

杭州电子科技大学

硕士学位论文

题 目: PathBubbles: 生物通路的交互式可
视化的研究与开发

研 究 生 竺涌楠

专 业 计算机应用技术

指导教师 方景龙(研究员)

陈健(助理教授)

潘志庚(研究员)

完成日期 2015 年 3 月

杭州电子科技大学硕士学位论文

**PathBubbles:生物通路的交互式可视化
的研究与开发**

研 究 生: 竺涌楠

指导教师: 方景龙 研究员

陈 健 助理教授

潘志庚 研究员

2015 年 03 月

Dissertation Submitted to Hangzhou Dianzi University

for the Degree of Master

**The Research and Development of
PathBubbles—an Interactive
Visualization System for Biological
Pathway Analysis**

Candidate: Yongnan zhu

Supervisor: Prof. Jinglong Fang

Assistant Prof. Jian Chen

Prof. Zhigeng Pan

March, 2015

杭州电子科技大学

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名： 日期： 年 月 日

学位论文使用授权说明

本人完全了解杭州电子科技大学关于保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属杭州电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署单位名称仍然为杭州电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。
(保密论文在解密后遵守此规定)

论文作者签名： 日期： 年 月 日

指导教师签名： 日期： 年 月 日

摘 要

随着生物技术的研究进展,产生了大量的生物通路数据和相关的实验数据。如何整合已有的生物信息数据来探索未知的生物通路信息,并且挖掘各种生物通路、各类物种之间的关系,正变得越来越重要。人类可以迅速地解释视觉信息,用综合的方式来解释这些数据的可视化形式将极具价值。然而,人类的记忆力是有限的。目前的可视化工具也仅仅只是为用户提供有限的可以用来分析大而复杂的数据的浏览和操纵的能力,例如:许多的数据分析中间过程繁多,这就需要应用不同的窗口来展示不同的信息。因此,仍然不能有效地降低生物学家的记忆和工作负担。现存的生物信息数据库,譬如:Reactome、KEGG等尽管提供了丰富的数据,不过缺少一个用户友好的交互式可视化分析界面。另外,现有的许多生物通路的可视化分析工具,只能提供最为具体的生物通路的可视化结果,无法提供全局的视图,这导致生物研究者无法把握数据分析过程中全局的观念。

针对上述问题,本文设计并实现了一个基于Web的关于生物通路的交互式可视化的分析系统——PathBubbles, PathBubbles系统最大的特点是把每一个窗口(功能视图)隐喻成一个气泡(Bubble),并把静态的多视图展示扩展到气泡隐喻界面。气泡之间不会重叠只会合组(Group)和离组(UnGroup),这样能够帮助用户方便且快速地切换视图,降低生物学家的记忆负担。针对生物通路的特点,该系统设计了生物通路图的可视化功能气泡, Sunburst 可视化功能气泡, 表格气泡, 该系统能够帮助生物学家探索和解释生物通路中的潜在的信息。具体的可视化功能如下:

(1) 生物通路图可视化气泡: PathBubbles 中的生物通路图不仅仅只是用节点和线来表示的网络图,而且创造性的在生物通路图中加入细胞区室的结构信息。这有利于生物学家找出生物通路中生物分子在细胞中的具体位置。另外,本文设计了一个在线编辑生物通路图的平台,方便生物学家编辑和改善生物通路图。

(2) Sunburst 可视化气泡: PathBubbles 以 Sunburst(环状层次图)的可视化形式整合 Reactome 现有的全部生物通路(417条)到一个 Bubble 中,并用颜色来映射生物学家感兴趣的信息,方便比较和查询。Sunburst 可视化能够给生物学家提供一个生物通路图的全局关系信息,具体的每一条通路信息可通过查询生成新的子 Sunburst 可视化。这种从全局到局部的数据选择和处理,与生物学家的数据分析过程一致。

(3) 交互式表格的可视化气泡：生物通路的具体信息用生物学家最为熟悉的图标形式展示。图标的查询和排序的交互式操作可以帮助提高分析效率。

综上所述，本文在可视化系统的气泡隐喻界面设计上，生物通路图的排列上，生物通路的可视化展现形式等方面取得了创新性进展。不同数据集之间的动态交互设计提高了生物学家对数据的探索和分析。

关键词： 生物通路，生物通路可视化，交互式可视化，界面设计

ABSTRACT

Due to the rapid advance in biological technique, lots of pathway data are generated. With the development of bioinformatics, a huge amount of biological experimental data is accumulated. As we know, human can interpret visual information rapidly, it is vital to describe the data in a comprehensive and interactive visualization manner. However, human has limited memory. Current visualization tools just provide a way to navigate and manipulate the data, mental workload can not be reduced. For example: most of data analysis process is various, which requires many different application windows to display different information processing result. In addition, although existing biological databases, such as Reactome and KEGG, provide an abundant of information, they lack a user-friendly and interactive visualization analysis interface.

In the views of those problems, we present PathBubbles, a web-based interactive biological pathway visualization system. A key aspect of PathBubbles is extending static multiple views to a metaphorical interface of Bubbles, each of which is treated as a functional unit view. Bubbles do not overlap, they can be grouped and ungrouped. This bubbles metaphor interaction interface will help users switch the views quickly and conveniently and reduce their mental workload, which can help biologists to explore and interpret the underlying information of the pathways. More specific introductions are described as below.

(1) Pathways visualization Bubble: In PathBubbles, pathway can be represented by a Node-link diagram, some compartment information is also be integrated. So we present a hybrid algorithm, which uses the combination of hierarchical pathway layout algorithm and force direct layout algorithm to arrange the pathway. Besides, an online pathway platform is also developed for biologists to edit and improve biological pathway.

(2) Sunburst visualization Bubbles: PathBubbles integrates all the existing pathways of Reactome in a single Sunburst, and uses color to map interesting information, which is convenient to compare, query and validate hypothesis. Sunburst can provide a global view in a first glance, detailed biomolecules of each pathway can also be visualized as a child-Sunburst. The data selection and processing way is from global to local, which is consistent with biological data analysis progress.

(3) Interactive Table visualization Bubble: Specific information of biological pathway are displayed in the form of a interactive table, which is the most familiar way for biologist to do data analysis.

In a word, we focus on the research on following aspects: design of interactive visualization system, automatic pathway diagram arrangement algorithm, and the visualization design of pathway analysis. The dynamic interaction between multi-view datasets visualization improves the data analysis and exploration.

Keywords: biological pathway, biological pathway visualization, interactive visualization, interface design

目 录

摘 要.....	I
ABSTRACT.....	III
目 录.....	V
图目录.....	VII
第一章 绪论.....	1
1.1 研究背景.....	1
1.1.1 生物通路.....	1
1.1.2 生物通路数据库.....	2
1.2 国内外研究现状.....	3
1.2.1 生物通路可视化.....	3
1.2.2 生物通路可视化工具.....	4
1.3 PATHBUBBLE 系统开发的必要性.....	6
1.4 本文工作与论文组织.....	7
第二章 PATHBUBBLES 系统设计.....	9
2.1 生物数据分析.....	9
2.2 需求分析.....	10
2.3 PATHBUBBLES 系统整体设计.....	11
2.3.1 PathBubbles 气泡隐喻界面的设计思想.....	11
2.3.2 PathBubbles Javascript 库的设计.....	12
2.3.3 PathBubbles 生物通路功能模块设计.....	14
2.4 本章小结.....	17
第三章 PATHBUBBLES 系统实现.....	18
3.1 相关技术介绍.....	18
3.2 界面设计的实现.....	19
3.2.1 虚拟桌面的实现.....	19
3.2.2 Bubble 之间交互算法的实现.....	19
3.3 功能模块的可视化实现.....	22
3.3.1 生物通路图可视化气泡.....	22
3.3.2 Sunburst 可视化气泡 以及交互式表格气泡.....	26

3.4 专家评估.....	31
3.4.1 可视化的有效性.....	31
3.4.2 系统交互设计的有效性.....	32
3.4.3 用户体验的总结.....	32
3.4.3 PathBubbles 系统可扩展性的建议.....	33
3.4 本章小结.....	33
第四章 在线排列和编辑生物通路图.....	34
4.1 设计思想.....	34
4.2 文档设计.....	35
4.2.1 生物通路图排列标准.....	35
4.2.2 排列生物通路图的一些指导准则.....	36
4.3 界面和交互设计.....	36
4.3.1 界面设计.....	36
4.3.2 交互设计.....	38
4.4 本章小结.....	38
第五章 总结与展望.....	39
5.1 工作内容总结.....	39
5.2 进一步工作的展望.....	40
致谢.....	42
参考文献.....	46
附录.....	50

图目录

图 1.1 Cytoscape 可视化结果.....	5
图 1.2 BioTapestry 界面.....	6
图 2.1 Codebubble, VisBubble 的界面.....	12
图 2.2 PathBubbles Javascript 库的结构设计.....	13
图 2.3 PathBubbles 场景树结构设计.....	14
图 2.4 PathBubbles 中生物通路可视化图的一些符号.....	16
图 2.5 Sunburst(环状层次机构图)可视化文件系统中的文件结构.....	17
图 3.1 PathBubbles 虚拟桌面的实现.....	19
图 3.2 PathBubbles 每个 Bubble 的菜单 Group 功能介绍.....	20
图 3.3 Path Bubbles 中 Bubble Group 算法流程图.....	20
图 3.4 PathBubbles 系统界面.....	22
图 3.5 改进的层次排列算法设计图.....	23
图 3.6 实时排列生物通路效果图.....	24
图 3.7 生物通路图在 Path Bubbles 系统中的可视化.....	25
图 3.8 Sunburst 显示 Reactome 的 1417 条生物通路信息.....	26
图 3.9 Sunburst 中生物通路的 Cross-talking 层次的切换.....	27
图 3.10 PathBubbles 中不同物种的生物通路分析.....	28
图 3.11 基因表达功能菜单.....	29
图 3.12 输入基因表达数据后的显示变化.....	30
图 3.13 基因数据的分析效果图.....	30
图 4.1 PathBubbles 中生物通路图质量的处理管线设计流程图.....	33
图 4.2 生物通路图的在线排列界面设计.....	37

第一章 绪论

随着生物技术的高速发展和广泛的生物研究，产生了大量的生物数据。和其他学科一样，生物学的研究也进入了一个采集数据的能力远远超过了将这些得到的数据转化为知识的时代。特别是高通量测序技术的产生，它可以提供大量的关于多个物种基因表达模式的信息。如何整合现有的数据；如何可视化这些数据；如何展示可视化结果；如何提高交互操作的有效性，从而帮助生物学家挖掘潜在的信息和探索数据间的关系，正变得越来越重要。在很多情况下，生物实验的目的主要是比较两个或者多个不同的生物状态，如疾病和控制，从而来帮助生物学家理解基因表达变化的原理。在很多的实例研究中，对基因，基因组，转录组和蛋白质数据的分析，必须在信号传导和新陈代谢的生物通路中进行，以提供可供测试的科学假设。人类可以迅速地解释视觉信息，那些可以帮助生物学家用综合的和层次的方式来解释这些数据的可视化方式将极具价值。然而，人类的记忆力是有限的。目前的可视化工具也仅仅只是为用户提供有限的可以用来分析大而复杂的数据的浏览和操纵的能力，因此，仍然不能有效地降低生物学家的记忆和工作负担。针对这些问题，本文在生物通路图的排列上，生物数据可视化，生物数据的可视化分析设计上，和交互式系统界面的设计上做了一些的研究工作。

1.1 研究背景

1.1.1 生物通路

生物通路(Pathway)是细胞内的生物分子(包括基因，基因产物和化合物等)经过一系列的生化级联反应来完成某一种生物过程^[1]。这样的生物通路可以加工生成某些新的生物分子，比如说基因，蛋白质，脂肪。生物通路在生物体中的作用是无可替代的，生物通路的变化(譬如：基因的差异表达)或多或少会对生物体产生某些影响。比如，某些生物通路是用来排毒的，某些会在受精卵变成胚胎的发育过程中起着至关重要的作用。因此，生物通路的正常工作与否在某种程度上预示着生物的健康程度^[2]。进一步说，如果某条生物通路出现问题，那么会导致相关的疾病，例如，细胞癌，肿瘤。

在生物体中，有许多种生物通路，按照反应的不同类型可以分为新陈代谢通路，基因调节通路，信号传递通路和蛋白质相互作用通路。

(1) 新陈代谢通路的主要功能是参与代谢的化学反应。

(2) 基因调节通路的主要功能是打开和关闭基因的表达。它能控制蛋白质和基因之间的相互作用。

(3) 信号传递通路的主要功能是参与信号传递的级联放大反应^[3]。

(4) 蛋白质相互作用通路的主要功能是蛋白质间或蛋白质——基因间的化学反应。

当前对生物通路的研究主要是研究生物分子在生物体中的功能行为表现、基因的表达情况以及各自发生的作用。生物研究者主要通过探究基因和蛋白质在实验中的表现情况为基础,来推断各种生物分子(基因,蛋白质,DNA,RNA,生物小分子等)之间的关系,揭示各种实验假设。尽管这不能揭示生物系统中的全部关系,但是展示了生物体中相关基因在生物通路中表现的局部现象,并反应了生物体里面生物分子间的关系,展示各种生命活动的规律。通过对生物通路的深入研究,生物研究者可以识别某条通路中的基因,蛋白质,和其他分子等,从而通过这些已有的线索来帮助疾病的诊断。通过对造成疾病的某条生物通路的识别和分析,能够进行更加专业的治疗和对某种疾病的预防。同时通过对病理的研究,也有助于新药的研发。

目前人体中所发现的许多重要的生物通路都是通过培养的细胞,细菌,果蝇,小鼠和其他生物体上做实验发现的^[2]。但是,许多潜在的生物通路还有待发现。通常生物学家需要花大量时间,去识别生物分子之间的复杂关系,理解各种生物通路是如何协同工作的,各种物种之间的生物通路的相似性,以及互相之间的功能变化。

1.1.2 生物通路数据库

随着对生物信息研究的继续,产生了大量的生物信息数据。这迫切需要数据库管理系统来存储、管理这些数据,同时也方便数据资源的共享。接下来选取一些常用的生物学数据库作一些简单的介绍。

(1) KEGG(Kyoto Encyclopedia of Genes and Genomes)^[4]: KEGG 是由日本京都大学的生物信息中心和东京大学的人类基因组中心联合开发的商用生物信息数据库。它是一个关于基因和酶促反应以及其他生物化学物质的在线数据库,它包含多个子数据库:关于生物分子信息的基因与化学信息数据库,功能信息与系统信息数据库和生物通路的通路数据库。KEGG 里面的生物通路图是人工绘制的,这很耗时,而且当出现新的通路,或一条通路中发现新的信息时,更新信息很费时间。最大的问题是它缺少一个用户友好的可编辑的交互分析界面。

(2) Reactome^[5,6]: Reactome 是一个由生物学专家撰写、编辑,并且经过同行评阅的关于人类生物通路的数据库。它包含新陈代谢通路,基因调节通

路, 信号传递通路和一些处理过程。它的数据和相关软件是开源共享, 并且免费使用的。目前还收录了其他物种的数据: 包括水稻, 果蝇, 鸡等等。通过 Reactome 官网(<http://www.reactome.org/>) 可以浏览生物通路, 也可使用其提供的特定的分析工具, 并提交实验数据。Reactome 数据库中的生物信息, 内容还综合多个其他数据库, 生物通路图虽然是人工绘制, 但是“专家撰写, 同行评审”的模式在某种程度上也提高了效率。

(3) BioCyc^[7]: BioCyc 是一个收集了 5500 种生物通路和基因组的数据库, 它提供一系列的软件。BioCyc 由一系列的数据库组成, 基于它的质量可分成三大类: 第一类包括 EcoCyc, MetaCyc, HumanCyc, AraCyc, YeastCyc, LeishCyc。这些包含的信息是最准确的, 因为需要经过至少一年的专家审阅。第二类是由 PathoLogic 程序产生的, 用来预测新陈代谢生物通路的。例如: vchocyc, shigellacyc, mtbrvcyc 等等, 也是经过一些专家审阅的。第三类也是由 PathoLogic 程序产生, 用来预测用于新陈代谢的通路, 不过由于计算量大, 并没有经过人为的审阅。在 BioCyc 官网(<http://biocyc.org/>), 可以利用它提供的一系列的可视化工具去浏览、可视化和分析基因组和生物通路信息, 利用官网提供的一系列的比较分析工具去探索和验证实验假设。

(4) Wikipathways^[8], Wikipathways 的主要设计思想扩展自 Wikipedia (维基百科), 采用“专家撰写, 同行评审的模式”, 并提供了生物通路信息的编辑, 生物通路图的编辑和可视化的相关工具。这样遍布在世界各地的生物通路专家可以利用该平台提供的工具进行生物通路的创建和编辑, 方便了对生物通路信息的更新和维护。Wikipathways 整合了现有一些数据库, 譬如 KEGG, Reactom 等里面的生物通路, 以一个开放的生物通路社区为基础, 方便了生物学家 (学生和老师) 进行交流和學習。

1.2 国内外研究现状

1.2.1 生物通路可视化

生物通路图(又称生物通路网络)是一系列的化合物和反应所组成的复杂网络的一种可视化表示, 经常以图的形式来表示^[9,10]。这样, 一条生物通路可以表示成节点和边的集合, 节点可以用来表示参与反应的反应物和产物, 例如: 蛋白质, 小分子化合物, DNA, RNA 等; 边可以表示各种相互作用关系, 例如: 反应和规则^[11]。可以说生物通路图的排列问题就是图的排列的问题。生物通路图以直观的网络图谱的方式来显示相关的生物学信息, 能够方便生物学家认清各种生物反应的网络图谱, 了解生物通路中反应生物分子之间的关系, 生物分子的功能作用, 以及各自的反应方式, 同时也能够方便生物学家对生物实

