



5

FUNDAMENTOS SOBRE CIÊNCIA DE DADOS GRAFOS

QUE TODO MUNDO
DEVERIA SABER

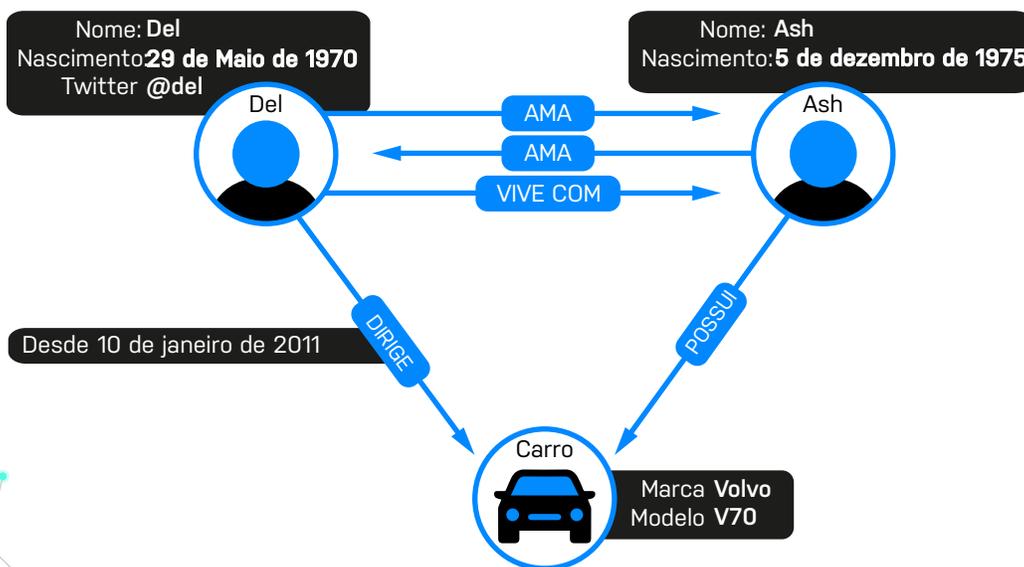
Por que Ciência de Dados grafos e por que agora?

As aplicações comerciais da ciência de dados grafos são novas e a maioria dos especialistas em dados ainda está se atualizando sobre a melhor forma de usá-las em suas organizações. Enquanto alguns especialistas em dados já estudaram a teoria dos grafos, outros ainda não tiveram muita exposição a ela. A ciência de dados grafos reúne análise de grafos, estatísticas e técnicas de IA e ML para melhorar seus modelos preditivos e prescritivos. Este artigo orienta você nos fundamentos da ciência de dados grafos para que você se sinta confiante e saiba quando usá-la em seu trabalho diário.

1. O que é um Grafo?

Antes de entender a ciência de dados grafos, você precisa entender os grafos. Em sua forma mais fundamental, um grafo é simplesmente uma maneira diferente de estruturar dados.

Em vez de linhas e colunas como em uma tabela de banco de dados relacional tradicional ou dataframe, os grafos usam nós (substantivos) e relacionamentos (verbos) como sua estrutura primária. Em um grafo, nós são substantivos – pessoas, lugares, coisas, organizações. As relações entre eles são verbos: tem relação de amizade com, trabalha para, gosta e assim por diante.



Os grafos representam dados através de relacionamentos.



Os nós representam uma entidade em um grafo.



Relacionamentos (bordas, links) conectam nós entre si.



As propriedades descrevem um nó ou relacionamento, nome, idade, altura..

2. Como a ciência de dados grafos é diferente das análises tradicionais?

Ao analisar dados de forma tabular, como em um banco de dados relacional, você tenta entender os pontos de dados sem ter uma maneira coerente de analisar suas conexões. O cientista de dados Ben Squire comparou os métodos tradicionais com "tentar resolver um cubo mágico apenas olhando um lado". O grafo oferece aos especialistas em dados a capacidade de observar, entender e analisar as conexões entre cada ponto de dados. Isso fornece contexto aos dados, o que é impossível obter em um modelo de dados tabular. Você pode ver quão fortes são as conexões, onde os grupos de conexões se formam, a importância de cada conexão e como as conexões influenciam umas às outras.

A ciência de dados grafos reúne estatísticas de grafos, análises e ML para contextualizar os dados e responder a perguntas urgentes.

