

# Cytoscape用户手册 中文版

The Cytoscape Collaboration

该中文版由陈钢 ([www.gossipcoder.com](http://www.gossipcoder.com)) 翻译

中文翻译项目网站: <http://code.google.com/p/cytoscape-cn/>

March 1, 2010

# Contents

0.1	开发	4
0.2	授权	4
0.3	2.6版本的更新	4
<b>1</b>	<b>启动 Cytoscape</b>	<b>6</b>
1.1	系统要求	7
1.2	入门	7
1.2.1	安装 Java	7
1.2.2	安装 Cytoscape	7
1.2.3	启动程序	7
<b>2</b>	<b>Cytoscape快速入门</b>	<b>11</b>
2.1	菜单	13
2.1.1	File	13
2.1.2	Edit	14
2.1.3	View	14
2.1.4	Select	14
2.1.5	Layout	14
2.1.6	Plugins	15
2.1.7	Help	15
2.2	网络管理	15
2.2.1	网络窗口排列	16
2.3	网络概览窗口	16
<b>3</b>	<b>命令行参数</b>	<b>18</b>
<b>4</b>	<b>Cytoscape 设置</b>	<b>20</b>
4.1	管理属性	21
4.1.1	设置属性的缺省值	21
4.2	书签的管理	22
4.2.1	代理服务器的设置	22
<b>5</b>	<b>创建网络</b>	<b>23</b>
5.1	导入确定格式的网络文件	24
5.1.1	从本地计算机上导入网络	24
5.1.2	从远程计算机上导入网络 (URL导入)	24
5.2	导入格式灵活的表格文件	25
5.2.1	支持的文件类型	25
5.2.2	基本操作	26
5.2.3	导入不带边的节点列表	27
5.2.4	高级选项	27
5.2.5	修改属性名称和类型	28
5.3	从 Web 服务导入网络	28
5.4	编辑新网络	28

<b>6</b>	<b>所支持的网络文件格式</b>	<b>29</b>
6.1	SIF 格式 . . . . .	30
6.2	GML 格式 . . . . .	31
6.3	XGMML 格式 . . . . .	31
6.4	系统生物学标记语言 (Systems Biology Markup Language) . . . . .	31
6.5	BioPAX (Biological PATHways eXchange)格式 . . . . .	31
6.6	PSI-MI 格式 . . . . .	31
6.7	纯文本表格和Excel表格 . . . . .	32
6.8	Cytoscape 中节点的命名 . . . . .	32
<b>7</b>	<b>节点和边的属性</b>	<b>34</b>
<b>8</b>	<b>加载基因表达 (属性矩阵) 数据</b>	<b>41</b>
8.1	数据文件格式 . . . . .	42
8.2	基本过程 . . . . .	42
8.3	实例 . . . . .	42
8.4	文件格式的详细说明 (高级用户) . . . . .	43
<b>9</b>	<b>从外部数据库导入网络和属性</b>	<b>45</b>
<b>10</b>	<b>浏览和布局</b>	<b>46</b>
<b>11</b>	<b>可视化风格</b>	<b>47</b>
11.1	什么是视觉风格? . . . . .	48
11.2	VizMapper用户界面简介 . . . . .	54
11.3	Introduction to Visual Styles . . . . .	56
11.4	Visual Attributes, Graph Attributes and Visual Mappers . . . . .	57
<b>12</b>	<b>节点和边的查找与过滤</b>	<b>73</b>
<b>13</b>	<b>编辑网络</b>	<b>74</b>
<b>14</b>	<b>插件和插件管理器</b>	<b>75</b>
<b>15</b>	<b>CytoPanel</b>	<b>76</b>
<b>16</b>	<b>渲染引擎</b>	<b>77</b>
<b>17</b>	<b>注释</b>	<b>78</b>
<b>18</b>	<b>链接</b>	<b>79</b>

# Cytoscape 2.6用户手册

本文档遵守Creative Commons License, 2006

作者: The Cytoscape Collaboration

中文翻译: 陈钢 Gang Chen chengang@gossipcoder.com

如需要最新的L<sup>A</sup>T<sub>E</sub>X代码, 请访问: <http://code.google.com/p/cytoscape-cn/>

Cytoscape项目由以下单位合作:

- 加州大学圣地亚哥分校
- 系统生物学研究中心
- Memorial Sloan-Kettering癌症研究中心
- Pasteur研究中心
- 安捷伦科技公司
- 加州大学旧金山分校

Cytoscape的资金来自NIH的美国国家通用医学研究中心 (NIGMS), 资金编号为: GM070743-01。整体资金通过来自Unilever PLC的合同提供。

# 引言

Cytoscape 项目致力于为用户提供一个开源的网络显示和分析软件。软件的核心部分提供了网络显示、布局、查询等方面的基本功能。软件的核心可以通过插件架构进行扩展，这样就能快速地开发出新的功能。

Cytoscape源自系统生物学，用于将生物分子交互网络与高通量基因表达数据和其他的分子状态信息整合在一起。虽然Cytoscape也能适用于其他分子构件和相互作用，但其最强大的功能还是用于大规模蛋白质-蛋白质相互作用、蛋白质-DNA和遗传相互作用的分析。各种物种，包括人类，的这方面的实验数据都在迅速增加。通过Cytoscape，用户可以在可视化的环境下将这些生物网络跟基因表达、基因型等各种分子状态信息整合在一起，还能将这些网络跟功能注释数据库链接在一起。

Cytoscape 的核心是网络（图），其中的节点（node）是基因、蛋白质或分子，其中的连接则是这些生物结构之间的相互作用。

## 0.1 开发

Cytoscape 是 Institute for Systems Biology（Leroy Hood 实验室）、加州大学圣地亚哥分校（Trey Ideker 实验室）、Memorial Sloan-Kettering 癌症研究中心（Chris Sander 实验室）、Pasteur研究院（Benno Schwikowski实验室）、安捷伦科技（Annette Adler实验室）和加州大学旧金山分校（Bruce Conklin实验室）的合作项目。

详情请访问<http://www.cytoscape.org>。

## 0.2 授权

Cytoscape 受 GNU LGPL（Lesser General Public License）的保护。在本手册的附录中找到该授权，同时可以访问 <http://www.gnu.org/copyleft/lesser.txt>。Cytoscape 还是用了其他的一些开源程序库，详情见本手册的致谢。

## 0.3 2.6版本的更新

Cytoscape 2.6 中增加了很多新功能，在性能和软件的易用性上也有提升。包括：

- Web Service Client Manager框架能将Web服务客户端集成到Cytoscape中。
- 通过Web服务客户端插件，可以从PathwayCommons、IntAct和NCBI Entrez Gene下载网络数据。
- 通过Web服务插件，可以从BioMart导入注释信息。这主要是用于ID的翻译和名称映射。
- Cytoscape主题。 Dynamic filters.
- 动态过滤。 Network Manager supports multiple network selection.
- 网络管理器支持多网络选取。 Label Positioning has been improved.
- 改进了标签的位置。 Session saving occurs in memory.
- 将会话保存在内存中。 XGMML Improvements.
- 改进了XGMML。 Network loading improvements.
- 网络加载得到了改进。
- Linkout integrated with attribute browser.

- 通过可视化属性，引入了更多的Visual Style。
- 修复了不计其数的bug。

## Chapter 1

# 启动 Cytoscape

Cytoscape 是一个Java程序，能在 Linux、Windows 和 Mac OS X 上运行。对于其他能安装Java 5的操作系统平台，比如以Solaris和FreeBSD为代表的UNIX，Cytoscape也能运行，但官方并对此提供支持。

## 1.1 系统要求

Cytoscape 对系统的具体要求取决于所加载、查看和操作的网络的大小。

	小型网络查看	大型网络分析和查看
处理器	1GHz	尽可能的快
内存	512MB	2GB以上
显卡	板载集成显卡	高端独立显卡
显示器	XGA(1024 × 768)	高分辨率或双显示器

表 1.1:

## 1.2 入门

### 1.2.1 安装 Java

如果计算机上还没有安装 Java，那么首先要下载并安装 Java SE 5 或 6。Cytoscape 从2.5版开始就不能在 Java 1.4 上运行。必须安装 Java SE 5 或 6 !!! Java SE 5 和 Java SE 6可以从这里下载:

- Java SE 5
- Java SE 6

一般情况下，Java SE 6 的运行速度要快一些。所以，如果您的计算机兼容 Java SE 6 的话，请尽量使用 Java SE 6。

### 1.2.2 安装 Cytoscape

Cytoscape 供下载的版本很多，安装方法也不尽相同。所有的版本都可以从 <http://cytoscape.org> 网站下载。

- Windows、Mac OS以及Linux平台上的自动安装包
- 压缩发行版
- 从源代码编译
- 从Subversion源中提取最新版的软件

Cytoscape的安装目录（无论是什么平台）中包含表 1.2 中的文件。

### 1.2.3 启动程序

双击安装程序创建的图标，或是在命令行中运行 `cytoscape.sh` (Linux或Mac OS X)，也可以双击 `cytoscape.bat` (Windows) 就能启动 Cytoscape。还可以在命令行中将这个jar文件以命令行参数的形式传递给命令 `java -Xmx512M -jar cytoscape.jar -p plugins`。-Xmx512M 标志是告诉 java 给 Cytoscape 分配较多的内存，-p plugins 则是告诉 Cytoscape 加载 plugins 目录中所有的插件。插件对于 Cytoscape 是非常重要的，诸如布局 (layout)、过滤器 (filter) 和属性浏览器之类的功能都是由插件提供的。关于命令行参数的详细信息请阅读命令行参数一章。在 Windows 中，只需要双击 jar 文件就能启动 Cytoscape。不过，这样就不能使用命令行参数了（比如制定插件目录的位置）。

成功地启动 Cytoscape后，就会看到如图 1.1 所示的窗口（图 1.1 来自Mac OS 10.4）。

#### 注意内存使用量

随着用用户所加载的网络的规模的增加，Cytoscape所需的内存也会增加。内存的使用量取决于网络对象（节点和边）的数量以及属性的数量。表 1.3 和 1.4是对内存需求量的粗略估计。

文件	描述
cytoscape.jar	Cytoscape 的主程序（Java 压缩包）。
cytoscape.sh	从命令行运行 Cytoscape 的脚本（用于 Linux 和 Mac OS X）。
cytoscape.bat	运行Cytoscape的脚本（用于 Windows）。
LICENSE.txt/html	Cytoscape GNU LGPL 授权。
lib/	Cytoscape运行所需的jar库。
docs/	各种格式的用户手册。也就是你正在阅读的东西。
plugins/	jar格式的Cytoscape插件。
sampleData/	
	galFiltered.gml – 分子相互作用网络数据示例*。
	galFiltered.sif – Simple Interaction 格式的同个网络*
	galExpData.pvals – 基因表达矩阵文件示例*。
	galFilteredAttrTable.xls – 微软 Excel 格式的节点属性文件示例。
	galFiltered.sif – 用上面的数据库和多个注释数据库创建的会话示例*。
	BINDyeast.sif – BIND数据库中酵母的蛋白质相互作用网络，2006年12月**。
	BINDhuman.sif – BIND数据库中人类的蛋白质相互作用网络，2006年12月**。
	yeastHighQuality – 分子生物相互作用的示例文件***。
	interactome_merged_networkTable.gz – 制表符分割格式的人类相互作用网络****。
	sampleStyle.props – 附加的 Visual Sytle 示例。

