

ML Y GRAFOS

GRAPHS ARE EVERYWHERE

GRAFOS PARA LOGRAR EL MÁXIMO RENDIMIENTO EN ANALYTICS, ML & IA

Los grafos pueden estar en cualquier lugar al que demos un vistazo. Los usamos a diario y pueden llegar a pasar inadvertidos. Son herramientas excepcionales para comprender la realidad y darle un sentido a elementos que se encuentran vinculados en un sistema. Los grafos en esencia son una expresión de datos abstractos que combinan dos elementos básicos para constituirse. Los nodos y vértices son las piezas que conforman a los grafos.

Un grafo es una composición de un conjunto de objetos conocidos como nodos que se relacionan con otros nodos a través de un conjunto de conexiones conocidas como aristas. Esta poderosa combinación de elementos presenta oportunidades analíticas increíbles que han cambiado radicalmente el mundo. No exageramos. En realidad han cambiado la forma en la que entendemos el mundo.

Estos elementos nos permiten comprender a profundidad las relaciones de los datos. Relaciones que no son simples de ver, pero que existen.

Gracias a esto diversos elementos que utilizamos todos los días han mejorado e influido directamente en nuestra vida. Es increíble el avance que han tenido las redes sociales y su impacto gracias a las potencialidades de los grafos.

Inclusive podemos encontrar la incidencia de los grafos en el desarrollo de los sistemas

de atención al cliente de las empresas y la evolución de los sistemas de recomendación de productos.

Adicionalmente los grafos son implementados cada vez más en sistemas de prevención de fraude bancario y de prevención del blanqueo de capitales. Estos permiten establecer patrones de comportamiento sospechoso y buscar a nivel de detalle la existencia de relaciones que demuestren operaciones fraudulentas.

Su alcance ha llegado a aportar elementos muy interesantes en áreas científicas innovadoras como la Inteligencia Artificial y el Machine Learning. Los grafos ayudan a estos modelos avanzados de aprendizaje automatizado a comprender de forma más rápida las relaciones existentes entre datos para que el conocimiento sea cada vez más y de mayor calidad.

En este e-book haremos un recorrido por la potencia de los grafos y sus propiedades analíticas, descubriremos los aportes de los grafos a los proyectos de machine learning, profundizaremos sobre las potencialidades de Neo4j para estos proyectos y sus capacidades en proyectos más avanzados como son los desarrollos de Inteligencia Artificial. Acompáñanos a descubrir dentro de este universo profundo y brillante de posibilidades como los grafos pueden ser de utilidad para lograr el **máximo rendimiento en Analytics, ML & IA.**

VENTAJAS DE USAR NEO4J COMO BASE DE DATOS DE GRAFOS

Abordemos un poco los aspectos técnicos de este software que concentra muchas características positivas que podemos destinar al combate contra el blanqueo de capitales.

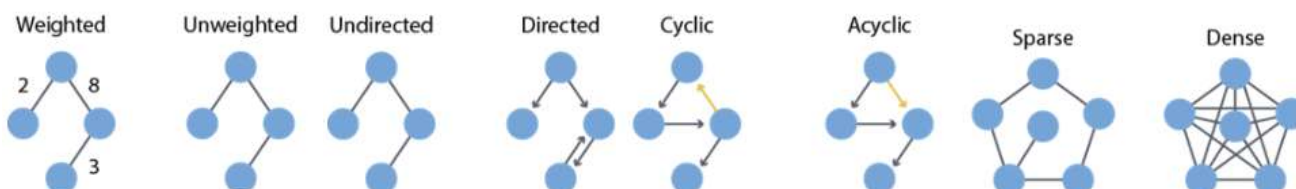
LAS RELACIONES SON EL MEJOR PREDICTOR DEL COMPORTAMIENTO

Las principales empresas del mundo ya lo sabían de hace años, y ese ha sido su valor diferencial para poder crecer, pero ahora este conocimiento ya está presente en el mercado y hay múltiples estudios que así lo confirman.

Autores como James Fowler en su libro *Connected*, dicen cosas como “Estamos aprendiendo que puedes hacer mejores predicciones sobre las personas al obtener toda la información de sus amigos y amigos de sus amigos que con la información que tienes sobre la persona en sí”. Esto es, utilizar el Contexto es un valor diferencial gigante, no quedarnos en el dato discreto, sino ir más allá y conseguir valor del dato conectado, de la relación entre los datos, ese debe ser el objetivo.



Los grafos poseen propiedades increíbles. Primero debemos recorrer algunos aspectos importantes sobre ellos para comprender su alcance y gran versatilidad. Estas poderosas representaciones de datos en nodos y vértices representan múltiples formas de estructuración debido a que existen diversos tipos de grafos.



Cada uno de estos tipos de grafos responde a necesidades de representación de datos diferente. Una de las ventajas más importantes que nos ofrecen los grafos sin duda alguna es su capacidad de representación visual. Los seres humanos entendemos de una forma más profunda el comportamiento de los datos si somos capaces de apreciarlos. Nuestro proceso de comprensión de elementos complejos gráficos es superior y es probable que podamos encontrar más y mejores soluciones a los retos que puede demandar la complejidad de los datos del mundo de hoy.

