# 1月

# 2月

## Distributionally Robust Removal of Malicious Nodes from Networks

Sixie Yu1;Yevgeniy Vorobeychik

https://arxiv.org/pdf/1901.11463.pdf

分布式的安全移除网络中的有害节点

原问题NP难，使用基于二元半定过程的近似方法(solving it approximately based on duality combined with Semidefinite Programming relaxation)。

MINT算法，移除良性节点的直接损失，剪断良性子图的间接损失，维持良性节点和恶性节点连接的间接损失

## Discovering Archetypes to Interpret Evolution of Individual Behavior

https://arxiv.org/pdf/1902.05567.pdf

使用高斯隐马尔科夫模型（G-HMM）簇来表示用户在社交网络中演化模式的原型

## Finding Nearest Neighbors in graphs locally

<https://arxiv.org/pdf/1902.05638.pdf>

不需要遍历整个图就可以找到最近邻

引文较少，只有十篇

传统方法：随机行走，可以获取单步之外的信息，捕捉到整个网络的结构。分为两类：到达节点的概率（Personalized-pagerank）和到达节点所需的步数。

随机行走不是很适合最近邻问题，最近邻问题只需要知道局部信息，不需要知道整张图的结构

提出使用局部近似方法，理论基础：节点不应该被远处节点影响

简单的设定：图中的节点允许与相邻节点交互

最短路径不能表征路径的多样性。——两个有多条路径相连的节点被很好的连接

度数越高的节点，邻接点成为最近邻的机会越小

度量距离的三个特征：路径数、对称性、度数

节点间的电阻距离：使用边权重作为电阻

方法类似马尔科夫链（设置适当的参数即变为马尔科夫链。相当于是马尔科夫链的扩展？）

## Network Embedding for Cross-network Node Classification

<https://arxiv.org/ftp/arxiv/papers/1901/1901.07264.pdf>

不及node2vec和deepwalk经典

网络嵌入是学习低维节点向量表示同时保留原始网络结构及属性的有效方法。 已有的网络嵌入方法多用于单个网络，不能学习跨网络的泛华和可比较的特征表示。本文解决跨网络的节点分类问题，使用源网络中有充分标注的信息来帮助对目标网络中未标注的节点进行分类。跨网络深度网络嵌入（CDNE），将源网络和目标网络的节点嵌入到一个统一的低维隐空间

该模型整合了深度网络嵌入和领域适应来学习标签区分和网络不变节点向量表征。

网络结构、节点属性和节点标签共同用来学习相同网络和不同网络中相似节点的相似隐向量表示

采用SAE\_s和SAE\_t两种层叠自编码器将源网络和目标网络的节点嵌入到统一的低维潜伏空间中。

一方面，使用网络结构捕获同一网络中的节点邻近性。首先使用SAE\_s和SAE\_t通过重构对应的源网络和目标网络的网络结构相似形矩阵学习低维节点向量表示。此外，成对约束已被合并到SAE\_s和SAE\_t中，以在潜在空间中的每个网络内嵌入更强连接的节点。另一方面，可以利用带属性的和标记过的信息来捕获跨网络的节点之间的邻近性。 SAE\_s首先被用于对源网络节点进行标签区分特征向量表示，通过将属于同一类的节点映射的较近，而属于不同的类的节点距离较远。然后，SAE\_t用于通过最小化跨网络边际和类条件分布差异来学习目标网络节点的网络不变特征向量表示。结果，目标网络节点将具有与相同标签相关联的源网络节点类似的潜在向量表示。请注意，在SAE\_s中，已经在嵌入空间中单独映射了不同类别的源网络节点。因此，通过匹配SAE\_t中的跨网络类条件分布，与目标网络或网络中的相同标签相关联的节点将具有类似的潜在向量表示，而与目标网络内的完全不同标签和跨网络相关联的节点会有相当不同的潜在向量表示。这些属性产生标签判别和网络不变的节点向量表示，这显着有利于跨网络节点分类问题。

领域适应的目的是迁移之前从源领域学习到的知识来帮助解决目标领域里的相同任务，广泛应用于NLP和CV领域，在社交网络中的应用较少

SAE stacked auto-encoder

PPMI度量节点间的结构相近性。

## \* DeepWalk: Online Learning of Social Representations

<https://arxiv.org/pdf/1403.6652.pdf>

(deepwalk)

经典方法。用于网络分类，异常检测。属于深度学习（无监督特征学习）

第一次将深度学习技术引入到网络分析中。

DeepWalk是KDD 2014的一篇文章，彼时word2vec在文本上的成功应用掀起来一波向量化的浪潮，word2vec是根据词的共现关系，将词映射到低维向量，并保留了语料中丰富的信息。DeepWalk算法思路其实很简单，对图从一个节点开始使用random walk来生成类似文本的序列数据，然后将节点id作为一个个「词」使用skip gram训练得到「词向量」。

DeepWalk是一种学习网络中顶点潜在 表示的方法，这些潜在表示将社会关系在连续向量空间中编码。DeepWalk将词序列的语言建模和无监督特征学习（或者说深度学习）方面的进步推广到了图。

DeepWalk将行走视为句子的等价，使用从截断的随机行走中获取的局部信息来学习潜在表示。通过对短的随机行走流建模学习图中顶点的社会表示（social representations）。社会表示是顶点的潜在特征，捕获了邻居相似性和社区成员。潜在表示以相对少的维度将社会关系在连续向量空间中加密。DeepWalk推广了神经语言模型，来处理由随机生成的步的集合构成的特殊的语言

DeepWalk的输入为一张图，并以潜在表示为输出。

**主要贡献：**

将深度学习作为工具引入到图的分析中，，并建立了适合统计模型的鲁棒表示。DeepWalk学习短随机行走中呈现的结构规律。

**问题定义**

**学习社会表示**

**方法**

## Large scale information network embedding

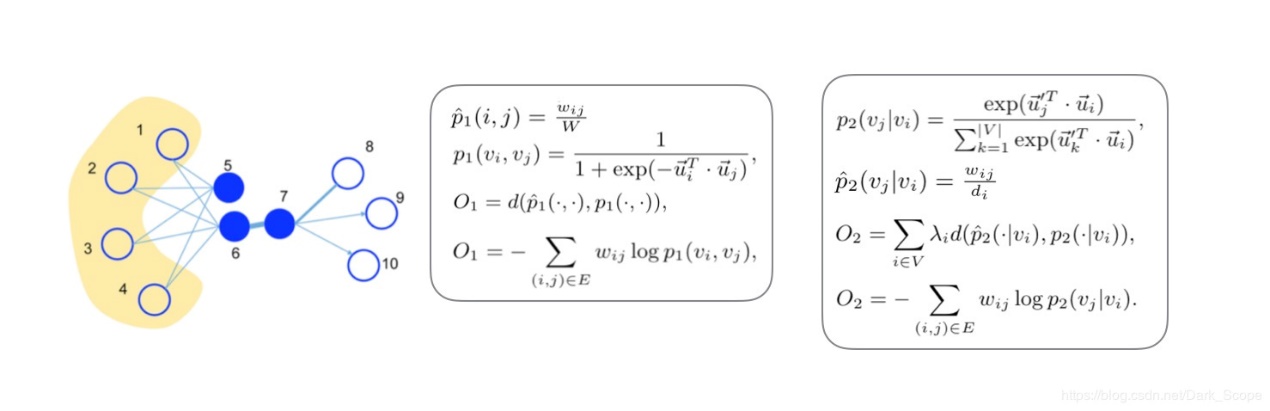
LINE

LINE分析了1st order proximity和2nd order proximity，其中一度相似性就是两个点直接相连，且边权重越大说明两个点越相似，如下图中的6和7；而二度相似性则是两个点之间共享了很多邻居，则它们的相似性就很高，如下图的5和6。

文章中非常简单的方式构造了一个目标函数，能同时保留二者的信息。以一度相似性为例，节点i和j相连的经验概率就是和归一化后的权重，即pˆ1(i,j)=wij/W

p^​1​(i,j)=wij​/W，而通过向量计算这个概率值是p1(i,j)=11+exp(−uTiuj)p1​(i,j)=1+exp(−uiT​uj​)1​，目标函数就是让这两个分布距离最小，选择KL散度作为距离衡量函数就得到了最后的损失函数O1O1​。

其中还有个优化的trick，edge-sampling algorithm：因为边的weight差异很大，直接用SGD效果不好，所以有个edge的采样，按照边的weight采样，然后每条边当做binary的算。



## 小结

**图（graph）**是一种数据格式，它可以用于表示社交网络、通信网络、蛋白分子网络等，图中的节点表示网络中的个体，连边表示个体之间的连接关系。许多机器学习任务例如社团发现、链路预测等都需要用到图结构数据，因此图卷积神经网络的出现为这些问题的解决提供了新的思路。

图上的神经网络，最早由Joan Bruna于2014年在论文Spectral Networks and Locally Connected Networks on Graph中提出。图卷积网络（简称GCN），由Thomas Kpif于2017年在论文Semi-supervised classification with graph convolutional networks中提出。它为图（graph）结构数据的处理提供了一个崭新的思路，将深度学习中常用于图像的卷积神经网络应用到图数据上。

图上的卷积网络从卷积方式上可以分为两种：1.谱（spectral）卷积，2.空间域卷积。Thomas Kpif的这篇论文就是属于谱卷积，即将卷积网络的滤波器与图信号同时搬移到傅里叶域以后进行处理。而Mathias Niepert的论文Learning Convolutional Neural Networks for Graphs中的图卷积网络就属于空间域的卷积，即让图中的节点在空间域中相连、达成层级结构，进而进行卷积。

非欧几里德域的数据（图就属于非欧数据）相关的几个概念：图嵌入（graph embedding）、网络嵌入（network embedding）、网络表示学习（network representation learning），这三个概念从原理上来说其实表达的是同一件事，核心思想就是“通过深度学习技术将图中的节点（或边）映射为向量空间中的点，进而可以对向量空间中的点进行聚类、分类等处理”。

## Learning Convolutional Neural Networks for Graphs

<https://arxiv.org/pdf/1605.05273.pdf>

## Community Detection in Attributed Graphs: An Embedding Approach

<https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17142/15705>

AAAI-18

复旦 香港浸会大学

社区：节点紧密连接，社区内节点同质。允许社区重叠。

提出基于结构嵌入的模型，同时考虑社区结构和节点属性。

基于对密度连接结构的观测提出新的社区结构嵌入方法，来使用潜在的社区成员对社区固有结构编码。然后利用节点属性信息和社区结构嵌入，将属性社区检测用公式表示为一个非负矩阵分解最优化问题，并设计更新规则保证找到收敛解。

**相关工作**

结构图聚类

文章关注生成模型。该算法认为边缘生成的概率依赖于它们的社区成员关系。

另一种结构图聚类方法流派利用非负矩阵分解技术，认为邻接矩阵可以分解为社区成员关系矩阵的线性组合。

属性图聚类

典型方法是提出一个区分模型，通过选择权重向量把节点属性合并到网络结构中

网络嵌入

深度行走：将随机行走看作句子的等效值来学习节点的潜在表示

**问题定义：**社区划分的两个层面，结构上，社区内的点紧密连接，社区间的点链接稀疏；属性上，社区内的点具有同质性，不同社区的点性质不同。允许/不允许社区重叠。

**非负矩阵分解：**

表示第i个节点属于第j个社区的可能性

将社区检测问题转化为最小化重构代价。SNMF算法。缺陷：只通过直接分解邻接矩阵利用了观测到的网络拓扑结构。对相连的节点对，无论它们是否属于同一个社区，邻接值都为1.

**社区结构嵌入方法**

**社区结构嵌入：**提出新的社区结构嵌入方法来量化节点的结构相近性，根据节点潜在的社区成员相近性。然后，根据相似性度量，采用负抽样skip-gram（SGNS， Mikolov等人。2013）来探索网络结构并描绘潜在的社区结构。最后，得到编码了固有社区结构的社区结构嵌入矩阵。

**社区成员相似：**因为社区内的节点紧密相连，每对相连的节点都有落入相同社区的趋势。相似性度量函数 Sigmod函数

**社区结构嵌入矩阵：**本文的社区结构嵌入试图最大化相连节点的Sigmod函数，同时最小化随机选择的节点对的Sigmod函数。现实世界中，大型网络非常系数，随机选择的节点对相连的概率非常低，基于SGNS，设计了节点对的负抽样目标函数——j从经验主义的均匀分布中抽样。最后，对任意节点对得到最佳的社区结构嵌入。

社区结构嵌入矩阵可以轻松的扩展到有向/无向/带权/无权网络

**基于优化的结构嵌入：**使用社区结构嵌入矩阵取代SNMF中的邻接矩阵。此时不考虑属性

**CDE模型** 使用结构信息和节点属性来处理社区检测问题

**社区属性：**不同的社区倾向于具有不同的属性。定义社区属性矩阵。表示第i个社区在节点第r维（不同社区有不同的节点属性序列）属性上的偏好。将社区成员和社区属性的发现用公式表示为非负矩阵分解问题，通过分解节点属性矩阵同时优化社区成员矩阵和社区属性矩阵——移除（对社区划分）不重要的节点属性

**CDE的统一目标函数：**在属性图G中，CDE的目标是找到社区内节点紧密相连且具有同类属性值的k个社区。通过包含社区结构嵌入和节点属性的目标函数，定义统一目标函数。（之前的两个目标函数加权。）

**通过U矩阵识别社区：** U 社区成员矩阵。

对于无重复社区，寻找最大的,对于有重复社区，寻找大于阈值的

**CDE的迭代更新规则：**统一目标函数非凸，故定义迭代更新规则（基于Majorization-Minimization framework）。对每一次迭代，固定C更新U，然后固定U更新C。

需要证明收敛性。

## Network Embedding as Matrix Factorization: UnifyingDeepWalk, LINE, PTE, and node2vec

<https://arxiv.org/pdf/1710.02971.pdf>

skip-gram模型（word2vec）促进了网络嵌入的研究。

DeepWalk：经验化的生成网络的标准拉普拉斯矩阵的低秩转化

LINE：顶点内容的大小设为1时，为DeepWalk的特例。

PTE：LINE的扩展。可以看做多网络拉普拉斯的联合分解。

node2vec：分解与二阶随机游走的平稳分布和转移概率张量相关的矩阵。

这些使用负抽样的模型都可以统一到矩阵分解框架中。证明了基于skip-gram的图网络嵌入算法与图拉普拉斯理论的联系.