表 1.2: \* 来自 Ideker et al, Science 292:929 (2001); \*\* 来自 [http://www.blueprint.org/bind/bind\\_downloads.html](http://www.blueprint.org/bind/bind_downloads.html); \*\*\* 来自 Mering et al, Nature, 417:399 (2002) Lee et al, Science 298:799 (2002); \*\*\*\*来自 Cytoscape 教程网页。原始数据可以从 [http://cytoscape.org/cgi-bin/moin.cgi/Data\\_Sets/](http://cytoscape.org/cgi-bin/moin.cgi/Data_Sets/)上由 Andrew Garrow、Yeyejide Adeleye和 Guy Warner (Unilever, Safety and Environmental Assurance Center) 创建的“A merged human interactome”中下载。

对象的数量（节点和边）	建议内存
0 – 70,000	512M（默认值）
70,000 – 150,000	800M

表 1.3: 无视图情况下的建议内存大小

## Cytoscape的整体内存需求

可以通过命令行参数增加Cytoscape的内存大小。例如，如果要给Cytoscape分配1G的内存，可以在命令行输入：

```
java -Xmx1GB -jar cytoscape.jar -p plugins
```

### 堆栈尺寸 (stack size)

这是另一个跟内存分配有关的选项。Cytoscape的部分功能需要较大的堆栈空间（某些操作所需的临时内存，比如Layout）。由于这个堆栈的大小是独立于前面的Xmx值的，所以有时候Layout算法会因为内存不足而失败。为了避免这种情况，可以用-Xss制定更大的堆。如果对大型网络布局时失败，可以尝试下面的命令：

```
java -Xmx1GB -Xss10M -jar cytoscape.jar -p plugins
```

选项-Xss10M的意思就是将堆的尺寸设置为10MB。在大部分情况下，这能解决由于内存不足导致的Layout问题。

注意：有些web服务客户端是多线程程序，其中没有线程都是使用-Xss选项所指定的内存大小。如果web服务客户端应为内存不足而失败，请调低堆栈的大小，然后再试一次。

更多的信息请参看[http://cytoscape.org/cgi-bin/moin.cgi/How to increase memory 20for Cytoscape](http://cytoscape.org/cgi-bin/moin.cgi/How_to_increase_memory_20for_Cytoscape)

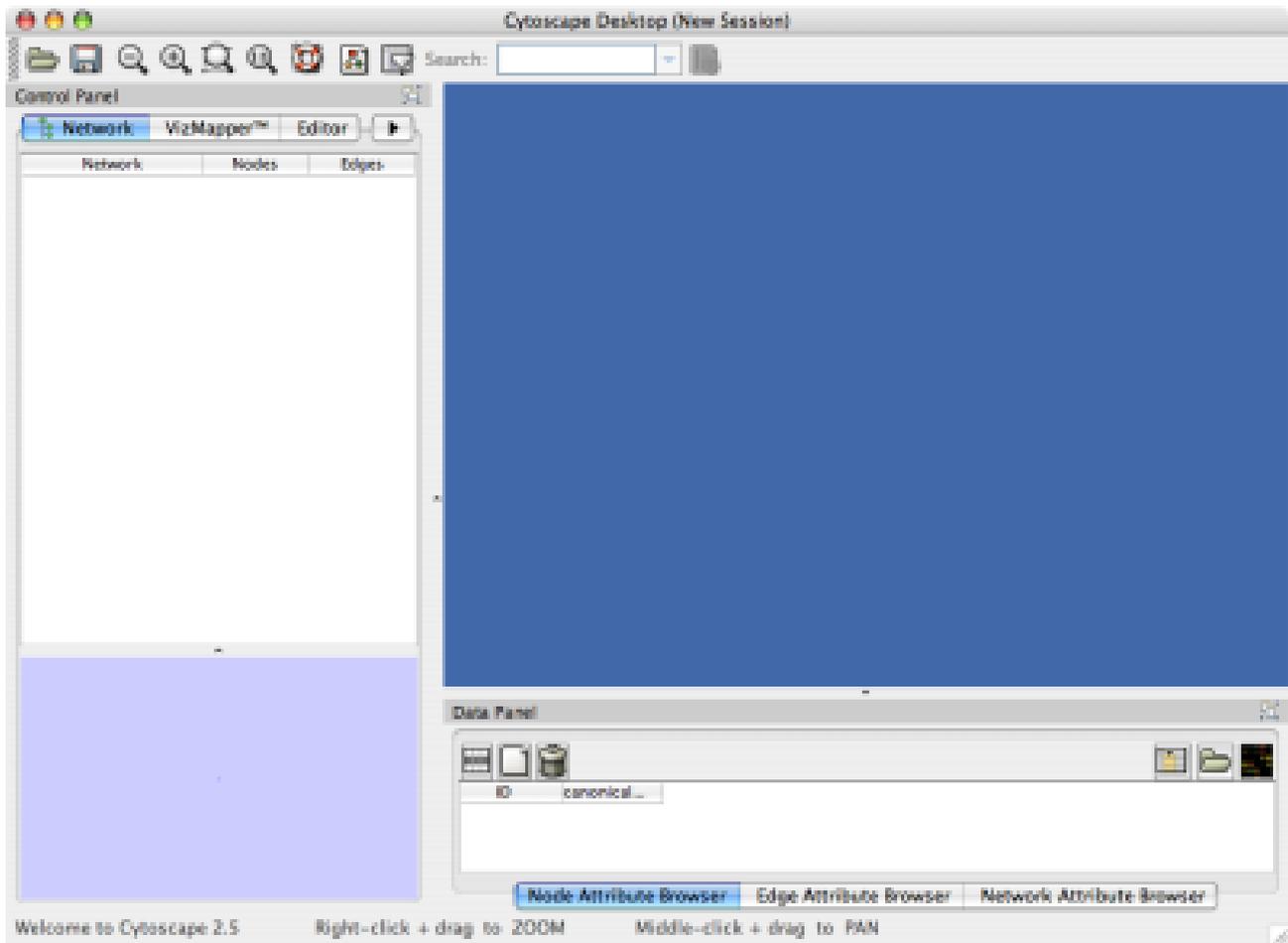


图 1.1: Cytoscape启动界面

对象的数量 (节点和边)	建议内存
0 – 20,000	512M (默认值)
20,000 – 70,000	800M
70,000 – 150,000	1G

表 1.4: 有视图情况下的建议内存大小

### 注意目录位置

为了能让程序正常运行，在解压后，所有的文件位置都不能随便改动。Cytoscape的核心程序会按照默认的目录结果去查找其所需的库。如果你确实手痒痒想自己改一下，可以通过修改\$CLASSPATH或是cytoscape.jar的manifest文件，这样在任何地方都能启动Cytoscape。

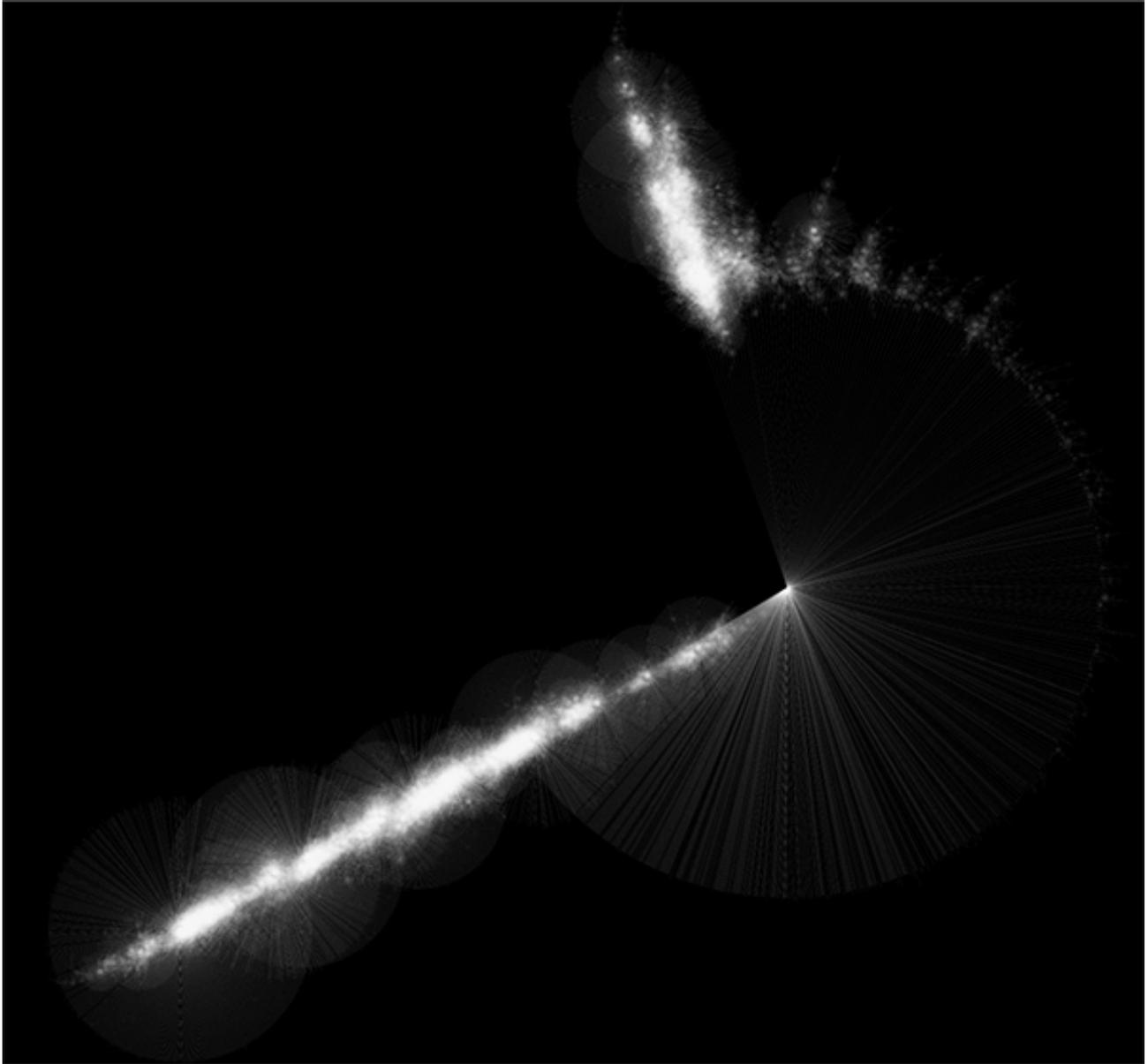


图 1.2: 随机生成的scale-free网络, 含有500K个节点和500K条边: 如果正确地设置了内存参数, 就能看到这个庞大的网络。在这个例子中, Cytoscape 大约使用了 5GB 内存。堆栈的大小是10MB。如果要使用大内存 (大于4GB), 就需要64位的操作系统和64位的Java SE 5或6。