Desde siglos atrás hemos podido aprovechar las ventajas de observar los datos. Un ejemplo claro de esto es uno de los casos que sentó las bases de la teoría de grafos.

Recordemos que Leonhard Euler logró resolver el mítico problema de los puentes de Königsberg ordenando sus pensamientos y diferentes soluciones de forma gráfica para llegar a una conclusión.

En la actualidad para lograr conclusiones eficientes a problemas complejos podemos apoyarnos en la analítica de grafos. Esta es una rama en constante y veloz crecimiento dentro del análisis de datos. Con la analítica de grafos podemos visualizar grandes volúmenes de información de forma clara para lograr revelar hasta las más pequeñas relaciones ocultas entre los datos.

LOS GRAFOS HAN CAMBIADO LA FORMA EN LA QUE VEMOS EL MUNDO





Cada vez es más común conocer de proyectos de investigación de mercado, análisis de datos de redes sociales para diseño de productos o campañas de mercadeo, análisis crediticios, detección de fraude bancario y prevención de legitimación de capitales que implementan el análisis de grafos. Esto se debe a que las estructuras de grafos no solo benefician a los analistas, también entregan numerosas ventajas a los sistemas.

Existen numerosos sistemas computarizados dotados de algoritmos de análisis muy potentes que pueden buscar, detectar ciclos, buscar rutas cortas y recorren los grafos más complejos que logremos imaginar en cuestión de segundos. Estos softwares superan ampliamente las capacidades o los límites de análisis del ojo humano.

Con el apoyo de estos algoritmos podemos detectar, explorar y extraer información valiosa de conjuntos gigantescos de datos.

Sin embargo, lo más interesante es que gracias a las propiedades especiales de estos algoritmos que se incorporan a nuestros proyectos de analítica basada en grafos podemos detectar las relaciones más discretas entre los datos. Relaciones ocultas, imperceptibles a simple vista, pero que no pueden ocultarse por completo.

Ahora podríamos preguntarnos, ¿pueden las máquinas utilizar este tipo de análisis a los algoritmos diseñados para generar conocimiento? La respuesta es bastante clara. Como hemos dicho antes, las computadoras, pueden aprovechar las ventajas que plantean los proyectos analíticos basados en grafos. Si las dotamos de algoritmos inteligentes, estas, pueden fácilmente recorrer, estudiar y analizar grafos de dimensiones descomunales para extraer conocimiento de gran utilidad.



¿QUÉ APORTAN LOS GRAFOS A PROYECTOS DE MACHINE LEARNING?

El potencial evidente de los grafos es aprovechable en diversos tipos de proyectos. En el área de aprendizaje automatizado o machine learning, no es la excepción. Es importante destacar que el Machine Learning no está sujeto a una estructura particular de representación de datos y ciertamente no es tarea sencilla. Aplicar técnicas de aprendizaje automatizado a grafos puede resultar una tarea un poco compleja, pero no imposible.

RETOS Y CONSIDERACIONES

Analizar grafos inmensamente grandes y alimentar a los algoritmos con sus datos es un reto importante. Las máquinas comparan con los seres humanos la dificultad de comprender visualizaciones gigantescas. Sus propiedades analíticas son ampliamente superiores.

Para aprovechar al máximo las propiedades de los grafos en modelos de aprendizaje automatizado contamos con los conocidos marcos de incrustación. Estos realizan un levantamiento o mapeo entre los nodos que conforman un grafo y nos permiten entender dicha área como un vector.

Existen diversos tipos de marcos de incrustación diseñados para codificar la información en vectores de números reales de baja longitud y longitud fija.

Es importante señalar que esta es la forma en la que funciona la gran mayoría de algoritmos de ML.

Estos marcos de incrustación permiten establecer áreas para hacer recorridos y analizar a profundidad los datos. Uno de los marcos más populares e implementado por sus características es el "node2vec". Este método aprende características de forma de permitir la exploración de vecindarios de nodos a través de recorridos aleatorios a través de determinadas secciones del grafo.





BENEFICIOS

Los modelos de aprendizaje automatizado se ven enormemente favorecidos por los grafos ya que aportan propiedades importantes para que el conocimiento generado sea de utilidad. Hagamos un recorrido rápido por un ejemplo común en estos días como es un buen tiempo de películas en Netflix.

Las bases de datos de esta plataforma recaban datos para alimentar los sistemas de recomendación de contenido. Entre los datos más importantes pueden destacarse la edad del usuario, su género favorito de visualización, horarios habituales, tiempo de reproducción e inclusive podemos tomar en cuenta las últimas películas que vio para establecer un patrón de comportamiento. Pero este patrón a simple vista posee una carencia importante que puede conducir a errores de recomendación importantes que pueden perjudicar la eficiencia de la recomendación y por consecuencia, disminuiría la satisfacción del usuario con la plataforma. Este elemento faltante es el contexto.