验数据进行观察,记录,分析。当前生物信息学领域中,如何清晰地可视化生物通路图谱,正变得越来越重要。

正如上节所知,许多生物数据库中的生物通路图都是由一些经过专业训练的生物学家来手动绘制的,因而非常耗时。许多生物学者也揭示了这种方式的不利之处,例如:费时且效率低,缺少数据探索的交互,编辑困难^[9,11-13]等等。伴随着生物数据的极具增长,手工绘制已经难以适应数据的增长速度。因此,近年来,生物研究者尝试应用各种图论的方法,来自动地排列组织生物通路图^[9,11-16],并附加一些交互手段来帮助探索生物通路数据。这些算法主要可以分成三大类:力导向算法,分层算法,网格排列算法。以下对这些算法作一个简单的介绍:

(1) 力导向(Force-Direct)算法:它是在生物通路排列中最为流行的算法之一。因为该算法可以产生很有吸引力的结果:节点之间均匀分布和更加少的边交叉。许多生物学家^[9,14-17]用它来产生图中节点位置的初时结果。力导向算法把节点看作是在物理系统中具有相互排斥力的粒子,并将边看作节点之间具有相互吸引的力。利用一个物理系统去找到一个能量最小的临界条件来确定系统平衡状态中的结点的位置。

(2) 层次排列算法:它是一种用来产生分层结构的排列算法。Itoh^[16]等人尝试提供一种结合的空间填充算法和力导向算法来可视化有多个种类关系的图,通过应用层次聚类算法和空间填充算法对节点进行归类,然后将某个特定种类的基因进行归类。Tsay^[18]等人提出采用分层的方法来组织可视化由许多子通路组成的生物通路,这样用层次关系来表明生物通路之间的关系。Barsky^[19]等人指出将节点排列在隔离的细胞区室(Compartment)里,用嵌套关系来表示层次关系需要额外的解释。因此建议把节点放在一层一层的结构上,每一层代表一个独立的细胞结构。

(3) 网格排列算法^[13]:该算法能够有效的组织节点,并产生整齐的排列。通过将节点放到网格上来避免重复,并且利用网络的拓扑结构关系来确定节点之间力的相互作用关系。不过主要问题是计算量大,耗时。基于这个算法,Kojima等人做了一系列的研究^[10,17]并提出了改进的SCCB-grid算法,利用力导向算法来产生初始结果,并且考虑各种生物信息属性(譬如子细胞结构信息)。然而,这个算法主要是针对信号传递通路和基因调节通路,并且层次结构关系被表示成嵌套关系。

1.2.2 生物通路可视化工具

为了方便生物学家可视化和分析生物数据,设计良好的生物数据可视化系统也是一个重要的内容。随着科技的发展,越来越多的生物学家利用可视化系

统来理解，分析生物数据和实验数据，这促使越来越多的研究者专注于可视化系统的研究。接下来对一些较为成熟的可视化工具作一些介绍。

(1) Cytoscape^[20]: Cytoscape 是一个关于生物通路信息的可视化和分析的开源平台。它通过生物分子之间的相互作用，并结合基因表达数据和其他的分子状态信息。它最大的作用是能够可视化人类和其他物种的大量的蛋白质-蛋白质，蛋白质-DNA，遗传交互信息。该软件的内核提供了一系列的数据分析和可视化功能，包括基本的网络排列，查询。它最大的特点是支持插件开发，支持主流的网络数据格式 (GML, PSI-MI, GraphML, KGML (KEGG XML), SBML, OBO, and Gene Association 等)，提供多种数据导出格式 (PDF, PS, SVG, PNG, JPEG, and BMP files 等)，集成较多的排列算法 (力导向排列算法，分层排列算法等)，并且提供多种方便开发的 API。

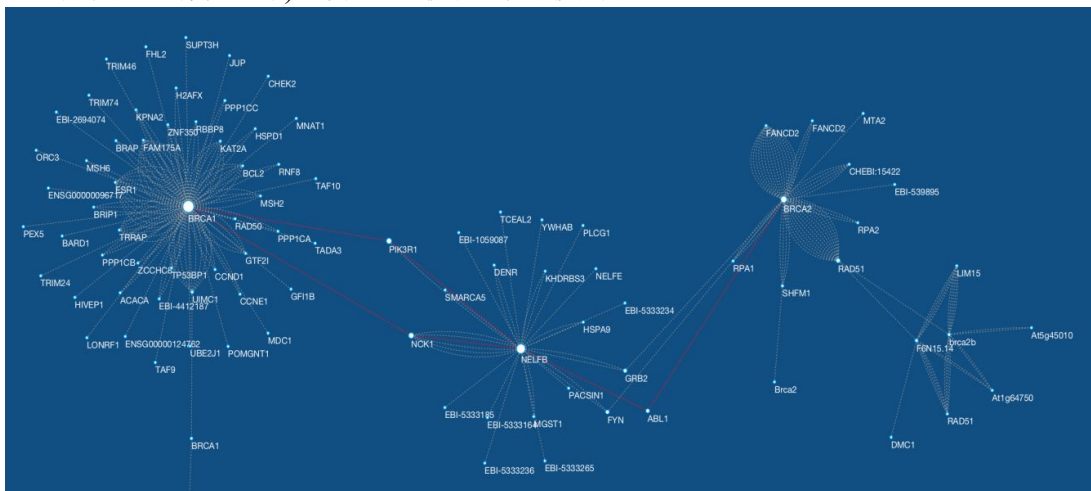


图 1.1 Cytoscape 可视化结果

(2) MEGA (Molecular Evolutionary Genetics Analysis)^[21]: MEGA 是一个执行基因序列的排列，推断进化树，估计差异时间，推断祖先序列的集成工具。许多生物学家利用它来重建物种的进化历史，推断基因选择在物种进化中的作用。

(3) BioTapestry^[22]: BioTapestry 是一个用于重构，可视化，模拟基因调节网络的开源的，免费的交互式系统。生物学研究者可以利用它来可视的动态的分析复合物的在基因调节网络中的时空分布。

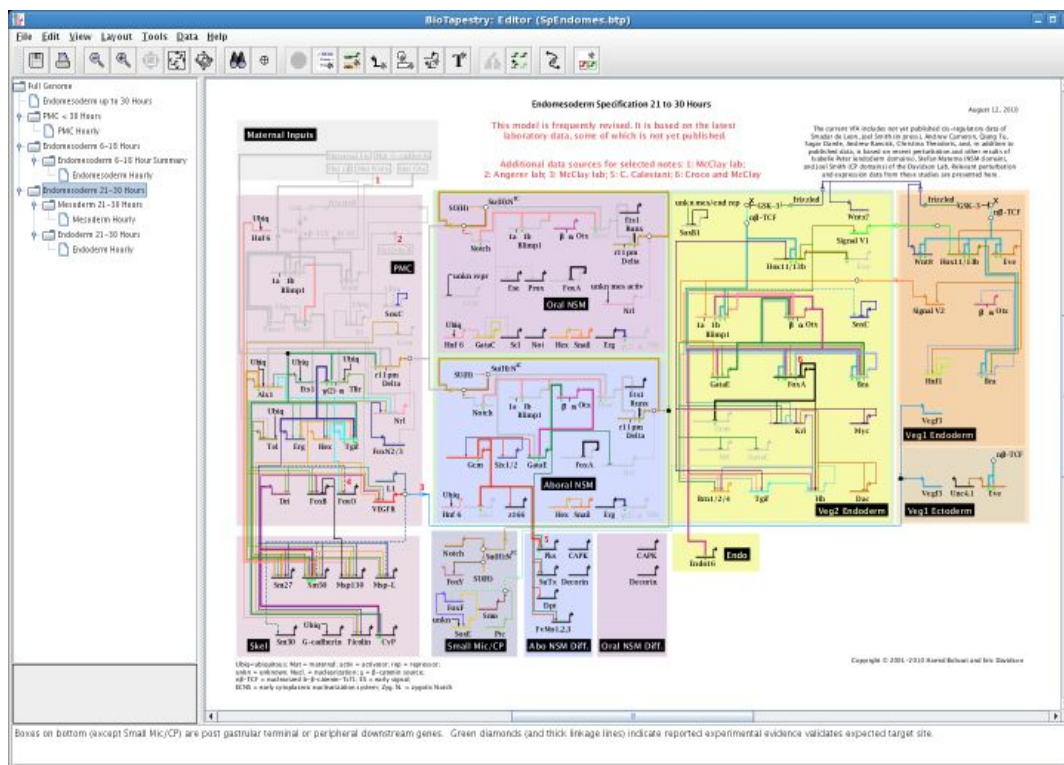


图 1.2 BioTapestry 界面

(4) GenMapp^[23]: GenMapp 是通过将一系列的基因映射到生物通路上来可视化基因表达数据。其官网还提供了基于 Cytoscape 的集成可视化与分析的软件 GenMAPP-CS，和显示和编辑生物通路的软件 PathVisio，通过 GenMAPP-CS 可以进行基因表达分析，通过 PathVisio 可以浏览和编辑生物通路图。

1.3 PathBubble 系统开发的必要性

当前生物学领域产生了大量的实验数据，这迫切需要一种合适的手段来分析这些数据。其中典型的情况是分析某物种在实验环境下的基因表达情况，从而在这里面找到一些有关联的模式，来帮助分析物种间的关系，物种间内部生物通路的关系，基因表达情况的关系。在许多情况下，生物学家会分析实验受体在实验条件和正常条件下基因的变化情况。前面提到的许多可视化的数据库和许多可视化工具^[4,5,20]都采用可视化的方式来显示这些数据，这样能够帮助生物学家数据的探索和实验假设的验证。

现有的许多工具都提供很强大的功能机制^[20,22,23]，主要的问题是要么缺少一个有用的用户友好的界面，要么很难做自定义基因表达数据在生物通路层面的分析。还有一些可视化工具，比如说 Lists2Networks^[24]，VisANT^[25]，BioVeen^[26]，CHIBE2^[27]，MIMO^[28]，and Kerfuffle^[29]等，也只是简单的列出数据集之间的关系，或者将基因表达情况直接显示在每条生物通路具体的生物分子上，这

样当解释分析实验数据的时候，缺少一个对基因表达情况在生物通路中的全局认识。其他还有一些工具也提供全局生物通路数据的可视化，可是缺少生物通路之间的关系的展示。例如，iPath^[30]虽然也提供基因调节通路的全局表达情况，但是项目的合作者(资深生物研究人员)认为，当解释生物通路的表达情况时，用户选择感兴趣的通路时，还是需要加入许多的人工干预，这样对实验数据的分析不是很方便。Cellular Overview^[31]也提供特定物种的新城代谢的全局生物通路图，但是和其他的工具一样，做基因表达情况的分析时，也是显示在每条生物通路的生物分子上，这样还是缺少全局的表达情况的对比和分析。许多的生物通路数据库，就像 Reactome^[5]生物通路尽管也是采用树型的结构来表示生物通路的层次关系，但是并没有提供一个全局的界面来显示基因的表达情况，这样在数据分析时，缺少了生物通路的上下文信息和全局的宏观把握。基于这样的背景分析，本文考虑开发一个提供全局的分析工具来帮助生物学家对实验数据进行分析。

由于计算机屏幕的大小限制，可视化结果的显示区域是有限的，而许多的软件都是将可视化的结果放在有限的固定窗口界面(例如：Cell Overview^[31], Reactome^[5])，这就导致许多的软件采用可放缩的用户界面来解决这种问题，但是问题是这样丢失了部分的上下文信息。当分析多个数据集时，又是很难记忆和比较的。本文采用 Bubble 的可扩展性界面，将当前工作视图缩小显示在导航栏上，这样方便了用户对所有工作数据的记忆。导航栏中工作视图之间的的滑动切换，方便了功能视图的管理。气泡之间的组合和拆分，方便了数据集之间的对比。

目前，生物通路图可视化工具主要分为两种：桌面安装版和基于 Web 版本。桌面安装版的特点是方便导入本地大规模数据，缺点是需要安装应用程序，不具备跨平台性；基于 Web 版本就较方便，不需要安装本地应用程序，例如，基于 Flash 版本仅依赖于 Flash 插件，基于 HTML5 完全不用安装就能浏览和分析数据。本文的 PathBubbles 的设计前提主要是基于 Web 的分析工具。

1.4 本文工作与论文组织

本文通过对生物学家关于生物通路的研究过程和分析流程进行了研究分析，并且对于生物通路图的可视化提出了一种混合的排列算法和一个在线人工优化的平台，开发设计了一个基于 Web 的交互式可视化分析系统 PathBubbles，它是一个关于生物通路研究的分析环境，集数据可视化，交互和分析于一体。在界面设计方面，PathBubbles 应用 VisBubbles 和 CodeBubbles 的思想，把每一个 Bubble 看作一个功能视图，并把静态的多视图展示扩展到气泡隐喻界面。气泡之间不会重叠只会 Group 和 UnGroup，这样能够帮助方便用户快速地切换功能视图。在生物数据方面，PathBubbles 以 Reactome 数据库为基

础，并整合 eGIFT，鸡等其他数据库，并且能够导入用户实验数据，帮助生物学家进行各类生物信息的比较分析，从而达到验证实验假设的目的。在可视化方面，PathBubbles 以 Sunburst(环状层次图)可视化的形式，整合 Reactome 现有的所有生物通路到一个 Sunburst 可视化气泡(Bubble)中，并用颜色来映射相关感兴趣的信息，方便比较和查询；以 Table(表格)的形式来展示每条生物通路的具体信息；以节点和线的图来表示每一条生物通路内部生物分子的相关信息。在界面设计方面，PathBubbles 提供可扩展性的 API，可整合 Canvas 和 SVG 并可以集成 D3.js，是一个托管于 Github 的开源的 Javascript 库。

在内容的具体安排上，首先，在第一章，对生物通路的基本概念以及研究生物通路的作用，以及近年来对于生物通路的可视化，生物通路可视化分析软件的必要性，关于集成生物通路可视化软件作了相关的介绍。第二章，分析了当前生物数据形式，基于对生物合作者的面谈，对系统的设计做了一些需求分析。之后针对需求分析以及设计目的介绍了 PathBubbles 的系统设计，针对 PathBubbles 基于 Bubble 的界面设计的可扩展性介绍了如何设计 PathBubbles Javascript 库，然后针对需求的相关功能分析，设计了相关生物通路可视化。第三章，首先对 Ajax，Mysql，PhP，Canvas，SVG 等相关技术，数据可视化库 D3.js 作了简单介绍，然后介绍了气泡隐喻界面的设计实现，以及对于气泡的 Group 和 Ungroup 算法的作了简单介绍，之后介绍了系统的主要功能模块的可视化设计实现，以及生物学家对本系统的相关测评结果和建议。第四章，基于本文提出的生物通路排列算法和生物研究者对系统评估后的建议，本文提出一种采用众包(Crowdsourcing)的方式来完善和编辑 PathBubbles 中的生物通路图。这样可以利用众人的力量来排列生物通路图，并且可以把排列后的结果输入到 PathBubbles 程序中去。最后在第五章作了全文工作的总结，并对未来可扩展性的工作作出了相关的展望。

第二章 PathBubbles 系统设计

本文的交互式可视化工具 PathBubbles 应用于对生物通路数据的分析和研究。在开发设计 PathBubbles 气泡隐喻界面中所产生的 Javascript 库将会成为用于基于 Bubble 的界面设计的开源框架。本章首先对生物数据做了简单的分析，对生物学家研究生物数据做了深入的需求分析，最后做了 PathBubbles 的系统设计。

2.1 生物数据分析

PathBubbles 系统的数据下载自 Reactome 数据库，它是表示人类的生物通路数据，之后被转成一种适应 PathBubbles^[32] 桌面版程序的特定的格式^[33]。Reactome 把每一条生物通路最基本的单元看作一个反应^[6]。一个反应是任何将输入转成输出的过程，它可以用有向边来表示，它的输入和输出代表生物分子。例如：蛋白质、复合物、小分子、Dna 和 Rna 等等。另外，除了输入和输出，一个反应在生物通路图中也可以包含其他信息，例如细胞的位置信息(或者细胞区室(Compartment))。如果一个反应的输入和输出来自两个不同的细胞区室(Compartment)，这就意味着这两个细胞区室之间是有关系的，也可以说在这个数据中有两种类型的层次关系信息。

(1) 对于一些复杂的生物通路，Reactome 会切分成一些子生物通路，根据需要这些子生物通路还可以被切分成更小的子生物通路。这是其中的一种层次关系：生物通路(Pathway)和子生物通路(SubPathway)。

(2) 对于每一条特定的生物通路，里面可能包含许多个细胞区室(Compartment)，而每一个细胞区室(Compartment)也会包含一些生物分子。所以这里包含另外一种层次关系，生物通路(Pathway)，细胞区室(Compartment)，生物分子(Biomolecules)。

如果不同生物通路之间有共享蛋白质基因，那么可以认为这些生物通路之间是具有 Cross-talking 的关系。Cross-talking 的蛋白质是表明一条生物通路对另外一条生物通路有影响的重要标志^[34]。

特拉华大学的生物合作者提供了鸡的基因信息表，鳄鱼的基因信息表和乌龟的基因信息表等数据，表中包含了蛋白质的相关信息。PathBubbles 系统以 Reactome 中人类生物通路的数据为基础，其他物种基因信息表中的一系列基因符号(蛋白质)，与之进行对比，可以检测不同物种之间的关系，以及每一条生物通路包含其他物种的基因的关系。那里的生物学家还根据特定的物种做了一些基因表达相关的研究实验，然后得到一些关于实验环境下的基因表达的实验数据表。通过对这些实验条件下数据的分析，可以研究特定物种中的哪些生物

通路中的基因在实验环境中被正调控了，哪些被负调控了，哪些没有多大改变。

针对以上数据的分析，本文的可视化设计需要考虑到以下几点：

- (1) 保持生物通路和子通路之间的层次关系。
- (2) 保持每一条生物通路内部的细胞区室(Compartment)，以及里面的生物分子的层次关系。
- (3) 保持 Reactome 数据库中每条生物通路的核心关系，也就是强调每个反应的输入和输出关系。
- (4) 区分输入和输出中不同的生物分子。
- (5) 表明不同的生物通路之间共享蛋白质的 Cross-talking 关系。
- (6) 能方便比较不同物种之间所共享的同祖基因信息。
- (7) 能方便查询和比较哪些基因被表达。

