社会信息网络中的社区分析

程学旗 沈华伟
中国科学院计算技术研究所

关键词：社会信息网络 社区 聚集 中观结构

新型信息网络的一个突出特点是社会交互性强，例如人们用博客发表日志，用社交网站与朋友联络，用微博交流新闻动态等，这类信息网络又称为社会信息网络[1~2]。在使用社会信息网络的过程中，人们在网络空间结成了种种关系，比如，微博中的关注关系、社交网络中的好友关系、在线商店中因共同购买或评论产品结成的共同兴趣关系等。经验分析表明，社会信息网络中的社会关联关系普遍存在着局部聚集特性，而大量具有局部聚集特性的关系数据使我们可以从中观尺度（社区）来度量信息网络，并从中分析信息的传播规律，挖掘信息的利用价值。基于这一思路，本文将围绕社会信息网络中关系的局部聚集特性，分析社会信息网络中的社区结构，通过回顾社区发现的进展，介绍我们在该问题上的研究工作，探讨社区分析的研究趋势和应用前景。

社会信息网络和社区结构

按系统的观点来看，自然界和人类社会都是由不同的网络系统组建而成的，包括社会网络、信息网络和生物网络等。一个网络系统可以抽象为个体和关系两类基本要素。个体通常对应网络用户或者包含内容的主体，关系则对应用户间的交互与联系（如图1所示）。社会信息网络兼有信息网络和社会网络的特性，既是用户间社会关系的反映，也是用户间进行信息交互的载体。研究社会信息网络对我们认识当代人类社会的组织结构、群体演化特点、信息传播规律以及信息控制机制等都有着重要的现实意义。

社会学中对社会网络的研究由来已久，人们发现了很多有价值的现象，如小世界现象、度幂律分布等，并且已经产生了一些重要的理论，如同质性理论（homophily）[3]、结构平衡理论（structural balance）[4]和弱连接理论（weak tie）[5]等等。然而，传统的社会学研究往往使用调查问卷的形式获得数据，所研究的数据规模较小，并且难以得到个人完整的信息行为记录。因此，传统研究的成果更多的是来源于直观认识，缺乏基于大规模真实数据的实验验证。社会信息网络的繁荣大大改变了这一现状，它不仅记录了用户的个人信息和用户间的关系信息，而且记录了用户创造、传播和消费信息的过程，这些数据可以帮助研究人员验证已有的理论、发现新的规律。

社会信息网络中，个体按照不同类型的关系

图1 社会信息网络示意图

结成群体（如图1所示，不同颜色的节点标示着不同的群体），如社交网络中的圈子、小组等。群体体现了关系的局部聚集特性，同一个群体内部的个体之间关系相对密切，不同群体的个体之间关系相对疏远。在网络系统研究中，人们把个体视为节点，关系视为边，群体视为社区（community）。社区通常由功能相近或性质相似的网络节点组成，在一定程度上反映了个体自发、无序行为背后的局部弱规则性和全局有序性。因此，社区成为研究网络结构和功能（包括信息扩散、网络演化、导航和搜索等）关系的“脚手架”和切入点。例如，在万维网（World Wide Web）中，社区对应着通过超链接紧密关联的网页，同一个社区的网页具有相近的话题，这一特征是PageRank[6]、HITS[7]等链接分析算法的基础假设之一；在基于协同过滤的推荐系统中，用户间因对产品的打分关系结成的隐式社区是很多推荐算法的基础。