Estatísticas de grafos, consultas de grafos (queries) e visualização impulsionam a exploração e os insights. **Estatísticas de grafos fornecem medidas básicas sobre um grafo**, como o número de **nós** e a distribuição de relacionamentos. As **consultas de grafos** respondem a qualquer pergunta, não importa a profundidade, seja com 6 ou 600 níveis de separação. A **visualização de grafos** permite que especialistas em dados vejam seus dados e explorem padrões que merecem uma investigação mais aprofundada.

A **análise de grafos** se baseia nas estatísticas de grafos, respondendo a perguntas específicas e ganhando insights de conexões entre dados existentes ou históricos. As consultas de grafos e os algoritmos são normalmente aplicados juntos em "receitas" durante a análise de grafos e os resultados são usados diretamente para análise.

ML aprimorado por grafos é a aplicação de dados grafos e resultados de análise para treinar modelos de ML ou apoiar decisões probabilísticas dentro de um sistema de IA. Estatísticas e análises de grafos são frequentemente usadas em conjunto para responder a certos tipos de perguntas sobre sistemas complexos e insights subsequentes, aplicados para melhorar o ML.

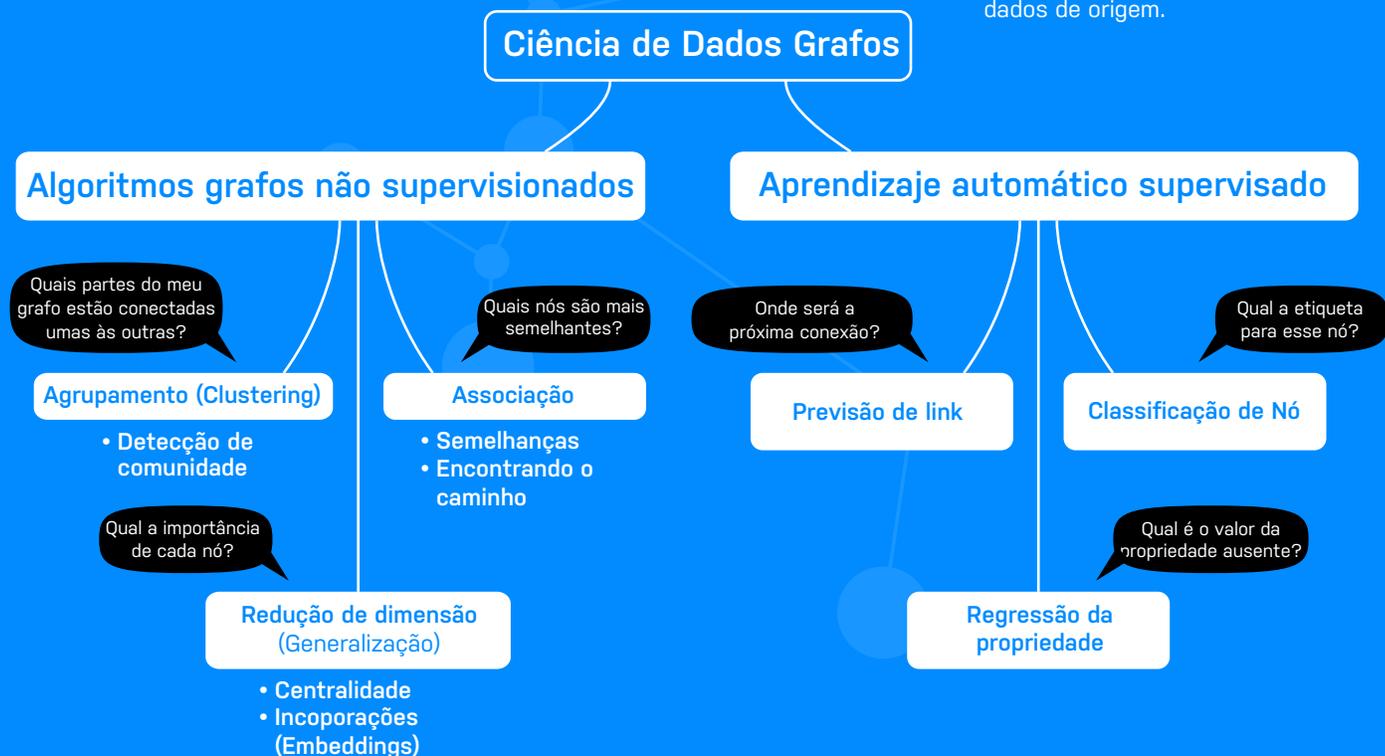
Visualização rápida: técnicas de ML em grafos que fazem o trabalho para você

O interesse em aprendizado de máquina baseado em grafos está explodindo. Até 2025, a Gartner projeta que a tecnologia de grafos alimentará 80% das inovações de dados e análises, frente aos 10% em 2021. Aqui estão algumas técnicas avançadas da ciência de dados grafos que sua equipe pode querer explorar.

