功能需求文档

古诗人时空流动性分析系统

# 系统简介

古诗人时空流动性分析系统，受到“唐宋文学编年地图”的启发，旨在通过利用古诗人的人生轨迹，结合现代轨迹嵌入技术进行分析与可视化展示。不同于“唐宋文学编年地图”，只将侧重点集中在了诗人人生轨迹的可视化与地理位置诗词的统计，本系统更侧重于通过轨迹数据进行数据仓库的建立和数据挖掘与OLAM数据分析，是结合前沿科技技术的对古诗人时空流动性模式进行挖掘和分析的系统。

本系统旨在提供一种通用的古诗人时空轨迹OLAM计算方法与可视化平台，基于此，任何领域的人群人生轨迹数据都可以通过此套流程进行分析和可视化。本系统以中南民族大学王兆鹏的精标注唐宋数据为例，用作系统的主要案例进行分析和可视化展示。

总体上，系统可以分为4（或5）个页面板块：

* 原始轨迹数据展示和搜索
* 人生轨迹数据立方体的OLAM计算分析
  + 时变动态数据（隶属于OLAM分析的一部分，但可以单独成为一个页面进行展示）
* 基于轨迹嵌入技术的时空流动性分析
* 空间规模性迁徙时变可视化

除此之外，文档静态站点主要从系统设计目的、系统分析结果、系统使用说明角度进行编写，其内容部分将与最终成品论文高度耦合，所以在本文档中将不会再赘述。

## 最终定型

系统页面分为5个页面板块，并取消文档静态站点。

***注：本文档所有“最终定型”部分均为开发完成后对原文档的修订。***

# 布局框架部分

## 简介

布局框架部分使用的是Nuxt Layout级别组件，此级别组件包含了基础的页面布局，主要是Banner和侧边栏的部分，而基于页面级别的路由则挂在到主要内容部分。

## 界面框图

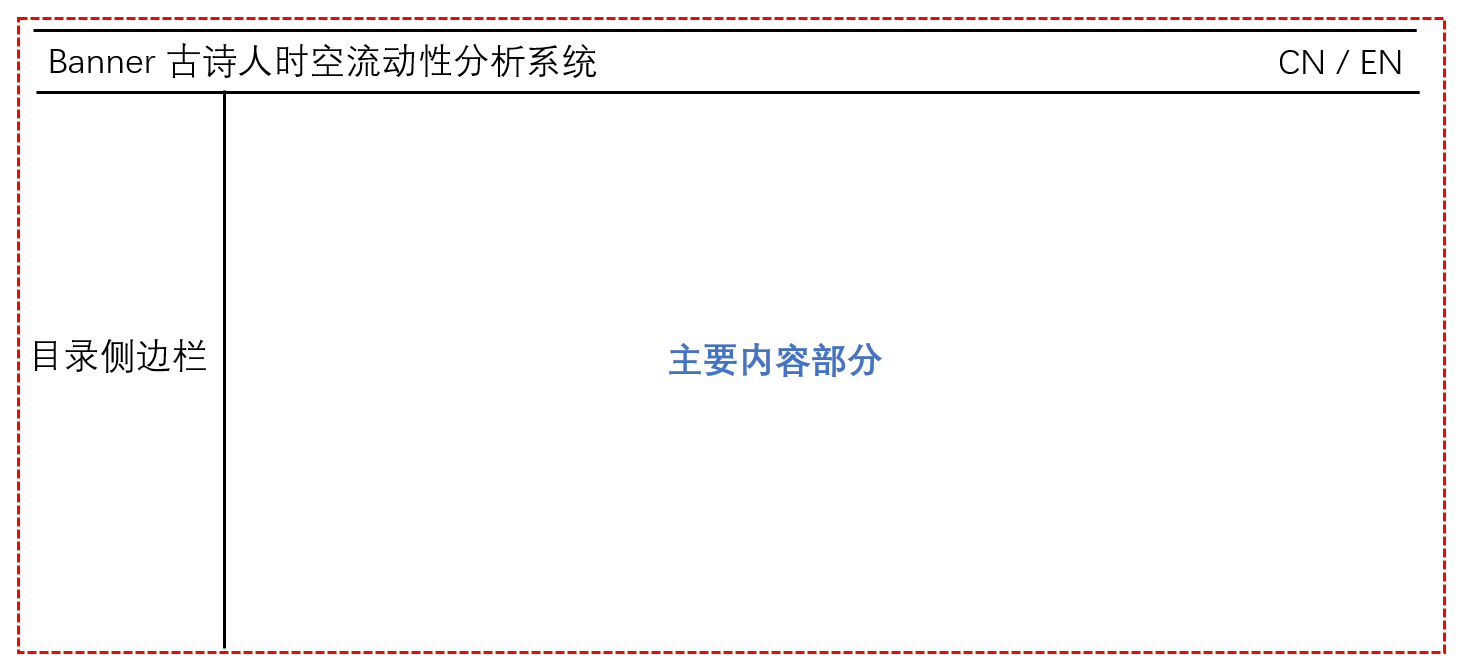


图 1布局界面框图

## 界面功能解析

### Banner横栏部分

Banner横栏主要包含两部分内容：

* 第一部分是系统标题，位于左侧显示，可以使用纯文字/图标Banner
* 第二部分内容是中英文切换部分，可以通过点按按钮切换中英文的页面（只切换翻译）

### 目录侧边栏

由于目录侧边栏的位置默认情况是基于前端路由目标定位的，相比较于使用tab切换不同页面，不需要一次性渲染出所有的DOM节点，这对于前端性能优化是有好的。侧边栏包含有对四个功能模块的跳转逻辑，跳转切换的部分为“主要内容部分”，而整体布局部分不会跟随一起切换。

### 主要内容部分

主要内容部分是Vue Router的主要挂载点，即其他的四个功能性页面都会挂载到这个区域，而对于布局框架部分的视图并不会产生影响。

## 最终定型

Nuxt被替换为Vue 3的Vue CLI，手工编写总体布局框架。侧边栏被取消，采用Banner中的导航按钮取代。中英文切换改为一个按钮动态变化。

图形用户界面, 文本, 应用程序

描述已自动生成

图 2实际框架图示

# 原始轨迹数据展示和搜索

## 简介

本模块是原唐宋诗人人生轨迹展示系统全方位重构升级版本，在修正了原有系统的Bug的同时，大幅提升了系统的美观性和可操作性，极大强化了内容排版以及搜索能力，使其具备有极强的可交互能力。