作为包括社会信息网络在内的很多复杂网络所普遍具有的一种结构特性[8]，社区结构在过去十年内得到了多个学科领域的广泛关注和深入研究。针对社区结构的研究大体上可以分为社区发现、社区演化分析和社区结构与网络动力学三个类型。其中，社区发现直接关系到网络系统的中观度量与对应的共性规律，是一项基础性问题，在过去十多年内吸引了多个领域学者的关注，形成了很多社区发现方法[9]。众多社区发现方法的提出为我们提供了充分的选择自 由度，同时由于缺乏各种方法间优劣的对比，也使得我们在根据具体应用需求选择相应方法时，面临着诸多困惑。另外，目前研究人员对网络社区认识上仍然存在差异。由于社区一开始仅具有一个定性的定义，长期以来，人们针对社区的定量定义进行了大量的尝试，但仍没有哪一种能得到广泛的认可。网络社区的重叠性、多尺度、异质等特性使得社区结构的研究变得更加复杂，也进一步拓宽了社区结构的理论研究与应用分析的范围。同时，人们对社区结构的研究逐步从具有单一类型节点、单一类型连边关系的网络扩展到具有多种类型节点和连边关系的网络上，包括有向网络、带权网络、二部图网络、多部图网络、具有多样性关系的网络甚至超图等。将网络社区结构扩展到更一般意义上的网络中观结构及其规则性（regularity）是目前重要的研究趋势。

与其他的复杂网络系统类似，社会信息网络的社区结构具有一些普适的共性规律和信息传播与社会交互的特例现象。本文后续内容将重点介绍社会信息网络中社区结构分析相关的研究进展，具体包括：（1）层次重叠社区发现：作为社会行为交互和信息扩散的媒介，社会信息网络所固有的社区结构既具有层次化现象、又高度重叠。这种层次重叠特性，一方面使得信息扩散得以高效进行，另一方面使得社会信息网络自身实时演化，并使得其在结构和信息传播方面出现大量的突发涌现现象；（2）社区结构和网络动力学：与其他网络系统类似，社会信息网络的结构与其所具有的功能与性能之间相互影响，互为因果。因此，探索与信息扩散等动力学过程具有密切关系的网络社区结构是社会信息网络中社区结构分析的关注重点之一；（3）异质网络多尺度社区发现：社会信息网络具有多尺度的社区结构，不同尺度的社区结构体现着网络不同的功能，小到只有几个节点的互动关系，大到具有数百个节点聚集起来的群组。由于社会信息网络具有服从幂律分布的节点度（也被称为节点度的异质性），给多尺度社区结构的发现带来了严峻挑战。结合社会信息网络的结构特征进行社区分析，对于我们认识社会信息网络上信息扩散、群体突发涌现等现象背后的机理具有重要意义。

社区发现

社区发现旨在识别出网络固有的社区结构，即按照节点间的连边关系把节点划分成若干节点组，使得节点组内部的连边相对稠密，不同节点组之间的连边相对稀疏。这一目标体现着“物以类聚、人以群分”的朴素思想，与图划分、传统聚类问题在

方法论上有着千丝万缕的联系。实际上，早期的社区发现方法大多借鉴图划分或聚类，譬如图划分的代表性方法NCut、RatioCut等和传统聚类中的层次聚类方法。

2002年，社区发现由格文（Girvan）和纽曼（Newman）正式提出[8]，使用的方法是一种分裂式层次聚类方法。在该方法中，作者为网络中的每条边定义了一个被称为边介数（edge betweenness）的量，一条边的边介数定义为网络的所有最短路径中经过该边的路径数目占最短路径总数的比例。边介数反映了相应的边在整个网络中的桥接作用，因此，当按照边介数由高到低依次删除边时，网络分裂的速度要远快于随机删除连边。格文等人提出的社区发现方法是按照边介数依次删除边，删边的过程对应着一个树图（dendrogram）。随后，纽曼和格文分别基于最短路径、随机行走、电流三个视角给出了不同的边介数，采用分裂式层次聚类算法进行社区发现。

归纳来看，分裂式层次聚类方法是自顶向下的过程，先将整个网络看成一个完整社区，然后按照某种策略逐个删除边，进行社区分割，直到每个顶点都孤立为止。与分裂式方法相反，聚合式层次聚类方法是自底向上的，先将网络中每个节点视为一个单独的社区，然后依据某种策略选择社区进行归并，直到所有节点都属于同一个社区为止。两种方法都最终得到一个树图，选择该树图的任何一层对树图进行切割都获得网络的一个划分，划分中的每个分量视为一个社区。两种方法面临的共同困难在于切割位置的选择，即选择合并或分裂停止的时机。该问题的本质在于缺乏一种度量网络划分质量的手段。

