本科毕业论文

论文题目：　基于结构的群组嵌入方法设计及在企业分析中的应用

姓　　名：王炜越　学　　号：15307130349

院　　系：计算机科学技术学院

专　　业：计算机科学与技术

指导教师：熊赟　职　　称：教授

单　　位：复旦大学

完成日期：2019 年 5 月 18 日

基于结构的群组嵌入方法设计及在企业分析中的应用

王炜越
学号：15307130349
专业：计算机科学与技术

摘要：目前已有很多关于社交网络分析的应用，如信贷反欺诈、保险反欺诈等等，这些应用中社交网络分析使用的都是网络中个体之间的联系来进行分析和判断。类似地，企业之间的相关信息也可以构成网络，使用相关社交网络分析框架,如群组嵌入框架 Community Embedding[1]（后简称为 ComE 框架），对企业网络可以进行企业分析，如有非正常关联的企业等等，用来给投资者提供参照作用。此外，企业网络具有较强的结构同一性，存在距离较远的节点拥有相似的结构特征的情况，所以不能仅仅通过具有相同的邻居来给节点分类，还需要仔细考虑节点的结构特征，但 ComE 中使用的 DeepWalk 得到的路径文件没有考虑重点考虑节点的结构特征。与 DeepWalk 相对，Struc2Vec 是一种重点研究节点的结构特征网络嵌入算法。因此，本文受启发于群组嵌入框架 ComE 和结构嵌入算法 Struc2Vec，提出一种新的企业群组分析框架：基于结构的群组嵌入框架----Struc_ComE，Struc_ComE 可以更好地帮助具有结构同一性的网络进行群组检测和节点分类，在企业分析情况上效果更好。

本文的主要贡献有：1) 根据 ComE 使用 DeepWalk 得到路径文件用来后续进行节点嵌入、群组检测和群组嵌入学习的思想，基于 Struc2Vec 有利于改进结构同一性网络的路径文件的启发，本文设计了基于结构的群组嵌入方法 Struc_ComE 框架；2) 将 Struc_ComE 框架在多种真实数据集上进行实验，与 ComE 框架实验结果进行对比分析；3) 在证明 Struc_ComE 框架比 ComE 框架能更好地识别结构特征之后，本文进一步研究社交网络分析算法在企业分析上的应用：使用 ComE 框架进一步对不同地区的企业网络进行实验分析以及进行步数研究。

关键字：Struc_ComE，ComE，结构同一性，结构特征，群组检测，节点分类。

Abstract: At present, there are many applications of social network analysis, such as credit anti-fraud, insurance anti-fraud and so on. In these applications, social network analysis uses the relationship between individuals in the network to analyze and judge. Similarly, the relevant information between enterprises can also form a network, and using the relevant social network analysis framework could do some analysis on these enterprises. For example, Community Embedding (after refer Community Embedding as ComE) framework can be used to analyze the enterprise network to get some information, like enterprises with abnormal associations, and so on, which can be used as a reference for investors. In addition, the enterprise network has strong structural identity, which means that the nodes with long distances have similar structural characteristics. Thus, we can not only classify the nodes by having the same neighbors, but also need to carefully consider the structural characteristics of the nodes. However, the path file obtained by DeepWalk used in come does not consider the structural characteristics of nodes. Compared with DeepWalk, Struc2Vec is a structural feature network embedding algorithm for nodes. Therefore, this paper enlightens the community embedding framework ComE and Struc2Vec, to propose a new enterprise community analysis framework: structure-based community embedding framework-Struc_ComE, Struc_ComE can better help the network with structural identity for community detection and node classification, and the effect is better in the case of enterprise analysis.

The main contributions of this paper are as follows: 1) According to the idea that ComE uses DeepWalk to get path files for subsequent node embedding, community detection and community embedding learning, based on Struc2Vec, it is beneficial to improve the path files of structural identity networks. This paper proposes a new Struc_ComE framework of community embedding method based on structure. 2) The Struc_ComE framework is tested on a variety of real datasets and compared with the experimental results of ComE framework. 3) After proving that the Struc_ComE framework can better identify the structural features than the ComE framework. In this paper, the application of social network analysis algorithm in enterprise analysis is further studied: the ComE framework is used to further analyze the enterprise network in different regions and study the values of parameters.

Keywords: Struc_ComE, ComE, Structure identity, Community detection, Nodes classification.

目录

第一章 引言.....5
1.1 研究背景和意义.....5
1.2 国内外研究现状.....6
1.2.1 图嵌入相关算法.....6
1.2.2 群组嵌入框架 ComE.....6
1.2.3 基于结构的网络嵌入算法 Struc2Vec.....8
1.3 本文贡献总结.....8
第二章 企业群组分析框架.....10
2.1 ComE 框架--群组嵌入框架.....10
2.2 Struc2Vec 算法.....11
2.3 基于结构的群组嵌入方法 Struc_ComE 框架.....12
2.4 小结.....15
第三章 实验工作.....16
3.1 实验流程.....16
3.2 数据处理.....16
3.2.1 节点为高管的网络.....16
3.2.2 节点为公司的网络.....16
第四章 Struc_ComE 结果对比分析.....18
4.1 节点为高管的网络结果分析.....18
4.1.1 浙江地区.....18
4.1.2 江浙沪地区.....19
4.1.3 东北三省地区.....20
4.2 节点为公司的网络结果分析.....21
第五章 ComE 结果分析.....23
5.1 地区研究.....23
5.1.1 广东地区上市公司.....23
5.1.2 江苏地区.....24
5.1.3 上海地区.....25
5.2 参数研究.....25
第六章 结论与未来改进.....27
6.1 结论.....27
6.2 未来改进.....27
第七章 致谢.....28
参考文献.....29

第一章 引言

目前社交网络分析算法已经在很多领域中有很好的应用效果，类似地，社交网络分析算法也可以应用在企业信息网络中，但企业信息网络具有结构同一性，分析过程中需要重点处理网络中的结构特征。本文在结构嵌入 Community Embedding 框架 ComE 和基于结构的网络嵌入算法 Struc2Vec 的基础上，提出企业群组分析框架：基于结构的结构嵌入框架 Struc_ComE，该框架中使用 Struc2Vec 算法代替 ComE 中的 DeepWalk 算法得到路径文件用于后续的节点嵌入、群组检测和群组嵌入任务处理。本章首先介绍相关研究背景和意义，再介绍国内外相关研究现状，最后陈述本文的主要贡献。

1.1 研究背景和意义

社交网络分析算法已经广泛应用于社交人物影响力计算、商品推荐、金融等分析领域，效果良好。社交网络就是一个可以用来表示人与人（节点与节点）之间联系的网络，社交网络分析就是分析人与人之间的联系，同样的，社交网络分析算法就是研究节点和节点之间的关系，通过研究和梳理这些关系，可以对节点进行归类，使其聚成团。

在金融领域中，社交网络分析算法已经被应用在如保险、信贷反欺诈等等。在信贷反欺诈中，对于通讯录中有相同的联系人的两个信贷者，具有高度相似性，就认为其之间的联系增强，其违约风险也就会较高。使用社交网络分析算法就可将这些具有高违约风险的人聚成的团筛选出来。

相似地，企业之间或者不同公司的高管之间也会形成一个网络，当一个公司高管同时任职不同公司时，就存在高管滥用职权来达成自己目的的危险，使得企业之间易存在非正常关联。当企业之间存在非正常关联时，那么其违反相关规定、具有不佳业绩或不良记录的可能性也会增大。及时挑选出这些具有非正常关联、存在投资风险的企业，给投资者进行提醒就很有必要。本文使用群组嵌入网络分析方法 ComE 进行群组检测和节点分类。

然而，与一些网络不同，企业网络具有结构同一性，很多性质特点相似的节点并没有相同的邻居或距离很近，但其与其他公司的关联形状与一些远距离的公司相同，因此这些公司应该被分到具有特殊性质的一类。但是 ComE 中使用 DeepWalk 得到路径文件，其无法有效捕捉节点结构特征，分析网络的过程中会忽略结构同一性，使得在企业网络中群组检测和节点分类效果变差。

