

图引擎服务

# 用户指南

文档版本 07

发布日期 2018-10-23



### 版权所有 © 华为技术有限公司 2018。 保留一切权利。

非经本公司书面许可,任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部,并不得以任何形式传播。

# 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。 本文档提及的其他所有商标或注册商标,由各自的所有人拥有。

# 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束,本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定,华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因,本文档内容会不定期进行更新。除非另有约定,本文档仅作为使用指导,本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为技术有限公司

地址: 深圳市龙岗区坂田华为总部办公楼 邮编: 518129

网址:<a href="http://www.huawei.com">http://www.huawei.com</a>客户服务邮箱:<a href="mailto:support@huawei.com">support@huawei.com</a>

客户服务电话: 4008302118

# 目录

1 简介	
2 入门	3
2.1 如何查询和分析图	
3 管理	10
3.1 图数据的格式	10
3.2 图管理	14
3.2.1 概览	14
3.2.2 创建图	
3.2.3 启动图	
3.2.4 停止图	
3.2.5 删除图	18
3.2.6 升级图	18
3.2.7 导出图	18
3.2.8 绑定 EIP	
3.2.9 解绑 EIP	20
3.3 备份与载入	20
3.3.1 备份图	20
3.3.2 载入图	
3.3.3 删除备份	23
3.4 元数据管理	23
4 访问和使用图引擎	27
4.1 访问图引擎编辑器	27
4.2 查询编辑器页面介绍	
4.3 Gremlin 查询	35
4.4 使用算法分析图	37
4.5 在绘图区分析图	39
4.6 条件过滤	41
4.7 查看运行记录	42
4.8 查看查询结果	43
5 算法参考	44
5.1 算法一览表	44
5.2 PageRank 算法	46

5.3 PersonalRank 算法	47
5.4 k 核算法(k-core)	48
5.5 k 跳算法 (k-hop)	48
5.6 最短路径(Shortest Path)	49
5.7 全最短路(All Shortest Paths)	50
5.8 关联路径(n-Paths)	51
5.9 紧密中心度(Closeness Centrality)	52
5.10 标签传播(Label Propagation)	52
5.11 Louvain 算法	53
5.12 关联预测(Link Prediction)	54
5.13 Node2vec 算法	55
5.14 实时推荐(Real-time Recommendation)	56
5.15 共同邻居(Common Neighbors)	57
5.16 单源最短路(SSSP)	58
5.17 联通分量(Connected Component)	58
5.18 度数关联度(Degree Correlation)	59
5.19 三角计数(Triangle Count)	59
5.20 聚类系数 (Cluster Coeffcient)	60

1 简介

图引擎服务(Graph Engine Service,简称GES),是针对以"关系"为基础的"图"结构数据,进行查询、分析的服务。广泛应用于社交关系分析、推荐、精准营销、舆情及社会化聆听、信息传播、防欺诈等具有丰富关系数据的场景。

# 产品优势

# ● 大规模

高效的数据组织,让您更有效的对百亿节点千亿边规模的数据进行查询与分析。

### ● 高性能

深度优化的分布式图形计算引擎,为您提供高并发、秒级多跳的实时查询能力。

# ● 查询分析一体

查询分析一体化,提供丰富的图分析算法,为关系分析、路径规划、精准营销等业务提供多样的分析能力。

### ● 算法丰富

提供丰富的基础图算法、图挖掘算法和图指标算法,供不同领域客户不同场景使用。

### ● 简单易用

提供向导式、简单易用的可视化分析界面,所见即所得;支持Gremlin查询语言,兼容您的使用习惯。

# 功能介绍

### 支持大规模图的查询和计算

支持百亿节点千亿边超大规模图,支持点查询、边查询和属性查询,秒级响应;提供丰富的基础图算法、图挖掘算法和图指标算法,满足各场景需求。

### ● 实例管理

支持不同规格不同数据的图引擎实例管理,用户按需收费。

### ● 标准图查询语言

支持标准图查询语言Gremlin,快速上手。

# ● 可视化

提供向导式探索环境,即学即用;支持查询结果可视化,所见即所得。

### ● 开放API

支持开放API,让二次开发和项目迁移更容易。

# $2_{\lambda i}$

# 2.1 如何查询和分析图

本章节介绍如何使用图引擎服务(GES)对示例图数据进行查询和分析的过程。分析的流程如下所示:

- 1. 创建图
- 2. 访问图
- 3. 查询和分析图
- 4. 查看分析结果

# 创建图

**步骤1** 进入图引擎服务管理控制台,单击"创建图"。

步骤2 在创建图页面,选择"区域"和"可用区"。

"区域":集群工作区域,可在左上方菜单栏选择,目前工作区域仅支持华北-北京一。

"可用区":可用区是使用独立电源和网络资源的物理区域。通过内部网络互联,再以物理方式进行隔离,提高了应用程序的可用性。建议您在不同的可用区下创建集群。选择集群工作区域下关联的可用区。

步骤3 填写"图名称"。例如, demo。

图名称的长度需满足如下条件。

- 4~64个字符, 且以字母开头。
- 字母不区分大小写。
- 只能包含字母、数字或下划线,不能包含其他特殊字符。

步骤4 "GES软件版本":显示当前的版本号。

步骤5 选择网络信息,包含"虚拟私有云"、"子网"和"安全组"。

● "虚拟私有云": VPC即虚拟私有云,是通过逻辑方式进行网络隔离,提供安全、隔离的网络环境。

选择需要创建集群的VPC,单击"查看虚拟私有云",可进入VPC服务查看已创建的VPC名称和ID。

若没有VPC,请先创建一个新的VPC。

● "子网":通过子网提供与其他网络隔离的、可以独享的网络资源,以提高网络安全。子网在可用区内才会生效。

选择需要创建集群的子网,可进入VPC服务查看VPC下已创建的子网名称和ID。

● "安全组":安全组是一个逻辑上的分组,为同一个VPC内具有相同安全保护需求并相互信任的弹性云服务器提供访问策略。单击"查看安全组"可了解安全组详情。

**步骤6** 选择公网访问的方式,包含"暂不使用"、"现在购买"和"使用已有",用户根据自身需求选择。

- "暂不使用":不使用弹性IP的图实例并且不能与互联网互通,仅可通过私有网络中已部署的弹性云服务器连接当前图实例使用。
- "现在购买":服务将自动为图实例分配独享带宽的弹性IP,以支持通过弹性IP 从互联网对图实例进行访问。同时,服务将使用租户权限在当前项目下自动创建 以ges\_agency\_default为前缀的授权委托以支持弹性IP的绑定处理。
- "使用已有":为图实例绑定指定的弹性IP,通过弹性IP可以从互联网对图实例进行访问。

# 步骤7 填写图相关参数。

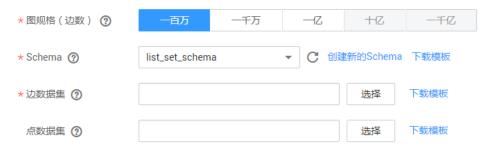
只支持从OBS中选择Schema、边数据集、点数据集等数据,建议您提前将数据文件存储至OBS中,导入OBS的操作请参见OBS《控制台指南》。图数据(包含Schema、边数据集、点数据集)需按照模板中的格式导入。模板中已包含一份电影数据的数据,可下载后直接导入。

- 图规格(边数):以边为单位,可选择一百万、一千万、一亿、十亿和千亿。根据用户当前的配额,系统会提示可创建的图的数量以及边数。
- Schema: 图的Schema信息。如果没有Schema,可单击"创建新的Schema"创建,创建Schema的操作指导请参考元数据管理。
- 边数据集: 所有描述边的结构、标签和属性的信息构成了边数据集。
- 点数据集: 所有描述点的ID、标签和属性的信息构成了点数据集。

### □ 说明

点数据集和边数据集当前仅支持英文路径和文件夹格式。

### 图 2-1 创建图的相关配置



步骤8 单击"立即创建"。

**步骤9** 在"订单确认"页面,确认信息后,单击"提交订单"开始创建图。

步骤10 系统将自动跳转到"图管理"页面,您创建的图将展现在图列表中。

### ----结束

# 访问图

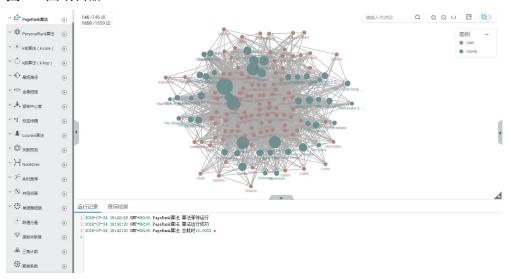
步骤1 在图管理页面,选中需访问的图,即demo。在操作列单击"访问"。

### 图 2-2 图管理列表



**步骤2** 页面进入图编辑器页面,您可以在此页面对当前图进行查询分析。如图2-3所示。

### 图 2-3 图编辑器



----结束

# 查询和分析图

- 通过Gremlin命令查询。
  - a. 在页面上方的Gremlin输入框中,输入查询命令,如g.V().limit(100)。

# □ 说明

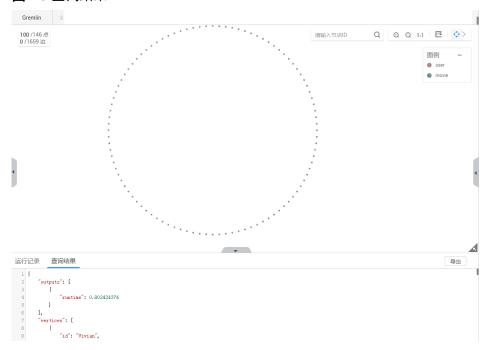
为了避免返回数据过大导致系统查询耗时过长,强烈建议您增加limit参数;并且,limit参数设置在1000以内展示效果更佳。