A **previsão de link (Link prediction)** preenche os espaços em branco em seus dados e prevê mudanças na estrutura do seu grafo. Previsão de links é uma tarefa comum de aprendizado de máquina aplicada a grafos: treinar um modelo para aprender onde os relacionamentos devem existir entre pares de nós em um grafo. Você pode pensar na previsão de links como a construção de um modelo para prever relacionamentos ausentes em seu dataset, ou relacionamentos que são prováveis que se formem no futuro.

Com a ciência de dados grafos, você pode treinar modelos de ML supervisionados, baseados nos relacionamentos e propriedades do nó em seu grafo para prever a existência e a probabilidade de relacionamentos. A **incorporação de nó (Node embedding)** transforma a topologia e os recursos de seu grafo em uma representação vetorial de baixa dimensão de cada nó. Esses vetores, também chamados de **incorporações (embeddings)**, podem ser usados para análise exploratória de dados, medições de similaridade e ML. Incorporações de nó podem agregar informações sobre a posição de um nó no grafo, seus vizinhos locais, sua centralidade e influência, e, em alguns casos, outras propriedades numéricas do nó.

Os **modelos de classificação de nó (node classification)** preveem a classe do nó em seu grafo. Uma classe pode ser um indicador binário, como no caso de uma conta de usuário estar envolvida em uma fraude, ou um indicador de valor múltiplo, como indicador do segmento de mercado ao qual um cliente pertence. Os modelos de classificação de nós podem ser treinados para prever a quais classes de nós (incluindo novos nós) eles pertencem. A classificação de nós pode incorporar uma ampla gama de recursos de entrada, incluindo a estrutura de rede do seu grafo e propriedades de seus dados de origem.



3. Por que os grafos são importantes?

Em uma organização, todos os departamentos podem se beneficiar da ciência de dados grafos para responder a perguntas como quem ou o que é importante, o que devo fazer a seguir e o que há de incomum nisso? Enquanto os casos de uso para a ciência de dados grafos abrangem setores e linhas de negócios – de ciências biológicas a manufatura – alguns **casos de uso** estão se tornando rapidamente os mais populares entre os cientistas de dados.



Detecção de anomalias e fraudes

A detecção de anomalias em redes corporativas pode ajudar a identificar ataques de segurança cibernética e evitar a perda de dados. A mesma estratégia usada para identificar os agentes de ameaças em um contexto de segurança cibernética pode ser usada para detectar fraudes em programas bancários, de seguros e governamentais, analisando relacionamentos e comportamentos em seu grafo.

Aqueles que usam a ciência de dados grafos para conter fraudes observaram melhorias de detecção em mais de 300%, economizando milhões de dólares a cada ano.

Cliente 360

Em todo o mundo, as empresas tentam entender melhor seus clientes e melhorar o valor da vida útil (LTV). Com a ciência de dados grafos, o conhecimento do cliente pode se tornar mais preciso e completo por meio de resolução da entidade. Esse processo analisa todas as entradas do banco de dados e identifica duplicatas. A criação de uma entrada de banco de dados mestre completa para cada cliente, em vez de várias entradas incompletas, melhora o LTV e aprofunda o conhecimento do cliente, permitindo programas e ofertas de marketing otimizados.

Aqueles que usam ciência de dados grafos para uma visão 360 do cliente podem aumentar o conhecimento sobre ele em 30%.

Mecanismos de recomendação

Os mecanismos de recomendação tornaram-se conhecidos por meio da Netflix e das experiências de compras on-line. No entanto, os mecanismos de recomendação podem ser úteis para uma empresa em diversos níveis e para diferentes equipes: desenvolvimento de produtos, recursos humanos para retenção de funcionários por meio de treinamento e qualificação, e mecanismos de recomendação para potencializar algumas das partes mais importantes do negócio.

Gestão do Supply Chain

Melhorar a cadeia de suprimentos leva a economias, não apenas em reais, mas também em emissões de carbono. Otimizar a rota significa: tempo perfeito, entrega perfeita, clientes mais felizes e menos desperdício ao longo do tempo em infraestrutura e emissões. A ciência de dados grafos ajuda a otimizar a rota da cadeia de suprimentos, encontrando o melhor caminho, equilibrando custos e eficiência com a satisfação do cliente e sustentabilidade.