## Chapter 2

# Cytoscape快速入门

加载了网络之后，Cytoscape 的界面如图 2.1 所示。

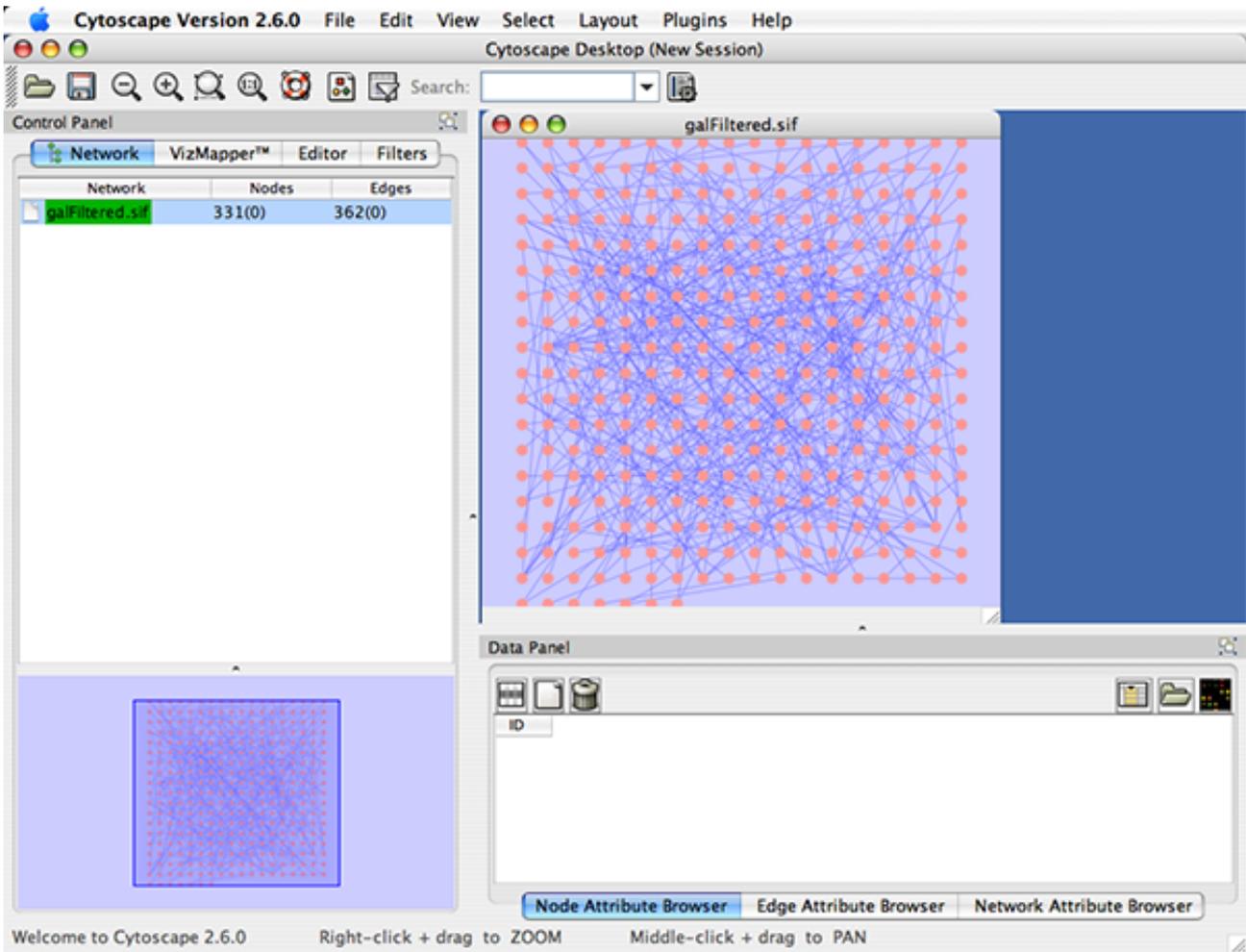


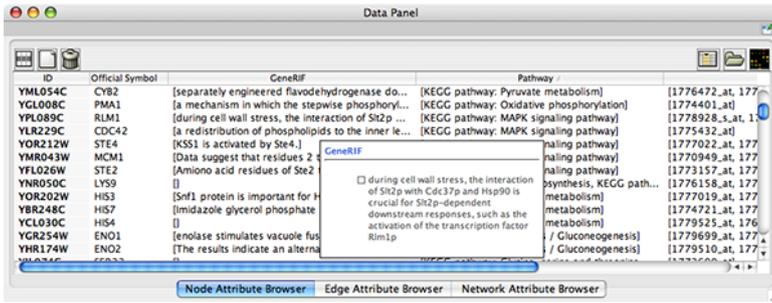
图 2.1: Cytoscape 加载网络后的界面截图

主界面由多个组件构成，包括：

- 顶部的菜单（稍后会对各个菜单做详细介绍）。
- 工具栏，包含了各种常用功能。从菜单中也能使用这些功能。鼠标指针在这些图标上停留片刻就能看到有关的提示。
- 网络管理面板（左上方的面板）。其中有可关闭的网络全局浏览面板（左下方）。
- 网络查看主窗口，网络就显示在这个窗口中。
- 属性浏览器面板（底部的面板），显示所选择的节点和边的属性。在这个面板中还可以对这些属性的值进行修改。

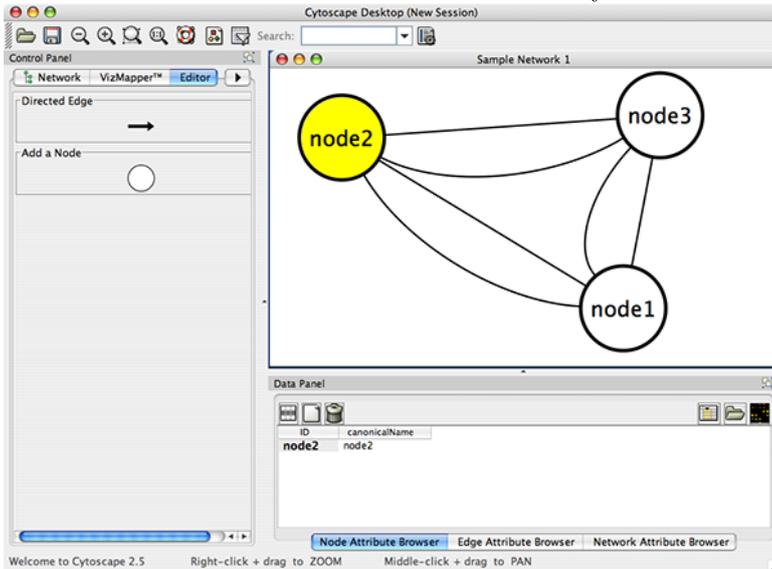
网络管理和属性浏览器面板是可以拖拽的标签面板，被称为 CytoPanels。通过点击 CytoPanel 右上角的浮动窗口 (Float Window) 控件  可以将这些面板设为浮动状态。

如果选择了这个控件，例如属性浏览器面板上的这个控件，就会出现两个 Cytoscape 的窗口，一个是主窗口，另一个是名为 CytoPanel 2 的新窗口，如下图所示。当鼠标指针指向单元格时，会看到弹出信息。



在图中可以看到，CytoPanel 2 现在有一个 Dock Window 控件。如果点选这个控件，这个窗口就会回到主窗口中。

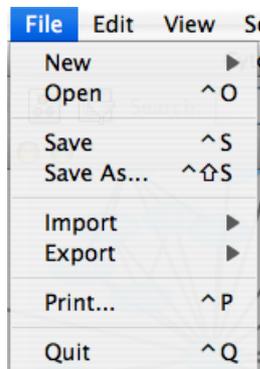
Cytoscape 还提供了一个用于构建和编辑网络的编辑器，把面板中的节点和边拖拽到主网络视图窗口中就能创建和编辑网络。用 Visual Style 可以定义节点的形状以及边的箭头。只需要选择 CytoPanel 1 中的 Editor 标签就能编辑网络。下图展示了一个编辑器，Visual Style 是 BioMoleculeEditor。



## 2.1 菜单

### 2.1.1 File

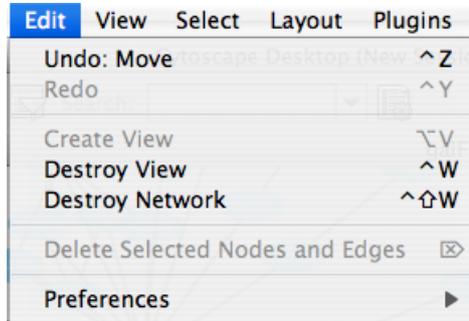
File 菜单中含有最基本的文件操作功能：File → Open 用于打开 Cytoscape 会话文件；File → New 用于新建空白网络，也可以从现用的网络创建新网络；File → Save 用于保存会话文件；File → Import 用于导入网络或属性数据；File → Export 用于导出数据和图片。File → Print 用于打印，File → Quit 则是关闭所有的 Cytoscape 窗口，并推出程序。



### 2.1.2 Edit

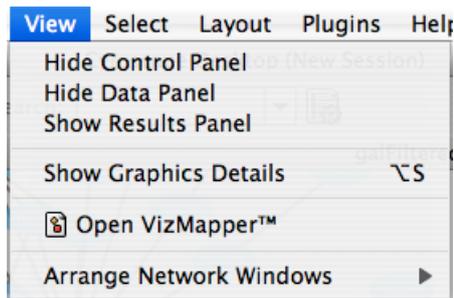
Edit菜单中提供了用于属性浏览器、网络编辑器和布局的撤销（Undo）和重做（Redo）功能。

还有创建和销毁视图（网络的显示方法）和网络（网络的原始数据，并不可可视化的）的菜单项，以及用于从当前网络中删除所选节点和边的菜单项。通过 Edit → Undo 可以恢复被删除的节点和边。配置和插件的设置可以通过Edit → Preferences → Properties来编辑。



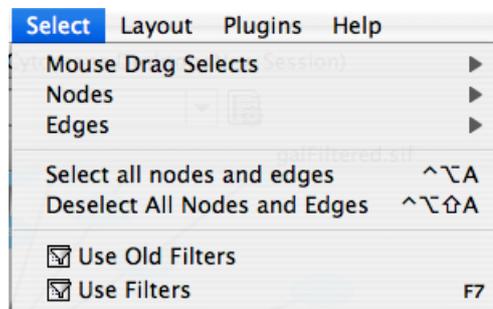
### 2.1.3 View

在 View 菜单中可以控制网络管理面板（CytoPanel 1）、属性浏览器（CytoPanel 2）、网络概览（在CytoPanel 1中）和VizMapper的显示和隐藏。



