TigerGraph

中亦科技
China Etek Service

图分析技术助力于金融业务转型升级

锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

1

图分析技术的分类

2

图技术在金融行业的应用

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图分析技术的分类

锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图分析技术的分类

1. 图查询
——使用图数据库的查询语言进行点边的搜索

2. 图算法
——中心性算法（Centrality）
——社区算法（Community）
——路径算法（Path）
——相似度算法（Similarity）

3. 图深度学习
——图嵌入（Graph Embedding）
——图卷积神经网络（Graph Convolution Networks）
——图注意力网络（Graph Attention Networks）
——图自编码器（Graph Autoencoders）

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图查询的介绍及应用

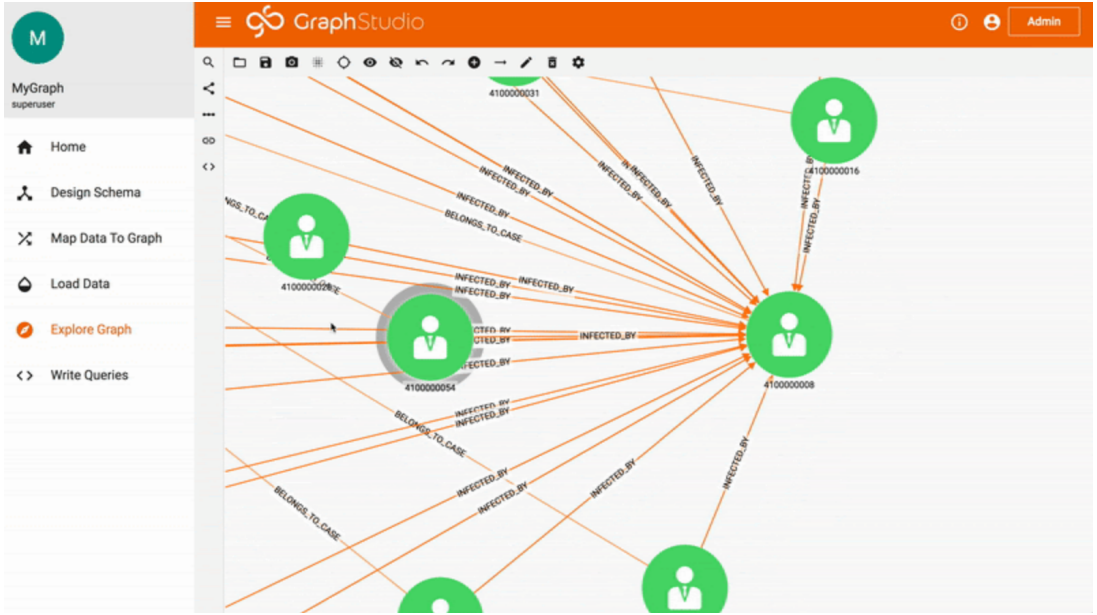
锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图查询的介绍

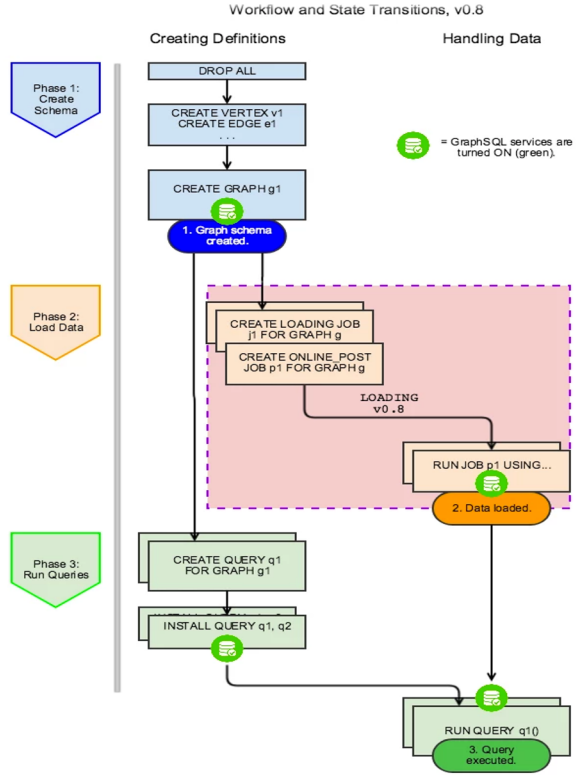
图查询：

使用图数据库的查询语言进行点边的关联查询，可以快速完成传统数据库难以完成的多度点边关联。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图数据库工作流程（TigerGraph）：



设计Schema → 根据自身数据仓库的数据情况和实际需求，设计Schema（图结构）。

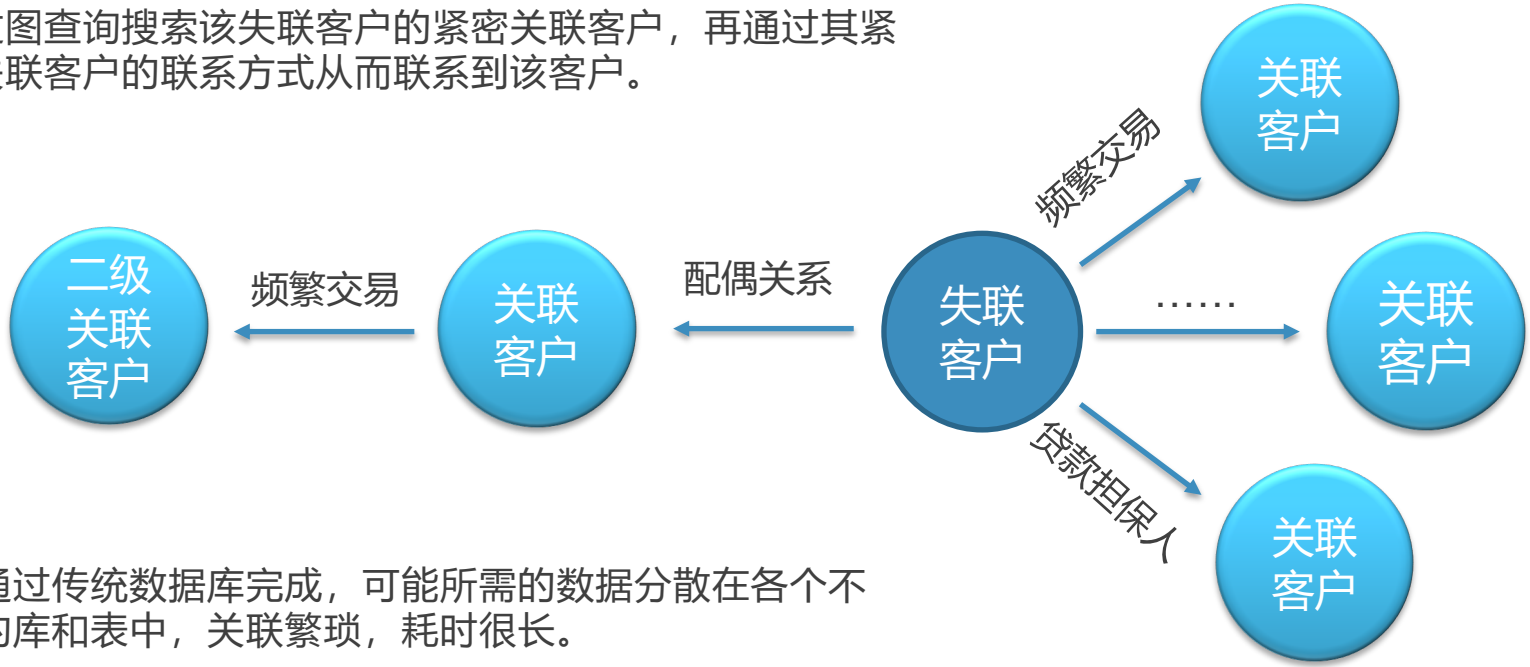
导入数据 → 准备数据，建立二维表的数据到图中点/边的映射关系，将数据导入图数据库。

编写图查询 → 根据业务需求编写图查询，利用图数据库平台进行数据分析

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图查询——个贷失联客户找回

通过图查询搜索该失联客户的紧密关联客户，再通过其紧密关联客户的联系方式从而联系到该客户。



若通过传统数据库完成，可能所需的数据分散在各个不同的库和表中，关联繁琐，耗时很长。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图算法的介绍及应用

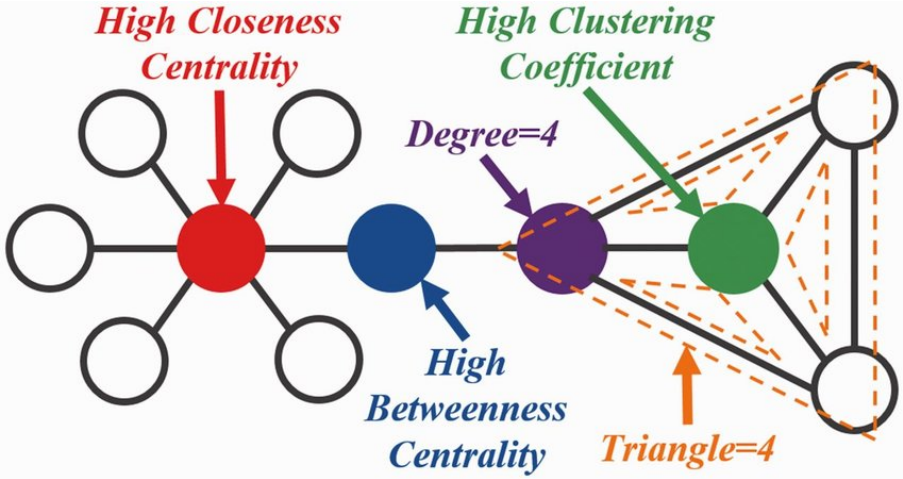
锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

中心性算法

中心性算法：衡量节点在图结构的不同维度的重要程度

——度中心性（Degree Centrality）
——接近中心性（Closeness Centrality）
——中介中心性（Betweenness Centrality）
——页面排名（PageRank）



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

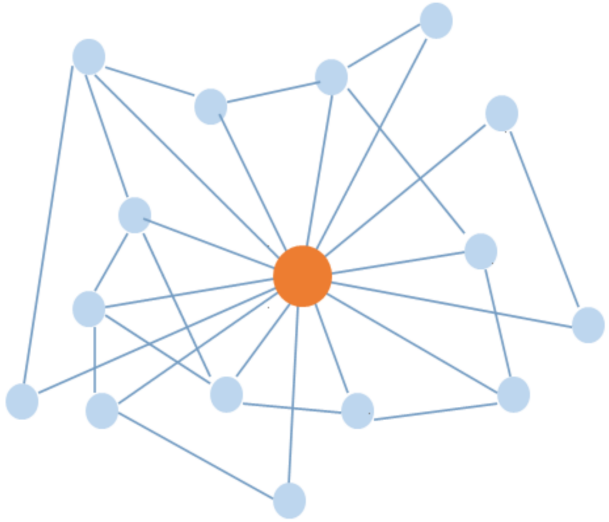
——度中心性（Degree Centrality）

【介绍】

度中心性是刻画节点中心性的最直接度量指标。一个节点的度越大就意味着这个节点的度中心性越高，该节点在网络中就越重要。对于有向图，还要分别考虑出度/入度/出入度。需要注意的是，由于不同的关系网络由于规模不同，在互相比较时，度中心性还要考虑去规模化等问题。

【计算】

直接统计节点的出度、入度即可。

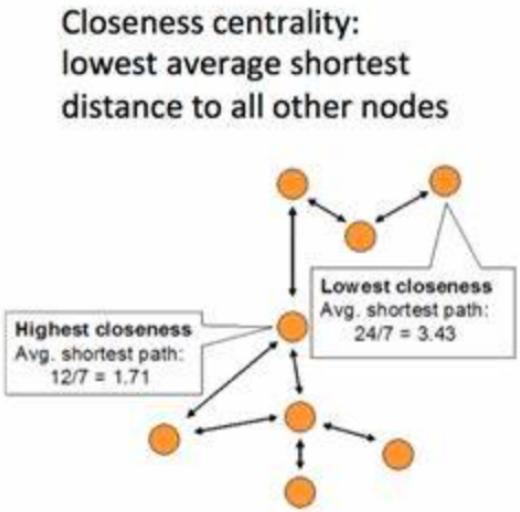


© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——接近中心性（Closeness Centrality）