## 界面框图

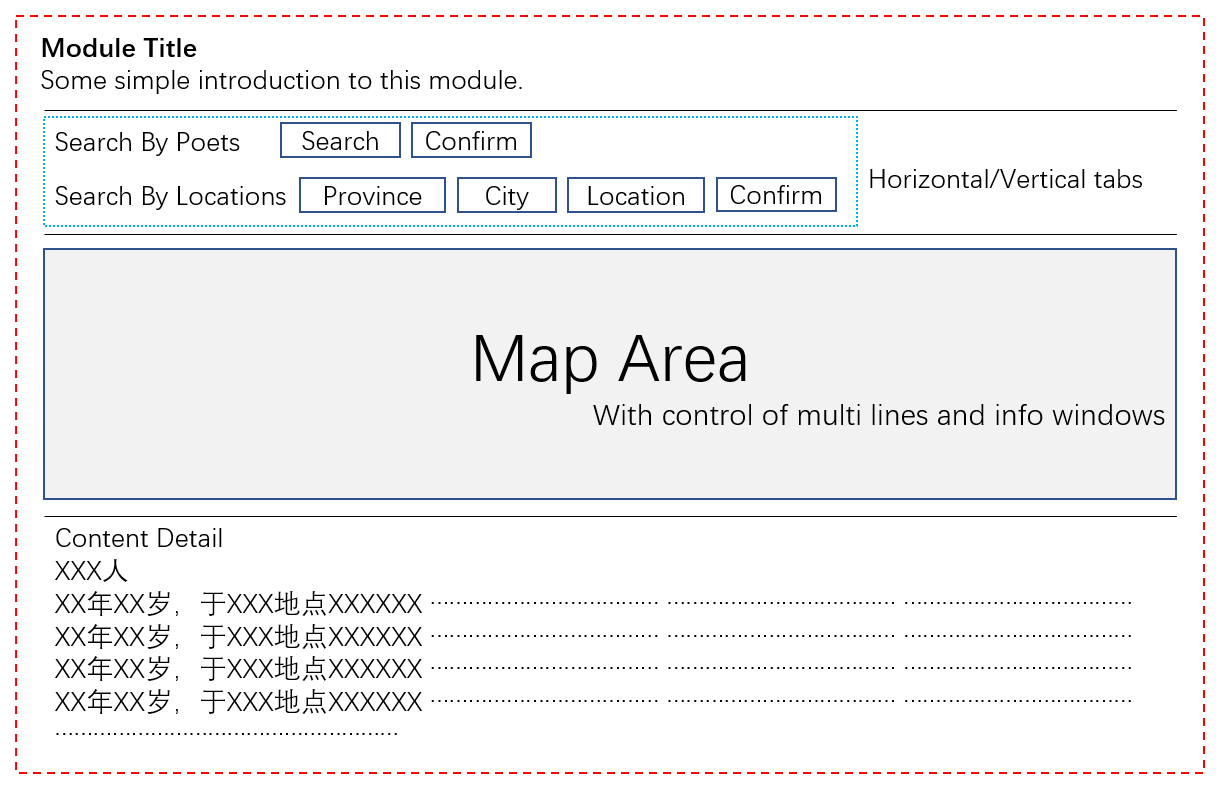


图 3原始轨迹数据展示和搜索界面框图

## 界面功能解析

### 标题部分

标题部分主要包含两部分文字的内容，第一部分文字为标题内容，说明本页面的标题内容，即“原始轨迹数据展示与搜索”。第二部分文字为简介文字，用于简单介绍本页面的：设计目的、功能作用、使用方法。内容不宜过长。

### 搜索部分

搜索部分为可以交互的内容部分，其主要提供了两种搜索方式，分别是按照“诗人”、“地点”搜索。

搜索方式“诗人”、“地点”是互斥的，即只能使用一种搜索方式进行搜索。所以对于两种搜索方式，使用横向排列的Tabs组件进行容纳（实际开发过程中，可以视情况调整为竖向Tabs）。无论是哪种搜索方式，均需要使用Confirm按钮进行确认。

* Confirm按钮：
  + 确认提交当前的搜索Tab的搜索内容
  + 从后端获取到对应的数据，并在前端可视化。
* Tab诗人：
  + 提供Search框下拉按钮，可供下拉选择或者搜索选择诗人
  + 页面初始化时，应当提供诗人的默认值供按钮初始选中
* Tab地点：
  + 提供三级索引（省份、城市、地点）方式定位地点，按钮同样为可搜索的下拉框
  + 由于实际上的省份、城市和地点的内容相对恒定（恒定列表），且复用率偏高，可以考虑直接持久化到Vuex中
  + 页面初始化时，应当提供省份、城市、地点的初始化默认值供按钮选中

### 地图部分

地图部分对应搜索部分的不同，有两种不同的响应方式：

* 搜索内容为诗人时：
  + 清空地图现有的内容
  + 调整地图可视化范围的中点和缩放
  + 将诗人的人生轨迹用多折线的方式绘制在地图上
  + 对轨迹上不同的地点，使用信息窗口的形式附加此诗人在此地点的经历
* 搜索内容为地点时：
  + 清空地图的现有内容
  + 调整地图的可视化范围的中点为此地点，缩放级别相比搜索为诗人时更大
  + 以信息窗体的形式，将所有游历过此地点的诗人的经历附加到信息窗体上（时间顺序）
  + 可以用圆形的方式覆盖此地点以提供更为直观的可视化效果

### 信息部分

信息部分对应搜索部分的不容，同样也有两种不同的响应方式：

* 搜索内容为诗人时：
  + 清空现有内容，将地图部分所有信息窗体的内容按时间顺序，在信息部分排版展示
* 搜索内容为地点时：
  + 清空现有内容，将地图部分信息窗体中的所有内容，按时间顺序在信息部分排版展示

## 最终定型

搜索区域使用两个横向Tab合并为一体，地理索引变为四级（国家、省份、城市、地点），不提供默认值选中；信息部分使用树形组件支持，内容扩充，不按时间组织；地图组件大体不变。

图形用户界面, 文本, 应用程序

描述已自动生成

图 4搜索区域

背景图案

低可信度描述已自动生成

图 5信息区域

# 人生轨迹数据立方体的OLAM计算分析

## 简介

对于诗人的人生轨迹序列数据，理论上讲，可以构成以诗人、时间（最小单位为年）、地理位置的三维数据立方体，基于数据立方体的OLAP操作上卷、下钻、切片，可以划分特定的数据立方体抽样，本模块的目的是针对特定的数据立方体抽样实时可视化数据挖掘结果，达到OLAM分析的目的。