学习网络的潜表示（如网络嵌入），已经得到广泛研究以自动发现和映射网络结构属性到潜空间。

## Representation Learning for Scale-free Networks

<https://arxiv.org/pdf/1711.10755.pdf>

2018 AAAI

网络嵌入（即网络表示学习）的目的是在保留网络结构和内在属性的同时学习网络中顶点的低维表示。现有的网络嵌入方法集中在保留微观结构，如节点的一阶和二阶近似，而宏观的无标度属性很大程度上被忽视了。无标度属性描述这样的事实：节点的度服从重尾分布（少数顶点具有很高的度）。

首先分析了欧几里得空间中嵌入和重构无标度网络的困难，将问题转化为球包装问题。然后，为无标度属性保留网络嵌入算法提出“度罚金”原则。引入该原则的两种实现：利用谱技术和skip-gram模型

网络表示：最直接的方式是使用网络的邻接矩阵，但是数据稀疏；最近兴起网络表示学习。

网络嵌入：

网络重建：基于节点在潜空间中的距离重建网络边。

节点度的密度函数。节点度

其中，为指数参数，C为归一化项。应用中，上述幂律形式只适用于du大于一个特定的最小值的顶点。

拉普拉斯特征映射（Laplacian Engenmap LE）

**重建无标度网络**

。选择特定的，对于嵌入向量为的顶点，我们看作所有点落入以为中心以为半径的封闭的球中，

General idea

DP-Spectral

DP-Walker

## Correspondence Analysis Using Neural Networks

<https://arxiv.org/pdf/1902.07828.pdf>

对应分析

对应分析（CA）是一种用于可视化和解释依赖关系的多变量统计工具。 CA已经在流行病学和社会科学等领域找到了应用。但是，用于执行CA的当前方法不能扩展到大的高维数据集。通过使用称为主要惯性分量的信息理论工具重新解释CA中的目标，我们证明了执行CA等效于解决两个随机变量的有限方差函数空间上的功能优化问题。我们证明了这个优化问题，反过来，可以通过神经网络有效地近似。由此产生的配方，称为对应分析神经网络（CA-NN），使CA能够以前所未有的规模进行。

之前不适用于大数据，现在适用了

主要惯性分量 principal inertia components ， PICs

PIC提供了X和Y的统计依赖性的细粒度分解，完全确定了主要函数（PF），即从Y（反之亦然）可靠地估计X的正交集的正则方差函数。

我们证明了CA产生的低维投影恰好是PIC理论中的主要功能。 反过来，主要函数可以通过求解X和Y的有限方差函数空间上的二次优化问题来确定。 乍一看，解决任意变量的这种优化是不可行的。 然而，通过限制我们的多层神经网络可表示的搜索功能，我们演示了如何有效地近似离散和连续（可能是高维）随机变量的主函数。 总之，首先根据基于PIC的优化程序制定CA，然后使用神经网络逼近该程序，我们能够以前所未有的规模执行CA.

主要贡献：

PIC和主要函数如何用于对应分析

引入对应分析神经网络CA-NN来估计PIC和主要函数，是的CA在离散和连续（高维）数据上可扩展

使用人造数据证明CA—NN找到的主要函数与理论预测值匹配

## Semi-Supervised Classification with Graph Convolutional Networks

基于图卷积网络的半监督分类

<https://arxiv.org/pdf/1609.02907.pdf>

## \*Graph Attention Networks

<https://mila.quebec/wp-content/uploads/2018/07/d1ac95b60310f43bb5a0b8024522fbe08fb2a482.pdf>

ICLR 2018

使用其中的注意力机制。

图注意力网络

我们提出图注意力网络（GATs），一种对图形结构数据进行操作的新型神经网络结构，利用蒙板自注意层来解决基于图形卷积或相似的现有方法的缺点。 通过堆叠节点能够参与其邻域特征的层，我们（隐式地）指定邻域中不同节点的不同权重，而不需要任何类型的计算密集矩阵操作（例如反转）也不需要预先知道图结构。 通过这种方式，我们同时解决了基于谱的神经网络的几个关键挑战，并使我们的模型适用于诱导和转换问题。

方法是通过参与邻接点并使用自注意力机制计算图中每个节点的隐表示，

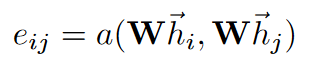
GAT Architecture

**Graph Attentional Layer 图注意力层**

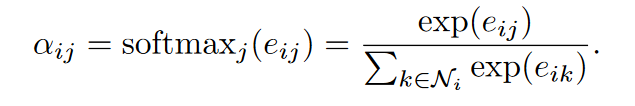
该层的输入为节点特征集合，输出是可能带有不同基数的新的节点特征集合

N为节点数，F为节点特征数。

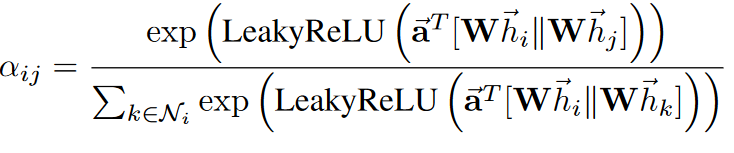
为了将输入特征转换为高层次特征，需要至少进行一次可学习的线性转换。为此，对每个节点进行以权重矩阵为参数的共享线性转换，然后在节点上执行自注意力——共享的注意力机制a计算注意力系数（表示节点j的特征对节点i的重要性）。



在最一般的形式中，允许每一个节点参与每一个其他节点权重系数的计算，将会丢弃结构信息。我们通过执行有遮蔽注意力图结构注入注意力机制——只对i的邻居节点集合中的部分点计算注意力系数（实验中计算i的一阶邻居。）。为了方便比较不同节点的注意力系数，对所有j使用softmax进行归一化。

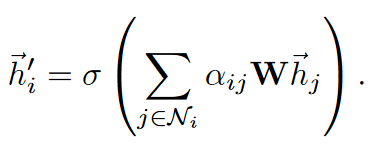


在实验中，注意力机制为以权重向量为参数的单层前馈神经网络，并使用LeakyReLU非线性化。

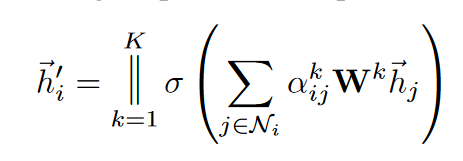


||表示串联。

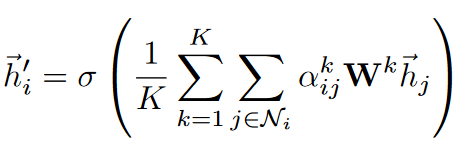
获取到归一化注意力系数之后，就计算对应特征的线性组合。每个节点的输出特征：



为了稳定自注意力机制的学习过程，使用多头注意力扩展注意力机制，即使用多个不同的注意力机制执行上述计算，然后将结果进行串联。



表示第k个注意力机制，表示对应的线性转移权重矩阵。但是这种情况下，最终的返回值中，每个节点将包含个特征。因此，只在网络的最后一层执行多头注意力multi-head attention，不再进行串联，而是使用求平均值。



**与相关工作的比较**

计算的高效性

不同于GCN,允许给对应同一邻居的不同节点赋予不同重要性，

注意力机制是图中的所有边共享的，因此不需要提前获取图的结构或者所有节点的特征。

大图上的推导表示学习（Hamilton）对每个节点抽样固定大小的邻居，以保持计算开销连续。

GAT可以形式化为MoNet的特例。

改进方向：2.2描述的问题，处理更大规模的批尺寸

## \*LINE: Large-scale Information Network Embedding

## Structural Deep Network Embedding

## Graph convolutional networks

GCN 半监督学习

CNN是图像上的。

GCN目标是在简单的格子之外泛化卷积。利用节点特征/属性

问题：对于子图，如何得到标准的节点序列

**为什么研究GCN？**

1.CNN无法处理Non Euclidean Structure的数据，学术上的表达是传统的离散卷积在Non Euclidean Structure的数据上无法保持平移不变性。也就是说在拓扑图中每个顶点的相邻顶点数目都可能不同，那么也就无法用一个同样的尺寸的卷积核来进行卷积运算。

2.CNN无法处理Non Euclidean Structure的数据，又希望在拓扑图上有效的提取空间特征来进行学习，所以GCN成为研究重点。

3.数据不是拓扑结构的网络时，也会用到GCN。广义上来讲任何数据在赋范空间内都可以建立拓扑关系，谱聚类就是这样的思想。

## Information diffusion backbones in temporal networks

信息传播骨干网络

基于传染病（SI）模型

SIR的话，看历史激活数量累积？

时间维的获取，通过划分时间窗口。

时变网络的表示，使用三维邻接矩阵，第三个维度为时间维，不同时间两个点可能连接，也可能不连接。

## Identifying influential spreaders in complex networks

2011

SIR模型

种子节点影响的人口存在重复。

## Theories for influencer identification in complex networks

2018

基于中心性选择种子节点

Collective influence

**社交网络中的信息传播**

## Conditions for Viral Influence Spreading through Multiplex Correlated Social Networks

2014年

## Event Detection in Twitter Stream usingWeighted Dynamic Heartbeat Graph Approach

<https://arxiv.org/pdf/1902.08522.pdf>

## 26日

### Efficient Path Prediction for Semi-Supervised and Weakly Supervised Hierarchical Text Classification

<https://arxiv.org/pdf/1902.09347.pdf>

分层文本分类有很多现实应用，而标注大量文本代价高昂，可以使用半监督学习或若监督学习来减少分类开销。提出开销敏感的学习算法来利用结构化信息，并更加充分的利用未标注和弱标注信息。

通过类的路径对文本分类

实现的是基于概率框架的分层文本分类，将来目标discrimative（区分）学习模型

### Wasserstein-Wasserstein Auto-Encoders

<https://arxiv.org/pdf/1902.09323.pdf>

深度生成模型

### Modularity as a Means for Complexity Management in Neural Networks Learning

<https://arxiv.org/pdf/1902.09240.pdf>

AAAI 2019 有点意思

训练带有很多参数或者复杂架构神经网络会使优化过程复杂化。为此，我们提出模块化方法来进行神经网络的设计，其中神经网络被分解为一个控制模块和服务函数模块，实现简单的操作。

本文将模块化的先验集成到神经网络的设计过程中。 我们通过在可以基于一组基本操作来解决问题域的假设下工作，对模块化进行了初步的简化方法。 我们的目标是促进在NN领域内构建复杂但可管理的系统，同时使各种模块实现能够共存。

主要贡献：

我们提出了一种用于复杂NN设计和训练的通用模块化架构的初始方法。

•我们证明了设计时考虑模块化的NN可以在更短的时间内进行训练，也是整体模型的竞争选择。

•我们提供了将我们的方法转移到其他问题的技巧和指导。

### Harmonizing Maximum Likelihood with GANs for Multimodal Conditional Generation

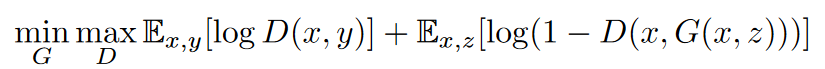
ICLR 2019

<https://arxiv.org/pdf/1902.09225.pdf>

图-图翻译的进展，很大程度依赖于GAN。通常GAN损失和重建损失进行优化。缺点：输出缺少多样化。为了实现训练稳定和多峰输出生成，提出使用新的损失moment reconstruction losses(瞬时损失)替代重建损失的训练方案。

证明了GAN损失和重建损失的明显不匹配，因此，模型不能同时在两个损失上达到最优。

定义为在输入x的条件下最小化真实的和生成的数据分布的距离测度



条件GAN中最常见的重建损失为和，



这两种损失分别起源于参数在拉普拉斯分布和高斯分布下的最大释然估计。

### Learning Segmentation Masks with the Independence Prior

AAAI 2019戴松民

<https://www.researchgate.net/publication/328900087_Learning_Segmentation_Masks_with_the_Independence_Prior/comments>

使用独立先验学习分割遮蔽（mask）？

基于GAN的独立先验。生成器生成具有多个特定类别实例提供者，布局模块和组合模块的图像。首先，每个提供者独立地输出具有软掩码的特定类别实例图像。然后布局模块校正提供的实例的姿势。最后，合成模块将这些实例组合成最终图像。针对遮蔽区域的对抗性损失和惩罚进行训练，每个提供者学习一个尽可能小但仍足以涵盖完整类别特定实例的遮蔽。弱监督语义分割方法广泛使用分组线索来模拟图像部分之间的关​​联，这些图像部分是人工设计的或者用昂贵的分割标签学习或仅仅在局部对上建模。与它们不同，我们的方法自动建模任何部分之间的依赖关系并学习实例分割。我们在两种情况下应用我们的框架：（1）使用盒级注释对类别特定图像进行前景分割。 （2）仅用一个同质对象簇（HOC）图像的无监督学习实例外观和掩模。我们在两个任务中都得到了吸引人的结果，这表明独立性先验对于实例分割是有用的，并且可以无人监督地仅用一个图像来学习实例掩码。

**独立先验**

将图像分层分解为三个级别的表示：部分（像素，特征，补丁等视觉基元），实例（单个对象）和图像。 我们对具有分层潜在因素的图像使用相应的描述：像素级别的外观（apperance，出现还是外观？）（外观和表面掩模），实例级别的对象类别和姿势以及图像级别的场景。 在不失一般性的情况下，我们假设自然场景图像是根据场景特定布局及其可见顺序合成N个实例的结果。 具有M <N个实例的图像可以被视为具有N-M个不可见实例的特殊情况。