### 图 2-4 Gremlin 查询



b. 按"回车"执行输入的Gremlin命令。查询结果分别在绘图区和结果区显示。

图 2-5 查询结果



# ● 通过选取算法、配置参数分析。

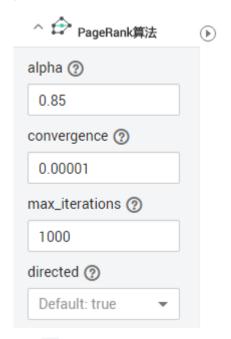
a. 在图编辑器页面左侧,展示当前系统支持的算法列表。从中选取需要分析的 算法。

图 2-6 算法列表



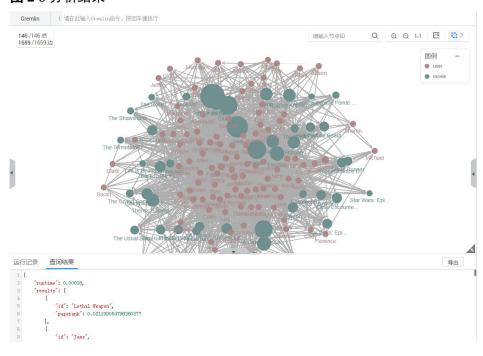
b. 选中需要使用的算法,然后单击 一展开算法参数配置,根据提示范围输入对应的取值。以PageRank算法为例,alpha表示权重系数,取值0.85; covergence 为收敛系数,取值0.00001; max\_iterations为最大迭代次数,取值1000; directed为是否考虑边的方向,默认为true。

# 图 2-7 设置算法参数



c. 单击 D 执行算法。分析结果分别在绘图区和结果区显示。

# 图 2-8 分析结果



# 查看分析结果

您可以在结果展示区查看运行记录和查询结果,或者单击右上角"导出"按钮,下载分析结果。

# **3** 管理

# 3.1 图数据的格式

在导入图数据之前,您需要了解GES中支持的图数据格式。GES支持载入具有标准csv格式的原始图数据,如果您的原始数据并不符合指定的格式,需要将数据整理为GES支持的格式。

GES支持的图数据格式包含三部分:点文件、边文件以及Schema。点文件和边文件分别存放了点数据和边数据,Schema用于描述点文件和边文件中的数据格式。

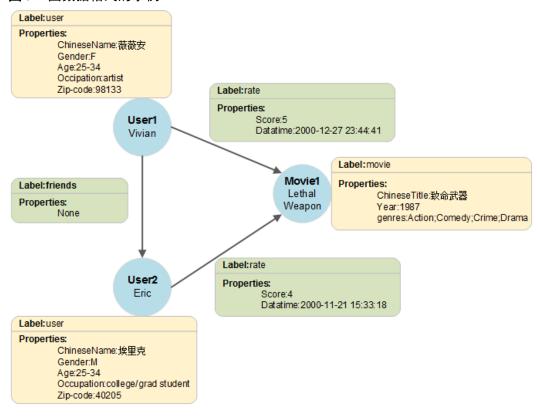
# 概念说明

GES是基于属性图模型导入图数据,因此,您需要了解属性图(Property graph)的基本概念。

1个属性图是由点、边、标签(Label)和属性(Property)组成的有向图。点也称作节点(Node),边也称作关系(Relationship)。点和关系是最重要的实体,Schema是用于描述点或边的属性信息,Schema由多个标签组成,每个标签又由1个或多个属性组成。为点或边设置标签,那么拥有相同标签的点属于一个分组,是一个集合。点和边只能设置一个标签。

如下示例中,此图数据由3个点和3条边组成,Vivian、Eric和Lethal Weapon表示点,(Vivian,Eric)、(Vivian,Lethal Weapon)和(Eric、Lethal Weapon)表示边,user和movie表示点的类型(标签),rate和friends表示关系类型(标签)。

# 图 3-1 图数据格式的示例



### Schema

GES的Schema用于定义点和边的属性信息,为XML格式的文件。

在Schema中包含了标签(Label)和属性(Property)。

### ● 标签 (Label)

Label是属性(Property)的集合,描述了一个点或边拥有的所有Property的数据格式。需要注意的是,在不同的Label中,如果定义了相同的Property name,则定义的cardinality和dataType需要跟已定义的一致。

### ● 属性 (Property)

Property指的是单个属性的数据格式,包含3个字段。

- Property name: 属性的名称,值得注意的是,同一个label中不能包含相同的 Property name。
- cardinality:数据的复合类型,取值为single、list和set。single是单值类型,表示该Property的数据是一个单值,如一个数字或一个字符串。list和set是集合类型,表示该Property的数据由多个值组成,不同的值用分号分隔。

### □说明

list和set类型不支持char array类型。

- dataType: 数据类型,支持的的数据类型如下表所示。

# 表 3-1 支持的数据类型

类型	描述	
string	不定长字符串类型。	
char	字符。	
char array	定长字符串(需指定长度)。	
float	float浮点类型(32位浮点)。	
double	double浮点类型(64位浮点)。	
bool	bool类型,取值(0/1)或者(true/false)。	
long	长整数类型(取值范围-2^63 to 2^63-1)。	
int	整数类型(取值范围-2^31 to 2^31-1)。	
date	日期,目前支持格式如下所示:	
	YYYY-MM-DD HH:MM:SS	
	YYYY/MM/DD HH:MM:SS	
	DD.MM.YY HH:MM:SS	
	YYYY-MM-DD	
	MM/DD	
	HH:MM	
enum	枚举类型(需指定枚举类型的个数以及每个枚举 值)。	

Schema的示例如下所示:

### 图 3-2 Schema 示例

```
<?xm1 version="1.0" encoding="ISO-8859-1"?>
                        <PMML version="3.0"
                         xmlns="http://www.dmg.org/PMML-3-0"
                         xmlns:xsi="http://www.w3.org/2001/XMLSchema_instance">
                           <label name="default">
           Label:
                           </label>
           default
                           <label name="movie">
                             properties>
                               cproperty name="ChineseTitle" cardinality="single" dataType="string"/>
                              cproperty name="Year" cardinality="single" dataType="int"/>
           Label:
                               cproperty name="Genres" cardinality="set" data Type="string"/>
           movie
                            properties>
                           </label>
                           <label name="user">
                             coroperties>
                               cproperty name="Chines eName" cardinality="single" data Type="string" />
                               cproperty name="Gender" cardinality="single" dataType="enum" typeNameCount="2"
                                                typeName1="F" typeName2="M"/>
Labels-
                               cproperty name="age" cardinality="single" dataType="enum" typeNameCount="7"
                                                typeName1="Under 18" typeName2="18-24" typeName3="25-34"
           Label:
                                                typeName4="35-44" typeName5="45-49"
           user
                                                typeName6="50-55" typeName7="56+"/>
                              cproperty name="Zip-code" cardinality="single" dataType="char array" maxDataSize="12"/>
                             </properties>
                           </label>
                           <label name="rate">
                             properties>
                               cproperty name="Rating" cardinality="single" data Type="int" />
            Labe
                               cproperty name="Datetime" cardinality="single" dataType="Date"/>
            rate
                             properties>
                           </label>
                         </lab els>
                         </PMMT>
```

# 点文件

点文件罗列了各个点的数据信息。一行为一个点的数据。格式如下所示,id是点数据的唯一标识。

```
id, label, property 1, property 2, property 3, ...
```

### 示例:

```
Lethal Weapon, movie, 致命武器, 1987, Action; Comedy; Crime; Drama
Vivian, user, 薇薇安, F, 25-34, artist, 98133
Eric, user, 埃里克, M, 18-24, college/grad student, 40205
```

# 边文件

边文件罗列了各个边的数据信息,一行为一条边的数据。GES中图规格是以边的数量进行定义的,如一百万边。格式如下所示,id 1、id 2一条边的两个端点的id。

```
id 1, id 2, label, property 1, property 2, ...
```

示例:

Vivian, Lethal Weapon, rate, 5, 2000-12-27 23:44:41 Eric, Lethal Weapon, rate, 4, 2000-11-21 15:33:18 Vivian, Eric, friends

# 3.2 图管理

# 3.2.1 概览

系统总览页面展现了图状态、图规格和图备份信息,可以帮助您快速了解已有图的信息。

图状态展现了处于6种不同状态的图数量。

# 表 3-2 状态说明

状态	说明
运行中	表示正在运行的图,处于该状态的图可访问。
准备中	表示正在创建的图,正在加载数据。
失败	表示创建失败的图。
己停止	表示已停止的图,处于该状态的图不可继续访问。停止状态的图可以重新启动。
停止中	表示正在停止的图。
故障	表示出现故障的图,处于该状态的图不可访问。
己冻结	表示用户的账户和资源被冻结。详情可参考 <b>账号中心</b> 。 <b>说明</b> 用户账户被冻结后,只能做删除操作。