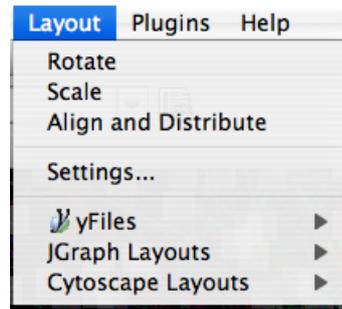
### 2.1.4 Select

在 Select 菜单中含有选择节点和边的各种选项。还有Select → Use Filter选项，用于根据节点或边的某些属性创建自动的过滤器。



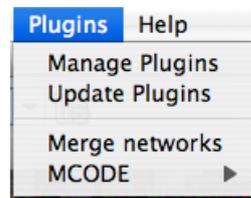
### 2.1.5 Layout

通过 Layout 菜单可以控制网络显示的形式。该菜单上半部分（Rotate, Scale, Align and Distribute）是用于控制网络的显示。菜单的下半部分是用于自动布局网络的各种布局算法。



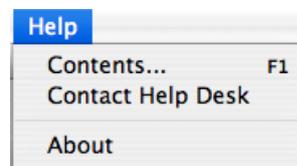
### 2.1.6 Plugins

Plugins 菜单中提供了插件管理功能（安装、升级和删除），以及插件所添加的选项，比如 Agilent Literature Search 或 Merge Networks。具体的内容取决于所加载的插件。<sup>1</sup>



### 2.1.7 Help

在Help菜单中可以打开在线帮助，浏览本手册内容。“About...”则会显示正在运行的 Cytoscape 的版本信息。



## 2.2 网络管理

Cytoscape 2.3之后的版本都可以同时加载多个网络，有无视图都可以。网络中存放着用户加载所有节点和边，以及用户做指定的视图。同一个网络可以有多个视图。网络（以及相应的视图）可以有序地组织在一起。

下图中展示了多个网络同时加载，并按层次结构组织的例子：

网络管理器（CytoPanel 1右上方的树状显示窗口）展示了当前所加载的网络。点击其中的网络就会在主窗口里面显示该网络的视图。在网络管理器中会显示各个网络的名称和尺寸（节点和边的数量）。如果网络是从文件中加载的，网络的名称就是文件名。

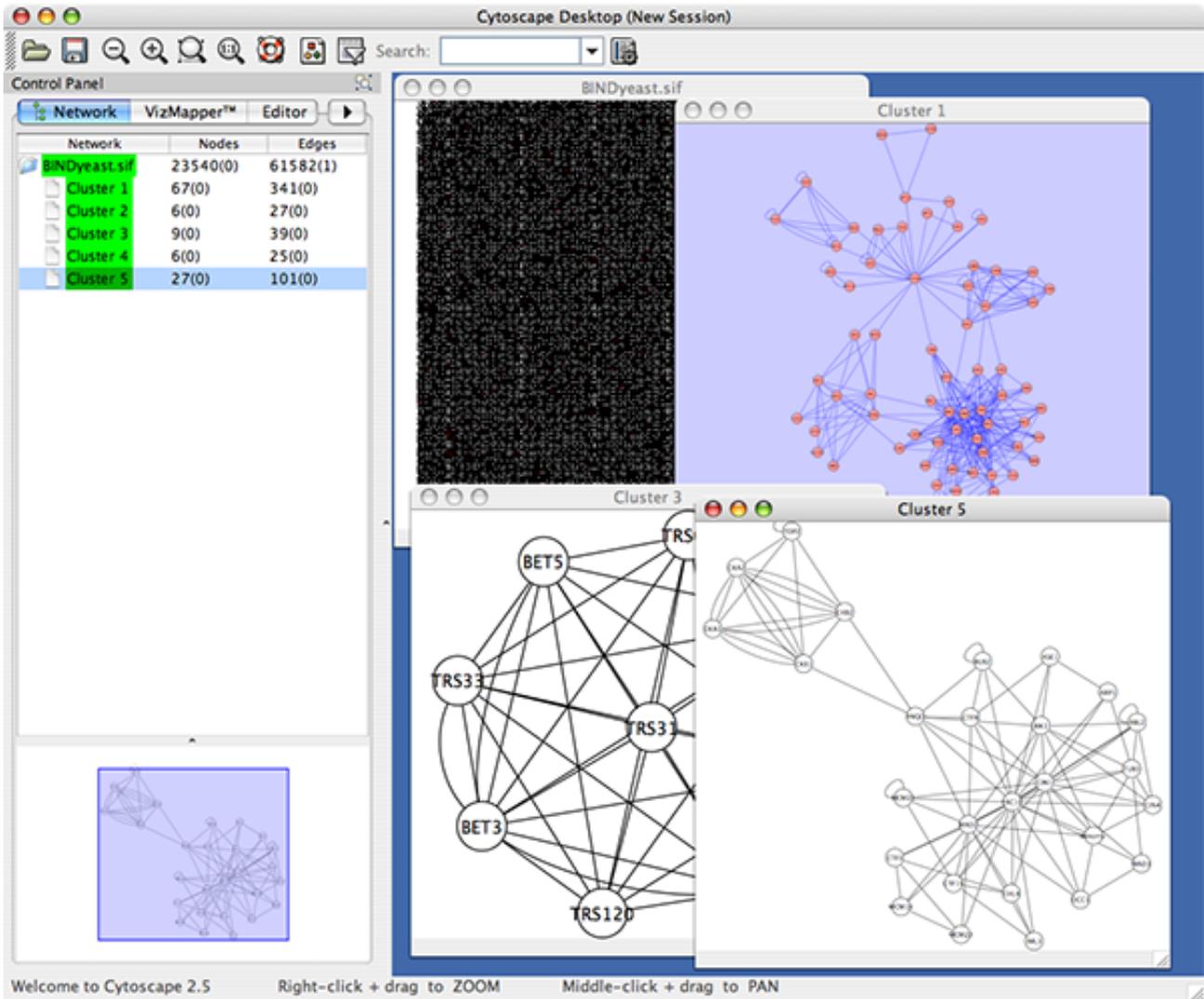
对于那些规模较大的网络（数千个点和边），需要较长的时间才能显示出来。正因为如此，Cytoscape中的网络可以不包含“视图”（View）。带有视图的网络以绿色高亮显示，而没有视图的网络则以红色高亮显示。在网络管理器中网络的名称上点击鼠标右键，就能创建或是销毁视图，通过Edit菜单也能实现同样的操作。用同样的方法可以销毁已加载的网络。在上图中，加载了七个网络，其中六个绿色的是有视图的，而红色的那个则是没有视图的。<sup>2</sup>

Cytoscape中的某些特定操作会创建出新的网络。如果新网络是从原有的网络中创建的，例如选择原网络中的部分节点，将这些节点复制到新网络中（用File → New → Network option），在这种情况下新网络就会显示为原网络的子网络。Cytoscape的这种显示方法使得用户能很快地看出网络间的关系。上图中树顶部的那些网络就是这样生成的。

在网络视图窗口中，所有网络视图都会以窗口的形式显示出来。通过窗口右上角的窗口标准控件可以最大化、最小化或是销毁网络视图。

<sup>1</sup>注意：在 <http://cytoscape.org/plugins2.php> 上有Cytoscape 插件的介绍。

<sup>2</sup>译者注：图中总共只有六个网络，都是绿色的，似乎原文有误。

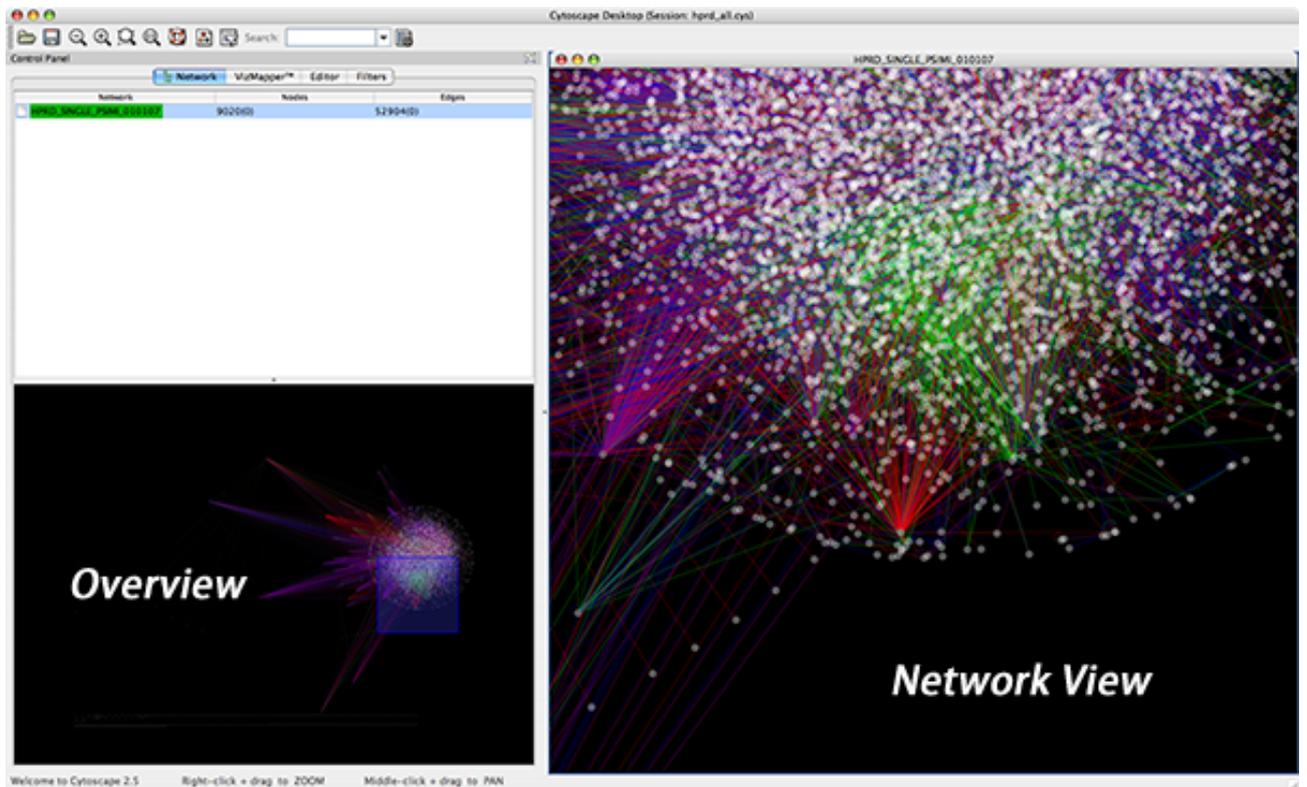
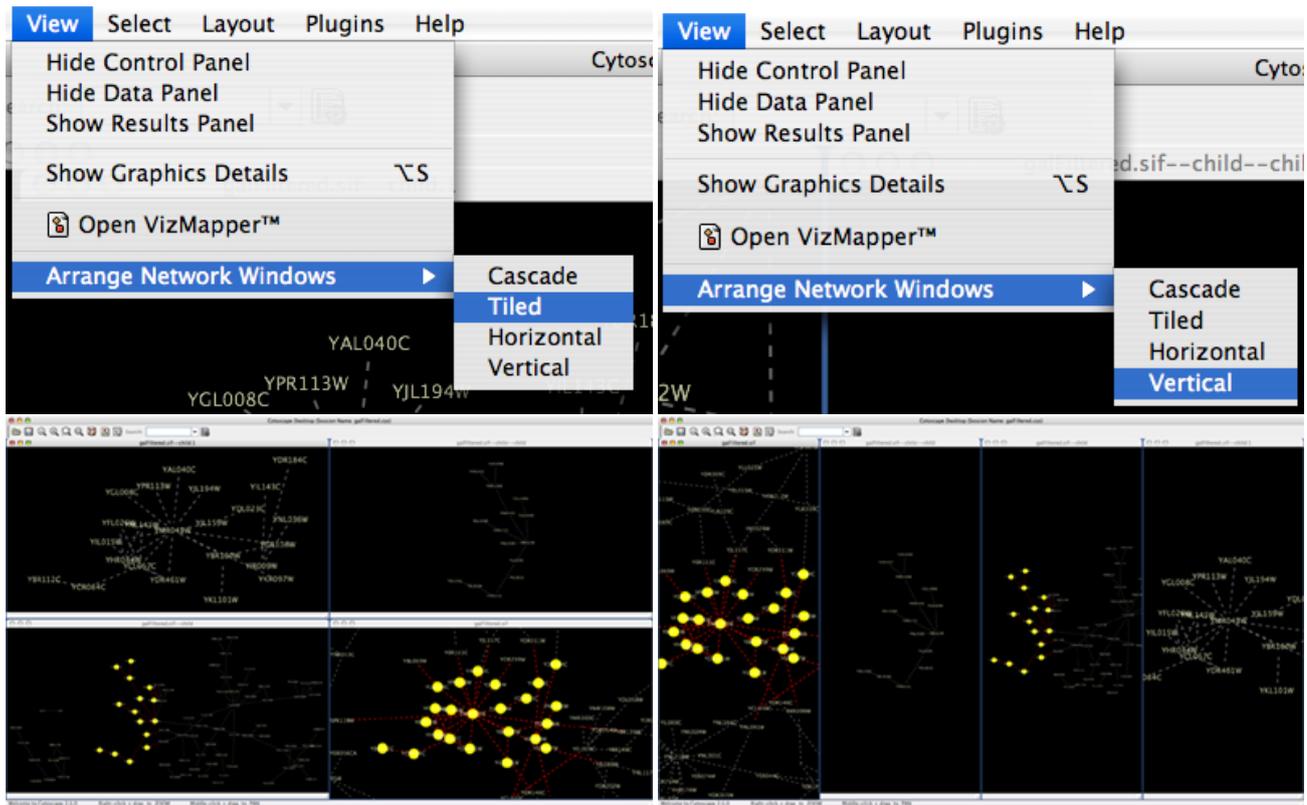