***有点像自然语言处理的生成模型（主题模型，pLSA、LDA）。主要工作就是对图像的层次划分和独立性假设。***

我们可以总结以下三个关于这些因素之间依赖关系的先验：

•图像部分和实例类别之间的依赖关系：对象部分的外观取决于其对象类别。

•图像部分之间的依赖性：对象间部分的外观在给定对象类别时彼此独立，但是对象内部部分仍然彼此相关。直观地，对象内部部分与对象间部分之间存在更高的亲和力。

•实例构成之间的依赖关系：特定场景的对象通过几何共现关联彼此相关，例如眼镜和人脸，鱼和水，汽车和道路等的共存。

这些属性导致以下因素主要影响合成图像的逼真度：

•外观：图像部件的外观应遵循与其类别对应的统计关联，以便我们无法从中找到任何工件。

•分割蒙版：图像合成实例的蒙版不应破坏其曲面的完整性，以便在雕刻曲面与具有独立外观的新曲面组合时，不会违反图像部分之间的依赖关系。

•几何关系：根据实例之间的几何共现关联，实例应放置适当的姿势。

**框架：**

通过使用GAN生成特定场景的逼真合成图像来学习对象外观和分割蒙版。我们需要避免图像层与彼此通信并相互补偿，以便在最终的合成图像中有效地显示每个实例的伪像。并且鉴别器不应该通过找出由布局引起的伪像而容易地获胜，这阻止了鉴别器发现由实例掩码引起的更多伪像。我们的模型通过合成多个类别c1的实例来生成特定于场景的图像。 C2; :::; cN具有合适的布局。为确保对象外观彼此独立，用于合成的实例图像独立于特定于类别的实例提供程序进行采样。为了将对象放置在合适的几何关系中，几何翘曲用于例如图像以根据场景特定的布局先验校正出现在最终合成图像中的姿势。这些先验可以基于人类观察来概括，或者可以从场景图像的边界框注释中进行调整。

## 27日

### Information Diffusion on Social Media During Natural Disasters

自然灾害期间的社交媒体

IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 5, NO. 1, MARCH 2018

选择两次地震作为微博内容。

文中选择中国的两次地震作为新浪微博的社会背景。在提出一个正式的微博信息流模型来表示在微博上传播的信息后，从三个主要方面研究信息传播：个体特征，交互参与者之间的社会关系类型，以及真实交互网络的拓扑。定量分析得出以下结论。首先，**Dunbar数字**的阴影在“公开的朋友/追随者（存在关注关系）”分布中很明显，参与地震信息传播的每个参与者的朋友/追随者的数量显示出典型的幂律分布，表明“富者越富”。其次，个人的追随者数量是用户影响力最关键的因素。陌生人是在地震发生后传播实时新闻的重要力量。第三，两种类型的真实互动网络（微博and twitter?）共享无标度和小世界属性，但组织结构较为宽松。此外，不同影响群体之间的相关性表明，与其他在线社交媒体相比，微博的讨论主要受到认证用户（大V）的支配和影响。

Dunbar数字？

**Dunbar's number** is a suggested cognitive limit to the number of people with whom one can maintain stable social relationships—relationships in which an [individual](https://en.wikipedia.org/wiki/Individual) knows who each person is and how each person relates to every other person.

数据使用微博API获取

### Information Diffusion in Social Networks: Friendship Paradox based Models and Statistical Inference

社交网络中的信息传播：基于模型和统计推导的友谊悖论

本文讨论了与社交网络中信息传播有关的三个重要方面：

（i）名为友谊悖论（图论理论后果）和单亲传染（朋友朋友的影响）的观察偏见如何影响信息扩散动态。

（ii）社交网络如何根据信息传播的状态调整其结构连通性。

（iii）如何估计信息传播引起的网络状态。

信息扩散是指社交网络（图形）中各个节点的意见（状态）如何随时间演变。

考虑在无向网络上的SIS模型的离散时间版本，其在每个时刻涉及两个步骤： 在第一步中，来自群体的随机抽样的个体（代理）m观察到随机选择的代理（m的邻居）的d（m）（m的度）数。 在第二步中，基于d（m）观察，代理m的状态在概率上演变为两种可能状态之一：感染或易感。

本文中，底层网络被建模为反应网络：随机图过程，其在每个时刻的转移概率取决于信息扩散过程的状态。 当网络是根据信息扩散的状态随机演变的反应网络时，网络的集体动态和扩散过程可以通过常微分方程（ODE）近似（在某些假设下）。 代数约束。 从统计建模和机器学习的角度来看，这一结果的重要性依赖于这样一个事实：它提供了复杂系统的集体随机动力学的简单确定性近似（随机图上的SIS过程，在同一时间尺度上演化））。

太长，先不看。

### Epidemic Processes over Adaptive State-Dependent Networks

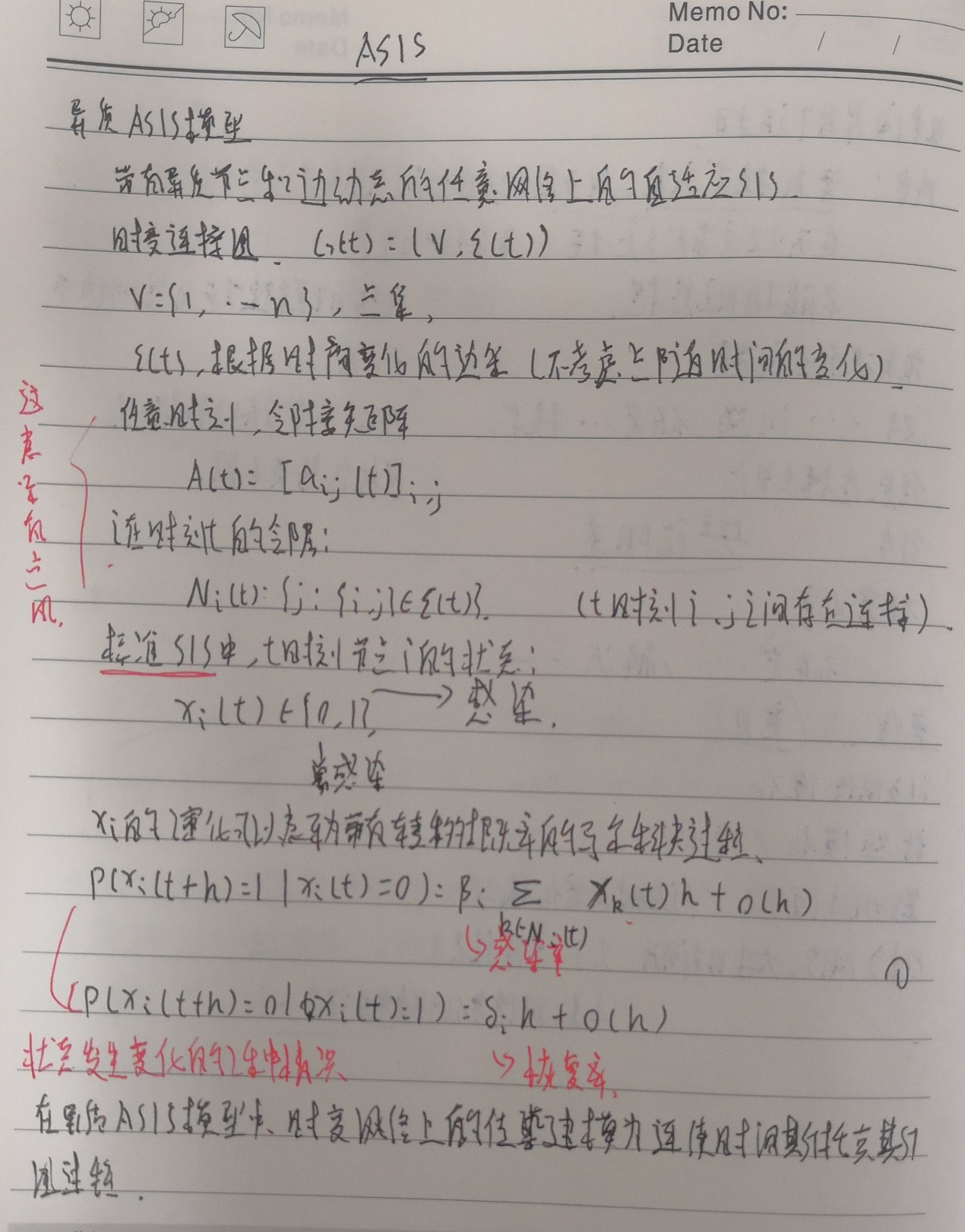
<https://arxiv.org/pdf/1602.08456.pdf>

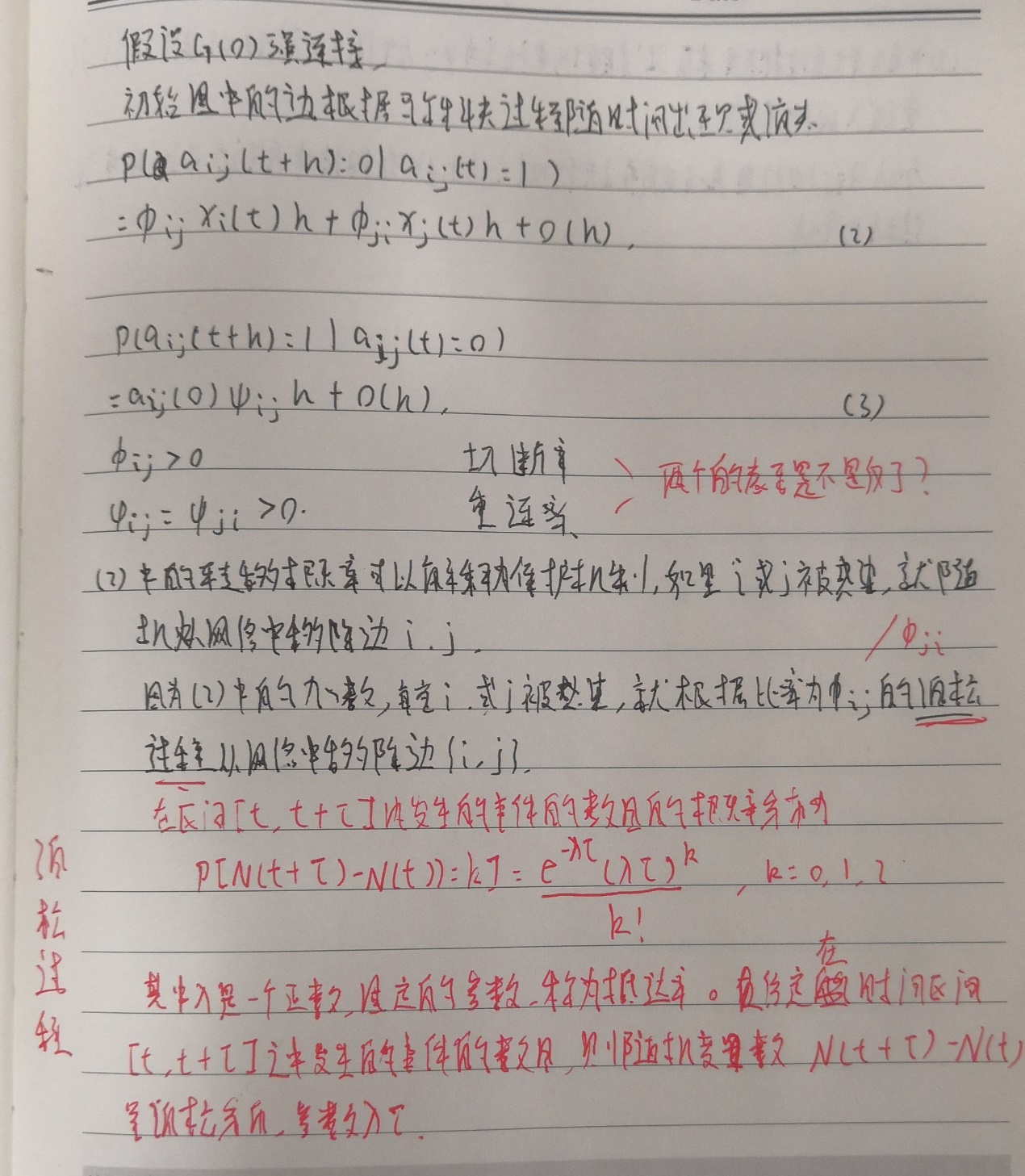
允许易感染节点临时切断与感染节点的链接以阻止感染的传播

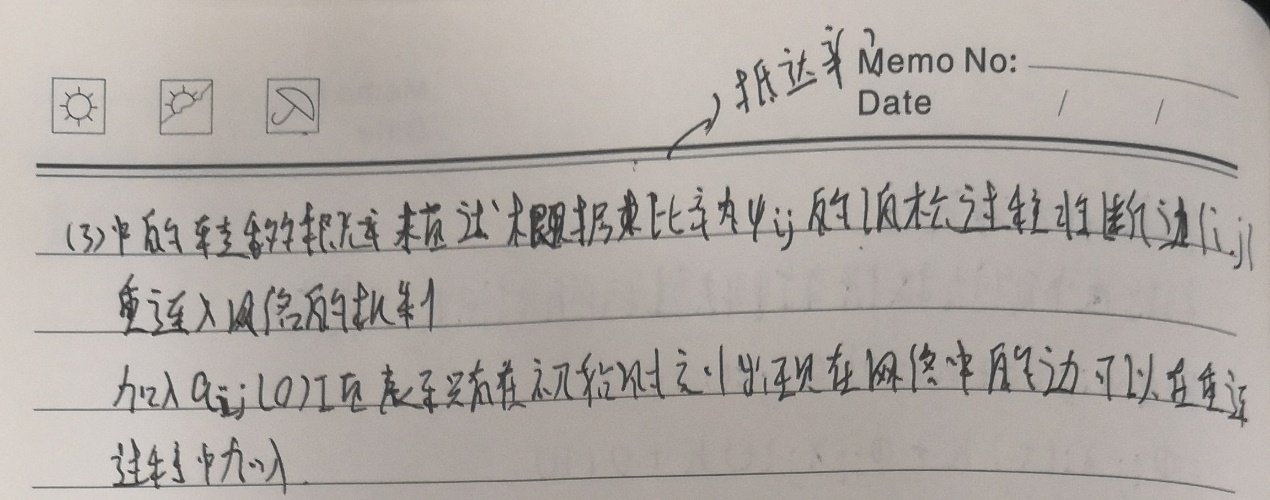
本文研究了在任意拓扑的自适应网络中发生的流行病过程的动态。我们将研究重点放在自适应易感染-感染-易感染（自适应SIS，ASIS）模型上，允许健康个体暂时切断连接到感染节点的边，以防止感染的传播。在本文中，我们得到了一个闭合形式的表达式，用于ASIS模型在具有异构节点和边缘动力学的任意网络中的传染阈值的下限。对于具有同质节点和边缘动态的网络，所得到的下界与标准SIS模型相对于静态网络的流行阈值成比例，其比例常数取决于适应性。此外，根据我们的结果，我们提出了一种有效的算法来优化调整适应率，以消除任意网络中的流行病爆发。我们通过几个数值模拟确认了所提出的下界的紧密度，并将我们的最佳适应率与流行的中心性度量进行了比较。