El contexto es lo que nos permite entender mejor las necesidades del usuario. Conocer esto es posible gracias a los grafos.

Los grafos son elementos totalmente dinámicos que podemos modelar cada vez que intuimos que pueden existir puntos de mejora para ofrecer mayor calidad al usuario.

La contextualización de las solicitudes del usuario es clave para ofrecer una experiencia excepcional. Debemos tener la posibilidad de identificar y clasificar en comunidades para que los impactos de nuestras recomendaciones sean altos. Con una base de datos relacional es imposible lograr estos resultados. Las bases de datos relacionales han aportado y aportan un valor fundamental, pero deben ser complementadas para que los algoritmos puedan tener en cuenta los elementos adicionales.

Ahora conozcamos un poco más sobre Machine Learning.

MACHINE LEARNING

El aprendizaje automático es una rama científica de la inteligencia artificial que, implementando algunos algoritmos, permite a estos sistemas aprender y generar conocimiento de forma progresiva a partir del análisis de datos. En la actualidad los modelos de aprendizaje automatizado se han convertido en herramientas indispensables para las empresas y para las personas.

Día a día disfrutamos de sus avances en nuestra vida. La gran mayoría de nosotros puede interactuar con filtros anti spam en las bandejas de correos electrónicos utilizamos sistemas de recomendación de contenido, software de reconocimiento

facial e inclusive el mercado está siendo abordado por coches con conducción automática, algo que hace pocos años atrás parecía sacado de una película de Hollywood.

El Machine Learning nace con el propósito de que las máquinas puedan generar conocimiento de forma similar a como aprendemos los seres humanos. En la actualidad estos modelos han avanzado de forma significativa permitiendo construir modelos predictivos de sucesos y de tomas de decisiones gerenciales con un nivel de eficiencia muy alto.

TIPOS DE ML

Existen diferentes tipos de aprendizaje automatizado que podemos aprovechar para nuestros proyectos. El aprendizaje supervisado que genera conocimiento partiendo de datos etiquetados es generalmente implementado en filtros detectores de spam, imágenes en captchas o aplicaciones de reconocimiento de voz o escritura. Particularmente este tipo de modelo aprende de estos resultados e incorpora ajustes en los parámetros interiores para poder adaptarse a datos nuevos que ingresan al sistema.

El aprendizaje de tipo no supervisado es el en que se incluyen dentro del modelo de análisis datos sin etiquetar y de los que no se tiene conocimiento previo de su estructura. Con este aprendizaje se busca conocer información importante de los datos sin tener referencia de posibles variables de salida. Este tipo de ML es muy útil para complementar estrategias de segmentación de mercadeo ya que permite establecer grupos de características similares.



También es necesario mencionar el Aprendizaje profundo o Deep Learning, que tiene como finalidad construir modelos que incrementen el rendimiento del conocimiento, a partir del resultado de cada interacción realizada. Es decir, que el modelo, al generar un resultado, lo incorpore automáticamente como un nuevo parámetro de ajuste para resultados nuevos.

MACHINE LEARNING

Los diferentes tipos de machine learning como hemos mencionado funcionan y pueden entregarnos maravillosos resultados gracias a **complejos algoritmos**. A continuación descubrirás un poco más sobre los principales tipos de algoritmos para aprendizaje automatizado.

ALGORITMOS DE REGRESIÓN

Dentro de los algoritmos principales de machine learning contamos con los algoritmos de regresión. Estos buscan estimar y determinar dentro del conjunto de datos o área de estudio las relaciones existentes entre diferentes variables. El análisis de regresión de variables se concentra en fijar como

dependiente una variable y estudiar su comportamiento cuando interactúa con otra serie de variables independientes. Con este tipo de análisis los modelos de aprendizaje automatizado pueden crear procesos predictivos eficientes.

ALGORITMOS BAYESIANO

Como su nombre bien lo indica, estos algoritmos toman como punto referencial para su funcionamiento el teorema de Bayes. Apoyándose en esta teoría realizan clasificaciones de los datos asumiendo cada uno de estos valores como variables independientes de cualquier otro.

Su aplicación dentro de proyectos de machine learning permite construir modelos probabilísticos efectivos para predecir clases o categorías dentro los conjuntos de datos.



ALGORITMOS DE AGRUPACIÓN

Los algoritmos de agrupación son un tipo importante de algoritmos que son de gran utilidad para proyectos de ML. Estos permiten establecer categorías dentro de datos no etiquetados para poder ordenar datos aunque estos no pertenezcan a grupos definidos.

Estos algoritmos realizan búsquedas dentro del conjunto de datos, estableciendo una variable de representación dentro del grupo y posteriormente de forma iterativa asigna en cada punto de datos esa variable, según las características que determinemos.

ALGORITMOS DE ÁRBOL DE DECISIÓN

Los algoritmos de machine learning que utilizan las técnicas de árboles de decisión realizan un esquema en el que podemos representar los posibles resultados para poder inclinarnos hacia una opción. Estos algoritmos son muy importantes ya que permiten trazar estrategias comerciales y de evolución de diferentes tipos de proyecto.