2.2 需求分析

为了能够帮助生物学家完成特定的数据分析任务，本文作了相关的需求分析，并以此来指导 PathBubbles 程序的功能设计，主要通过采用交互的方法不断地用户交流，保证系统的实用性，并为用户提供可用的功能。

首先，我们与合作者作了深入的沟通和交流，了解到他们的实际需求以及数据分析的过程和目的。主要采用面谈和问卷调查两种方式，面谈对象是特拉华大学的合作者：Carl Schmidt 教授和他的同事，他们是研究鸡的生物通路的专家(Reactome 数据库的一个分支)。然后根据面谈的结果做了一份详细的需求相关的问卷调查，了解哪些核心问题是生物研究者目前较为关心的。之后将问卷调查的结果和生物合作者进行确认，之后在询问相关的生物系的学生和一些生物学家。

根据最后的分析，可以发现生物学家目前最为关心的是四类需求 (这与前面的数据分析一致)：(1)生物通路图，(2)不同物种之间的比较 (同祖关系)，(3)不同生物通路之间的关系(Cross-talking)，(4)某些物种的实验数据结果分析(例如，在某种实验条件下的基因表达情况)

由于 Reactome 生物通路的层次结构简洁明了，PathBubbles 系统的人类生物通路数据也来自 Reactome，所以程序的可视化设计展现应该以 Reactome 为基础，尽量保留生物通路的层次结构关系。根据前期的面谈结果和之后对合作者提供数据分析的结果，本文做了一份关于生物学家进行生物通路相关的研究所关心主要问题的一个需求调查表，并根据需求做了需求类别归类的设计：生物通路图，Cross-talking，同祖关系，基因表达^[35]。

表 2.1 关于生物学家研究生物通路的一些调查问题

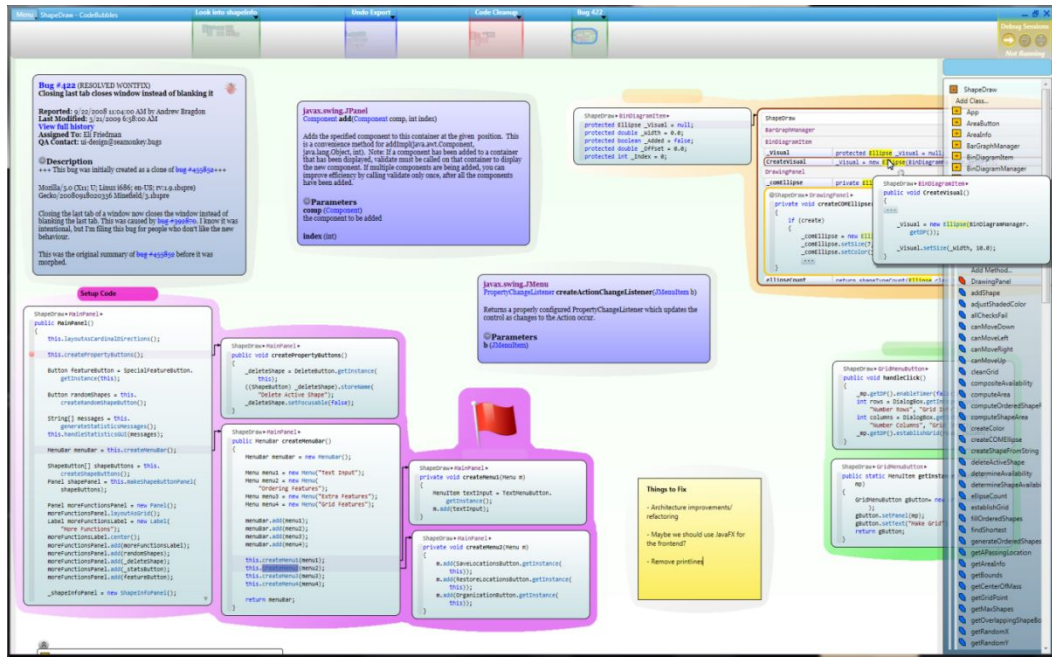
问题		需求类别
Q1	生物通路中包含 Cross-talking 的蛋白质的数量?	Cross-talking
Q2	在一条生物通路中的生物分子是否被表达了?	生物通路图
Q3	人类中的生物通路中的某条生物通路包含其他物种的基因的完整程度?	同祖关系
Q4	哪些蛋白质被正表达了?	基因表达
Q5	连接两个蛋白质之间的通路是什么?	生物通路图
Q6	两个细胞区室(Compartment)之间有共享的蛋白质吗?	基因表达
Q7	不同生物通路之间有共享的蛋白质吗?	Cross-talking
Q8	某种蛋白质的 Cross-talking 的生物通路的数量?	Cross-talking
Q9	同一物种不同生物通路之间的表达程度?	基因表达
Q10	两个蛋白质之间的可能的路径是什么?	生物通路图

2.3 PathBubbles 系统整体设计

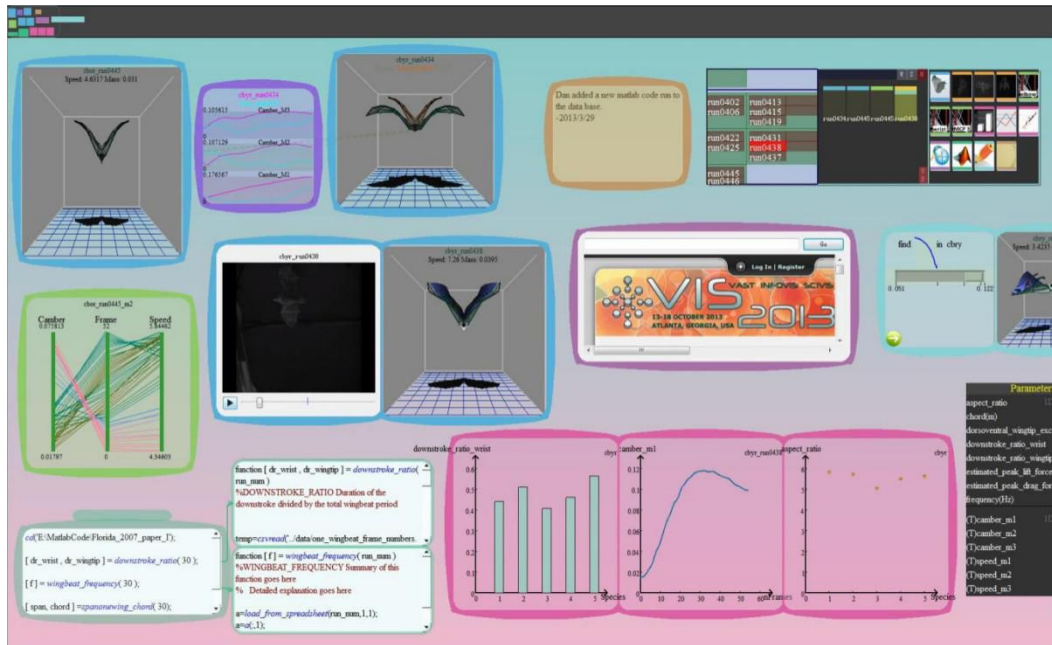
本文 PathBubbles 是一个基于 Web 的集可视化, 交互和数据分析于一体的系统。目前本系统侧重生物数据的分析, 但是为了实现系统的可扩展性, 可以在框架搭建时设计一个设计良好的接口, 这样就可以将 PathBubbles 的气泡隐喻界面设计思想扩展用于其他交互式 Web 系统的开发。本小节将会展开讨论三部分问题的设计: (1)PathBubbles 的气泡隐喻界面设计思想, (2)PathBubbles 界面的可扩展性基础库(Javascript 库)的设计, (3)生物通路功能模块的设计

2.3.1 PathBubbles 气泡隐喻界面的设计思想

气泡隐喻界面的设计扩展自 VisBubbles^[36], CodeBubbles^[37] (见图 2.1), 主要将气泡(Bubble)看作是一个关于可视化的功能单元或者功能视图, 在气泡里可以创建特定的可视化, 也可以展示各种可视化结果。与传统的基于 Window 的界面窗口不同点在于, 气泡之间不会重叠, 只会 Group 和 UnGroup。通过将多个功能气泡放在一个界面中, 方便信息的查找, 并简化窗口切换, 减少对许多窗口功能的记忆。将类似功能的 Bubble 放到一个 Group 方便比较和移动。另外重要的一点是虚拟桌面空间设计思想: 这种桌面将传统的桌面扩展成可保存当前工作状态的支持“多窗口”的视图, 在顶部的导航栏中, 可以显示所有工作状态的总体视图, 并根据相关的视图进行来回的切换。顶部导航栏中的视图是每一个当前窗口视图的按比例缩放, 这样有利于用户对当前工作状态的记忆, 方便工作状态的来回切换。



(a) CodeBubbles 界面



(b) visBubbles 界面

图 2.1 CodeBubble, VisBubble 的界面

2.3.2 PathBubbles Javascript 库的设计

良好的结构组织对一个系统的设计有着深远的影响。PathBubbles 是基于 HTML5 并结合 Canvas 和 SVG 的 2D 绘图的 Javascript 库。目前在 HTML5 中较为流行的一个 3D 开源的 Javascript 库是 Three.js [38], 许多 PathBubbles [32] 中的设计思想(譬如, 场景树的设计, 渲染的流程)都来自对 Three.js 源码的解读,

并且 PathBubbles 也可方便的集成 Three.js 在未来做一些 3D 相关的渲染工作。以下对整个库的结构作一个说明。(见图 2. 2)

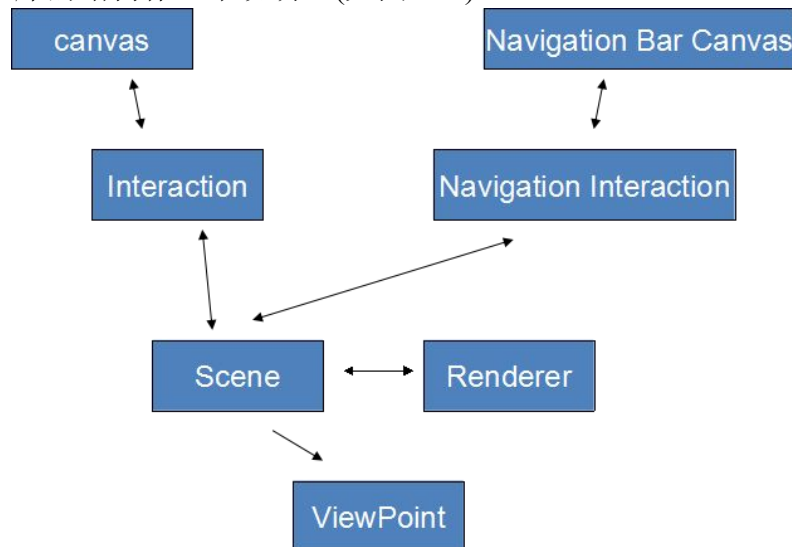


图 2. 2 PathBubbles javascript 库的结构设计

Canvas, Navigation Bar(Canvas):两个用于绘制图形(场景中的物体)的画布。对于 Sunburst 和 Table 的可视化设计, 主要利用 D3.js 库, 所以在 PathBubbles 的绘制中应该整合 SVG, 主要通过将 SVG 元素封装到 Div 中, 并放到 Canvas 上面进行渲染。

Scene:(图 2. 3)可以将整个场景看作一个场景树来管理, 所以在场景中建立的物体都继承 Object2D 这个对象(见代码 2. 1)。它包含一些最为基本的信息: 当前物体在整个场景中的 Id, 当前物体的名字, 它的孩子数组, 它在场景相对于其父亲的相对偏移位置, 如果没有父亲节点那都是 0。整个场景也继承自 Object2d 这个对象, 所有他的子对象都放到它的孩子数组里面。场景的作用是用来管理所有在场景中建立的物体的相对位置和保留一些基本信息。

代码 2. 1

```

PATHBUBBLES.Object2D = function () {
  this.id = PATHBUBBLES.Object2DIdCount++;
  this.name = "";
  this.parent = undefined;
  this.children = [];
  this.offsetX = 0;
  this.offsetY = 0;
  this.x = 0;
  this.y = 0;
};
  
```

Renderer: 用来绘制场景中的所有物体到 Canvas 上, 并按相同比例缩放渲染到导航栏的 Navigation Bar Canvas 上。

ViewPoint: 用来管理场景中的当前视口在导航栏中的位置。

Interaction: 用来管理 Canvas 上场景中的物体的交互。

Navigation Interaction: 用来管理导航栏上场景中的物体的交互。

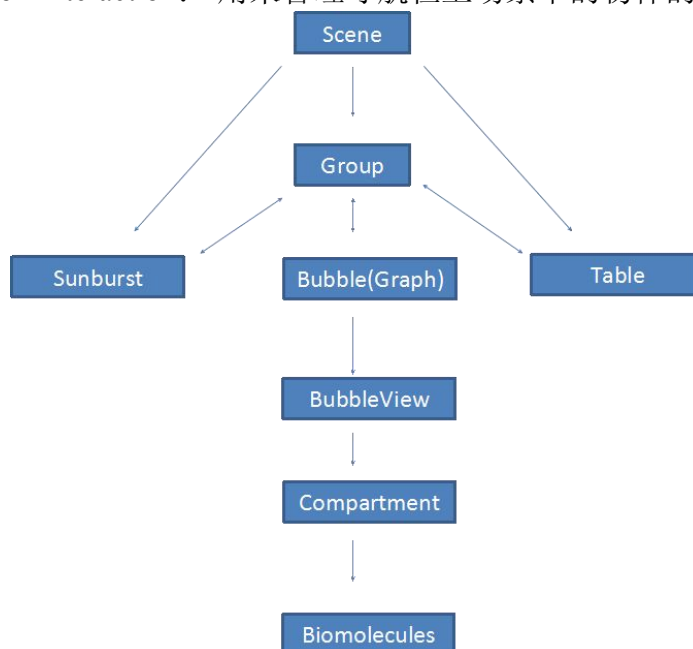


图 2.3 PathBubbles 场景树结构设计

根据以上 PathBubbles Javascript 库的整体结构设计，关于新建一个场景的设计可以用如下示例代码(代码 2. 2)

代码 2. 2

```
canvas = $("#bgCanvas")[0]; //获取 canvas
navCanvas = $("#navCanvas")[0]; //获取 navigation bar canvas
canvas.width = window.innerWidth; //canvas 宽度
canvas.height = window.innerHeight; //canvas 高度设置
navCanvas.height = 50; //导航条高度
navCanvas.width = window.innerWidth; //导航栏宽度
scene = new PATHBUBBLES.Scene(); //新建 scene
viewpoint = new PATHBUBBLES.ViewPoint(); //新建当前导航栏中 viewpoint
viewpoint.h = 50; //viewpoint 高度
viewpoint.w = window.innerWidth / window.innerHeight * 50; //宽度缩放比*50
var renderer = new PATHBUBBLES.Renderer(); //新建渲染器
navIntercection = new PATHBUBBLES.NavInteraction(renderer); //导航栏交互
interaction = new PATHBUBBLES.Interaction(renderer); //场景交互
function render() { //自动刷新渲染函数
    requestAnimationFrame(render);
    renderer.render();
}
render(); //调用渲染
```

2.3.3 PathBubbles 生物通路功能模块设计

根据前期的需求分析和数据分析, 将当前 PathBubbles 系统中的 Bubble 功能视图模块分成三类: 生物通路图的可视化, Sunburst(环状层次图)可视化, Table(表格)的交互式可视化。以下分别对这些功能模块作相关的介绍:

(1) 生物通路图的可视化气泡

前面提到生物通路图是一种以图的形式可视化的生物通路, 主要通过将生物分子以及相关的反应映射到图中来反应这些关系。为了考虑需求分析和数据分析中的相关需求, 作如下几方面的设计。(见图 2.4)

1) 节点: 在生物通路图中, 每个节点代表一种生物分子。图中的节点有许多相关的属性可供选择设置, 例如: 形状, 背景颜色, 大小, 边框颜色, 边框层次, 边框的粗细等等。在 PathBubbles 中的生物通路图中, 主要采用形状、颜色和大小来区分不同的生物分子, 用边框颜色和层次来表示当前生物通路中的蛋白质在他的子生物通路中出现, 然后把属于同一种子通路的生物分子画上相同的颜色边框。

2) 边: 在生物通路图中, 用边来表示各种相互作用关系, 反映类型, 生物学功能等等。边可分为有向边和无向边。边也可以采用不同的形状, 附加采用不同的颜色、粗细。本系统主要强调每一个反应的输入和输出关系, 所以系统中的边采用有向边。对于复杂的网络, 传统的箭头很难区分方向, 所以应该采用锥形箭头(Tapered arrow)^[39]来表示方向。由于有时两个结点之间有可能有多种关系, 需要将其中的生物学功能关系(箭头)表示成由渐变圆点组成的虚线, 这样当箭头重叠起来时, 还能看到下面的箭头被覆盖的箭头类型。

3) 图的排列算法: 项目合作者(特拉华大学的生物学家)特别提到要将细胞区室(Compartment)隔离出来。而目前为止现有的排列算法并不能满足合作者的需求。所以本系统设计了一种结合层次排列算法和力导向算法的混合排列算法。主要通过层次排列算法来排列细胞区室(Compartment), 用力导向算法来排列细胞区室(Compartment)内部的生物分子。具体会在第四章实现部分讲到。