当前系统支持5种规格,展示了不同规格的图数量。

# 表 3-3 规格介绍

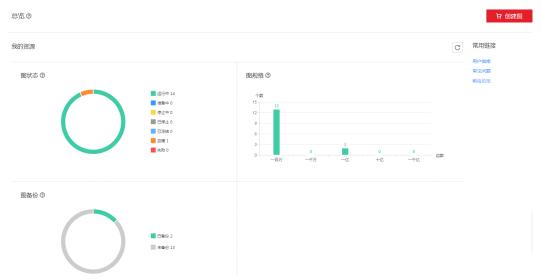
No a wild take	
规格	说明
一百万	表示允许创建的图的边数不超过一百万。
一千万	表示允许创建的图的边数不超过一千万。
一亿	表示允许创建的图的边数不超过一亿。
十亿	表示允许创建的图的边数不超过十亿。
一千亿	表示允许创建的图的边数不超过千亿。

为防止数据丢失,系统支持对图数据进行备份。当前页面展示了已备份的图数量,以 及未备份的图数量。

# 表 3-4 备份状态介绍

备份状态	说明
已备份	己完成备份的图数量。
未备份	未备份的图数量。

# 图 3-3 总览信息



# 3.2.2 创建图

# 操作场景

在开始使用图引擎服务时,您必须先创建一个图。

# 操作步骤

步骤1 登录图引擎服务管理控制台,单击"创建图"。

步骤2 在创建图页面,选择"区域"和"可用区"。

"区域":集群工作区域在左上方菜单栏选择,目前工作区域仅支持华北-北京一。

"可用区":可用区是使用独立电源和网络资源的物理区域。通过内部网络互联,再以物理方式进行隔离,提高了应用程序的可用性。建议您在不同的可用区下创建集群。选择集群工作区域下关联的可用区。

# 步骤3 填写图名称。

图名称的长度需满足如下条件。

- 4~64个字符,且以字母开头。
- 字母不区分大小写。
- 只能包含字母、数字或下划线,不能包含其他特殊字符。

### 图 3-4 基本信息



步骤4 选择网络信息,包含"虚拟私有云"、"子网"和"安全组"。

● "虚拟私有云": VPC即虚拟私有云,是通过逻辑方式进行网络隔离,提供安全、隔离的网络环境。

选择需要创建集群的VPC,单击"查看虚拟私有云"进入VPC服务查看已创建的 VPC名称和ID。如果没有VPC,需要创建一个新的VPC。

● "子网":通过子网提供与其他网络隔离的、可以独享的网络资源,以提高网络安全。

选择需要创建集群的子网,可进入VPC服务查看VPC下已创建的子网名称和ID。

● "安全组":安全组是一个逻辑上的分组,为同一个VPC内具有相同安全保护需求并相互信任的弹性云服务器提供访问策略。单击查看安全组可了解安全组详情。

**步骤5** 选择公网访问的方式,包含"暂不使用"、"现在购买"和"使用已有",用户根据自身需求选择。

- "暂不使用":不使用弹性IP的图实例并且不能与互联网互通,仅可通过私有网络中已部署的弹性云服务器连接当前图实例使用。
- "现在购买":服务将自动为图实例分配独享带宽的弹性IP,以支持通过弹性IP 从互联网对图实例进行访问。同时,服务将使用租户权限在当前项目下自动创建 以ges agency default为前缀的授权委托以支持弹性IP的绑定处理。
- "使用已有":为图实例绑定指定的弹性IP,通过弹性IP可以从互联网对图实例进行访问。

### **步骤6** 填写图相关参数。

只支持从OBS中选择Schema、边数据集、点数据集等数据,建议您提前将数据文件存储至OBS中,导入OBS的操作请参见。图数据(包含Schema、边数据集、点数据集)需按照模板中的格式导入。模板中已包含电影数据的数据,可下载后直接导入。

- 图规格(边数):以边为单位,可选择一百万、一千万、一亿、十亿或千亿。根据用户当前的配额,系统会提示可创建的图的数量以及边数。
- Schema: 图的Schema信息。
- 边数据集:所有描述边的结构、标签和属性的信息构成了边数据集。
- 点数据集:所有描述点的ID、标签和属性的信息构成了点数据集。

### □说明

点数据集和边数据集当前仅支持英文路径和文件夹格式。

### 图 3-5 参数配置



步骤7 单击"立即创建"。

步骤8 在"订单确认"页面,确认信息后,然后单击"提交订单"开始创建图。

步骤9 系统将自动跳转到"图管理"页面,您创建的图将展现在图列表中。

一开始处于"准备中"状态,当创建成功后,状态显示为"运行中"。如果创建失败,创建的图将不在图列表中呈现,请重试或联系华为云客服。

----结束

# 3.2.3 启动图

# 操作场景

图列表中关闭了已停止的图,如需要重新启用,您可以启动图,使得图重新可被访问和分析。

处于"运行中"状态的图,无法执行启动操作。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

**步骤3** 在图管理列表中,选择需启动的图,在"操作"列选择"更多">"启动"。

**步骤4** 等待几分钟, 当启动成功后, 状态将切换为"运行中"。

如果启动失败,请稍后再试,若还是失败,请联系华为云客服。

----结束

# 3.2.4 停止图

# 操作场景

当某张图不需使用时, 您可以停止图。停止图后, 您无法继续访问使用。

□□说明

停止图并不会释放资源

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

**步骤3** 在图管理列表中,选择需停止的图,在"操作"列选择"更多">"停止"。此时,图 状态处于"停止中"。

步骤4 等待几分钟, 当停止成功后, 状态将切换为"停止"。

----结束

# 3.2.5 删除图

# 操作场景

如果用户已完成图数据的分析,您可以删除图以释放资源。

# □说明

图删除后,数据无法恢复,请谨慎操作。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

步骤3 在图管理列表中,选择需删除的图,在"操作"列选择"更多">"删除"。

**步骤4** 在弹出的确认提示框中,单击"确定"完成图删除。

----结束

# 3.2.6 升级图

# 操作场景

由于GES软件版本不断升级,旧版本的图可以通过升级操作升级为新版本的图。

# □说明

当前仅支持1.0.3以上的图版本的升级功能。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

**步骤3** 在图管理列表中,选择需升级的图,在"操作"列选择"更多">"升级"。

**步骤4** 在弹出的确认提示框中,单击"确定"完成图升级。

----结束

# 3.2.7 导出图

# 操作场景

用户如果需要把图数据导出到本地使用,您可以执行导出图的操作。

### 川说明

当前仅支持1.0.3以上的图版本的导出功能,1.0.3版本之前的图版本请升级后再执行此操作。

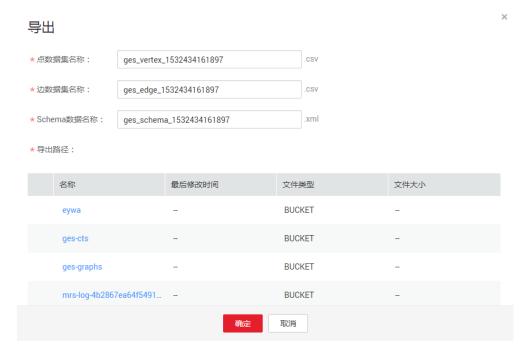
# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

步骤3 在图管理列表中,选择需导出的图,在"操作"列选择"更多">"导出"。

### 图 3-6 导出图



**步骤4** 在弹出的导出图页面下方,选择存储路径。

**步骤5** 单击"确定",系统开始执行导出操作,过一段时间后,可以到选择的OBS路径下查看是否导出成功。

# ----结束

# 3.2.8 绑定 EIP

# 操作场景

如果用户需要通过公网访问GES服务,您可以通过绑定弹性公网IP(简称EIP)来实现。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

步骤3 在图管理列表中,选择需绑定EIP图,在"操作"列选择"更多">"绑定EIP"。

**步骤4** 在弹出的"绑定EIP"页面中,选择可用EIP。若无可用EIP,请单击"创建EIP",购买并创建成功后再选择使用。

# 图 3-7 绑定 EIP



步骤5 单击"确定",完成绑定。

----结束

# 3.2.9 解绑 EIP

# 操作场景

当您无需继续使用EIP,可通过解绑EIP来释放网络资源。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏,选择"图管理"。

步骤3 在图管理列表中,选择需解绑EIP的图,在"操作"列选择"更多">"解绑EIP"。

步骤4 在弹出的确认框中,单击"确定"完成操作。

----结束

# 3.3 备份与载入

# 3.3.1 备份图

# 操作场景

为确保数据安全,您可以选择将图数据备份,以便后续出现故障或错误时,可以使用备份数据进行恢复操作。

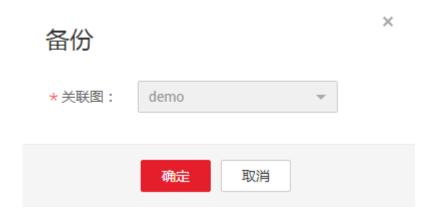
# 操作步骤

备份操作的入口有两个: 图管理页面和备份管理页面。

● 图管理页面入口

- a. 登录图引擎服务管理控制台。
- b. 在左侧导航栏选择"图管理"。
- c. 在图管理列表中,选中需要进行备份的图,在"操作"列单击"备份"。
- d. 在备份页面,确认关联图的名称,单击"确定"开始备份。 当系统中只有一个图时,关联图选项不能更改。