Um ROI de dezenas de milhões para uma organização usando ciência de dados de grafos para otimização de rotas não é nada comparado às 60.000 toneladas de emissões de carbono eliminadas usando essas rotas otimizadas.

4. Quais são as grandes perguntas que a ciência de dados grafos ajuda a responder?

A ciência de dados grafos ajuda você a responder grandes perguntas para tomar decisões comerciais importantes. Se você ouvir perguntas como estas, você pode se beneficiar do uso da ciência de dados grafos para respondê-las:

O que é incomum?

Para entender anomalias e padrões ocultos em seu grafo, considere o uso da **detecção de comunidades**. As comunidades são clusters dentro do seu grafo e [algoritmos de detecção de comunidade](#) podem ser usados para descobrir e identificar esses clusters. A detecção de comunidades ajuda a descobrir padrões incomuns, prever comportamentos similares, encontrar entidades duplicadas ou simplesmente preparar dados para outras análises.

O que devo recomendar?

Para construir um mecanismo de recomendação, considere o uso de **similaridade**. A similaridade identifica padrões de repetição em seu grafo. [Algoritmos de similaridade](#) empregam comparações de conjunto e distância para pontuar quão semelhantes os nós individuais são baseados em seus vizinhos ou propriedades. Esta abordagem é utilizada em aplicações tais como recomendações personalizadas e desenvolvimento de hierarquias categóricas.

Qual conteúdo é o mais importante? Quem é o mais influente? Onde está nosso maior risco?

Para entender o que é importante, use algoritmos de **centralidade**. As métricas de centralidade como o [PageRank](#) ajudam a identificar o que é importante. Os [algoritmos de centralidade](#) revelam quais nós são importantes com base na topologia de grafos. Eles identificam nós influentes com base em sua posição na rede, incluindo suas conexões. Estes algoritmos são usados para inferir dinâmicas de grupo, como credibilidade, vulnerabilidade em cascata e pontes entre grupos.

Qual é a melhor rota?

Para entender a otimização da rota, use o **pathfinding**. Os algoritmos de localização de rotas encontram as melhores rotas através de seus dados conectados. [Os algoritmos de localização de rotas](#) são fundamentais para a análise de grafos e encontram os caminhos mais eficientes ou mais curtos para atravessar entre os nós. Eles podem ser usados para compreender dependências complexas e avaliar rotas para usos como logística física e roteamento de chamadas de menor custo ou IP.

5. Quem usa ferramentas de ciência de dados grafos?

Qualquer pessoa que trabalhe com dados pode tirar proveito da ciência de dados grafos para ajudar a encontrar respostas para perguntas difíceis. Aqui estão algumas das funções que atualmente usam a ciência de dados grafos.



Os **cientistas de dados** geralmente são os principais usuários de ferramentas de ciência de dados grafos, porque são os profissionais de ciência de dados que costumam possuir uma compreensão profunda de algoritmos e modelos.



Os **engenheiros de aprendizado de máquina (ML)** trabalham para dimensionar, melhorar, integrar ou operar modelos de ML desenvolvidos por cientistas de dados.



Os **engenheiros de dados** identificam tendências em conjuntos de dados, criam pipelines de dados e utilizam algoritmos de grafos para transformar e enriquecer os dados grafos e torná-los mais úteis para a organização e para outros especialistas em dados.



Cientistas de dados informais aplicam técnicas de ciência de dados grafos sem a necessidade de análise profunda e experiência em ML, usando uma experiência com pouco código/sem código



Os **analistas de dados de negócios** visualizam, desenvolvem e relatam a análise realizada por um cientista de dados para usuários de negócios.



Desenvolvedores de aplicativos e arquitetos de software que aprendem sobre bancos de dados grafos geralmente usam consultas de grafos usando a [linguagem de consulta Cypher](#). Eles aprendem algoritmos de grafos enquanto tentam analisar padrões em seus grafos e usam essas técnicas em suas aplicações (a ciência de dados grafos geralmente é o ingrediente secreto em aplicações de diferenciação).

