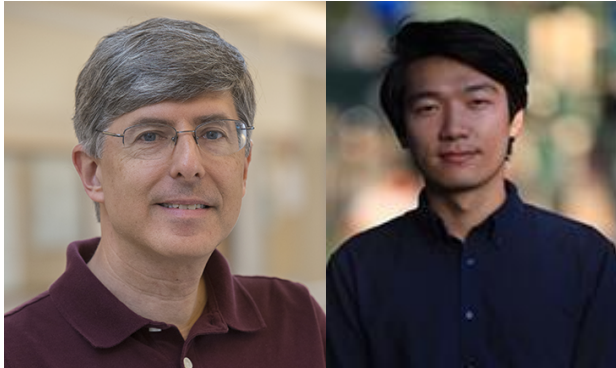


[Centro de Información de COVID \(CIC\): Charlas científicas de relámpago](#)



Transcripción de una presentación de Steven Skiena y Xingzhi Guo (CUNY Stonybrook), febrero de 2022

Título: Evolución de Incorporación de Knowledge Graph para COVID-19

Financiado por la Oficina de Infraestructura Cibernética Avanzada de la NSF, Dirección de Informática e Información e Ingeniería (OAC/CISE) a través del Programa del Fondo de Semillas del

Northeast Big Data Innovation Hub.

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de febrero 2022](#)

Editora de la Transcripción: Saanya Subasinghe

Editora de la Traducción: Isabella Graham Martínez

Transcripción

Xingzhi Guo:

Diapositiva 1

Así que gracias por tenernos aquí. Mi nombre es Xingzhi. Soy estudiante de cuarto año de Ph.D. trabajando con el Profesor Steven Skiena en la Universidad de Stony Brook y el Profesor Skiena es también el Director del Stony Brook AI Institute. Y esta presentación es acerca de cómo la incorporación de gráficos de conocimiento cambia durante COVID a medida que nuestro mundo cambió. Y el contenido principal se basa en un artículo reciente publicado en SIGKDD 2021, en colaboración con Boajian [Zhou] cuando era postdoc aquí.

Diapositiva 2

Bueno, vamos a seguir. Así que lo primero es lo primero. ¿Qué es la incrustación de gráficos u otros pueden llamarlo una incrustación de red o incrustaciones de nodos? Así que básicamente, es una función que asigna un nodo en el gráfico a un vector numérico que llamamos vector de incrustación. Como muestra el siguiente ejemplo, podríamos asignar cada nodo de este gráfico a este espacio 2D. Así

que la incrustación podría capturar el significado del nodo en el gráfico original, pero el hecho de que los nodos más cercanos en el gráfico pueden tener vectores similares en el espacio de incrustación. Así que en este ejemplo, puedes ver que los nodos con el mismo color también se unen entre sí en este plan 2D. Y lo más importante, por este vector de incrustación numérica de baja dimensión, podríamos aplicar algoritmos de aprendizaje automático existentes para muchas tareas posteriores. Por ejemplo, la clasificación de nodos, el clúster de nodos para el descubrimiento de comunidad o la detección de valores atípicos. Y, sin embargo, este ejemplo es un gráfico estático donde no hay nueva arista, no hay nuevo nodo, todo está fijo, pero en nuestra vida real nuestro mundo siempre está cambiando. También lo hacen los gráficos del mundo real y los nodos.

Diapositiva 3

Bueno, echemos un vistazo a la gráfica cambiante del mundo real - o podemos llamarla gráfica dinámica. Así que ahora en el gráfico de conocimiento de enlaces de Wikipedia, cada nodo es artículos de Wiki que generalmente describen las entidades del mundo real y cada borde es el hipervínculo que conecta dos artículos, algo así como la cita que hacemos cuando estamos escribiendo un artículo. Así que este gráfico es - es a gran escala. Tenemos millones de nodos y cientos de millones de bordes y sigue escalando como se puede ver en esta pequeña figura. Y ciertamente algunas entidades pueden cambiar muchas y han sido capturadas por este gráfico de conocimiento dinámico. Y quiero mostrar un ejemplo específico de la ciudad de Wuhan.

