

Social Network Analysis

孙才俊

Beijing University of Posts and Telecommunications, Beijing, China
<https://www.iovi.com/about>

January 14, 2016

社交网络分析的核心问题

- 1 社交网络的结构特性研究
- 2 社交网络中群体互动研究
- 3 社交网络中的信息传播研究
- 4 概念整理
- 5 References

1. 社交网络的结构特性研究

要从一段无标注的自然文本中学习出一些东西，无非就是统计出词频、词的共现、词的搭配之类的信息。

在社交网络的结构特性研究方面,已有的研究工作可以概括为社交网络的结构分析与建模、社区发现、社交网络演化分析3个主要方面。

- 网络结构分析与建模

社交网络结构分析是通过统计分析方法来分析网络中节点度的分布规律、关系紧密程度、相识关系的紧密程度；社交网络建模采用结构建模的方法来研究产生这些特性的机制，以此来深刻认识社交网络的内在规律和本质特征。

- 社区发现

社区结构是在线社交网络的一种典型的拓扑结构特征。通过挖掘社区可以发现用户联系的紧密情况，获得用户之间的社交关系以及社会角色，并进一步结合社区内用户观点/行为等分析，有助于理解网络拓扑结构特点、揭示复杂系统内在功能特

- 社交网络演化分析

社区具有动态演化性，需对演化机理进行分析与识别。

1.1 网络结构分析与建模

在社交网络结构建模方面，图论方法得到了广泛的应用，很多学者都尝试运用图论对社交网络进行定量分析。结构特点：“六度分隔”、小世界现象、无标度、幂律分布和结构鲁棒性等。

- 单向关系

有向图(Directed Graph)模型刻画社交网络中的单向关系，并可使用有向图描述网络的中心(centrality)

- 个体-组织等归属关系

为描述个体-组织等归属关系，通常采用采用二部图(Bipartite Graph)和多部图模型(Multipartite Graph)，并可计算个体间接近性

- 多人合作关系模型

超图(Hypergraph)建模方法，认为一条边可以连接两个以上的节点；

- 针对个体横跨多社交关系模型

在无向图的基础上，提出基于节点唯一标识的多图交叉(Cross-Graph)模型,并且在多图并集上使用可控的启发式算法(Heuristic Algorithm)发现符合特定要求的网络结构

1.2 社区发现

在社会学领域,社区是一群人在网络上从事公众讨论,经过一段时间,所形成的人际关系网络。

- 不重叠的社区结构发现
通过寻找使得社区的模块度最大的网络划分来发现网络社区的合成聚类算法(Agglomerative Hierarchical Clustering), 但出模块度优化方法存在分辨率限制问题, 使得基于模块度优化的方法无法识别出一些较小的社区。
- 重叠的社区结构发现基于K-完全子图的重叠社区发现方法(K-Clique-Community Finding Algorithm), 该方法的优点在于能够揭示网络社区间的重叠现象, 不足之处在于其参数选择缺乏有效的理论指导。
- 网络层次化社区(Hierarchical Community Structure)的发现[1]
针对拟合函数的局部优化实现(Based on the Local Optimization of a Fitness Function).

1.3 社交网络演化分析

从社交网络社区结构是在线社交网络的一种典型的拓扑络演化中的统计规律展开研究,并提出了面向不同类型社交网络的演化模型。

- 时间和结构相结合的社区演化模型
[2]利用K平均(K-Means)和合成聚类(Agglomerative Hierarchical Clustering)给出了时间和结构相结合的社区演化模型
- 测度社区结构和社区进化
美国NEC实验室的迟(Chi)等人扩展了相似性计算方法,用图分割(Graph Cut)作为测度社区结构和社区进化的指标,第一个提出进化谱聚类(Spectral Clustering)算法

2. 社交网络中群体互动研究

社交网络中的群体互动方面,已有的研究工作可以概括为社交网络中群体行为建模及特征分析、群体情感建模与行为互动2个方面。

- 社交网络上的用户行为分析

在线社交网络上的用户行为包括展示自我、与陌生人建立关系、分享兴趣和信息、发布信息、搜索信息、浏览信息和推送信息;可以围绕各种话题与不同人群进行互动;可以构建兴趣社区、学习社区和娱乐社区,共享知识、学习交流并分享快乐。

- 群体情感建模与行为互动

情感分析是针对主观性信息(“支持”、“反对”、“中立”)进行分析、处理和归纳的过程,主观性信息表达了人们的各种情感色彩和情感倾向。社交网络中每个人情感状态不同,影响力也会不同。

2.1 社交网络上的用户行为分析

- 社交网络选择行为
采用博弈论的方法，通过制定个体选择策略和收益量化函数对个体行为进行建模和分析，研究表明社交网络的社会资本(Social Capital)收益、用户自尊心以及生活满意度等心理变量共同影响用户社交网络选择行为。
- 社交网络异构性及用户在其间的迁移行为
2011年，美国亚利桑那州立大学库玛 (Kumar) 等人针对Twitter、Facebook、Youtube等7种社交网络，研究了它们的异构性及用户在其间的迁移行为，建立了个体迁移和群体迁移模式，表征用户在社交网络之间的迁移方式。
- 挖掘群体中的个体行为特征
利用个体属性中多维性抽取和个体行为特征密切关联的属性取值，从而达到自动学习个体喜好和行为规律的目标；基于关联分析的属性独特性度量方法，用于挖掘群体中的个体行为特征。