ALGORITMOS DE REDES NEURONALES

Las redes neuronales son uno de los tipos de proyectos de ML más interesantes que podemos implementar en la actualidad. Estos son programaciones complejas que buscan funcionar como capas de procesamiento de datos y se conectan a su vez con otras capas tratando de simular el procesamiento de información del cerebro humano.

Los algoritmos para este tipo de ML son implementados con la finalidad de establecer relaciones no lineales entre los datos o para detectar la relación existente de los datos que ingresan a un sistema con datos pre-existentes.





ALGORITMOS DE REDUCCIÓN DE DIMENSIÓN

En segmentos anteriores mencionamos algunos detalles sobre dificultades comunes que pueden tener los proyectos de ML para analizar volúmenes gigantescos de datos. Los algoritmos de reducción de dimensión cumplen de forma específica con la reducción del número de variables que deben considerar para lograr una solución concreta. Gracias a estos algoritmos se pueden optimizar los tiempos de procesamiento de datos.

ALGORITMOS DE APRENDIZAJE PROFUNDO

Para los proyectos de Deep Learning o aprendizaje profundo se deben utilizar algoritmos especiales que pueden ejecutar datos a través de múltiples capas de redes neuronales. Estos algoritmos aprenden de forma progresiva sobre el objeto o la imagen que estudian y elevan progresivamente su nivel de complejidad.

ALGORITMOS DE GRAFOS APLICADOS A ML

Los algoritmos de Machine Learning se nutren de características que provienen de un análisis ejecutado previamente en una base de datos. A medida que los algoritmos de ML tengan mayor capacidad de asimilar e incorporar características o variables, también llamadas features en inglés, su rendimiento será superior.

Hasta el momento la gran mayoría de las

variables que los algoritmos de ML analizan provienen de los resultados que arrojan bases de datos relacionales, esto significa que las características sobre las que desarrollan su funcionamiento son discretas. Estas variables tienen una gran relevancia, más sin embargo necesitan ser complementadas por variables que permitan comprender el contexto que las acompaña.

Un ejemplo que podría ilustrar mejor la necesidad de utilizar variables complementarias que agreguen contexto a los datos puede ser el cálculo de un promedio de edad. Estas cifras por sí solas aportan algunos elementos analíticos, pero al incorporar características provenientes de un análisis de grafos sobre cuántos objetos están relacionados con este promedio nos ayudará a comprender a profundidad la utilidad y funcionalidad de los datos.

Las características del grafo se pueden recabar ejecutando algoritmos de diferentes tipos para mejorar las predicciones.

Los modelos actuales de ML ignoran la estructura de red y las relaciones complejas. La vinculación de los grafos dentro de estos modelos de aprendizaje supervisado agregan características predictivas a dichos modelos para ejecutar predicciones basadas en relaciones.

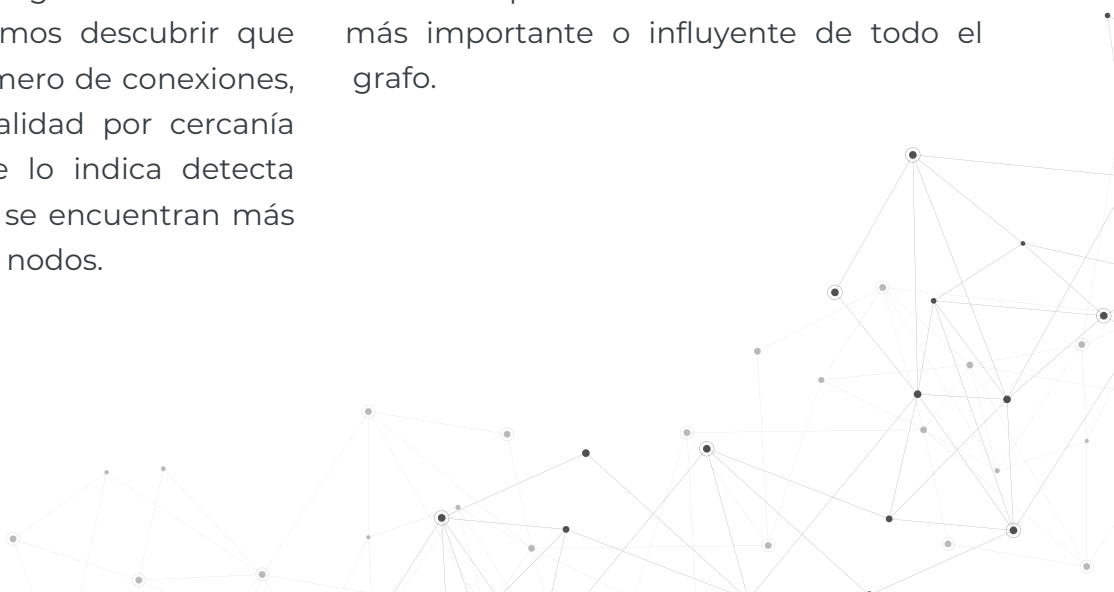
A continuación conocerás los cinco tipos principales de algoritmos de grafos aplicables a estos proyectos de aprendizaje automatizado.

