A Local Algorithm for Structure-Preserving Graph Cut

现如今，大规模的图数据可以从社交网络、共同作者网络、蛋白质交换网络、公共交通网络等应用中产生。许多在图像挖掘方面的工作以一阶马尔科夫链为基础模型，研究重点为图中的节点和边。但是这些工作在应用到更重要的领域中的高阶图的结构时却不行，比如在银行的用户身份鉴定信息网络中，星型的网络结构对应于一组组合身份；在财务传输网络中，环形结构常常意味着洗钱行为的发生。在这篇文章中，我们注重于挖掘***特定用户***的高阶网络结构以及希望找到一种能够在不用破坏很多原有边的基础上将之从原图中分离出来的子图，称为***structure-rich***的子图。

在寻找这样的子图过程中最重要的是避免过高的计算成本。为了解决这个问题，受局部图聚类算法族（用来高效定位低导通性的分割图且不用探索整个图）的启发，我们提出通过规范化这个关键思想来对高阶网络结构建模。In particular，我们从高阶导通性的一般定义开始，定义了高阶扩散核，这是以用户指定的高阶网络结构引发的高阶随机游走为基础。然后我们提出一个全新的High-Order Structure-Preserving LOcal Cut(HOS-PLOC)算法，该算法运行时间为图中的边的数量的polylogarithmic time。算法从一个种子节点开始，迭代地探索它的邻域，直到找到一个具有小的高阶导通性的子图。此外，我们从效率和效率两方面分析了它的表现。在合成图和实际图上的实验结果证明了我们提出的HOSPLOC算法的有效性和有效性。

A Temporally Heterogeneous Survival Framework with Application to Social Behavior Dynamics

社会行为动力学是理解和建模复杂的社会动态现象（如信息传播，意见形成和社会动员）的核心基石之一。尽管近年来提出了多种社会行为动力学模型，但社会行为动力学的基本成分和最小模型仍然没有得到解答。在这里，我们通过探索大规模的社交沟通数据集发现人类互动行为动力学在响应时间维度和自然时间维度上展现了丰富的复杂性。为了应对这一挑战，我们开发了一个temporal Heterogeneous Survival framework，其中响应时间维度和自然时间维度的规律可以有机地结合在一起。将我们的模型应用在两个在线社交沟通数据上，结果表明我们的模型能够成功完成对社交沟通数据集中交互模式的再现，并且所提出的方法可以显着超越其他最先进的baseline。同时，学习的参数和发现的统计规律可能导致多种潜在的应用。

Ego-Splitting Framework: from Non-Overlapping to

Overlapping Clusters

我们提出了ego-splitting，一种新的用于复杂网络中的聚类检测框架，通过利用ego-net来对重叠的聚类进行解耦。Ego-splitting是一个高度可扩展和灵活的框架，具有可证明的理论保证，将复杂的重叠聚类问题简化为一个更简单，更易于处理的非重叠（分区）问题。我们可以将社区检测扩展为具有数百亿边缘的图，并且比以前的基于ego-net的其他解决方案表现更好。

更准确来说，我们的框架按照以下两个步骤运行：一个局部的ego-net分析阶段，和一个全局的图像分区阶段。在局部的阶段，我们首先使用分区算法对节点的ego-net进行分区，然后使用计算的群集将每个节点拆分成代表其社区中节点的实例化的角色节点。最后，在全局的步骤中，我们对新创建的图进行分割，以获得原始图的重叠聚类

FORA: Simple and Effective Approximate Single-Source

Personalized PageRank

给定一个图G，一个源节点s以及一个目标节点t，节点t关于节点s的personalized PageRank(PRR)就是从s开始结束于t的随机游走的概率。一个single-source PPR（SSPPR）查询枚举G中的所有节点，并返回关于源节点s中具有最高PPR值的前k个节点。SSPPR在网络检索核社交网络中具有重要的应用，比如Twitter的“关注谁”的推荐服务。但是，SSPPR的计算代价十分昂贵，应用中无法满足实时性。目前的解决方法要么使用没有保证质量的启发式算法，或者使用十分昂贵的高计算能力的数据中心。