# 图 3-8 备份数据



e. 在左侧导航栏选择"备份管理",您可以在备份管理列表中查看到刚备份的数据。

一开始, "备份状态"显示为"备份中", 请耐心等待几分钟, 当"备份状态"显示为"成功"时,表示备份成功。

# 图 3-9 备份管理

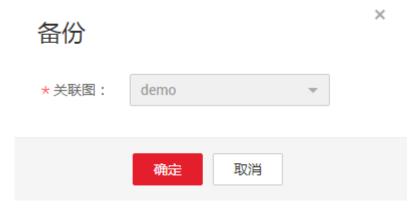


# ● 备份管理页面入口

- a. 登录图引擎服务管理控制台。
- b. 在左侧导航栏选择"备份管理"。
- c. 在"备份管理"页面右上角,单击"备份"。
- d. 在备份页面,选择"关联图",即当前用户下创建的图,单击"确定"开始备份。

当系统中只有一个图时, 关联图选项不能更改。

# 图 3-10 备份数据



e. 您可以在备份管理列表中查看刚备份的数据。

一开始,"备份状态"显示为"备份中",请耐心等待几分钟,当"备份状态"显示为"成功"时,表示备份成功。

# 图 3-11 备份管理



# 3.3.2 载入图

# 操作场景

如果当前编辑的图数据存在问题,需要获取之前备份的数据进行分析时,您可以将备份数据载入,以恢复图数据。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏选择"备份管理"。

步骤3 在"备份管理"页面,选择备份数据对应行,在"操作"列单击"载入"。

**步骤4** 在"载入"页面,确定载入的备份数据信息,勾选"载入操作将覆盖关联图。载入操作启动后,关联图将重新启动。",单击"确定"。

# 图 3-12 载入数据



**步骤**5 页面提示载入成功后,您可以在"图管理"页面,访问关联图,获取载入后的数据。

----结束

# 3.3.3 删除备份

# 操作场景

当备份数据不再使用时,可以根据情况删除备份数据。

# 操作步骤

步骤1 登录图引擎服务管理控制台。

步骤2 在左侧导航栏选择"备份管理"。

步骤3 在备份列表中,选择需要删除的备份数据,在"操作"列,单击"删除"。

步骤4 在弹出的对话框中,单击"确定"删除数据。

数据删除后无法恢复,请谨慎操作。

----结束

# 3.4 元数据管理

元数据指图数据中的Schema。在图引擎服务的管理控制台,您可以从本地或OBS导入元数据、创建元数据、编辑元数据或删除元数据。

# 导入元数据

如果您已有元数据文件,您可以选择将元数据文件导入到图引擎服务中,以便后续创建图使用。

1. 在图引擎管理控制台,单击左侧导航栏的"元数据管理"。

- 2. 在"元数据管理"页面,单击"导入"。
- 3. 在"导入"页面,您可以从本地或OBS中导入元数据。
  - a. 从本地导入
    - "选择本地文件": 单击"上传",选择本地的文件。文件格式必须为xml格式。
    - "名称":您可以在文件名称文本框中定义元数据文件的名字。
    - "存储路径":选择元数据文件存储的OBS路径。

### 图 3-13 从本地导入元数据



# b. 从OBS导入

"选择文件路径": 单击 从OBS中选择Schema文件,文件格式必须为xml。您需要提前将元数据文件上传至OBS中。

"名称":元数据在图引擎服务中的文件名称。

# 图 3-14 从 OBS 导入元数据



4. 单击"确定"完成操作。导入完成后,元数据文件将展示在元数据管理页面中。

# 创建元数据

如果本地或OBS都没有元数据文件,您还可以手动创建元数据文件。

- 1. 在"元数据管理"页面,单击右上角"创建"。
- 2. 在创建元数据文件页面,输入元数据的名称,文件格式默认为xml。
- 3. 存储路径:选择OBS存储路径,请选择一个文件夹来存储数据。
- 4. 定义:元数据文件中详细的Label定义。一个Schema文件可定义多个Label,您可单击"添加lable"按需增加。

在Label定义中,您需要定义Label名称,并在对应Label下添加Property,单击Lable 名称下的"添加",增加属性,还可通过"上移"、"下移"操作对Property进行排序。Property设置参数信息如表3-5所示,其他元数据详情请参考图数据的格式。

表 3-5 Property	参数
----------------	----

参数名称	描述
Property名称	属性的名称,长度限制为1到64位,且不能包含<> &。
基数	数据的复合类型。 <ul><li>● 单值</li><li>● 多值: 勾选是否允许重复值。</li></ul>
数据类型	属性的数据类型。当前支持char、float、double、bool、long、int、date、enum、string。
操作	删除: 删除不需要的属性。

5. 单击"确定"完成操作,当文件创建完成后,元数据文件将展示在元数据管理页面。

# 查找元数据

在元数据管理页面,您可以在搜索框中输入元数据文件的名称进行查找。

# 复制元数据

当编辑元数据时,会覆盖之前的元数据文件,为避免原始元数据文件丢失,建议您在 编辑元数据之前,先复制一份元数据。

- 1. 在"元数据管理"页面,在对应的元数据文件的"操作"列,单击"复制"。
- 2. 定义元数据文件名称以及存储路径,然后单击"确定"。
  - "名称": 复制后的元数据文件名称,仅需要设置名称,文件格式默认为xml。
  - "存储路径":元数据文件存储的OBS路径。

文件复制成功后,新的元数据文件将呈现在元数据管理页面。

# 图 3-15 复制元数据文件



# 编辑元数据

如果您导入或创建的元数据文件不符合业务诉求,需要进行变更时,您可以通过编辑元数据对label和Property进行修改。

编辑元数据文件完成后,将覆盖之前的元数据文件。为避免数据丢失,建议您在编辑前,先复制一份元数据。

- 1. 在"元数据管理"页面,在对应的元数据文件的"操作"列,单击"编辑"。
- 2. 在"编辑"页面,您可以添加Label、添加Property、修改Label名称、通过上移和下移调整Property排序等操作。
- 3. 修改完成后,单击"确定"保存修改。

# 删除元数据

当元数据文件失效后,您可以在"元数据管理"页面,元数据文件对应操作列中单击 "删除",删除元数据文件。

# □□说明

数据删除后无法恢复,请谨慎操作。

# 4 访问和使用图引擎

# 4.1 访问图引擎编辑器

# 操作场景

您可以通过图引擎编辑器对图进行分析和查询。其内置的丰富的算法,供不同领域客户在不同场景使用;兼容Gremlin查询语言,支持开放的API接口。简单易用,让零基础用户快速上手。

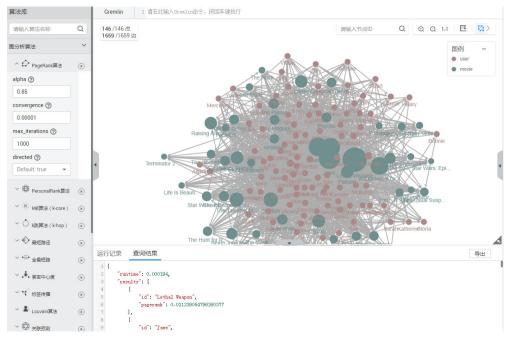
# 操作步骤

步骤1 在图引擎服务管理控制台,选择"图管理"。

步骤2 在"图管理"页面,选择需要访问的图,单击"操作"列的"访问"。

**步骤3** 图引擎编辑器页面如图4-1所示。您可以在编辑器页面对图数据进行分析。

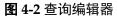


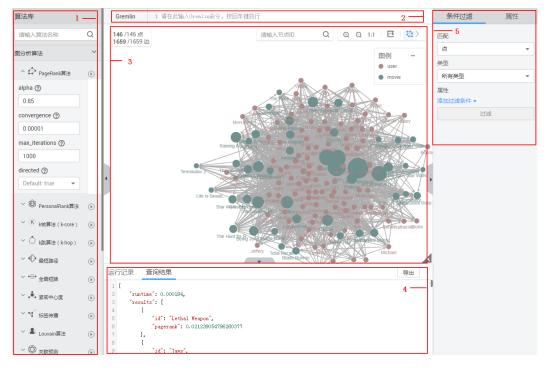


----结束

# 4.2 查询编辑器页面介绍

查询编辑器页面分为绘图区、结果展示区、算法区、Gremlin查询区、属性及过滤区。





# 表 4-1 区域说明

序号	区域名称	说明	
1	算法区	罗列GES服务支持的所有算法,且可在此区域设置各算法相关的属性。算法区功能介绍如表4-2所示。	
2	Gremlin查询区	Gremlin搜索框,可以再查询区输入查询语句执行操作。	
3	绘图区	图数据的可视化展示区。绘图区预置了快捷操作,方便您对图数据进行分析。 绘图区详细功能介绍如 <b>表4-3</b> 所示。	
4	结果展示区	两个页签分别展示运行记录以及查询结果。	
5	属性及条件过滤 区	<ul><li>属性页签展示选中点或边的属性信息。</li><li>条件过滤页签可以设置条件属性,对图数据进行筛选分析。</li></ul>	

