

Trabajo #2 Comparación con modelos nulos. Luis Luarte. Redes sociales. DCCS. UDD.

Estadísticas descriptivas.

	Neuronas (nodos)	Sinapsis (links)	Diámetro (estimación)	Longitud de camino medio (estimación)	Densidad	Distribución de grado [primer, tercer cuartil]	Clusterizacion media
Red completa	32 172	2 258 042	28	4.33	0.002	[6, 93] _{in} , [2, 98] _{out}	0.41
MB	11 690	1 277 773	34	3.65	0.009	[11, 175] _{in} , [6, 192] _{out}	0.44
CX	7124	419 503	15	3.13	0.008	[5, 94] _{in} , [0, 95] _{out}	0.63
SEZ	13 357	372 425	20	4.38	0.002	[3, 29] _{in} , [2, 35] _{out}	0.26

Medidas de centralidad.

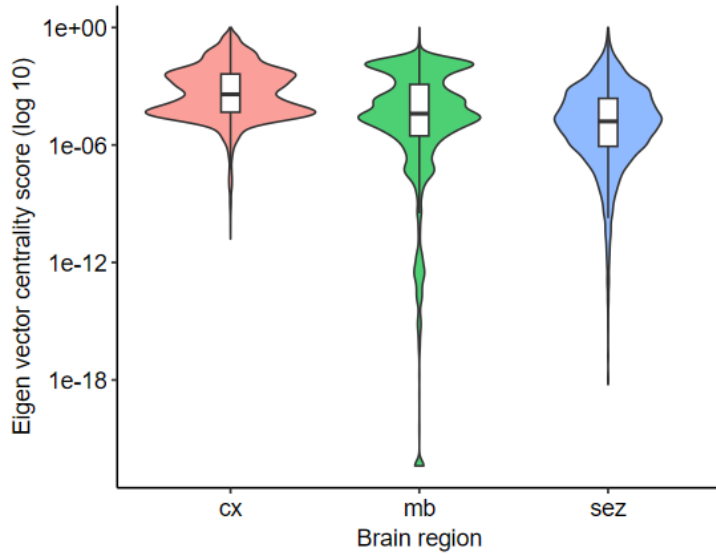
Centralidad de auto-vector: La relevancia de la centralidad de auto-vector esta directamente vinculada a las demandas del algoritmo de aprendizaje de diferencia temporal. La regla de actualización del algoritmo TD(0)

$$V(s_t) \leftarrow V(s_t) + \alpha \cdot \delta_t$$

Donde,

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

Requiere de un proceso de integración temporal, donde un valor almacenado, $V(s_t)$, esta continuamente siendo ajustado por la señal de error δ . Esto requiere de un sustrato capaz de mantener una representación persistente pero actualizable de la función de valor. La centralidad de auto-vector es una métrica relevante para identificar este sustrato en tanto considera el núcleo recurrente de un red-una estructura con una arquitectura ideal para sostener una actividad “auto-sustentada”, que logra sostener el valor de $V(s_t)$ en el tiempo. Por lo tanto, esta métrica provee de un vinculo cuantitativo entre la estructura física del circuito y su capacidad de realizar una computación basada en memoria requerida por algoritmo TD. Debido a que mi hipótesis (generada en la entrega anterior) conceptualiza una $f_{integracion}(W)$ y que esta se encuentra en alguna de las subredes, el análisis se enfoca por subred el cual debería informar la potencial clasificación de cada red dada su funcionalidad estimada.