1.2 国内外研究现状

1.2.1 图嵌入相关算法

社交网络分析算法分析和学习的是网络中各个节点的低纬度潜像表示，然后这些表示可以作为特征来帮助执行图的可视化、分类和聚类等各种任务。传统意义上的图嵌入是一个降维过程，用一个低维的矩阵来表示原始一个较高维的矩阵。目前图嵌入相关算法情况如**表 1** 所示。

表 1：当前国内外图嵌入相关算法简介

名称	算法特点
Unsupervised Network Embeddings：	
DeepWalk[3]	将节点视为单词并生成短随机游走作为句子来弥补网络嵌入和单词嵌入之间的差距。DeepWalk 按需生成随机游走，可扩展
LINE[4]	采用广度优先搜索生成上下文节点
Node2Vec[5]	DeepWalk 的扩展，其引入一个偏向随机步行程序。
Walklets[6]	学习多尺度网络嵌入
GraRep[7]	将图形邻接矩阵提升到不同的幂来利用不同尺度的节点共现信息。
SDNE[8]	学习节点表示，通过深度自动编码器保持两跳邻居的接近度。
DNGR[9]	也为基于深度神经网络的网络嵌入学习方法。
Attributed Network Embeddings：	
TADW[10]	研究节点与文本特征关联情况。
CENE[11]	公同模拟节点中的网络结构和文本内容。
HSCA[12]	基于归因图的网络嵌入方法，同时模拟同因、拓扑结构和节点特征。
Heterogeneous Network Embeddings：	
HEBE[13]	嵌入大规模异构事件网络
EOE[14]	用于耦合异构网络，两个同构网络间边缘连接

表 1

1.2.2 群组嵌入框架 ComE

传统处理方式中，图嵌入的任务主要聚焦在为图中的每个节点输出一个向量表示，这样图中相似的两个点在低维空间中就会拥有相似的向量表示，这种节点嵌入在保留网络结构上很成功，并且显著改进了大量应用的效果，包括节点分类[17,18]，节点聚簇[19,20]，连接预测[21,22]和图可视化[23,24]等等。

ComE 框架主要研究的是群组嵌入，群组嵌入探索的是群组在低维空间中的一种潜像表示。其不仅对如图可视化等群组层面的应用有帮助效果，而且对群组检测和节点分类也有益处。为了学习群组嵌入，ComE 将群组嵌入、群组检测和节点嵌入形成了一个闭循环（**图 1**）：节点嵌入通过输出可以更好帮助群组嵌入

的优质群组来协助改进群组检测效果，群组嵌入又可通过引入一个高阶群组意识的近似来优化节点嵌入。基于该闭循环，[1]中提出了 ComE 框架：一个同时解决群组嵌入，群组检测和节点嵌入三个任务的群组嵌入框架。

图 1：ComE 框架中节点嵌入、群组检测和群组嵌入闭循环

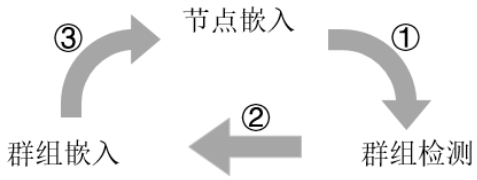


图 1

因为一个群组是一组联系紧密的节点的组合，所以群组嵌入可以特征显示出它的内部节点成员在低维空间中都是如何分布的。因此，群组嵌入不能简单地用一个向量来表示，而应该使用一个在低维空间中的分布来代替。因为每个群组都是一组联系紧密的节点的组合，该算法中使用了高斯混合模型[25]来将每个群组嵌入视为二维空间中的一个多元高斯分布。通过将群组嵌入定义为一个多元高斯分布，增强由节点嵌入结果得到群组检测的准确性。高斯混合模型有效地检测群组和从图中推断群组嵌入分布。已知群组任务和群组嵌入的情况下，该算法中扩展了 DeepWalk[3]和 LINE[4]的神经网络，来同时保留一阶、二阶和高阶（群组意识）高阶。群组中两个节点的联系可能是高阶的，该算法中认为群组嵌入是介绍了一个群组意识的高阶接近给节点嵌入，因此群组嵌入不需要两个节点有直接的联系或分享相同的“上下文”来整明其具有相似性。

使用经典数据集 Karate Club Graph 复现 ComE，证明 ComE 的在社交网络分析中的可应用性。结果如**图 2** 所示。Karate Club Graph 有 34 个节点，78 条边，已知该图中含有两类团体，一类由教练（节点 1）领导，一类由俱乐部管理者（节点 34）领导。其中有一些俱乐部成员被分到两类中的哪类都可以，对分类没有决定性作用。

图 2：ComE 框架在 Karate Club Graph 数据集上的实验结果

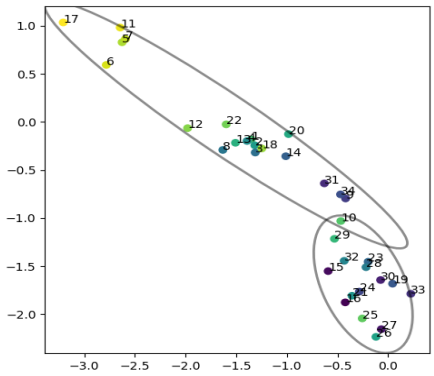


图 2

1.2.3 基于结构的网络嵌入算法 Struc2Vec

结构同一性是一种对称概念，该概念中网络节点是根据网络结构和其与其他节点之间的关系来定义的。[2]提出的 Struc2Vec 是一个学习具有结构同一性的节点的潜像表示的框架。

几乎在所有的网络中，节点一般有一个或更多帮助决定其在系统中的角色的功能，类似于每个人在一个社交网络中有不同的社会角色或社会位置[26,27]。本质上来说，不同节点可能具有相似功能，如公司中的实习生，生物化学反应中的催化剂等等，因此节点通常根据其这些功能被分到不同类别中，而这些功能的界定和识别通常使用的是节点和边的性质。但是当这些功能仅仅由网络结构定义而不是位置信息时，再使用节点和边的性质来学习节点的功能就失去了准确性。简言之，“邻居”相似的节点应该有相似的潜像表示，但是“邻居”是一个由网络中一些渐近概念定义的局部概念，因此当两个节点的“邻居”结构相似、距离遥远时，无法被分到一类。

为了更好学习节点的结构特征，Struc2Vec 有以下几个关键步骤：

- 不使用节点或边的信息来评估节点的相似性，与其邻居的位置或节点标记无关，且不需要网络必须是连接的。
- 建立一个层级来衡量不同规模的节点相似性，允许对结构相似概念的逐步严格，具体来说，在层级的最底层，节点之间的结构相似性仅取决于他们的度数，而在层级的最顶层，结构相似性取决于整个网络。
- 使用一个多层次图来编码结构相似性，并为节点生成结构性上下文，这些节点是通过遍历多层图（不是原始网络）的加权随机游走观察到的结构相似的节点序列。因此，经常出现在相似上下文中的两个节点可能具有相似结构。

Stru2Vec 能够更好地捕捉结构同一性来寻找节点潜像表示，因此 Struc2Vec 在更多依赖于结构同一性的分类任务上表现更好。

1.3 本文贡献总结

社交网络算法目前已经在如信贷反欺诈、保险反欺诈中得到了良好应用，类似的，企业之间的联系也可形成一个网络，如一个高管同时任职于不同公司，因此将社交网络算法应用到企业网络上进行群组检测和节点分类，可帮助检测出特殊性质企业群体，如有非正常联系的企业、家族企业等等。

本文使用群组嵌入框架 ComE 对企业网络进行企业分析--群组检测和节点分类，ComE 利用群组嵌入、群组检测和节点分类闭合循环来帮助增进群组检测和节点分类效果。但 ComE 应用过程中存在的问题是，企业关联网络具有较强的结

构同一性，直接使用 ComE 进行群组检测和节点分类会忽略相距较远但是结构类似的节点的相似性，无法将这类相似节点分到同一类别中。基于此问题，本文提出了一个新的企业群组分析框架 Struc_ComE—基于结构的群组嵌入框架，Struc_ComE 将 ComE 框架中由 DeepWalk 得到的路径文件替换为由 Struc2Vec 得到，Struc_ComE 进一步考虑了节点的结构特征独立于局部性、点和边性质，详细分析不同节点之间的结构相似性，寻找结构形状类似的节点分到一类。Struc_ComE 在结构同一性较强的企业网络上进行企业分析----群组检测和节点分类任务时，改进效果明显。

