《Higher-order organization of complex networks》

论文解读

21821152 朱策

## 简介

网络是一种用于理解以及衡量物理学，生物学，神经学以及工程，社会学等系统的十分基础的工具，对于低阶层（lower-order）的网络我们在科学界已经被广泛的应用，而对于复杂网络，这个领域还是几乎未知，没有被人探索过的。因此本文主要介绍了一种使用高阶层（high-order）的连接方式对网络进行聚类的通用方法。这种框架展示了一种包括信息在神经网络中的反向传播以及网络运输的中心结构在内的一种高阶层组织方式。

## 序列与网络

对于每一个序列（见图1A）同一种网络都有一种不同的高阶层聚类方式（图1B），这意味着不同的聚类方式主要取决于所选择的序列。

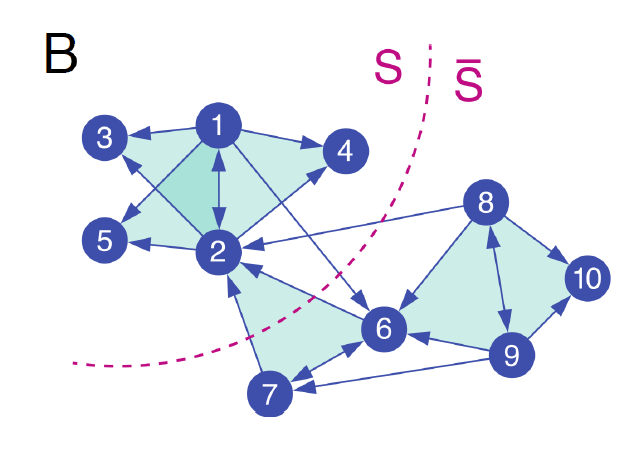
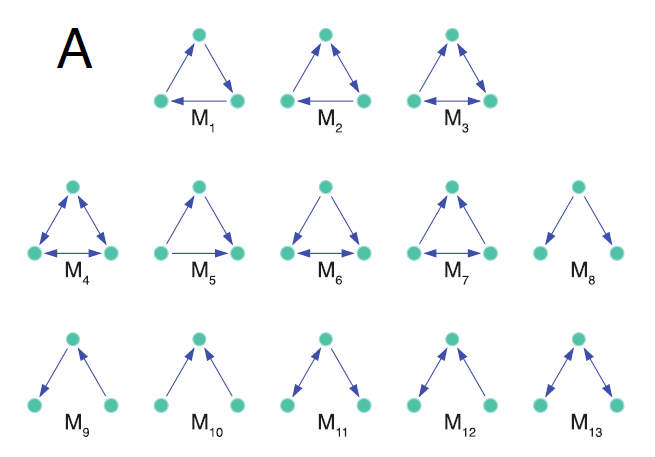


图1 序列与网络

## 算法介绍

概念性的，给定一个序列 ，我们以两个目标来寻找一个节点的集合 。第一，在中的节点应该参与M的大量的实例，第二，集合S应该避免去切断M的实例（比如只有部分M的节点在集合S中）。那么给定一个序列网络M，高阶层的聚类框架的目标就是寻找一个由个节点集合S定义的聚类使得下述比值（式1）最小化：

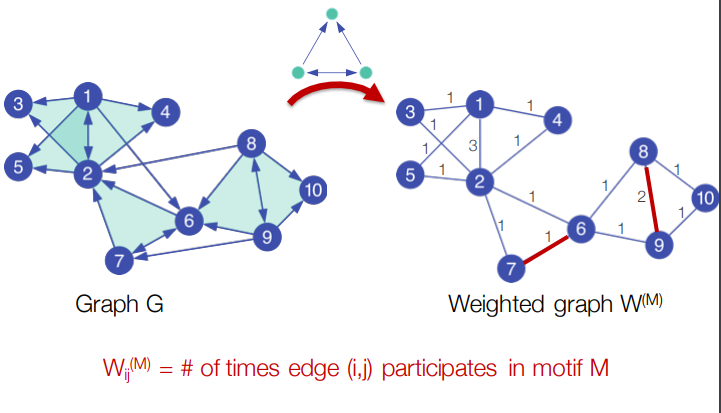
式1

此处 代表S集合中的剩余节点， 代表至少在S中有一个节点，在 中有一个节点的序列网络M 的数量， 代表M的实例中属于S的节点数目。

因此对于该方法有效确定节点的集合S的步骤如下：

步骤1，给定一个感兴趣的网络与序列M，给出针对节点i与j的该序列网络M的邻接矩阵，其计算方法为（式2），其主要样例如图2：

= 包含节点 与 节点的边参与到序列M的实例数量



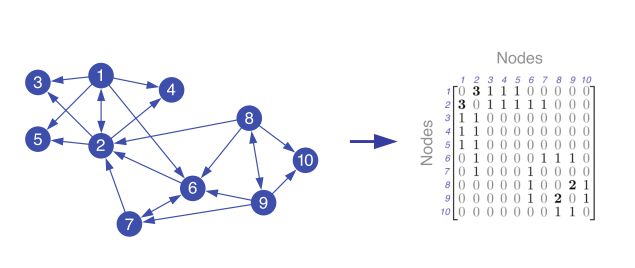


图2邻接矩阵计算示意图

步骤2，基于矩阵的拉普拉斯算子（Laplacian）计算矩阵M的费德勒（Fiedler vector） ,

步骤3，如图3，根据其费德勒（Fiedler vector） 的大小对于所有的节点进行排序，对f}并且计算每一个的导率（conductance），最后取其导率最低的集合 即为最好的高阶层集合。