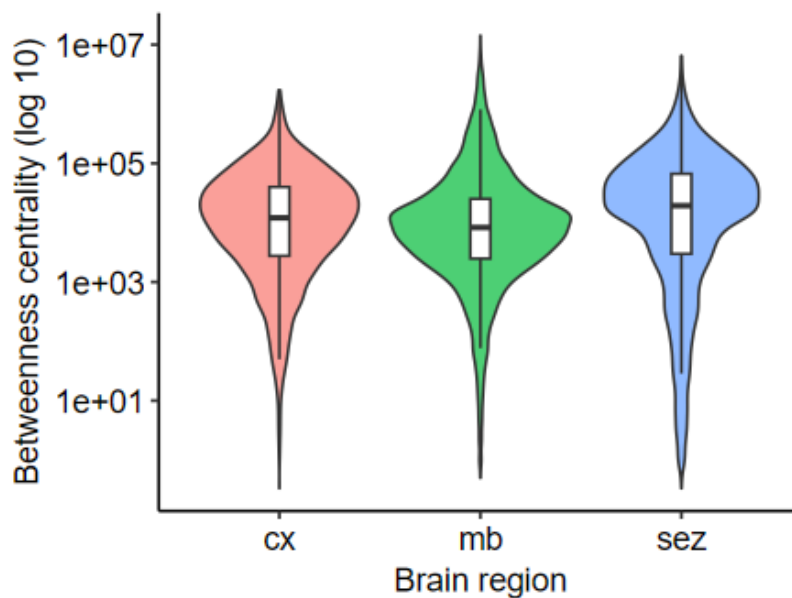


Evaluando la hipótesis de que el MB esta estructuralmente especializado para implementar la función de valor V_{π} . Inicialmente (Removiendo neuronas con centralidad de auto-vector cero que se encontraban desconectadas), no se aprecia una diferencia sustancial en la centralidad de auto-vector entre el MB y las restantes áreas. Sin embargo, se puede apreciar que el MB posee una distribución muy diferenciada, donde existe una población neuronal sustancial con muy baja centralidad (la cola larga) la cual convive en la red con otra de alta centralidad. Por lo anterior, observamos en MB una “cámara de eco” donde los nodos de alta influencia están conectados a otros nodos de alta influencia, que podría sustentar una circuitería reverberante capaz de sostener una representación estable. La población periférica de baja centralidad podría ser las “Kenyon cells” las cuales cumplen un rol de separación de patrones, el cual sería procesado y mantenido en línea por el grupo de neuronas MBON (las de alto grado de centralidad). Lo anterior habilita para realizar un testeo vía secuenciación genética de la correlación entre la centralidad de auto-vector y el tipo de población neuronal.

Centralidad de intermediación. La relevancia de la Centralidad de Intermediación proviene de la necesidad de modular el balance entre exploración y explotación en el marco del aprendizaje TD, un proceso gobernado por un parámetro de temperatura (τ) en una política softmax. Esto requiere un mecanismo neural capaz de controlar globalmente el flujo de información, ya sea restringiéndolo a rutas establecidas y de alto valor (explotación) o permitiendo el acceso a rutas novedosas (exploración). La Centralidad de Intermediación es, por lo tanto, la métrica ideal para esta función, ya que identifica con precisión los nodos que sirven como "puentes" o "controladores" críticos en los caminos de comunicación más cortos a través de la red. Al controlar estos pocos nodos de alta intermediación, el circuito puede abrir o cerrar eficientemente avenidas de comunicación completas, proporcionando un sistema centralizado para cambiar

globalmente la dinámica de la red. Así, esta métrica cuantifica la aptitud estructural de un circuito para implementar el control del flujo de información necesario para modular la estrategia de comportamiento de un agente.

Para el análisis de esta métrica se ocupó una metodología ligeramente distinta, en vez de computar la métrica a nivel de cada subred, se realizó inicialmente una computación a nivel de la red completa (se removieron neuronas con peso sináptico menor a 5 para acelerar la computación) y luego se mapearon a nivel de neurona la subred a la cual pertenecían.



Los resultados preliminares a nivel visual parecen mostrar una distribución de la métrica hacia valores mas altos en el SEZ comparados con las demás áreas. Esto se podría constituir como evidencia inicial a esta funcionalidad de neuronas “gatekeeper” para controlar las dinámicas de la red mediante las vías de comunicación centrales. Lo anterior, entonces podría apoyar la posibilidad de el control del parámetro τ asociado a la exploración-explotación, sea efectivamente implementado en el SEZ, tal como se menciono en la primera entrega como un regulador de la ganancia del sistema. Realizando un análisis estadístico simple con la métrica en \log_{10} como dependiente arroja que SEZ tendría efectivamente mayores valores de centralidad de intermediación (contraste de medias marginales bajo un modelo lineal, SEZ/CX, ratio = 1.161, SE = 0.07, $p = 0.037$; SEZ/MB, ratio = 1.348, SE = 0.06, $p < 0.01$)

Modelos nulos.