自适应网络？

异质ASIS模型







传染性阈值

传染病扑灭的最经济适应

### Networked SIS epidemics with awareness

Paarporn, K., Eksin, C., Weitz, J.S., Shamma, J.S.: Networked SIS epidemics with awareness.IEEE Transactions on Computational Social Systems **4**(3), 93–103 (2017)

<https://arxiv.org/pdf/1607.02502.pdf>

[73]研究了静态接触网络上的SIS过程，其中节点具有关于流行病状态的部分信息，并且当他们认为流行病目前普遍存在时，通过限制他们与邻居的相互作用来做出反应。  
我们在静态接触网络上研究SIS流行病过程，其中节点具有关于流行病状态的部分信息。当他们认为这种流行病目前很普遍时，他们通过限制与邻居的互动来做出反应。节点的意识由其社交网络中受感染邻居的分数以及整个网络中受感染节点分数的全局广播加权。基准（无意识）和意识模型的动力学由离散时间马尔可夫链描述，其中衍生出了领域近似（MFA）。 MFA的状态被解释为节点被感染的可能性。我们显示存在“亚稳态”的充分条件，该感知模型的状态与该基准模型的状态一致。此外，我们使用耦合技术在两个链（马尔科夫链）之间进行完整的随机比较分析，作为MFA分析的概率类比。特别是，我们表明，增加意识可以减少对样本路径空间的任何感染度量的预期，例如：根除时间或完全感染。我们的特点是在耦合分布方面降低了预期。在模拟中，我们评估社交距离对来自不同随机图族（几何，Erd̋os-Renyi和无标度随机网络）的接触网络的影响。

MFA对马尔科夫链进行邻域近似以求得近似结果。

### Epidemic processes over time-varying networks

Pare, P.E., Beck, C.L., Nedi ´ c, A.: Epidemic processes over time-varying networks. IEEE ´Transactions on Control of Network Systems **5**(3), 1322–1334 (2018)

Time-varying network时变网络的另一种表述？

Temporal network嘞？

https://arxiv.org/pdf/1609.05128.pdf

分析了时变网络上传染过程的稳定性，并提供了充分的收敛条件

### Information diffusion backbones in temporal networks

### **Adaptive Gradient Methods with Dynamic Bound of Learning Rate**

论文地址：<https://openreview.net/pdf?id=Bkg3g2R9FX>

项目地址：https://github.com/Luolc/AdaBound

AdaBound 最重要的属性是「和 Adam 一样快、和 SGD 一样好」

## 小结：

负抽样：

Skip-gram:

Softmax

ReLU

Attention

LSTM

谱技术

拉普拉斯矩阵：

图拉普拉斯（graph Laplacian）

拉普拉斯特征映射

自动编码器

**Network Motif:**

Network motifs are sub-graphs that repeat themselves in a specific network or even among various networks. Each of these sub-graphs, defined by a particular pattern of interactions between vertices, may reflect a framework in which particular functions are achieved efficiently. Indeed, motifs are of notable importance largely because they may reflect functional properties.

揭示网络结构的自相似，相同结构重复出现。检测存在困难，有许多motif属性的检测方法。

# 3月

## 1

### on-line adaptative curriculum learning for GANs

https://arxiv.org/pdf/1808.00020.pdf

生成性对抗网络（GAN）可以成功地对概率分布进行近似并产生逼真的样本。然而，诸如充分的融合条件和模式崩溃等悬而未决的问题仍然存在。在本文中，我们在该领域的现有工作的基础上，提出了一个新的框架，用于训练生成器对抗一个鉴别器网络，这可以被视为一个学生/多个教师的环境。我们在全信息对抗性强盗框架中形式化这个问题，我们评估算法选择混合鉴别器的能力，以便在学习期间为生成器提供反馈。为此，我们提出了一个奖励函数，它反映了生成器的进展，并动态更新分配给每个鉴别器的混合权重。我们还在我们的算法和随机优化方法之间建立联系，然后展示使用文献中的多个鉴别器的迭代方法可以从我们的框架中恢复。我们认为，较少的表达式判别器更平滑，并且对模式映射具有一般的粗略视图，这强制了发生器覆盖了大部分数据分布支持。另一方面，高度表达的鉴别器确保样本质量。

**鉴别器混合**

专家混合问题，full-information adversarial bandit setting。

### - One-class adversarial nets for fraud detection

<https://arxiv.org/pdf/1803.01798.pdf>

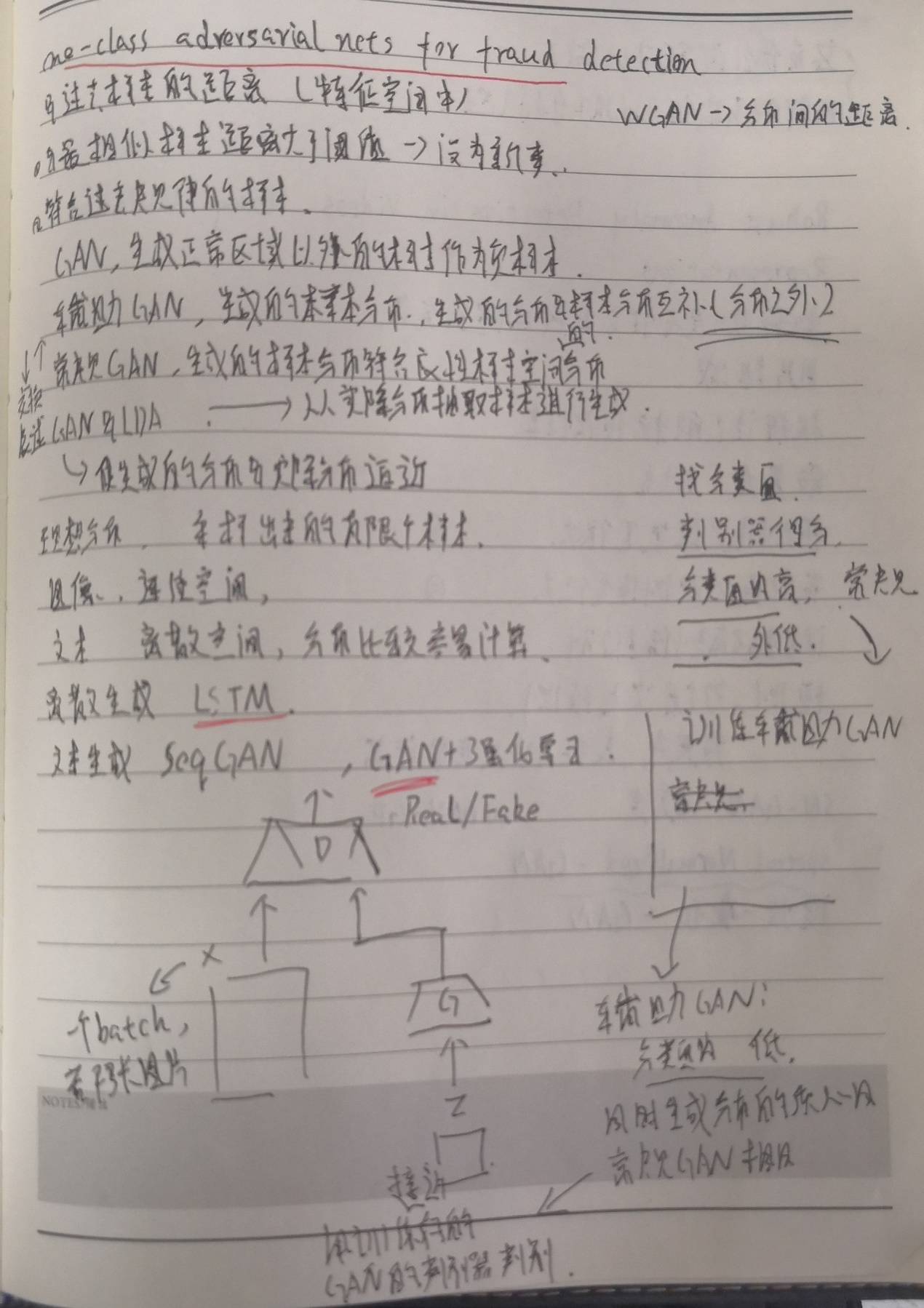
诈骗检测，不需要负样本。

生成负样本。

目前，大多数欺诈检测方法需要包含良性和恶意用户记录的训练数据集。 但是，在实践中，通常没有或很少有恶意用户的记录。 在本文中，我们开发了用于欺诈检测的一类对抗网（OCAN），其中，训练数据只有良性用户。 OCAN首先使用LSTM-Autoencoder在线活动序列中学习良性用户的表示。 然后，它通过与常规GAN模型不同的辅助GAN模型来训练鉴别器检测恶意用户。

**框架概览**

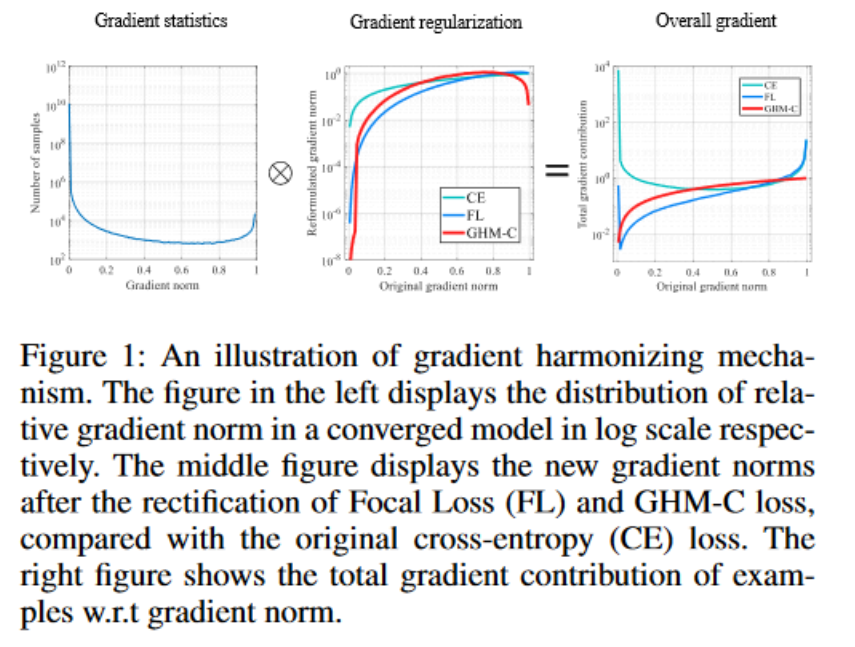
OCAN在 训练期间包含两个阶段。第一阶段是学习用户表示。采用LSTM-Autoencoder来学习良性用户活动序列的良性用户表示。 LSTM-Autoencoder模型是一个序列到序列模型，它分别由两个LSTM模型和编码器和解码器组成。编码器计算输入的隐藏表示，并且解码器基于隐藏表示来计算构造的输入。训练的LSTM-Autoencoder可以捕获用户活动序列的显着信息，因为目标函数是使得构造输入接近原始输入。此外，经过训练的LSTM-Autoencoder的编码器在部署用于欺诈检测时，由于良性和恶意用户的活动顺序不同，因此有望将良性用户和恶意用户映射到连续特征空间中相对独立的区域。给予用户表示，第二阶段是使用可以清楚地区分良性和恶意用户的鉴别器来训练补充GAN。互补GAN的生成器旨在生成在良性用户的低密度区域中的补充样本，并且鉴别器旨在分离真实和互补的良性用户。然后，鉴别器具有检测位于与良性用户不同的区域中的恶意用户的能力。图1右侧显示了 训练互补GAN用于欺诈检测的框架。训练OCAN的伪代码显示在算法1中。给出了包含Nbenign用户的活动序列特征向量的训练数据集Mbenignt，我们首先训练LSTM -Autoencodermodel（第3-9行）。在训练LSTM-Autoencoder之后，我们采用LSTM-Autoencoder模型中的编码器来计算良性用户表示（第11-14行）。最后，我们使用良性用户表示来训练互补的GAN（第16-20行）。简单来说，我们用小批量大小为1编写算法，即在训练数据集中对每个用户进行训练以训练LSTM-Autoencode和GAN。在实践中，我们对样本良好的用户进行抽样，并使用生成器在一个小批量中生成完整的样本。