## 界面框图

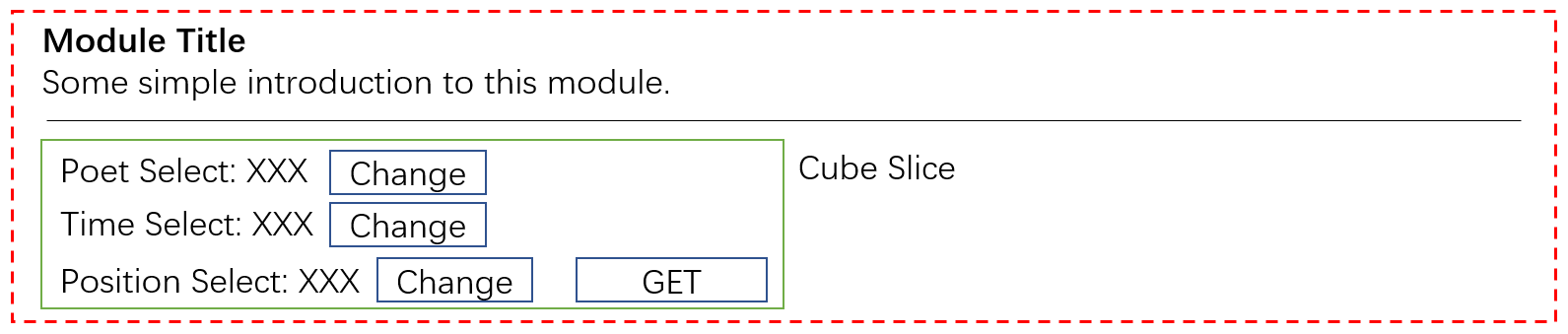


图 6人生轨迹数据立方体OLAM分析界面框图

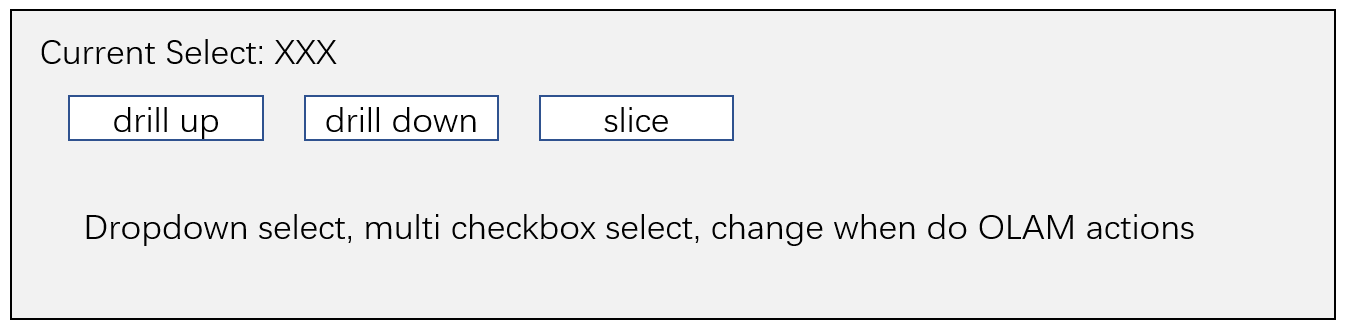


图 7数据选择部分底部表单弹框框图

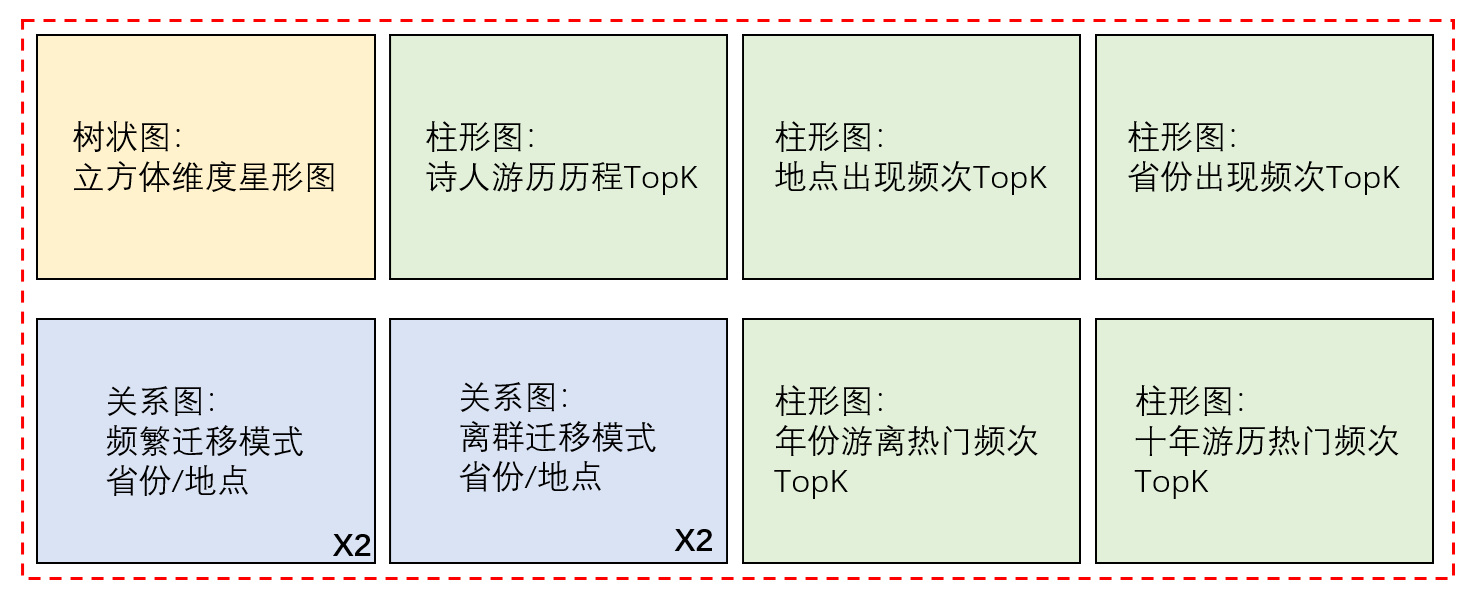


图 8人生轨迹数据立方体OLAM分析界面框图（续）

## 界面功能解析

### 标题部分

标题部分主要包含两部分文字的内容，第一部分文字为标题内容，说明本页面的标题内容，即“诗人人生轨迹数据立方体OLAM分析”。第二部分文字为简介文字，用于简单介绍本页面的：设计目的、功能作用、使用方法。内容不宜过长。