# 表 4-2 算法区介绍



# 表 4-3 绘图区介绍

<b>表</b> 4-3 绘图区介绍	
界面元素	说明
144 /146 点 1642 /1659 边	第1行: 144 表示当前图上的点数据,146 表示整张图的所有点数据。 第2行: 1642 表示当前图上的边数据,1659 表示整张图所有边的数据。
请输入节点ID Q	输入节点ID,例如2,按回车或者单击查询图标,可快捷搜索出对应的点数据并渲染到图展示区域。 可支持多个顶点ID输入,例如2,3,4。ID间以英文字符逗号分隔。
Q	放大图,最大放大至6倍。
Q	缩小图,最小缩小至0.05倍。
1:1	适配当前屏幕。 当图界面中展示的点边与当前操作的浏览器窗口 大小失调时,单击此按钮,可快速自动适配窗口 大小。
푠	选择是否显示图例。
> 0 m × A № H	快速切换布局。从左往右分别为:圆形布局、网格布局、核心单节点布局、分层布局、力引导布局、自动分群布局、核心双节点布局。展示效果如图4-3、图4-4、图4-5、图4-6、图4-7、图4-8、图4-9所示。  说明 核心双节点布局必须要选中两个点才可以生效。
绘图区快捷键操作	框选: Ctrl+ 鼠标左键拖动 框选区域内所有点边被选中并高亮,效果如下所示。

界面元素	说明
	正选/反选: Ctrl + 鼠标左键 ctrl+鼠标左键选中点或者边,可以选中并高亮, 按住ctrl键再次点击选中点或者边,可以取消选中 并去除高亮状态。
	全选: Ctrl + A 选中所有点和边,并高亮。
	选择关联点边: Ctrl + E 选中点之后按ctrl+E,可以将有关系的点和边选中 高亮。
	<b>删除: Delete</b> 快捷删除边或者边。
	<b>适配:</b> Alt + F 根据当前屏幕宽度和高度自动缩放所有点和边。
	<b>缩小:</b> - 键盘对应 - 键,缩小当前画布显示。
	<b>放大:</b> = (+)
	<b>取消选中: Esc</b> 所有选中的点和边取消选中,高亮状态消失。
	<b>放大缩小:滑动鼠标滚轮</b> 滑动鼠标滚轮可以放大缩小画布显示。

图 4-3 力引导布局展示效果

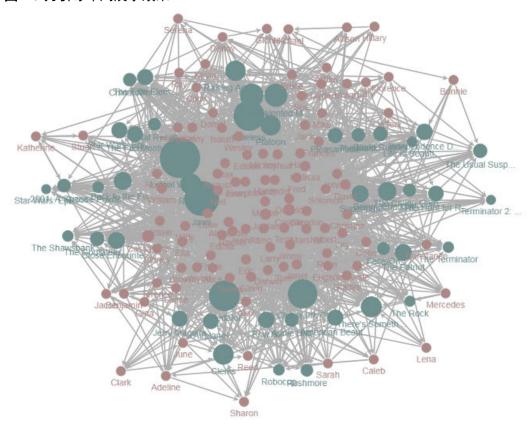


图 4-4 圆形布局展示效果



图 4-5 网格布局展示效果

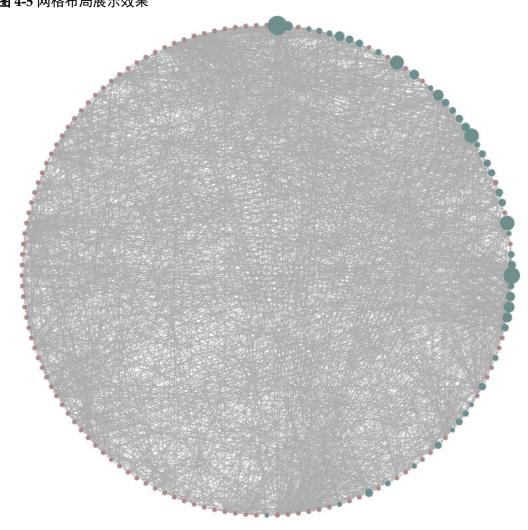


图 4-6 核心单节点布局展示效果

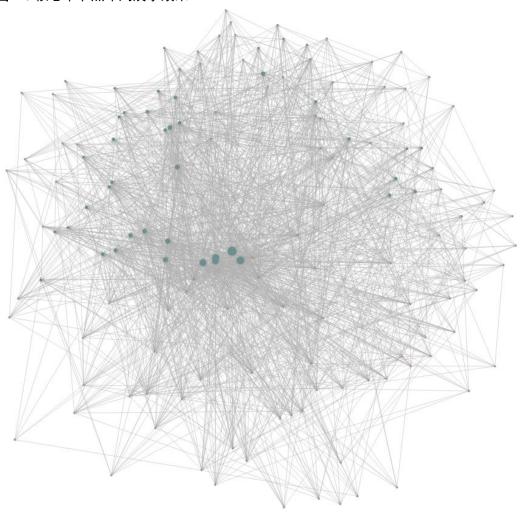


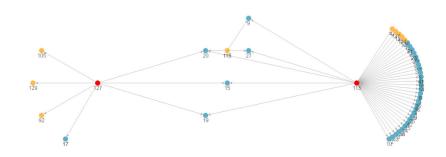
图 4-7 分层布局展示效果



图 4-8 自动分群布局展示效果



#### 图 4-9 核心双节点布局展示效果



## 4.3 Gremlin 查询

## 操作场景

Gremlin是Apache Tinkerpop框架中使用的图遍历语言,使用Gremlin可以很方便的对图数据进行查询,进行图的修改、局部遍历和属性过滤等。

## 操作步骤

**步骤1** 进入图引擎编辑器页面,详细操作请参见访问图引擎编辑器。

**步骤2** 在Gremlin查询区,输入查询语句,按"回车"键执行操作。 常用的查询语句如下所示。

#### ● 点查询

g.V().limit(100): 查询所有点,但限制点的返回数量为100,也可以使用range(x, y)的算子,返回区间内的点数量。

g.V().hasLabel('movie'): 查询点的label值为'movie'的点。

g.V('11'): 查询id为'11'的点。

#### □说明

不推荐使用 "g.V()" 语法, 由于点过大时, 这种查询方式影响展示效果。

### ● 边查询

g.E(): 查询所有边,不推荐使用,边数过大时,这种查询方式不合理,一般需要添加过滤条件或限制返回数量。

g.E('55-81-5'): 查询边id为'55-81-5'的边。

g.E().hasLabel('rate'): 查询label为 'rate' 的边。

g.V('46').outE('rate'): 查询点id为 '46' 所有label为 'rate' 的边。

#### ● 属性查询

g.V().limit(3).valueMap(): 查询点的所有属性(可填参数,表示只查询该点,一个点所有属性一行结果)。

g.V().limit(1).label(): 查询点的label。

g.V().limit(10).values('userid'): 查询点的name属性(可不填参数,表示查询所有属性,一个点每个属性一行结果,只有value,没有key)。

#### ● 新增点

方式1:

a = graph.addVertex(label,'user',id,'500','age','18-24'): 新增点,Label为user,ID为500,age为18-24。

方式2:

g.addV('user').property(id,'600').property('age','18-24'): 新增点,Label为user,ID为500,age为18-24。

- 删除点
  - g.V('600').drop(): 删除ID为600的点。
- 新增边
  - a = graph.addVertex(label,'user',id,'501','age','18-24')
  - b = graph.addVertex(label, 'movie', id, '502', 'title', 'love')
  - a.addEdge('rate',b,'Rating','4'): 新增边,边的两个点ID分别为501、502。
- 删除边
  - g.E('501-502-0').drop(): 删除ID为"501-502-0"的边。

#### ----结束

## 参考信息

GES服务中的Gremlin与开源的差异点如表4-4所示。

#### 表 4-4 差异点

差异点	说明			
Vertex and Edge IDs	Edge id是由source vertex的id,target vertex的id,以及区分重边的index,通过'-'字符连接,sid-tid-index。 Edge id和vertex id必须是String类型。			
User Supplied IDs	只有点id允许用户提供,不能带'-'字符。			
Vertex Property IDs	和边的属性一样,点属性没有id,返回的id为点的id。			
Vertex and Edge Property	GES的vertex 和 edge property由schema文件定义,所以没有真的增加以及删除property的方法,只有修改以及清除的操作,类似property(),remove()等方法,都是修改属性的值,property()设置的值由参数决定,remove(),会把string类属性,变为空字符串,数字类属性变为0。List属性变为空list。			
Variables	GES graph structure不支持variables特性。			
Cardinality	GES支持single 和list cardinality,GES的点的属性的value 类型由schema文件定义,所以设置propety的值的时候, 不会增加新的属性,之后修改对应的定义的属性。			
Transactions	GES的Gremlin实现不支持显式地使用Transactions。			