因此，我们提出了FORA，一种简单而有效的基于索引的对SSPPR过程的近似解决方法。FORA的基本思想是以一种简单而又平凡的方式将两种现有的方法Forward Forward（快速但不保证质量）和Monte Carlo Random Walk（准确但缓慢）结合起来，从而产生一种快速和准确的算法。除此之外，FORA包含了一个简单而有效的索引方案以及一个具有高度可裁剪性质的top-k选择模块。大量的实验表明，FORA比其主要竞争对手的效率高出数个数量级。值得注意的是，在一个数十亿的Twitter数据集上，FORA使用一台商用服务器，在5秒钟内的除了top-500的近似SSPPR查询。

Improved Degree Bounds and Full Spectrum Power Laws in Preferential Attachment Networks

考虑网络演进的一个随机优先模型，它允许节点和边缘到达。从任意非空图开始，在每个时间步，有两个可能的事件：以概率p> 0，新节点到达并在新节点和现有节点之间增加一个新的边，以概率为1-p在两个现有节点之间添加一条边。在这两种情况下，所涉及的现有节点都是根据preferential attachment随机选择的，即与其度成正比的概率。已知产生幂律分布的网络，也就是度为k的节点所占的比例与成正比，在这里在(2,3]区间内。

将在k时刻度为k的节点数记为,我们显著提高了一些long-standing的结果。具体来说，我们证明的值以偏差为)集中在它的均值附近且与k独立。我们还将期望与加性误差结合在一起且与t独立。这些新的结合让我们能够在与以往相比更大的k值上进行对进行估计。反过来，也使我们能够估计其他的一些重要的数值。

最后，我们引入了一个新的广义模型，它通过允许节点和边到达的概率为时变的来扩展，以及形成新的分量。我们表明，扩展模型可以产生任何指数在范围内(1, ∞）的幂律网络。此外，G（p）中建立的集合边界也适用于。

Inferring the Strength of Social Ties:A Community-Driven Approach

在线社交网络正在不断壮大，变得更加密集。一个特定用户的社会关系具有很高的可变性：从亲密的朋友，亲戚到熟人，到陌生人。推断社会联系的强度是模拟网络中的用户交互并理解他们的行为的重要组成部分。而且，这个问题在计算社会科学，病毒式营销和人们的推荐等方面都有应用。

在本文中，我们研究推断给定网络中社会关系的力量的问题。我们的工作是受到最近的一个方法的启发，它利用strong triadic closure（stc）原理，一个植根于社会心理学的假设。为了指导我们的推理过程，除了网络结构之外，我们还将一个紧凑社区的集合作为输入，这些是我们期望通过强关系连接的节点集合。

我们考虑了两个相关问题的形式化，这些形式化反映了我们设定的假设：违反STC的行为是少量的以及输入社区的强连接的连接性。我们发现这两个问题的公式化是NP难的，而且第一个问题的公式化很难通过近似得到，而针对第二个问题我们开发了一个近似保证的算法。我们在现实世界数据集上通过分别与针对违反STC的优化方法和社区连通性的baseline进行比较来验证提出的方法。

Unsupervised Feature Selection in Signed Social Networks

社交媒体服务的快速增长，以前所未有的速度带来大量的高维社交媒体数据。特征选择在通过查找相关特征的子集、准备高维数据方面功能非常强大。现有的绝大多数用于社交媒体数据的特征选择算法关注于链接实例（如友谊和用户关系）之间的正相关互动。但是，在现实世界的许多社交网络中，实例也可能是负相关的。最近的工作表明，负面的联系在推进许多学习任务方面比正面的联系更有价值。在本文中，我们研究了在Signed Social Network中的无监督特征选择问题，并提出了一个新的框架SignedFS。具体而言，我们提供了一个原则性的方式来模拟用户潜在表征学习的正面和负面的联系。然后，当标签信息不可用时，我们将用户的潜在表示嵌入到特征选择中。此外，我们重新审视了Signed Social Network中的同质性和平衡理论的原理，并将signed graph的正则化纳入到特征选择框架中，以捕捉Signed Social Network中用户之间的一阶和二阶相似度。在两个现实世界中的Signed Social Network的实验证明了我们提出的框架的有效性。当然也需要进一步的实验来了解SignedFS中的不同组件的影响。