Tipo de red. Se espera que la red cerebral de la *Drosophila* sea una red de mundo pequeño, que es un tipo de red que esta altamente agrupada a nivel local y tiene rutas de comunicación cortas a nivel global. Se cree que esta estructura surge de una presión evolutiva para un procesamiento de información eficiente (caminos cortos) mientras se minimizan los costos de cableado (agrupamiento local). Además, la presencia de neuronas “hub” que hemos visto en los gráficos de centralidad sugiera que puede tener propiedades de red libre de escala, que podrían formarse por un mecanismo de crecimiento donde nuevas neuronas se conectan preferentemente a las ya importantes.

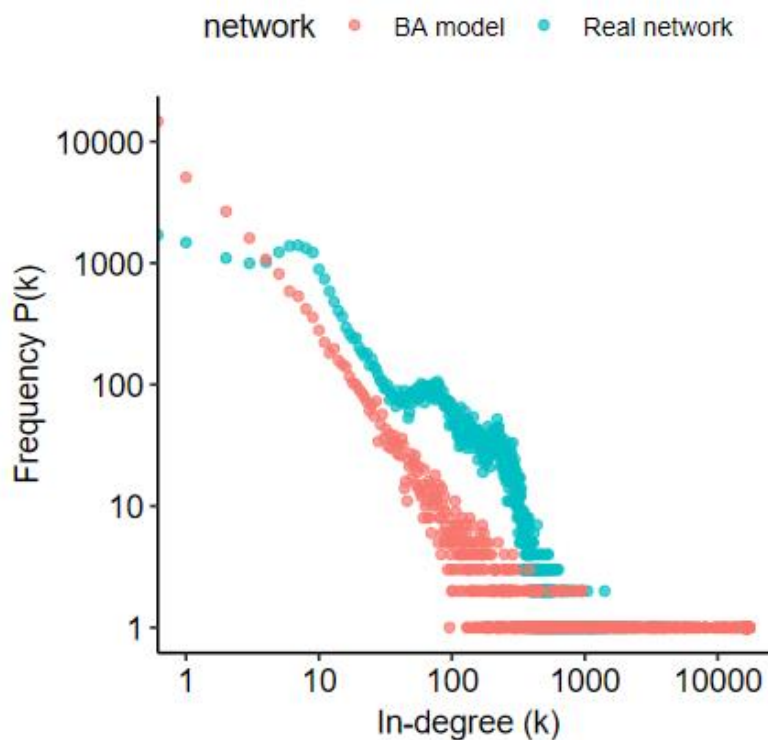
Hipótesis. La hipótesis central es que el conectoma de la *Drosophila* no es un grafo aleatorio, sino una estructura altamente organizada. La red cerebral de la *Drosophila* tendrá un coeficiente de clustering significativamente mayor y una longitud de camino promedio mas larga que un grafo aleatorio de Erdős-Rényi comparable con el mismo número de nodos y enlaces. La combinación en la red de un alto coeficiente de clustering y una corta longitud de camino promedio (en términos generales, no comparado con una red aleatoria Erdős-Rényi) es consistente con las propiedades de un modelo de mundo pequeño de Watts-Strogatz. Finalmente, la distribución de la red se aproximará a una ley de potencias, un sello distintivo de las redes libres de escala, y será fundamentalmente diferente de la distribución en forma de campana de un grafo aleatorio. Esto es consistente con una red formada por un modelo de conexión preferencial de Barabasi-Albert. Propongo que este último principio topológico es un resultado plausible del aprendizaje Hebbiano, donde la conectividad es impulsada por la actividad neuronal correlacionada, donde una neurona que es frecuentemente activa y contribuye de manera efectiva a la activación de sus vecinas fortalece sus conexiones sinápticas, lo que la hace mas “atractiva” como objetivo de nuevas conexiones.

