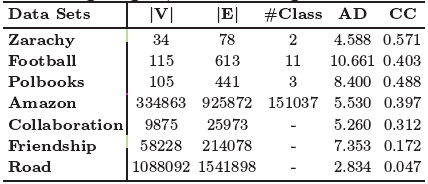
# 吸引子数据集

表1：现实世界数据集的统计，其中AD：平均程度; CC：聚类系数



**数据来源：**

* UCI网络数据存储库（https://networkdata.ics.uci.edu/index.php）
* 斯坦福大型网络公司公开的性能数据集合（http://snap.stanford.edu/data/）

现实世界网络

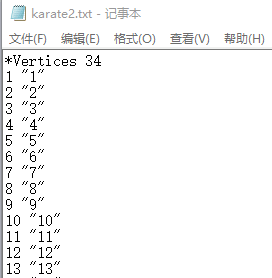
## [Zachary’s Karate Club](https://networkdata.ics.uci.edu/data.php?id=105)

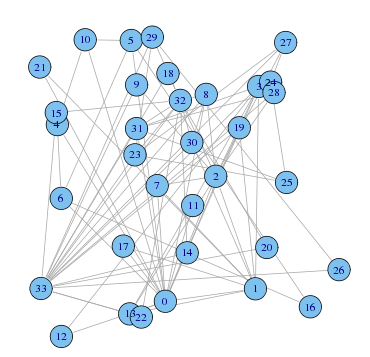
### **1.1 数据集说明**

1970年美国大学空手道俱乐部34名成员之间的友谊社交网络。

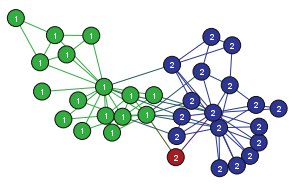
由扎克里对扎伊尔空手道俱乐部的观察而得名的网络反映了这些会员之间的友谊关系。特别地，网络可以分为两个社区，这反映了管理员和教练之间的不一致。具体来说，使用吸引子算法成功找到了两个社区，除了一个成员被视为噪声（node'10'）。还有一点很有意思的是，这个成员位于两个社区之间，分别与两个社区的中心节点连接。在现实情况下，确定其社区属性也是不利的。实际上，更有可能将这个节点分配给两个社区，这是我们在本研究中不会讨论的重叠聚类。

### IMG_256**1.2 数据集内容截图**





### **1.3 运行结果**



空手道俱乐部网络吸引力。节点颜色表示不同的检测到的社区。

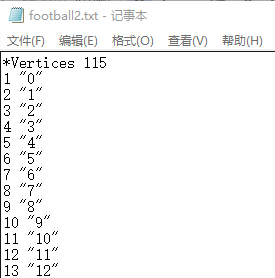
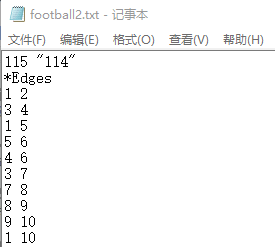
## [American College Football](https://networkdata.ics.uci.edu/data.php?id=5)

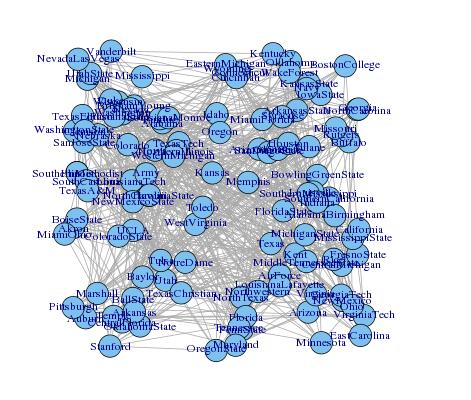
### **2.1 数据集说明**

2000年秋季，美国各大学院之间的美式足球赛网络。

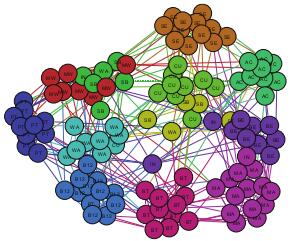
来自美国足球赛的网络是在常规赛2000赛季的第一赛季，其中图中的115个顶点代表队，613个边代表了他们连接的两队之间的常规赛。 团队分为12个会，每个会约8-12个团队，因此，真正的社区结构已经是已知的。有趣的是，吸引力自动发现12个社区的质量（NMI = 0.923，ARI = 0.897，Purity = 93.0％）。 大多数团队都被正确分配到相应的社区。