### 数据选择（立方体切片）部分

对于诗人、时间、地点三个维度，可以分别进行上卷、下钻、切片操作（视图见图 6），所有维度的选取操作均通过底部弹出表单的形式进行选取（每个维度对应不同的底部弹出表单，视图见图 7）。此操作步骤可以视为选取切片立方体，各个具体维度选择解释如下：

* 诗人
  + 当前处于诗人级别：
    - 当前选中：具体的单个诗人/切片（外显“切片”）
    - 允许进行：
      * 上卷（到此诗人所对应的朝代）
      * 切片（多选诗人）
  + 当前处于朝代级别：
    - 当前选中：具体的单个朝代/切片（外显“切片”）
    - 允许进行：
      * 下钻（到此朝代的默认诗人处）
      * 切片（多选朝代）
* 时间
  + 当前处于单年级别：
    - 当前选中：具体的单个年份/对应的范围年份
    - 允许进行：
      * 上卷（上卷到此年份所对应的单年）
      * 切片（提供范围年份选取）
  + 当前处于十年级别：
    - 当前选中：某个10年/对应的范围年份
    - 允许进行：
      * 上卷（上卷到此十年所对应的百年）
      * 下钻（下钻到此十年所对应的第一年）
      * 切片（提供范围年份切片，十年级别）
  + 当前处于百年级别：
    - 当前选中：某个百年/对应的范围年份
    - 允许进行：
      * 下钻（下钻到此百年的第一个十年区间）
      * 切片（提供范围年份切片，百年级别）
* 地点
  + 当前处于地点级别：
    - 当前选中：某个具体地点/外显“地点切片”
    - 允许进行：
      * 上卷（上卷到本地点对应的城市）
      * 切片（范围选取地点）
      * 切换（切换同城市下的地点）
  + 当前处于城市级别：
    - 当前选中：某个具体的城市/外显“城市切片”
    - 允许进行：
      * 上卷（上卷到本城市对应的省份）
      * 下钻（下钻到城市对应的默认地点）
      * 切片（范围选取城市）
      * 切换（切换同省份下的城市）
  + 当前处于省级别：
    - 当前选中：某个省份/外显“省份切片”
    - 允许进行：
      * 上卷（上卷到地点对应的国家）
      * 下钻（下钻到省份的默认城市）
      * 切片（对省份范围选取）
      * 切换（切换同国家下的不同省份）
  + 当前处于国家级别：
    - 当前选中：某个国家
    - 允许进行：
      * 下钻（下钻到国家对应的默认省份）
      * 切片（范围选取国家）
      * 切换（切换不同的国家）

### 图表可视化汇总部分

图表可视化汇总为本模块主体界面的续接，具体形式上，包含有10个由柱形图和关系图构成的图表，如图 8所示，具体的，在页面排列时可以按照每行两个进行排列。

*注：根据实际开发的排版布局，可以增设“城市”级别的图表（城市频次TopK、频繁离群的城市迁徙模式），在此章节则不赘述。*

#### 树状图：立方体维度星形图

此图表为三叉树状图，总根节点名为“Cube”，三个子树分叉分别为三个维度。按照切片立方体的选取条件不同，此关系图不同叉臂的展示如下：

* 诗人：永远衔接“诗人/朝代”作为二级节点，之后依据所处级别不同，后可衔接具体的诗人、朝代等，以星型放射作为三级节点连接。当切片涉及三级节点选取不超过三个节点时，可全部展示，若超过三个三级节点，则使用前三个三级节点和“等”节点（共四个三级节点）
* 时间：永远衔接“年份/十年”作为二级节点，之后依据所处级别的不同，其后可以衔接具体的年份/十年，以星型放射作为三级节点连接。当切片涉及三级节点选取不超过三个节点时，可全部展示，若超过三个三级节点，则使用前三个三级节点和“等”节点（共四个三级节点）
* 地点：永远衔接“地点/城市/省份/国家”作为二级节点，之后依据所处级别的不同，其后可以衔接具体的地点、城市、省份、国家，以星型放射作为四级节点连接。当切片涉及四级节点选取不超过三个节点时，可全部展示，若超过三个四级节点，则使用前三个四级节点和“等”节点（共四个四级节点）