本文贡献概括来讲有如下几点：

1) 根据 ComE 使用 DeepWalk 得到路径文件用来后续进行节点嵌入、群组检测和群组嵌入学习的思想，基于 Struc2Vec 有利于改进结构同一性网络的路径文件的启发，本文设计了基于结构的群组嵌入方法 Struc_ComE 框架：将 ComE 中使用 DeepWalk 得到的路径替换为由 Struc2Vec 得到的路径文件，进一步考虑节点结构特征，减少对位置信息的偏重，改进企业网络的群组检测和节点分类效果。

2) 将 Struc_ComE 框架在多种真实数据集上进行实验，与 ComE 框架实验结果进行对比分析，结果表明 Struc_ComE 框架在具有结构同一性的企业网络上具有更好的群组检测和节点分类效果。

3) 在证明 Struc_ComE 框架比 ComE 框架能更好地识别结构特征之后，本文进一步研究社交网络分析算法在企业分析上的应用：使用 ComE 框架进一步对不同地区的企业网络进行实验分析，发现在不同地区企业分析的差异性，如广东、上海、江苏地区；使用 ComE 框架进行参数步长和步数的研究，发现同时适当增大步长和步数效果更好。

第二章 企业群组分析框架

本章主要介绍本文提出企业群组分析框架：基于结构的群组嵌入框架 Struc_ComE 的原因、过程及改进效果。Struc_ComE 是一种基于结构的群组嵌入框架，可更好地处理具有结构同一性的企业网络。Struc_ComE 的提出过程来自对群组嵌入框架 ComE 和网络嵌入算法 Struc2Vec 的研究启发：ComE 可以通过节点嵌入、群组嵌入和群组检测三者的闭循环有效进行群组嵌入学习，但是 ComE 中的路径文件由 DeepWalk 得到，DeepWalk 无法有效利用节点的结构特征；与 DeepWalk 相对应，Struc2Vec 是基于节点结构的网嵌入算法，可有效处理节点的结构特征。因此结合 ComE 和 Struc2Vec 二者的优点可有效改进对结构同一性网络的学习分析效果：使用 Struc2Vec 替换 ComE 中的 DeepWalk 算法得到新的路径文件，更好地用于节点嵌入、群组检测和群组嵌入的闭循环。

2.1 ComE 框架--群组嵌入框架

正如上文 1.2.2 中所讲，1) ComE 中良好利用节点嵌入、群组检测和群组嵌入的互相促进关系，将三者形成一个闭合循环，如**图 1**，帮助增进群组检测和节点分类效果。ComE 中还使用一个低维空间的分布来表示群组嵌入，而不是简单的一个向量；2) ComE 承认节点之间的高阶接近，不仅仅是一阶和二阶，因此不需要两个节点有直接的联系或分享相同的“上下文”来证明其具有相似性，距离较远的情况下，只要有相似的结构也可被分为一类；3) ComE 中使用了高斯混合模型[25]来将每个群组嵌入视为二维空间中的一个多元高斯分布。通过将群组嵌入定义为一个多元高斯分布，增强由节点嵌入结果得到群组检测的准确性。

ComE 的主要工作流程为：

1) 在输入的图文件上使用 DeepWalk 得到路径文件；
2) 使用上述路径文件得到节点嵌入和内容嵌入；
3) 控制该节点嵌入和内容嵌入，进行群组检测和群组嵌入最优化处理；
4) 控制上述得到的最后一个群组检测和群组嵌入，进行节点嵌入最优化处理。

使用本文创建的模拟企业网络数据集 1（如**图 3**）进行测试 ComE 结果如**表 2** 所示，准确率为 100%。其中，1 和 2（或 11 和 12）属于同一公司，但是其又同时（或不同时）属于不同的公司，这样二者在不同的公司中又会和不同的高管成为同事，即有关联性，如 1、2 和 3，具体表示如**图 3** 所示（**图 3** 中使用连线来表示节点之间有关联）。在**图 3** 所示的例子中，可以发现 1 和 2 与 11 和 12 都有导致公司有非正常关联的可疑性：因 1 和 2（或 11 和 12）互相之间有关联，

且又同时与其他不同高管在不同公司中有关联，而这些高管之间却没有这样复杂的联系，因此 1 和 2 以及 11 和 12 为特殊高管群体。

图 3：本文创建的模拟企业网络 1

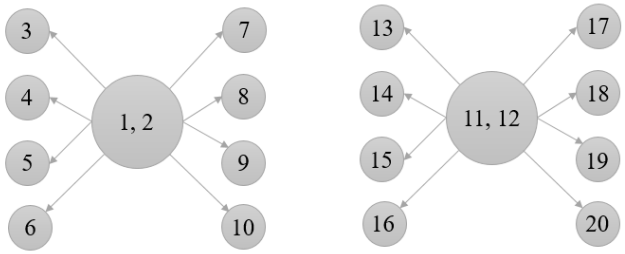


图 3

表 2：ComE 在本文创建的模拟的企业网络 1 上的实验结果

类别	样本
类别 1	1~2，11~12
类别 2	3~10，13~20

表 2

再使用本文创建的模拟企业网络数据集 2（如**图 4**），结果如**表 3** 所示，却呈现一般情况，正确率降为 50%，可见 ComE 在企业网络上无法有效捕捉结构同一性，导致当样本量增大时，正确率大幅降低。

2.2 Struc2Vec 算法

目前大多数节点嵌入算法，如 DeepWalk[3]、LINE[4]和 Node2Vec[5]等都在分析节点和边信息上赋予了过多权重，而忽略节点结构特征，无法有效地捕捉结构同一性。但是这些算法在目前的一些应用中却呈现良好效果，原因为：在大多数现实网络中，许多节点特征显示很强的同质性，即属于一类的节点确实就有相似的邻居。但是对于具有结构同一性的网络，这些方法就无法准确捕捉到潜像表示和节点分类情况。企业关联网络就是具有结构同一性的网络，且当样本范围和数量增加，或该地区家族企业数目较多、区别不明显时（如浙江等地），结构同一性质会大幅增加，此时再使用 DeepWalk[3]或 Node2Vec[5]等算法，分类准确性就会大幅下降。因此引入 Struc2Vec 框架，Struc2Vec 可捕捉网络的结构同一性，更侧重考虑节点的结构特征，而不是位置信息。Struc2Vec 在不借鉴节点位置信息的情况下考虑节点结构的相似性、建立层次来衡量结构相似性、为节点生成随机内容，以此来评定节点的功能性质，用于后续分类。使用 Struc2Vec 算法替换 ComE 中的 DeepWalk 得到路径文件可改进群组检测和节点分类效果。

具体步骤为：

1) 确定图中每个顶点对之间的结构相似性，其周围邻域大小不一定相同；
2) 构造一个加权多层图，其中网络中的所有节点都存在于每一层，并且每一层对应于衡量结构相似性的层次中的一个水平；
3) 使用 2) 中的多层次图为每一个节点生成上下文；
4) 从节点序列给定的上下文中学习潜像表示

使用本文手动创建的模拟企业网络数据集 2（如**图 4**）验证使用 Struc2Vec 之后的 Struc_ComE 对结构同一性更好的捕捉能力，图 4 中包含了类似**图 3** 中的 6 个团体，分类结果如**表 3** 所示，相比 ComE 分类正确率为 50%，Struc_ComE 的分类准确率仍为 100%（将 1, 2, 11, 12, 21, 22, 31, 32, 41, 42, 51, 52 分在一类，其他分在一类），可见 Struc_ComE 框架对结构同一性的良好捕捉。