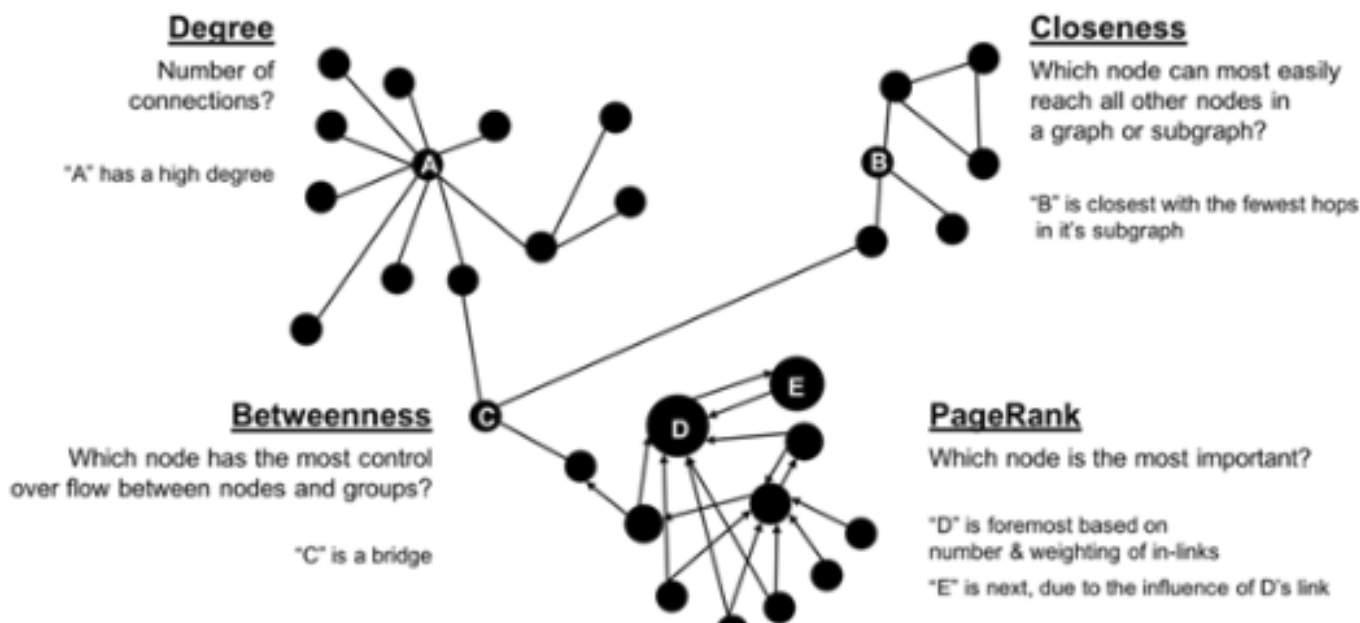
ALGORITMOS DE CENTRALIDAD

Los algoritmos de medida de centralidad son parte fundamental de los proyectos de ML acelerados con grafos. Estos algoritmos buscan calcular o estimar una medida o valor que poseen los nodos que componen un grafo. Con este cálculo se puede determinar la relevancia que este nodo representa.

Dentro de este tipo de algoritmos podemos destacar el algoritmo de grado de centralidad con la que podemos descubrir que grafo posee mayor número de conexiones, el algoritmo de centralidad por cercanía que como su nombre lo indica detecta cuáles pares de nodos se encuentran más cerca de otros pares de nodos.

Adicionalmente es importante mencionar el algoritmo de centralidad de intermediación que se encarga de determinar qué nodo sirve como conector entre otros pares de nodos. También se encuentra en esta categoría el reconocido algoritmo Page Rank. Este algoritmo es el que rige el posicionamiento de los contenidos dentro del motor de búsqueda más utilizado de internet. Es útil para determinar cuál nodo es el más importante o influyente de todo el grafo.





