN: | |
| 1. Física Aplicada | |
| 2. | |

| DISCIPLINA CIENTÍFICA (Marque X, solamente una opción) | |
|---|-------------------------------------|
| Ciencias Naturales y Exactas; | <input checked="" type="checkbox"/> |
| Ingeniería y Tecnologías; | <input type="checkbox"/> |
| Ciencias Médicas; | <input type="checkbox"/> |
| Ciencias Agrícolas; | <input type="checkbox"/> |
| Ciencias Sociales; | <input type="checkbox"/> |
| Humanidades | <input type="checkbox"/> |

| OBJETIVO SOCIOECONÓMICO (Marque X, solamente una opción) | |
|--|-------------------------------------|
| Exploración y explotación del medio terrestre; | <input type="checkbox"/> |
| Ambiente; | <input type="checkbox"/> |
| Exploración y Explotación del espacio; | <input type="checkbox"/> |
| Transporte, telecomunicaciones y otras infraestructuras; | <input type="checkbox"/> |
| Energía; | <input type="checkbox"/> |
| Producción y tecnología industrial; | <input type="checkbox"/> |
| Salud; | <input type="checkbox"/> |
| Agricultura; | <input type="checkbox"/> |
| Educación; | <input type="checkbox"/> |
| Cultura, ocio, religión y medios de comunicación; | <input type="checkbox"/> |
| Sistemas políticos y sociales, estructuras y procesos; | <input type="checkbox"/> |
| Defensa; | <input type="checkbox"/> |
| Avance general del conocimiento: I+D financiada con los Fondos Generales de Universidades (FGU); | <input type="checkbox"/> |
| Avance general del conocimiento: I+D financiados con otras fuentes. | <input checked="" type="checkbox"/> |



| | |
|---|---|
| 1 | Proyecto de Investigación |
| | Título: Desarrollo de un algoritmo computacional para contar ciclos primarios en redes complejas. |
| | Resumen del proyecto (máximo 200 palabras) <p>Desde el inicio del estudio moderno de las redes complejas, hace unos 20 años, se han desarrollado varias medidas para caracterizar topológicamente las redes. Ejemplos de estas medidas son el "shortest average path length", el "clustering coefficient", el "degree distribution", la función de distribución "degree-degree correlation", la maximización de la "modularidad" para cuantificar la estructura de comunidades, etc.</p> <p>A pesar de la gran cantidad de medidas y conceptos desarrollados durante estos 20 años, una caracterización general de una red compleja a través de este conjunto finito de medidas no se ha logrado todavía. En particular, no tenemos una manera práctica de caracterizar la estructura de ciclos (camino cerrado que no pasan más de una vez por cada uno de los nodos que definen el camino) dentro de una red compleja.</p> <p>El objetivo de este proyecto es desarrollar un algoritmo computacional que sea capaz de obtener la estructura de "ciclos primarios" de una red compleja cualquiera. Es decir, un algoritmo que nos diga cuantos circuitos primarios de longitud "n" existen en la red, para todo "n". Esta estructura de ciclos primarios se presume ser una medida topológica importante que podría dar pie incluso a clasificar las redes complejas bajo nuevas categorías.</p> |
| | Palabras clave (4-6): redes complejas, topología de redes, ciclos |



| | |
|---|--|
| 2 | Objetivos, relevancia, productos y resultados esperados de esta propuesta de investigación |
|---|--|

2.1 Objetivos

2.1.1 Objetivo General

- Desarrollar un algoritmo computacional capaz de medir la distribución de ciclos primarios de una red compleja cualquiera y utilizarlo para clasificar redes complejas reales bajo los criterios de esta distribución de ciclos.

2.1.2 Objetivos Específicos

- a. Introducir dentro de campo de las redes complejas el concepto de ciclos primarios, mostrando la importancia que éstos pueden tener en la caracterización de las redes complejas. (Se explicará más adelante porqué es importante introducir este tipo de ciclos).
- b. Desarrollar un algoritmo que sea capaz de computar la distribución de ciclos primarios de una red cualquiera. Es decir, el algoritmo nos tiene que dar un "cycle sequence", de la misma manera que se tiene el "degree sequence" de una red.
- c. Aplicar el algoritmo a distintos tipos de redes reales (biológicas, sociales, tecnológicas, etc) y obtener las distribuciones de ciclos típicas para ellas.
- d. Establecer una clasificación de las redes reales en función de las características de las distribuciones de ciclos primarios obtenidas.