图 4：本文创建的模拟企业网络 2

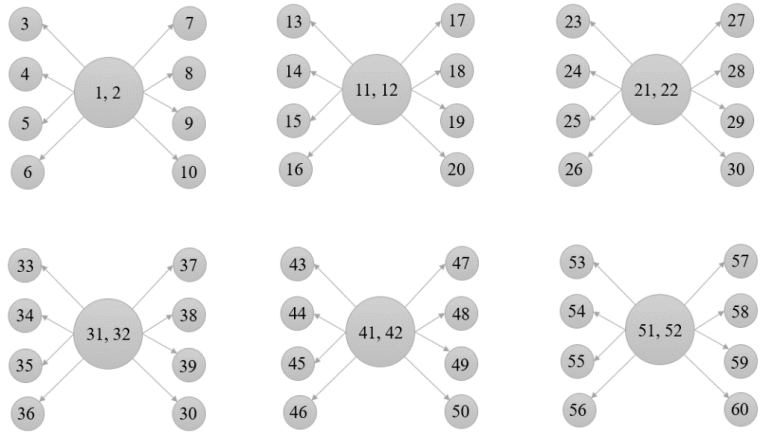


图 4

表 3：Struc_ComE 和 ComE 在本文创建的模拟的企业网络 2 上的实验结果

类别	ComE	Struc_ComE
类别 1	1~10, 11~20, 41~50	1~2, 11~12, 21~22, 31~32, 41~42, 51~52
类别 2	21~30, 31~40, 51~60	3~10, 13~ 20, 23~30, 33~40, 43~ 50, 53~ 60
正确率	**50%**	**100%**

表 3

2.3 基于结构的群组嵌入方法 Struc_ComE 框架

由 2.1 和 2.2 节可知，群组嵌入方法 ComE 可同时完成节点嵌入、群体检测和节点分类任务，但其无法有效捕捉结构同一性，在节点嵌入步骤中没有深入研究节点结构特征，当输入网络为具有结构同一性的企业关联网络时，群组检测和

节点分类准确率会大幅下降。Struc2Vec 恰好是一种考虑节点结构特征的框架，在具有结构同一性的网络中，或想根据节点结构特征类似进行节点类别划分的任务中，可以表现良好。

为了在企业网络上得到更好地捕捉结构同一性，通过研究节点结构特征来对节点进行更好的分类和群组检测，本文提出企业群组分析框架 Struc_ComE----基于结构的群组嵌入方法设计：使用 Struc2Vec 替换 ComE 中的 DeepWalk，将结构特征考虑到 ComE 进行群组检测和节点分类的任务中，提高分类效果和准确率。

具体 Struc_ComE 的主要工作流程为：

1) 在输入的图文件上使用 Struc2Vec 框架，得到路径文件；
2) 使用上述路径文件得到节点嵌入和内容嵌入；
3) 控制该节点嵌入和内容嵌入，进行群组检测和群组嵌入最优化处理；
4) 控制上述得到的最优群组检测和群组嵌入，进行节点嵌入最优化处理。

与 ComE 中使用 DeepWalk 不同，Struc_ComE 使用 Struc2Vec 生成路径文件，这样保证在整个框架最开始阶段、进行节点嵌入、群组检测和群组嵌入三个任务之前，就已将节点结构特征考虑进来，确保后面任务完成的过程中，都是在分析过节点结构特征的基础上进行的，有效提高群组检测和节点分类任务中对节点结构特征的捕捉效果，和分析权重。使得在应用到具有结构同一性的企业网络中时，群组检测和节点分类效果大大提高，分类更清晰，减少正常企业被分为有非正常关联企业的类别中。

使用经典数据集 Karate-Mirrored Network 进行 ComE 和 Struc_ComE 的对比分析，结果如**图 5** 和**图 6** 所示，Struc_ComE 更能有效地根据结构特征将镜面节点对挑选出来----镜面节点之间的距离较小。本文通过计算，得到如图所示的结果，ComE 的镜面节点对总距离为 252.42，Struc_ComE 的镜面节点对总距离为 3.69，比 ComE 改进了 67.5 倍，可见其对结构特征的有效识别。

图 5：ComE 和 Struc_ComE 在 Karate-Mirrored 数据集上的结果(左 ComE 右 Struck_ComE)

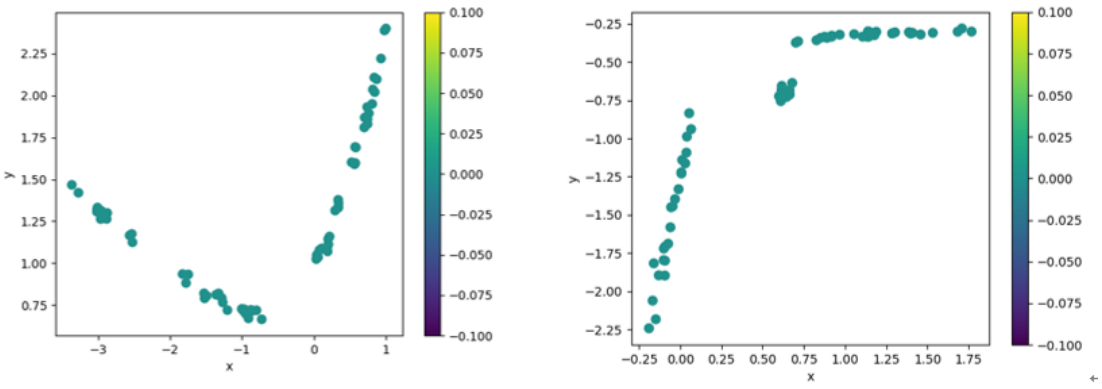


图 5

图 6：ComE 和 Struc_ComE 在 Karate-Mirrored 数据集节点对距离情况

ComE							Struc_ComE						
ID	x	y	ID	x	y	距离	ID	x	y	ID	x	y	距离
1	0.1884	1.0967	37	-0.9077	0.6716	1.3821	1	0.7019	-0.3736	37	0.7156	-0.3649	0.0003
2	0.3393	1.3777	39	-1.3226	0.8201	3.0727	2	0.8909	-0.3410	39	0.8531	-0.3421	0.0014
3	0.5254	1.6014	38	-1.7445	0.9333	5.5990	3	0.8902	-0.3410	38	0.8764	-0.3354	0.0002
4	0.2965	1.3140	59	-1.2802	0.7935	2.7568	4	1.0557	-0.3181	59	1.1899	-0.3027	0.0182
5	0.1086	1.0882	63	-0.9495	0.7005	1.2699	5	0.6416	-0.7158	63	0.6153	-0.6836	0.0017
6	0.0766	1.0771	50	-0.8576	0.7155	1.0036	6	1.3880	-0.3052	50	1.1783	-0.3248	0.0443
7	0.0406	1.0488	55	-0.9151	0.7072	1.0301	7	1.4587	-0.3201	55	1.3036	-0.3061	0.0243
8	0.3367	1.3292	43	-1.2653	0.7666	2.8826	8	1.1072	-0.3340	43	1.1401	-0.2994	0.0023
9	0.5743	1.6921	41	-1.4892	0.8060	5.0436	9	1.2870	-0.3129	41	1.1294	-0.3332	0.0252
10	0.5773	1.5968	58	-2.5603	1.1659	10.0300	10	-0.0990	-1.7208	58	-0.0550	-1.4506	0.0749
11	0.0596	1.0358	53	-0.8735	0.7222	0.9690	11	0.6638	-0.6845	53	0.6799	-0.6373	0.0025
12	0.1923	1.0705	67	-0.7238	0.6659	1.0028	12	-0.0095	-1.3325	67	0.0075	-1.2220	0.0125
13	0.2029	1.1126	62	-0.9778	0.7174	1.5504	13	-0.0590	-1.5819	62	0.0666	-0.9397	0.4282
14	0.3451	1.3589	60	-1.3575	0.8126	3.1972	14	1.2868	-0.3129	60	1.1479	-0.3033	0.0194
15	0.7658	1.8932	64	-2.9618	1.2631	14.2919	15	-0.1891	-2.2426	64	-0.0921	-1.8004	0.2050
16	0.8749	2.0959	66	-3.0066	1.3320	15.6494	16	-0.1295	-1.8962	66	-0.0726	-1.6893	0.0460
17	0.0349	1.0246	52	-0.7967	0.7198	0.7844	17	-0.0909	-1.7076	52	0.0382	-1.0932	0.3942
18	0.2195	1.1592	40	-1.0019	0.7278	1.6778	18	-0.0310	-1.3973	40	0.0549	-0.8352	0.3233
19	0.8122	1.9489	61	-3.0127	1.3203	15.0245	19	-0.0997	-1.7966	61	0.0122	-1.1412	0.4421
20	0.3440	1.3361	54	-1.2007	0.7209	2.7645	20	0.6166	-0.6567	54	0.5997	-0.7242	0.0048
21	0.8471	2.0190	46	-3.0146	1.3067	15.4202	21	-0.0927	-1.8975	46	-0.1475	-2.1835	0.0848
22	0.1996	1.1435	49	-0.9676	0.7261	1.5364	22	-0.0454	-1.4448	49	0.0411	-0.9883	0.2159
23	0.8213	2.0348	47	-2.8793	1.2637	14.2887	23	-0.1669	-2.0618	47	0.0077	-1.2300	0.7222
24	0.9341	2.2191	35	-3.3636	1.4667	19.0364	24	1.7093	-0.2807	35	1.4098	-0.3093	0.0905
25	0.7476	1.8292	44	-2.9511	1.3092	13.9502	25	0.6546	-0.7171	44	0.6461	-0.7285	0.0002
26	0.7468	1.8573	57	-2.9598	1.3141	14.0340	26	0.6647	-0.7127	57	0.6551	-0.7089	0.0001
27	0.7019	1.8084	68	-2.8717	1.2994	13.0292	27	-0.1598	-1.8176	68	0.0307	-1.1634	0.4643
28	0.7423	1.9303	45	-1.7759	0.8823	7.4398	28	1.5327	-0.3092	45	1.3987	-0.3146	0.0180
29	0.5661	1.5893	56	-1.5257	0.8210	4.9662	29	0.6407	-0.7239	56	0.6533	-0.6907	0.0013
30	0.8410	2.1058	36	-3.2715	1.4198	17.3830	30	1.6837	-0.3017	36	1.7679	-0.3011	0.0071
31	0.5857	1.6897	65	-1.5175	0.7911	5.2309	31	0.6107	-0.7560	65	0.6200	-0.7362	0.0005
32	0.7076	1.8676	48	-1.8236	0.9355	7.2757	32	1.1426	-0.3348	48	1.1415	-0.3146	0.0004
33	0.9830	2.3863	51	-2.5208	1.1246	13.8689	33	0.9679	-0.3214	51	0.9201	-0.3284	0.0023
34	1.0037	2.3980	42	-2.5296	1.1748	13.9803	34	0.8271	-0.3577	42	0.9186	-0.3345	0.0089
					总距离	252.4221						总距离	3.6876
												改进倍数	67.4519