【介绍】

接近中心性反映在网络中某一节点与其他节点之间的接近程度。用一个节点到所有其他节点的最短路径距离累加求和后取倒数来表示接近性中心性。接近中心性考虑的本质是每个节点到其它节点的最短路径的平均长度，因此对于一个节点，它距离其他所有节点越近，那么它的接近性中心性越大。



【计算】

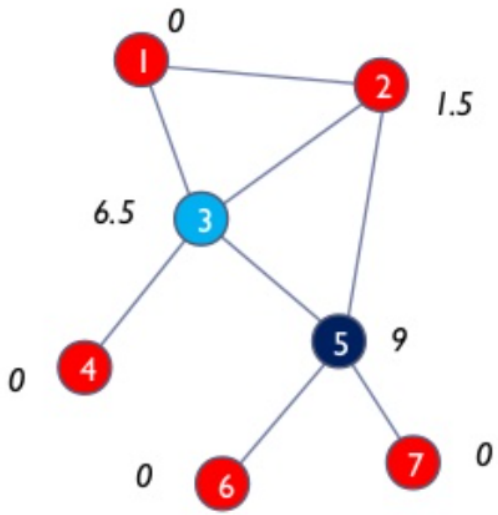
节点到所有其他节点的最短路径求和的倒数

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——中介中心性（Betweenness Centrality）

【介绍】

中介中心性以经过某个节点的最短路径数目来刻画该节点的重要性。中介中心性的本质是一个节点担任其它任意两个结点之间最短路径桥梁的次数。一个节点充当“中介”的次数越高，它的中介中心度就越大。若考虑标准化的问题，则需要用一个节点承担最短路桥梁的次数除以所有的最短路径的数量。



【计算】

经过某节点的最短路径条数 / 图中所有最短路径的条数

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

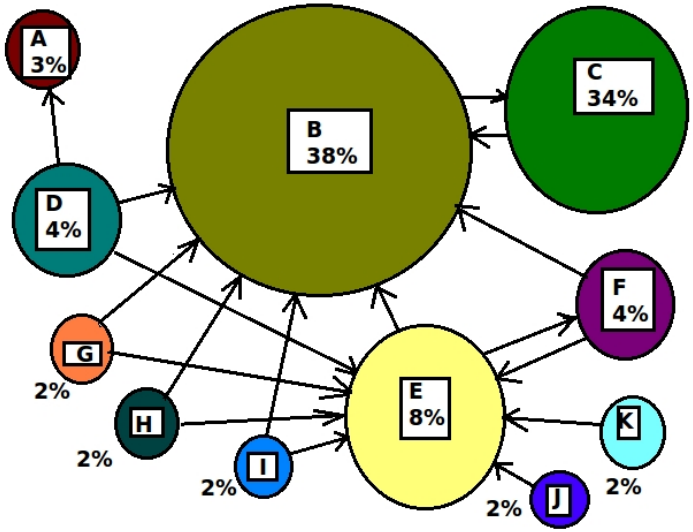
——页面排名（PageRank）

【介绍】

一个页面的“得分数”由所有链接向它的页面的重要性来决定，到一个页面的超链接相当于对该页投一票。一个页面的最终PageRank是由所有链向它的页面的重要性经过递归得到的，因此一个有较多链入的页面会有较高的等级。

【计算】

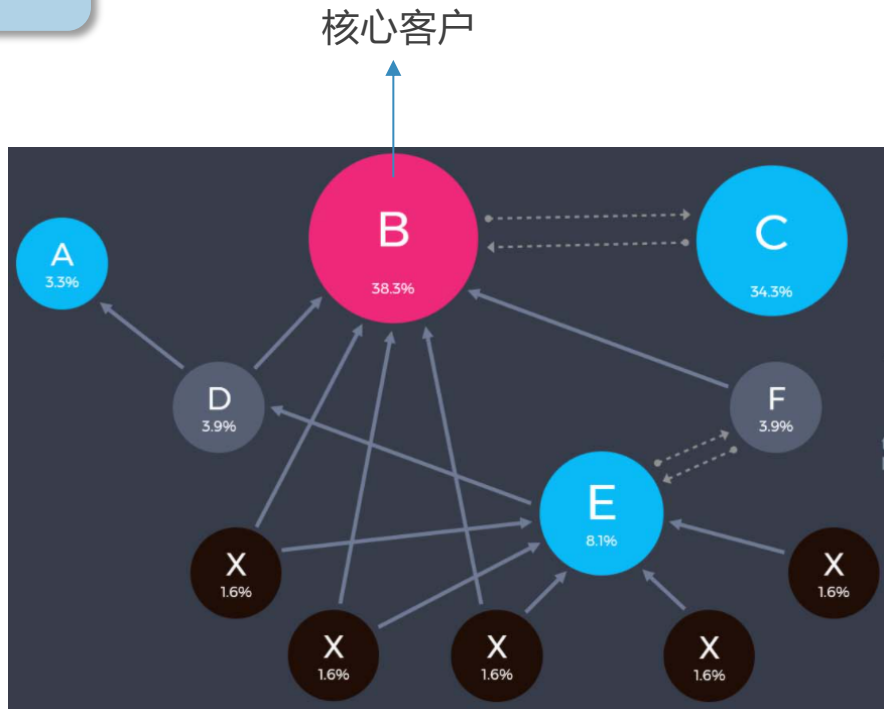
为每个节点赋予初始Rank值，然后给定阻尼系数（不通过边传播Rank，而是随机跳到一个节点的概率），进行Rank的加权传递。迭代此过程直到各节点的Rank稳定。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

PageRank——核心客户识别

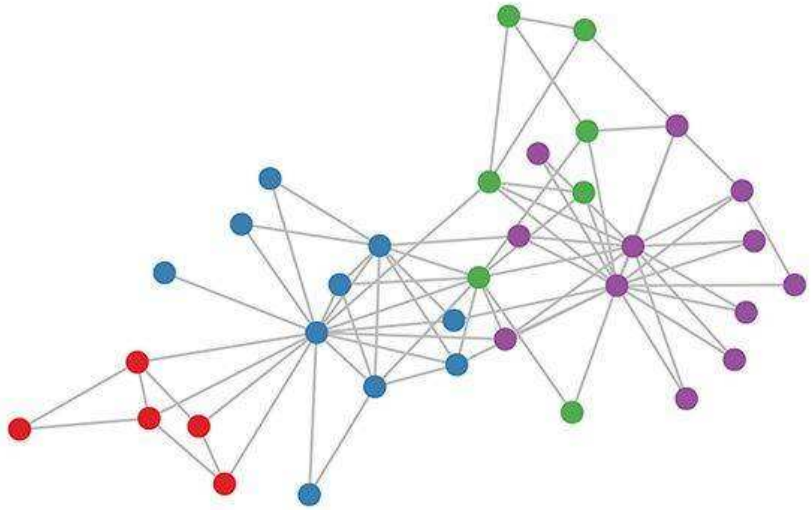
根据信用卡客户的所有转账交易，通过给定聚合函数的方式计算客户之间的交易亲密度，在交易亲密度的边权重基础上利用PageRank算法计算每个客户的PageRank值，通过PageRank值的大小识别重要核心客户。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

社区算法：将图数据中的节点进行社区划分，得到在某些维度具有高度相似性的节点集合

——连通分量（Connected Components）
——标签传播（Label Propagation）
——鲁汶算法（Louvain Algorithm）
——三角计数（Triangular Count）



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

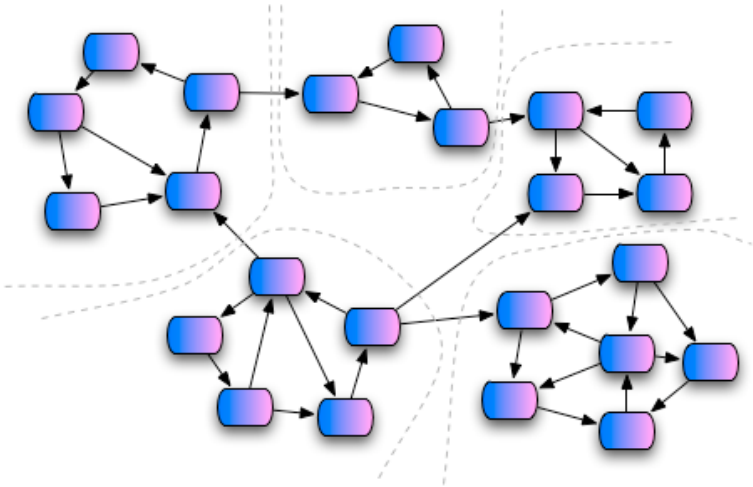
——连通分量（Connected Components）

【介绍】

两个节点如果有路径（点-边-点-...-点-边-点）相连，则称这两个节点连通。一个连通分量中任意两个节点都联通，因此一个连通分量就是一个社区。

【计算】

将有边相连的节点归入同一个社区即可。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——标签传播（Label Propagation）

【介绍】

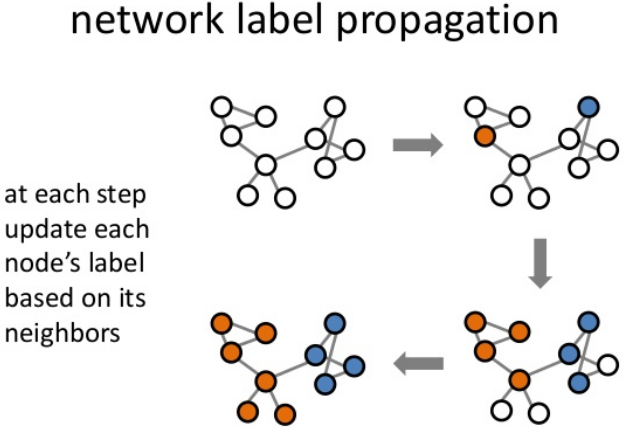
利用关系网络中的边，将节点的标签进行互相传播，通过迭代得到每个节点稳定的标签，从而获得社区划分。

【计算】

(1)为所有节点指定一个唯一的标签。

(2) 对于某个节点，考察其所有邻居节点的标签并进行统计，将出现个数最多的那个标签赋给当前节点。当个数最多的标签不唯一时，随机选一个。

(3) 依上述规则，逐轮刷新所有节点的标签，直到节点标签不再变化为止。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——鲁汶算法（Louvain Algorithm）

【介绍】Louvain算法是基于模块度的社区算法，它的优化目标为最大化整个数据的模块度，模块度计算如下：

$$Q = \frac{1}{2m} \times \sum_{i,j}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j)$$

其中m为图中边的总数量；
ki表示所有指向节点i 的边权重之和；
Aij 表示节点i，j之间的边权重；
ci 表示节点i所属的社区；
δ函数表示：当p=q时δ(p,q) 为 1，否则为0。

【计算】
(1) 将图中的每个节点看成一个独立的社区，此时社区的数目与节点数相同；
(2) 对每个节点i，依次尝试把节点i分配到其每个邻居节点所在的社区，计算分配前与分配后的模块度变化δ(Q)，把节点i分配给δ(Q)最大的那个邻居节点所在的社区 (3) 重复(2)，直到所有节点的所属社区不再变化；
(4) 对图进行压缩，将所有在同一个社区的节点压缩成一个新节点，两个新节点之间边的权重是两个新节点内所有原始顶点之间相连的边权重之和；
(5) 重复(1)-(4)直到整个图的模块度不再变化或变化很小。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