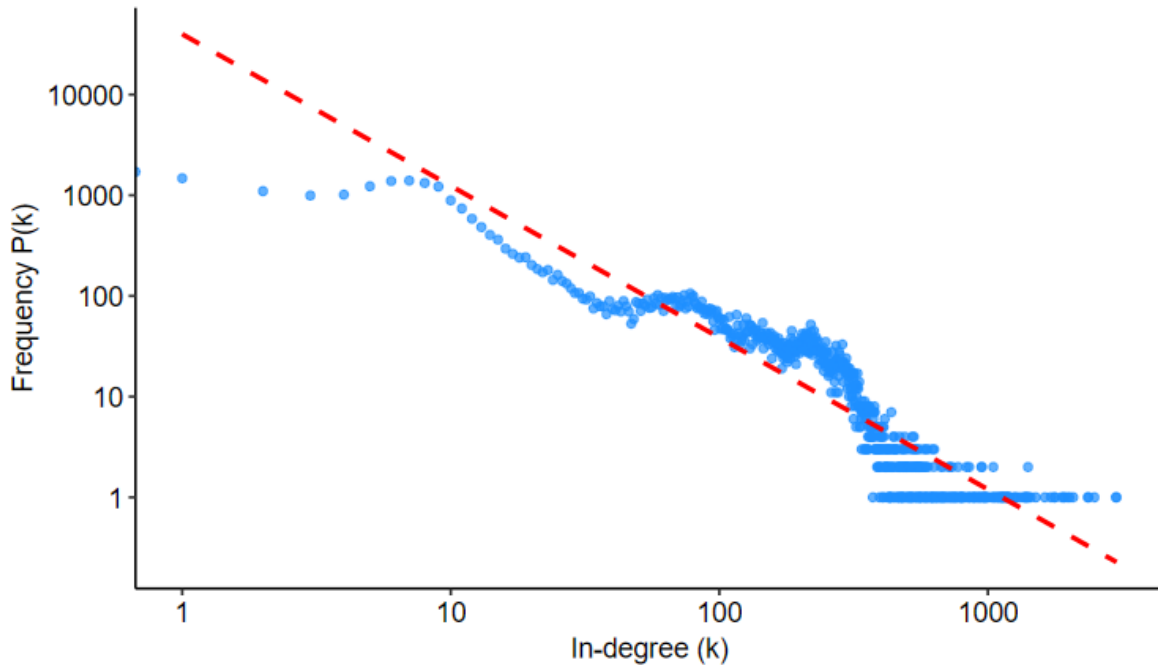
Erdős-Rényi (ER). Esta es la línea de base para “aleatoridad”. El punto aquí es mostrar que las propiedades de la red cerebral son muy diferentes del grafo ER. Como se esperaba, el coeficiente de clustering de la red cerebral se encuentra fuera (mayor) de la distribución esperada para una ER ($p = 0$), y una longitud de camino mayor ($p = 0$), lo que es esperado en tanto las señales en sistemas biológicos se espera tengan un flujo jerarquizado, pasando por secuencias específicas de áreas cerebrales lo que naturalmente aumenta la longitud del camino.

Small world (SM). Adicionalmente a los resultados anteriores, la computación del clustering real versus el ER $\gamma = 95.2$ mostrando un alto grado de estructura local, para la longitud de camino $\lambda = 1.71$, indicando un $\sim 71\%$ mas largo que una red aleatoria implausible mente eficiente en la transmisión de información. Finalmente, el coeficiente de SM $\sigma = \frac{\gamma}{\lambda} = 55.68$ muestra evidencia de una propiedad de mundo pequeño, esto es,

alto coeficiente de clustering con una alta eficiencia global (relacionada al largo de camino).

Barabasi-Albert (BA). Al visualizar las distribuciones del grado de entrada tanto de la red real como de un modelo comparable de BA en un gráfico log-log, es evidente que el conectoma de *Drosophila* exhibe claras características de red libre de escala, lo que apoya la hipótesis. La tendencia lineal descendente observada en la red real es consistente con la ley de potencias del modelo BA, sugiriendo que un mecanismo de crecimiento tipo "conexión preferencial" es un principio organizador clave en la formación de la red. Sin embargo, una comparación visual directa de las dos distribuciones graficadas revela que el modelo BA puro es una descripción incompleta, como lo demuestra la distintiva "protuberancia" en la distribución de la red real para grados de rango medio, una característica no predicha por el modelo simple. Estas desviaciones sugieren que, si bien el principio general de "los ricos se hacen más ricos" se mantiene, restricciones biológicas más complejas también juegan un papel crucial en la configuración de la arquitectura final. Todo lo anterior, se muestra de manera análoga mostrando la red real con el ajuste de una regresión lineal en la escala log/log.





Conclusiones.

