



实时深度链接分析： 图分析的下一个里程碑

作者：TigerGraph首席执行官 许昱



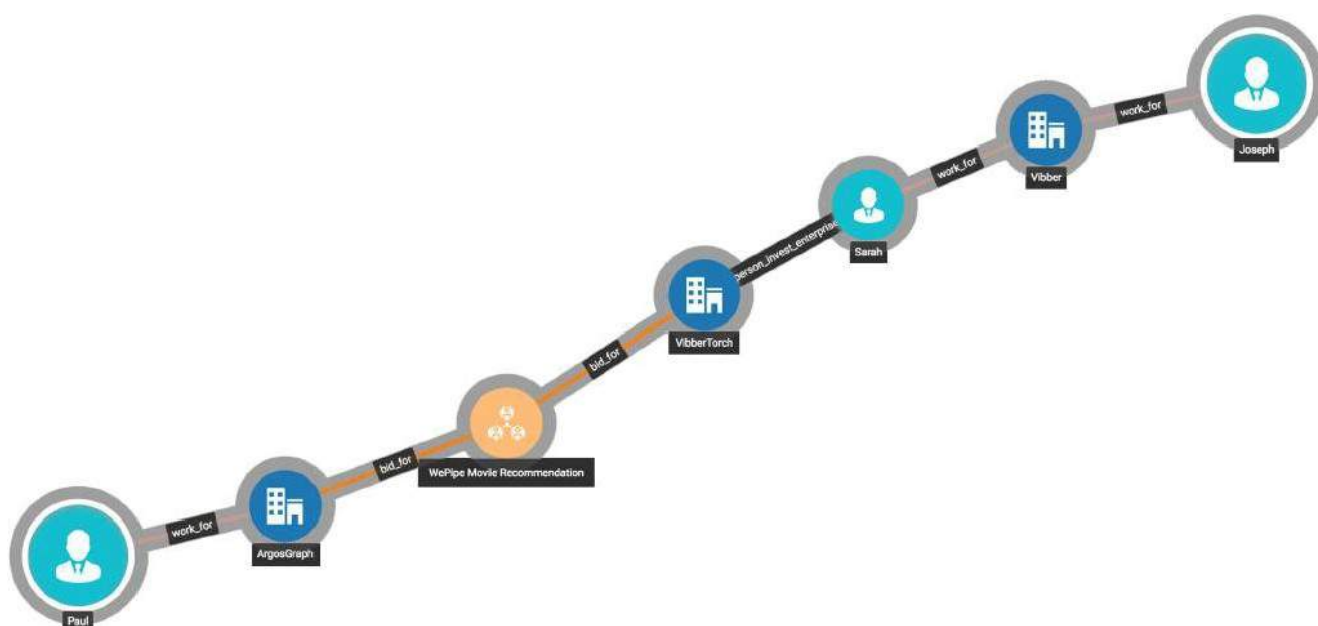
图数据库的最大优点之一在于它能够分析和回答海量数据中复杂的数据关联问题。与传统的关系数据库应用程序相比，早期图技术在这种关系分析方面已经显现出明显优势。但是，面对当今企业生态系统中普遍日益增长的实时数据时，早期图技术则显得运行缓慢，分析能力有限。

这是因为现有的图技术难以加载海量数据，无法提供令人满意的图遍历速度，并且无法导入快速到达的实时数据。因此，当前的图技术在大图中通常限于两步的实时遍历（即最多走两步）。

让我们一起来看看这个经典个性化的推荐问题：“那些与你喜好相同的顾客还购买了哪些商品？”。首先，从一个人开始，该查询首先确定你浏览/点赞/购买过的商品。其次，它确定浏览/点赞/购买过这些商品的其他顾客。第三，找出这些人购买过的其他商品。