Diapositiva 4

Vale, así que antes de COVID, Wuhan es probablemente menos famoso para la gente de todo el mundo. Y a finales de 2019 y principios de 2020, creo que la mayoría de la gente lo conocía como el primer lugar del brote de COVID. Y creo que - así que creo que este es un buen ejemplo del cambio de la o la entidad cambiada. Así que aquí la figura muestra un artículo de Wikipedia de Wuhan. Y destaco los hipervínculos que tiene. Y los dos primeros párrafos son los geo[lógicos] o algunos eventos históricos relacionados con Wuhan. Pero en 2019, de repente, vemos que se crearon muchos nuevos enlaces relacionados con COVID. Así que volviendo a la perspectiva de incrustación de gráficos, la pregunta es ¿cómo rastrear eficientemente esas incrustaciones de nodos en este gráfico masivo dinámico para que podamos detectar el movimiento de incrustación del nodo y compararlo a través del tiempo para que podamos ver cómo esto cambia?

Diapositiva 5

Bueno, a continuación, esta pregunta nos motiva a diseñar un nuevo algoritmo que puede manejar un problema que llamamos un subconjunto de nodos incrustados en gráficos dinámicos grandes. Así que usando este algoritmo podemos rastrear las incrustaciones de varios nodos predefinidos. Así que es un subconjunto de nodos en lugar de los cuatro nodos en el gráfico que nos interesan, ya que el gráfico sigue evolucionando. Así que la idea clave es utilizar el rango de página personalizado que es algoritmo muy exitoso utilizado por la búsqueda de Google y una ventaja sobre otros métodos es que podríamos calcular solo lo que necesitamos para los nodos del subconjunto de - para los nodos del subconjunto, mientras que la mayoría de los otros algoritmos tienen que calcular todas las incrustaciones de cada

nodo a través de cada vez. Pero finalmente utilizan solo una parte de ellos, por lo que el resto solo se desperdician. Así que nuestro algoritmo es más eficiente y rápido, es muy adecuado para este problema, pero para más detalles, consulte nuestro documento, que están aquí [<https://arxiv.org/abs/2106.01570>]. Y como este ejemplo ilustra el concepto, nuestro método puede calcular las incrustaciones de un nodo específico, en este caso Wuhan, a través de diferentes años. Y podemos esperar un gran movimiento de integración de 2019 a 2020.

Diapositiva 6

Bueno, así que a continuación - así que vamos a ver el resultado del experimento. Primero, recopilamos la instantánea del gráfico de Wikipedia en inglés todos los días en 2020. Y como he mencionado antes es un gráfico dinámico muy grande. Se puede ver en esta tabla que aproximadamente tenemos 30.000 nuevos bordes insertados cada día y liberamos los datos en este repositorio de GitHub por lo que es muy fácil de acceder. Y seguimos la pista de Wuhan junto con otros cientos de ciudades chinas. Y esta figura muestra cómo los cambios que detectamos en el espacio de incrustación para que como pueden ver, la curva de Wuhan es prominente. Tienen un pico enorme y destacamos varios picos con anotación y encontramos que todos están correlacionados con, correlacionados con la línea de tiempo de COVID. Y hay otro pico - se puede ver que detectamos otro pico en la ciudad de Chengdu que refleja la tensión diplomática entre Estados Unidos y China cuando EE.UU. decidió cerrar el consulado en Chengdu. Así que tiene la casualidad que descubrimos en este estudio.

Diapositiva 7

Así que, oops, así que, uh - tenemos otro experimento donde queremos ver las ciudades más cambiadas en diferentes períodos de tiempo. Así que clasificamos los movimientos de incrustación del nodo rastreado en cada momento y encontramos que, como pueden ver, Wuhan generalmente es el más cambiado entre otros a medida que COVID evoluciona. Y destacamos el título de las noticias en ese período de tiempo para que podamos tener una idea de lo que sucedió en ese lugar.

Diapositiva 8

De acuerdo. Así que las principales conclusiones son: tenemos un algoritmo de incrustación de nodos muy eficiente que puede capturar las incrustaciones en gráficos dinámicos muy grandes. Luego investigamos la evolución del gráfico de conocimiento de Wikipedia en 2020 y encontramos los interesantes cambios de nodos durante la COVID. Para más detalles, por favor refiérase a nuestro documento y vea los recursos liberados. [<https://arxiv.org/abs/2106.01570> y <https://github.com/zjlxgz/DynamicPPE>] Y sí, gracias, gracias!