Las estadísticas descriptivas por sí solas revelan profundas especializaciones estructurales entre las tres regiones cerebrales analizadas. El Mushroom Body (MB) se destaca como el módulo más grande y densamente interconectado, conteniendo más de la mitad de las sinapsis totales de la red y las neuronas con los grados de entrada y salida más altos. Esto, combinado con su gran diámetro, sugiere una arquitectura diseñada para un procesamiento complejo a través de largas vías internas. En contraste, el Complejo Central (CX) es el módulo más compacto, eficiente y, notablemente, el más clusterizado localmente, con el coeficiente de clusterización más alto de todos (0.63). Esta firma estructural es consistente con un circuito diseñado para una computación altamente local y competitiva, como la selección de acciones. Finalmente, la Zona Subesofágica (SEZ) se presenta como la red más dispersa, con la densidad, los grados y el coeficiente de clusterización más bajos. Su baja eficiencia interna (largo de camino de 4.38) refuerza la idea de que su importancia no radica en el procesamiento local, sino en su rol como un modulador global, cuya influencia se ejerce a través de conexiones de largo alcance entre módulos.

El análisis comparativo revela que el conectoma de la *Drosophila* es una estructura altamente organizada y no aleatoria, tal como se planteó en la hipótesis central. Específicamente, la red cerebral es fundamentalmente distinta a un grafo aleatorio de ER, presentando un coeficiente de clustering significativamente mayor y una longitud de camino promedio también mayor ($p = 0$), lo que es consistente con un sistema biológico que requiere un flujo de información jerarquizado a través de rutas

específicas. Además, los resultados confirman de manera robusta la propiedad de mundo pequeño. Con un coeficiente γ de 95.2 y un λ de 1.71, el coeficiente de mundo pequeño (σ) resultante de 55.68 demuestra la coexistencia de una alta estructura local con una alta eficiencia global. Finalmente, el conectoma exhibe claras características de red libre de escala, consistentes con el principio de conexión preferencial del modelo Barabási-Albert. Sin embargo, el modelo BA puro es una descripción incompleta, como lo demuestra la distintiva "protuberancia" en la distribución de la red real, lo que sugiere que restricciones biológicas más complejas juegan un papel crucial en la configuración de la arquitectura final⁶. Este mecanismo de crecimiento podría tener una base biológica en el aprendizaje Hebbiano.

El análisis de modelos nulos establece que el conectoma posee una arquitectura global sofisticada (mundo pequeño y libre de escala), la cual es necesaria para un procesamiento de información eficiente y robusto. Sobre esta base, el análisis de centralidad revela firmas estructurales específicas a nivel de subred que se alinean con las demandas computacionales del aprendizaje por diferencia temporal (TD). La distribución única de la centralidad de auto vector en el MB, con su vasta periferia de baja centralidad (Kenyon cells) y su núcleo de alta influencia (MBONs), apunta a su capacidad para sostener la actividad reverberante necesaria para una función de valor basada en memoria. Simultáneamente, la mayor centralidad de intermediación observada en el SEZ, confirmada visual y estadísticamente, proporciona la evidencia estructural para su rol como un modulador global del flujo de información, necesario para balancear la exploración y explotación. En conjunto, la estructura global de la red habilita la computación compleja, mientras que las propiedades locales de las subredes sugieren una especialización funcional consistente con el marco teórico propuesto.

Estos hallazgos abren varias líneas de investigación para análisis futuros. Una primera hipótesis, basada en la distribución de la centralidad de auto vector, es que las neuronas con los puntajes más altos en el MB corresponden predominantemente a las neuronas de Salida del MB (MBONs), formando el núcleo computacional recurrente, lo cual podría verificarse cruzando los datos de centralidad con anotaciones de tipo celular. Una segunda línea podría investigar la identidad de los nodos "puente" en el SEZ, postulando que las neuronas con la mayor centralidad de intermediación conectan de manera crítica los tres módulos y es probable que incluyan las neuronas que expresan "hugina", responsables de modular el estado global del cerebro. Finalmente, la desviación del modelo BA puro sugiere una tercera hipótesis: que la "protuberancia" en la distribución del grado de entrada corresponde a una clase específica de neuronas, posiblemente dentro del Complejo Central, que sigue reglas de cableado distintas y optimizadas para un tipo diferente de computación.