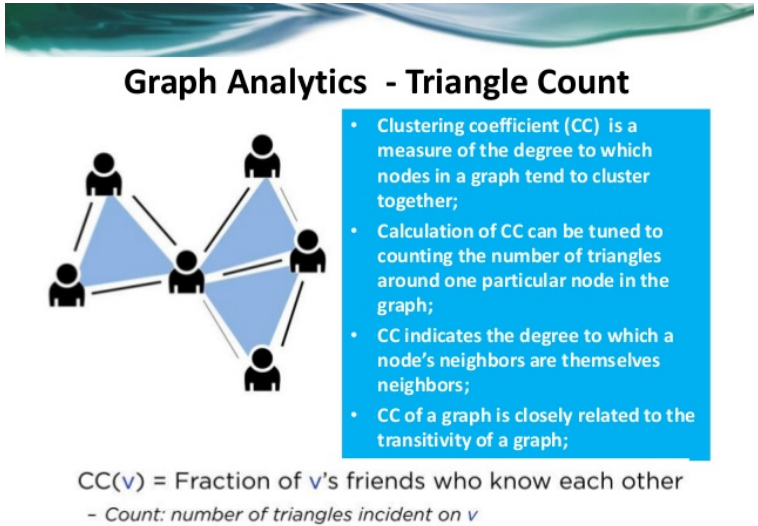
——三角计数（Triangular Count）

【介绍】

一个节点集合中，具有三角形关系（有向或无向）节点的三角形个数。三角形关系数量越多，代表该点集的节点之间关系越紧密。也可以三角计数作为社团分类的标准。

【计算】

从任意节点出发，经过两度边后回到原来节点，即一个三角形。遍历所有节点后计算三角形之和除以3即可。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

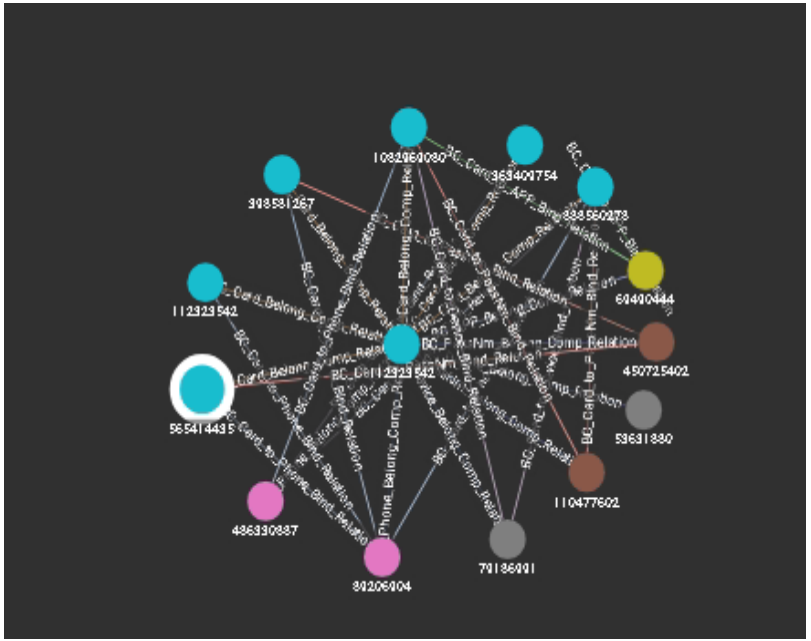
连通分量算法——持卡网络实体解析

根据某银行现有数据，包括大量手机号、卡号、身份证号、设备号、APP账号等之间的关联事件，利用这些关联事件构建持卡人关系网络图。
使用连通分量算法来解析不同的持卡人实体。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

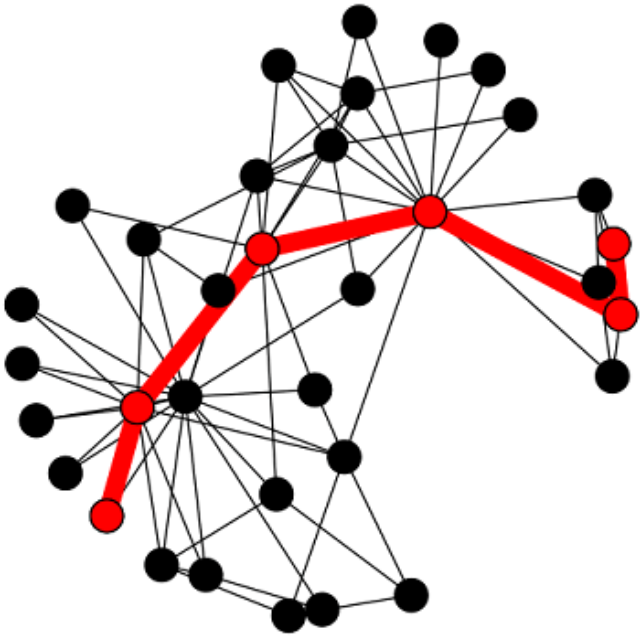
连通分量算法——持卡网络实体解析

一个连通分量实例中，可能有一张身份证，多个设备、多个手机号、多张借记卡、多张信用卡。我们通过连通分量，认定同一个连通分量中的节点都属于同一个持卡人实体。下一步做数据探索和场景应用的时候（例如反洗钱，反欺诈，营销推荐等），就把这些节点认定是同一个持卡人（利益共同体）。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

路径算法：搜索给定的图结构/图路径
——环路探测（Cycle Detection）
——最小生成树（Minimum Spanning Tree）
——最短路径（Shortest Path）



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

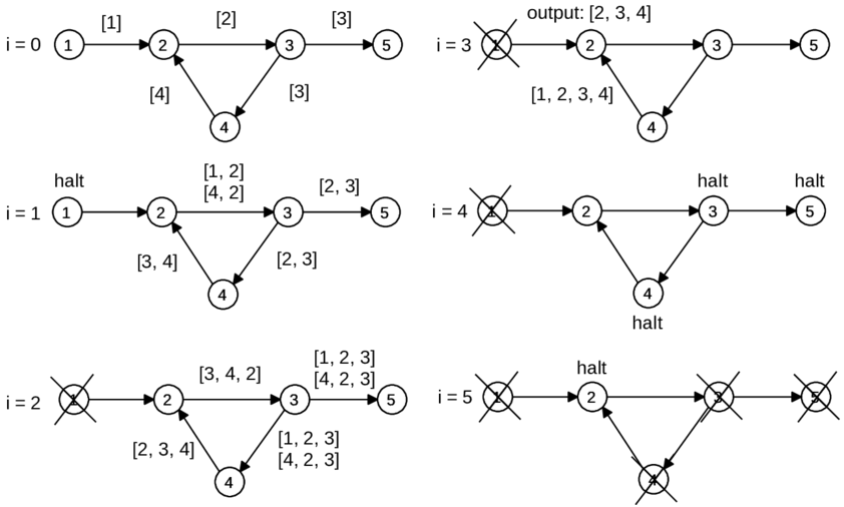
——环路探测（Cycle Detection）

【介绍】

计算一个节点集合中的环路条数。环路是首尾相连的路径，不同点、边的类型构成的环路，具有不同的实际含义。环路探测可以为实际场景提供分析帮助。

【计算】

Rocha–Thatte cycle detection algorithm



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——最小生成树（Minimum Spanning Tree）

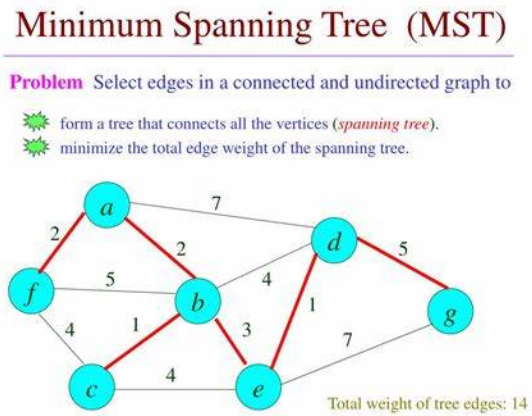
【介绍】

生成树：一个连通图的生成树是指它的一个连通子图，其含有连通图中全部的n个节点，但只有足以构成一棵树的n-1条边。一颗有n个节点的生成树有且仅有n-1条边，如果生成树中再添加一条边，则必定会成环。

最小生成树：在连通图的所有生成树中，所有边的权重之和最小的生成树，称为最小生成树。

【计算】

Kruskal算法，Prim算法等。

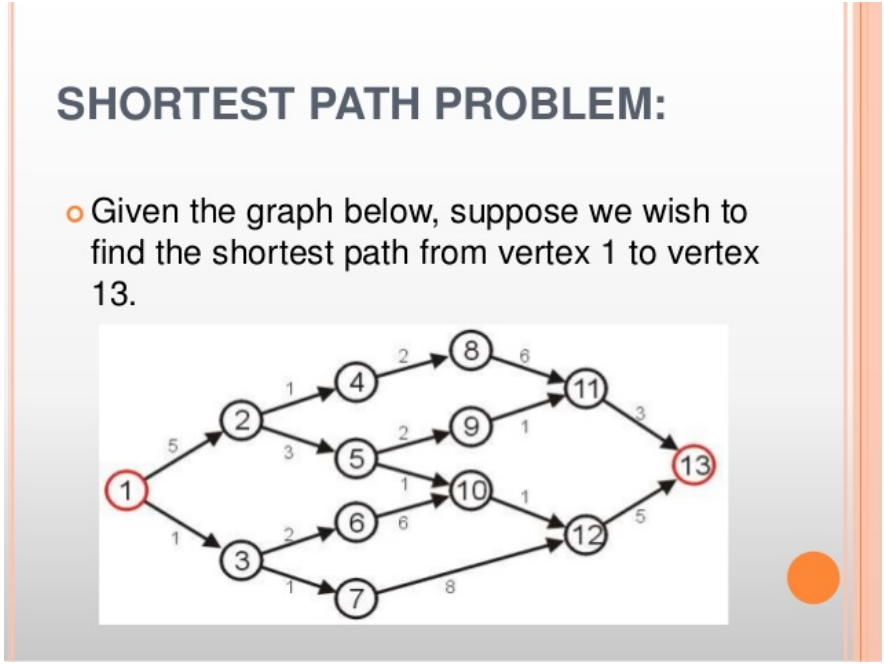


© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——最短路径（Shortest Path）

【介绍】
计算图中任意两节点之间的最短路径。

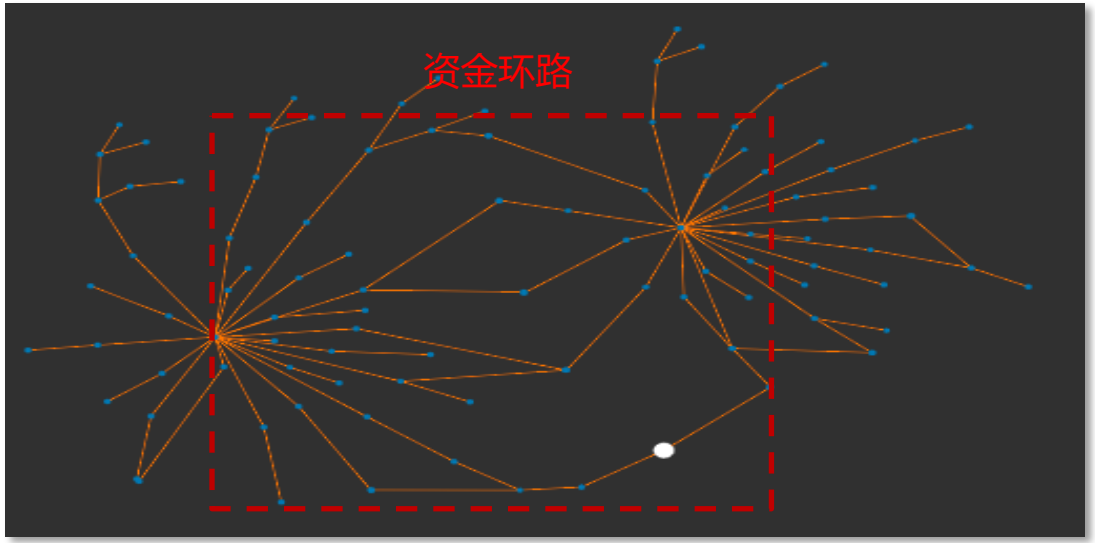
【计算】
Dijkstra算法，Floyd算法，SPFA算法，Bellman-Ford算法等。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