2.2 Detalle de los resultados esperados (con relación a los objetivos)

- a. Definición, ejemplos y descripción de la importancia de los ciclos primarios.
- b. El algoritmo para obtener los ciclos primarios de una red cualquiera.
- c. Un conjunto de distribuciones de ciclos típicas correspondientes a un conjunto de redes reales.
- d. Una clasificación de las redes reales en función de la distribución de ciclos primarios.

| | |
|---|---|
| 3 | Relevancia de la propuesta de investigación y su relación con la(s) líneas de investigación |
|---|---|

Las redes complejas se han convertido hoy en día una base metodológica que permite entender la estructura subyacente de muchos sistemas complejos. De hecho, muchos sistemas reales, tanto naturales como hechos por el hombre, son básicamente redes: pensemos en la "gene co-expression network" (una red que describe la estructura del sistema de coexpresión de genes), la "protein-protein interaction network" (la red que modela la estructura de interacción entre proteínas), etc.

La estructura, o topología, de una red describe no sólo la base subyacente de un sistema complejo sino que también es clave para entender los procesos dinámicos que tienen lugar sobre ella. Pensemos, por ejemplo, en el área de la epidemiología. Hoy en día tenemos el caso de la propagación del Corona Virus hacia distintos países del mundo. Esta propagación del virus tiene una descripción estocástica en función de parámetros como la probabilidad de infección, tiempo de incubación, etc. Pero todo este proceso dinámico, que se modela estocásticamente, corre sobre una red de contactos entre humanos, y la estructura de esta red de contactos es clave en la propagación del virus.

Así pues, caracterizar de la forma más completa posible una red compleja es esencial para entender tanto la estructura de la misma como para entender los procesos dinámicos que ocurren sobre ella. El gran avance que se ha hecho en los últimos 20 años en la comprensión de la interacción entre proteínas, en cómo se expresan los genes dentro de un célula, etc, se ha debido a la introducción de nuevas medidas y conceptos topológicos capaces de caracterizar cada vez mejor la estructura de las redes complejas.

La motivación para este proyecto es precisamente caracterizar de forma cada vez más completa la estructura topológica de una red, con el fin de entender mejor aquellos sistemas que se pueden



describir mediante redes y comprender en mayor profundidad los procesos dinámicos que ocurren sobre ellas.

Los métodos de la física, en particular los del área de los sistemas complejos y la física estadística, aplicados a este tipo de estudios, es decir, la física aplicada, han mostrado ser muy capaces de aportar al entendimiento de la topología de redes y, a través de este conocimiento, aportar al entendimiento de muchos sistemas reales.

| | |
|---|---|
| 4 | Productos esperados (marcar con una "X" al menos uno de los productos no señalados) |
|---|---|

| Tipo de Producto: | Marcar con una "X" |
|--|--------------------|
| a. Disertación a la Comunidad Politécnica (obligatorio); | X |
| b. Presentación de un artículo en formato de la Revista Politécnica (obligatorio) | X |
| c. Proyecto de Titulación; | |
| d. Aplicación tecnológica construida o implementada; | |
| e. Patente presentada; | |
| f. Perfil de proyecto de mayor impacto científico, técnico, pedagógico o de innovación. | |
| g. Publicaciones científicas indexada en SCIMAGO-SCOPUS/WoS/SCIELO/Latindex Catálogo o un artículo en congreso indexado en SCOPUS. | |

| | |
|---|---|
| 5 | Descripción y metodología y diseño del proyecto |
|---|---|

5.1 Descripción, metodología y diseño del proyecto (Máximo dos carillas)

El problema de establecer una nueva medida topológica relacionada con una distribución de ciclos en redes es que encontrar todos los posibles ciclos de una red es sumamente difícil, principalmente, porque hay muchos ciclos que se pueden ver como "envoltura" de, a su vez, varios ciclos. Como ejemplo, pensemos en dos ciclos que compartan algunos ejes; en este caso, un nuevo ciclo se puede formar por los ejes no comunes de los dos ciclos.

