Tarea 3

Observaciones:

- Cada pregunta vale 40 puntos; ustedes escojan **2** preguntas para contestar, de modo que el puntaje sea de 0 a 80, y eso se lo sumamos a la tarea previa para hacer dos tareas con un total de 205 puntos.
- El archivo que se ocupa en la P3 lo dejé en la misma carpeta que tenía los archivos de la T2.
- **Incluyan** todo su código en la entrega.
- Dondequiera que vea un "comente" o "interprete", en esta tarea y las que sigan hágalo. Nos es de yapa; por lo general es lo más importante de una pregunta. Sobre todo en preguntas como la 1 o la 2, lo que cuenta es el análisis (y si se les ocurre enriquecerlo con otras cosas que medir, además de las que se sugieren, bienvenidas sean).

1. Juguemos con arena. Genere dos redes, una Erdös-Renyi y otra Barabási-Albert, ambas con 1000 nodos y grado promedio 2 (o puede ser 3, en realidad no es crucial). Sobre cada una de ellas haremos el siguiente experimento. Primero que nada, supondremos que cada nodo contiene un "balde" capaz de alojar una cantidad de granos de arena idéntica a su grado. Inicializamos cada balde a una cantidad de granos al azar (o sea, un número al azar entre 0 y el grado del nodo). Luego, durante 10000 iteraciones, hacemos lo siguiente:

   - Se agrega un grano de arena a un nodo escogido al azar.
   - Si la cantidad de granos rebalsa el balde, tenemos una avalancha. Se reparte un grano a cada uno de los vecinos del nodo. Si con eso alguno de ellos rebalsa, ese también reparte, y así sucesivamente. El tamaño de la avalancha se mide como la cantidad total de granos transferidos.
   - Considere además en cada transferencia de grano una probabilidad de $1/1000$ de que el grano se pierda (nunca llegue a su destino). Esto es para evitar que el sistema se sature.

   Grafique la distribución de los tamaños de las avalanchas, para cada tipo de red. Si parece ser una distribución conocida (¡debería!) estime sus parámetros[1].

   En alguna iteración pasada algunos alumnos tuvieron que agregar consideraciones especiales para casos límites (por ejemplo, para evitar avalanchas infinitas que les aparecían en componentes conexas pequeñas). Esos casos dependen bastante de la forma en que decidan implementar las avalanchas, así que no agrego reglas extra acá; ustedes vean si las necesitan, y de ser así déjenlas documentadas de algún modo.

2. Se han propuesto diversos modelos para abordar un problema crucial para la humanidad: ¿Sobreviviríamos a un apocalipsis zombie? La variedad de modelos es necesaria, porque no conocemos los detalles de las interacciones, y las películas muestran bastante diversidad.

   Aquí propondré un modelo simple, que separa la población en tres tipos de agentes: humanos, zombies, y destruidos. En cada interacción entre un humano y un zombie, habrá una probabilidad α de que el zombie muerda al humano y lo zombifique, una probabilidad β de que el humano logre curar al zombie (convirtiéndolo en humano), y una probabilidad $1 - \alpha - \beta$ de que el humano destruya al zombie (que pasara a destruido). Además, en cada iteración, todo zombie tiene una probabilidad γ de destruirse espontáneamente (por accidentes o trampas

[1] Considere que lo que importa es como *cae* la curva; puede que haya un tramo inicial que despreciar.

puestas por los humanos; no son muy astutos, estos zombies). Nos interesa determinar los rangos de parámetros que permiten la supervivencia de la humanidad.

Para simplificar sus vidas, uno de los parámetros lo fijaremos: $\beta = 0.1$, obtenido tras largos experimentos de curación de zombies[2]. De modo que lo que buscamos es, en el plano (α, γ), determinar las zonas en las que la humanidad se extingue y en las que no lo hace.

La forma de lograrlo es simular esta dinámica en una red de 5000 nodos, con grado promedio 4, en la que inicialmente hay un 5% de zombies. En cada iteración se escoge un humano al azar, y un vecino zombie al azar (si es que existe) e interactúan según se indicó antes, cambiando de estado alguno de los dos según el resultado. Además, como dijimos antes, cada zombie de la red tiene una probabilidad γ de pasar a destruido en cada iteración.

Determine la frontera entre la vida y la muerte de nuestra especie para tres posibles topologías de la red: Erdös-Renyi, Barabási-Albert, y SBM con 50 comunidades ("refugios") de 100 nodos[3].¿Depende su resultado de la topología?

Comente sus resultados.

Nota 1. La forma natural de visualizar aquí es mediante un heatmap, indexado por αy γ, y partir moviendo ambos parámetros en un rango amplio. Sin embargo, de ahí conviene hacer "zoom" hacia la zona en que ocurre el cambio de resultado; algún alumno del pasado, por ejemplo, sólo estudió el rango con valores de $\gamma < 0.0005$, pues con más de eso la muerte espontánea de zombies le hacía muy fácil la cosa a los humanos. Dependerá un poco, en todo caso, de la forma exacta en que implementen la dinámica.

Nota 2. Más que extinción total (que a veces ocurre) conviene mirar el porcentaje final de humanos; a veces sobreviven en alguna componente conexa pequeña, por ejemplo.

3. Baje el archivo "primaryschool.csv". Son datos del proyecto *sociopatterns*, en que se registró la proximidad física de los alumnos y profesores de una escuela primaria en EEUU cada 20 segundos, durante dos días. Cada fila contiene (separados por tabs) la etiqueta temporal (en segundos), los IDs de las personas (como 240 en total, con IDs entre el 1426 y el 1922), y la clase a la que pertenecen (dato que ignoraremos). Escriba el código necesario para digerir los datos y haga lo siguiente:

   (a) Asuma que el tiempo completo se corta en ventanas de 10 minutos (sin traslape). Para cada ventana genere la red agregada de las interacciones que allí ocurrieron, y a partir de esta calcule el grado promedio, la transitividad (coeficiente de clustering) y la modularidad (tras aplicar el algoritmo de comunidades que prefiera). Grafique la evolución de estos tres valores a lo largo del tiempo. Comente.
   Luego repita esto mismo pero para ventanas de 2 horas.

   (b) Considerando como origen al alumno con ID 1695, del primero B, considere su conjunto de influencia $I(T, d)$ consistente en todos los nodos que se pueden alcanzar a partir de él, respetando el orden temporal de las conexiones, entre los tiempos T y $T + d$. Grafique el tamaño de $I(T, d)$ en función de ambos parámetros (es decir, en función de ambos ejes a la vez, no por separado, puede ser un heatmap o un gráfico 3d). Use la granularidad que le parezca razonable para obtener una buena visualización. Comente.

[2] Un complejo proceso que incluye pociones y manzanas doradas.

[3] Stochastic Block Model; aquí ustedes elijan una probabilidad de conexión intra-comunidad y otra intercomunidades, de modo que el grado promedio sea el deseado. Asegúrense de que la intra sea al menos el triple que la inter.