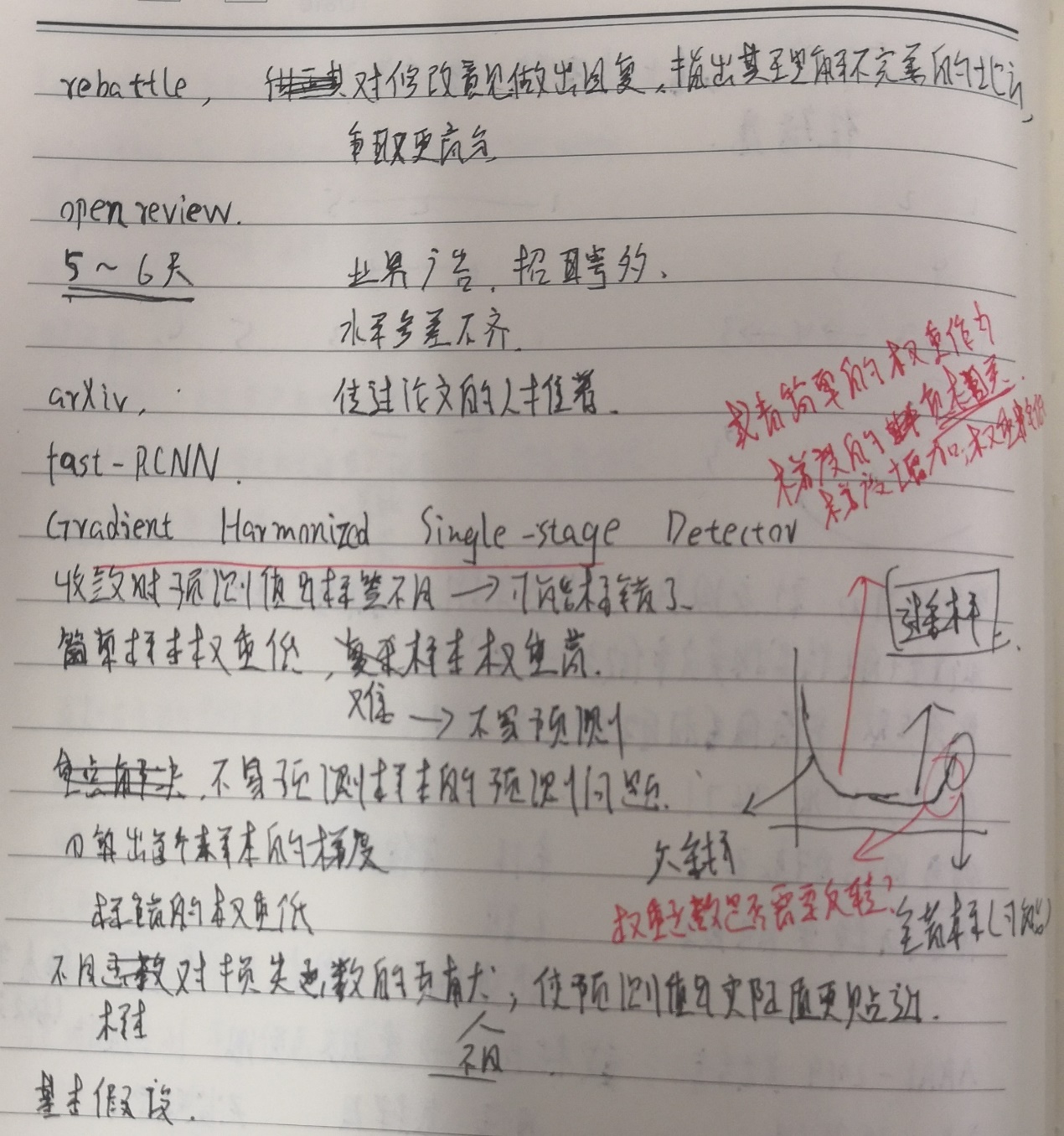


### - Gradient Harmonized Single-stage Detector

<https://arxiv.org/pdf/1811.05181v1.pdf>

给不同样本赋予不同权重。





## 2

### \*Generative Adversarial Nets

<http://www.github.com/goodfeli/adversarial>

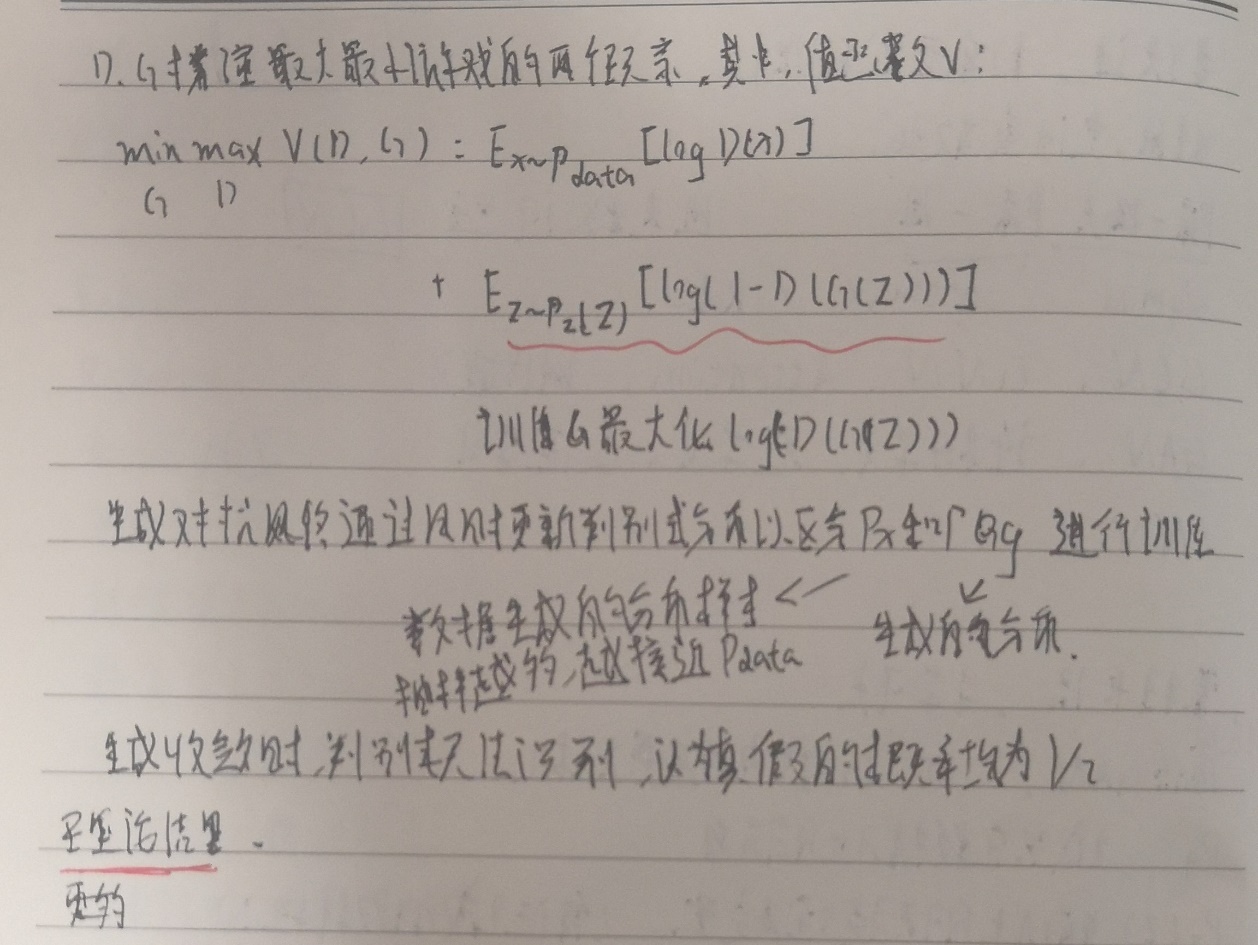
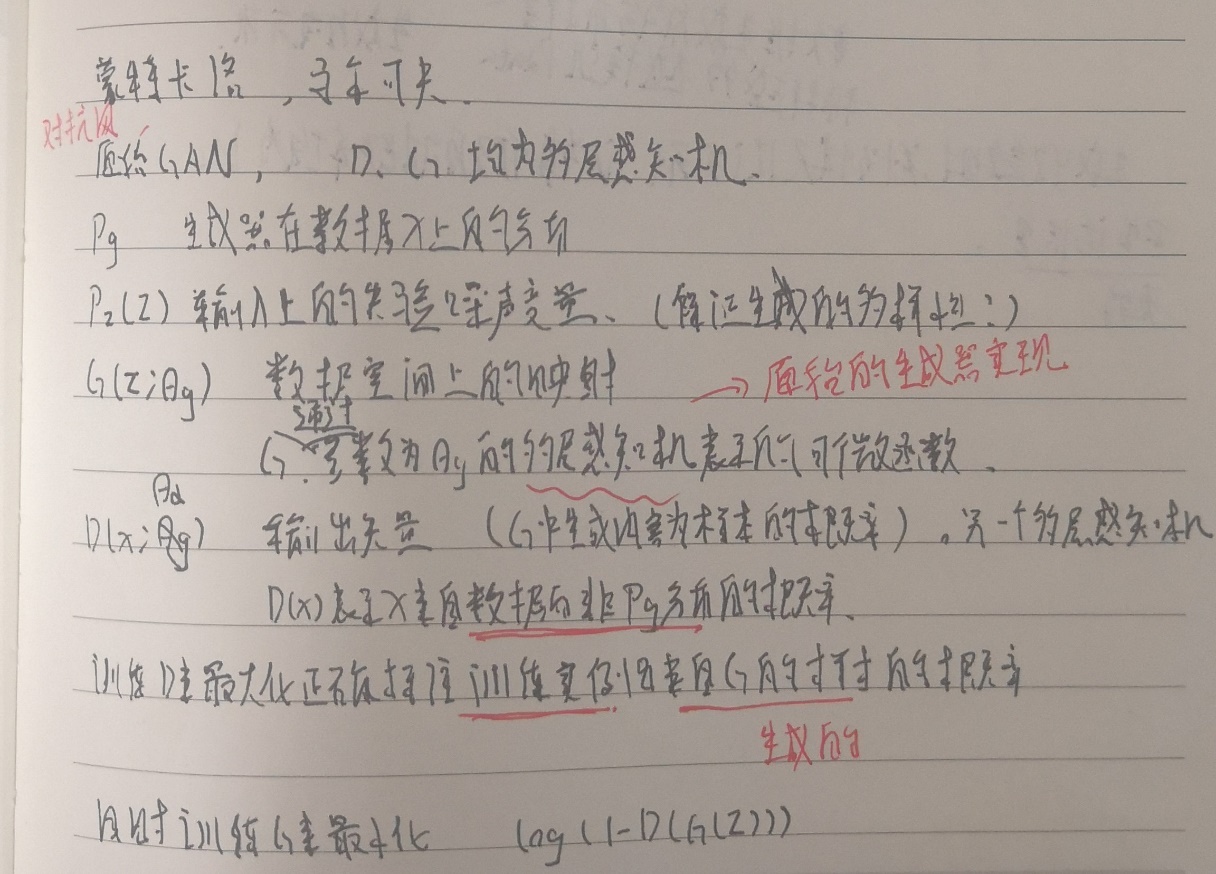
大名鼎鼎的GAN

一个通过对抗过程估计生成模型的新框架，同时训练两个模型：捕获数据分布的生成模型G和估计样本来自训练数据的概率的判别模型D. G的训练程序是最大化D犯错误的概率。 这个框架对应于极小化极大算法双人游戏。 在任意函数G和D的空间中，存在唯一的解决方案，其中G恢复训练数据分布并且D在任何位置都等于1/2。 在G和D由多层感知器定义的情况下，整个系统可以用反向传播进行训练。在训练或生成样本期间不需要任何马尔可夫链或展开的近似推断网络。

（提出GAN之前。）深度生成模型的影响较小，因为难以近似在最大似然估计和相关策略中出现的许多难以处理的概率计算，并且由于难以在生成环境中利用分段线性单元的好处。

在对抗性网络框架中，生成模型与对手进行对比：一种判别模型，用于学习确定样本是来自模型分布还是数据分布。生成模型可以被认为类似于造假者团队，试图生产虚假货币并在没有检测的情况下使用它，而歧视模型类似于警察，试图检测伪造货币。在这个游戏中的竞争促使两个团队改进他们的方法，直到不能从真正的物品中识别出赝品。

该框架可以为多种模型和优化算法产生特定的训练算法。在本文中，我们探讨了生成模型通过多层感知器传递随机噪声来生成样本的特殊情况，并且判别模型也是多层感知器。我们将这种特殊情况称为对抗性网络。在这种情况下，我们可以仅使用非常成功的反向传播和丢失算法[17]来训练两个模型，并且仅使用前向传播来生成来自生成模型的样本。不需要近似推理或马尔可夫链。



### Attention is All You Need

注意力机制。

NIPS 2017

不是对Attention的介绍，而是如何只利用Attention实现卷积实现的功能。使用注意力机制实现的序列转换模型，实验环节在机器翻译中做了验证。

编码器解码器的作用？

Softmax主要用来做的工作？

如果不用激励函数，每一层输出都是上层输入的线性函数，无论神经网络有多少层，输出都是输入的线性组合。  
如果使用的话，激活函数给神经元引入了非线性因素，使得神经网络可以任意逼近任何非线性函数，这样神经网络就可以应用到众多的非线性模型中。

## 4

### ATM: Adversarial-neural Topic Model

东南大学 2018年论文

主题模型广泛用于文本中的主题结构发现。 但传统的主题模型通常需要针对特定任务的专用推理程序。 而且，它们不是为了生成单词级语义表示而设计的。 为了解决这些局限性，我们提出了一种基于生成性对抗网（GAN）的主题建模方法，称为对抗 - 神经主题模型（ATM）。 提出的ATM使用Dirichlet先验对主题建模，并使用生成器网络捕获潜在主题之间的语义模式。 同时，生成器还可以生成单词级语义表示。 为了说明将ATM移植到主题建模以外的任务的可行性，我们将ATM应用于开放域事件提取。

单词级语义表示的意义呢？

主要贡献：

在主题模型上使用对抗训练。

不仅能从文本中抽取主题，还能生成编码了词间语义关系的词嵌入。

**对抗神经主题模型**

**表示映射**

文档表示为V维归一化向量，使用TF-IDF获取权重

**网络架构、**

**训练**

**主题生成**

### Don’t Settle for Average, Go for the Max: Fuzzy Sets and Max-Pooled Word Vectors

<https://openreview.net/pdf?id=SkxXg2C5FX>

ICLR 2019

可以用来在缺少语料的条件下（只有用户属性标签）进行用户和话题的对比。

最近的文献表明，平均单词向量后跟简单的后处理优于语义文本相似性任务的许多深度学习方法。此外，当平均的单词向量在大量的释义上受到监督时，它们在标准STS基准标记上获得了最先进的结果。受这些见解的启发，我们甚至可以进一步突破单词嵌入的极限。我们提出了一种新颖的模糊词袋（FBoW）表示文本，其中包含词汇表中的所有单词，但同时具有不同的隶属度，隶属度从单词向量之间的相似性得出。我们证明了最大池化（max-pooled）单词向量只是一个模糊BoW的特例，应通过模糊Jaccard指数而不是余弦相似度进行比较。最后，我们提出了DynaMax，一种完全无监督和非参数的相似性度量，它根据这些对，动态地提取和最大化良好的特征。该方法既高效又易于实现，但是在STS任务上大大优于当前基线，并且甚至与经过训练直接优化余弦相似性的监督词向量竞争。

词袋都无视了词的顺序关系。

SoW，将句子看作词的集合。

BoW，SoW的扩展，考虑不同词频。

（Fuzzy set）模糊集同时包括词汇表中的所有词，只是具有不同的成员度。 这概括了经典集合，其中单词属于集合或不属于集合。（解决同义词问题）。内容不同但语义相近的词袋比较时相似度不再为0.不是向量，所以不能使用余弦相似度，需要使用Jaccard