### 2.2.1 网络窗口排列

在同时处理多个网络时，可以用View→Arrange Network Windows菜单项中的命令来重新排布窗口的排列方式。

## 2.3 网络概览窗口

网络概览窗口中会显示网络的概况（或者说是“鸟瞰”）。这样就能比较方便地浏览规模较大的网络。蓝色矩形表示当前网络视图窗口中所显示的范围，用鼠标拖动这个矩形框就能浏览网络的其他部分。放大或缩小网络时，矩形框的大小也会自动的做出相应的变化。



## Chapter 3

# 命令行参数

Cytoscape 可以接受很多命令行参数，包括指定网络文件、属性文件和会话文件。下面的内容是用 Cytoscape 的“-h”或“-help”运行参数得到的：

```
usage: java -Xmx512M -jar cytoscape.jar [OPTIONS]
-h,--help          Print this message.
-v,--version       Print the version number.
-s,--session <file> Load a cytoscape session (.cys) file.
-N,--network <file> Load a network file (any format).
-e,--edge-attrs <file> Load an edge attributes file (edge attribute format).
-n,--node-attrs <file> Load a node attributes file (node attribute format).
-m,--matrix <file> Load a node attribute matrix file (table).
-p,--plugin <file> Load a plugin jar file, directory of jar files,
                  plugin class name, or plugin jar URL.
-P,--props <file> Load cytoscape properties file (Java properties
                  format) or individual property: -P name=value.
-V,--vizmap <file> Load vizmap properties file (Java properties format).
```

在指定文件时，可以是本地文件的路径，也可以是URL。例如，可以这样指定本地的网络文件（假设myNet.sif位于当前工作目录）：`cytoscape.sh -N myNet.sif`。还可以用URL指定网络：`cytoscape.sh -N http://example.com/myNet.sif`。

### 参数 用途

**-h,--help** 显示上面已经展示过的帮助信息，然后退出程序。

**-v,--version** 显示 Cytoscape 的版本号，然后退出程序。

**-s,--session <file>** 这个选项用于加载指定的会话文件。由于会话文件只能在特定的时间加载，所以这个选项在命令行中只能使用一次。要给这个选项指定一个.cys的Cytoscape会话文件。不过，cys后缀并不是必须的。

**-N,--network <file>** 这个选项用来加载各种类型的网络文件，包括SIG、GML和XGMML的各种格式的文件都可以通过-N选项加载到Cytoscape中。在一个命令中可以同时加载多个网络。

**-e,--edge-attrs <file>** 这个选项用来指定边属性文件。在一个命令中可以同指定多个边属性文件。

**-n,--node-attrs <file>** 这个选项用于指定节点属性文件。在一个命令中可以同指定多个边属性文件。

**-m,--matrix <file>** 这个选项用于指定数据矩阵文件。如果是用于生物网络分析，这个数据矩阵的内容就是基因表达数据。所有的数据矩阵文件都会以节点属性的形式读入Cytoscape。在一个命令中可以同指定多个数据矩阵文件。

**-p,--plugin <file>** 这个选项用于将指定的Cytoscape插件(.jar)文件加载到Cytoscape中。老版本中的“资源插件选项(resource plugin option)”已经合并到了这个选项中<sup>1</sup>。如果插件位于Cytoscape的CLASSPATH路径中，还可以用类名称来制定插件。例如，假设能在CLASSPATH中找到名为MyPlugin的类，那么就可以这样指定插件：`cytoscape.sh -p MyPlugin.class`。最后一种指定插件的方面就是指定一个文本文件，在这个文本文件中列出所有需要加载的插件的jar文件名。

**-P,--props <file>** 这个选项可以用来指定Cytoscape的各项属性。可以用过属性文件（格式是标准的Java属性格式），也可以用等号设置单个属性的值。例如，指定缺省的物种：`cytoscape.sh -P defaultSpeciesName=Human`。如果属性中有空格，加上引号就可以了，例如：`cytoscape.sh -P "defaultSpeciesName=Homo Sapiens"`。这个选项可以用来代替以前的-noCanonicalization,-species和-bioDataServer选项。现在的命令可以写成这样的：`cytoscape.sh -P defaultSpeciesName=Human -P noCanonicalization=true -P bioDataServer=myServer`。

**-V,--vizmap <file>** 这个选项用来指定可视化属性文件。

上面介绍的这些属性（包括插件）都可以在正在运行的Cytoscape的GUI中加载。

<sup>1</sup>这句话的翻译不确定，请有了解老版本Cytoscape的同学指正

## Chapter 4

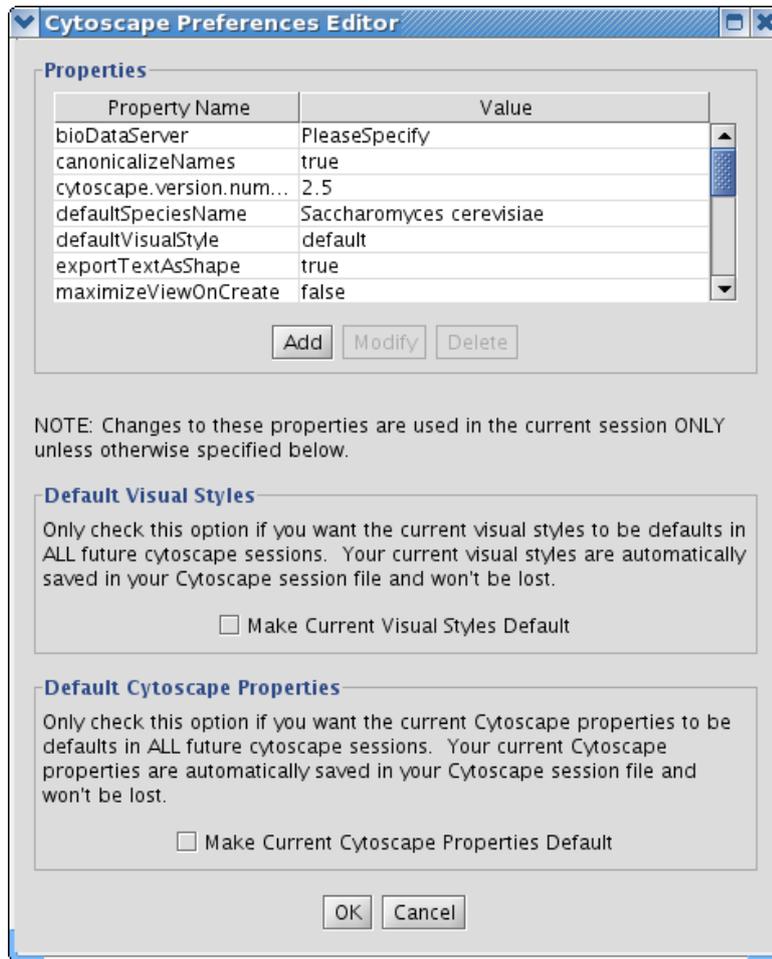
# Cytoscape 设置

## 4.1 管理属性

1

用菜单上的Edit → Preferences → Properties...可以打开Cytoscape的属性编辑器，在这里可以修改各种属性。对于一般的属性，都是保存在Cytoscape的会话文件中，所以只对当前会话有效。只有将其设为缺省属性，或是用File → Export将其导出，才能在其他的会话中生效。

可以通过 Add（添加）、Modify（修改）和 Delete（删除）操作来配置 Cytoscape 的属性。



下面介绍一些常见的属性。

属性名称	缺省值	有效值
viewThreshold	10000	大于0的整数
secondaryViewThreshold	30000	大于0的整数
viewType	tabbed	tabbed
defaultWebBrowser		指向系统默认浏览的路径。只有当 Cytoscape 找不到系统的浏览器时才需要设置这个属性。