使用feature函数可以看到当前支持的Gremlin特性,显示false表示GES服务不支持此特性,显示为true表示GES服务支持此特性,特性详情可参考Gremlin官网。

gremlin> graph.features()
==>FEATURES

## 4.4 使用算法分析图

### 操作场景

社交网络、电商推荐等场景下,常常可以使用图算法做关系分析、社团发现。举例来说,使用PageRank算法分析社交网络中的关键人物,使用最短路径算法找到人物之间的关系路径、做好友推荐,使用k-core算法做小圈子发现。

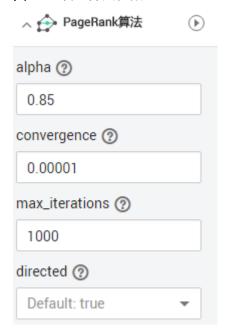
### 操作步骤

步骤1 进入图引擎编辑器页面,详细操作请参见访问图引擎编辑器。

**步骤2** 在算法分析区, 你可以选择算法, 并设置参数。

图引擎服务支持的算法如**算法一览表**所示,详细算法介绍请参见**算法参考**。

#### 图 4-10 设置算法参数



**步骤3** 单击 ▶ 运行算法分析,您可以在如下界面获取查询结果。

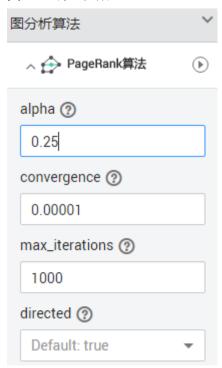
#### 图 4-11 查询结果

```
运行记录
             查询结果
 1 {
 2
       "runtime": 0.00034,
 3
       "results": [
 4
           {
               "id": "Lethal Weapon",
 5
 6
               "pagerank": 0.021239054796260377
 7
          },
 8
               "id": "Jaws",
```

**步骤4** 以模板中的电影数据为例,调整参数后得到的PagRank值不同,但TOP排序不会有明显差异。

挑出了最有影响力的两本电影,分别是ID Lethal Weapon 和ID Jaws。

#### 图 4-12 调整参数

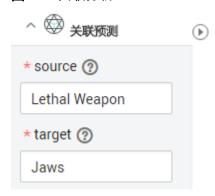


#### 图 4-13 调整参数后的查询结果

```
查询结果
运行记录
 1 {
 2
       "runtime": 0.000106,
       "results": [
 3
           {
 4
               "id": "Lethal Weapon",
 5
               "pagerank": 0.011853634993846037
 6
 7
           },
 8
               "id": "Jaws",
 9
```

**步骤5** 分析一下他们的关联程度,关联度达到0.35,说明还是有不少人同时看过这两本电影的。

#### 图 4-14 关联分析



#### 图 4-15 分析结果

----结束

## 4.5 在绘图区分析图

#### 操作场景

绘图区直观地展示了图数据的情况,您也可以通过绘图区对数据进行分析。

绘图区的快捷键以及界面元素的功能,请参见表4-3。

## 操作步骤

步骤1 进入图引擎编辑器页面,详细操作请参见访问图引擎编辑器。

步骤2 在绘图区,您可以选中一个点,单击右键。

#### 图 4-16 右键选项



#### ● 查看属性

选中查看属性, 您可以在属性页签查看选中节点的属性信息。

#### 图 4-17 查看属性



#### ● 扩线查询

扩展出当前点有关联的点,可以选择OUT(沿出边)、IN(沿入边)或者ALL(双向)。

#### ● 导出图片

导出当前绘图区显示的图。

#### ● 路径查询

选中两个点,然后右键选中路径查询。只有当选中两个点时该选项才可用,否则置灰。

运行结束后将清空绘图区,返回点边数据并渲染绘图区,返回数据中根据选中的两个点会描绘一条路径。

#### ----结束

## 4.6 条件过滤

## 操作场景

为了方便用户对图数据的分析,可以通过设置条件过滤,对图数据进行进一步的过滤分析。

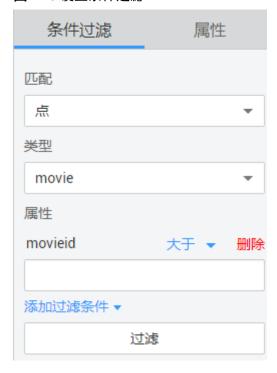
## 操作步骤

步骤1 进入图引擎编辑器页面,详细操作请参见访问图引擎编辑器。

步骤2 在"属性及条件过滤"区,设置条件,然后单击"过滤"。

- 匹配:默认选中点,表示要过滤的是点或者边。
- 类型:默认选中所有类型,下拉框中显示点边所有类型,类型由用户上传的schema文件中定义。
- 添加过滤条件:属性由用户上传的schema文件中定义。选中一个属性,并选择约束条件(包含小于、大于、等于、不等于、在范围、存在、不存在、大于或等于、小于或等于)。此处可添加多条过滤条件,也可单击"删除",删除已设置的过滤条件。





**步骤3** 运行结束后,过滤结果将呈现在绘图区以及结果查询区。

----结束

## 4.7 查看运行记录

### 操作场景

系统记录用户执行操作记录,方便用户在分析数据时了解执行进度和执行完成时间。

#### 操作步骤

**步骤1** 进入图引擎编辑器页面,详细操作请参见访问图引擎编辑器。

**步骤2** 在执行Gremlin查询或算法分析之后,在"运行记录"页签下,展示运行情况。运行记录会显示本地时间,发送的Gremlin命令及命令内容。

#### 图 4-19 Gremlin 运行记录

```
运行记录 查询结果

1 2018-08-02 19:59:23 GMT+08:00 Gremlin命令: g.V().limit(100) 运行
2 2018-08-02 19:59:24 GMT+08:00 Gremlin命令: g.V().limit(100) 总耗时: 0.1607 s
```

在左侧算法执行后,运行记录会显示本地时间、执行的算法名称及运行状态,总共三个状态,分别是等待运行、运行中、运行成功。

#### 图 4-20 算法执行的运行记录

#### Running Record Query Result

```
1 2018-08-02 20:01:13 GMT+08:00 PageRank The algorithm is waiting to run.
2 2018-08-02 20:01:15 GMT+08:00 PageRank Algorithm ran successfully.
3 2018-08-02 20:01:15 GMT+08:00 PageRank Total duration:0.0003 s
```

----结束

## 4.8 查看查询结果

## 操作场景

数据分析结束后,您可以直接在绘图区查看结果或者在"查询结果"页签获取结果信息。

### 操作步骤

步骤1 进入图引擎编辑器页面,详细操作请参见访问图引擎编辑器。

步骤2 在执行Gremlin查询或算法分析之后,在"查询结果"页签下,展示运行情况。

当返回结果很大,绘图区和结果区无法完全展示时,单击右上角"导出"按钮,可下载分析结果。

#### 图 4-21 查询结果

#### 运行记录 查询结果

----结束

# **5** 算法参考

## 5.1 算法一览表

为满足用户各种场景需求,图引擎服务提供了丰富的基础图算法、图挖掘算法和图指标算法。算法简介如下所示。

### 表 5-1 算法一览表

算法	介绍			
PageRank算法	又称网页排名,是一种由搜索引擎根据网页(节点)之间相互 的超链接计算的技术,用来体现网页(节点)的相关性和重要 性。			
PersonalRank算法	PersonalRank算法又称Personalized PageRank算法。该算法继承了经典PageRank算法的思想,利用图链接结构来递归地计算各节点的重要性。与PageRank算法不同的是,为了保证随机行走中各节点的访问概率能够反映出用户的偏好,PersonalRank算法在随机行走中的每次跳转会以(1-alpha)的概率返回到source节点,因此可以基于source节点个性化地计算网络节点的相关性和重要性(PersonalRank值越高,对source节点的相关性/重要性越高)。			
k核算法(k- core)	k-core是图算法中的一个经典算法,用以计算每个节点的核数。 其计算结果是判断节点重要性最常用的参考值之一,较好的刻 画了节点的传播能力。			
k跳算法(k-hop)	从起点出发,通过宽度优先搜索(BFS),找出k层与之关联的所有节点。找到的子图称为起点的ego-net。k跳算法会返回ego-net中节点的个数。			
最短路径 (Shortest Paths)	用于解决图论研究中的一个经典算法问题,旨在寻找图中两节 点之间的最短路径。			
全最短路(All Shortest Paths)	用于解决图论研究中的一个经典算法问题,旨在寻找图中两节 点之间的所有最短路径。			