**2.2 数据集内容截图**





### **2.3 运行结果**



美式足球网络吸引力。

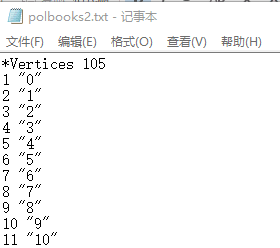
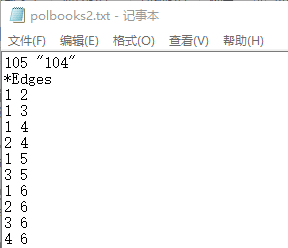
## 三、[Books about US Politics](https://networkdata.ics.uci.edu/data.php?id=8)

### **3.1 数据集说明**

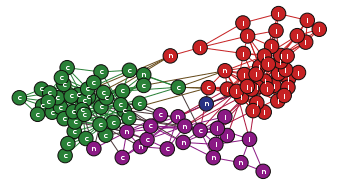
由网络书商Amazon.com出售的关于美国近期政治的书籍网络。 图书之间的边代表同一买家频繁地购买图书。该网络由V. Krebs编辑，未发布，但可以在Krebs的网站上找到。

在2004年总统选举之前出版的关于美国政治的政治书籍的这个网络由105个节点和441个边组成。节点代表网络书商Amazon.com销售的图书。边代表同一买家频繁地共同购书。根据纽曼阅读亚马逊书籍的描述和评论，每本书都标有“l”，“n”或“c”，以表明他们是“自由主义”，“中立”还是“保守”。吸引力允许将这些书分类到这些类别，其中两个群体分别很好地代表相应的自由和保守的书籍。

### **3.2 数据集内容截图**



### **3.3 运行结果**



政治书籍网络的吸引力。

## [Amazon network](http://snap.stanford.edu/data/amazon0601.html)

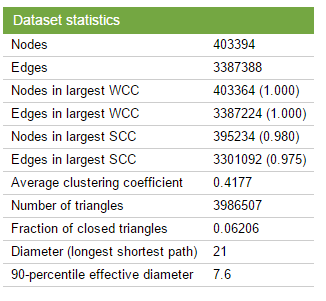
### **4.1 数据集说明**

该网络由334,863个节点和925,872个边组成，每个节点代表亚马逊网站上的一个产品。 每个产品根据亚马逊提供的类别分类到相应的社区，在[20]中调查了最高质量的前5000名社区。