2.2 群体情感建模与行为互动

- 群体感情建模

针对博客采用向量的表示方法对多元化情绪进行建模，但其基于表情符号来构建向量，并未利用内容信息进行分析研究。

- 群体行为互动建模

采用博弈论方法，对网络中的级联行为(Information Cascading)，个体效应与群体效应的相互影响，进行了建模和分析

- 传播的倾向性

先对文本的倾向性进行分析，再结合社区发现方法发现有相同倾向性的群体，用户会把自己的倾向性传播给具有连接关系的其他用户，使他们逐渐持有相同或相似的主观感受

- 个体从众行为分析

根据“追随领袖”模型中的领袖信用，建立集成社交网络信息的协同过滤推荐模型，该研究表明，基于意见领袖的推荐模型能够更加准确预测消费者的选择行为。

3. 社交网络中的信息传播研究

在社交网络的信息传播方面,已有的研究工作可以概括为社交网络的信息及其能量、社交网络信息传播模型和社交网络信息传播影响3个方面

- 社交网络的信息及其能量

在社交网络的信息及其能量方面,已有研究主要集中在信息的符号表示与意义,以及信息传播能量及演化方面。

- 社交网络信息传播模型

社交网络信息传播特指以社交网络为媒介进行的信息传播过程。

- 社交网络信息传播影响

社交网络信息传播影响是社交网络信息传播领域的关键研究热点,其目的是发现社交网络中最有信息传播影响力的节点集合,从而经过信息在社会网络中的传播,最终能够最大化信息的传播范围。

3.1 社交网络的信息及其能量

- 社交网络的信息
信息和信息熵是统计信息理论中最基本的概念和量，不仅用作信息量的度量，还用来表示自然和社会各种系统的有序度和无序度
- 在信息传播过程中的能量及演化分析
用信息传播状态熵刻画信息在社交网络的扩散状态及势能,指出信息传播过程中，随着多种异源异构信息互为补充不断演化，增强了信息最后的能量。
- 社交网络话题演化
基于联合概率推理，利用高斯条件随机场模型进行文本内容、影响力与话题演化的统一分析；

3.2 社交网络信息传播模型

已有研究主要集中在传染病模型、网络拓扑图模型以及基于统计推理的信息传播模型，之后出现了很多变种，例如SIR、SIS 以及类似于SI模型的级联模型

- 传染病模型SIRS

传染病模型认为，当感染个体对某个未感染个体的传播率大于某一临界值时，感染个体会将病毒传播给该未感染个体，这个过程会持续到整个网络感染个体总数处于某一稳定状态。

- 基于网络拓扑图的社交网络信息传播模型

在线社交网络的网络拓扑结构的演化特性，包括度分布聚类系数、群聚系数、顶点度相关系数等属性。

- 基于统计推理的社交网络信息传播模型

提出了BRPT算法，利用动态概率网络的结构分解来减小抽样分布的维数，利用部分样本对局部分布进行近似以提高计算效率

- 基于桶消元的信息传播算法

基于联合树的静态信息传播算法，可有效计算中等规模网络的信息传播；动态信息传播的条件BK（CBK）算法，通过引入局部结构的条件独立性提高动态信息传播的计算精度

3.3 社交网络信息传播影响

国内外对社交网络信息传播学影响的研究主要有概率论方法、经济学方法和传播学方法。

- 概率论方法

概率论方法是基于概率图模型对研究信息传播的敏感度和影响力的方法，可以对概率图参数、结构的重要性进行量化分析。基于蒙特卡罗的信息传播计算方法，但是这种类型的影响力分析方法不能有效揭示非线性关系。同时，研究者探讨了概率结构模型的影响力分析方法，如美国匹斯堡大学的王海琴等人[52]研究发现，贝叶斯网络对于参数概率的变化非常敏感，并论证了敏感度分析对贝叶斯网络参数分析非常有效，但这些贝叶斯网络敏感度分析方法还仅涉及到单个参数；将贝叶斯网络灵敏性分析延伸到多个参数，利用动态概率结构模型进行影响力的敏感度分析等。

3.3 社交网络信息传播影响

- 经济学中的效用理论

基于效用函数的信息传播影响力分析模型是利用函数评估信息传播对于系统所产生的影响，且易于实现信息传播影响力的度量。基于熵的网络信息传播的影响力评估函数方法，通过在强连接树中进行传播信息的影响力推理估算；基于多Agent因果概率图模型给出了一种利用部分信息的影响力分析方法

- 传播学领域

“沉默螺旋”等理论，指出个人在表达自己观点时受环境影响，当发现自己属于优势意见时，倾向于表明自己观点；反之，则转向沉默或附和。魔弹理论，把媒介对人的刺激看作是魔弹打入大脑，能迅速地被受众所接受。知识沟理论，是指由于社会经济地位，高者通常能比社会经济地位低者更快地获得信息，大众媒介传送的信息越多，这二者之间的知识鸿沟也就越有扩大的趋势。

- 启发式算法
- 超图
- Hierarchical Agglomeration Clustering
- 社会资本：指个体或团体之间的关联——社会网络、互惠性规范和由此产生的信任，是人们在社会结构中所处的位置给他们带来的资源



Andrea Lancichinetti, Santo Fortunato, and János Kertész.
“Detecting the overlapping and hierarchical community structure in complex networks”. In: *New Journal of Physics* 11.3 (2009), p. 033015.



Deepayan Chakrabarti, Ravi Kumar, and Andrew Tomkins.
“Evolutionary clustering”. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2006, pp. 554–560.