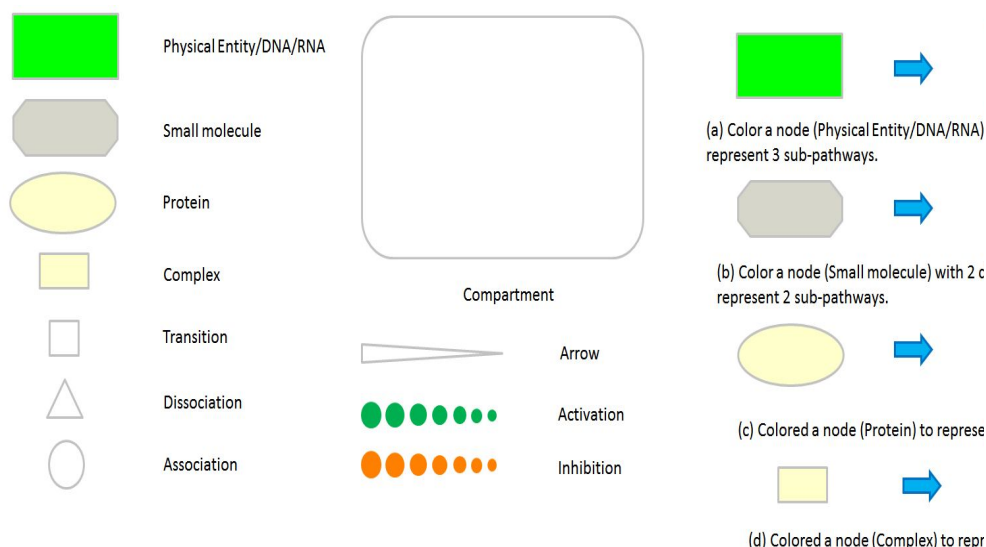


图 2.4 PathBubbles 中生物通路可视化图的一些符号

(2) Sunburst^[40]可视化气泡

层次结构最直观的表达是树结构，用节点和连线来表示父子关系。但是当父子关系中的节点与其他节点有联系时，很难再用线来表示。而且在有限节点上附加其他信息也是有限的。层次关系还可以用 TreeMap 来表示，尽管每一个区域之间，可以用线连起来表示节点之间的联系，但是层次关系并不是很明确。圆环径向排列的方式(Sunburst)，也可以表示层次结构(图 2.5 Sunburst 可视化文件结构^[40])，在圆环上也能映射其他信息，圆环上的 Arc 之间的关系可以用线连接。具体的来说，在 Sunburst 中，每一个孩子节点是用 Sunburst 中的一个 Arc 来表示，孩子节点是父节点 Arc 的细分，角度与孩子的数量有关，深层的子节点画在远离圆心的地方。在每一个 Arc 上可以用相关的颜色来表示其他信息，还可以在 Arc 画其他形状来表示其他信息。

对于生物通路数据的层次信息，Sunburst 是一个很好的可视化选择(每一个 Arc，表示一条生物通路)；对于不同物种间生物通路的同祖信息，可以在每一个 Arc 上用颜色来表示(完全，不完全，空的)；对于不同通路之间的 Cross talking 关系，可以用一些经过 Cluster 的线来表示；对于在同一物种上作相关的基因表达研究分析时，可以用颜色来表示每条生物通路上基因的表达情况；对于具体每一层中不同生物通路中，物种间每条生物通路相同的蛋白质的个数多少或者在做基因表达分析时，对于被表达的基因的数量，可以分别在 Sunburst 的 Arc 上画柱状图表示。

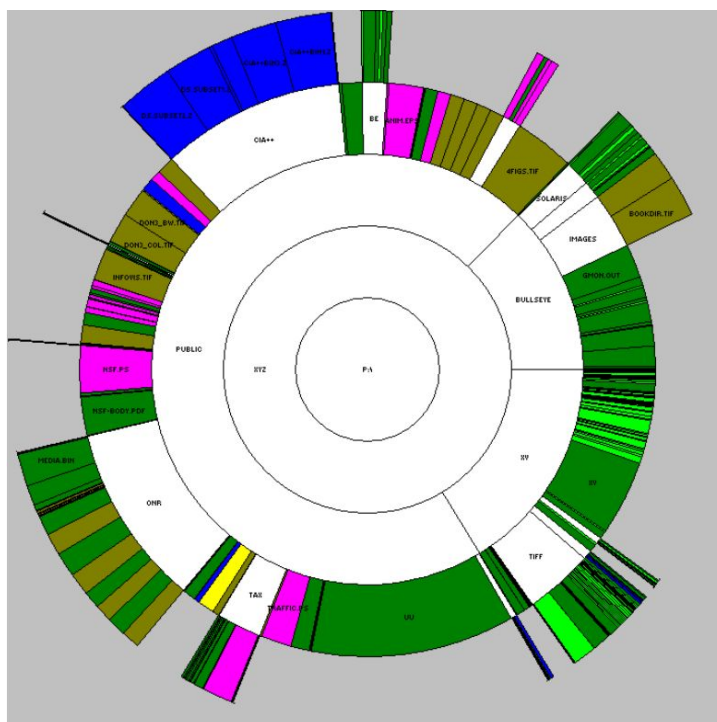


图 2.5 Sunburst(环状层次机构图)可视化文件系统中的文件结构

(3) Table(表格)的交互式可视化气泡

在早期的交谈中，了解到生物研究人员比较倾向于用 Table 来显示原始信息，因为研究人员已经习惯用 Table 做简单的比较。由于 Sunburst 可视化是一个分层结构，每一个 Arc 都可以包含相关的信息，所以每一个 Arc 是可以搜索的，当关心原始数据时，就可以用 Table 的可视化来显示这个生物通路相关的原始信息。这个 Table 应该是可以交互的，例如，点击 Table 的表头，可以排序。这种设计能方便对变化大的数据的比较、查找和分析。

2.4 本章小结

本章首先讨论了 PathBubbles 系统中生物数据相关的分析，并针对生物合作者对生物通路的研究分析过程和要求作了相关的需求分析。之后介绍了 PathBubbles 气泡隐喻界面的设计思想，PathBubbles Javascript 库的设计思想以及其可扩展性方面的考虑。最后针对生物学家四个方面的需求提出了三种可视化设计。

第三章 PathBubbles 系统实现

根据第二章的系统设计，本文实现了交互式生物通路可视化系统。本章主要对实现 PathBubbles 系统的相关技术作了简单介绍；对于界面的实现，和主要功能模块的实现做了相关的介绍。

3.1 相关技术介绍

PathBubble 是基于 Web 的，在 HTML5 Canvas 上绘制图形，利用 Ajax 进行后台数据的读取，用 MySQL 数据库管理生物通路数据，用 PHP 进行数据存储以及与后台数据库之间的交互的集生物通路可视化与分析于一体的可视化系统。PathBubbles 主要用 Javascript 操纵数据的处理交互，以及图形的绘制。在数据可视化方面，PathBubbles 可以利用 2D Canvas 自定义设计可视化，也可集成 D3 做相关的可视化操作。

Canvas 是 HTML5 发布的新特性，它的主要的特点就是可以直接利用客户机的图形资源，创建 2D，3D 的图形。它需要包含 Canvas 标签，并设置 Width 和 Height 属性。利用 Javascript 调用 HTML5 Canvas API，可以在 Canvas 上面渲染想要的场景。PathBubbles 基本的框架是利用 Canvas API 绘制的。

Ajax 全称是异步 Javascript 和 XML。利用 Ajax，web 应用程序可以在不需要终端当前页面的显示的情况下，发送数据给服务器，也可以从服务器获取数据。通过 Ajax，可以操纵 Text，XML 和 Json 等数据格式。PathBubbles 中的可视化数据是由 Ajax 获取的。

MySQL 是一个广泛使用的开源的关系型数据库管理系统。使用数据库能够方便的管理和操纵(查询)数据。PathBubbles 利用数据库管理原始的各种生物通路的信息，方便后期的查询。对于前期的可视化，主要使用 Json 和 XML 数据格式。

PHP (Hypertext Preprocessor) 全称是超文本预处理器，它是一种开放源代码的脚本语言。它可以嵌入 HTML 将静态的网页动态化。PHP 是一种运行于服务器的脚本语言，具有强大的跨平台性。PathBubbles 应用 PHP 来进行与数据库之间的交互。

D3^[41] 全称是 Data-Driven Documents。Data 表示用户提供的数据，Documents 表示任何能够被浏览器渲染的元素包括 HTML 和 SVG。D3 是连接数据和渲染的桥梁^[42]。D3 是一个包含多种可视化技术的库。PathBubbles 提供

连接 D3 的接口，这样方便人们利用已有的可视化技术做进一步的数据分析设计。

3.2 界面设计的实现

气泡隐喻界面是根据 VisBubbles 和 CodeBubbles 的界面设计思想扩展产生的。在场景中每一个 Bubble 都是有边界的(包括 Group)，Bubble 之间不会重叠只会 Group。所以当有一个 Bubble 出现在另外 Bubble 上面时，应该考虑 Group 的问题。当一个 Bubble 插入到一组 Bubble 中去时，需要考虑插入位置的问题。当从同一种 Bubble 中新建的一个子 Bubble 时，应该放在一个组。本节首先介绍虚拟桌面界面的实现以及 Bubble 之间关系切换算法的实现。

3.2.1 虚拟桌面的实现

前面提到在 PathBubbles 中，用到两个 Canvas：整个场景的 Canvas 和导航栏上面的 Canvas。导航栏上面的 Canvas 和场景的 Canvas 宽度一致，只是导航栏上面的 Canvas 高度固定是 50。那么导航栏中当前窗口大小就按比例缩： $\text{viewpoint.w} = \text{window.innerWidth} / \text{window.innerHeight} * 50$ 。同时也可以从这里看出虚拟屏幕空间尽管扩大了当前屏幕空间但并不是无限的。在导航栏用户可以随意的切换视口来选择当前感兴趣的工作区间。(见图 3.1)

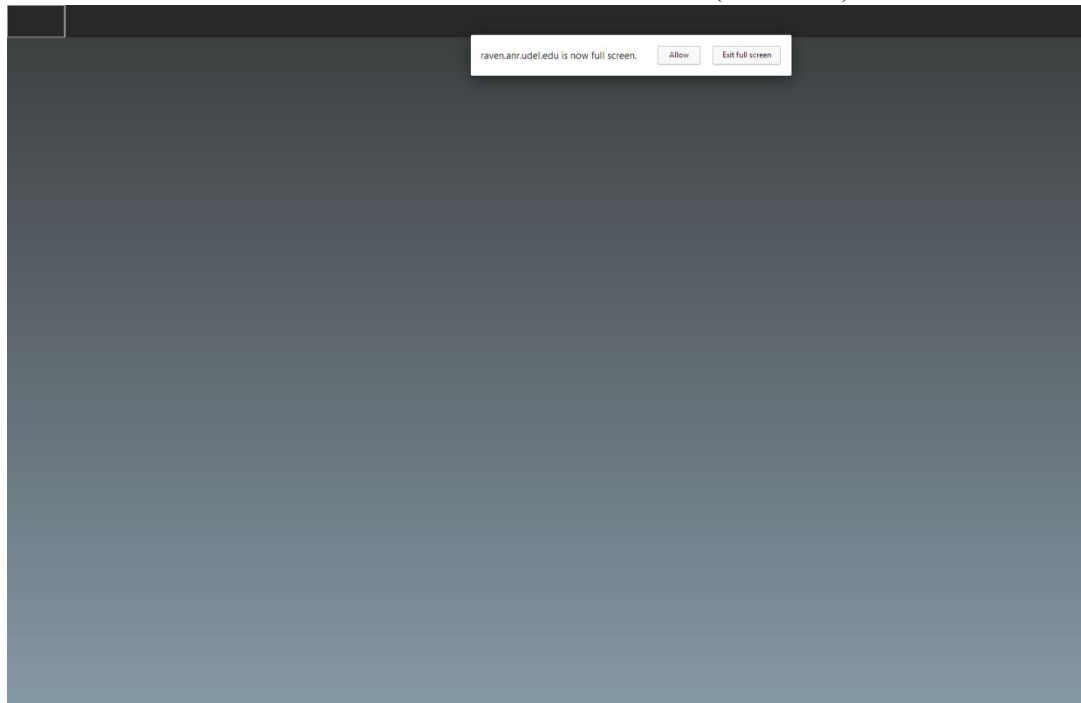


图 3.1 PathBubbles 虚拟桌面的实现

3.2.2 Bubble 之间交互算法的实现

在 PathBubbles 中，每个 Bubble 右上角菜单栏中(见图 3.2)有三个键：分别是 U/G，X，M，默认值是 U。当显示为 U 时，(见图 3.2 C)表示按键功能是

ungroup, 该 Bubble 是可以被 Group 的。当显示为 G 时表示按键功能是 group (见图 3.2 A, B), 表示该 Bubble 当前是不能被 group, 它放在其他 Bubble 上面, 是不能与其他 Bubble Group 的(见图 3.2 D)。在前期的交谈中, 了解到生物研究者更倾向与将所有 Bubble 往右放, 新建的 Bubble 放在当前 Bubble 的右边。所以对于 Bubble 的 Group 算法的设计作如下流程考虑(见图 3.3)。在此, 需要明确的一点是: Bubble 之间的 Group 是 Bubble 之间的重叠引发的, 如果两个 Bubble 不重叠, 那么就不会 Group。

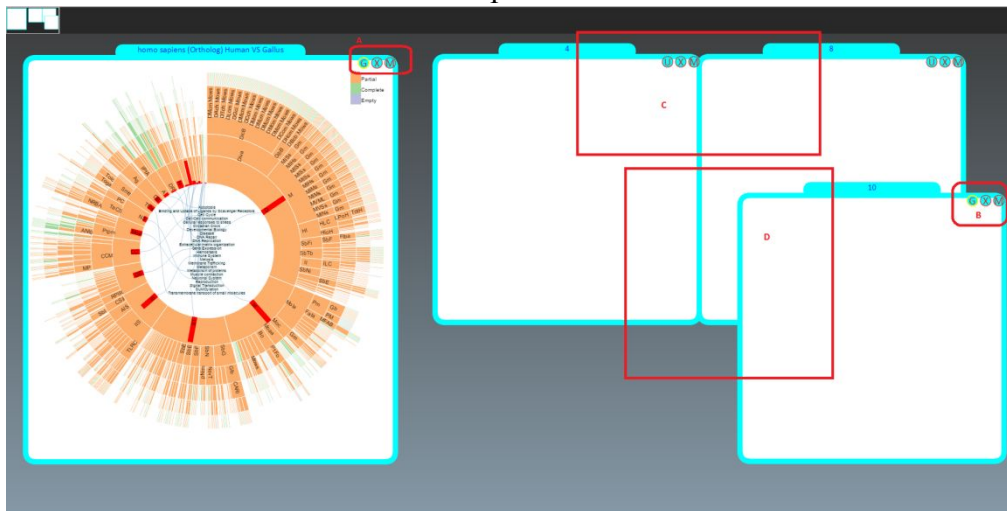


图 3.2 PathBubbles 每个 Bubble 的菜单 group 功能介绍

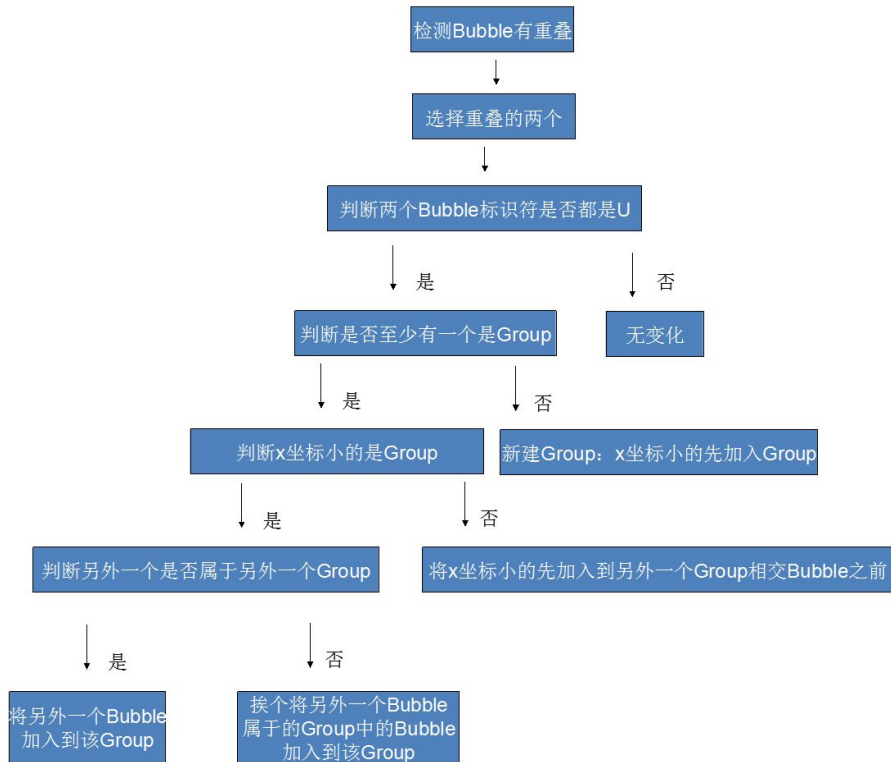


图 3.3 Path Bubbles 中 Bubble Group 算法流程图

Group (组合): 场景的 Interaction 会自动检测 Bubble 是否有重叠。当检测到有气泡之间有重叠时，首先任意选择重叠的两个 Bubble，并检测各自的 Group 标识符，如果是 U，那么被自动 Group，否则标识符是 G，那么是不会被 Group 的。当标识符都是 U 时，需要判断是否两个都各自有 Group，Interaction 会自动检测物体的 GROUP 属性以及 parent 属性。如果都没有 Group，那么新建一个 Group Bubble 放到场景中，并依次按 x 由小到大将这两个重叠的 Bubble 加入到 Group 中去。如果其中有一个是 Group，另外一个也是 Group，那么 x 大的那个 Group 中的所有 Bubble 依次插入到 x 小的那个 Group 中去。如果其中一个是 Group 另外一个不是，那么需要判断判断 x 小的那个是否是 Group，如果是的话，将另外一个 Bubble 加入到该 Group 中去；如果不是的话，那么将该 Bubble 加入到另外 Bubble Group 中。这样 Group 中的 Bubble 在数组 children 中的顺序正好是按 x 从小到大排列的。在 Group 中当有新元素插入时，自动将从 children 数组中第一个 Bubble 开始按宽度加上第一个坐标位置从左到右排列(见代码 3.1 refresh 函数)。那么根据以上规则，当新建一个 Bubble 时，可以将该 Bubble 放在当前 Bubble 的位置加上当前 Bubble 的宽度减 10 的位置，也就是制造新 Bubble 与当前 Bubble 重叠，那么它会自动放在当前 Bubble 的右边。