算法	介绍			
关联路径(n- Paths)	该算法用于寻找图中两节点之间在k层关系内的n条路径。适用于关系分析、路径规划、网络规划等场景。			
紧密中心度 (Closeness Centrality)	紧密中心度是一个节点到所有其他可达节点的最短距离的平均,该指标可以用来衡量信息从该节点传输到其他节点的时间长短。节点的"Closeness Centrality"越小,其所在图中的位置越中心。			
标签传播(Label Propagation)	一种基于图的半监督学习方法,其基本思路是用己标记节点的标签信息去预测未标记节点的标签信息。利用样本间的关系建图,节点包括已标注和未标注数据,其边表示两个节点的相似度,节点的标签按相似度传递给其他节点。标签数据就像是一个源头,可以对无标签数据进行标注,节点的相似度越大,标签越容易传播。			
Louvain算法	基于模块度的社区发现算法,该算法在效率和效果上都表现较好,并且能够发现层次性的社区结构,其优化目标是最大化整个社区网络的模块度。			
关联预测(Link Prediction)	给定两个节点,根据Jaccard度量方法计算两个节点的相似程度,预测他们之间的紧密关系。			
Node2vec算法	通过调用word2vec算法,把网络中的节点映射到欧式空间,用向量表示节点的特征。Node2vec通过回退参数P和前进参数参数Q来生成从每个节点出发的随机步,它带有BFS和DFS的混合,回退概率正比于1/P,前进概率正比于1/Q,每个节点出发生成多个随机步,反映出网络的结构信息。			
实时推荐(Realtime Recommendation )	一种基于随机游走模型的实时推荐算法,能够推荐与输入节点相近程度高、关系或喜好相近的节点。该算法可以基于历史购买或浏览数据进行相近商品推荐,也可以针对人进行相近喜好的潜在好友推荐。			
共同邻居 (Common Neighbors)	是一种常用的基本图分析算法,可以得到两个节点所共有的邻居节点,直观地发现社交场合中的共同好友、消费领域共同感兴趣的商品,进一步推测两个节点之间的潜在关系和相近程度。			
单源最短路 (SSSP)	图论中的经典问题,给定一个节点(称为源),该算法给出从该源节点出发到其余各节点的最短路径长度。			
联通分量 (Connected Component)	联通分量代表图中的一个子图,当中所有节点都相互连接。考虑路径方向的为强联通分量(strongly connected component),不考虑路径方向的为弱联通分量(weakly connected component)。 <b>说明</b> 本算法计算得到的是弱联通分量。			
度数关联度 (Degree Correlation)	度数关联度算法计算所有边上起点和终点度数之间的Pearson关 联系数,常用来表征图中高度数节点是否和高度数节点相连。			
三角计数 (Triangle Count)	不考虑边的方向,统计图中三角形个数。三角形越多,代表图中节点关联程度越高,组织关系越严密。			

算法	介绍
聚类系数(Cluster Coeffcient)	聚类系数是表示一个图中节点聚集程度的系数,证据显示,在 现实的网络中,尤其是在特定的网络中,由于相对高密度连接 点的关系,节点总是趋向于建立一组严密的组织关系。

## 5.2 PageRank 算法

## 概述

PageRank又称网页排名,是一种由搜索引擎根据网页(节点)之间相互的超链接计算的技术,用来体现网页(节点)的相关性和重要性。

- 如果一个网页被很多其他网页链接到的话说明这个网页比较重要,也就是 PageRank值会相对较高。
- 如果一个PageRank值很高的网页链接到一个其他的网页,那么被链接到的网页的 PageRank值会相应地因此而提高。

## 适用场景

适用于网页排序、社交网络重点人物挖掘等场景。

## 参数说明

表 5-2 PageRank 算法参数说明

参数	是否 必选	说明	类型	取值范围	默认值
alpha	否	权重系数。	Double	0~1,不包括0和 1。	0.85
convergenc e	否	收敛精度。	Double	0~1,不包括0和 1。	0.00001
max_iterati ons	否	最大迭代次 数。	Int	1~2000。	1000
directed	否	是否考虑边 的方向。	Bool	true或false。	true

## 注意事项

收敛精度设置较大时,迭代会较快停止。

#### 示例

输入参数alpha=0.85,coverage=0.00001,max\_iterations=1000,directed=true,计算结果中的top节点组成的子图会展示在绘图区,节点大小根据PageRank值的大小来区别,JSON结果会展示在结果区。

## 5.3 PersonalRank 算法

#### 概述

PersonalRank算法又称Personalized PageRank算法。该算法继承了经典PageRank算法的思想,利用图链接结构来递归地计算各节点的重要性。与PageRank算法不同的是,为了保证随机行走中各节点的访问概率能够反映出用户的偏好,PersonalRank算法在随机行走中的每次跳转会以(1-alpha)的概率返回到source节点,因此可以基于source节点个性化地计算网络节点的相关性和重要性。(PersonalRank值越高,对source节点的相关性/重要性越高)。

### 适用场景

适用于商品、好友、网页推荐等方面。

## 参数说明

表 5-3 Personal Rank 算法参数说明

参数	是否必 选	说明	类型	取值范围	默认值		
source	是	节点的ID。	String	-	-		
alpha	否	权重系数。	Doubl e	0~1,不包括0和1。	0.85		
converg	否	收敛精度。	Doubl e	0~1,不包括0和1。	0.00001		
max_iter ations	否	最大迭代次 数。	Int	1~2000。	1000		
directed	否	是否考虑边 的方向。	Bool	true或false。	true		

## 注意事项

收敛精度设置较大时, 迭代较快结束。

## 示例

输入参数source=Lee, alpha=0.85, converage=0.00001, max\_iterations=1000, directed=true, 计算结果中的top节点组成的子图会展示在绘图区, 节点大小根据PersonalRank值的大小来区别, JSON结果会展示在结果区。

## 5.4 k 核算法(k-core)

#### 概述

k-core是图算法中的一个经典算法,用以计算每个节点的核数。其计算结果是判断节点重要性最常用的参考值之一,较好的刻画了节点的传播能力。

### 适用场景

适用于社区发现、金融风控等场景。

## 参数说明

#### 表 5-4 k-core 算法参数说明

参数	是否必 选	说明	类型	取值范围	默认值
k	是	核数。 算法会返回核数大 于等于k的节点。	Int	大于等于0。	-

## 注意事项

无。

#### 示例

输入参数k=10, 计算结果中核数大于等于10的节点组成的子图会展示在绘图区, 节点颜色根据核数来区别, JSON结果会展示在结果区。

## 5.5 k 跳算法(k-hop)

#### 概述

从起点出发,通过宽度优先搜索(BFS),找出k层与之关联的所有节点。找到的子图称为起点的"ego-net"。k跳算法会返回ego-net中节点的个数。

## 适用场景

该算法适用于关系发现、影响力预测、好友推荐等场景。

### 参数说明

表 5-5 k-hop 算法参数说明

参数	是否必 选	说明	类型	取值范围	默认值
k	是	跳数。	Int	1~10.	-
source	是	节点的ID。	String	-	-
mode	否	方向。	String	OUT, IN, ALL.	OUT

## 注意事项

k值越大,覆盖的点越广。根据六度空间理论,社交网上6跳可以覆盖到所有人。BFS按边搜索。

## 示例

计算从Lee节点出发三跳关系组成的子图。

输入参数k=3,source=Lee,mode=OUT。子图会展示在绘图区,JSON结果会展示在结果区。

## 5.6 最短路径(Shortest Path)

## 概述

用以解决图论研究中的一个经典算法问题,旨在寻找图中两节点之间的最短路径。

## 适用场景

适用于路径规划、网络规划等场景。

## 参数说明

表 5-6 最短路径算法参数说明

参数	是否必 选	说明	类型	取值范围	默认值
source	是	输入路径的 起点ID。	String	-	-
target	是	输入路径的 终点ID。	String	-	-
weight	否	边上权重。	String	空或字符串。  ● 空: 边上的权重、距离默认为"1"。  ● 字符串: 对应的边上的属性将作为权重,当某边没有对应属性时,权重将默认为1。  说明  边上权重应大于0。	-
directed	否	是否考虑边 的方向。	Bool	true或false。	false

## 注意事项

本算法只返回一条最短路径。

### 示例

计算从Lee节点到Alice节点的一条最短路径。

输入参数source=Lee,target=Alice,weight=weights,directed=false。最短路径会展示在绘图区,JSON结果会展示在结果区。

## 5.7 全最短路(All Shortest Paths)

## 概述

全最短路径用以解决图论研究中的一个经典算法问题,旨在寻找图中两节点之间的所有最短路径。

## 适用场景

适用于路径规划、网络规划等场景。

## 参数说明

表 5-7 All Shortest Paths 参数说明

参数	是否必 选	说明	类型	取值范围	默认值
source	是	输入路径的起 点ID。	String	-	-
target	是	输入路径的终 点ID。	String	-	-
directed	否	是否考虑边的 方向。	Bool	true或false。	false

## 注意事项

无。

## 示例

输入参数source =Lee, target =Alice, directed=false。计算结果会展示在绘图区,JSON结果会展示在结果区。

## 5.8 关联路径 (n-Paths)

### 概述

该算法用于寻找图中两节点之间在层关系内的n条路径。

## 适用场景

适用于关系分析、路径规划、网络规划等场景。

## 参数说明

表 5-8 n-Paths 算法参数说明

参数	是否必选	说明	类型	取值范围	默认值
source	是	输入路径的起点 Strin g		-	-
target	是	输入路径的终点ID	、路径的终点ID Strin g		-
directed	否	是否考虑边的方 向。	Bool	true或false	false

参数	是否必选	说明	类型	取值范围	默认值
n	否	路径个数。	Int	1~100	10
k	否	层数	Int	1~10	5

#### 示例

输入参数source=Lee, target=Alice, n=10, k=5, directed=false, 计算结果会展示在绘图区, JSON结果会展示在结果区。