环路探测算法——反洗钱资金环路挖掘

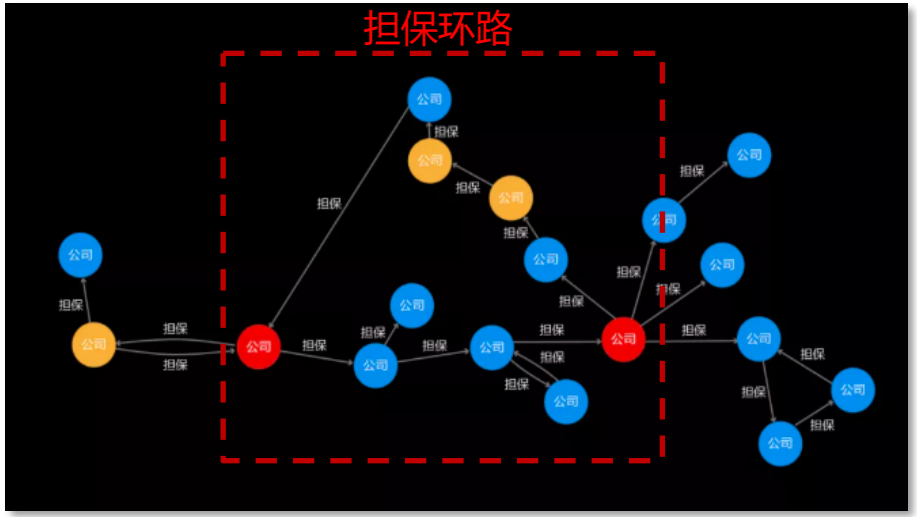
背景：在银行的反洗钱风控业务中，经常遇到洗钱团伙经过多手资金周转进行洗钱的行为，增加了侦测难度。
算法应用：在客户的转账关系网络中，利用环路探测算法，搜索资金环路，挖掘到潜在的洗钱团伙的资金链路，以及链路中的详细流水情况，为后续业务排查提供更多参考。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

环路探测算法——反洗钱资金环路挖掘

背景：中小企业常常通过企业担保来获得贷款，而随着担保链条的不断延长，会形成担保环路，大大增加了对公贷款的系统性风险。
算法应用：在客户的担保关系网络中，利用环路探测算法，搜索客户担保环路，分析担保环路中的企业相互关系，为对公贷款业务风控提供参考。

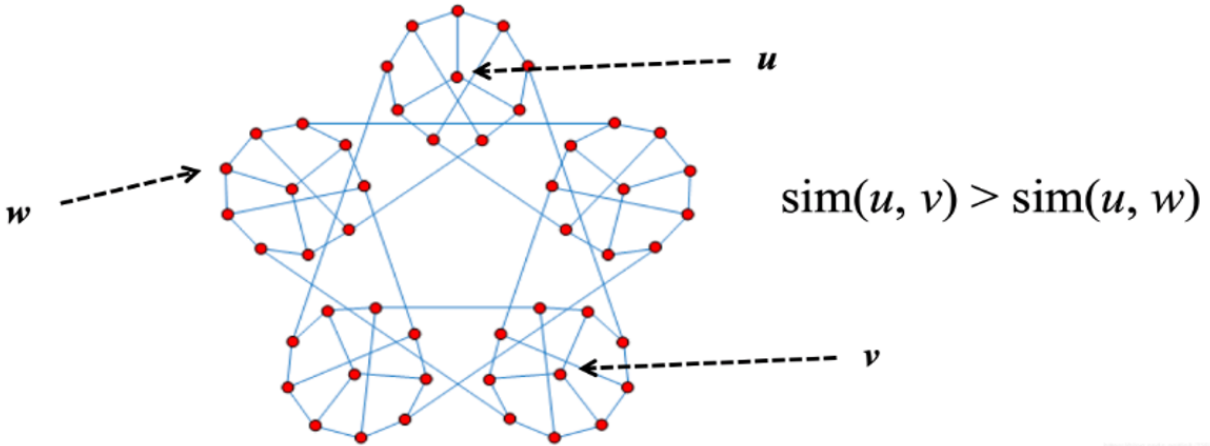


© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

相似度算法：衡量图中两个节点的相似程度

——余弦相似度（Cosine Similarity）

——杰卡德相似度（Jaccard Similarity）



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

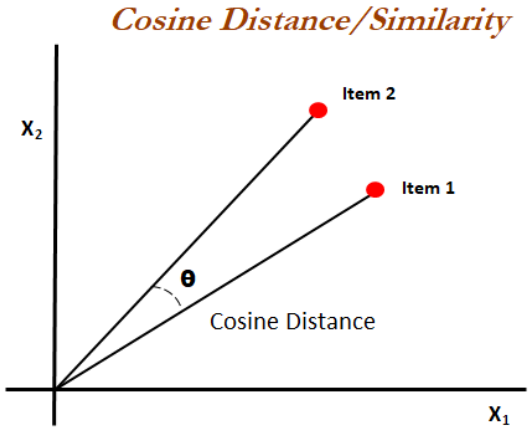
——余弦相似度（Cosine Similarity）

【介绍】

节点A和节点B的余弦相似度定义为：（A dot B）/（||A|| * ||B||）
向量的点乘就是向量各对应元素两两相乘的和，而图中节点的点乘需要另行定义，可以是边权重序列，也可以是邻居权重序列。对于不等长向量的相似度分析，可以参考动态时间规整法（Dynamic Time Warping）。

【计算】

对于两个节点，计算点乘（内积）和各自范数，相除即可。

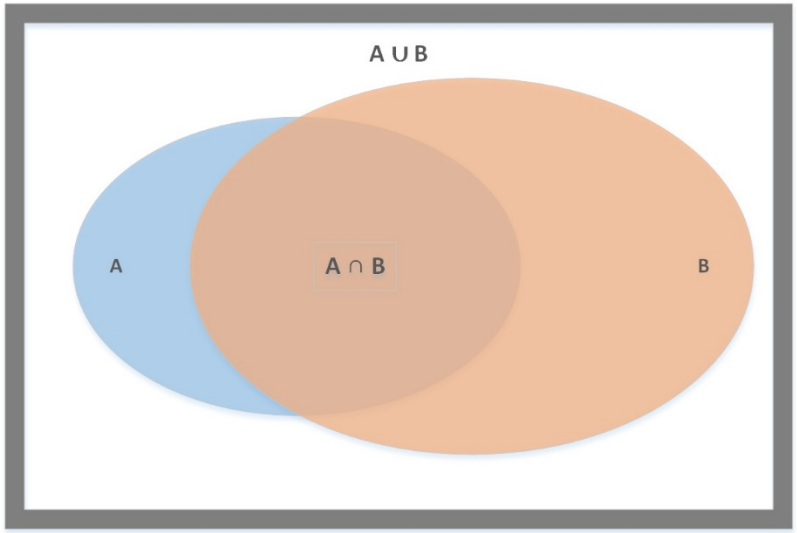


© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

——杰卡德相似度（Jaccard Similarity）

【介绍】
节点A和节点B的杰卡德相似度定义为节点A邻居和节点B邻居的交集节点数量除以并集节点数量。Jaccard系数计算的是两个节点的邻居集合的重合程度，以此来衡量两个节点的相似度。

【计算】
计算两个节点邻居集合的交集和并集，然后相除即可。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

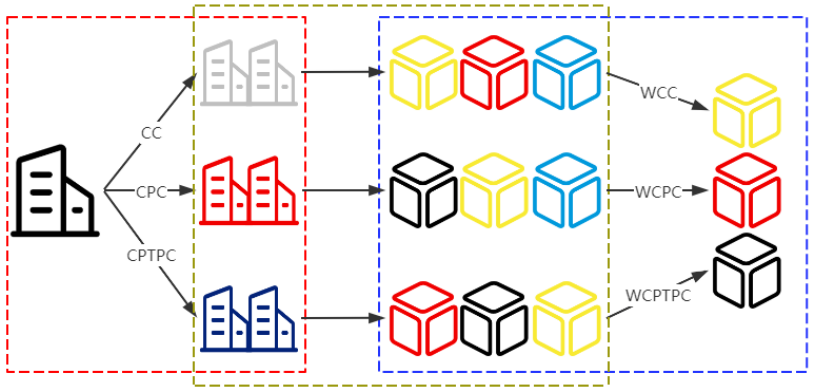
基于元路径的相似度算法——营销产品推荐

主要思想：

1.根据相似性度量基于不同元路径计算不同公司之间的相似度；

2.根据相似公司已拥有的产品预测公司对于不同结算产品的偏好；

3.根据不同的权重对不同元路径下得到的结果进行加权平均，融合成为最终的结果。



语义关系表		
元路径	语义关系	推荐模式
CC	目标公司的子母公司	基于公司关系推荐
CPC	与目标公司拥有相同结算产品的公司	基于协同过滤的推荐
CPTPC	与目标公司拥有同类结算产品的公司	基于产品内容推荐

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

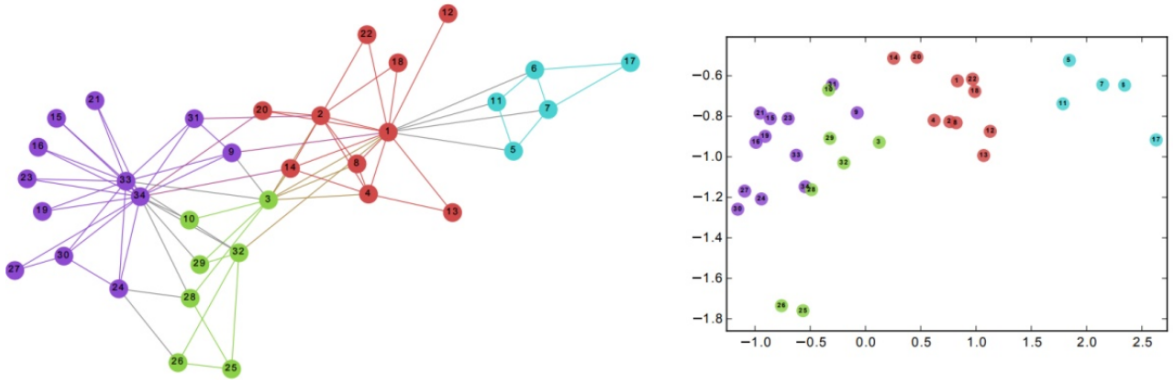
图深度学习的介绍及应用

锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图嵌入

1. 图嵌入的数学本质是一个映射 f: X —> Y，将一个空间的点映射到另一个空间。
2. 图嵌入所代表的映射通常是从高维到低维的（如果要完整记录图结构的所有信息，需要N*N的邻接矩阵，而图嵌入向量矩阵只有N*K，且K远小于N）。
3. 一般图嵌入映射到的低维空间表示具有分布式稠密表示的特性。



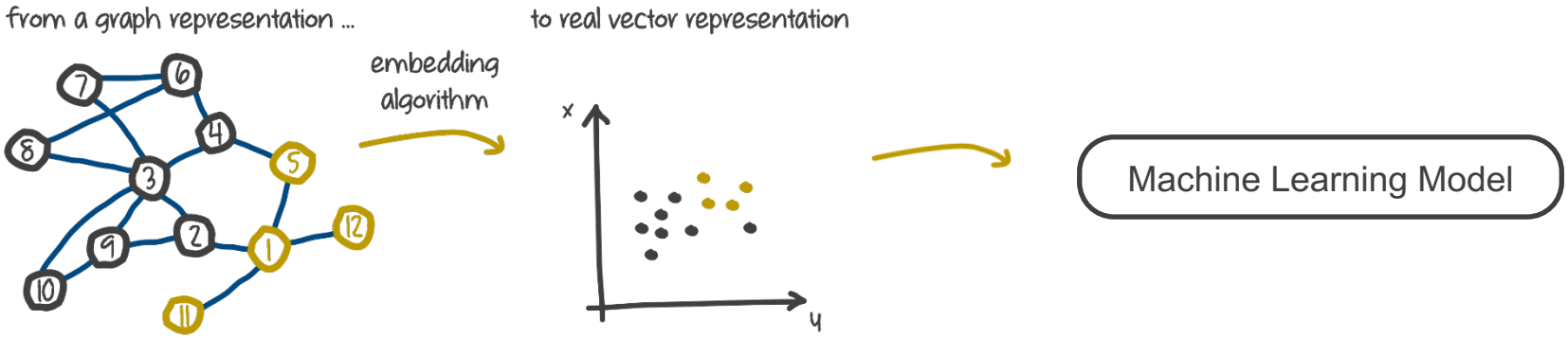
(a) Input: Karate Graph　　(b) Output: Representation