#### 柱形图：诗人游历TopK

在对应的切片立方体选取范围之下，按照游历的地点数量作为统计标准（前闭后闭区间），统计所有诗人的游历TopK排名（K默认可取20）。

#### 柱形图：地点频次TopK

在对应的切片立方体的选取范围之下，按照地点被所有诗人造访的频次作为统计标准（前闭后闭区间），统计所有地点的被造访TopK排名（K默认可取20）。

#### 柱形图：省份频次TopK

在对应的切片立方体的选取范围之下，按照省份被所有诗人造访的频次作为统计标准（前闭后闭区间），统计所有省份的被造访TopK排名（K默认可取20）。

#### 关系图：频繁迁徙模式（地点）

统计对应的切片立方体的选取范围之下，按照地点为节点，统计频次被定义为频繁（前35%）的迁徙模式，并按照地点为节点，进行关系图绘制。

#### 关系图：频繁迁徙模式（省份）

统计对应的切片立方体的选取范围之下，按照省份为节点，统计频次被定义为频繁（前35%）的迁徙模式，并按照省份为节点，进行关系图绘制。

#### 关系图：离群迁徙模式（地点）

统计对应的切片立方体的选取范围之下，按照地点为节点，统计频次被定义为离群（后20%）的迁徙模式，并按照地点为节点，进行关系图绘制。

#### 关系图：离群迁徙模式（省份）

统计对应的切片立方体的选取范围之下，按照省份为节点，统计频次被定义为离群（后20%）的迁移模式，并按照省份为节点，进行关系图绘制。

#### 柱形图：游历热门年份TopK

统计对应的切片立方体的选取范围之下，按照年份为统计单位，统计年份内所有游历的频繁频数排名的TopK（K默认可取20）并进行绘制。

#### 柱形图：游历热门十年份TopK

统计对应的切片立方体的选取范围之下，按照十年为统计单位，统计各个十年内所有游历的频繁频数排名的TopK（K默认可取20）并进行绘制。

## 最终定型

OLAP选取部分的基本逻辑不变，UI表现上略有区别：

图形用户界面, 应用程序

描述已自动生成

图 9 OLAP选取区域

时间维度的上卷和下钻替换为尺度增大和尺度缩小：

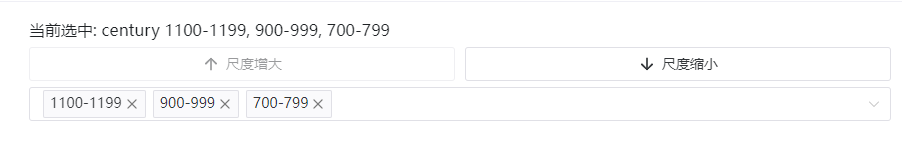


图 10时间维度OLAP抽屉

图表部分树形图表现形式略简化，其他图表大体不变，柱形图均增设底部的缩放拖动交互条，增设游历热门百年图表：

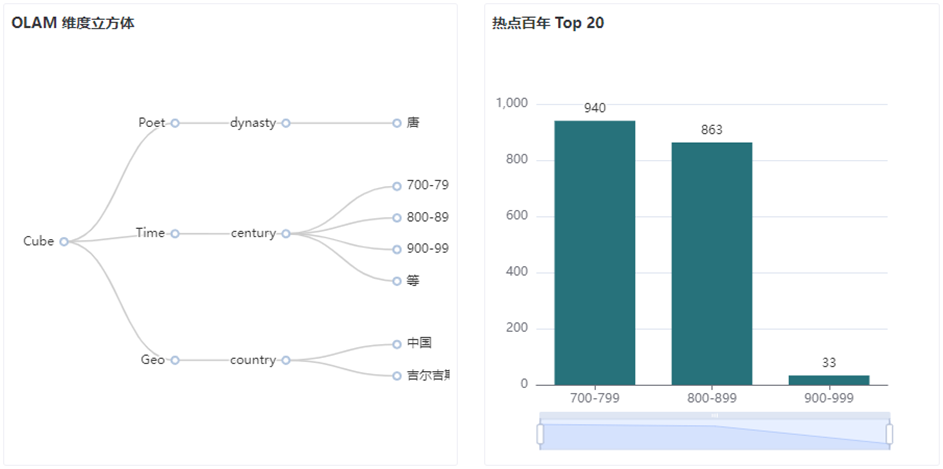


图 11树形图和柱形图示例

OLAM界面会使用预置数据初始化来填充首次加载的内容。

# 时变动态数据可视化

## 简介

为了便于总览分析时变范围内的所有常见数据信息，时变动态数据可视化页面将按照时间变化为顺序，动态展示各类目子数据项的实时排名变化情况。按照随时间的增长，可以可视化的内容包含：诗人游历数量的可视化、地点造访频次的可视化、省份造访频次的可视化、游历频繁十年段的可视化、两地迁徙总数量统计可视化。

## 界面框图

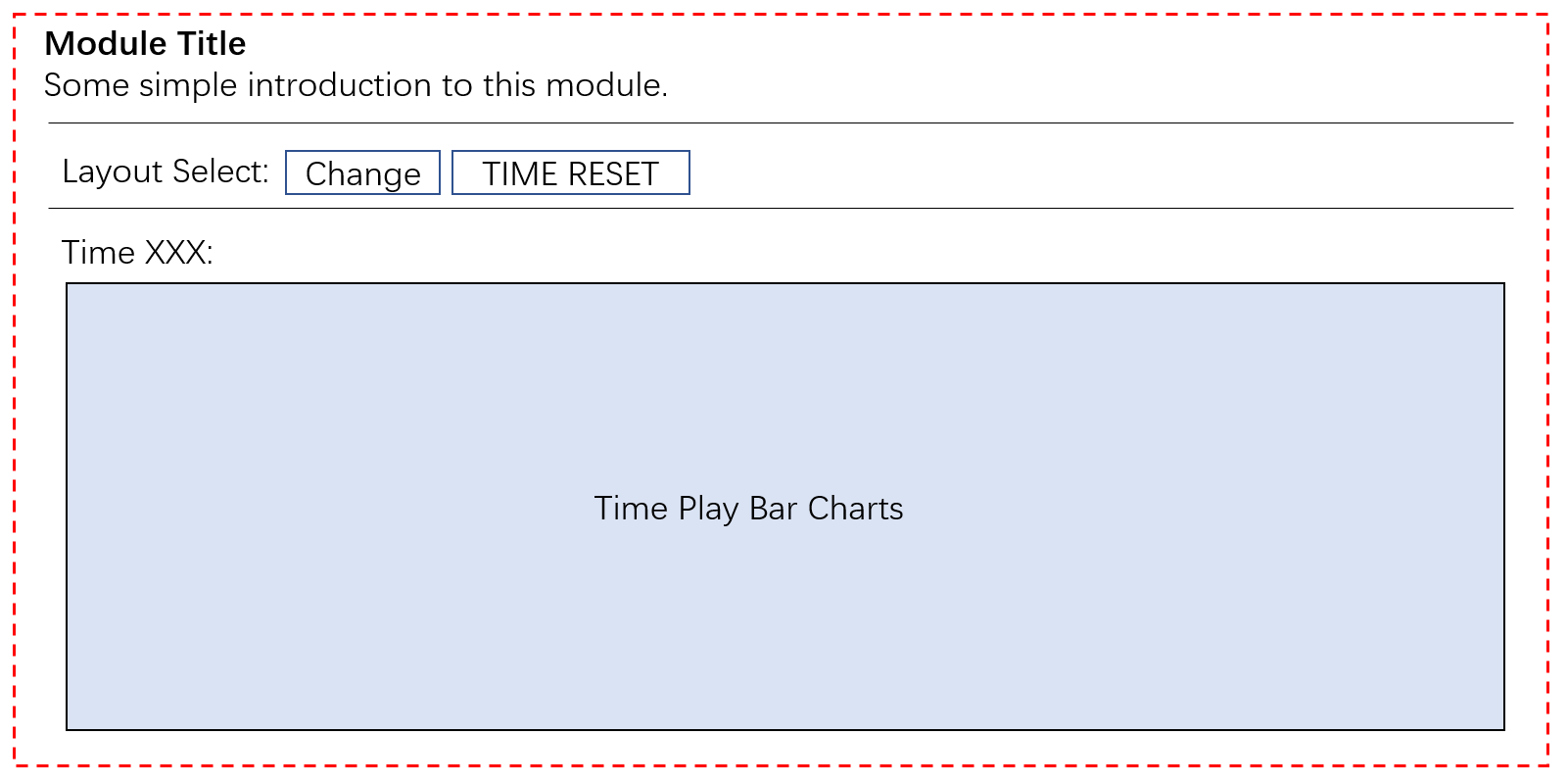


图 12时变动态数据可视化界面框图

## 界面功能解析

### 标题部分

标题部分主要包含两部分文字的内容，第一部分文字为标题内容，说明本页面的标题内容，即“时变动态数据可视化”。第二部分文字为简介文字，用于简单介绍本页面的：设计目的、功能作用、使用方法。内容不宜过长。