Network Inference via the Time-Varying Graphical Lasso

许多重要的问题可以被建模为一个相互联系的实体系统，其中每个实体都记录时间相关的观测值或测量值。为了发现趋势，检测异常，并解释这些数据的时间动态，理解不同实体之间的关系以及这些关系随着时间的推移是非常重要的。在本文中，我们引入了time-varying graphical lasso（TVGL），一种从原始时间序列数据推断时变网络的方法。我们通过估计一个稀疏时变逆协方差矩阵来解决这个问题，这个矩阵反映了动态网络之间的相互依赖关系。由于动态网络推理是一个计算量很大的任务，因此我们推导了一种基于Alternating Direction Method of Multipliers (ADMM)（ADMM）的可扩展消息传递算法，以高效的方式解决这个问题。我们还讨论了几个扩展，包括更新模型的流式算法和实时合并新的观察。最后，我们对真实和合成数据集上的TVGL算法进行评估，获得可解释的结果，并在准确性和可扩展性方面优于最新的baseline。

Relay-Linking Models for Prominence and Obsolescence in Evolving Networks

不断发展的社交网络中的节点获取链接（朋友，引用）的速度显示出复杂的时间动态。使用Preferential attachment 和 link copying 模型虽然可以支持便捷优美的分析，但是只是捕捉到了“rich get richer”这一效应，而没有aging和decline的过程。最近的提出的aging model大多都很复杂，而且参数很多，大部分模型估计每个节点都有1-3个参数。这些参数是本征的：它们解释了同一个节点过去事件的衰落，也没有用网络来解释连接注意力可能转移的地方。我们认为传统的关联动力学描述不足以判断模型的忠实性。我们提出了一个演化图的新时间草图，并介绍了网络时间动态的几个新特征。然后，我们提出了一个frugal aging models的模型族，这些模型没有单个节点的参数，只有两个全局参数。我们的模型是基于a surprising inversion or undoing of triangle completion, where an old node relays a citation to a younger follower in its immediate vicinity。尽管参数很少，但是新的模型族显示出了更好的与真实数据拟合性。在结束之前，我们分析了各个研究团体的时间特征，进一步了解他们的比较动态。为了便于重复研究，我们将尽快将所有的代码和处理过的数据集提供给公众。

The Co-Evolution Model for Social Network Evolving and Opinion Migration

几乎所有现实世界的社交网络都是动态的，随着时间的推移而不断变化，新的链接可能形成，旧的链接可能会消失，这些行为很大程度上取决于社会行动者（即网络中的节点）的同质性。与此同时，社会行为主体的（潜在）性质也随着时间的推移而发生变化，部分原因是从社交网络中受到影响，这种影响也会反过来影响网络结构。社会网络演化和节点属性迁移通常被视为两个正交的问题，并已经分别进行了研究。在本文中，我们提出了一个共同演化模型，通过将两个现象一起建模来完成闭环，其中包含两个主要部分：（1）当节点属性已知时的网络生成模型; （2）当社会网络结构已知时的属性迁移模型。模拟表明，我们的模型有几个很好的性质：（1）它可以模拟一些常见的现象，如意见收敛和基于社区的意见分歧; （2）通过社会影响范围，意见领袖和噪音水平等因素来控制演化。最后，我们的模型的有效性可以通过在国会对共同赞助的立法议案进行预测来体现，它的表现胜过了几个最先进的baseline。

When is a Network a Network? Multi-Order Graphical Model Selection in Pathways and Temporal Networks

我们介绍了一个用于对在网络中观察到的不同长度的捕获pathways的顺序数据的建模框架。这样的数据是重要的，例如，在Web中学习点击流，交通系统中的旅行模式，社交网络中的信息级联，生物路径或时间标记的社交交互。虽然通常将图分析和网络分析应用于这些数据，但最近的研究表明，与时间相关的特性可能会使这些方法的结果无效。这提出了一个基本的问题：什么时候对顺序数据进行网络抽象是合理的？为解决这个问题，我们提出了一个框架，将多个更高阶的马尔可夫链组合成一个多层图形模型，同时捕获多个长度尺度上的路径中的时间相关性。我们还开发了一个模型选择技术来推断这样模型的最佳层数，并显示它胜过baseline马尔可夫秩序检测技术。