图 6

潜像表示的应用之一就是网络节点分类，Struc_ComE 可以更多利用结构特征来完成这个任务，为了证明 Struc_ComE 的这种功能，本文使用[2]中巴西和欧洲机场数据集进行 ComE 和 Struc_ComE 分类结果对比分析，如**表 4** 和**表 5** 所示。该机场交通网络为节点表示机场、边表示航班的无方向连通图。机场被根据其活动水平进行标签标记，活动水平由航班和乘客进行衡量。

- 巴西机场交通网络：从国家民航管理局（ANAC）收集的从 2016 年 1 月至 12 月的数据，网络中有 131 个节点，1038 条边。机场活跃性由相应时间段里起飞和降落的总数衡量。
- 欧洲机场交通网络：从欧洲联盟（欧统局）收集的从 2016 年 1 月至 11 月的数据，网络中有 399 个节点，5995 条边。机场活跃性由相应时间段里起飞和降落的总数衡量。

每个机场中，有四个标签用来表示机场活跃性，本文使用分类准确率来评估 ComE 和 Struc_ComE 的表现，使用 90%节点进行训练，使用 10%节点进行测试，结果如**表 4、5** 所示。显而易见，Struc_ComE 效果比 ComE 好。

表 4：ComE 和 Struc_ComE 对巴西和欧洲机场数据集的分类准确率结果

分类准确率	ComE	Struc_Come
Brazil	48.5%	85.7%
europe	38.1%	57.2%

表 4

表 5：ComE 和 Struc_ComE 对巴西和欧洲机场数据集的分类准确率结果

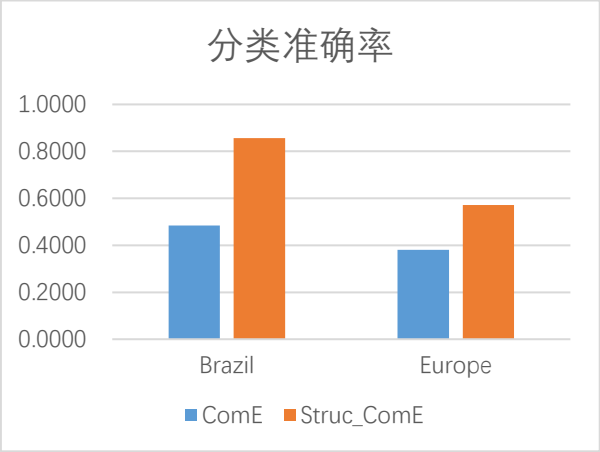


表 5

2.4 小结

本文使用群组嵌入网络学习方法 ComE 对企业网络进行群组检测和节点分类任务，但企业网络具有结构同一性，而 ComE 中使用 DeepWalk 得到的路径文件侧重使用节点的位置信息，而忽略节点的结构特征，无法从结构特征分析的角度对网络进行学习、根据结构相似性得到群组检测和节点分类，导致分类结果还有很大提升空间。考虑到 Struc2Vec 框架是一种使用节点结构特征进行节点嵌入任务的框架，其可以很好地捕捉结构同一性，将结构相似的节点划分到具有相似性质的一类中。因此，本文进行了基于结构的群组嵌入设计----提出 Struc_ComE 框架：通过使用 Struc2Vec 替换 ComE 中的 DeepWalk，使 ComE 中的路径文件由 Struc2Vec 生成，在框架运行的第一步就考虑结构特征，将其作为接下来所有要处理的任务：节点嵌入、群组嵌入和群组检测任务的基础条件，保证接下来的每一步都是在考虑结构特征的前提条件下进行的，提高在企业网络上群组检测和群组嵌入的准确性。

第三章 实验工作

本章主要陈述实验的流程和相关的数据处理，数据处理中包括节点为高管的网络和节点为公司的网络。

3.1 实验流程

1) 数据处理初始文件得到可作为输入的图存储文件；
2) 使用 ComE 或 Struc_ComE 对企业网络进行企业分析；
3) 使用 Mean Shift 算法将结果可视化；
4) 查阅资料检验筛选出的公司是否确实具有潜在不正常关联嫌疑。

3.2 数据处理

3.2.1 节点为高管的网络

数据为从国泰安数据库[1]上获取的 2018 年公司董监高个人特征的数据，其中包括公司的证券代码、董监高 ID 和名字。

数据处理分为以下几步：

1) 删除数据中的重复条目；
2) 根据董监高 ID 重新给董监高由 1 到 N 编号；
3) 然后将所有董监高任职信息按照公司证券代码从小到大的顺序排序，排序之后将属于同一个公司的董监高分到一个数据条目里（一行）；
4) 遍历 3)得到的数据文件，属于同一个数据条目中的高管为在同一公司任职的，因此彼此之间有关联，为其两两加入一条边来表示关联性。这样形成一个图的邻接表，存入.adjlist 文件中，读入该文件就可以得到一个图的结构。

3.2.2 节点为公司的网络

初始数据为从国泰安数据库[2]上获取 2018 年全部 A 股上市公司董监高个人特征的数据，其中包括公司的证券代码、董监高 ID 和名字。

[1] http://cn.gtadata.com/
[2] http://cn.gtadata.com/

数据处理分为以下几步：

1) 删除数据中的重复条目，删除董监高中文名字一列→文件①；
2) 将数据文件①中的数据条目根据公司证券代码重新给公司由 1 到 N 编号→文件②；
3) 将数据文件②中的数据条目按照董监高 ID 从小到大的顺序进行排序，排序之后将同一个董监高参与的所有公司分到一个数据条目里（一行），得到新的数据文件→文件③；
4) 删除数据文件③中长度为 1 的数据条目，这些条目表示一个高管只参与了一个公司的正常现象，其所参与的公司不和其他公司有联系。而长度大于 1 的数据条目，表示在一个高管同时参与不同的公司，本文认为这种情况下，这些公司我们认为其互相之间有联系，为公司网络中的边→文件④；
5) 遍历数据文件④，为属于同一个数据条目中的节点两两之间加入关联边，形成表示图结构的邻接表，存入.adjlist 文件中→文件⑤。