### 4.1.1 设置属性的缺省值

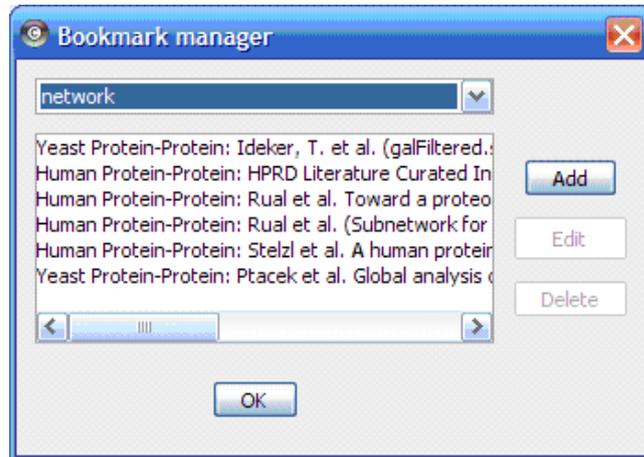
Cytoscape 的缺省属性是可以修改的。

<sup>1</sup>重要：如果你用过老版本的 Cytoscape 就会发现属性的处理方式已经发生了 改编。其中，最重要的变化是属性不再是存放在当前目录或主目录的 cytoscape 目录中，而是 保存在 Cytoscape 的会话文件中（后缀是.cys）。在 cytoscape 目录下肯定会有一个 cytoscape.props 文件，但这个文件只有当用户要求将当前属性保存为缺省属性时才会被修改。除非实在是有什么特别的原因，否则还是用缺省设置比较好。

在 Edit→Preferences→Properties... 中编辑属性，选上 Make Current Cytoscape Properties Default 。这样，当前属性就会保存到.cytoscape目录中，对每个Cytoscape会话都会生效。否则，这些属性就只会保存到当前会话的.cys文件中，新的会话依然会使用缺省属性。

## 4.2 书签的管理

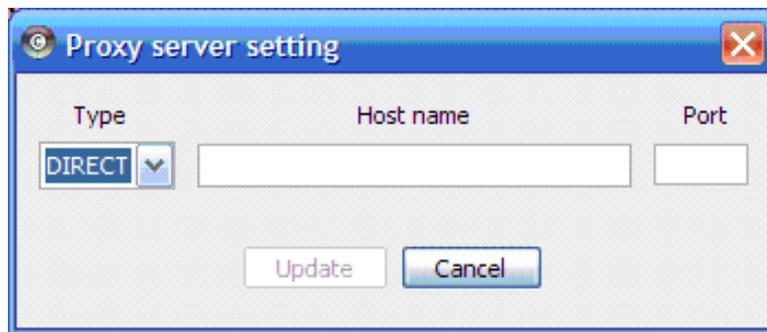
Cytoscape中预设了一些指向Cytoscape服务器上的演示网络文件的书签。用户可以通过Edit→Preferences→Bookmarks... 启动书签管理器来添加、修改和删除书签。



书签分为两类：网络书签和注释书签。网络书签是指向互联网上网络文件的URL。这些网络都可以加载到Cytoscape中。注释书签是指向ontology注释文件的URL。只有在导入ontology的时候才会用到注释书签。

### 4.2.1 代理服务器的设置

在Edit→Preferences→Proxies... 中可以设置Cytoscape的代理服务器。



设置了代理服务器之后，通过URL加载网络时的网络数据就会通过代理服务器。其他插件也可以使用这个功能。

## Chapter 5

# 创建网络

在Cytoscape中有四种创建网络的方法:

1. 导入预设的格式化网络文件。
2. 导入预设的未格式化的文本文件或Excel文件。
3. 从Web Service导入网络。
4. 创建一个空网络，然后手动地添加节点和边。

## 5.1 导入确定格式的网络文件

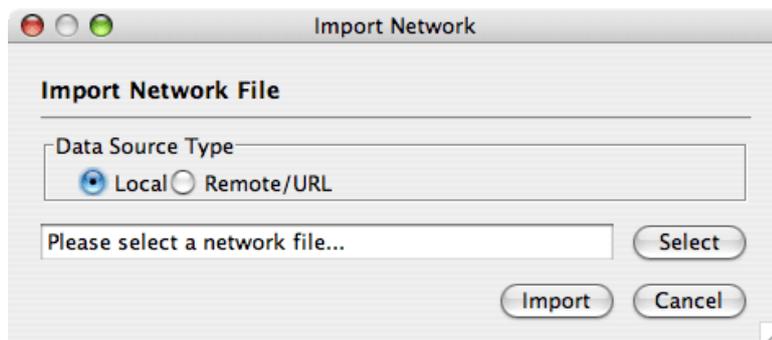
在6中介绍的所有网络文件格式都可以导入到Cytoscape中。点击File → Import → Network (multiple file types)，在弹出的“Import Network”窗口中就可以把网络文件导入到Cytoscape中。可以是本地计算机上的网络文件，也可以是远程计算机上的（用URL）。

### 5.1.1 从本地计算机上导入网络

在缺省情况下，Cytoscape会从本地计算机加载网络。

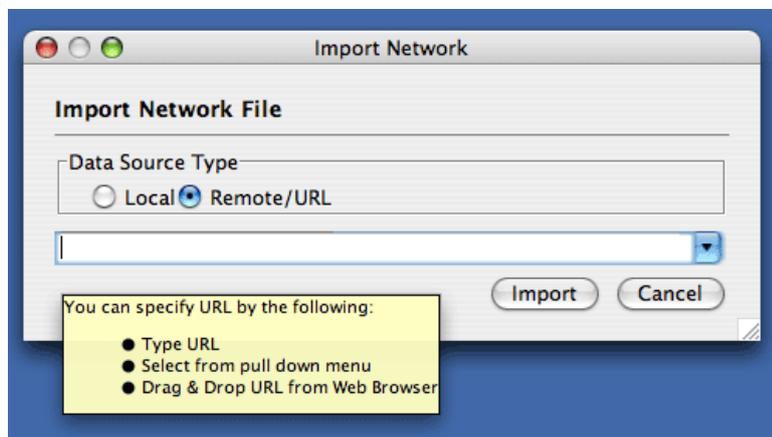
在Import Networks中会显示缺省的“Data Source Type: Local”，意思就是从本地计算机导入网络文件。点击Select按钮选择要加载的网络文件（只有Cytoscape能识别的文件才会显示出来），然后点击Import按钮把网络加载到Cytoscape中。在Cytoscape的sampleData的文件夹中找到一些不同类型的网络文件。

SIF、GML和XGMML格式的网络文件也可以在命令行中用-N选项直接导入。



### 5.1.2 从远程计算机上导入网络（URL导入）

在Import Networks中还可以使用URL导入网络文件。首先把Data Source Type设置成Remote，然后手动填写或是用书签插入合适的URL。点击文本域右边的箭头就能访问收藏的URL。（书签管理器的详细信息参阅Preferences中的Bookmark Manager一节。还可以把浏览器中的URL拖拽到URL文本框中。指定好URL后，点击Import按钮就能加载网络。

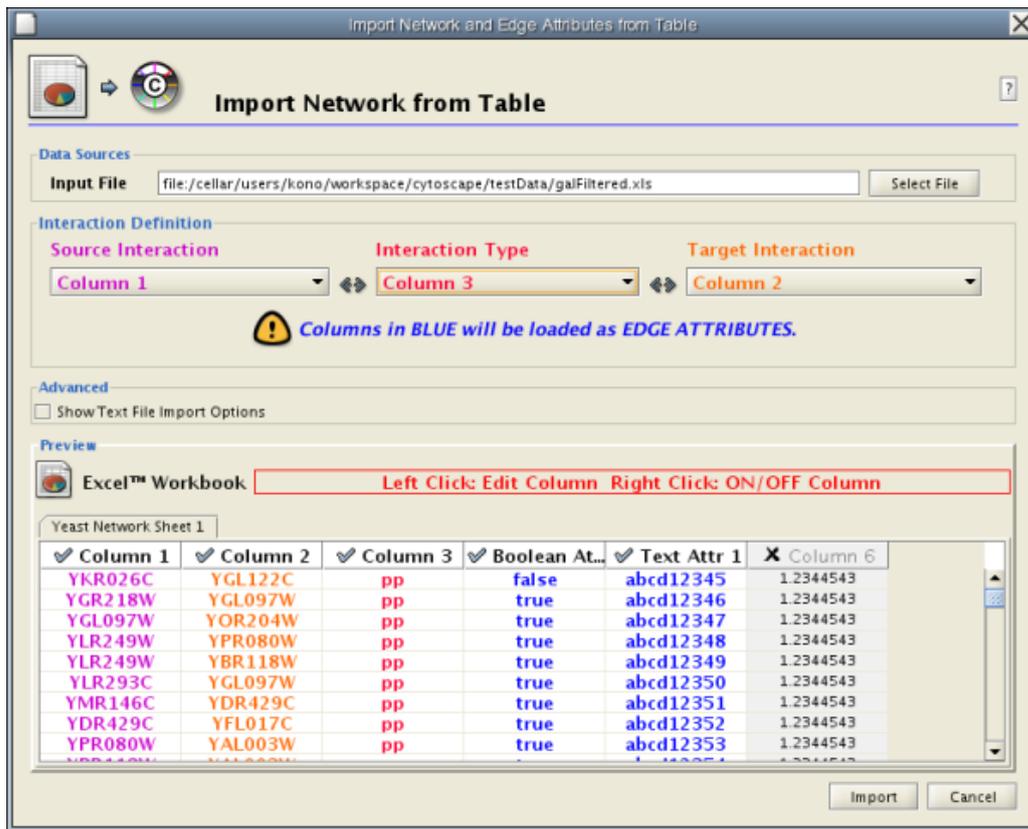


从URL地址导入网络一定要注意一点。由于Cytoscape主要（但不是完全）是根据文件的后缀来判断文件的类型，所以如果URL中的文件的后缀不合适的话，在导入网络时就有可能遇到麻烦。如果Cytoscape没能在URL识别出有意义的文件名和后缀，它就会根据MIME类型来猜测文件的类型。如果所有的导入句柄就无法识别MIME类型，导入就会失败。

在导入网络时有可能遇到的另一个问题就是防火墙，导致无法访问文件。为了解决这个问题，Cytoscape支持使用代理服务器。在 Edit → Preferences → Proxy Server...中就可以设置代理服务器。在Preference有这方面的更多信息。

## 5.2 导入格式灵活的表格文件

从2.4版开始，Cytoscape就能通过Edit → Import → Network from Table (Text/MS Excel)...从文本文件和Excel文件中导入网络。在弹出的窗口中可以设置处理文件的各种选项。还能预览当前设置对文件的处理结果。修改设置时，预览窗口也会自动更新。除了设置文件的处理方法，用户还必须指定Source节点和Target节点，还可以这是边的类型。



### 5.2.1 支持的文件类型

“Import Network from Table”功能支持按特定符号分割的文本文件，和单工作表的微软Excel文件。下面就是一个例子：

source	target	interaction	boolean attribute	string attribute	floating point attribute
YJR022W	YNR053C	pp	TRUE	abcd12371	1.2344543
YER116C	YDL013W	pp	TRUE	abcd12372	1.2344543
YNL307C	YAL038W	pp	FALSE	abcd12373	1.2344543
YNL216W	YCR012W	pd	TRUE	abcd12374	1.2344543
YNL216W	YGR254W	pd	TRUE	abcd12375	1.2344543