人 → 产品 → （其他）人 → （其他）商品

该查询要求在图上实时走三步，因此它超出了当前图技术对于海量数据集2级跳的限制。如果再增加一层关系很容易将查询扩展到4步或更多步。



当今的企业因为业务需要，往往要求图技术能够实时做到3到10步以上的浏览和分析，从而发掘和预测错综复杂的关系。每增加一步，即可揭示更多连接和隐含的关系。

这就是**实时深度链接分析**，它具有以下性能属性：

性能	指标
可连接性高	3到10步以上
图遍历速度快	每个服务器每秒遍历超过1亿个顶点或边
数据更新快	每个服务器每秒更新10万次以上

实时深度链接分析应用案例

实时深度链接分析在当今企业的许多重要的数据分析应用中，成为一把利器。这里我们给出三个案例：实时反欺诈、个性化推荐、和优化供应链。



风险 / 欺诈控制

实时深度链接分析通过辨识高风险交易来打击金融犯罪。例如，从一个新的信用卡交易开始，该交易与其他实体发生关联的方式可以辨识如下：

新交易→信用卡→持卡人→（其他）信用卡→（其他）不良交易

该查询使用4步来寻找与传入交易只有一卡之隔的连接。现今的欺诈者们试图在自己与已知不良行为或不良行为者之间保持迂回连接，藉以掩盖自己的不法活动。任何单条连接路径可能看起来清白无辜，但是如果把从A到B的多条路径联合起来分析，很多欺诈行为被暴露出来的机会就大大增加了。因此，需要更多步才能找到相隔两个或多个交易的链接。这种遍历模式适用于许多其他用例。在这些用例中，你可以简单地使用网页点击事件、电话记录或汇款事件替换交易。通过实时深度链接分析，可以揭示多个隐藏的连接，从而最大程度发掘和及时阻止欺诈行为。



多维个性化推荐

为了找到类似的顾客，或者为产品提供个性化推荐，需要进行实时深度链接分析。请见以下推荐路径示例，其中所有路径均超过3步：

人→购买事件→产品→（被包含在其他的）购买（中）→

（其他）人→（其他）产品

人→购买事件→产品→产品的分类→（其他）产品

人→点击事件→产品→产品的分类→（其他）产品

由于当前系统的局限性，今天我们看到的多数推荐功能都是使用离线计算。这些推荐是在后台提前计算好的，不能利用最新数据执行实时的按需分析，从而丧失很多推荐机会或造成过时的推荐。



电力流 / 供应链物流 / 道路交通优化



这些应用调整各个实体和/或各个连接，直到建立一个动态平衡或最优的状态。例如，在国家或区域电网中，根据实际使用情况，动态调整每台发电机的电力输出。

所有实体

→与相邻实体的关系和对相邻实体的影响

→更新所有实体

〈重复直到到达稳定状态〉。

为了得到适当的值，需要在图数据上进行迭代计算（类似于PageRank），直到其中的度量值收敛为止。在这里，每个顶级组件（例如，发电机）是多个从属部件的网络枢纽，从而导致多层关联。在电力分配网络，关联程度通常需要6步。

发电机→变压器→控制单元→下层

变压器→下层控制单元→电表→用电设备

对于这类案例，只有实时深度链接分析才能达到这个级别的计算要求。



实时深度链接分析与企业人工智能

实时深度链接分析是企业人工智能的关键组成部分。由于其技术局限性，早期的数据分析技术人为地打散实体和关系，并将它们作为数据行或键值对，分散存储在多个表/行/列中。从而导致一些关系丢失，或者增加处理难度。实时深度链接分析可以处理数据实体之间的所有连接，使用户能够更准确、更快速和更深入地浏览、发现和预测关系。这些功能对于个性化推荐、欺诈防范、供应链物流优化、公司知识图谱构建和查询等企业人工智能应用都是至关重要的。

介绍 TigerGraph

TigerGraph是第一个也是唯一的原生并行实时图分析平台；TigerGraph是第一个也是唯一一个具有实时深度链接分析能力的图数据库。TigerGraph可以实现每台计算机每秒遍历上亿个顶点/边，并遍历三到十步以上，所有遍历速度比传统方式快几个数量级。它还包括每台服务器每秒10万次以上的快速数据更新。TigerGraph旨在结合原生并行图的存储和计算，达到支持实时图更新和提供内置并行计算。如欲了解有关原生并行图的更多信息，请访问www.tigergraph.com阅读白皮书“原生并行图：图数据库演进的下一个里程碑”。

关于TigerGraph

TigerGraph是基于原生并行图（NPG）技术的全球首个实时图分析平台。TigerGraph通过为具有复杂和海量数据的企业提供实时深度链接分析支持，实现图平台的真正承诺和好处。TigerGraph的成熟技术已经被蚂蚁金服、VISA、软银、中国国家电网公司、Wish、Elementum等客户所采用。

TigerGraph由许昱博士于2012年创立，并获得启明创投、百度、蚂蚁金融、AME云创投、莫拉多风险投资公司、佐德·纳齐姆、丹华资本和DCVC风投基金公司的投资。TigerGraph的总部位于加利福尼亚州红木市。如欲了解更多信息，请访问www.tigergraph.com。



©2017 TigerGraph