Una manera de resolver el problema es definir lo que podríamos llamar ciclos primarios, es decir, ciclos que no "contengan" otros ciclos. En el fondo, estos ciclos primarios son los que importan (al menos en una gran cantidad de aplicaciones). Por ejemplo, en "network epidemiology" [1], muchos de los resultados teóricos que se obtienen en relación a si una epidemia se puede dar o no asumen que, en una red infinita, los ciclos de menor longitud (que corresponden a los primarios que hemos definido), son de longitud infinita. Se asume también que, para redes finitas, los ciclos más importantes a tener en cuenta serían los primarios.

Resulta, por tanto, muy importante tener una medida de la distribución de ciclos primarios en una red, esto es, saber cuantos ciclos primarios de una longitud dada, "n", hay en la red bajo estudio.

El algoritmo computacional que se va a desarrollar para, de forma totalmente general, contar cuantos ciclos primarios de una longitud dada existen en una red, se basa en un algoritmo ya establecido en el campo. El algoritmo en cuestión [2] se basa a su vez en la idea de la tomografía de redes, y se utiliza usualmente para el cálculo del shortest average path length y otras propiedades topológicas. La idea para el desarrollo del nuevo algoritmo es que una pequeña variación del algoritmo



tomográfico, aplicado en modo "forward" and "backward" (y esto es clave), permite encontrar, y por tanto contar, los ciclos primarios.

Una vez implementado el algoritmo, se pretende aplicarlo sobre varios tipos de redes reales (redes neuronales, proteínicas, de expresión de genes, sociales, tecnológicas, etc [3-6]) para obtener la distribución de ciclos típica asociada a cada tipo de red. Esto se va a hacer con dos objetivos en mente: 1) Establecer una clasificación de las redes reales en función de la distribución de ciclos primarios que tengan. 2) Mostrar que la definición de ciclo primario utilizada es útil a nivel práctico, porque, de hecho, nos permite clasificar redes reales.

El proyecto es puramente teórico y computacional, por lo que sólo se requiere acceso a poder computacional.

Proyectos muy similares a este ya han probado su interés en el área. Baste decir que Mark Newman introdujo, a inicios del presente siglo (unos 18 años atrás), una medida básica para medir la correlación entre los grados de los nodos de una red [7,8]. Esta medida, a pesar de haber sido ya superada [9], le permitió, tras aplicarla a unas pocas redes reales de la época, establecer una clasificación de redes reales que hoy en día es muy utilizada [10]: redes asortativas y redes disortativas. Él vió que redes sociales suelen ser asortativas y que redes biológicas y tecnológicas son disortativas [10].

A parte del posible interés a nivel epidemiológico, donde, como hemos dicho, se presume que los ciclos primarios juegan un papel importante, la nueva medida podría acabar abriendo una subarea dentro de las redes complejas (es, por cierto, lo que acabó pasando con las correlaciones de Mark Newman). Por ejemplo, dentro del área la neurociencia computacional, aún no se conoce muchas propiedades de una red de neuronas tan compleja como el cerebro [11]. Conocer la distribución de los ciclos primarios de un cerebro puede resultar muy importante.

En resumen, la metodología para el proyecto estaría basada en el algoritmo de tomografía de redes y el diseño del proyecto tendría, básicamente, dos partes: 1) implementación computacional del algoritmo basado en el de tomografía y 2) aplicación del mismo para el estudio de redes complejas reales.

[1] Leon Danon et al. (2011). Networks and the Epidemiology of Infectious Disease. *Interdisciplinary Perspectives on Infectious Diseases*, Volume 2011, Article ID 284909, 1-28, doi:10.1155/2011/284909.

[2] R. Xulvi-Brunet et al. (2003). Correlations in Scale-Free Networks: Tomography and Percolation. *Physical Review E*, 68, 036119. DOI:<https://doi.org/10.1103/PhysRevE.68.036119>.

[3] Alan Mislove et al. (2007). Measurement and analysis of online social networks. *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. October 24-26, 2007, San Diego, California, USA.