第四章 Struc_ComE 结果对比分析

本章进行 Struc_ComE 和 ComE 实验结果的对比分析，分别使用节点为高管的网络和节点为公司的节点进行实验。在节点为高管的网络中选取的实验地区为：浙江地区、江浙沪地区以及东北三省地区，选取浙江地区因浙江地区的家族企业比较多，结构同一性比较强，有利于分析 Struc_ComE 框架和 ComE 框架对结构特征信息的处理效果。江浙沪地区和东北三省地区均为地区范围扩大，样本数量增加，结构同一性增加，也有利于分析 Struc_ComE 框架和 ComE 框架对结构特征信息的处理效果。

4.1 节点为高管的网络结果分析

本节为使用节点为高管的企业信息网络进行企业群组分析，实验地区有浙江地区、江浙沪地区以及东北三省地区。在使用高管作为节点的网络中，为了得到特殊企业群组，本文在得到特殊高管群体之后进行了高管 ID 与公司证券代码的匹配来得到特殊企业群组。从三个地区的实验结果中可看出，当结构同一性增强时，基于结构的群组嵌入方法 Struc_ComE 框架比群组嵌入方法 ComE 有更好的群组检测效果。

4.1.1 浙江地区

浙江地区民营企业较多，而且民营企业中家族企业占八九成，比例超过全国平均水平，这样的特征导致浙江地区的企业高管网络结构同一性进一步加强。如**表 6** 所示，基于结构的群组嵌入框架 Struc_ComE 与群组嵌入框架 ComE 实验结果对比明显，群组分类更清晰，聚簇效果更明显。并且可发现步长对于 Struc_ComE 框架结果的影响大于对 ComE 框架结果的影响。得到的非正常关联企业案例如**表 7** 所示。

表 6：ComE 和 Struc_ComE 对浙江地区实验结果对比

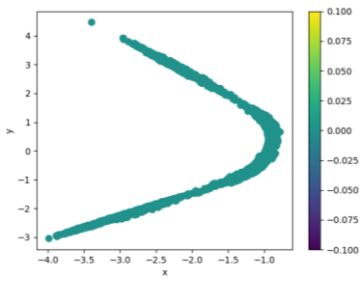
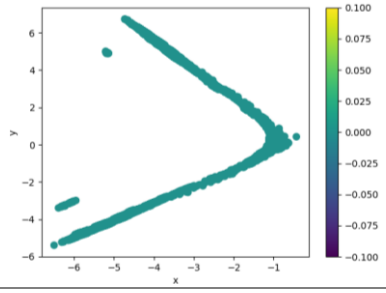
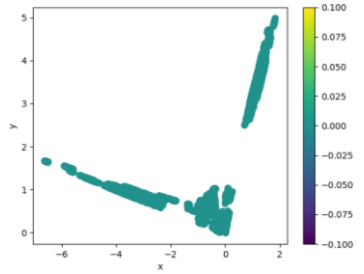
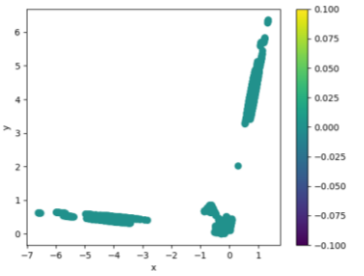
	步数：10 步长：20	步数：10 步长：30
ComE		
Struc_ComE		

表 6

表 7：浙江地区存在非正常关联的企业案例

公司代码	公司名称	备注
603877	太平鸟	家族企业：张江平家族
603789	星光农机	董事长持有绝大股份
603701	德宏股份	家族企业：张元园家族
603611	诺力股份	家族企业：丁氏家族
603520	司太立	家族企业：胡氏家族

表 7

4.1.2 江浙沪地区

江浙沪地区：范围扩大，样本数量增加，这意味着在距离更远、但具有相似结构特征的节点数量增加，结构同一性进一步增加，使用 ComE 得到遍历到的图文件会增加分类和勘测的不准确性。具体结果如**表 8** 所示。

由**表 8** 可知，使用 Struc_ComE 进行实验之后，可视化结果中显示的分类情况更明显。特殊公司例子如**表 9** 所示。

表 8：ComE 和 Struc_ComE 对江浙沪地区实验结果对比

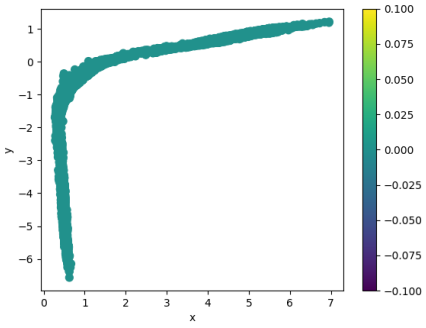
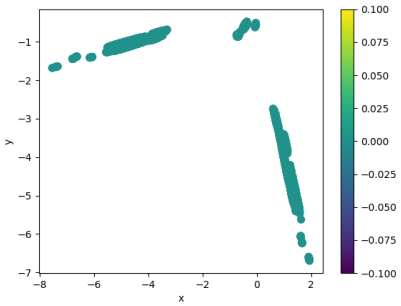
	ComE	Struc_ComE
可视化结果		

表 8

表 9：江浙沪地区存在非正常关联的企业案例

公司代码	公司名称	备注
600366	宁波韵升	家族企业：竺韵德家族
603088	宁波精达	家族企业：郑朗才家族
002570	贝因美	家族企业：谢宏家族，或引入国资自救
600628	新世界	家族企业：郑裕彤
603009	北特科技	高管违规持股

表 9

4.1.3 东北三省地区

东北三省地区同样存在：范围变大，样本数量增加，网络结构同一性增加的的特点，使用 Struc_ComE 框架进行实验之后可发现分类效果变好，结果如**表 10** 所示。

具体公司案例如**表 11** 所示。由**表 11** 所知，该案例中出现*ST 类型企业，ST 与*ST 标记均有“特别处理”含义，该标记针对的对象是出现财务状况或其他状况异常的公司。在股票前加上 ST 标记，是给市场一个警示，表示该股票存在投资风险，起到警告作用，股票投资风险大收益也大；*ST 标记是表明该公司连续三年亏损，有退市风险，给市场一个警惕提醒。因此此类公司也是具有非正常关联嫌疑，需要提醒消费者谨慎投资的企业。

表 10：ComE 和 Struc_ComE 对东北三省地区实验结果对比

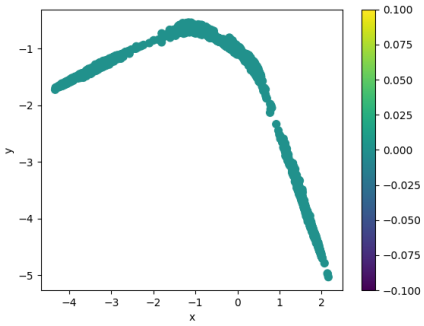
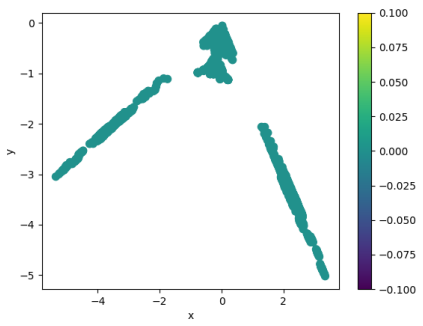
	ComE	Struc_ComE
可视化结果		

表 10

表 11：东北三省地区存在非正常关联的企业案例

公司代码	公司名称	备注
002447	晨鑫科技	家族企业：刘晓庆、刘永辉大额套现
600891	*ST 秋林	ST*标记，证监会用来提醒投资者的特殊企业
002698	博实股份	高管大幅增持
603609	禾丰牧业	家族企业：金氏家族
000587	金洲慈航	重大非正常资产重组解决债务问题

表 11

4. 2 节点为公司的网络结果分析

该部分实验中，将全部 A 股上市不同公司作为图中的节点，认为同一高管参与的不同公司之间存在关联，用边来表示该关联性。在由此构成的图上运行社交网络分析算法，直接得到有嫌疑具有特殊性质的企业公司群体。