### 操作按钮部分

本页面共有两个操作按钮，第一个操作按钮为选择需要可视化的项目内容，可选项包含：诗人游历数量的可视化、地点造访频次的可视化、省份造访频次的可视化、游历频繁十年段的可视化、两地迁徙总数量统计可视化。

第二个按钮为时间重置按钮，用于将时间重置到初始时间并重新开始演示。

### 时变柱形图部分

随着时间由初始时间到终止时间的演变，柱形图的排名也开始动态进行变化，当切换到当前页面时，默认加载默认选中数据（诗人游历数量可视化）。每次切换数据系列时，随机生成一个默认颜色供当前图表的所有系列使用，同时所有框选进入加载状态，从后端获取数据，数据获取完毕后，从当前时间位置进行数据更新，删除旧的计时器，并启动新的计时器。

#### 诗人游历数量

按年份时变，可视化展示诗人的游历数量的排名。

#### 地点造访频次

按年份时变，可视化展示地点被诗人造访频次的数量变化排名。

#### 省份造访频次

按年份时变，可视化展示省份被诗人造访频次的数量变化排名。

#### 游历频繁十年段

按每十年段内的游历数量进行年份时变排名。

#### 两地迁徙总数量统计

在年份时变范围内，统计两地迁徙点对的总数量统计排名。

## 最终定型

选择区域改为下拉框加确认按钮，时间重置按钮移动到动态排序图右下角。同时，演示内容增加到诗人游历时变排名、游历地点时变排名、游历城市时变排名、游历省份时变排名、游历年份时变排名、游历十年时变排名、游历百年时变排名、地点迁徙点对时变排名、城市迁徙点对时变排名、省份迁徙点对时变排名共十个。诗人游历时变排名为默认初始化内容。动态柱形排序图增加侧边缩放交互条。