Fique à frente na curva

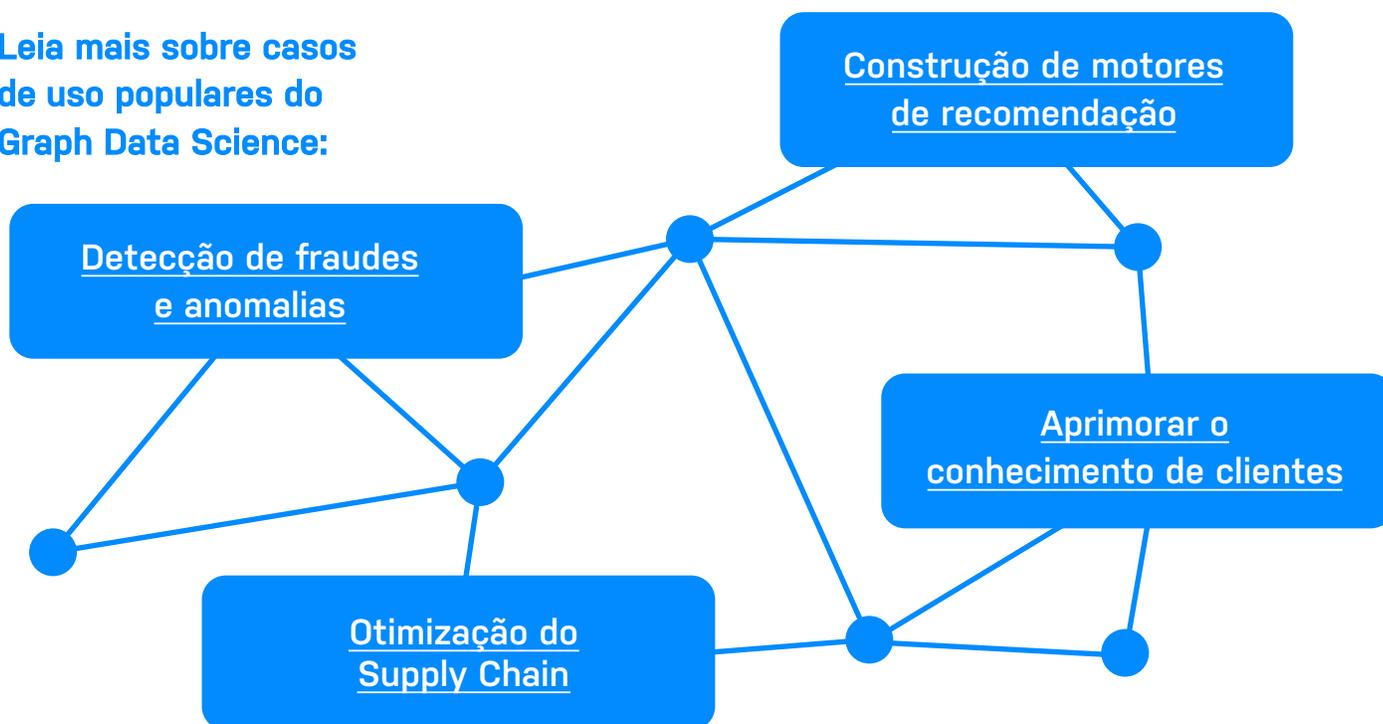
As aplicações comerciais da ciência de dados grafos são novas e os especialistas em dados ainda estão tentando entender a melhor forma de usar a ciência de dados grafos em suas organizações. Ao considerar casos de uso, especialistas em dados e cientistas de dados devem ter em mente que a ciência de dados grafos ajuda a responder perguntas essenciais, como o que é importante, o que é incomum e o que vem a seguir. Usando essa estrutura, é mais fácil identificar oportunidades de usar a ciência de dados grafos para melhorar modelos e fazer previsões.

Organizações de todos os tamanhos, todos os setores e dentro de todos os departamentos estão usando a ciência de dados grafos para fazer recomendações, identificar anomalias, encontrar fraudadores, melhorar o conhecimento do cliente e otimizar as cadeias de suprimentos.

Então, o que você fará com a ciência de dados grafos?

Pronto para experimentar? [Acesse o sandbox gratuito do Neo4j Graph Data Science](#) com dados para cenários comuns.

Leia mais sobre casos de uso populares do Graph Data Science:



A Neo4j é líder em tecnologia de banco de dados grafos. Seu banco de dados grafos é o mais implementado do mundo. Desta forma, ajudamos marcas globais – incluindo **Comcast**, **NASA**, **UBS**, e **Volvo Cars** – a revelar e prever como pessoas, processos e sistemas estão inter-relacionados.

Com uma abordagem que prioriza as relações entre os dados, as aplicações criadas com a Neo4j lidam com os desafios dos dados conectados, como **análise e inteligência artificial**, **detecção de fraudes**, **recomendações em tempo real** e **grafos de conhecimento**. Veja mais em neo4j.com.

Dúvidas sobre a Neo4j?
Entre em contato de qualquer lugar do mundo:

info@neo4j.com
neo4j.com/contact-us