UnGroup(分离): 当鼠标点击已经 Group 的 Bubble 的 U 时会触发 Group 中 children 数组中 Bubble 的减少。当鼠标点击没有 Group 的 Bubble 的 U 时, 并不会触发当 children 数组元素一有变化会触发 children 数据中的 Bubble 按 x 由小到大排列。(见代码 3.1 refresh 函数)

代码 3.1

```
function refresh () {  
  for (var i = 0; i < this.children.length - 1; ++i) {  
    this.children[i + 1].x = this.children[i].x + this.children[i].w;  
    this.children[i + 1].y = this.children[i].y;  
  }  
}
```

3.3 功能模块的可视化实现

根据上一章的 PathBubbles 系统设计思想, 本节首先介绍生物通路图的生成算法, 之后介绍几个功能模块的实现。图 3.4 是交互式可视化分析系统 PathBubbles 实现的一个界面。

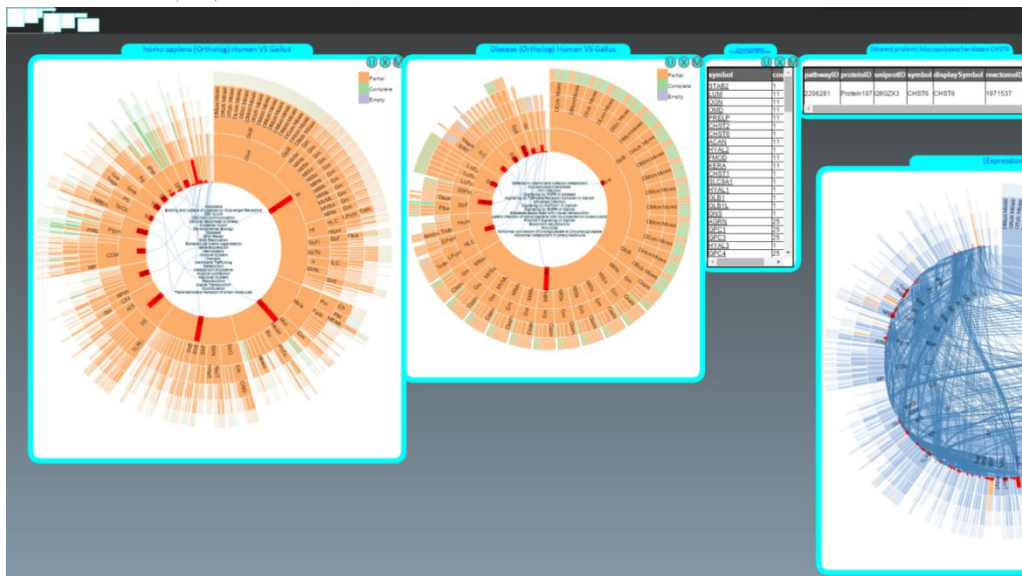


图 3.4 PathBubbles 系统界面

3.3.1 生物通路图可视化气泡

为了保持每一条生物通路的层次信息, 并且明确表示其内含的细胞区室信息, 本文设计并提出了一种混合的排列算法: 用层次排列算法来排列生物通路中的细胞区室(Compartment), 用力导向算法来排列细胞区室(Compartment)内部的生物分子。

层次排列算法: 为了保持合适的节点密度, 程序自动设定细胞区室(Compartment)的大小与其内部节点数成正比。当排列细胞区室时, 将每一个细胞区室(Compartment)看作是图中的一个节点, 把连接两个细胞区室中的生物分

子的反应关系看作是细胞区室(Compartment)之间的连线。这样可以抽象成只包含细胞区室的 Node-Link 图。可以采用改进的“最长路径层优先排列”^[43]的算法(见图 3.5 (b)) 来将细胞区室(Compartment)一层一层的排列。由于反应关系可以在任何的细胞区室(Compartment)存在, 很有可能一个有向图会包含圈。所以需要采用“贪心算法”去除已有的圈^[43]。(见图 3.5 (a)) 该算法会将有圈的有向图转化成无圈的有向图。具体的算法实现见算法 3.1。

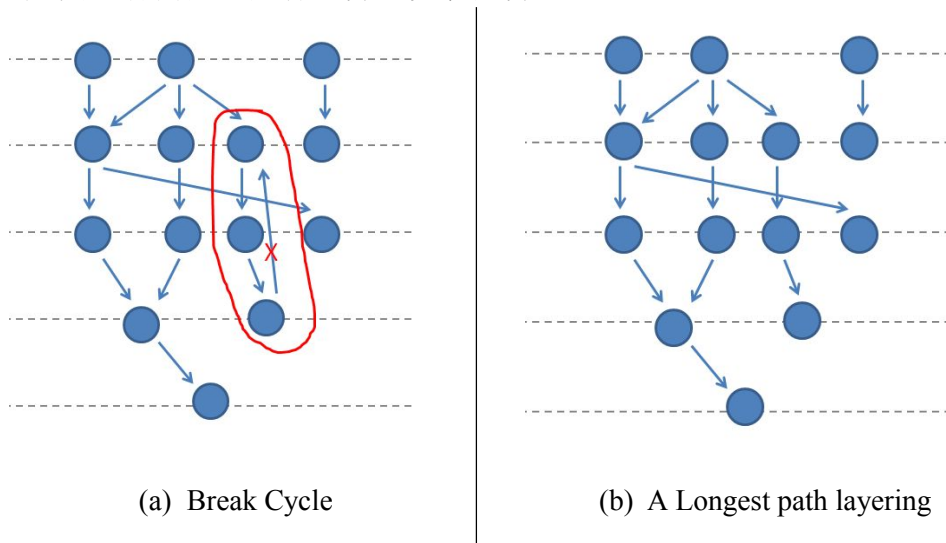


图 3.5 改进的层次排列算法设计图

算法 3.1 改进的图的层次排列算法

输入: 图 $G = (V, E)$

输出: 层次排列好的图 G' .

1. 找到图 G 的所有子树. 找到子树中的根节点, 用深度优先算法遍历它的所有子树. 如果中间遇到圈就破圈,
2. 找到每一个子树中的所有路径, 根据每一条路径的长度排序(节点的数量)
3. 取最长的那条路径
4. 创建数组: `levels[]`, 数组长度是更具最长路径中节点的数量
5. 把最长路径中找到的节点放到数组: `levels`, 然后按这种方式将余下的这个子树的其他节点方法数组的合适的位置。
6. 遍历所有其他子树, 把子树中路径的节点放到数组 `levels` 的合适位置.
放置原则如下: 对于每一条路径, 如果至少有一个节点已经在数组中, 那么这条路径的其他节点也按已有节点的位置选择相应的位置.
例如, 有路径 $[A, B, C, D, E, F]$ 已经被放到数组 `levels` 中, 于是 `levels[0] = {A}`, `levels[1] = {B}`, `levels[2] = {C}`, `levels[3] = {D}`, `levels[4] = E`, `levels[5] = F`. 如果另外有一条路径: $[T, H, C, N, E]$ 需要排列. 可以发现 `C` 已经被放置了, 难么目前这天路径的其他节点可以根据已经放置的 `C` 放置, 结果是 `levels[0] = {A, T}`, `levels[1] = {B, H}`, `levels[2] = {C}`, `levels[3] = {D, N}`, `levels[4] = {E}`, `levels[5] = {F}`.
6. 由于把每个 compartment 看作是一个节点, compartment 是有大小的. 应该考虑 compartment 的长宽, 因此, 应该计算每一层的最大宽度, 最大高度。
7. 将排列好的结果放置到 canvas 画布上.
放置原则: 当图中最长层的宽度比屏幕大小还要大的话, 那么所有 compartment 从左到右放置. 如果所有的宽度都比屏幕宽度要小, 那么就排好的结果移到中心。

力导向算法：对于每一个细胞区室(Compartment)内部的生物分子，采用力导向算法排列。在力导向算法中，将生物分子看作物理系统中的粒子，粒子间有排斥力，粒子间的连线(边)表明吸引力。但是应用到生物通路图的可视化排列中，应该做一些设计改进，由于每一个生物分子都是有大小的，为了保证节点之间不会重叠，在设计算法时，需要考虑节点的大小。所以当计算节点间的距离时，应该考虑边界之间的距离，并且在计算时，应该把每一个节点都看作是圆，圆的半径就是节点的宽度和长度的最大值。具体算法见算法 3.2。在图 3.6 展示采用混合排列算法排列生物通路的实时过程。图中，每一个细胞区室(Compartment)采用一个矩形来表示，细胞区室(Compartment)内部的生物分子也是各种大小不一的矩形，用箭头来表示节点之间的关系。如果总的细胞区室(Compartment)的空间大小比屏幕空间大的话会超出屏幕，需要用滚动条拉动查看。图中细胞区室(Compartment)之间的连线，因为细胞区室(Compartment)的位置已经排好了。

算法 3.2 每一个 compartment 内部力导向算法

输入: 图 $G = (V, E)$

输出: 力导向算法排列好的图 G' .

1. 随机分布初始节点位置
 2. 计算每次迭代局部区域内两两节点间的斥力所产生的单位位移(一般为正值)
 3. 计算每次迭代每条边的引力对两端节点所产生的单位位移(一般为负值)
 4. 步骤 2、3 中的斥力和引力系数直接影响到最终态的理想效果，它与节点间的距离、节点在系统所在区域的平均单位区域均有关，需要开发人员在实践中不断调整
 5. 累加经过步骤 2、3 计算得到的所有节点的单位位移
 6. 迭代 n 次，直至达到理想效果
-

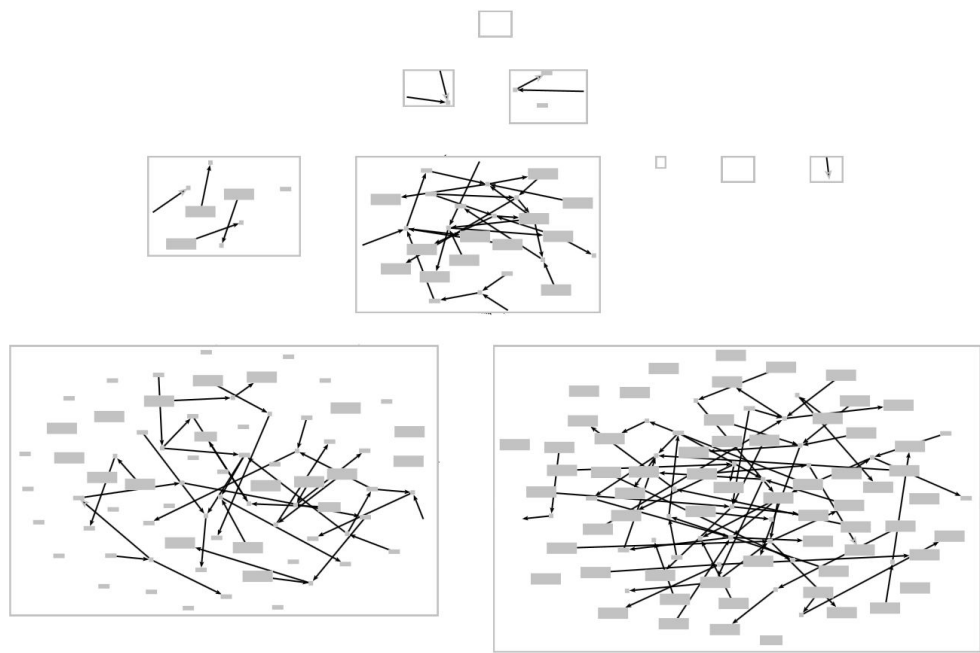


图 3.6 实时排列生物通路效果图

在 PathBubbles 中所有的生物通路图都是经过预处理的，用户可以按本系统制定的 xml 或者 json 格式 load 转换生物通路图，也可以直接从 PathBubbles 选择特定的生物通路图。图 3.7 生物通路图在 PathBubbles 中的显示：(A) Abacavir transport and metabolism 通路 (B) Peroxisomal lipid and metabolism 通路。

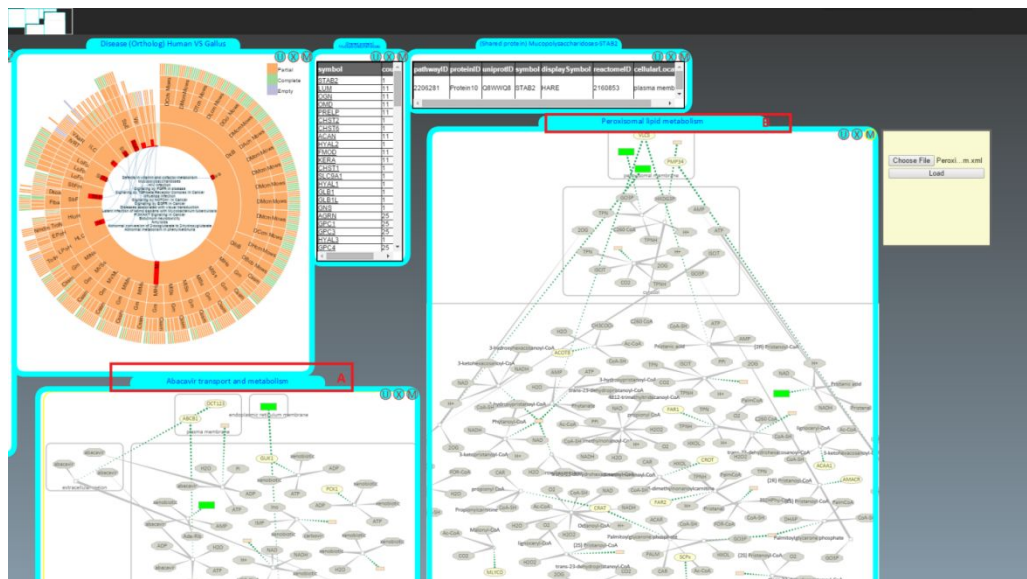
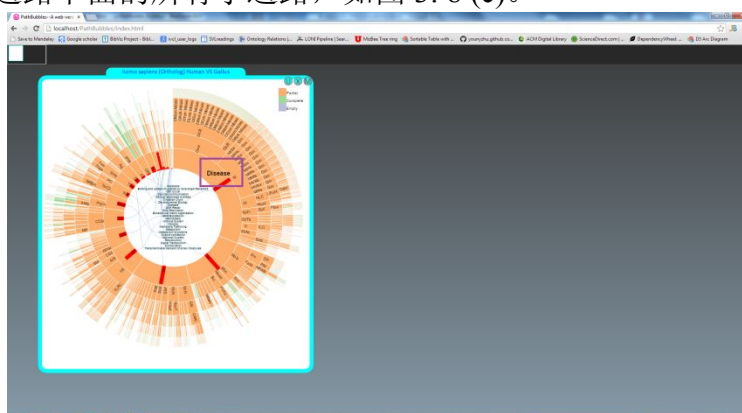


图 3.7 生物通路图在 Path Bubbles 系统中的可视化

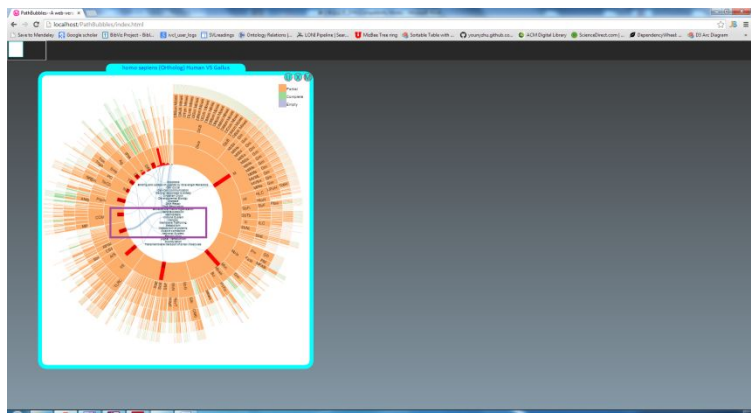
3.3.2 Sunburst 可视化气泡 以及交互式表格气泡

正如第二章所讲，除了生物通路图，生物学家还关心生物通路的层次关系，生物通路之间的 Cross-talking 关系，各种物种之间的同祖关系，同一物种在某种实验条件下基因的表达情况等。本节将要讲述 PathBubbles 系统是如何实现这些功能的。

(1) 生物通路层次关系：因为 Reactome 中的数据中大概有 1417 条生物通路，用很难用层次关系显示所有通路。在第二章分析中也讲到生物合作者希望在此基础上映射其他信息，基于多种信息的放置，程序的设计选择了 Sunburst 可视化形式。在实际设计中，发现每一条生物通路的名字很长，很难在 Sunburst 里面细小的 Arc 中放下。为了解决这个问题，设计成当鼠标悬浮在这个 Arc 上面时，鼠标指针会变成手的形状，并且会显示它的名字。如图 3.8 (a) 鼠标放在 Disease 通路上面(鼠标指针没有在截图里)。但是生物学家要比较生物通路时，每次去一个个鼠标悬浮比对名字又很费时间，所以采用取每一条生物通路的名字中的首字母(最多取 4 个) 显示到一个 Arc 中间，对于有些过于细小的 Arc，就隐藏生物通路名称的首字母。生物合作者认为，首字母也是很费时间去记忆的，最好的方式是只显示最顶层生物通路的名称。所以在最顶层的中间只排列显示第一层所有生物通路的名称，并且用线与相应的通路相连。当鼠标悬浮在通路上面时，这个名字相连的线会变粗，如图 3.8 (b) 在 Sunburst 可视化图中显示了生物通路的层次关系，其实每一个生物通路都有层次关系。在 PathBubbles 中所有的通路是可以连续查询的，这样的设计是当鼠标左键点击一个 Arc，相应会新建一个 Sunburst，并且与当前的成为一个 Group，里面显示这个选择的通路下面的所有子通路，如图 3.8 (c)。