N维临接矩阵（高维、稀疏）　→　K维嵌入向量（低维、稠密）

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

担保网络图嵌入——对公贷款违约模型

构建对公客户的担保关系网络，使用Node2Vec算法对担保关系网络进行图嵌入，得到每个节点（客户）的嵌入向量（嵌入向量反应了该节点在关系网络中的拓扑结构）。将图嵌入向量作为特征，结合客户的行为特征和属性特征，训练客户贷款违约模型，预测客户在未来潜在的违约情况。



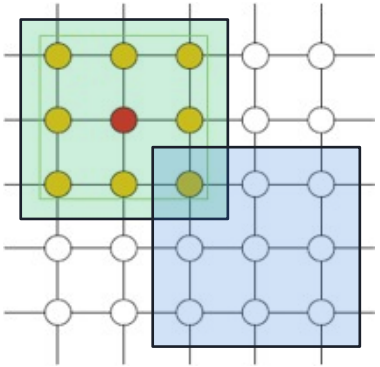
© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图卷积神经网络

——欧式空间的结构化数据

固定的数据维度、局部数据有序

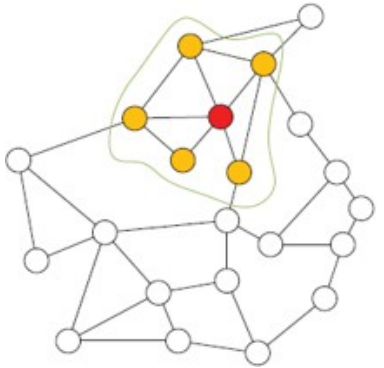
语音、图像、视频（规则结构）满足以上要求



——非欧空间结构的数据

局部维度可变

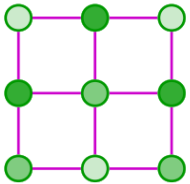
局部排列无序



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图卷积神经网络

经典卷积

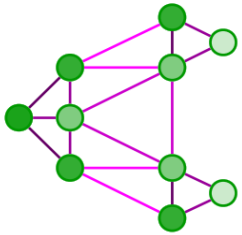


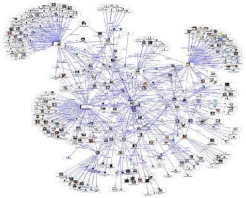
语音

图像

视频

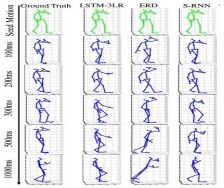
图卷积





社交网络

分子结构

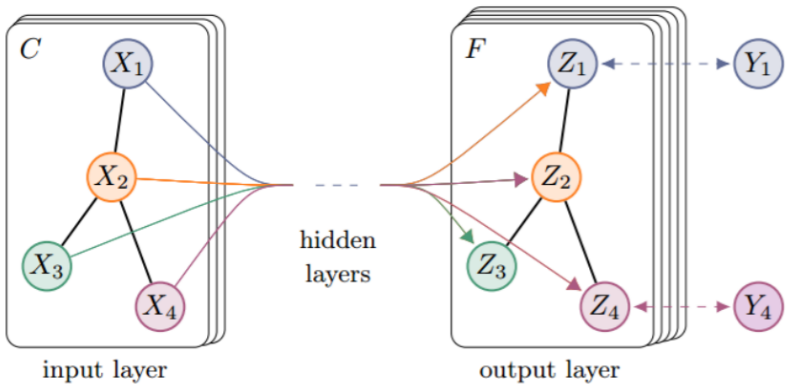


人体骨架

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图卷积神经网络应用

$X \in \mathbb{R}^{n \times C}$ $Z \in \mathbb{R}^{n \times F}$



$X_1 \in \mathbb{R}^{1 \times C}$ $Z_1 \in \mathbb{R}^{1 \times F}$

$\mathcal{V}_t \in \mathbb{R}^{n \times c}$

节点回归：

1. 给定一个图结构（如使用对公客户之间的转账网络）

2. 使用[t - T' + 1，t]时刻的信号作为输入（如前一个月以来，客户每日的资金流入量）

3. 使用[t + 1，t + T]时刻的信号作为输出（如接下来一周，客户每日的资金流入量）

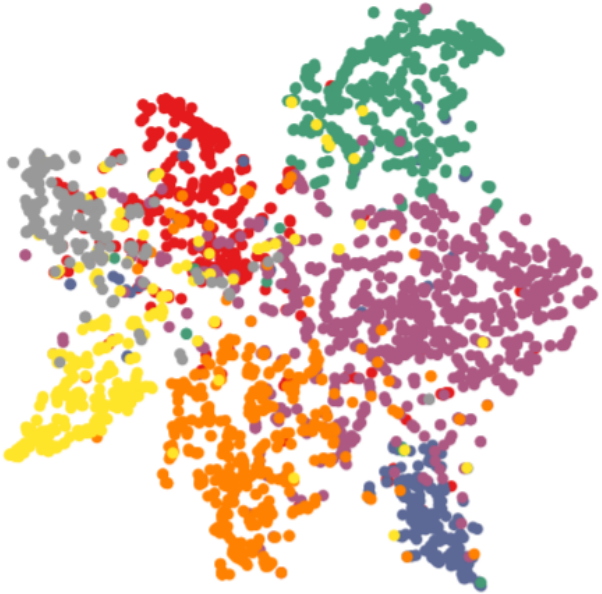
4. 输入样本进行模型训练，完成模型后可用以预测将来客户的资金变动情况，进行进一步的风控、营销的应用。

$$[\mathcal{V}_{t-T'+1}, \mathcal{V}_{t-T'+2}, \cdots, \mathcal{V}_t; \mathcal{G}] \xrightarrow{h(\cdot)} [\mathcal{V}_{t+1}, \mathcal{V}_{t+2}, \cdots, \mathcal{V}_{t+T}]$$

重点在于如何构建图结构、如何选取节点信号、后续模型的实际应用。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图卷积神经网络应用



节点分类：

1. 给定一个图结构（如企业担保关系网络）

2. 给部分贷款违约企业打上坏样本标签，部分信用良好的企业打上好样本标签，其余企业无标签（用于预测）。

3. 给每个企业节点添加上客户属性以及历史行为的特征。

4. 使用GCN进行半监督学习，预测无标签企业的违约概率。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图分析技术致力于银行应用场景

锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

中禾科技
China Etek Service

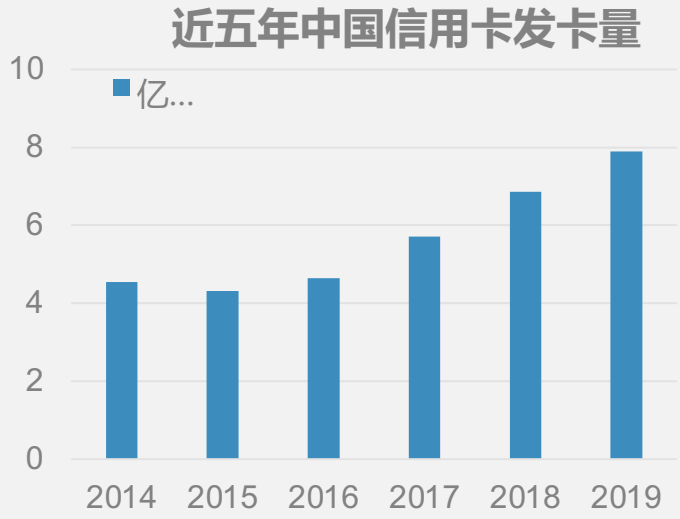
图分析技术致力于银行反套现

锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

信用卡套现

信用卡套现盛行导致信用卡不良率攀升，共债风险向银行传导。银行风控定价失效，产生过度授信，造成贷款市场不合理膨胀，严重情况会引发金融波动。

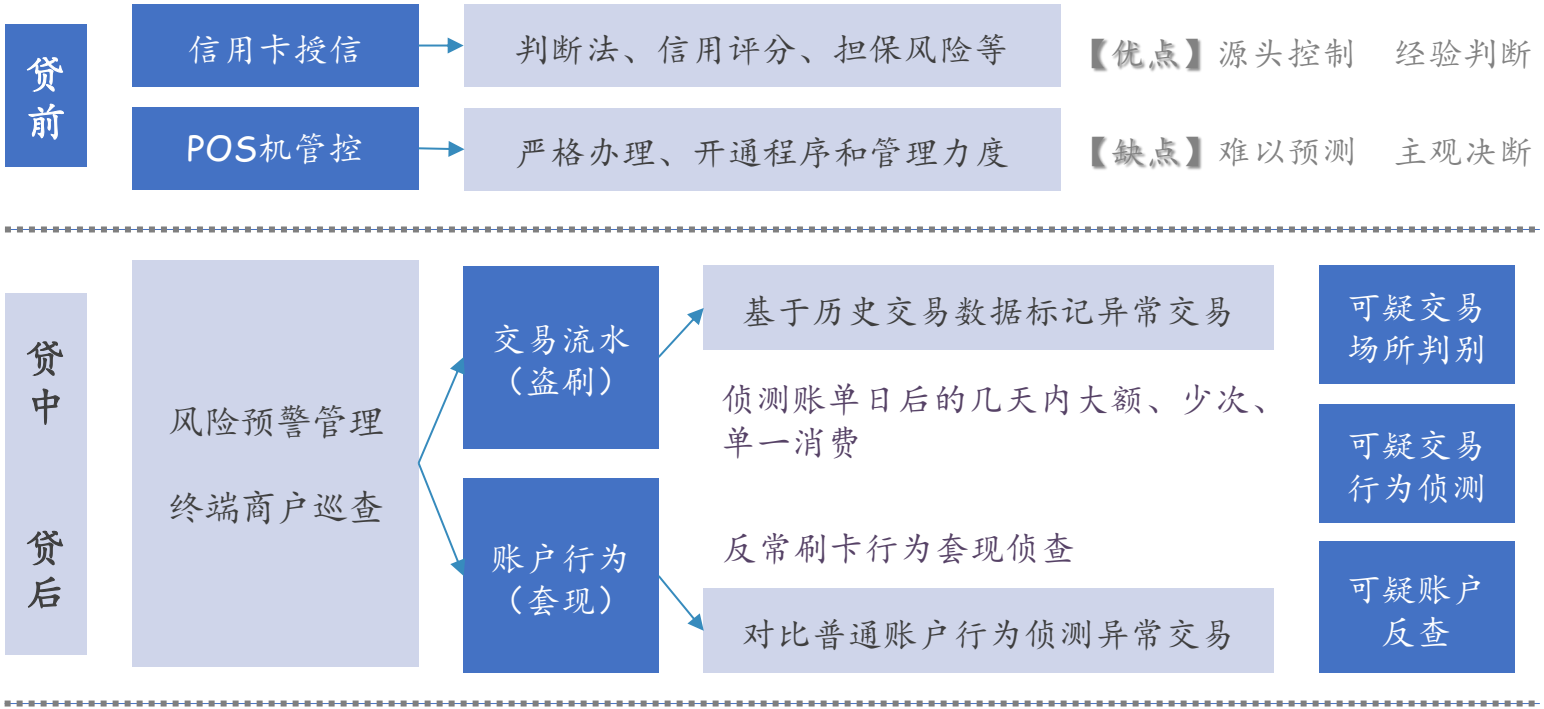


各大银行近五年信用卡不良率

银行	2014	2015	2016	2017	2018	2019
建设银行	0.85%	1.08%	1.28%	1.17%	1.09%	1.21%
交通银行	1.68%	1.82%	2.14%	1.98%	1.84%	2.49%
招商银行	0.94%	1.37%	1.21%	1.26%	1.14%	1.30%
中信银行	-	-	1.41%	1.30%	0.98%	1.74%
平安银行	2.77%	2.5%	2.15%	1.20%	1.19%	1.37%
浦发银行	1.47%	1.58%	1.33%	1.27%	1.58%	2.38%