通过最大池化获取词袋之间的相似关系。

词嵌入

## 5

### \*node2vec: Scalable Feature Learning for Networks

2016年

有代码

**特点：**

无监督

没有模型，只有训练，没有预测节点

直接学习节点协作

不能泛化到不可见节点；整个嵌入必须被重新训练

应用于大图时包含太多参数

**而推导式特征学习：**

可以用于大型网络

可以泛化到新节点

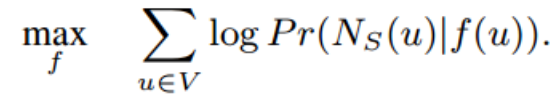
对顶点嵌入建立了模型：

训练阶段可能会很慢，但是预测阶段会很快

做节点的内容嵌入和结构嵌入。

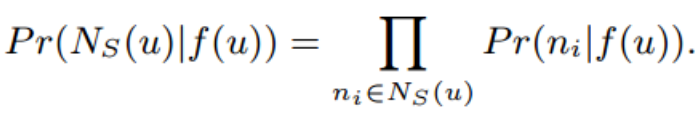
扩展了Skip-gram网络架构。

目标函数：最大化在u的特征表示为f时观测到网络邻居的对数概率。

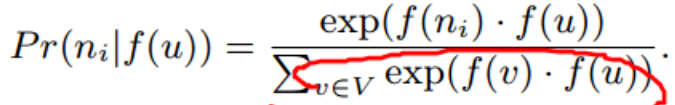


**两个假设**：

条件独立。假设给定源特征表示的条件下，观测到一个邻居节点的似然与观测到其他邻居节点的似然独立。（就可以直接相加，此处相当于做了简化）



特征空间对称。源节点与特征节点在特征空间中互相之间有对称效应。因此，我们将每个源节点对的条件似然建模为以他们特征的点积为参数的Softmax单元。



然后目标函数可以表示为



其中



在定义特征f的模型参数上使用随机梯度下降对进行优化，减少运算时间。

借鉴的Skip-gram，其中词的邻居窗口中的内容是有序的，但网络却是无序的，为此，设计一个随机过程抽取给定源节点的多个不同邻居节点（根据抽样策略，不一定是直接邻居节点）。

结论和过程不及图注意力网络优美。

对邻居节点进行抽样作为局部搜索的一种形式。

广度优先抽样：

获取节点邻居的微观视图。

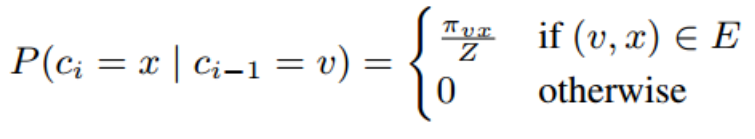
深度优先抽样：

获取节点邻居的宏观视图。

**Node2vec**

*随机行走*

给定源节点，随机行走固定步长。通过如下分布生成第i步节点（确定第i步去哪个节点）。

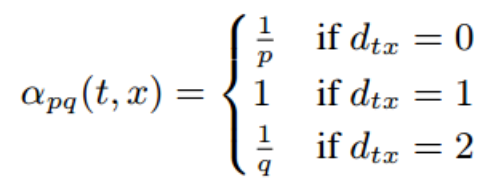


其中为v,x间的非归一化转移概率。Z为标准化常数。

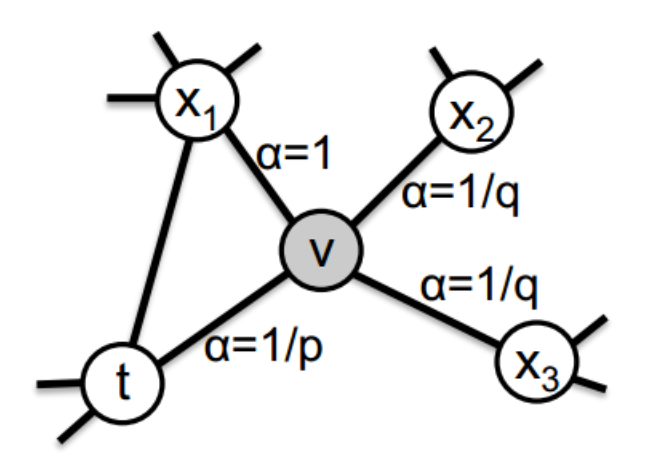
*搜索偏差*

最简单的方法是基于静态边权重抽取下一个节点。

定义了一个参数为p和q的二阶随机行走：考虑随机行走刚穿过边（t，v），现在停留在v节点。随机行走需要决定下一步，所以就评估从v开始的边（v,x）上的转移概率，将非标准化的转移概率设为



其中为t和x之间的最短路径距离（0,1,2）。



返回参数p：控制再访的可能性（再次访问刚刚访问过的节点），q越大，再访概率越低。

入出参数q：允许搜索区分内部点和外部点。如果q小于q，更倾向于访问离节点t较远的节点。

Node2vec算法

通过从每个节点开始模拟固定步长为l的r随机行走抵消起始节点带来的偏差。

3步：

预计算转移概率

随机行走模拟

使用SGD优化（目标函数）。

### \*Mutual Information Neural Estimation

https://arxiv.org/pdf/1801.04062.pdf

用神经网络估计互信息

### Dual Attention Networks for Multimodal Reasoning and Matching

https://arxiv.org/pdf/1611.00471.pdf

基于注意力的文本和图像相互匹配

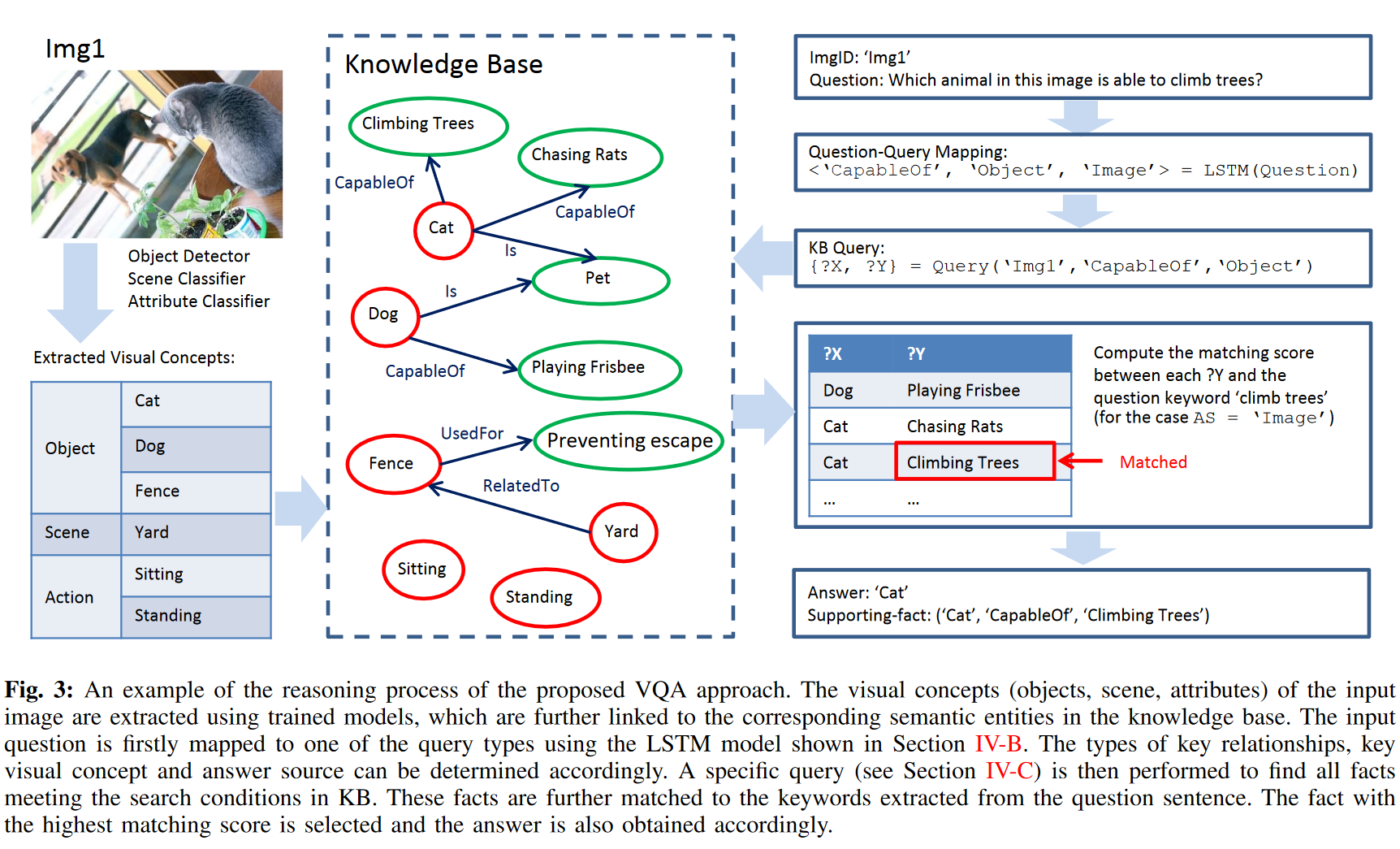
### FVQA: Fact-based Visual Question Answering

https://arxiv.org/pdf/1606.05433.pdf

图像问答

Supporting-fact 在外部知识库的知识的结构化表示。

{图像-问题-答案}三元组作为训练实例



### Relational Pooling for Graph Representations

<https://arxiv.org/pdf/1903.02541.pdf>

使用Relational Pooling(PR)泛化了GNN（之前的是基于Weisfeiler Lehman WL算法、 graph Laplacians、 graph diffusion kernels）

在这项工作中，我们介绍了关系池（RP），这是一种新颖的图神经网络框架，可以为任何图输入提供最大的表示能力。 当前的方法大多是自下而上的，通过图上的转换进行指定，RP采用自上而下的方法，为图指定理想化的最强大的表示，并提供了一个框架来开发算法近似这个理想。 RP理想情况下可以区分WL同构测试无法区分的图形，更令人惊讶的是，我们表明在RP框架下，可以使用基于WL-GNN的实用算法来区分不能通过相应的WL同构测试区分的图形。

数学太复杂。张量表示中引入了自己定义的一种操作。

知识点：

Graph Laplacians

Graph diffusion kernel

Recurrent Neural Network

Convolutional Neural Networks

### Meta-Leaning with Differentiable Closed-Form Solvers

<https://arxiv.org/pdf/1805.08136.pdf>

ICLR2019

元学习：学习如何学习。

由于标准微调程序的高计算要求，根据少数例子调整深度网络以适应新概念是具有挑战性的。因此，关于少样本学习（few-shot learning）的大部分工作集中于简单的适应学习技术，例如最近邻居或梯度下降。尽管如此，机器学习文献包含了大量有效学习非深度模型的方法。在本文中，我们建议使用这些快速收敛方法作为少样本学习的主要适应机制。主要思想是使用标准机器学习工具（例如岭回归）作为深度网络内部模型的一部分，使其能够快速适应新颖数据。这需要通过求解器步骤反向传播错误。虽然通常在这样的过程中涉及的矩阵运算的成本很高，但通过使用Woodbury特性，我们可以使少量的例子起作用。我们提出了基于岭回归和逻辑回归分量的闭合求解和迭代求解器。

### Learning Deep Representations by Mutual Information Estimation and Maximization

ICLR 2019

<https://arxiv.org/pdf/1808.06670.pdf>

提出Deep InfoMax(DIM),通过最大化深度神经网络编码器的输入和输出之间的相互信息来研究表示的无监督学习。 重要的是，我们表明结构很重要：将输入中的地理知识纳入目标可以显着提高表示对下游任务的适应性。 我们通过与对方的先前分布匹配来进一步控制表示的特征。

## 9

### Global Vectors for Node Representations

<https://arxiv.org/pdf/1902.11004.pdf>

<https://github.com/brochier/gvnr> 计划提供代码，尚未开源

大多数网络嵌入算法包括通过随机游走测量节点的共现，然后使用带有负抽样的Skip-Gram来学习嵌入。虽然它已被证明是一个相关的选择，但还有其他选择，例如GloVe，在网络嵌入领域的应用尚未被深入研究。即使SGG比GloVe更好地处理非共现，它的复杂性也更差。在本文中，我们提出了一种基于GloVe的网络嵌入矩阵因子分解方法，它可以更好地处理非共存的竞争时间复杂度。我们还展示了如何通过同时学习单词，节点和文档表示来扩展此模型以处理节点是文档的网络。

### \*Inductive Representation Learning on Large Graphs

<https://arxiv.org/pdf/1706.02216.pdf>

William L. Hamilton∗ wleif@stanford.edu

Rex Ying ∗rexying@stanford.edu

Jure Leskovec [jure@cs.stanford.edu](mailto:jure@cs.stanford.edu)

**GraphSage**

**NIPS 2017**

Inductive Representation Learning on Large Graphs

将GCN的思想调整为归纳式节点嵌入

泛化了简单卷积，

证明了这些泛化：

提升表现

允许模型学习局部结构

在从内容推荐到识别蛋白质功能的各种预测任务中，大图中节点的低维嵌入已被证明是非常有用的。 然而，大多数现有方法要求在嵌入训练期间图中的所有节点都要出现; 这些先前的方法本质上是传统的，并且不会自然地推广到看不见的节点。 这里我们介绍GraphSAGE，这是一种利用节点特征信息（例如，文本属性）的一般性归纳框架，可以有效地为以前看不见的数据生成节点嵌入。 我们学习了一种通过对节点的本地邻域中的特征进行采样和聚合来生成嵌入的功能，而不是为每个节点训练单独的嵌入。