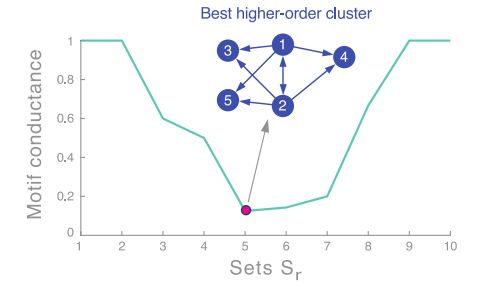


图3 步骤3示意图

## 算法优势

从结果的准确度来讲，该算法主要寻找到了一系列点的集合S，并且满足式3：

式3

其中 代表该算法找到的最佳集合S的序列导率，代表最佳集合的序列导率，换句话说，使用本算法计算得到的集合S被证明十分接近最佳集合。

从实现方式来讲，本算法的实现方法简单，并且易于扩展，对于带方向的，带记号的或者带权重的网络都可以进行拓展。

## 算法源码实例运行结果

对于celegans 前向网络【1】，其节点为神经原，其边即为神经原突触。

对具体的数据，其包括如下3个数据信息（图4）

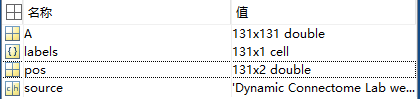


图4 celegans前向网络数据

其中其具体意义如下：

A：包含网络的边信息，即记录从节点到节点的单向边。

labels：包含节点的名称

pos：包含网络节点的所在位置

我们使用Bi-Fan 结构【1】（如图5）以及两种简单的的结构对该单向网络寻找其最优聚类，其结果如图6，并与论文中展示的结果（图7）进行对比：

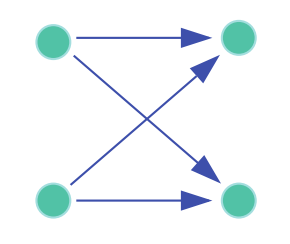


图5 Bi-Fan 结构