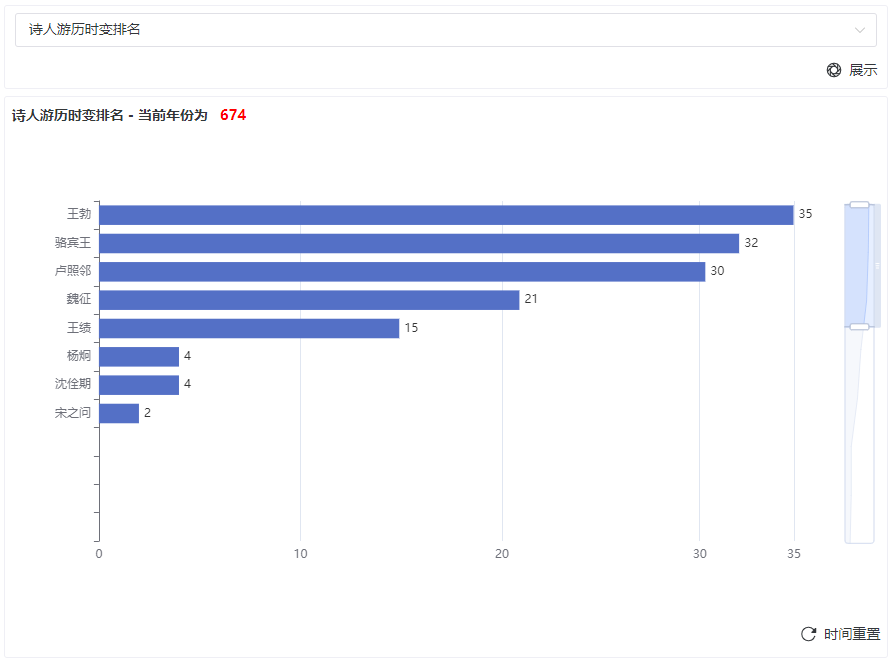


图 13界面示例

# 基于轨迹嵌入技术的时空流动性分析

## 简介

通过使用序列嵌入模型学习所有诗人的轨迹嵌入，再使用不同的聚类算法对所有的轨迹嵌入进行分析和汇总，并将分析结果进行可视化的展示与呈现。

## 界面框图

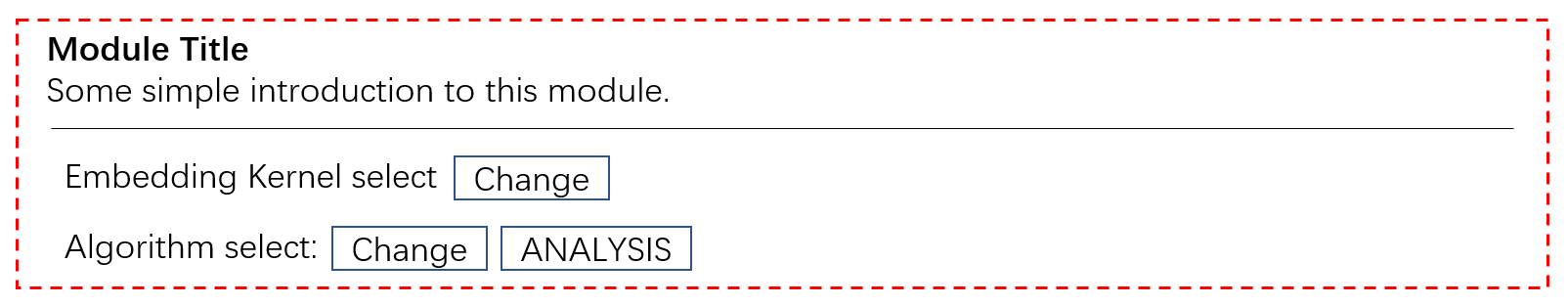


图 14基于嵌入技术的时空流动性分析模块界面框图

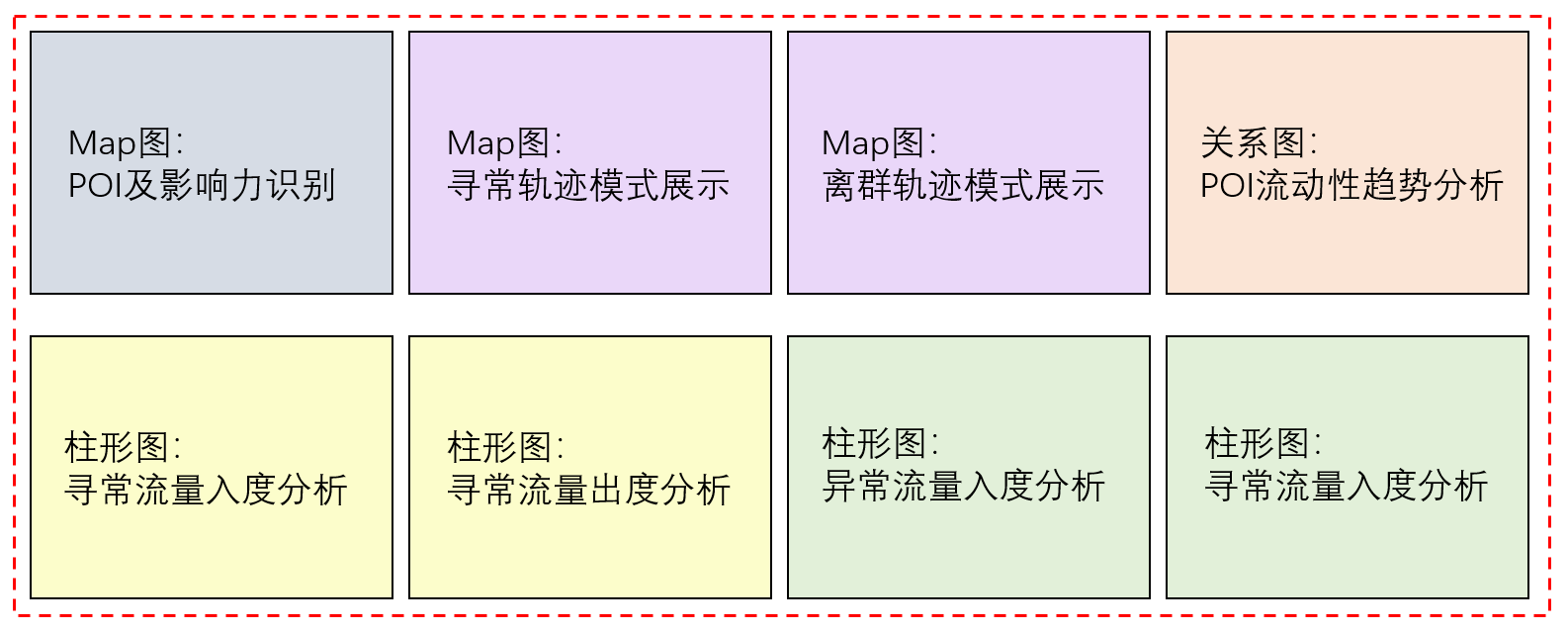


图 15基于嵌入技术的时空流动性分析模块界面框图（续）

## 界面功能解析

### 标题部分

标题部分主要包含两部分文字的内容，第一部分文字为标题内容，说明本页面的标题内容，即“基于嵌入技术的时空流动性分析”。第二部分文字为简介文字，用于简单介绍本页面的：设计目的、功能作用、使用方法。内容不宜过长。

### 嵌入内核与算法选择部分

嵌入内核选择部分提供嵌入算法的内核选取下拉框（目前只提供一个，且默认选中）。

算法选择部分可以操作共两个按钮，分别是算法选择下拉框，用于对算法进行选择；以及分析按钮，将选中算法的分析结果数据从后端取回并在前端进行展示。

### 图表可视化汇总部分

图表可视化汇总为本模块主体界面的续接，具体形式上，包含有8个由柱形图和关系图构成的图表，如图 15所示，具体的，在页面排列时可以按照每行两个进行排列。

*注意，在本部分，我们强调的不是分析的结果，而是提供了一个平台，数据层，嵌入层，算法层互相解耦合，并通过设计层层通信协议的方式，来达到各种可自由扩展的地步。*

#### Map图：POI及影响力大小

基于轨迹点嵌入技术（例如Word2Vec），通过对应的聚类算法，可以针对聚类算法的密度结果，逐级确定POI以及POI的影响力信息。并在Map地图上通过信息窗体（展示地理位置相关信息）和圆形覆盖物的大小来确定影响力的大小。

*由于需要额外进行轨迹点嵌入的分析，且和统计分析模块内容有所重合，所以本处初步预想实现的时候进行忽略，也可以减少一些工作量。*

#### Map图：寻常轨迹模式展示

基于轨迹序列嵌入结果，通过对应的聚类算法，聚类得出寻常轨迹模式和异常轨迹模式。对所有寻常轨迹模式在地图上以折线形式进行展示，并在轨迹的起点/终点用信息窗体的形式附加对应轨迹内容信息。

#### Map图：离群轨迹模式展示

基于轨迹序列嵌入结果，通过对应的聚类算法，聚类得出寻常轨迹模式和异常轨迹模式。对所有异常轨迹模式在地图上以折线形式进行展示，并在轨迹的起点/终点用信息窗体的形式附加对应轨迹内容信息。

#### 关系图：POI流动性趋势分析

根据在6.3.3.1得出的POI信息，结合原始轨迹序列信息，可以得到POI之间的流动性网络，并通过关系图的形式进行可视化展示。

*由于需要额外进行轨迹点嵌入的分析，且和统计分析模块内容有所重合，所以本处初步预想实现的时候进行忽略，也可以减少一些工作量。*

#### 柱形图：寻常流量入度分析

根据在6.3.3.2得出的寻常轨迹模式，对相关涉及到的所有轨迹地点的入度进行分析统计排名。

#### 柱形图：寻常流量出度分析

根据在6.3.3.2得出的寻常轨迹模式，对相关涉及到的所有轨迹地点的出度进行分析统计排名。

#### 柱形图：异常流量入度分析

根据在6.3.3.3得出的异常轨迹模式，对相关涉及到的所有轨迹地点的入度进行分析统计排名。

#### 柱形图：异常流量出度分析

根据在6.3.3.3得出的异常轨迹模式，对相关涉及到的所有轨迹地点的出度进行分析统计排名。

## 最终定型

内核选择即为选择轨迹嵌入表示的算法模型内核，算法选择则是选择用于半监督训练标注数据时的基准算法。分析结果区域大改：

* 只保留一个地图区域，并增加交互式的轨迹选择显示按钮，用红绿显示离群和频繁
* 增加树形图便于更直接查看整体的分析结果
* 其余图表均为柱形图，即离群、频繁的轨迹复杂网络下国家、省份、城市、地点的出入度