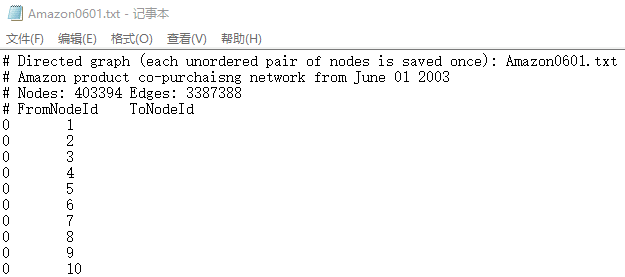
然而并没有找到所说明的数据集，此处选用的是如下数据集：

网络是通过爬网亚马逊网站收集的。它是基于购买此商品的客户也购买了亚马逊网站的功能。如果产品i经常与产品j共同购买，则该图包含从i到j的有向边。

数据于2003年6月1日收集。



### **4.2 数据集内容截图**



没有类信息的网络

## [Hepth collaboration network](http://snap.stanford.edu/data/ca-HepTh.html)

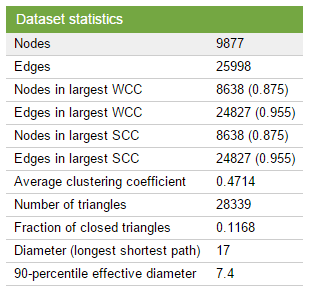
### **5.1 数据集说明**

该网络是一个9,875名作者，致力于高能物理学理论的协作网络。吸引力识别1384个社区，其结果是模块化= 0.579，ncut = 1179。

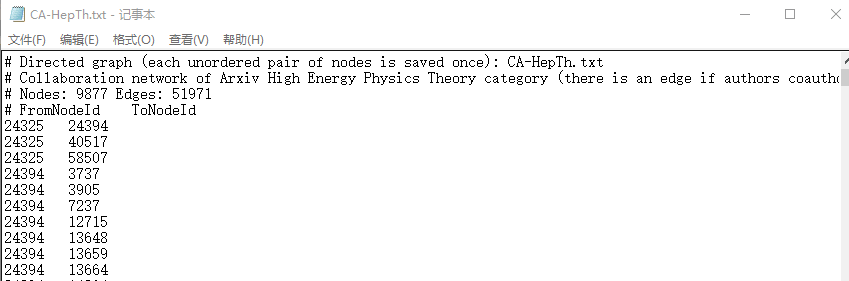
所下载的数据集说明如下：

Arxiv HEP-TH（高能物理 - 理论）协作网络来自电子打印arXiv，涵盖了提交给高能物理 - 理论类的作者论文之间的科学合作。 如果作者与作者j共同撰写了一篇论文，该图形包含从i到j的无向边。如果这篇论文是由k个作者共同撰写的，那么它会在k个节点上产生完全连接的（子）图。

数据涵盖1993年1月至2003年4月期间的文件（124个月）。 它从在ArXiv开始的几个月内开始，因此基本上代表了其HEP-TH部分的完整历史。



### **5.2数据集内容截图**

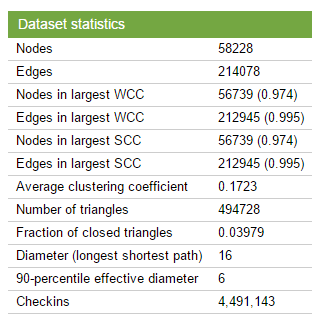


## [Brightkite friendship network](http://snap.stanford.edu/data/loc-brightkite.html)

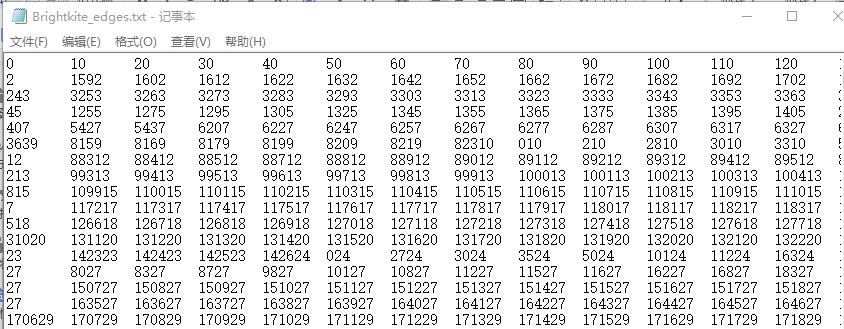
### **6.1 数据集说明**

该图是一个基于位置的友谊网络，包括58,228个节点和214,078个无向边。 吸引力发现8045个社区。

Brightkite曾经是一个基于地理位置的社交网络服务提供者,用户分享他们的位置被签入。友谊网络使用公共API收集,由58228个节点和214078个边。这个网络最初是定向的，但是我们已经建立了一个无定向的网络，在这两方面都有友谊。我们也收集了4491143签到这些用户在此期间2008年4月——2010年10月。



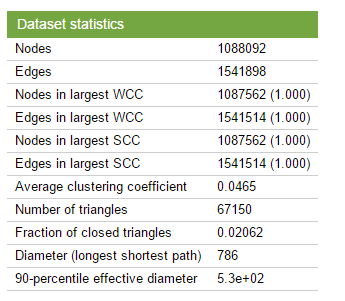
### **6.2 数据集内容截图**



## 七、[Pennsylvania road network](http://snap.stanford.edu/data/roadNet-PA.html)

### **7.1 数据集说明**

该网络反映了宾夕法尼亚州的路面结构，其中节点代表交叉路口或端点，边界代表连接这些交叉点或端点的道路。这里，我们为吸引力设置参数λ= 0.6，对于MCL设置i = 1.4，因为两个算法的默认值由于网络的稀疏性而不能导致良好的结果。吸引力最终识别59,919个群集，其中moduleity = 0.856，ncut = 25055。MCL实现了可比性能，比Metis算法更好。 Louvain和Infomap只发现少数同等社区（分别为492和208）。



### **7.2 数据集内容截图**