(a) 鼠标悬浮 Disease



(b) 鼠标悬浮中间 text



(c) 鼠标左键点击 Disease，新建子 Sunburst

图 3.8 Sunburst 显示 Reactome 的 1417 条生物通路信息

(2) 生物通路之间的 Cross-talking 关系：在生物通路中，如果两条生物通路有同一种蛋白质的话，就可以说这两天生物通路是有 Cross-talk 关系的。根据前一章的分析可以知道，生物学家关心的是：哪些生物通路有 Cross-talking 关系。由于生物通路之间的 Cross-talking 关系有可能很多种，如果在 Sunburst 中画出所有的 Cross-talking 关系，那么整幅图就会显得很混乱，所以只需要显示用户感兴趣的那一层的 Cross-talking 关系。用户可以在功能菜单中选择“Show CrossTalk”，然后在 Cross-talking 下拉菜单中选择 Cross-talking 的层次。(如图 3.9)



图 3.9 Sunburst 中生物通路的 Cross-talking 层次的切换

(3) 不同物种间生物通路的同祖信息比较：Reactome 是关于人类生物通路的数据库，但是生物学家可以通过将其他物种的基因数据与人类的生物通路进行对比来预测其他物种的生物通路，找出两者的关系。这样的预测都是以人的生物通路为基础的。生物合作者把这样的关系分为三类：Complete, Partial, Empty。Complete：表示其他物种的基因在人的某条生物通路中都能找到，那么就说这条生物通路是 Complete 的。Partial：表示其他物种的部分基因在人的某条生物通路中能找到，那么就说这条生物通路是 Partial 的。Empty：表示在其他物种中的所有基因在人的某条生物通路中都找不到，那么这条生物通路是 Empty 的。

在 PathBubbles 中 Sunburst 上的 Arc，分别用：黄(Partial)，绿(complete)，紫(empty) 等颜色来表示生物通路的同祖关系。默认情况下，打开一个 Sunburst 显示的是人类和鸡的生物通路的同祖关系。另外，程序还提供生物合作者的三个物种数据库(分别是鸡，鳄鱼，乌龟)中的基因信息作为可选数据的同祖实验数据，可以在菜单中的下拉列表选择。同时，程序还允许用户导入自己的基因数据，做同祖信息的分析。用户自定义基因数据格式是第一列基因 Id，第二列是基因符号，中间用 Tab 隔开。一个基因一行。在一个 Sunburst 中，要显示 1417 条生物通路，在有限的空间里面，最外层的生物通路将会变得很小，考虑到这个问题，系统提供选取子通路的功能。只要鼠标左键点击它的任意一个祖先通路的 Arc，就可以产生包含这个祖先的子通路。还可以在 Sunburst 的 Arc 中嵌入柱状图，用来对比物种间生物通路共享基因的个数。柱子越高，表示数量越多，生物学家通过查看柱子的高低就能找到感兴趣的生物通路。这些共享的基因的具体情况，可以在一个表格中列出来。在表中用户还可以排列表来找

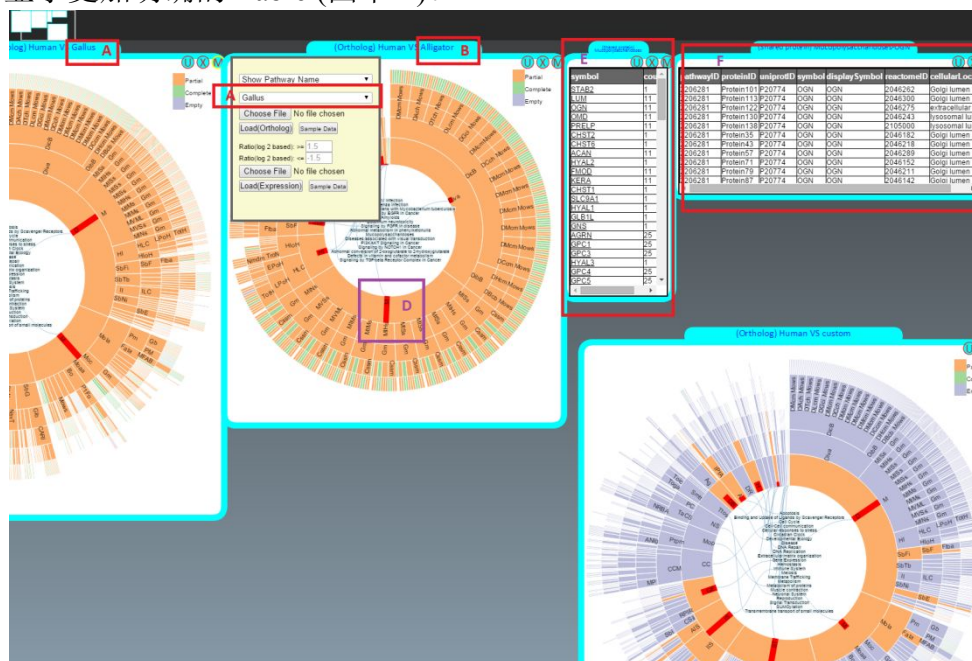


图 3.10 PathBubbles 中不同物种的生物通路分析

对于环境条件的不断变化，单细胞原核生物具备高度的自适应性，它们能够迅速的调节各种基因的表达水平。真核生物的基因调控远比原核生物复杂，不同的生物，不同的细胞，具有不同的表达机制^[44]。基因的差别表达是细胞分化和功能的核心。基因的差别表达包括正调控和负调控。根据与合作者的讨论发现生物研究人员所关心的是“对于某种试验环境下，生物通路的基因差异表达情况以及不同生物通路间基因差异表达的数量”。在 PathBubbles 的 Sunburst 可视化中，提供了分析特定物种的基因表达情况的功能。该系统允许用户上传实验数据(基因的表达情况表)，并且只要在程序中提供实验的正调控和负调控范围值，之后在 Sunburst 可视化中就会用不同的颜色来展示基因表达的情况。对于特定生物通路的基因表达情况，可以用柱状图的形式显示。当用户右键点

击具体柱状图时，可以查看具体的基因信息，用户可以点击不同基因符号来查询具体的蛋白质信息。

我们的生物合作者主要研究关于环境温度对于肉鸡的基因表达的影响情况^[45]。接下来，以这些数据为例，做一个简单的介绍。在前面做同祖信息时，提到程序提供三个物种(鸡，乌龟，鳄鱼)做为基准可供选择，如果用户提供其他物种的实验表达数据，首先应该提供该物种的基因信息，之后在接着进行基因表达情况的研究。首先介绍 **Sunburst** 可视化做基因表达分析要用到的相关功能菜单。(见图 3.11)图中用户可以自定义的输入，以 \log_2 为基准的范围。用户可以选择自定义基因表达数据。数据格式是第一列基因 **Id**，第二列是基因符号，第三列是 **ratio** (每个生物通路中的基因在实验中和正常情况下的差异表达的情况)，**ratio** 是生物研究中一个很重要的指标，比如：**ratio** 越大，该生物通路受到因不同实验环境导致的基因差异表达影响越大，该生物通路可能是生物学家下一步研究的重点。中间用 **Tab** 隔开。一个基因一行，基因间用换行符号隔开。接下来输入基因表达的实验数据。(图 3.12)图中对于不同的生物通路，在 **Sunburst** 可视化图上面显示不同的颜色来表示被差异表达的基因占这条通路的比例。同样，在中间的第一圈，用柱状图显示各个通路基因的数量。用户如果右键点击感兴趣生物通路的柱子(见图 3.13 A 点击 **Disease** 柱子)，就会得到关于基因表达信息的表格。(见图 3.13 B)对与表中的特定的基因，用户可以右键点击，查看这些蛋白质在人类中的基本信息。(见图 3.13 中 C)对于所有的表，如果点击表头都是可以排序的。

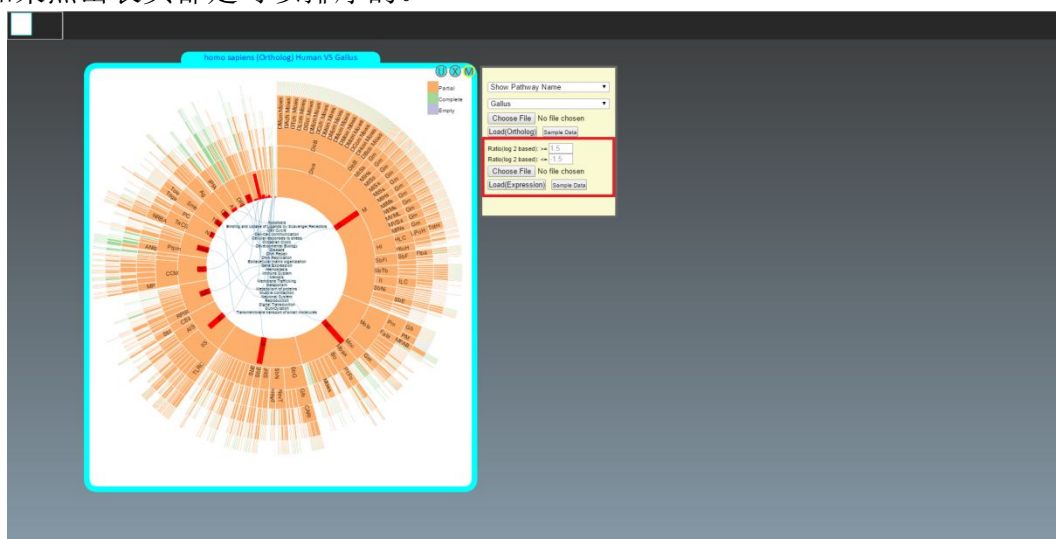


图 3.11 基因表达功能菜单

等)映射到 Sunburst 上 Arc 的颜色中。这种方式能够给生物学家提供一个对生物体的所有生物通路中的相关生物分子的变化全局观,有助于研究人员迅速的捕捉感兴趣的生物通路。

(3) 在同一个 Sunburst 可视化图中,同时展示多种生物信息,这是许多生物通路数据库可视化软件所缺失的。Sunburst 可视化填补了这方面的空白。

(4) 生物通路是有层次关系的,对每一个生物通路只需要点击上面的 Arc 就能产生新的 Sunburst 可视化,这方便了生物学家的操作和功能的记忆,同时也方便生物学家对生物通路上层 Sunburst 全局可视化和新产生 Sunburst 生物通路的“全局”可视化之间的对比。

3.4.2 系统交互设计的有效性

生物学家特别喜欢 PathBubbles 系统中提供的交互,只需要简单的点击,就可以把许多的信息连接到一块,这种简单而有用的方式,能够帮助并提高数据分析的效率。利用 Sunburst 可视化从全局到局部的信息探索方式,与生物研究人员在实际研究中数据分析过程是一致的。接下来,从以下几个方面总结研究人员的观点:

(1) 两个物种间同祖信息的比较,在 Sunburst 中提供基于生物通路的全局关系。这种方式能够帮助生物研究者迅速找到类似的生物通路。

(2) 生物通路间 Cross-talking 关系:只需要点击某条生物通路就能查找生物通路之间的关系以及所关联的生物通路,这有助于研究人员找到与当前感兴趣生物通路相关的其他生物通路,并且进行实验假设的验证。

(3) 基因表达信息, Sunburst 提供所有生物通路之间基因表达的情况,方便实验对生物通路影响情况的对比,尤其是上面的柱状图,更加具体的显示了生物通路中异常基因表达的数量情况。

(4) 可以排列的图表形式展示生物通路的具体信息,这是生物学家所熟悉的方式,大大减少了学习新的可视化工具的不必要性。用图表来显示查找到的感兴趣的生物通路的具体信息与研究人员的实验过程要求一致。

(5) PathBubbles 系统中提供的 Bubble(气泡)之间 Group(组合)和 Ungroup(分离)功能,方便生物研究者对对比实验数据进行记忆分析。系统中可扩展性的导航栏,方便生物研究人员在进行数据分析时,用功能视图进行来回切换比较。

3.4.3 用户体验的总结

总的来说,使用本系统的生物研究人员认为 PathBubbles 是一个很有用的用于生物通路信息浏览和分析的软件。与其他软件相比,它操作简单易用,很容易上手。全新的界面和交互方式减少了多窗口操作和多种数据集的记忆负

担。从 Sunburst 可视化中的层次关系中产生新的 Sunburst 可视化，保留了浏览信息的上下文，这减少了其他工具分析时来回切换界面的记忆负担。系统中集成多种数据库中的多个数据，提供了数据分析的良好的可扩展性。接下来主要从以下两个方面总结两个生物学家对 PathBubbles 系统的用户体验：

(1) 可扩展性：就目前来看，还没有一个工具可以整合 Reactome 数据库中所有人类生物通路数据到一个可视化中，并提供数据分析的全局视图。PathBubbles 中的 Sunburst 可视化整合所有生物通路，方便了研究人员对研究数据的全局比较。许多生物研究人员喜欢将多个相关联的 Bubble 气泡(功能视图)进行组合(多窗口性)来对比分析数据，这减少了数据分析时，不同窗口之间的来回切换，直接可以进行对比分析。

(2) 动态性：PathBubbles 支持对信息的动态分析，对自定义信息的导入。只是简单的拖放操作就可以对可视化结果进行重排，并且在分析的不同阶段，都可以随机的导入数据，切换默认数据。

(3) 同步性：由于人的记忆是有限的，当处理大量的数据时，更加困难。导航栏中对前一功能视图的缩小显示的方式，很好地减少了数据分析过程的记忆。组合的气泡附加相同的状态有助于寻找数据集的相似性。

3.4.3 PathBubbles 系统可扩展性的建议

生物合作者认为 PathBubbles 系统还应该可以分析其他类型数据，解决其他类型的问题。比如，用基因和染色体的信息来推断信号传导和新陈代谢的生物通路。比如，在生物通路分析中可以增加编辑和查找的功能，这样方便生物研究人员创建特定数据相关的生物通路图，以及对各种生物通路数据的保存。比如，在本系统中集合其他一些现有的数据库，可以减少生物研究者在不同数据库之间的来回切换。

3.4 本章小结

本章首先介绍了 PathBubbles 系统实现的相关技术，之后介绍了气泡隐喻界面的具体实现以及如何管理场景中的 Bubble，然后介绍了 PathBubbles 系统功能模块的实现，从生物通路图的排列，到生物通路数据的分析，比较等功能的进行了相关的可视化设计，如 Sunburst 可视化，交互式图表的可视化。最后请了两个生物学专家对 PathBubbles 系统的可操作性和实用性上作了评估，结果是一致认为“这是一个很有用的系统”，研究人员还针对 PathBubbles 系统的现有的特点提出了一些可扩展性的建议。PathBubbles 系统是一个开源项目，目前托管于 Github: <https://github.com/yongnanzhu/PathBubbles>。

详细的应用界面，操作文档和数据分析界面，请访问

<http://raven.anr.udel.edu/~sunliang/PathBubbles/>。

第四章 在线排列和编辑生物通路图

在许多生物通路数据库中，生物通路图都是由人工来排列的，这个过程是很耗时的。尽管 Reactome 采用了“由专业人员撰写，同行评阅”的模式，这种模式与众包类似，但是需要专业人员人工绘制。其实只要运用适当的质量控制方式和训练文档，这种图的排列的后期改进一般人还是可以完成的。本章设计实现了一个在线排列和编辑生物通路图的网站，从而为 PathBubbles 方便快捷的制作有用的生物通路图，也可方便生物研究人员在线浏览和编辑生物通路图。

4.1 设计思想

在第三章介绍，本系统运用混合排列的方式来排列产生生物通路图，但是还是存在一些边交叉，箭头流向混乱等问题。最近，亚马逊公司的土耳其机器人(Mturk)^[46]被广泛使用，这是一个工作平台，主要是针对那些人能做，但是计算机很难做到的任务。其实对于生物通路图，项目合作者认为前期混合排列的结果还是不错的，不过需要人工改进。针对这个想法，本文设计了一个在线排列生物通路图这个平台。图 4.1 介绍了排列产生更优的生物通路图的处理管线：从 Reactome 的原始数据处理到运用人工在线排列并得到结果。图 4.1 中混合的排列过程和 Coloring 的过程在第二章和第三章已经介绍过。本章主要讲述如何设计在线排列网站来得到良好的排列结果。

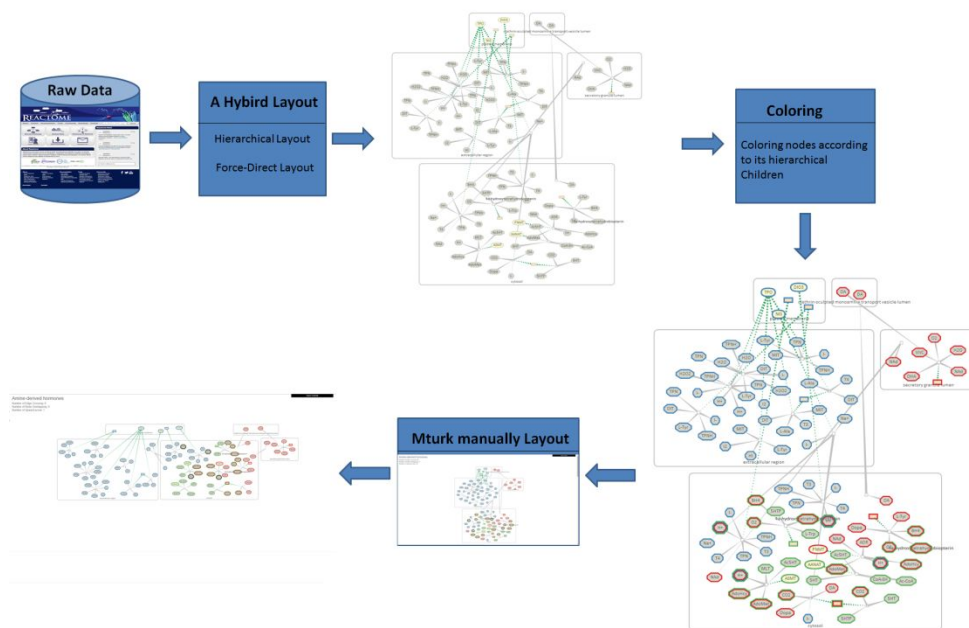


图 4.1 PathBubbles 中生物通路图质量的处理管线设计流程图

因为任何人都可以登录网站，去排列生物通路图，所以应该设计一个良好的用户文档。因为并不知道排列者是否懂生物通路图。所以在文档中应该对图的排列要求和相关的理论做一个简单而明确的介绍，另外一点就是质量控制，就是关于怎么训练人们去提高质量。为了提高制作速度，可以把这样的任务放到 Mturk 平台。