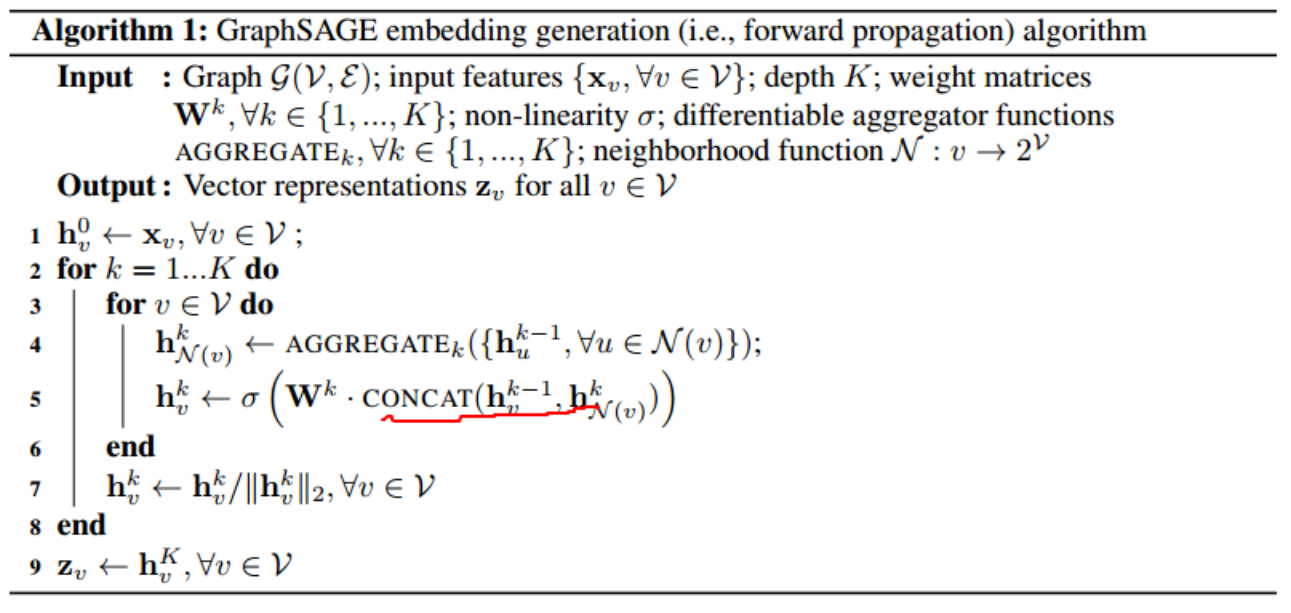
将GCN扩展到归纳式无监督学习。提出框架，泛化GCN方法以使用可训练的聚类函数（而不仅仅是简单的卷积）。

Sample and aggregate

不同于矩阵分解方法，利用了节点特征（如文本属性，节点配置文件信息，节点度）来学习泛化到不可见节点的嵌入函数。同时学习了节点邻居的拓扑结构和节点邻居的特征分布。

GraphSaGE不是为每个节点训练一个独特的嵌入向量，而是训练一组聚合器函数，该函数学习从节点的局部邻域聚合特征信息。 每个聚合器函数聚合来自给定节点的不同跳数或搜索深度的信息。

使用节点分类对算法进行评估。（节点嵌入经典操作）



**学习GraphSaGE的参数**

使用基于图的损失函数来输出表征，使用随机梯度下降调整权重矩阵及聚合函数的参数

基于图的损失函数鼓励相邻节点具有相似的表征



其中v为固定步长随机行走中在u附近共同出现的节点。是Sigmoid函数，为负抽样分布，Q定义负抽样的数目。

**聚合器架构**

节点的邻居没有天然的顺序关系。

平均集合器



LSTM聚合器：本质上是不对称的。

池化聚合器：对称和可训练的。



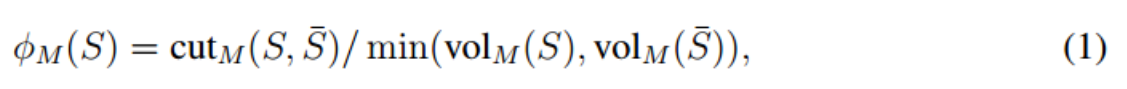
为非线性激活函数。

### Higher-order organization of complex networks

Austin R. Benson,1 David F. Gleich,2 Jure Leskovec

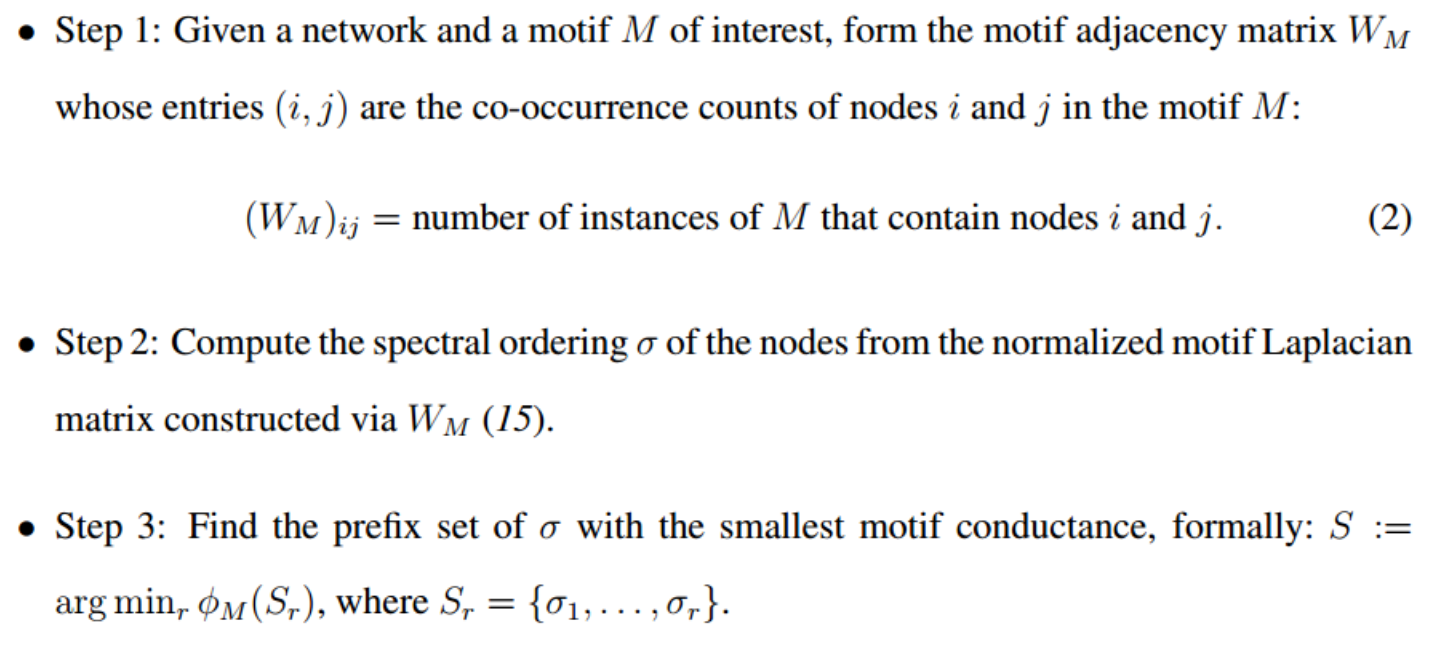
76页……

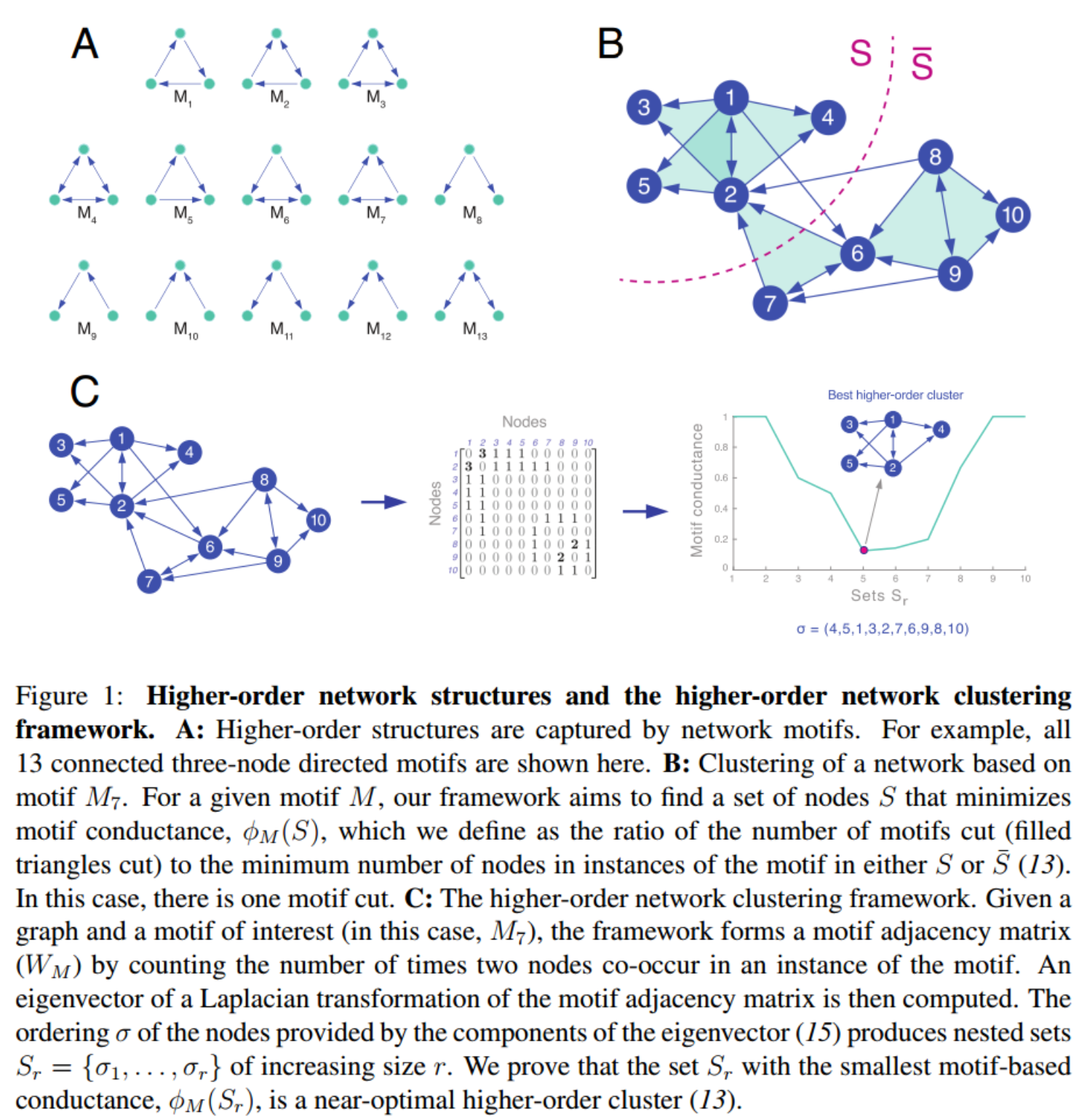
给定motif M，高阶聚类框架的目标是找到最小化下方比率的簇



表示剩余节点（S的补），表示至少有一个节点在S中且至少有一个节点在中的motif M实例的个数，表示驻留在S中的M的实例的个数。

算法通过如下三步有效的识别节点集合S中的簇：





13种motif

#### Derivation and analysis of the motif-based spectral clustering method

Motif-based embedding

文章看不懂。

就是不同的连接模式中边有不同的权重？

没看出价值。看看相关论文吧

## 11

### \*Mutual Information Neural Estimation

MINE

Mohamed Ishmael Belghazi 1 Aristide Baratin 1 2 Sai Rajeswar 1 Sherjil Ozair 1 Yoshua Bengio 1 3 4  
Aaron Courville 1 3 R Devon Hjelm 1 4

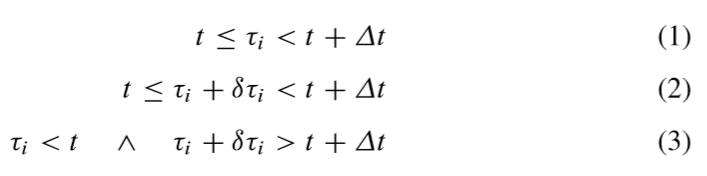
包含松估计器 紧估计器，可以得到任意精度。

认为高维连续变量的互信息可以通过神经网络上的梯度下降获取。

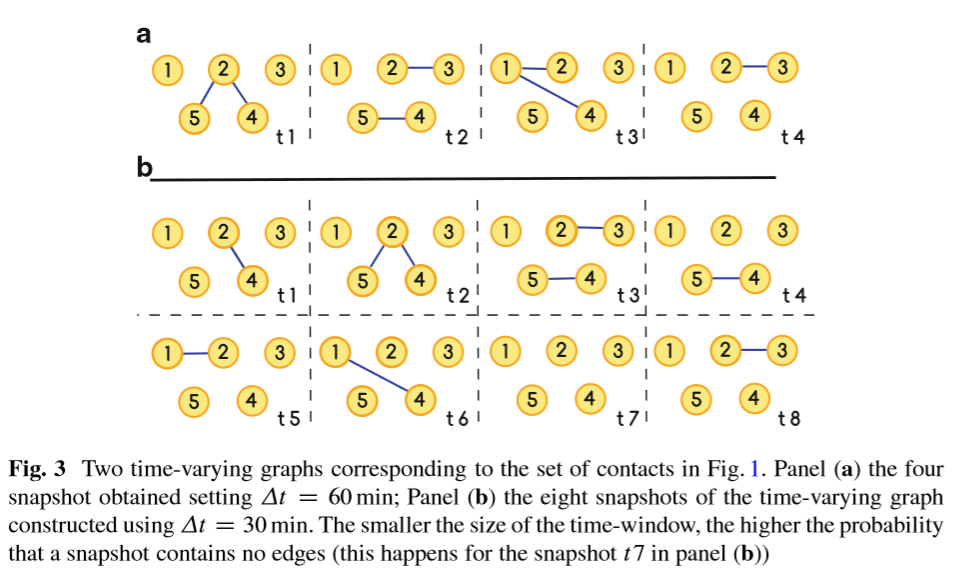
详见论文解读。

### Graph Metrics for Temporal Networks

2.2 考虑时间窗口内是否出现连接作为边。而不是某个时间点的连接状态。有效解决短时间连接在某个时间点捕捉不到的情况。



时变网络是在N个节点上定义的M个图的序列

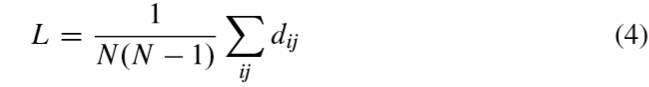


Maximal cliques of the affine graph

#### 4 distance, efficiency and temporal clustering

节点之间的临时最短距离：

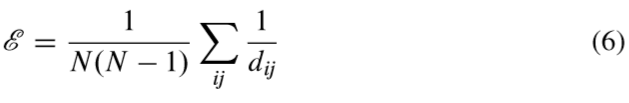
特征临时路径：定义为图中节点对之间平均临时距离。



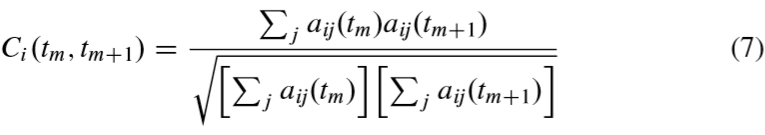
图的临时直径：



上面定义图直径的问题：时变图中存在大量暂时不相连的节点对，它们的距离通常定义为无穷。为了避免分歧，时变图的暂时全局效率（temporal global efficiency）定义为：