实验结果如**表 12** 所示。

表 12：ComE 和 Struc_ComE 在节点为公司的全部 A 股公司网络上的实验结果对比

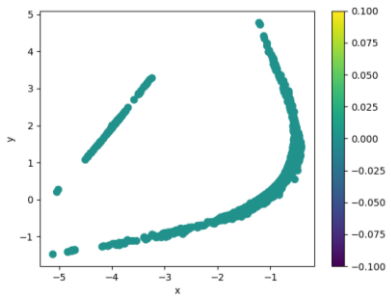
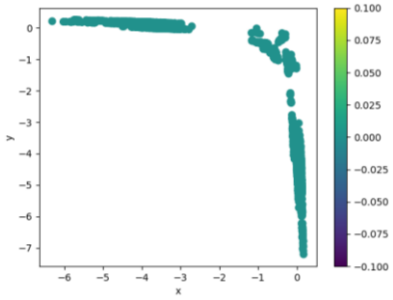
	ComE	Struc_ComE
可视化结果		

表 12

由**表 12** 中可视化结果可知，在 ComE 框架情况下，以公司为节点的网络可以被分出一些具有非利好投资标的性质嫌疑的公司群体，但使用 Struc_ComE 框架改进之后，分类结果比较明显，增加了一些在 ComE 中被遗漏的具有可疑非正常关联的企业。

具体公司实例如**表 13** 所示。

表 13：节点为公司的全部 A 股公司网络存在非正常关联的企业案例

公司代码	公司名称	备注
300417	南华仪器	董事长薪酬过高，利润大幅下滑
300504	天邑股份	家族企业：李氏家族
300629	新劲刚	家族企业：王刚家族
300314	戴维医疗	家族企业：陈再宏家族
300352	北信源	家族企业：林皓家族

表 13

第五章 ComE 结果分析

第四章中已分别使用节点为高管和节点为公司的企业信息网络的实验结果详细介绍了在具有结构同一性的网络上，本文提出的基于结构的群组嵌入框架 Struc_ComE 相比群组嵌入框架 ComE 的改进效果，本章将进一步研究社交网络分析算法在企业分析上的应用。本章继续使用群组嵌入框架 ComE 对不同地区企业和相关参数（步长和步数）进行研究，发现不同地区的企业特征有所差异，且在不同的步长和步数参数下企业群组检测效果不尽相同。

5.1 地区研究

本节首先对不同地区的企业具有不同的企业特征进行研究。

5.1.1 广东地区上市公司

使用上述方法对广州地区的上市公司进行检测之后，可视化结果如**图 7** 所示。从**图 7** 中可以发现，公司大体分成四个种类，其中坐标在(-5, -3)左右的群体数目最小，为我们想找的特殊群体。将代码中类别 k 设置为 k=4，再次运行，得到高管特殊群体，然后将其与公司证券代码匹配得到可疑公司的证券代码。

图 7：ComE 在广东地区的实验结果

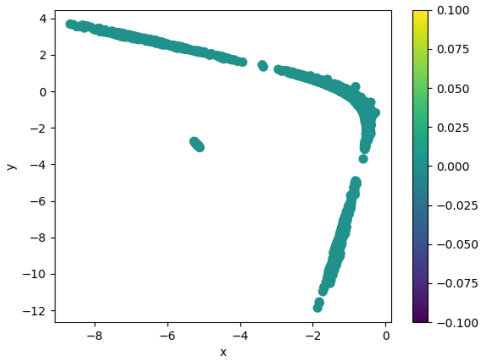


图 7

将可疑公司按照其证券代码排序整理，选取一些公司代码进行查询之后，发现一些公司确实为家族企业（或为有家族企业风的国企）或曾有不良记录。如 002121 科陆电子，该公司在 2018 年预亏最高为 11 亿；002774 快意电梯，为罗氏姐弟控制的家族企业，在 2018 年上市前陷入被财务部长敲诈；等等，见**表 14**。由此可见，该算法在公司性质分析、群组检测方面的可实用性。

表 14：广东地区存在非正常关联的企业案例

公司代码	公司名称	备注
002774	快意电梯	家族企业：罗氏姐弟，原财务总监蓄谋勒索董事长
002210	飞马国际	家族企业：黄壮勉家族
002121	科陆电子	2018 年预亏最高为 11 亿
002889	东方嘉盛	家族企业：孙卫平夫妇
300679	电连技术	陈氏家族一股独大

表 14

5.1.2 江苏地区

对江苏地区的上市公司的检测的可视化结果如**图 8** 所示，将其中的特殊小群体提取出来之后，匹配得到可疑公司代码，查询检验之后确为特殊公司的例子如**表 15** 所示。

图 8：ComE 在江苏地区的实验结果

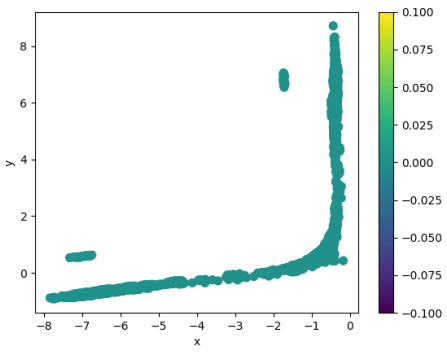


图 8

如**表 15** 所示，该五家均为企业中的高管互相之间有亲戚关系，一个家庭在公司中有较大的控制权的家族企业（这里的家族企业不仅包括传承几代的企业，也包括公司内部由一个家庭控权的企业）。

表 15：江苏地区存在非正常关联的企业案例

公司代码	公司名称	备注
002223	鱼跃医疗	家族企业：陈氏家族
002165	红宝丽	家族企业：芮氏家族
002778	高科石化	家族企业：许氏家族
603118	亚邦股份	家族企业：许氏家族
603016	新宏泰	家族企业：赵汉新家族

表 15

5.1.3 上海地区

对上海地区的上市公司的检测的可视化结果如**图 9** 所示，将其中的特殊小群体（(-5，-2.5）坐标附近）提取出来之后，匹配得到可疑公司代码，查询检验之后确为特殊公司的例子如**表 16** 所示。

如表 17 所示，上海地区的五家公司例子与江苏和广东均有所不同，其出现了国资委大量持股的企业：上工申贝。国资委大量持股的这些企业往往存在市场化不够、其余股东话语权较少或没有资格与第一大股东讨论甚至对话以及公司不具有持久竞争力和活力的问题，这样不利于企业永续发展，获得的成功只能是短期的，因此也是投资者在进行投资时需谨慎考虑的一类公司。

图 9：ComE 在上海地区的实验结果

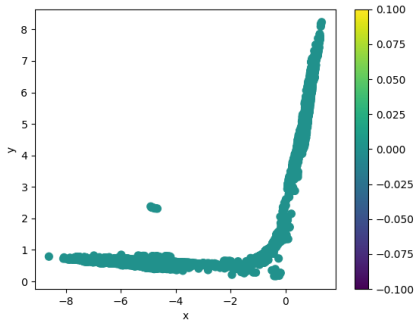


图 9

表 16：上海地区存在非正常关联的企业案例

公司代码	公司名称	备注
603157	拉夏贝尔	业绩不济，逐渐发展为家族企业
600843	上工申贝	浦东国资委大量持股
300017	网宿科技	高管内部关系复杂
300230	永利股份	高管内部关系复杂
300129	泰胜风能	高管内部关系复杂