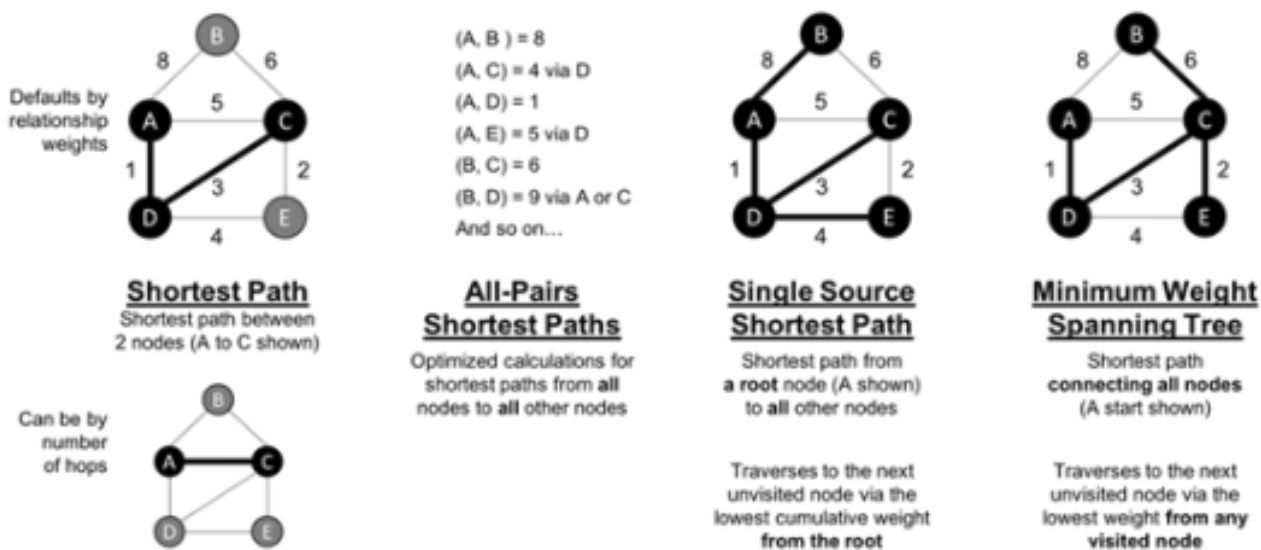
ALGORITMOS DE RUTAS ÓPTIMAS "SHORTEST PATH" Y BÚSQUEDA

Los algoritmos de detección de rutas óptimas son interesantes porque se encargan de detectar las rutas más eficientes para llegar de un nodo a otro dentro de un grafo. Dentro de este tipo de algoritmos se encuentra el algoritmo de rutas más cortas, que como su nombre lo indica, determina dentro de la estructura de un grafo la ruta más corta para recorrer el grafo de un nodo a otro.

También se encuentra dentro de esta categoría el algoritmo de rutas cortas en pares que se encarga de indicarnos si existen pares de nodos que no se relacionan de forma directa. El algoritmo de árbol de expansión de peso mínimo forma parte de los algoritmos de rutas cortas y búsqueda ya que partiendo de un nodo, se encarga de encontrar todos los nodos relacionados con el menor peso posible dentro del grafo.

El algoritmo de rutas aleatorias o random walk es uno de los más importantes de esta categoría porque se encarga de recorrer el grafo de forma aleatoria partiendo desde un nodo para conseguir durante su recorrido nuevas relaciones entre nodos.





DETECCIÓN DE COMUNIDADES

Los algoritmos de detección de comunidades cumplen con una función de gran importancia para el análisis de datos en grafos que contienen información compleja, e stos tomando en cuenta sus características, formas de relación y diferentes tipos de conexiones nos ayudan a establecer o determinar la existencia de comunidades complejas de datos.

Dentro de esta categoría encontramos el algoritmo de propagación de etiquetas, que permite encontrar comunidades dentro de un grafo de forma muy veloz.

También podemos destacar al algoritmo de modularidad que se encarga de detectar clusters de nodos dentro del grafo.

Existe adicionalmente un método de construcción de clusters dentro de los grafos conocido como Algoritmo de Louvain que detecta dichas estructuras recorriendo numerosas veces el grafo hasta que no se detectan nuevas conexiones posibles.



SIMILITUD

El algoritmo de similitud de nodos compara un conjunto de nodos en función de los nodos a los que están conectados. Dos nodos se consideran similares si comparten muchos de los mismos vecinos. Existen diversos algoritmos que se enfocan a conseguir determinar la similitud entre nodos, los más conocidos son los siguientes:

- Jaccard Similarity
- Conise Similarity
- Pearson Similarity
- Euclidean Distance
- Overlap Similarity
- Approximate Nearest Neighbors

Cada uno de estos algoritmos aporta un enfoque distinto y tiene en cuenta diferentes características a la hora de arrojar la semblanza entre los nodos, por lo que en muchas ocasiones se deben llevar a cabo ejecuciones de más de uno de ellos para determinar cuál es el que mejor aplica en el universo de datos en el que se está trabajando.

LINK PREDICTION

El último de los grupos de algoritmos nativo de grafos tiene que ver con la capacidad de predecir la aparición de nuevas relaciones entre nodos, es decir, interpretan la estructura del grafo y deciden la cercanía entre un par de nodos para dar como resultado un scoring de la probabilidad de que estos nodos se conecten.

Algunos de los algoritmos que se usan para este propósito son los siguientes:

- Adamic Adar
- Preferential Attachment
- Resource Allocation
- Common Neighbors



NEO4J UNA HERRAMIENTA EXCEPCIONAL

Los proyectos de ML requieren del estudio de complejos de cúmulos inmensos de datos. Neo4j es una poderosa herramienta que es el complemento que necesitamos. Este es un gestor de bases de datos NoSQL orientado a grafos que contiene un conjunto de prestaciones y características potentes. Con este poderoso gestor de BBDD relacionales podemos establecer, detectar y estudiar las relaciones entre los datos.