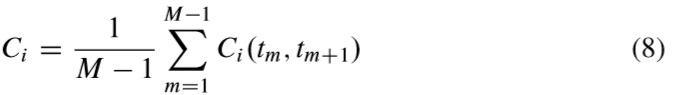


上的拓扑重叠：



表示什么？？？

节点i邻居的平均拓扑重叠，在所有可能的后来的暂时快照上的平均值：



平均拓扑重叠估计从节点i到它的邻居j的边保留超过穿过两个连续时间窗口（两个时间窗口中都出现）的概率。

在所有节点上计算的平均值，即时间相关系数：



#### 5 中间中心度 紧密中心度及谱中心度

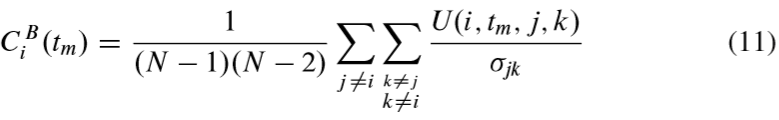
##### 5.1 中间中心度及紧密中心度

静态图上的中间中心度：



是从节点j到节点k的最短路径的数目，是从节点j到节点k，通过节点i的路径的数目。

暂时中间中心度：



在时刻通过



紧密中心度表示节点i与其他节点有多近。可以通过网络中i到其他节点平均距离的倒数获取。

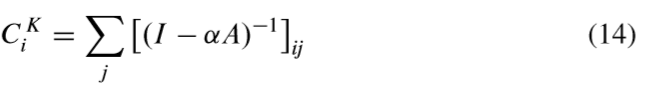


##### 5.2 谱中心度及沟通度

Katz中心度：节点i和节点j交流的倾向可以长度为1 2 3 的步的数目来量化。长度越短的步越重要。

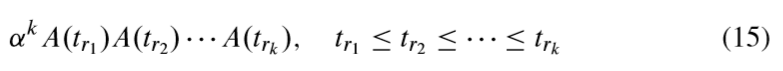
通过因子对步长进行缩放，

中的条目度量节点i同节点j交互的倾向。I为的单位矩阵。



节点i的Katz中心度S在第i行的和度量。

在时变网络中，为了计算任意长度的所有可能路径数，需要求所有可能形式的乘积总数：



如果，所有乘积的总数可以表示为：



Q称为通信矩阵，由此可以定义广播中心度：



广播中心度量化节点i在时变网络中到达所有其他节点的好坏。

和接收中心度：



#### 6 Meso-scale 结构

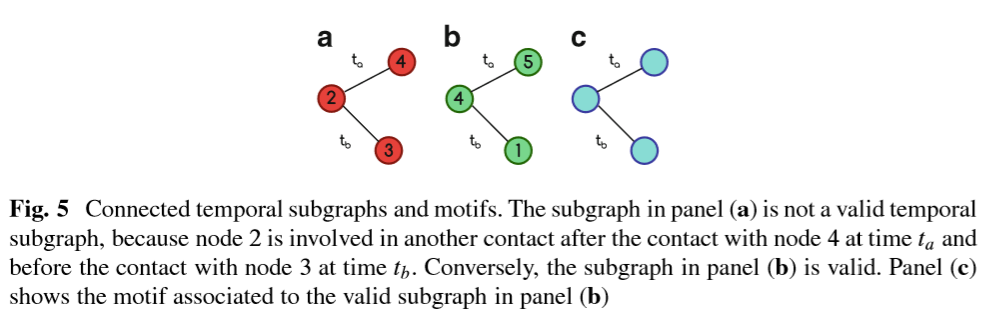
##### 6.1 时变Motif

先说静态图Motif

如果和是同构的，就存在置换矩阵P：



有效时变子图



这个没看懂，a与b有啥区别？？？

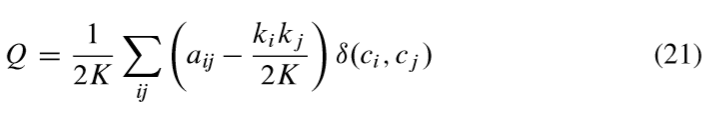
##### 6.2 时变社区和子模性

社区的无重叠划分可以通过子模函数度量



其中为邻接矩阵的元素，表示零模型图（null-model graph）中I j之间边数目的期望，为i所属的社区，如果节点i、j属于同一个社区，;K为图中边的总数。

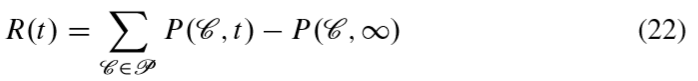
最简单的零模型图中，图中的所有顶点具有相同的度，但边是随机放置的。



其中表示节点i的度。

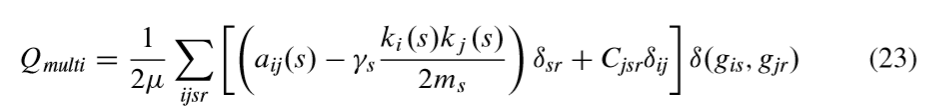
扩展到时变网络：

划分在t时刻的稳定性定义为：



假设随机行走达到稳定状态，表示随机行走者从社区中的一个节点出发，并在时刻t在中的一个节点发现（t步后仍在社区中）的概率。

多片图（multi-slice graph），



时间片s

表示片s中的节点i的度。表示片s中边的总数。

为Kronecker函数。？一个二元函数，克罗内克函数的[自变量](https://baike.baidu.com/item/%E8%87%AA%E5%8F%98%E9%87%8F/6895256)（输入值）一般是两个[整数](https://baike.baidu.com/item/%E6%95%B4%E6%95%B0/1293937)，如果两者相等，则其输出值为1，否则为0。（表示是否属于相同社区）

等于1，当时间片s的节点i与时间片r的节点j属于相同社区的条件下。

## 12

### CommunityGAN: Community Detection with Generative Adversarial Nets

上海交大 19年。

有代码<https://github.com/SamJia/CommunityGAN>

Motif该如何翻译？主题？

社区检测是指发现共享相似属性或功能的顶点组以便理解网络数据的任务。随着最近深度学习的发展，图形表示学习技术也被用于社区检测。但是，社区只能通过应用基于学习顶点嵌入的聚类算法来推断。像K-means和Gaussian Mixture Model这样的通用聚类算法不能输出很多重叠的社区，而重叠社区已被证明在许多现实世界的网络中非常普遍。在本文中，我们提出了CommunityGAN，这是一个新颖的社区检测框架，共同解决重叠的社区检测和图表示学习。首先，与传统图形表示学习算法的嵌入不同，CommunityGAN的嵌入表示顶点到社区的隶属强度。其次，采用特定设计的生成对抗网（GAN）来优化这种嵌入。通过主题级生成器和鉴别器之间的极小极大竞争，它们都可以交替地和迭代地提高它们的性能并最​​终输出更好的社区结构。

#### 知识点

Motif是什么？

Network motifs are sub-graphs that repeat themselves in a specific network or even among various networks.

clique是什么？（全连接子图）

In the [mathematical](https://en.wikipedia.org/wiki/Mathematics) area of [graph theory](https://en.wikipedia.org/wiki/Graph_theory), a **clique** ([/ˈkliːk/](https://en.wikipedia.org/wiki/Help:IPA/English) or [/ˈklɪk/](https://en.wikipedia.org/wiki/Help:IPA/English)) is a subset of vertices of an [undirected graph](https://en.wikipedia.org/wiki/Undirected_graph) such that every two distinct vertices in the clique are adjacent; that is, its [induced subgraph](https://en.wikipedia.org/wiki/Induced_subgraph) is [complete](https://en.wikipedia.org/wiki/Complete_graph).

Affiliation Graph Model（AGM，查看原论文）

### Community Preserving Network Embedding

### GraphGAN: Graph Representation Learning with Generative Adversarial Nets

2017年

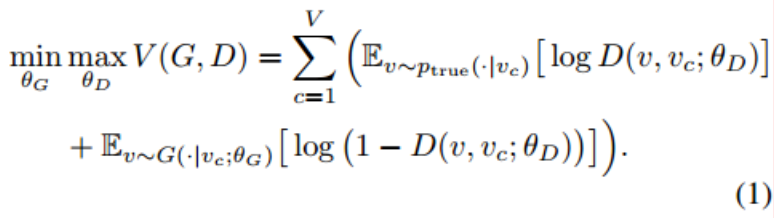
图形表示学习的目标是将图形中的每个顶点嵌入到低维向量空间中。现有的图表表示学习方法可以分为两类：学习图中潜在连通性分布的生成模型，以及预测一对顶点之间边存在概率的判别模型。本文提的GraphGAN统一了上述两类方法，其中生成模型和判别模型发挥了游戏理论极小极大的游戏。具体来说，对于给定的顶点，生成模型试图在其所有其他顶点上拟合其潜在的真实连通性分布，并产生“假”样本以欺骗判别模型；而判别模型试图检测采样顶点是否来自真实链接或由生成模型生成。随着这两种模型之间的竞争，它们都可以交替地和迭代地提高它们的性能。此外，在考虑生成模型的实现时，我们提出了一种新的图形softmax来克服传统softmax函数的局限性，可以证明它满足归一化，图形结构感知和计算效率的理想特性。

生成器根据什么生成？原始gan根据学习到的概率分布。

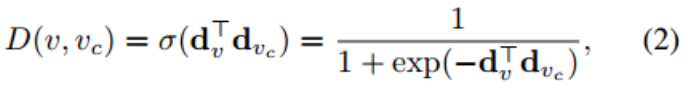
判别器根据什么判别？原始gan根据小批真实数据（图像）。

这里假设真实潜在连接分布，表示与图中所有其他顶点的连接偏好。

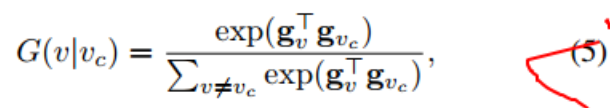
这里的目标函数：



判别器：两个输入顶点内积的sigmoid函数



生成器：直接方式，在其他所有顶点上的softmax函数



如何对生成器求解而不是再设计一个激活函数？

## 13

### Community-Affiliation Graph Model for Overlapping Network Community Detection

### \*Adversarial Network Embedding

ANE

<https://arxiv.org/pdf/1711.07838.pdf>

除了捕获网络结构属性的目标之外，多数节点嵌入方法都缺乏用于增强表征稳健性的额外约束。在本文中，我们的目标是利用生成对抗网络的优势来捕获潜在特征，并研究它在学习稳定和鲁棒图表示方面的贡献。具体而言，我们提出了一个逆向网络嵌入（ANE）框架，该框架利用对抗性学习原则来规范表征学习。它由两部分组成，即结构保留组件和对抗性学习组件。前一个组件旨在捕获网络结构属性，而后一个组件通过将潜在表示的后验分布与给定先验相匹配来帮助学习鲁棒表示。

详见解析

## 14

### \*Community-affiliation Graph Model for Overlapping Network Community Detection

AGM

Jaewon Yang  
Stanford Universitycrucis@stanford.eduJure Leskovec  
Stanford University[jure@cs.stanford.edu](mailto:jure@cs.stanford.edu)

网络中的社区常常重叠，因为节点可以同时属于多个社区。识别重叠社区对于理解真实世界网络的结构和功能至关重要。

尽管过去对网络中的社区结构进行了广泛的研究，但实际上所有研究都假设社区之间的重叠与非重叠部分本身的密集程度较低。在这里，我们在6个大型社交，协作和信息网络上验证了这一假设，其中节点明确说明了他们的社区成员资格通过检查这些地面真实社区，我们发现社区重叠与非重叠部分的联系更紧密，这与传统观念形成鲜明对比，即社区重叠与社区本身相比更为稀疏。

实际上，所有现有的社区检测方法都无法检测到具有密集重叠的社区。我们提出了Community-Affiliation Graph Model，这是一种基于模型的社区检测方法，建立在二分节点 - 社区从属网络之上。我们的方法成功地捕获了重叠的，非重叠的以及分层嵌套的社区，并且比从生物学到社会和信息网络的网络中的最先进方法更准确地识别相关社区。

在本文中，我们研究了真实社区的重叠，并发现节点共享边缘的概率随着它们共有的社区数量的增加而增加。 我们发现一对节点的共享社区数量与它们通过边缘连接的概率之间的关系越来越大。 这样做的直接后果是，社区重叠的网络部分往往比社区的非重叠部分更密集地连接（图1（c））。 这一观察结果与目前网络社区的结构定义形成鲜明对比，也意味着现有方法[1]，[2]，[23]无法正确识别这种社区重叠。 当前的社区检测算法要么错误地将重叠识别为单独的群集，要么将两个重叠的群落合并为单个群集。

详见解析。

## 15

### Community Detection in Networks with Node Attributes

### Overlapping Community Detection at Scale: A Nonnegative Matrix Factorization Approach

### \*Semi-Supervised Classification With Graph Convolutional Networks

<https://arxiv.org/pdf/1609.02907.pdf>

GCN

## 18

### Long Short-Term Memory

学习通过循环反向传播在延长的时间间隔内存储信息需要很长时间，主要是由于不充分的衰减误差。我们简要回顾一下Hochreiter 于1991年对这个问题的分析，然后通过引入一种新的，有效的，基于梯度的方法“长期短期记忆”（LSTM）来解决这个问题。通过截断不会造成伤害的梯度，LSTM可以学习通过在特殊单位内通过“恒定误差转盘”强制执行恒定误差，桥接最小时间滞后超过1000个离散时间步长。乘法门单元学习打开和关闭对恒定误差的访问。 LSTM是时空局部的;每个时间步长和重量的计算复杂度是O（1）。我们对人工数据的实验涉及局部，分布式，实值和噪声模式表示。通过与RTRL，BPTT，循环级联相关，Elman网和神经序列分块的比较，LSTM可以带来更多成功的运行，并且学得更快。 LSTM还解决了以前的循环网络算法从未解决的复杂的人工长时滞任务。

# 4月