4.2 文档设计

生物通路图排列的训练文档¹应该包括图排列任务的介绍，一个简单的图排列的理论，一系列帮助生物学家提高图排列质量的策略。对于里面的任务指令，应该尽可能的简单，明了，容易让人理解。为了得到一副生物学有用的生物通路图，通过查阅文献和与生物合作者讨论，总结了一系列排列生物通路图的标准。接下来首先来介绍这些标准，之后介绍如何将这些标准应用到排列生物通路图的设计任务中去。

4.2.1 生物通路图排列标准

生物通路图的排列准则应该能够帮助生物学家研究生物学的过程，包含生物学家研究所关心的信息。本文将这些标准分成两个大类：来自图论的美学标准和包含生物学相关知识的标准。

在过去的十多年，为了得到一副易读的图，人们提出了一系列的标准，这些标准是基于图的相关属性的^[43,47]，这些属性包括包括：边交叉，边长度，角度，绘制区域大小，图的对称性等等。以下列出生物通路图排列的标准：

- (1) 减小边与边的交叉。
- (2) 较小边与节点之间的交叉。
- (3) 减小边的长度。
- (4) 均匀分布节点。
- (5) 避免节点重叠。
- (6) 最大化屏幕利用率。

在计算机图形学的发展中，人们尝试利用优化图的绘制算法去改进图的绘制结果，通过检测输出图形是否遵循相关标准来评价这副图的好坏，从而来评价图排列算法的好坏^[48]。然而，在生物学领域，一副好的图需要附加一些生物学的信息^{[10], [11], [15], [19], [17], [49]}。以下列出一些需要被包含在图中的生物学信息：

- (1) 保留生物通路中反应信息，例如：有向性。
- (2) 强调生物信息流的大致方向，例如：从上到下。

¹ <http://www.csee.umbc.edu/~yongnan/mturkOld/>

- (3) 保留生物学的结构信息，例如：细胞区室(Compartment)位置信息，节点层次信息。
- (4) 尽可能的将属于同一种类的节点放在一起。同一个细胞区室(Compartment)内部节点尽可能放在一起。
- (5) 反映一个反应的输入与输出的关系，走向。

4.2.2 排列生物通路图的一些指导准则

在训练文档的设计中，将以上通路图的设计应用到图排列的指导准则中。表 4.1 列出了这些准则。为了方便理解，采用图文加指令的形式。

表 4.1 生物通路图排列准则

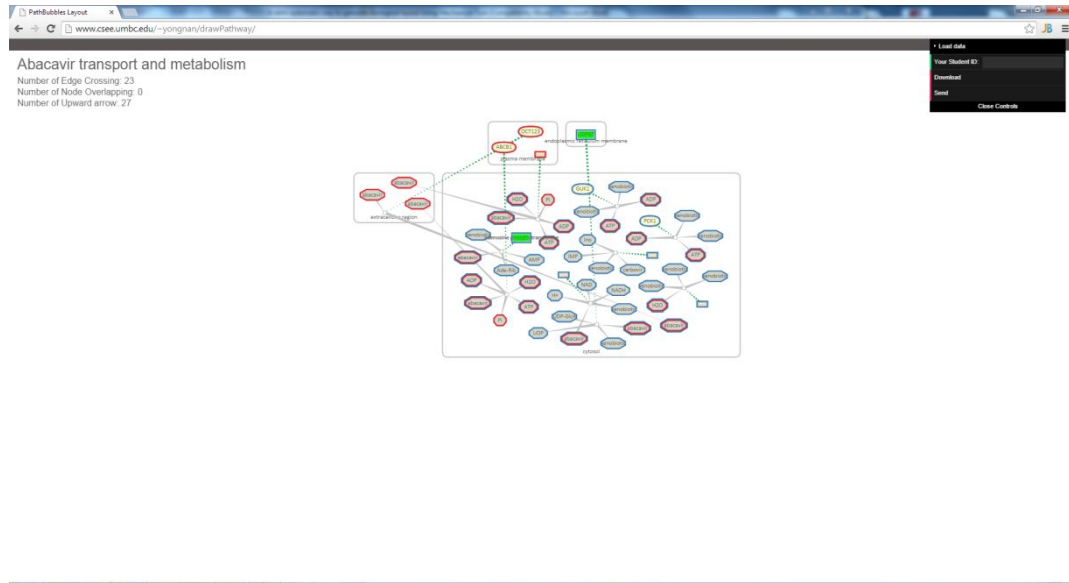
Rules	Operation	Input	Example Output
Avoid node overlapping	Drag the overlapping node so nodes won't overlap.		
Reduce edge crossing	Drag a node so edges won't intersect.		
Reduce edge length	There are some unnecessarily long edges in the graphs, please drag the node (or a group of nodes) to shorten the edges.		
Move a compartment	If a compartment has no connections to any other compartments, move it to the top of the graph.		
Resize a compartment to fit the width of its name	If the name of a compartment is too long which exceeds the width of this compartment, adjust the width of the compartment to fit it.		
Direction of the arrows	Move nodes and compartment to make sure most arrows are pointing downwards.		
Never ever change the containing relationships	In your final graph, never place a node outside of a compartment to which it belongs.		

4.3 界面和交互设计

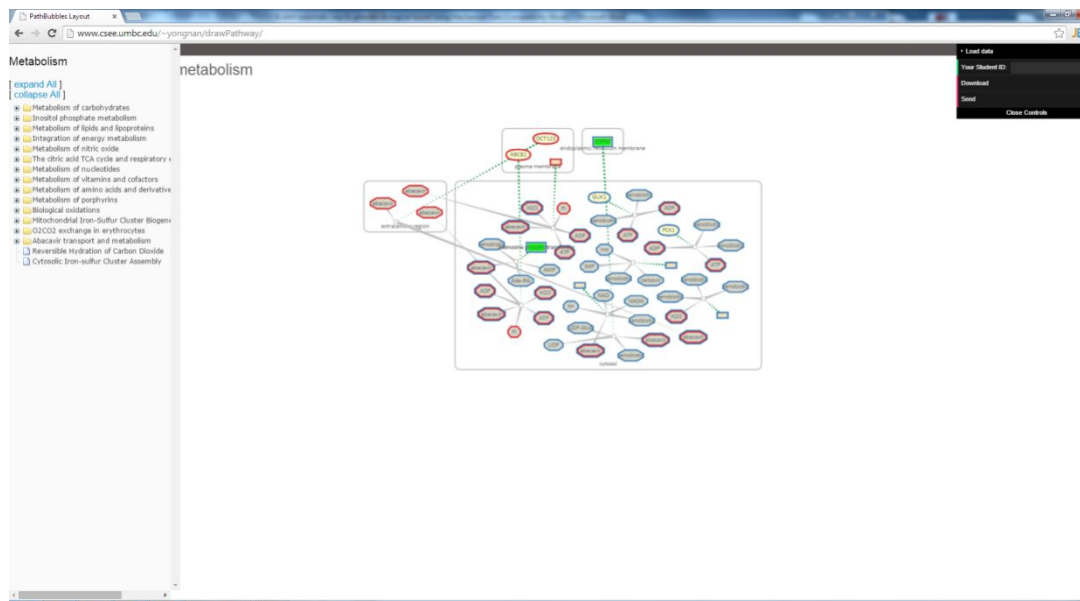
4.3.1 界面设计

生物通路图的在线排列应该包含生物通路的选择，生物通路图的放置，功能菜单的放置等等。图 4.2 是生物通路图编辑平台的界面，在中心显示要排列的生物通路图。左上角显示最重要的价格衡量指标：边交叉数量，节点重叠数量，箭头往上数量等等。界面右上角是功能菜单。对于生物通路的选择，采用

两种方式，第一种是当用户鼠标靠近屏幕左边时，(图 4.2 b) 会自动弹出一列包含生物通路的树，用户可以点击选取相应的生物通路排列。第二种是用户如果有本系统指定格式的生物通路图，那么可以通过功能菜单 Load data—Choose File—Load 导入相应的生物通路数据。在排列过程中，用户可以通过 download 按钮下载相应自己排列的中间结果。当排列完成时，只要点击 Send 就可以将数据保存到系统服务器。



(a)



(b)

图 4.2 生物通路图的在线排列界面设计

4.3.2 交互设计

对于生物通路图的排列，主要是通过鼠标单击的选择，移动。针对任务需求作以下交互操作的设计：

(1) 选择：鼠标左键单击可以选择任何一个节点(Node)或细胞区室(Compartment)。当被选中时，选中的那个节点的边界会变成黄色。

(2) 多选：

- 1) 在用鼠标左键选择时，如果同时按下“G/g”键可以进行多选操作。
- 2) 选择与一个反应相连的节点时，如果在鼠标选择该节点时，同时按下“Shift”键，就能选中与这个反应相连的所有节点时
- 3) 选择相同边框颜色的节点，当鼠标点击一个边框有颜色的节点时，同时按下“Alt”键，就能选择同时有相同边框的所有节点。

(3) 移动：任何选中的节点或细胞区室(Compartment)，都可以通过拖动鼠标来移动他。

(4) 改变大小：可以通过把鼠标放在细胞区室(Compartment)的边框来改变细胞区室(Compartment)的大小。

针对用户排列的过程，在屏幕左上角实时显示当前图的边线交叉数量，节点重叠数量和箭头朝上的数量。这样可以实时的提醒生物通路图的排列中如何考虑这些指标。针对以上这些，在训练文档中还需要提供一个简单的操作视屏，以供参与者迅速学习。完整的设计文档和实验网站，请访问：

<http://www.csee.umbc.edu/~yongnan/mturkOld/>和
<http://www.csee.umbc.edu/~yongnan/drawPathway/>。

4.4 本章小结

本章设计并实现了一个在线排列生物通路图的网站。为了得到有意义的生物通路图，设计了详细的排列指导准则。这样一个在线平台，无论是对生物通路图的优化，还是教学都是很有用的，所得到的图对于 PathBubbles 系统中生物通路图的优化作了补充。对于编辑生物通路图的网站，项目合作者特拉华大学的 Carl Schmidt 教授，还用于了生物通路的教学。在教学中，他让学生在在学习过程中排列生物通路图，取得的结果还是不错的。与传统的静态纸质生物通路图相比，这种方式更能让人快速理解生物通路图中包含的信息，以及里面包含的生物分子之间的关系。

第五章 总结与展望

随着生物信息学的不断发展，如何利用现有的数据，如何有效地提高生物数据的分析，正变得越来越重要。本文对生物数据可视化系统的界面和交互设计，生物通路数据的可视化做了一系列的研究，对于特定的生物数据和生物学家研究过程，开发设计了交互式可视化分析系统 PathBubbles，对于包含特定生物信息的生物通路图的绘制算法做了相关的研究，对于生物通路图的优化设计了一个在线的交互式排列平台。对于生物通路的可视化形式，设计了层次分明的 Sunburst 可视化图。本章主要分为两部分，首先总结本文的相关研究工作，然后对未来的进一步工作作出相关的展望。

5.1 工作内容总结

本文设计并实现了一个基于 Web 的交互式可视化系统 PathBubbles，对于气泡隐喻界面的可扩展性设计，开发了开源框架 PathBubbles Javascript 库。本文首先通过对生物学家进行了需求分析(面谈和问卷调查)，之后根据需求和提供的生物数据，做了相关的数据分析，总结了主要的功能。通过进一步对相关功能和可视化需求的分析，设计了三种可视化形式功能气泡：生物通路图气泡，Sunburst 可视化气泡和交互式表格气泡。对于生物通路图的进一步改进和完善，设计并实现了一个在线排列的平台。以下主要对系统中的界面设计，可视化设计进行展开。

(1) 界面设计：这是一个全新的基于网络的生物通路分析工具。本文将 VisBubbles，CodeBubbles 的气泡隐喻界面的设计思想应用于 PathBubbles 的设计。PathBubbles 把每一个气泡看成一个功能视图。将传统的桌面设计成可扩展的虚拟桌面，在上部的导航栏中显示当前当先场景中的物体，以及以前场景中的物体。所有场景保存于导航栏，极大的方便了生物学家进行场景之间的切换，和记忆多场景功能。支持多视图气泡窗口的创建，这样有利于数据之间的对比，以及数据分析状态的记忆。

(2) 可视化设计：本文对生物数据进行了分析，对生物学家的相关需求设计了问卷调查(表 2.1 关于生物学家研究生物通路的一些调查问题)。通过对这些需求的分析，设计并实现了提供全局数据分析的可视化视图 Sunburst，Sunburst 可视化视图集成了 Reactome 数据库中所有人类的生物通路(目前一共有 1417 条)，即保存了生物通路之间的层次关系，又用颜色显示实验数据中的生物通路中的基因变化情况。Sunburst 可视化气泡中每一个 Arc 又可以产生新的子

Sunburst 可视化气泡，这样方便用户对信息的记忆。Sunburst 可视化气泡还可以结合生物学家所熟悉的图表的形式来排列显示所有实验相关生物通路的蛋白质的信息。Sunburst 可视化气泡给予生物学家一个全局的分析数据的界面，方便浏览信息，颜色的显示极大的提高了生物研究人员对于有用信息的查找。Sunburst 可视化气泡集成和整合多种默认数据，方便用户查找和切换数据。PathBubbles 还提供用户输入自定义数据的接口，这样方便用户对自有物种实验数据的分析。无论是同祖信息比较分析还是基因表达情况的分析，这都提供了一个从全局到局部的分析方式，这与生物学家实际的数据分析过程是一致的。

5.2 进一步工作的展望

本文的可视化系统 PathBubbles 是一个基于 Web 的集可视化，交互和分析于一体的开源框架。对于扩展气泡隐喻的设计思想是一个很好的补充。该框架可以设计用于生物数据的研究分析，也可用于其他数据的研究分析。对于进一步的研究工作可从两方面展开：框架的可扩展性结构的优化和生物数据的进一步功能性可视化分析的探索。

框架的进一步可扩展性优化：(1)对于场景管理，PathBubbles 采用场景树的结构，然而对于场景中物体事件的监听，当前版本并没有采用场景树的结构。因为在 PathBubbles 中主要采用 HTML5 Canvas API 来渲染场景，但是在 HTML5 中所有的事件只能加到一个 Canvas 中，这样事件的选择还是要通过一种遍历机制，本文采用最简单的位置映射的方式来遍历事件，这种方式最大的问题是每增加一个事件就要考虑整个场景中事件的遍历。如果事件遍历也采用场景树，那么事件将于物体一致(直接加到物体上)，这样方便事件监听，也方便框架使用者的设计。(2)D3 是一个可视化库，里面集成多种可视化技术。PathBubbles 集成 D3，这样可以方便用户应用 D3 相关的可视化技术。主要问题是 D3 采用 SVG 做渲染媒介，尽管 PathBubbles 将 SVG 放到 Canvas 上面的设计思想可以解决两者的兼容问题，但是拖动 SVG 有时会出现“卡顿”的问题。该框架的进一步工作可以将 D3 产生的结果，直接渲染到 Canvas 上面，这样一方面集成了 D3，同时也解决了交互的问题。

生物通路数据可视化的进一步开发分析：本文设计并实现了一个在线的用于人工排列生物通路图的网站，当新的生物通路产生时，或者当一条通路中新的生物分子加入时，这里采用众人的力量优化生物通路图的思想。在 PathBubbles 中，可以可视化生物通路图，但是只有简单的排列功能，并没有编辑生物通路图的功能，在这方面可以做相关的扩展。在生物数据分析方面，目前 PathBubbles 可以用于同祖信息分析，基因表达分析，可以扩展加入其他生物数据分析的功能，整合更多的生物数据库，比如 PPI(Protein Protein

Interaction)数据库, RLIMS-P(Phosphorylation Database 数据库)。对于生物数据可视化方面, 在 PathBubbles 中主要采用 Sunburst 来显示所有生物通路信息, 还可以集成其他的可视化技术来做生物数据的相关功能分析。

致 谢

时光流逝，岁月不在。眨眼间，3年的硕士生涯已接近尾声，蓦然回首，不禁感慨万千。

我要感谢我的导师方景龙教授，方老师渊博的知识，严谨的治学态度，踏实的工作作风，不断地鞭策和激励着我，对我各方面地成长起到了很大的作用。特别感谢方老师3年来对我的支持和鼓励。

我要感谢美国马里兰大学巴尔的摩分校的陈健老师。本文的许多研究工作是在陈老师的悉心指导下完成的。从论文的选题，到最后的完成，每一个细节都得到了陈老师的细心指导和热情支持。3年来，陈老师为我的科研和成长，付出了许多宝贵的时间和精力。难以忘记第一次 Skype 会议时的对话——我说“我本科不是计算机的，目前好多东西还不会，但是我会努力学习的”，您说“没关系，慢慢来”，您的话语深深地激励着我，给予了我前进的动力和奋斗的方向。是您把我带入了数据可视化这个充满挑战的研究领域，使我找到了为之付出汗水的方向；是您在我科研学习中遇到问题时，给予我激励和指导；是您在我失落的时候，鼓励我要充满自信；是您给予我“开眼看世界”的机会，让我来马里兰大学巴尔的摩分校访学一年，这一年是我能力提升最快的一年。陈老师渊博的知识，严谨的治学态度，高度的责任感和夜以继日的工作作风，都给我留下了深刻的印象。陈老师传授给我的不仅仅只是知识，更为重要的是科研态度和科研方法。在此，我由衷地向您说一声“谢谢”。