银行信用卡风控场景

根据信用卡风险监管要求，银行在不断完善着信用卡套现侦测体系



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

央行发布《金融科技（FinTech）发展规划（2019-2021年）》

加强金融科技战略部署

强化金融科技合理应用

人工智能

- 统筹资产（数据、算法、算力）、**推动人工智能在金融领域的应用；**
- 探索应用路径和方法，**适应不同场景的业务特征；**
- 加强应用安全潜在风险研判和防范，制定监测预警机制与监管规则；
- 建立人工智能金融应用法律法规、伦理规范和政策体系；

增强金融风险技防能力

提升金融业务风险防范能力

- **完善风险防控体系：**运用机器学习等提高风险识别效能；
- **健全风险监测和早期干预：**构建动态评估体系和分级分类风控规则，将智能风控嵌入业务流程，提高防控的及时性；
- 组织建设统一金融风控平台
- 构建风险联防联控机制

提赋能金融服务提质增效

加大金融审慎监管力度

夯实金融科技基础支撑

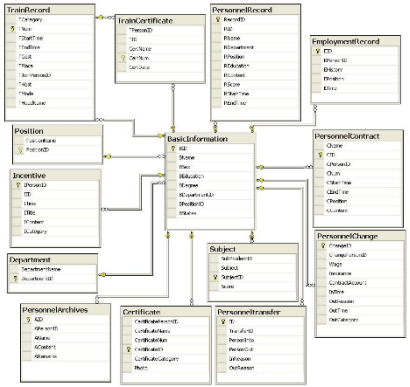
© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图数据库与图计算

图数据库与图计算是一种极为简洁的拓扑结构、可以表达任意维度的数据间形成的高维、复杂的网络拓扑构造的体系架构。

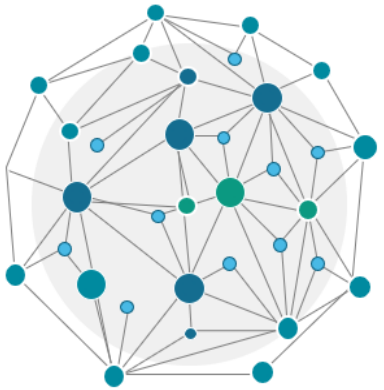
关系数据库

- **关系模型**是二维表模型，通过二维表关联所构成的一个数据组织；
- 二维表结构抽象层次更高；
- 对于网状等复杂结构，关系模型需要多度查询，开发效率低、运行耗时。



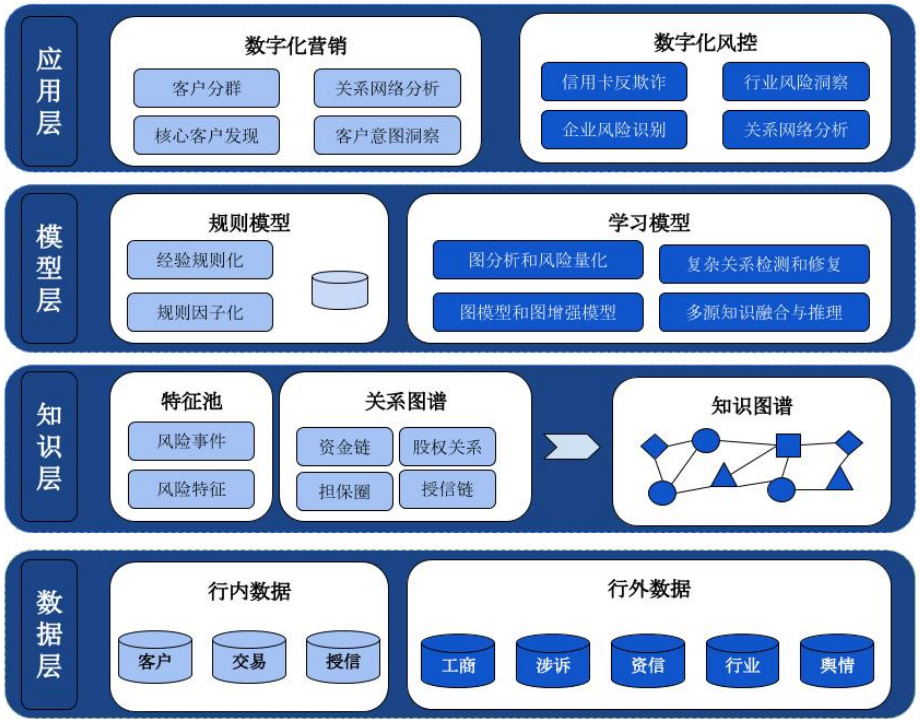
图数据库

- **图模型**是一种自然关系的表达，通过带属性的点和边描述；
- 图基于现实世界的实体和关系建模，更为直接，易解释；
- 特别适合多样的复杂关系组织和查询。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图数据库与图计算



图计算在风控、营销、大数据分析等领域的快速发展

针对数据关联性

图计算是大数据向快数据、深数据发展的必然趋势，它主要解决了数据之间的关联关系问题。

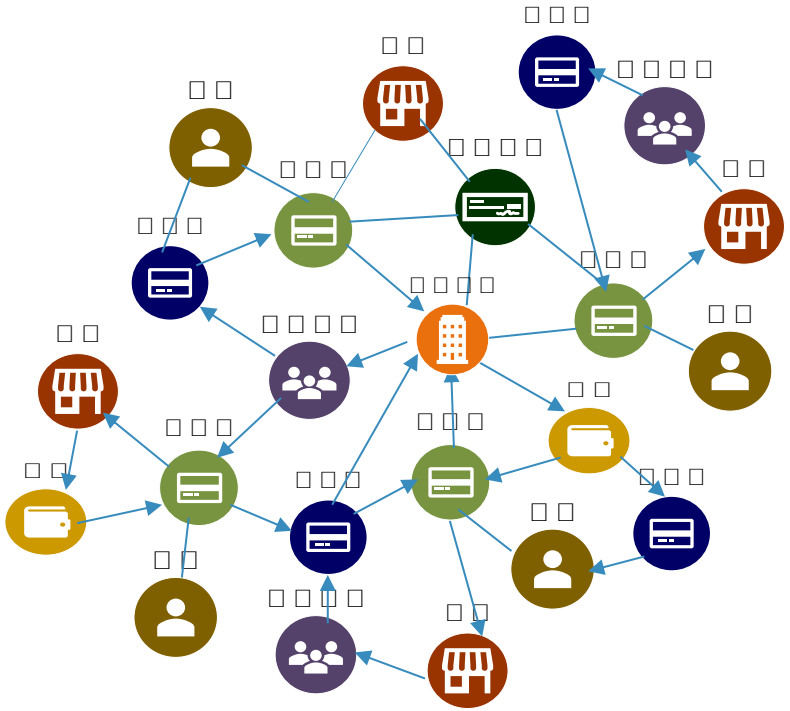
图的可解释性

图提供了一种通用表示范型，点和边的组合解释性广泛。图计算和人工智能的结合能够解决 AI 白盒化的问题。

增强机器学习

机器学习在某些应用上面临特征不够丰富和有效的问题，通过从图上抽取图特征可以补充大量特征，提供模型性能。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |



信用卡资金流向图

基于银行海量资金流水建立的异质信息网络。针对信用卡侧业务包括消费信贷，循环信贷和预借现金。

节点表示

- 客户层（对公、对私客户、信用卡商户）
- 卡片层（信用卡、借记卡、行外卡）
- 业务层（转账、消费、还款、分期）

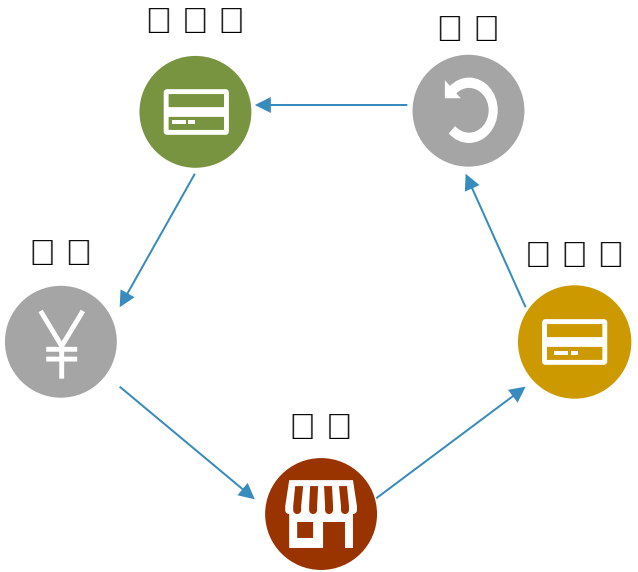
关系边表示

- 客户与卡的从属关系
- 对公客户与商户的从属关系
- 对私客户与商户的法人关系
- 业务关系（交易关系、还款关系、分期关系、消费关系等）

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

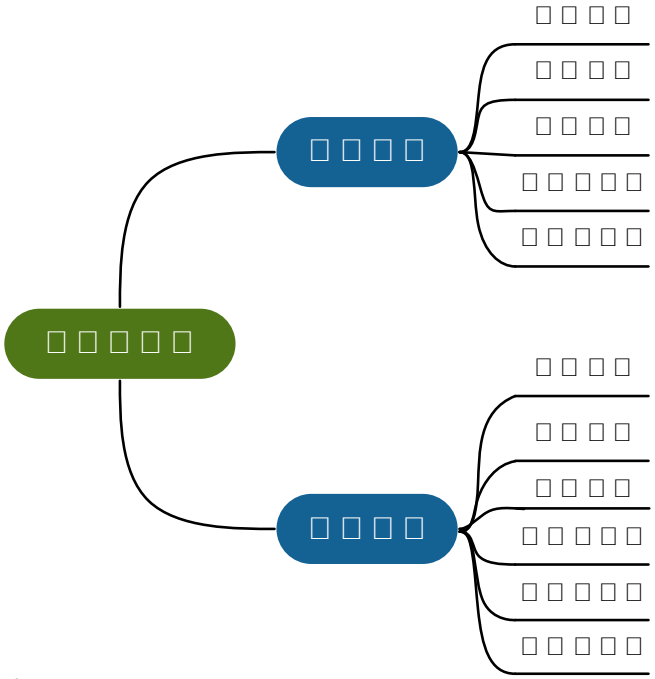
反套现图规则

【定义】在图上寻找信用卡资金流入与流出的闭环，其中资金流入指信用卡消费，资金流出包含借记卡转账和信用卡还款。



【分类】**循环信贷**特点：普通的信用额度，交易频率高；

消费信贷特点：专项分期业务，交易金额大。



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

银行信用卡反套现

- 套现灰黑产高速发展（套现手法多变）
- 银行沉淀的案例和经验不多
- 传统规则方法（覆盖率低和计算效率低）

- **标注数据困难**

无监督方法过滤大量正常样本，从而加速标注。

- **新套现模式**

从资金流水中侦测新的异常交易模式。

- **模型性能**

调整密度参数在召回率和误警率之间进行平衡。

- **可解释性**

图具有较强的表示性，使得结果可解释性较好。

- **海量流水**

图数据库处理海量数据复杂的拓扑结构关系。

- **黑样本稀缺**

采用无监督的方法，对黑样本的需求不高。

- **完美账单**

密集子图进行大数据统计学习，不依赖领域经验。

- **数据孤岛**

无法解决，银行之间的交易数据不共享。

密集子图方法

- 无监督的异常检测方法
- 图表示数据的内在关系
- 常用于反欺诈领域（比如，反洗钱）

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

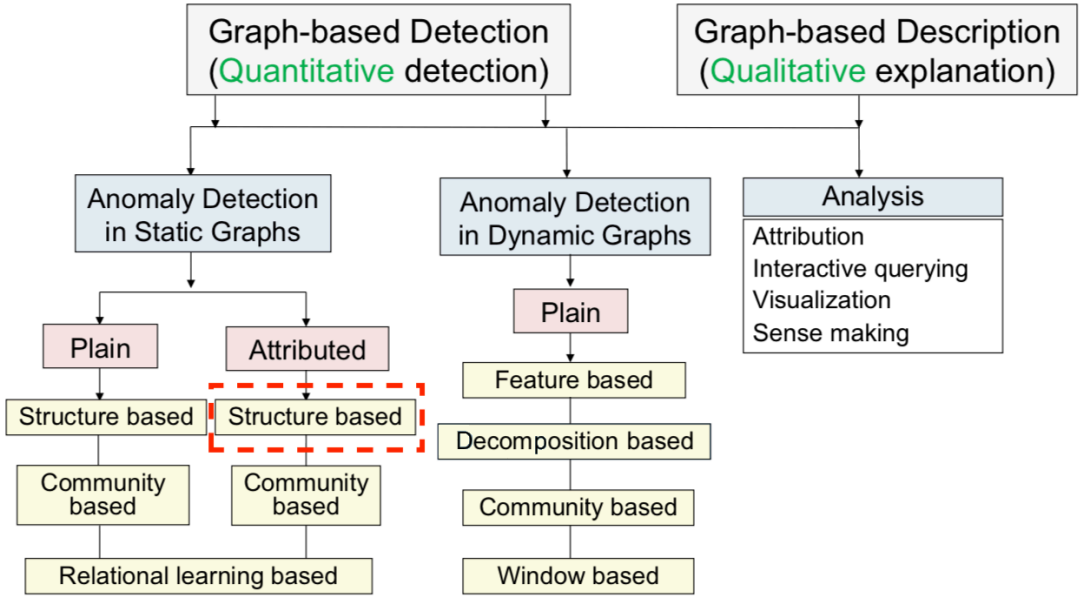
基于图的异常检测

采用图异常侦测技术对由银行资金流水所构建的特定网络进行挖掘，找到信用卡和商户所构成的疑似套现团伙。

【问题】识别套现信用卡和商户

【数据】信用卡交易和借记卡转账

【技术】密集子图异常检测



在图异常检测技术中，密集子图属于定量检测方法，主要针对属性图，基于结构的方法，通过建立基于联通和属性的测度函数侦测偏离正常模式的子图结构。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

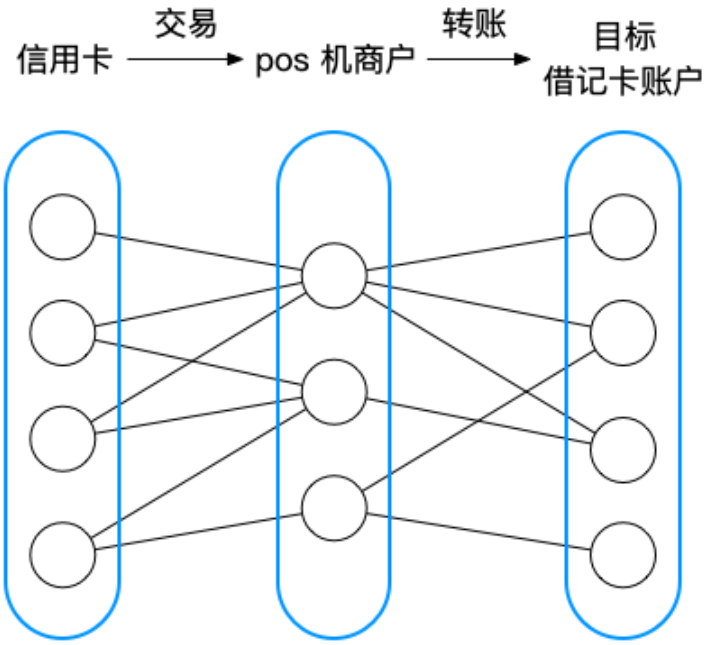
一般问题描述

1. 给定一个无向图
2. 设计一个密度函数
3. 从图上找一组节点使密度最大化

- given an undirected graph $G = (V, E)$
  and a density measure $f : 2^V \to \mathbb{R}$
- find set of vertices $S \subseteq V$
  that maximizes $f(S)$

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

构造图模型



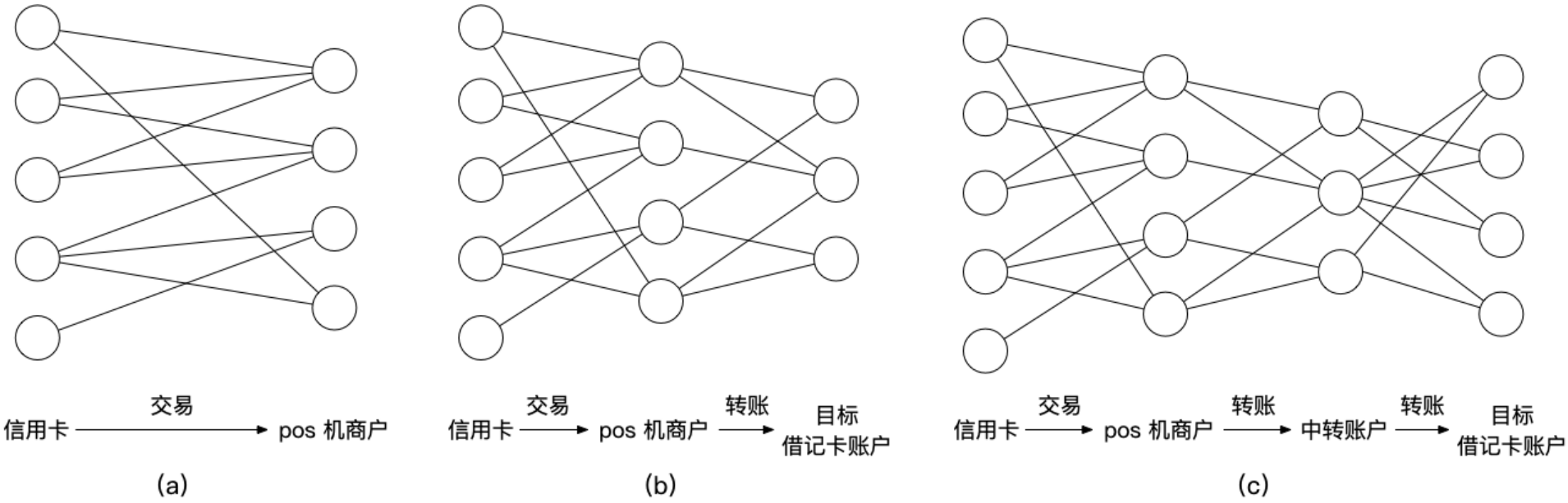
【多部图】将节点集分为多个两两互不相交的节点子集，每条边的两端分属于两个互不相交的子集，子集内节点不连接。

银行资金流向 k 部图

符号	说明
$G = (V, \mathcal{E})$	银行资金流向图（k 部图）
$V = X \cup Y_1 \cup \ldots \cup Y_{k-2} \cup Z$	节点集合，节点表示账户
$\mathcal{E} = \mathcal{E}_1 \cup \ldots \cup \mathcal{E}_{k-1}$	边集合，边表示交易关系
X	信用卡集合
Y_1	pos 机商户结算账户集合
$Y_2 \ldots Y_{k-2}$	资金中转账户集合
Z	资金流转至最终的借记卡账户
$\mathcal{E}_1$	信用卡与 pos 机商户的交易关系
$\mathcal{E}_2 \ldots \mathcal{E}_{k-1}$	pos 机商户结算账户、中间级资金中转账户与最终级资金流出账户之间的转账关系。
k	参数 $k \geq 2$ 表示图分成几个部分

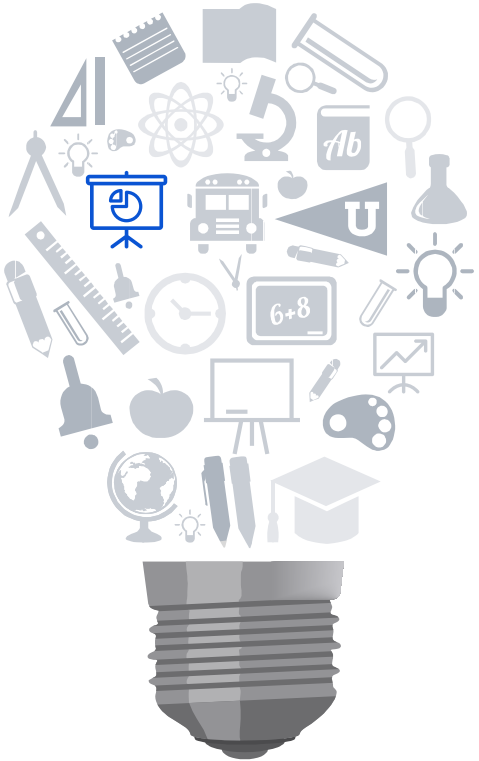
© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

构造图模型



不同 k 值对应不同的套现过程：（a）当 $k=2$ 时，图表示为二部图，pos 机商户结算账户就是资金目标账户；（b）当 $k=3$ 时，商户资金流向目标账户，没有资金中转；（c）当 $k=4$ 时，资金从商户流出会经过中转行户再流向目标账户。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |



子图结构

从信用卡账户通过 pos 机结算商户，经过中转到银行个人账户构成大额资金流动链。

问题描述

设计一个衡量套现疑似行为的密度函数，用于找出资金流向图上使密度函数最大化的子图。

套现行为特征

大资金流

信用卡套现往往产生高额的资金流动，从而构成密集的子图结构。

流转成本

资金流动过程中的中间账户的流入资金量和流出资金量相当。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

实验准备

某月行内全量商户的信用卡交易数据以及结算账户转账数据。

信用卡数	300 万
商户数	40 万
流出账户数	700 万
交易汇总边数	300 万
转账汇总边数	700 万

下发排查

- Tiger 运行输出嫌疑度排名前十的子图
- 商户关闭率35%，处置率75%

案例学习

对高风险的商户进行案例分析，总结已有规则体系未覆盖到的套现模式。

结果分析

从三个方面（商户嫌疑，外部和内部关联性）对结果进行后验分析，

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

外部关联性

信用卡套现客户与资金流出账户之间具有较强的关联性。比如，资金流出账户可能是套现客户的同名行外账户。

内部关联性

团伙信用卡套现团伙内部有较强的关联性，通过银行现有的数据可以发现这些人员是家人或者同事关系。

商户可疑度

商户套现指标，这里主要使用借贷比。信用卡套现商户的日常交易中，贷记卡占比相对高。甚至全部是贷记卡交易。

针对 61 家嫌疑商户和 81 组嫌疑案例，通过规则（比如，名字或地址匹配）进行评估。不同组合的风险评级如下表。

外部关联性	内部关联性	商户可疑性	套现风险
√	√	√	高
√	√	×	高
√	×	√	高
√	×	×	中
×	√	√	中
×	√	×	低
×	×	√	低
×	×	×	低