网络文件至少要有两列：一列是起点，一列是终点。相互作用的类型是可选的。所以，最简单的网络表格应该是这样的：

```

YJR022W YNR053C
YER116C YDL013W
YNL307C YAL038W
YNL216W YCR012W
YNL216W YGR254W
    
```

在网络文件中，每一行都是一条边以及这条边的属性。所以，网络文件是网络数据和边属性的组合。当然，表中的某些列可能并不是边的属性。遇到这种情况，就可以选择不导入这些列，在预览窗口中点击这一列的表头就是了。例如，在导入下面这种表时，这个功能有很有用：

```

Unique ID A Unique ID B Alternative ID A
Alternative ID B Aliases A Aliases B Interaction
detection methods First author surnames Pubmed
IDs species A species B Interactor types Source
database Interaction ID Interaction labels
Cross-references Associated Files Experiment
files Experiment labels Different techniques
Different Pubmed articles Different sources Weight

7205 5747 TRIP6 PTK2 Q15654 Q05397-1 vv|HPRD
Currently not available
14688263|15892868(Marcotte) Mammalia Homo
sapiens protein|protein HPRD|Marcotte 0 Thyroid
hormone receptor interactor 6-FAK-|PTK2-TRIP6
NA(HPRD)|NA(Marcotte)
HPRD/02859_psimi.xml|other/ORIGINAL_DATA_MARCOTTE.txt
vv(HPRD/02859_psimi.xml)|HPRD(other/ORIGINAL_DATA_MARCOTTE.txt)
17651(ExptRef)|Marcotte 2 2 2 2

4174 7311 MCM5 UBA52 P33992 P62987
neighbouring_reaction Currently not available
15608231(Reactome) Homo sapiens Homo sapiens
protein|protein Reactome 1 P33992-P62988
Reaction:68944<->Reaction:68946(Reactome)|Reaction:68946<->
Reaction:68944(Reactome) other/ORIGINAL_DATA_MARCOTTE.txt
neighbouring_reaction(other/REACTOMEhomo_sapiens.interactions.txt)
Reactome 1 1 1 1

7040 7040 TGFB1 TGFB1 P01137 P01137 nmr:
nuclear magnetic resonance Currently not available
8679613 Homo sapiens Homo sapiens protein|protein
BIND 2 TGFB1-TGFB1- 72085(BIND)
BIND/bind_taxid9606.1.psi.xml nmr: nuclear
magnetic resonance(BIND/bind_taxid9606.1.psi.xml)
NotAvailable 1 1 1 1

```

这份数据是用制表符分割的，含有网络数据（相互作用）、边的属性和节点属性。把这份数据导入到 Cytoscape 中，首先选择 Unique ID A 作为起点，Unique ID B 作为目标，Interactor type 作为相互作用类型。不要导入节点属性（Alternative ID A、species B 等等）。其它的列则作为边的属性导入。

网络导入功能无法导入节点属性，只能导入边属性。关于节点属性的导入，请阅读本手册的 Attributes 一节。

注意：

1. 这里的数据来自 Andrew Garrow、Yeyejide Adeleye 和 Guy Warner (Unilever, Safety and Environmental Assurance Center, 2006年10月12日)的 *A merged human interactome* 数据集。在 <http://www.cytoscape.org> [http://cytoscape.org/cgi-bin/moin.cgi/Data\\_Sets/](http://cytoscape.org/cgi-bin/moin.cgi/Data_Sets/) 上可以下载到真实的数据。

## 5.2.2 基本操作

导入文本表格或 Excel 表格中的网络，步骤如下：

1. 点击 File → Import → Network from Table (Text/MS Excel)...
2. 点击 Select File 按钮，选择一份表格。

- 从表格中选择相互作用的起始点和相互作用类型。如果将相互作用的类型设置为缺省类型，则所有的相互作用的类型都会被设为pp；这个参数可以再Advanced Options中修改（下面会介绍）。
- （可选）如果需要的话，可以定义边属性列。在表格中，除了网络数据，还可以有一些列来表示边的属性。
  - 激活或取消属性列——在预览表格中鼠标左键单击表头，就能激活或去表边的属性。如果表头及该列的内容时蓝色的，则该列将会作为边的属性导入。例如，下面表格中的第1到第3列将会作为网络数据，第4列不会导入，而第5、6两列则会作为边属性导入。

Column 1	Column 2	Column 3	Column 4	attr 1	attr 2
YKR026C	YGL122C	pp	FALSE	abcd12345	1.2344543
YGR218W	YGL097W	pp	TRUE	abcd12346	1.2344543
YGL097W	YOR204W	pp	TRUE	abcd12347	1.2344543
YLR249W	YPR080W	pp	TRUE	abcd12348	1.2344543
YLR249W	YBR118W	pp	TRUE	abcd12349	1.2344543
YLR293C	YGL097W	pp	TRUE	abcd12350	1.2344543
YMR146C	YDR429C	pp	TRUE	abcd12351	1.2344543
YDR429C	YFL017C	pp	TRUE	abcd12352	1.2344543
YPR080W	YAI 003W	nn	TRUE	abcd12353	1.2344543

- 改变属性名称和数据类型——在预览表格的表头上用鼠标右键单击，就能修改属性的名称和数据类型。有关信息，详见下面的“修改属性名称和类型”。

5. 点击 Import按钮。

### 5.2.3 导入不带边的节点列表

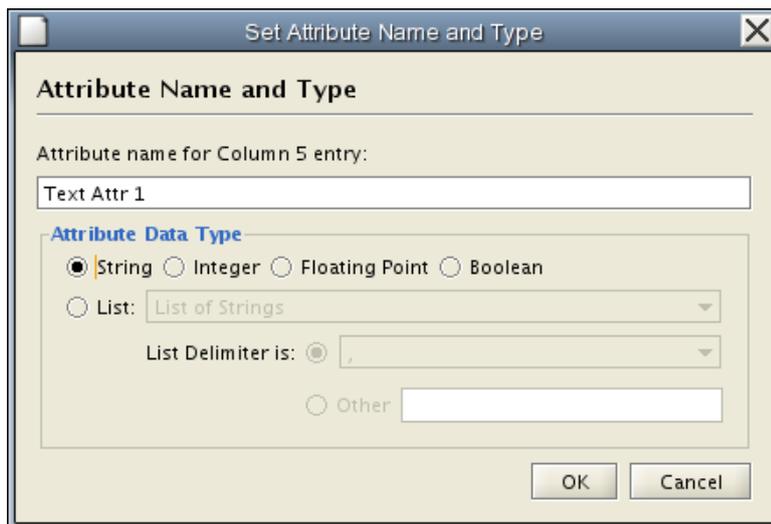
在 Table Import 中可以导入不带边的节点列表。如果只选择源节点列，就会创建一个没有相互作用的网络。在使用一些web服务客户端来扩展节点时，这个功能很有用。详细信息，请参阅“从外部数据库中导入网络”。

### 5.2.4 高级选项

选择 Show Text File Import Option 选项，可以对一些选项进行设置。

- 分隔符 (Delimiter)：可以为文本表格选择多个分隔符。缺省情况下，分隔符是制表符和空格。
- 预览选项 (Preview Options)：选好网络表格文件后，文件中的前一百行数据会显示在预览窗口中。修改这个选项的值，可以在预览窗口中显示更多的内容。如果要显示文件中所有的内容，选上“Show all entries in the file”即可。点击 Reload 按钮可以更新预览窗口中的内容。
- 属性名称 (Attribute Names)
  - 将第一行设为属性名称 (Transfer first line as attribute names)：选上这个选项，所有的边的属性就会根据表格第一行命名。
  - 起始导入行 (Start Import Row)：设置从表格中第几行开始导入。例如，如果要忽略文件中的前三行，就可以将这个选项设为4。
  - 注释行 (Comment Line)：以该字符开头的行将被视为注释，不会导入。这个选项可以用于跳过文件中的注释。
- 网络导入选项 (Network Import Options)：如果相互作用的类型被设为缺省相互作用，所有边的类型就会被设为这个值。

### 5.2.5 修改属性名称和类型



在这里可以修改属性的名称和数据类型。

- 修改属性名称 (Modify Attribute Name) : 输入新的属性名称, 然后点击OK即可。
- 修改属性数据类型 (Modify Attribute Data Type) : 支持以下数据类型:
  - String
  - Boolean (True/False)
  - Integer
  - Floating Point
  - List of (one of) String/Boolean/Integer/Floating Point

Cytoscape 有基本的数据类型探测功能, 能自动地根据该列数据的内容推测数据的类型。不过还是可以自行选择合适的数据类型。对于list, 必须要设置一个全局的分隔符 (例如, 表格中所有的单元格都必须使用相同的分隔符)。

## 5.3 从 Web 服务导入网络

从 2.6.0 开始, Cytoscape 有了一个新功能 **Web Service 客户端管理器 (Web Service Client Manager)**。有了这个功能, 用户可以访问各种类型的数据库。详见从 **外部数据库导入网络和属性**。

## 5.4 编辑新网络

在Cytoscape中还可以新建空网络, 然后手动地添加节点和边。点击 File → New → Network → Empty Network, 然后可以用CytoPanel 1中的Editor手动的添加网络节点和边 (详见Editor一章)。

## Chapter 6

# 所支持的网络文件格式

Cytoscape 可以读取以下格式的网络或路径文件:

- Simple interaction file (SIF or .sif format)
- Graph Markup Language (GML or .gml format)
- XGMML (extensible graph markup and modelling language).
- SBML
- BioPAX
- PSI-MI Level 1 and 2.5
- Delimited text
- Excel Workbook (.xls)

SIF 格式的文件只有节点和相互作用，而其它的格式都可以存储网络布局信息，还可以跟其它的网络软件和数据源交换数据。SIF 文件通常用于在新建网络时导入相互作用，因为用文本编辑器和电子表格软件能很方便的创建这种格式的文件。在导入了相互作用，应用了某种网络布局后，就可以将网络存为GML或XGMML格式，从而能去其它系统交换数据。所有的这些格式（Excel例外）都是文本文件，用普通的文本编辑器就能编辑和查看这些文件。

## 6.1 SIF 格式

这种简单的格式可以很方便地用于从相互作用列表构建网络。利用这种格式，还能很方便的把小网络组合在一起，或是在现有的数据中添加新的相互作用。但这种格式的缺点也是显而易见的，其中不包含布局信息，这使得Cytoscape不得不在每次加载网络时都重新计算网络的布局。

SIF 文件中的每行都由起点、相互作用类型（或边的类型）和一个或若干个重点构成。

```
nodeA <relationship type> nodeB
nodeC <relationship type> nodeA
nodeD <relationship type> nodeE nodeF nodeB
nodeG
...
nodeY <relationship type> nodeZ
```

下面是一个具体的例子:

```
node1 typeA node2
node2 typeB node3 node4 node5
node0
```

第一行是两个节点，node1和node2，以及两点间的typeA型的相互关系。第二行加入了三个新节点，node3、node4 和 node5，这一行中的node2跟第一行中的node2是同一个节点。第二行还设定了三条起点相同且类型也相同的相互作用。第三行说明了如何引入孤立的节点。 This form is not needed for nodes that do have relationships, since the specification of the relationship implicitly identifies the nodes as well.