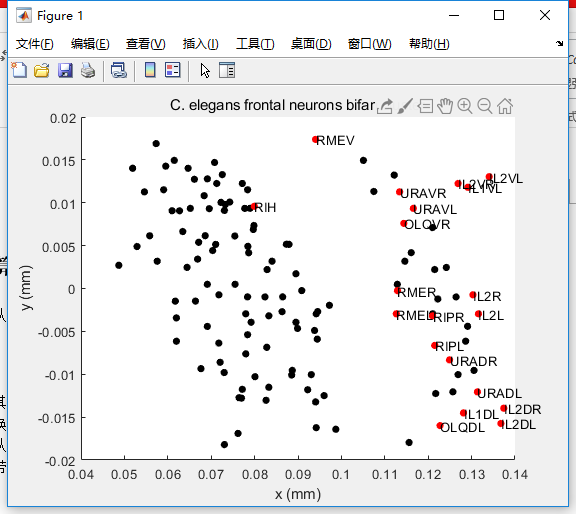


图6 使用Bi-Fan 结构对elegans 前向网络聚类结果

对比论文中的可视化结果（图7），可见结果与论文中的吻合

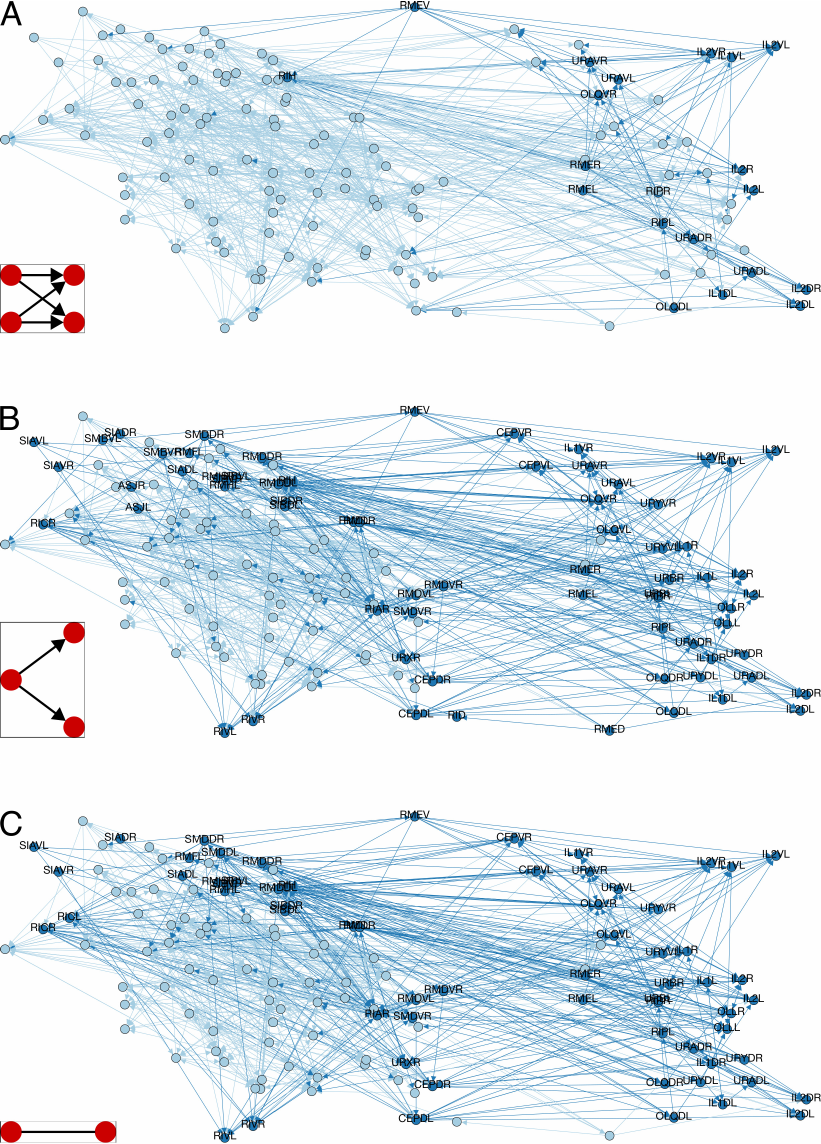


图7 三种结构对celegans 前向网络进行聚类的可视化

并且对于产生的结果进行进一步的可视化（图7），我们可以获得更多的信息：

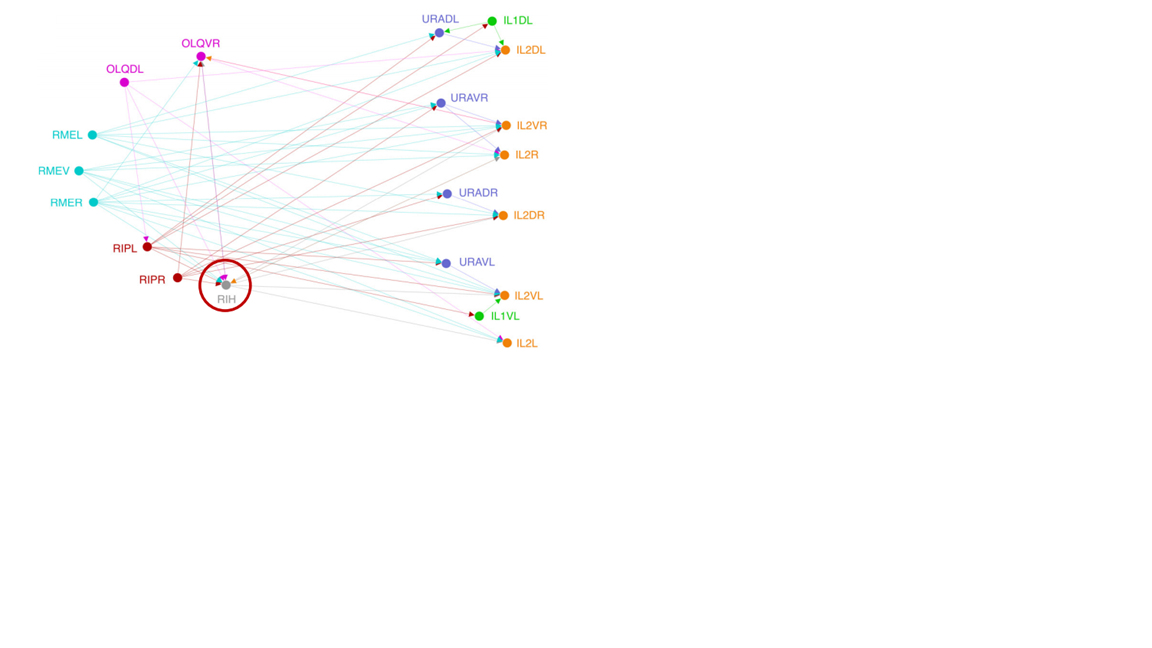


图7 使用Bi-Fan 结构对elegans 前向网络聚类结果

我们可以看出RME开头的神经元起到了输入的作用，IL2开头的神经元作为信息传递的终点，URA开头的神经元经常作为中间神经元，而RIH神经元是最多作为信息传递终点的神经元。

## 参考文献

【1】Ibrahim Z M, Tawfik A Y, Ngom A. Qualitative motif detection in gene regulatory networks[C]//2009 IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2009: 124-129.