地图

描述已自动生成

图 16内核和算法选择部分

地图

描述已自动生成

图 17地图轨迹选取示例

图示

描述已自动生成

图 18树形图结果示例

# 空间规模性迁徙时变可视化

## 简介

受Science一篇以计量社会学论文工作[[1]](#footnote-1)的启发，将诗人的历史人生历程进行动态的随着时间进行的可视化展示，直观展现随着时变的唐宋诗人群体空间规模性迁徙时变可视化。

## 界面框图

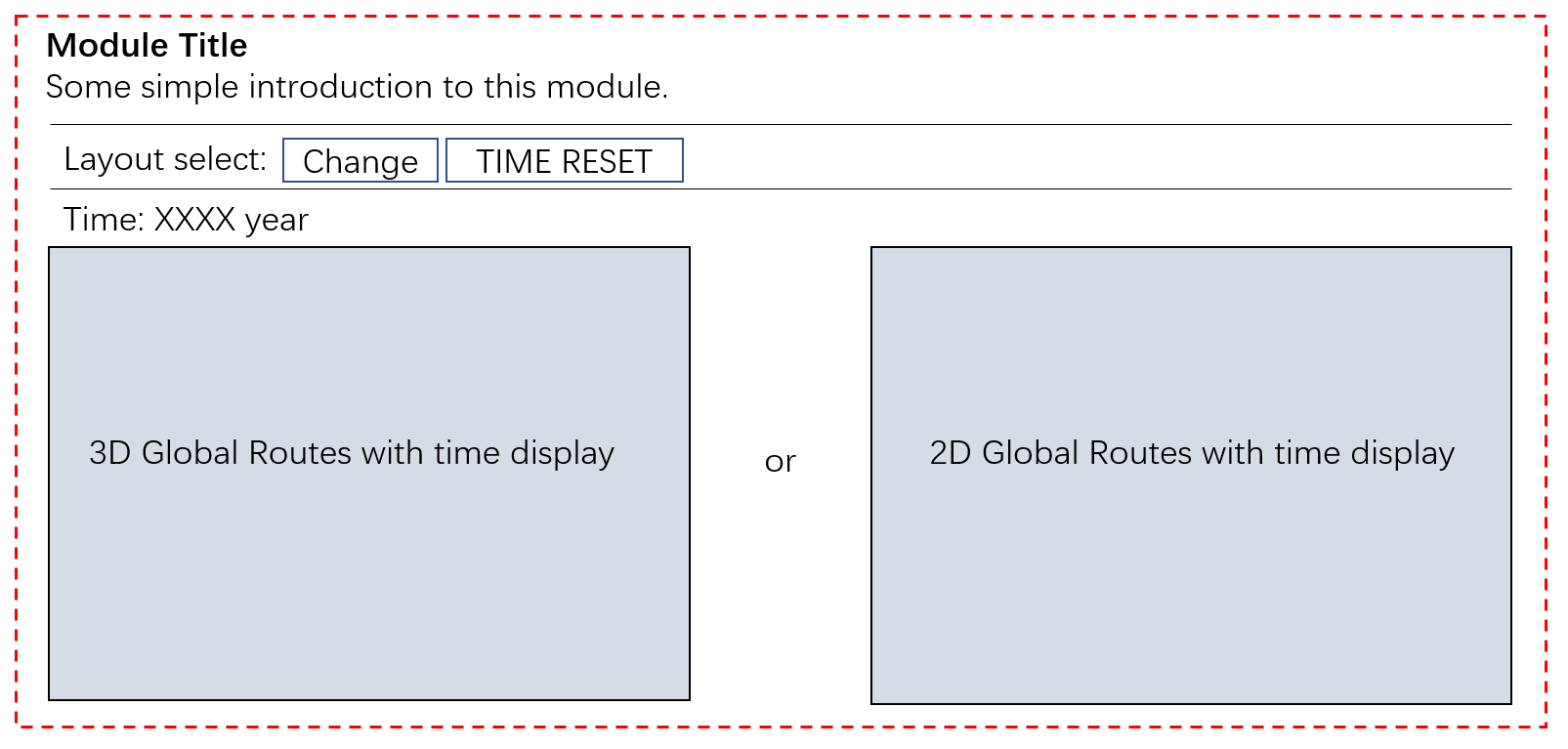


图 19空间规模性迁徙时变可视化模块界面框图

## 界面功能解析

### 标题部分

标题部分主要包含两部分文字的内容，第一部分文字为标题内容，说明本页面的标题内容，即“空间规模性迁徙时变可视化”。第二部分文字为简介文字，用于简单介绍本页面的：设计目的、功能作用、使用方法。内容不宜过长。

### 布局选择部分

可以选择两种地图布局方式（均为ECharts GL实例），分别为3D地球形式布局以及2D世界地图布局，可以在时间演进的过程中进行世界地图的布局更替。

点击时间重置按钮即可重置时间为初始时间并开始时间演进。

### 可视化展示部分

可视化展示的路径随着时间进行变化，时间变动从初始年份直到终止年份，每年份变动之后，地图数据也随着进行更新。页面刷新之后，数据在3秒的定时之后进行演变。数据演变直到年份达到终止年份为止，数据将不再变动。年与年之间的数据更新时间间隔可以根据实践进行设定（根据条件判断，在定时任务中反复进行setTimeout，或根据setInterval方便获取定时器）。

## 最终定型

定型使用高德地图Loca可视化（而非ECharts GL）。受到前端性能问题的影响（3D过于卡顿），布局仅支持一种（2D脉冲线+脉冲点），不可切换；时间重置按钮移动到地图的右下方。定时按照大概一秒三次来变化。

地图

描述已自动生成

图 20可视化示例

1. Schich M, Song C, Ahn Y Y, et al. A network framework of cultural history[J]. science, 2014, 345(6196): 558-562 [↑](#footnote-ref-1)