表 16

5.2 参数研究

本节对步长和步数对 ComE 的群组检测效果进行分析研究，发现同时适当增加步数和步长有利于得到更好的结果。

表 17：浙江地区在不同步长和步数参数值下的分类情况

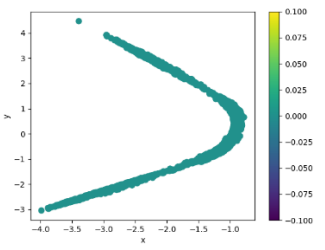
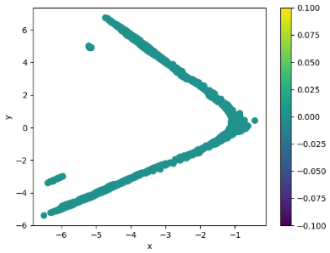
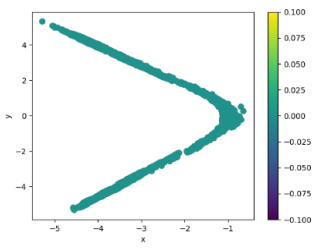
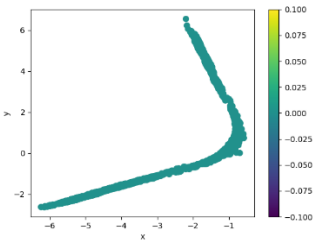
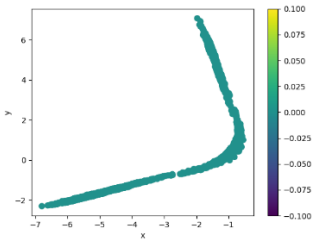
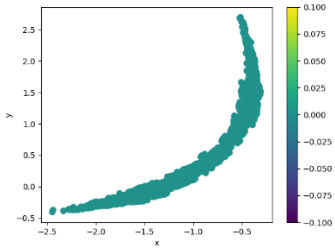
 **步数：10 步长：20**	 **步数：10 步长：30**	 **步数：10 步长：50**
 **步数: 20 步长: 20**	 **步数: 50 步长: 10**	 **步数: 20 步长: 10**

表 17

本文选取浙江地区进行参数研究，动态调整参数：步数和步长，将结果进行对比研究，如**表 17** 所示。

由**表 17** 可知，步数相同时，步长越长，图形的形状越精细，面积范围不会发生大幅变化；步长相同时，步数越多，节点越分散开来，面积范围变大。因此应适当同时增加步长和步数以达到更好效果。

第六章 结论与未来改进

6.1 结论

经过上述实验和分析，可发现使用社交网络分析方法可以对企业网络进行企业分析，如有非正常关联的企业等等，用来给投资者提供参照作用。对于企业网络具有较强的结构同一性的问题，本文启发于群组嵌入框架 ComE 和基于结构的网络嵌入算法 Struc2Vec，使用 Struc2Vec 替换 ComE 中得到路径文件的 DeepWalk 算法，提出企业群组分析框架：Struc_ComE----基于结构的群组嵌入框架。Struc_ComE 框架仔细考虑了节点的结构特征，减轻节点距离较远却拥有相似的结构特征的情况所带来的群组检测结果不准确的问题。通过 Struc_ComE 框架与 ComE 框架在多个真实数据集上的实验结果对比，表明 Struc_ComE 框架使群组检测和节点分类效果更清晰。其次，在证明 Struc_ComE 框架比 ComE 框架能更好地识别结构特征之后，本文进一步研究社交网络分析算法在企业分析上的应用：使用 ComE 框架进一步对不同地区的企业网络进行实验分析，发现在不同地区企业分析的差异性，如广东、上海、江苏地区；使用 ComE 框架进行参数步长和步数的研究，发现同时适当增大步长和步数效果更好。

6.2 未来改进

未来可进一步改进的地方有：

1) 目前在使用 Struc_ComE 框架时，时间代价比较大，运行全部 A 股的数据时耗费时间较长，后续可进一步改进，增加框架时间性能。
2) 对于参数值可进一步进行敏感性分析，找到每种情况下最佳状态值。
3) 进一步结合企业网络性质进行改进。

第七章 致谢

首先诚挚感谢熊赟老师在实验过程和论文撰写阶段的对我的指导、督促和鼓励。在论文的选题、研究、实验以及论文攥写过程中，我与熊赟老师时时保持进度的最新状态，老师对我进行进度督促、问题解答和心态鼓励。即使在出差路上，老师仍然及时与我进行讨论；老师在身体不舒服时也依然与我进行问题的沟通；当我完成阶段性进步时，老师给予我很大的鼓励，给了我继续坚持对下一问题深入研究的动力；在论文攥写阶段，老师对我的每一版都认真阅读、手把手指导，使我对如何书写论文有了更清晰的理解，而这些知识也将对我未来的学习中继续有所帮助。再次感谢熊赟老师在我整个毕业设计完成过程中的帮助和指导，给予我明确的方向和动力，以及在我迷茫和找不到问题原因时的指点和鼓励，帮助我顺利完成毕业设计整个过程。

其次，感谢我的父母，其在我整个毕设的完成过程中，经常给予我鼓励和支持，给予我很大的心理动力，帮助我坚持下去、相信自己，是我在毕业设计过程中的心灵慰藉。

最后，感谢我的同学：陈豪同学、段雷同学、丘丹磊同学和张宇虹同学，他们在我完成毕业设计的过程中，给予我课业上的帮助和直接的陪伴。当我忙于毕业设计时，他们主动帮我分担小组工作中的任务；当我受困于毕业设计中的问题时，他们给予我最直接的安慰和陪伴，以及对问题的讨论。感谢他们，在我完成毕业设计过程中对我的帮助和鼓励。

参考文献

[1] Sandro Cavallari, Vincent W.Zheng, Hongyun Cai, Kevin Chen-Chuan Chang and Erik Cambria: Learning Community Embedding with Community Detection and Node Embedding on Graphs. In CIKM'17, November 6-10, 2017, Singapore.

[2] Leonardo F.R. Ribeiro, Pedro H.P. Saverese and Daniel R. Figueiredo: struc2vec: Learning Node Representations from Structural Identity.

[3] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 701–710. ACM, 2014.

[4] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Largescale information network embedding. In Proceedings of the 24th International Conference on World Wide Web, pages 1067–1077. ACM, 2015.

[5] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 855–864. ACM, 2016.

[6] Bryan Perozzi, Vivek Kulkarni, Haochen Chen, and Steven Skiena. Don't walk, skip! online learning of multi-scale network embeddings. In 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE/ACM, 2017.

[7] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Grarep: Learning graph representations with global structural information. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pages 891–900. ACM, 2015.

[8] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1225–1234. ACM, 2016.

[9] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Deep neural networks for learning graph representations. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pages 1145–1152. AAAI Press, 2016.

[10] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information. In IJCAI, pages 2111–2117, 2015.

[11] Xiaofei Sun, Jiang Guo, Xiao Ding, and Ting Liu. A general framework for content-enhanced network representation learning. arXiv preprint arXiv:1610.02906, 2016.

[12] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. Homophily, structure, and content augmented network representation learning. In Data Mining (ICDM), 2016 IEEE 16th International Conference on, pages 609–618. IEEE, 2016.

[13] Huan Gui, Jialu Liu, Fangbo Tao, Meng Jiang, Brandon Norick, and Jiawei Han. Large-scale

embedding learning in heterogeneous event data. 2016.

[14] Linchuan Xu, Xiaokai Wei, Jiannong Cao, and Philip S Yu. Embedding of embedding (eoe): Joint embedding for coupled heterogeneous networks. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pages 741–749. ACM, 2017.

[15] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. GraRep: Learning Graph Representations with Global Structural Information. In *CIKM*. 891–900.

[16] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *KDD*. 701–710.

[17] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. GraRep: Learning Graph Representations with Global Structural Information. In *CIKM*. 891–900.

[18] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *KDD*. 701–710.

[19] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. In *KDD*.

[20] Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. 2016. Asymmetric Transitivity Preserving Graph Embedding. In *KDD*. 1105–1114.

[21] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. LINE: Large-scale Information Network Embedding. In *WWW*. 1067–1077.

[22] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural Deep Network Embedding. In *KDD*. 1225–1234.

[23] Hanyin Fang, Fei Wu, Zhou Zhao, Xinyu Duan, Yueting Zhuang, and Martin Ester. 2016. Community-Based Question Answering via Heterogeneous Social Network Learning. In *AAAI*. 122–128.

[24] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. 2016. Learning Convolutional Neural Networks for Graphs. In *ICML*. 2014–2023.

[25] V Blondel, A Gajardo, M Heymans, P Senellart, and P Van Dooren. 2004. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM review* (2004).

[26] Francois Lorrain and Harrison C White. 1971. Structural equivalence of individuals in social networks. *e Journal of mathematical sociology* 1 (1971).

[27] Lee Douglas Sailer. 1978. Structural equivalence: Meaning and de nition, computation and application. *Social Networks* (1978).