## 5.9 紧密中心度(Closeness Centrality)

### 概述

紧密中心度是一个节点到所有其他可达节点的最短距离的倒数,进行累积后归一化的值。该指标可以用来衡量信息从该节点传输到其他节点的时间长短。节点的"Closeness Centrality"越小,其所在图中的位置越中心。

## 适用场景

该算法常用于社交网络中关键节点挖掘等。

## 参数说明

#### 表 5-9 紧密中心度参数说明

参数	是否必 选	说明	类型	取值范围	默认值
source	是	输入需要计算的节 点ID。	String	-	-

### 示例

输入参数source=Lee, 计算Lee节点的紧密中心度, JSON结果会展示在结果区。

## 5.10 标签传播(Label Propagation)

#### 概述

标签传播算法是一种基于图的半监督学习方法,其基本思路是用已标记节点的标签信息去预测未标记节点的标签信息。利用样本间的关系建图,节点包括已标注和未标注数据,其边表示两个节点的相似度,节点的标签按相似度传递给其他节点。标签数据就像是一个源头,可以对无标签数据进行标注,节点的相似度越大,标签越容易传播。

### 适用场景

常用于资讯传播、广告推荐、社区发现等场景。

## 参数说明

#### 表 5-10 标签传播算法参数说明

参数	是否 必选	说明	类型	取值范围	默认值
coveragenc e	否	收敛精度。	Doubl e	0~1,不包括0和 1。	0.00001
max_iterati	否	最大迭代次 数。	Int	1~2000。	1000

## 注意事项

标签传播默认使用ID作为标签。

## 示例

输入参数coverage=0.00001, max\_iterations=1000, 计算得到带有不同标签的子图会展示在绘图区, 节点颜色根据不同标签来区别, JSON结果会展示在结果区。

## 5.11 Louvain 算法

## 概述

Louvain算法是基于模块度的社区发现算法,该算法在效率和效果上都表现较好,并且 能够发现层次性的社区结构,其优化目标是最大化整个社区网络的模块度。

## 适用场景

常用于社团挖掘、层次化聚类等场景。

## 参数说明

### 表 5-11 Louvain 算法参数说明

参数	是否必选	说明	类型	取值范围	默认值
coveragenc e	否	收敛精度。	Doub le	0~1,不包括0和1。	0.00001

参数	是否必选	说明	类型	取值范围	默认值
max_iterati	否	最大迭代次 数。	Int	1~2000。	100

## 注意事项

本算法只生成最后的社区结果,不保存层次化结果。

## 示例

输入参数coverage=0.00001, max\_iterations=100, 计算得到不同社区的子图会展示在绘图区, 节点颜色根据不同社区来区别, JSON结果会展示在结果区。

## 5.12 关联预测(Link Prediction)

### 概述

给定两个节点,根据Jaccard度量方法计算两个节点的相似程度,预测他们之间的紧密 关系。

## 适用场景

常用于社交网上的好友推荐、关系预测等场景。

## 参数说明

#### 表 5-12 关联预测算法参数说明

参数	是否必选	说明	类型	取值范围	默认值
source	是	输入起点ID。	String	-	-
target	是	输入终点ID。	String	-	-

## 示例

输入参数source=Lee, target=Alice, 计算两个节点之间的关联度, JSON结果会展示在结果区。

## 5.13 Node2vec 算法

## 概述

Node2vec算法通过调用word2vec算法,把网络中的节点映射到欧式空间,用向量表示节点的特征。Node2vec通过回退参数 P 和前进参数 Q 来生成从每个节点出发的随机步,它带有BFS和DFS的混合,回退概率正比于1/P,前进概率正比于1/Q。每个节点出发生成多个随机步,反映出网络的结构信息。

## 适用场景

适用于节点功能相似性比较、节点结构相似性比较、社团聚类等场景。

## 参数说明

表 5-13 Node2vec 算法参数说明

参数	是否必 选	说明	类型	取值范围	默认值
P	否	回退参数。	Doubl e	-	1
Q	否	前进参数。	Doubl e	-	1
dim	否	映射维度。	Int	1~200,包括1和200。	50
walkLen gth	否	随机步长。	Int	建议取1~100,包括1和 100。	40
walkNu mber	否	每个节点的随机 步长数。	Int	建议取1~100,包括1和 100。	10
iteration s	否	迭代次数。	Int	1~100,包括1和100。	10

## 注意事项

无。

## 示例

输入参数 P=1,Q=0.3,dim=3,walkLength=20,walkNumber=10,iterations=40,得到每个节点的三维向量表示。

## 5.14 实时推荐 (Real-time Recommendation)

## 概述

实时推荐是一种基于随机游走模型的实时推荐算法,能够推荐与输入节点相近程度高、关系或喜好相近的节点。

## 适用场景

该算法可以基于历史购买和浏览数据进行相近商品推荐,也可以为用户进行相近喜好的潜在好友推荐。

适用于电商、社交等多领域的推荐场景。

## 参数说明

#### 表 5-14 实时推荐算法参数说明

参数	是否必选	说明	类型	取值范围	默认 值
sources	是	节点的ID,可以是多个。	Strin g	source节点的个数不超过30个。	-
alpha	否	权重系数,其值越大,步长越长。	Doub le	0~1,不包括0 和1。	0.85
N	否	总的游走步数。	Int	1~200000。	1000
nv	否	游走过程提前结束参数: 候选推荐 节点访问次数的最小值。 说明 对于一个节点,如果其在随机游走过 程被访问到,且被访问到的次数达到 "nv",则该节点将记入候选推荐的 节点。	Int	1~10。	5
np	否	游走过程提前结束参数:候选推荐 节点个数。 说明 若某个source节点的候选推荐节点达 到"np",对于该source节点的随机 游走将提前结束。	Int	1~2000。	1000

参数	是否必选	说明	类型	取值范围	默认值
label	否	希望输出的点的类型。 说明      其值为空时,将不考虑点的类型,输出算法原始计算结果。     对其赋值时,将从计算结果中过滤出具有该"label"的点的返回。	Strin g	节点label。	-
directed	否	是否考虑边的方向。	Bool	true 或false。	true

## 注意事项

结束条件中,nv和np越小,算法提前结束越快。

#### 示例

输入参数sources =Lee, alpha=0.85, N=10000, nv=5, np=1000, label为空, directed=true。

计算结果中的top节点组成的子图会展示在绘图区,节点大小根据最终的score值的大小来区别,JSON结果会展示在结果区。

## 5.15 共同邻居(Common Neighbors)

#### 概述

共同邻居是一种常用的基本图分析算法,可以得到两个节点所共有的邻居节点,直观 地发现社交场合中的共同好友、以及在消费领域共同感兴趣的商品,进一步推测两个 节点之间的潜在关系和相近程度。

## 适用场景

适用于电商、社交等多领域的推荐场景。

## 参数说明

#### 表 5-15 共同邻居参数说明

参数	是否必 选	说明	类型	取值范围	默认值
source	是	输入起点ID。	String	-	-
target	是	输入终点ID。	String	-	-

### 注意事项

无。

### 示例

输入参数source=Lee, target=Alice, 计算结果会展示在绘图区, JSON结果会展示在结果区。

## 5.16 单源最短路(SSSP)

#### 概述

图论中的经典问题,给定一个节点(称为源),该算法给出从该源节点出发到其余各节点的最短路径长度。

### 适用场景

可应用于网络路由、路径规划等领域。

## 参数说明

#### 表 5-16 SSSP 算法参数说明

参数	是否必选	说明	类型	取值范围	默认值
source	是	节点的ID。	Strin g	-	-
directed	否	是否考虑边的方 向。	Bool	true或false。	true

## 示例

计算从Lee节点出发,到其余各节点的最短路径长度。

输入参数source=Lee, directed=true。

## 5.17 联通分量(Connected Component)

#### 概述

联通分量代表图中的一个子图,当中所有节点都相互连接。考虑路径方向的为强联通分量(strongly connected component),不考虑路径方向的为弱联通分量(weakly connected component)。本算法计算得到的是弱联通分量。

## 参数说明

无。

#### 示例

单击运行,计算各个节点所属的联通分量,JSON结果会展示在结果区。

## 5.18 度数关联度(Degree Correlation)

#### 概述

度数关联度算法计算所有边上起点和终点度数之间的Pearson关联系数,常用来表示图中高度数节点是否和高度数节点相连。

### 适用场景

该算法常用于衡量图的结构特性。

## 参数说明

无。

#### 示例

单击运行,计算图的度数关联度,JSON结果会展示在结果区。

## 5.19 三角计数(Triangle Count)

### 概述

统计图中三角形个数。三角形越多,代表图中节点关联程度越高,组织关系越严密。

## 适用场景

三角计数常用于衡量图的结构特性。

## 参数说明

无。

## 使用说明

不考虑边的方向以及多边情况。

#### 示例

单击运行,计算图的三角形个数,JSON结果会展示在结果区。

## 5.20 聚类系数 (Cluster Coeffcient)

## 概述

聚类系数表示一个图中节点聚集程度的系数。在现实的网络中,尤其是在特定的网络中,由于相对高密度连接点的关系,节点总是趋向于建立一组严密的组织关系。

## 适用场景

该算法常用于衡量图的结构特性。

## 参数说明

无。

## 使用说明

不考虑多边情况。

## 示例

单击运行,计算图的聚类系数,JSON结果会展示在结果区。