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

图分析技术致力于股权预警

锻造凝炼IT服务 助推用户事业发展
Perfecting IT service and favoring clients 'success

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

集团客户股权关系变动预警

目的

建立股权变动预警模型，有效识别集团企业股权控制链，提高对集团客户授信风险的控制能力。

算法

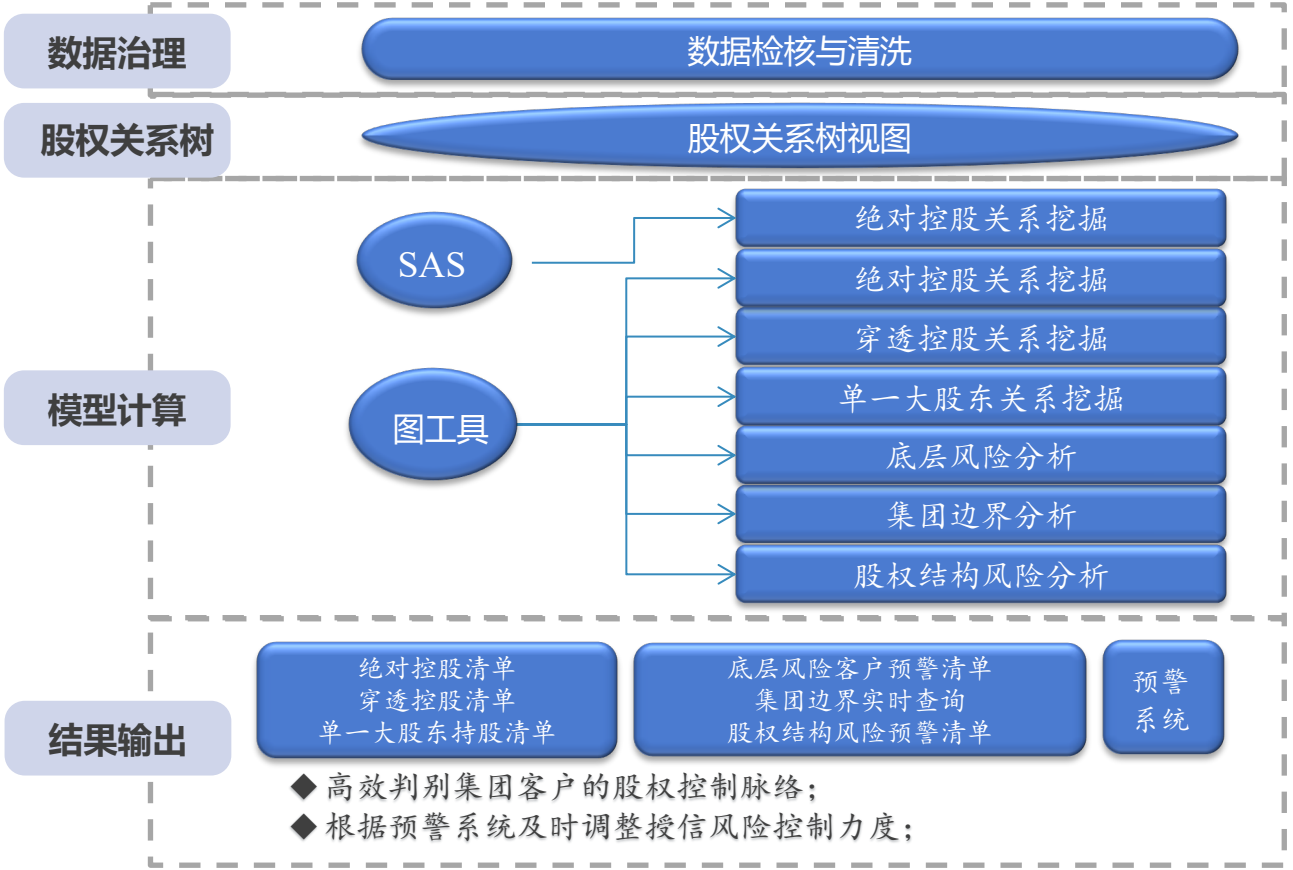
使用图数据库技术与大数据分析结合。

优势

1. **业务逻辑与技术逻辑**的高适配度；
2. 业务人员能够**高效判别集团客户的股权控制脉络**，根据变动**预警提示**，**及时调整授信**风险控制力度；
3. 图链路的索引和累加使得模型计算**响应更加及时快速**；
4. **可视化**特征便于业务和技术人员共同探索和分析**股权关系中存在的风险**；

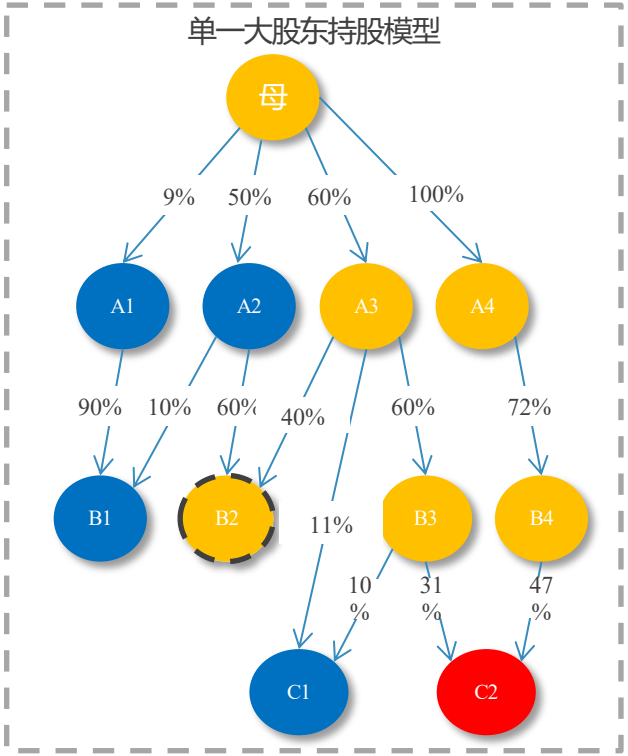
© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

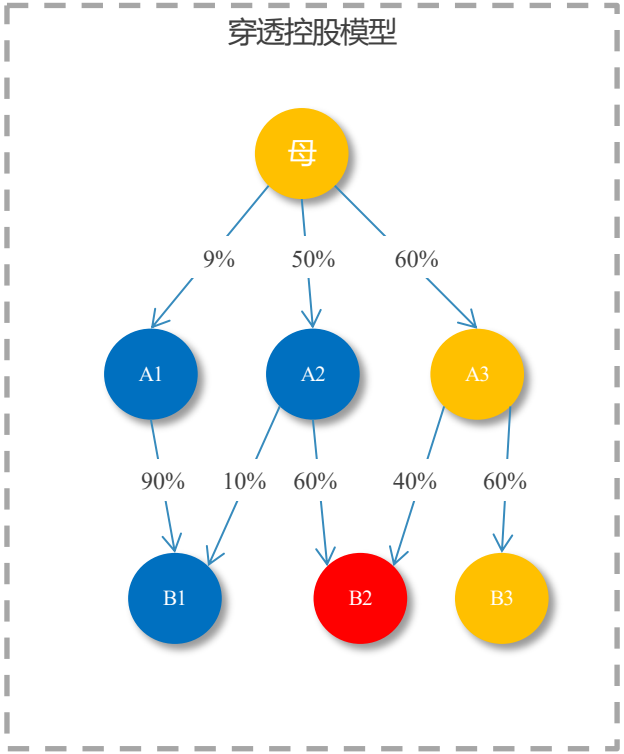
集团客户股权关系变动预警



© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

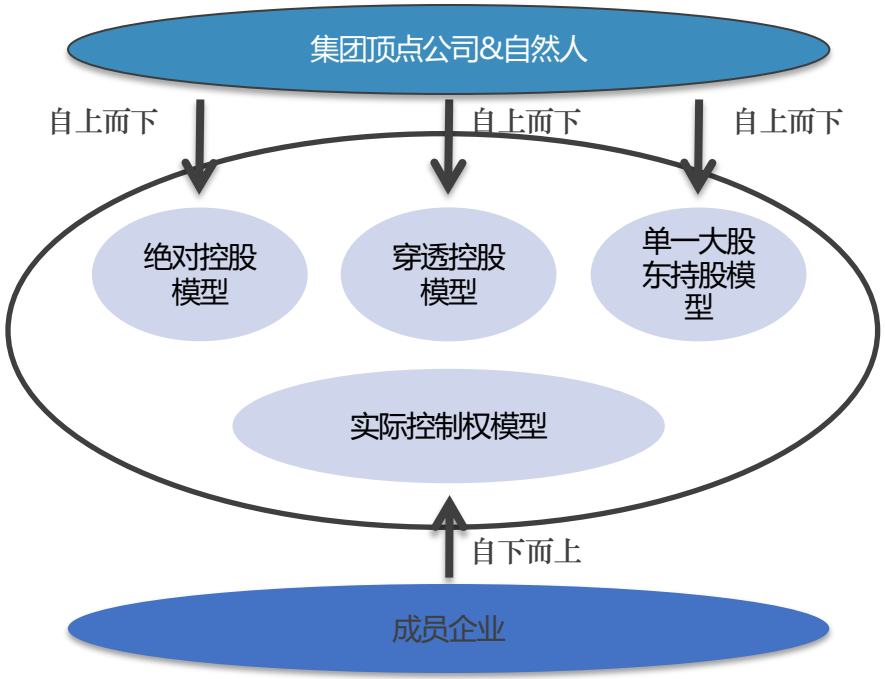
集团客户股权关系变动预警





© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

央企及全球客户股权关系变动预警



模型的直接功能

1. **自上而下模型**是为了探索央企大客户集团具有不同控制权的目标子公司，从而基于不同控制权，在业务场景中考虑对子公司采用不同决策。同时，可以探索到集团控制边界；

2. **自下而上模型**是为了探索出目标成员公司的实际控制集团或实际控制人，用于决策成员公司控制权所属发生变化的情况。

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

央企及全球客户股权关系变动预警

模型的扩展应用

集团关联关系识别

- **关联交易侦查**

结合企业银行流水来查找关联企业

- **战略投资关系发现**

战略投资持股平台查询结合供应链数字产品来查找企业客户的亲密关联上下游

- **担保关系发现**

结合担保圈数字产品来确定集团客户的综合授信风险

- **顶层风险排查**

向上追溯至实际控制的自然人和高管，再结合个人信用信息，对顶层自然人的家族内成员及成员控股企业进行风险排查

集团风险传导模型

- **集团风险传染模型**

检验集团内一家公司当年出现风险是否会在次年传染到集团成员公司

- **集团风险传染路径模型**

检验集团风险会通过什么渠道传染

- **集团风险管理与风险传染模型**

检验何种风险管理措施能够缓解或加剧风险传染

股权健康体检

- **股权结构维度分析**

包括股权集中度、股权制衡度等维度分析企业股权健康程度

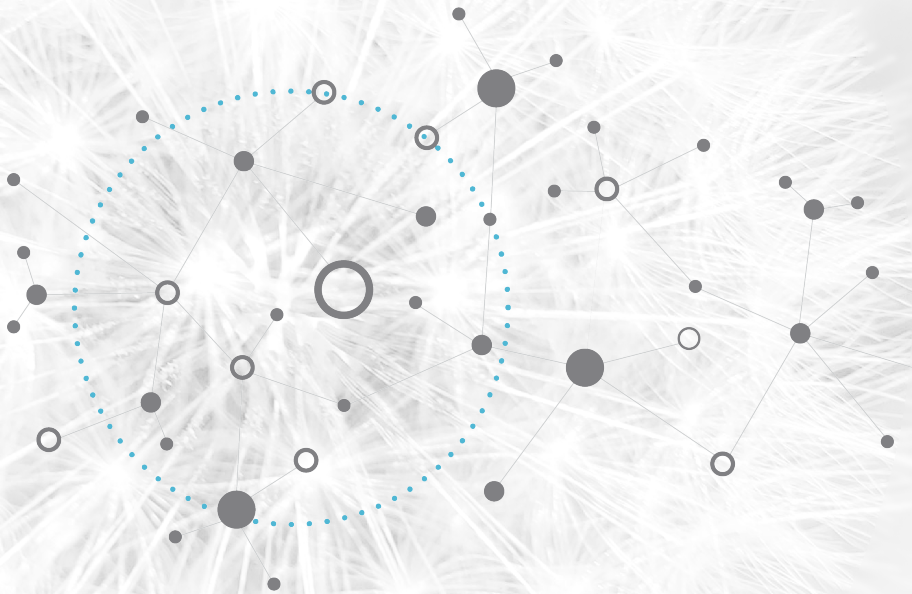
- **股权结构行业分析**

与同行业和全行业对比分析股权结构的健康程度

- **股权健康画像**

提供分析报告，从控制权，股权架构，股权激励，股权税筹，股权融资等多方面构建股权健康程度画像

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM | CONFIDENTIAL INFORMATION |

TigerGraph

END
谢谢大家