重复的条目被忽略。两个点之间的多条相互作用必须是不同的类型。例如，下面的node1和node2之间就有两条边，但类型分别是xx和yy。

```
node1 xx node2
node1 xx node2
node1 yy node2
```

节点的自相互作用是允许的:

```
node1 xx node1
```

Cytoscape中每个点和每条边都有标识符，通常在节点和边的属性数据中都有。节点的名称必须是独一无二的，名称相同的节点会被视为同一个节点。节点的名称缺省情况下就是文件中的名称，除非用Visual Mapper映射到了其他字符串。详见“视觉风格”一章。边的名称则是由边的起点和终点，加上相互作用的类型构成的，例如: sourceName (edgeType) targetName。

<relationship type>标签可以是任何字符串。完整的单词或是单词的组合都可以用来定义相互作用的类型，例如: geneFusion、cogInference、pullsDown、activates、degrades、inactivates、inhibits、phosphorylates、upRegulates等等。

在系统生物学领域，常见的相互作用类型有:

1. pp ..... protein → protein interaction
2. pd ..... protein → DNA (例如, 转录因子跟调控基因的上游结合。)

还有一些相对少见的相互作用类型:

1. pr ..... protein → reaction
2. rc ..... reaction → compound
3. cr ..... compound → reaction
4. gl ..... genetic lethal relationship
5. pm ..... protein-metabolite interaction
6. mp ..... metabolite-protein interaction

### 分隔符

在简单的相互作用文件格式中, 空白 (空格或制表符) 用来分割名称。但在有些情况, 节点名称或是边的类型中会含有空格。Cytoscape对分隔符的处理原则是这样的: 如果文件中有制表符, 那么就用制表符分割不同的字段, 而空格则被视为字段中的内容。如果文件中没有制表符, 那么所有的空格就都是分隔符 (也就是中字段中不会有空格)。

如果在导入网路后, 发现网络中没有边, 而且点的名称也有问题, 那很有可能是因为文件中存在制表符, 使得Cytoscape在导入网络时判断错误。另一方面, 如果网络中节点的名称只是正确名称的一部分, 那就很有可能是错误地将空格作为了分隔符, 实际上应该用制表符。

用简单的相互作用的形式存放网络的文件的后缀是.sif, Cytoscape能识别出文件夹中所含有的这类文件。

## 6.2 GML 格式

跟SIF格式不同, GML是一种富图格式语言 (rich graph format language), 很多网络可视化软件包都支持这种格式。在<http://www.infosun.fmi.uni-passau.de/Graphlet/GML/>上可以找到该格式的具体说明。

通常都不必直接修改GML文件的内容。当SIF格式的网络导入并实施布局后, 就可以以GML的形式保存和加载网络。GML文件中的视觉风格在加载GML文件后, 会保存为名为Filename.style的视觉风格。

## 6.3 XGMML 格式

XGMML是GML的XML扩展版, 它是基于GML定义的。除了网络数据, XGMML还包含节点、边和网络的属性。XGMML文件的规范在<http://www.cs.rpi.edu/~puninj/XGMML/>上。

由于XGMML继承了XML文件的灵活性, 所以XGMML比GML更受欢迎。如果不确定应该用哪种格式, 那就选择XGMML吧。

## 6.4 系统生物学标记语言 (Systems Biology Markup Language)

系统生物学标记语言 (Systems Biology Markup language) 是一种用于描述生化网络的XML格式。SBML文件格式规范: <http://sbml.org/documents/>

## 6.5 BioPAX (Biological Pathways eXchange)格式

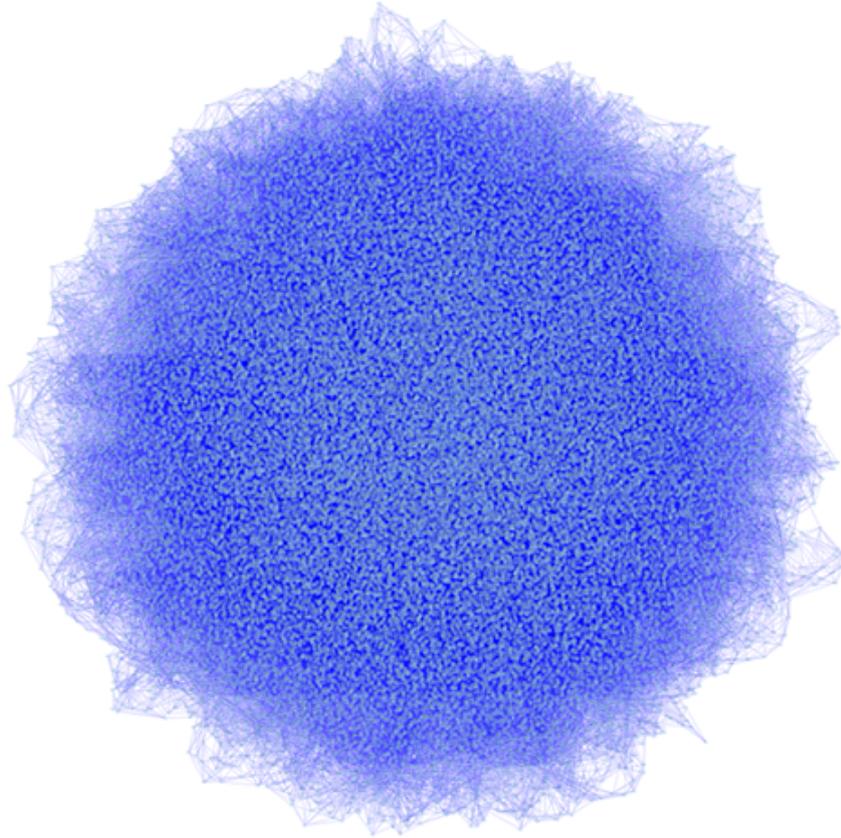
BioPAX是一种用于交换生物路径数据的OWL (Web Ontology Language) 文档。该格式的完整说明文档: <http://www.biopax.org/index.html>

## 6.6 PSI-MI 格式

PSI-MI格式一种用于描述蛋白质相互作用及有关数据的XML格式。PSI-MI XML的格式规范: <http://psidev.sourceforge.net/mi/xml/doc/user/>。

## 6.7 纯文本表格和Excel表格

Cytoscape 为微软的Excel文件 (.xls) 和纯文本表格提供了原生支持。可以用这些表格存储网络数据和边的属性。在导入文件的过程中，用户可以指定哪一列是源节点，哪一列是终点，以及相互作用的类型、边属性等等。其它的一些网络分析工具，例如igraph (<http://cneurocv.s.rmki.kfki.hu/igraph/>)，也可以将网络输出成简单的文本文件。Cytoscape 可以读取这些文本文件，并构建网络。具体信息请阅读5.2 节。



在Cytoscape中显示根据 igraph 的Watts-Strogatz小世界模型生成的网络（含有5万个节点，25万条边）

利用Table Import功能，可以读取其它软件生成的网络。

## 6.8 Cytoscape 中节点的命名

通常情况下，节点表示基因，节点间的边表示相互作用（或者是其他的生物关系）。为了紧凑，基因也可以表示成响应的蛋白质。节点还可以用来表示化合物和生化反应等各种东西，并不局限于基因。

如果想要将网络中的基因或蛋白跟GO注释或基因表达数据整合在一起，那各个数据文件中的基因名称必须完全一致。我们强烈建议用户用ORF系统名称或是标准的获取编号（accession number）来命名基因和蛋白质。而常用名称（common name）由于更适合显示在屏幕上，所以可以将其存放在注释目录或节点的属性文件中。Cytoscape的annotation/目录中有酵母的ORF到常用名称的对应文件。将来会逐步地支持其它物种。

为什么要建议使用标准的基因名称？所有 Cytoscape 能识别的外部数据格式都有专门的命名方式。例如，蛋白质相互作用网络会列出蛋白质的名称，相应的属性和表达数据也使用同样的命名方法。

但如果要把不用来源的数据结合在一起，问题就出现了：对于同一个东西，不同的数据源的命名可能是不一样的。例如，基因就可以有多个不同的名称，包括正式的“染色体位置”标识符，研究人员讨论时常用的若干个名称。此外，每个数据库可能还会有自己的一套给基因编号的方法（例如Swiss-Prot的蛋白质编号）。如果一个数据源使用的是正式名称，而另一个数据使用的是常用名称，那Cytoscape就必须要知道哪两个名称指的是同一个基因。

有两种方法来解决这个问题，一个较简单，另一个稍微复杂一些。简单的方法就是假设所有的数据源使用同样的命名方法。如果是这样，Cytoscape就能很方便的将来自不同数据源的数据结合在一起。

跟人工整合不同来源的数据一样，为了处理不同命名方式的数据，Cytoscape需要同义词信息（见“注释”一章）。同义词表提供了某个物种的每个对象的权威名称，以及相应的其它可以识别的名称。要注意，同义词表自身的数据就是按照这个“权威”名称命名的。例如，在酵母中，通常用ORF名称作为这个“权威”名称。

如果能提供这样信息，Cytoscape在缺省情况下会将所有的名称都转换成相应的权威名称。而不能识别的名称则保持不变。这样Cytoscape就能将不同来源的数据结合在一起，即使它们的名称不同也没关系—只要能在同义词表中找到就行。

除了上面的数据，还需告诉Cytoscape各个对象所属的物种，因为数据服务器需要物种信息才能找到相应的名称。通过Cytoscape的命令行选项-P (`cytoscape.sh -P "defaultSpeciesName=Saccharomyces cerevisiae"`)，或是编辑网络的属性 (`Edit → Preferences → Properties...`) 都可以设定网络的物种。

通过-P选项 (`cytoscape.sh -P "canonicalizeName=false"`) 或是编辑属性 (`Edit → Preferences → Properties...`) 以关闭名称的自动映射。在当前的版本中，名称映射对基因表达数据无效。基因表达数据应该使用跟其它数据源相同的命名方式，伙食使用同义词表中的权威名称。

## Chapter 7

# 节点和边的属性

## Chapter 8

# 加载基因表达（属性矩阵）数据

## Chapter 9

# 从外部数据库导入网络和属性

## Chapter 10

# 浏览和布局

## Chapter 11

# 可视化风格

## Chapter 12

# 节点和边的查找与过滤

## Chapter 13

# 编辑网络

## Chapter 14

# 插件和插件管理器

## Chapter 15

# CytoPanel

## Chapter 16

# 渲染引擎

## Chapter 17

### 注释

## Chapter 18

# 链接

感谢

## 附录A：老式的注释服务器格式

## 附录B: GNU Lesser General Public License

## 附录C：增加Cytoscape的内存