Neo4j en los últimos años ha evolucionado e incorporado potentes herramientas analíticas. En sus inicios Neo4j estaba más orientado a las soluciones transaccionales, pero en la actualidad ya incorpora algoritmos nativos de grafos para dar salida a las necesidades de ML que se requieren en todas las organizaciones que se orientan a los datos. El desarrollo constante del software y la complejidad de las consultas que se requerían ejecutar dentro de él, logró un proceso evolutivo dentro del desarrollo de Neo4j hasta el punto de incorporar una capa de algoritmos nativos de grafos.

Neo4j en la actualidad es una de las mejores opciones para enfrentar proyectos de analítica y puede ser de gran ayuda en proyectos de ML.

Es importante destacar que a pesar de ser una herramienta versátil y muy completa no puede asumirse que podemos resolver absolutamente todo con ella. Hay datos que deben ser analizados y consultados con bases de datos SQL.

Neo4j posee una biblioteca de algoritmos muy completa y que ha evolucionado de forma acelerada en los últimos años. En especial ha crecido la sección de algoritmos propio de una base de datos nativa de grafos en la que se concentran más de 35 algoritmos diferentes clasificados en cinco categorías importantes como son los algoritmos pathfinding, algoritmos de centralidad, algoritmos de detección de comunidades, de similitud y de predicción de relaciones.

Pathfinding & Search

- Parallel Breadth First Search & DFS
- Shortest Path
- Single-Source Shortest Path
- All Pairs Shortest Path
- Minimum Spanning Tree
- A* Shortest Path
- Yen's K Shortest Path
- K-Spanning Tree (MST)
- Random Walk

Centrality / Importance

- Degree Centrality
- Closeness Centrality
- CC Variations: Harmonic, Dargalchev, Wasserman & Faust
- Betweenness Centrality
- Approximate Betweenness Centrality
- PageRank
- Personalized PageRank
- ArticleRank
- Eigenvector Centrality

Community Detection

- Triangle Count
- Clustering Coefficients
- Connected Components (Union Find)
- Strongly Connected Components
- Label Propagation
- Louvain Modularity – 1 Step & Multi-Step
- Balanced Triad (identification)

Similarity

- Euclidean Distance
- Cosine Similarity
- Jaccard Similarity
- Overlap Similarity
- Pearson Similarity

neo4j.com/docs/graph-algorithms/current/

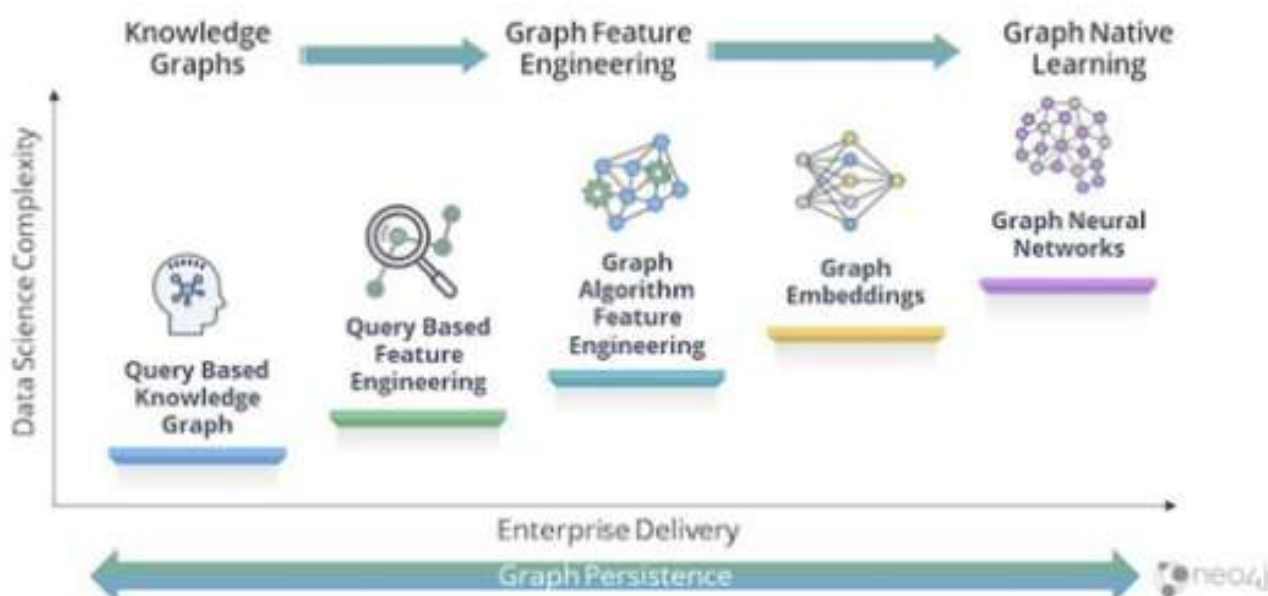
Link Prediction

- Adamic Adar
- Common Neighbors
- Preferential Attachment
- Resource Allocations
- Same Community

¿POR DÓNDE EMPIEZO MI PROYECTO DE ML CON GRAFOS?