[4] Laura Garton et al. (1997). Studying Online Social Networks. *Journal of Computer-Mediated Communication*, Volume 3, Issue 1, 1 June 1997, JCMC313, <https://doi.org/10.1111/j.1083-6101.1997.tb00062.x>

[5] U. Alon. (2003) Biological networks: the tinkerer as an engineer. *Science* 26 Sep 2003: Vol. 301, Issue 5641, pp. 1866-1867. DOI: 10.1126/science.1089072

[6] Justin Balthrop et al. (2004). Technological Networks and the Spread of Computer Viruses. *Science* 23 Apr 2004: Vol. 304, Issue 5670, pp. 527-529. DOI: 10.1126/science.1095845

[7] M. E. J. Newman (2002). Assortative Mixing in Networks. *Phys. Rev. Lett.* **89**, 208701

[8] M. E. J. Newman and Juyong Park (2003). Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122



- [9] R. Xulvi-Brunet and I.M. Sokolov (2005). Changing correlations in networks: assortativity and dissortativity. *Acta Physica Polonica B*, Vol. 36, 1431-1455.
- [10] M. E. J. Newman (2003). The Structure and Function of Complex Networks. *SIAM Rev.*, 45(2), 167–256.
- [11] David Meunier et al. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, Volume 4, 1-11. doi: 10.3389/fnins.2010.00200

6 Infraestructura, equipos y fondos adicionales.

6.1 Infraestructura y equipos

- *Laboratorio de Computación Avanzada del Departamento de Física, dotado de un servidor DELL de 32 cores. El servidor se encuentra en el edificio #12 en el laboratorio 507.*

| Infraestructura | Equipos | |
|-------------------------------------|---|--|
| Nombre del laboratorio | Nombre del Equipo | Ubicación del Equipo |
| Laboratorio de Computación Avanzada | Servidor DELL de la Facultad de Ciencias: 2 Procesadores Intel XEON Silver 2.1GHz, 32 cores. RAM 32 GB RDIMM Dual Rank, 224 GB SSD, 10 TB 7200 RPM SATA. Rack autocontenido, con UPS y refrigeración autónoma. | Room 507 del Edificio de Ciencias, Departamento de Física. |

6.2 Breve justificación del equipo requerido

- No aplica.

6.3 Fondos Adicionales

- No aplica.



PROYECTO DE INVESTIGACIÓN INTERNOS SIN FINANCIAMIENTO O
AUTOGESTIONADOS

ANEXO 4 - DECLARACIÓN

TIPO DE INVESTIGACIÓN

| | |
|----------------------|--------------------------|
| Investigación básica | Investigación aplicada X |
|----------------------|--------------------------|

TÍTULO DEL PROYECTO

Desarrollo de un algoritmo computacional para contar ciclos primarios en redes complejas.

DECLARACIÓN DEL DIRECTOR DEL PROYECTO

El equipo de investigadores, representado por el Director del Proyecto declara lo siguiente:

- Que el presente proyecto es una creación original de mi autoría y del equipo de investigadores, y por tanto asumimos la completa responsabilidad legal en caso de que un tercero alegue la titularidad de los derechos intelectuales del proyecto, exonerando a la EPN de cualquier acción legal que se derive por esta causa.
- Que el presente proyecto no ha sido presentado en ninguna convocatoria de otra institución pública o privada. El incumplimiento será causal para que el proyecto no sea tomado en consideración.
- Que todos los bienes adquiridos en proyecto permanecerán bajo la custodia y responsabilidad del director de proyecto durante la ejecución del mismo.
- Que si el proyecto genera algún producto o procedimiento susceptible de obtener derechos de propiedad intelectual, de los cuales se deriven beneficios, aceptamos que éstos serán compartidos entre los investigadores y la institución o las instituciones participantes en el proyecto, conforme a lo establecido en el COESC.
- Que el equipo de investigadores y/o instituciones participantes se comprometen a mantener la confidencialidad de la información si ésta podría ser susceptible de protección por patentes, y solicitar la valoración de propiedad intelectual respectiva previa a cualquier publicación o difusión.
- Que para el caso de derechos de autor otorgamos una licencia de uso exclusivo con fines académicos para la o las instituciones participantes en el proyecto.



ESCUELA POLITÉCNICA NACIONAL
VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y VINCULACIÓN



Firma del Director del Proyecto

Nombre: Ramon Xulvi Brunet

C.I.: 1757068646