我要感谢杭州师范大学的潘志庚教授。潘老师严格的管理方式和在科研工作中所表现出来的激情，使我学到了很多，真心地感谢您。研究生的前两年，潘老师为我在数字媒体与人机交互研究中心提供了良好的学习环境和科研氛围。难以忘记读研前的夏天，为方便我们去实验室学习，您腾出了家里的房间让我们居住；难以忘记每周一次的工作汇报，让我养成了“本周总结，下周计划”的工作习惯；难以忘记您在汇报中给我提出的宝贵意见，使我不断地改正学习进步；难以忘记您让我在国际会议中做志愿者帮忙，使我领略学术的魅力；难以忘记每年中心举行的年会和几年一次的 VRMM 大会，使我感受了家的温暖；

我要感谢吴克勤博士(师姐)对我学习，工作和生活中的指导和帮助，师姐的科研能力和乐于助人是大家有目共睹的，值得我用一生去学习。难以忘记，一年前寒冷的冬天，您来杜勒斯机场接我；难以忘记，您为我找到住处，并帮

我添置各种生活用品；难以忘记，每周的三四六，您让我学会了如何做菜；难以忘记，每周五，您载着我们去大中华，韩亚龙买菜的日子；难以忘记，您在百忙之中，带我和国浩去滑雪，让我知道美国的生活也是很丰富多彩的(尽管也就这么一次)；难以忘记，我刚到实验室的前几天，您破例在晚上十一点前载我回家(尽管最后我也养成了 1 点以后回家的习惯)；难以忘记，我刚接手 PathBubbles 时，您给我耐心的讲解(尽管最后我还是走了些许弯路)；难以忘记，7 月中旬，您在密西西比，我在实验室，我们通宵改程序，调代码(尽管最后实验还是失败了)；难以忘记，您已经去工作了，还开车将近一个小时，每天回来做 Study；难以忘记，您喜欢吃鱼，但是也吃鱼头和鱼尾；难以忘记，我们的其中一个对话，我说：“每年您到底休息几天”，您说“圣诞节出去玩几天”，那我说“其他时间呢”，您说“之后在家睡 3 天，就去实验室”。

我要感谢张国浩博士，同门如手足。在美国期间，国浩给予了我许多的帮助。刚来美国的前三天，国浩腾出了志光的地方，让我找到暂栖之地。在我的工作和学习中，国浩提出了许多建设性的意见和建议，每一次的讨论，使我受益匪浅。国浩同学的踏踏实实，不骄不躁，对我很是鼓舞。

我要感谢山东大学的徐延宁老师，在百忙之中，还通宵达旦的写代码，改程序，对此，我很是感动。

我要感谢王志光同学，孟庆雷同学，一个开车，一个提供车，载着我们出去烧烤。

我要感谢郑州大学的宋为老师，武汉大学的宋成芳老师，与他们的交谈使我学会了许多。

我要感谢赵赫楠博士，她在我心情郁闷的时候，给予了我一些建议和意见。

我要感谢 Alex 博士，尽管我们在 Pathbubbles 项目中的交集不多，但每次的交流还是让我学到了很多。

我要感谢我们的项目合作者——特拉华大学的 Carl Schmidt 教授，Carl 博士严谨地治学态度和科研精神深深地影响着我。总以为美国人周六日，应该是不工作的，但是突然而来的关于程序的意见，预示着他还是这么努力的在工作。每周二的例会，总是会给予我许多的意见和建议，有些还是很有用的，有些甚至是颠覆性的。

我要感谢特拉华大学的孙亮博士，本文的许多工作都是通过我与他的讨论完成的。对于生物学相关的知识，孙亮耐心地给我讲解，最后程序的设计阶段，我们几乎每天都有好几次地交流。与孙亮的讨论，使我学到得不只是这个项目相关的知识，还是对待事情的认真态度。

我要感谢 Van Gogh 实验室的 Marc Olano 教授, Yu Wang 博士, Weiss 博士, Ari Blenkhom 博士和 Yingying Zhu 博士对每周组会的付出, 每一次 Marc 的点评都会引发大家一系列的思考。感谢 Van Gogh 实验室和 IVCL 实验室我们这两个实验室的老师 and 同学, 我们共同营造了良好的科研氛围, 一起的讨论和交流, 给予了我许多的灵感, 在科研工作中也有许多的帮助, 这开阔了我的视野, 使我懂得了如何多角度看问题。

我要感谢我的房东 Liusong Ding 和 Yuwei Zhang, 我们是房东和房客的关系, 也是同乡的关系。一年来我们一起做饭, 一起吃饭, 互相包容。

我要感谢教会的冯舟, 周杨凌, 安娜等朋友和牧师, 我们一起查经, 一起游玩。尽管对于信仰, 我还是持保留意见, 但是不同思想的碰撞, 还是引起了我对人生的许多思考。

我要感谢杨长云教授, 杨老师是我见过的真正来美国学习, 交流的, 每天早上 9 点不到去学校, 晚上 11 点回去。这样的精神深深鼓舞着我去努力学习奋斗。无论是写论文, 还是平常对待生活的交谈, 我从他那里学到了许多。不是为了生活而生活, 而是为了享受生活而生活。尽管每天这么努力的学习工作(很忙的样子), 但是在周六日, 他还是会带着他的孩子出去玩; 尽管手头并不是很富裕, 但还是会花好多的钱买书。杨老师对生活的追求, 对工作的执着, 深刻地改变了我对待生活的想法和态度, 原来“学习, 工作, 家庭”可以如此平衡!

我要感谢在美国遇到的未知名的朋友, 相逢就是缘。

我要感谢邵伯仲, 王盛波, 张士诚, 严政, 孟庆如, 吴博, 陈晨阳同学和郭晶师姐, 读研之前的暑假, 我们一起学习, 一起爬山, 一起畅谈未来。之后的一年, 在数字媒体与人机交互实验室, 我们一起科研, 一起活动。

我要感谢张才章, 方沛毅, 邹雪, 陈攀, 商伟峰, 陶超, 巍峨, 陈娟娟, 陈龙飞, 孙偏偏, 郑俊鹏, 吕步云, 何强等同学, 我们一起玩耍, 一起打乒乓。

我要感谢张东生, 何晓文, 万季师兄和朱雪梅师姐, 感谢你们对我的照顾, 在你们毕业欢送之时, 我没能敬上一杯酒, 表示抱歉。

我要感谢 601 实验室和 603 实验室的同学, 我要感谢我的师弟师妹们, 同在我们的大实验室, 我们是有缘的。

我要感谢室友董紫辉, 袁国伟, 林通, 何强, 吴腾飞, 张超, 常若愚等。我们一起聚餐, 我们一起喝酒, 我们一起 KTV。

我要感谢吴勇, 我们在读研前就认识, 之后我们一起打牌, 一起分享电影。

我要感谢王兴起老师，一直每天以身作则，每天很早到实验室，下午很晚才走，激励着实验室的每一个人要如此工作学习。

我要感谢田晓秘书，张军辉和张晓琪辅导员老师，董紫辉党支部书记和顾靓班长，对于我不在学校的这一年，相关资料填写的帮助以及学校通知的提醒，我铭记于心。

我要感谢王毅刚，李黎，陈临强等老师在研一时给我打下的基础。

我要感谢父母，是他们给予了我生命，是他们养育了我，在二十多年的求学生涯中给予我无条件的支持和关爱。

我要感谢我的亲戚朋友以及所有帮助过我的人。

特别感谢王大全老师和吴克勤师姐在论文最后的修改提出的宝贵意见，本文的终稿修改得益于王大全老师的悉心指导。特别感谢吴腾飞、刘愉等人在我回国后找工作的相关指导、建议。特别感谢翠云姐和邵伯仲帮我投递阿里的岗位。

我要感谢陈健老师的 NSF 基金的支持，使我在美国访学的这一年，可以安心的学习，科研。

我要感谢三句话：“做该做的事，请做好本职工作”，“努力，不一定有所收获，但是不努力，一定什么都不会有”，“蓝海战略，我的蓝海在哪里？”。

努力工作，追求生活，享受生活。

竺涌楠

2014/11/17 凌晨 3 点

于 美国·巴尔的摩

参考文献

- [1] 百科. 生物学通路[EB/OL]. . <http://www.baike.com/wiki/生物学通路>.
- [2] Genome. Biological Pathways[EB/OL]. . <http://www.genome.gov/27530687>.
- [3] Huang D, Huang Y, Bai Y, et al. MyBioNet: interactively visualize, edit and merge biological networks on the web[J]. *Bioinformatics* (Oxford, England), 2011, 27(23): 3321–2.
- [4] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes[J]. *Nucleic acids research*, 2000, 28(1): 27–30.
- [5] Croft D, O’Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes[J]. *Nucleic acids research*, 2011, 39(Database issue): D691–7.
- [6] Joshi-Tope G, Gillespie M, Vastrik I, et al. Reactome: a knowledgebase of biological pathways[J]. *Nucleic acids research*, 2005, 33(Database issue): D428–32.
- [7] Caspi R, Foerster H, Fulcher C A, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases[J]. *Nucleic acids research*, 2008, 36(Database issue): D623–31.
- [8] Kelder T, Van Iersel M P, Hanspers K, et al. WikiPathways: building research communities on biological pathways[J]. *Nucleic acids research*, 2012, 40(Database issue): D1301–7.
- [9] Becker M Y, Rojas I. A graph layout algorithm for drawing metabolic pathways[J]. *Bioinformatics*, 2001, 17(5): 461–467.
- [10] Kojima K, Nagasaki M, Jeong E, et al. An efficient grid layout algorithm for biological networks utilizing various biological attributes[J]. *BMC bioinformatics*, 2007, 8: 76.
- [11] Schreiber F. High quality visualization of biochemical pathways in BioPath[J]. *In silico biology*, 2002, 2(2): 59–73.
- [12] Klukas C, Schreiber F. Dynamic exploration and editing of KEGG pathway diagrams[J]. *Bioinformatics* (Oxford, England), 2007, 23(3): 344–50.
- [13] Li W, Kurata H. A grid layout algorithm for automatic drawing of biochemical networks[J]. *Bioinformatics*, 2005, 21(9): 2036–42.
- [14] Genc B, Dogrusoz U. A Constrained , Force-Directed Layout[J]. 2004: 314–319.

- [15] Dogrusoz U, Giral E, Cetintas A. A Compound Graph Layout Algorithm[J]. 2004: 442–447.
- [16] Itoh T, Muelder C. A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs[J]. In Visualization Symposium, 2009. PacificVis' 09. IEEE Pacific, IEEE, 2009,121-128.
- [17] Kojima K, Nagasaki M, Miyano S. An efficient biological pathway layout algorithm combining grid-layout and spring embedder for complicated cellular location information.[J]. BMC bioinformatics, 2010, 11: 335.
- [18] Tsay J-J, Wu B-L, Jeng Y-S. Hierarchically organized layout for visualization of biochemical pathways.[J]. Artificial intelligence in medicine, 2010, 48(2-3): 107–17.
- [19] Barsky A, Munzner T, Gardy J, et al. Cerebral: visualizing multiple experimental conditions on a graph with biological context.[J]. IEEE transactions on visualization and computer graphics, 2008, 14(6): 1253–60.
- [20] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks.[J]. Genome research, 2003, 13(11): 2498–504.
- [21] Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.[J]. Molecular biology and evolution, 2007, 24(8): 1596–9.
- [22] Longabaugh W J R, Davidson E H, Bolouri H. Visualization, documentation, analysis, and communication of large-scale gene regulatory networks.[J]. Biochimica et biophysica acta, 2009, 1789(4): 363–74.
- [23] Genmapp. genmapp[EB/OL]. . <http://www.genmapp.org/>.
- [24] Lachmann A, Ma'ayan A. Lists2Networks: integrated analysis of gene/protein lists.[J]. BMC bioinformatics, 2010, 11: 87.
- [25] Hu Z, Mellor J, Wu J, et al. VisANT: an online visualization and analysis tool for biological interaction data.[J]. BMC bioinformatics, 2004, 5: 17.
- [26] Hulsen T, De Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams.[J]. BMC genomics, 2008, 9: 488.
- [27] Babur Ö, Dogrusoz U, Çakır M, et al. Integrating biological pathways and genomic profiles with ChiBE 2.[J]. BMC genomics, 2014, 15: 642.
- [28] Di Lena P, Wu G, Martelli P L, et al. MIMO: an efficient tool for molecular interaction maps overlap.[J]. BMC bioinformatics, 2013, 14: 159.

- [29] Aboukhalil R, Fendler B, Atwal G S. Kerfuffle: a web tool for multi-species gene colocalization analysis.[J]. BMC bioinformatics, BMC Bioinformatics, 2013, 14(1): 22.
- [30] Yamada T, Letunic I, Okuda S, et al. iPath2.0: interactive pathway explorer.[J]. Nucleic acids research, 2011, 39(Web Server issue): W412–5.
- [31] Latendresse M, Karp P D. Web-based metabolic network visualization with a zooming user interface.[J]. BMC bioinformatics, BioMed Central Ltd, 2011, 12(1): 176.
- [32] Zhu Y. PathBubbles[EB/OL]. <https://github.com/yongnanzhu/PathBubbles>.
- [33] Wu K, Chen J, Pruett W A, et al. Hummod browser: An exploratory visualization tool for the analysis of whole-body physiology simulation data[C]//2013 IEEE Symposium on Biological Data Visualization (BioVis). IEEE, 2013(c): 97–104.
- [34] Lex A, Partl C, Kalkofen D, et al. Entourage: visualizing relationships between biological pathways using contextual subsets.[J]. IEEE transactions on visualization and computer graphics, 2013, 19(12): 2536–45.
- [35] Keqin Wu, Liang Sun, Yongnan Zhu, Carl Schmidt J C. PathBubbles: A Dynamic, Scalable, and Synchronous Visualization Environment for Pathway Analysis(to be submitted)[R].
- [36] Li G, Bragdon A, Pan Z. VisBubbles: a workflow-driven framework for scientific data analysis of time-varying biological datasets (SAP-0281)[J]. SIGGRAPH Asia Posters. Hong Kong, China: ACM, 2011.
- [37] Bragdon A, Zeleznik R, Reiss S P, et al. Code bubbles[C]//Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. New York, USA: ACM Press, 2010: 2503.
- [38] Three.js. threejs[EB/OL]. . <http://threejs.org/>.
- [39] Holten D, Van Wijk J J. A user study on visualizing directed edges in graphs[C]//Proceedings of the 27th international conference on Human factors in computing systems - CHI 09. New York, USA: ACM Press, 2009: 2299.
- [40] Stasko J, Zhang E. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations[C]//IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings. IEEE Comput. Soc, 2000: 57–65.
- [41] Mbostock. D3.js[EB/OL]. . <http://d3js.org/>.
- [42] Murray S. Interactive Data Visualization for the Web[M]. “O’Reilly Media, Inc.,” 2013: 272.

- [43] Battista G Di, Eades P, Tamassia R, et al. Graph Drawing: Algorithms for the Visualization of Graphs[J]. Prentice Hall PTR, 1998.
- [44] 张磊. 第十四章 基因表达与调控[EB/OL].
<http://course.tjau.edu.cn/yichuanxue/html/wangluojiaoxue/wangluo-4-08.htm>.
- [45] Liang Sun*, Susan J. Lamont², Fiona McCarthy³, K. Vijay-Shanker⁵, Rachael M. DeRita¹, Max Rothschild², Chris Ashwell⁴, Michael E. Persia² C J S. Transcriptome response to heat stress in a chicken hepatocellular carcinoma cell line(To be submitted)[R].
- [46] Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk.[J]. Behavior research methods, 2012, 44(1): 1–23.
- [47] Purchase H C. Metrics for Graph Drawing Aesthetics[J]. 2002: 501–516.
- [48] Purchase H, Plimmer B. Graph drawing aesthetics in user-sketched graph layouts[J]. Proceedings of the Eleventh Australasian Conference on User Interface-Volume 106, 2010, 106.
- [49] Lambert a., Dubois J, Bourqui R. Pathway Preserving Representation of Metabolic Networks[J]. Computer Graphics Forum, 2011, 30(3): 1021–1030.

附录:

攻读硕士期间主要的研究成果

在读期间发表的学术论文:

- [1] **Yongnan Zhu**, Liang Sun, Carl Schmidt, Keqin Wu, Zhigeng Pan, Jinglong Fang, JianChen. *PathRings: an web-based tool for Exploration of Ortholog and Expression Data in Biological Pathways*[J], Bioinformatics (under the review).(SCI)
- [2] Liang Sun*, **Yongnan Zhu***, Catalina O. Tudor, Ashique Mahmood. Jia Ren, Vijay Shanker, Jian Chen, Carl Schmidt. *WebGIVI: A Web-based Gene Enrichment Analysis and Visualization Tool*[J], Bioinformatics (under the review).(SCI)
- *These authors contributed equally to this work
- [3] Keqin Wu, Liang Sun, **Yongnan Zhu**, Carl Schmidt, and Jian Chen. *PathBubbles: A Dynamic, Scalable, and Synchronous Visualization Environment for Pathway Analysis*[J], IEEE Transactions on Visualization and Computer Graphics. (under the review).(SCI)
- [4] Jian Chen, Judith E. Terill, Henan Zhao, Guohao Zhang, Keqin Wu, Alexander Garbrino, **Yongnan Zhu**. Interactive Visual Computing Laboratory Research[C].
- [5] Jian Chen, Andrew Bragdon, C, Guangxia Li, **Yongnan Zhu**, Zhigeng Pan, Sharon Swartz. *MotionTrail: An Exploratory Visualization Environment for Bat Flight Motion Data Analysis*[J], Visual Computing. (under the review).(SCI)
- [6] (In preparation - should be done in about three months) Jian Chen, **Yongnan Zhu**, Alexander P. Auchus. *A Review of Diffusion Tensor MRI Visualization Approaches: A semiotic Perspective*[J], IEEE Transactions on Visualization and Computer Graphics.(SCI)
- [7] (In preparation - should be done in about three months) Yanning Xu, Alexander P Auchus, Stephen Correia, **Yongnan Zhu**, Jian Chen, *Perceptually Motivated Global Illumination by, Multi-type Light Sources for Dense Streamribbon Visualization*[J], IEEE Transactions on Visualization and Computer Graphics.(SCI)

参加的科研项目:

- [1] Site-PI: ABI Development: PathBubbles for Dynamic Visualization and Integration of Biological Information, NSF DBI-1260795, 2012-2015