Son increíbles las oportunidades de evolución y crecimiento que representan para las empresas y proyectos gubernamentales la aplicación de proyectos de Machine Learning aplicando el conocimiento que se desprende de los grafos, pero a veces se hace complicado saber por dónde empezar y conseguir demostrar este valor en un mundo de presupuestos limitados, tiempo ajustado y proyectos analíticos ya en marcha. Aquí intentaremos dar un poco de luz sobre este tema.

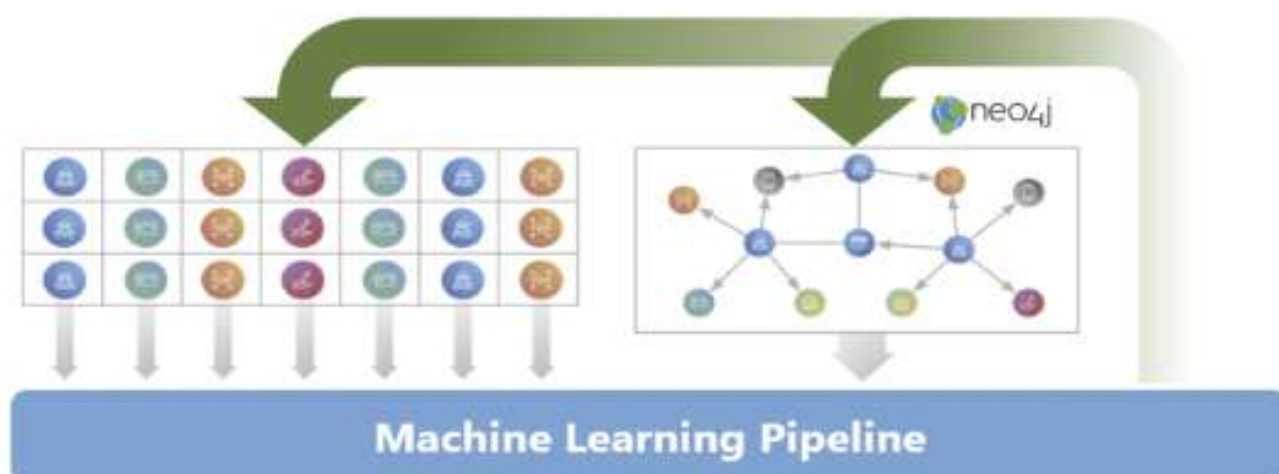
La imagen siguiente muestra lo que llamamos **“Graph Analytics Maturity Model”**.



En él podemos observar cómo existe un path bastante claro en lo que a la adopción de los grafos se refiere dentro de las organizaciones. Asumimos que ya existe cierta cultura de analítica avanzada, ML, ... dentro de la organización. En este caso, el primer nivel de evolución del uso de los grafos es realizarle consultas al mismo. Una vez tenemos los datos conectados dentro de un sistema gestor nativo de grafo como Neo4j

obtener conocimiento del grafo es sencillo. Usando consultas en Cypher (el SQL de los grafos) podemos interrogar a la base de datos y obtener cualquier información.

Pero a veces esto no es suficiente, y lo que queremos es alimentar nuestro Pipeline de Machine Learning con Features que provienen precisamente del grafo.



Estás features o variables, incorporan conocimiento de la estructura del grafo, como por ejemplo el número de vecinos. Esta información sería muy compleja de obtener de otro modo.

El siguiente nivel de evolución es el **“Graph Algorithm Feature Engineering”**, y aquí la alimentación del modelo de ML viene dada no sólo por consultas más o menos sencillas sobre el grafo, sino con datos provenientes de la ejecución de complejos algoritmos, como PageRank, Label Propagation, etc, dando como resultado Features relevantes para el cálculo del modelo en producción.

Los siguientes estadios son los más evolucionados, y requieren de mucha experiencia dentro de la organización. Graph **Embeddings**, de los que ya hemos hablado en este libro, así como las Redes Neuronales de grafos, ayudan a sacar el máximo partido del hecho de tener un grafo como parte de la arquitectura de datos de la organización. El grafo es ese repositorio de relaciones que permitirá a la organización dar un salto cualitativo en la capacidad de analizar nuestros datos como nunca lo habíamos hecho. Recordemos que:

LAS RELACIONES SON EL MEJOR PREDICTOR DEL COMPORTAMIENTO

En Graph Everywhere, como apasionados de los datos y los retos de desarrollo tecnológico, nos hemos sumado a las iniciativas de proyectos de ML que aspiran a crecer progresivamente a incorporar modelos de grafos. Esperamos que te animes tú también a vivir esta experiencia con nosotros, porque ya sabes:

GRAPHS
ARE
EVERYWHERE