为解决网络划分度量这一本质问题，纽曼等人提出了著名的模块度（modularity）的概念[10]。模块度采用一种被假定没有社区结构的网络（一般使用配置模型产生）作为参照网络，对于一个给定的网络划分，通过对比原有网络和参照网络中处于该划分的各个分量内部边的比例，给出一种度量网络划分质量的手段。模块度的定义为

$$Q=\frac{1}{2m}\sum_{ij}\left(A_{ij}-\frac{k_ik_j}{2m}\right)\delta\left(c_i,c_j\right)$$

这里，m表示网络中边的总数，邻接矩阵A的元素A_{ij}表示连接节点i和j的边数，$k_i=\sum_j A_{ij}$表示节点i的度，c_i表示节点i所属的社区，当且仅当$c_i=c_j$时$\delta\left(c_i,c_j\right)=1$。一般认为，对于同一个网络的不同划分，模块度越高，该划分越能体现网络固有的社区结构。如此一来，网络社区发现问题变成了一个模块度优化问题，即从网络的所有可能划分中寻找一个划分，使该划分具有最大的模块度，该划分的各个分量视为社区。模块度的提出很大程度上推动了社区发现的研究，研究人员开始探索基于模块度优化的算法。布兰德（Brandes）等人指出模块度优化问题是一个NP难问题[11]，因此，寻找启发式优化方法成为主要的研究思路。基于贪婪算法、模拟退火、极值优化、谱优化等算法，人们提出了一系列模块度优化方法。同时，研究人员开始把模块度的定义扩展到有权网络、有向网络和二部网络等其他类型的网络上。

模块度优化方法迅速成为应用最广泛的社区发现方法，在很多领域的网络上取得了成功应用。然而，进一步的研究发现模块度优化方法在处理各种具体网络时存在诸多缺陷：（1）文献[12]发现模块度所采用的参照网络由于随机波动也会呈现出伪社区结构，从而对模块度的参照网络提出了质疑，并基于统计显著性给出了相应的改进办法；（2）文献[13]指出模块度优化方法存在“分辨率限制”问题。对于给定的网络，存在一个固有的分辨率，使得模块度优化方法不能识别出该分辨率以下的社区，因此，一些小社区往往被大的社区吸纳，从而淹没在其他网络社区结构中而无法被识别出来；（3）本文作者在文献[14]指出模块度优化方法不能很好地处理节点度分布高度异质的网络，而真实网络的度分布往往服从幂率（power-law）分布，进而通过引入尺度（rescaling）变换，成功地解决了该问题。针对模块度优化方法在处理社会信息网络中

的社区结构时所面临的各种问题，我们进行了深入研究，在层次重叠社区发现、社区结构和信息扩散的关系、异质网络多尺度社区结构发现等方面取得了一系列进展。

层次重叠社区发现

基于模块度的社区发现方法把社区发现问题视为寻找网络中具有最大模块度的网络划分问题，因此，每个节点属于且仅属于一个社区，社区之间不允许重叠。然而，真实世界复杂网络的社区之间普遍存在重叠节点（如图2所示），这些节点在不同社区间起着桥梁作用，在社区中发挥着重要作用，是社区间信息扩散的媒介和社区演化的推手。例如，社会网络中，一个人可以同时属于多个社交圈子（社区），而同时属于多个社交圈子的人又是信息传递、社会交往中的关键因素。

鉴于上述问题，人们开始着眼于重叠社区结构的研究。最初，匈牙利科学院帕尔拉（Palla）等人在《自然（*Nature*）》上发表论文[15]，指出了社区重叠这一重要现象，并提出了一种基于完全子图渗流（clique percolation）的重叠社区发现方法。随后，该算法被扩展应用到有权网络、有向网络和二部网络等。截至目前，完全子图渗流算法仍然是重叠社区发现方法中应用最广泛的方法之一，被应用到生物、信息、社会等网络上。尽管基于完全子图渗流的方法在发现重叠社区方面取得了一定成功，但仍面临一些挑战。首先，该方法所发现的社区仅

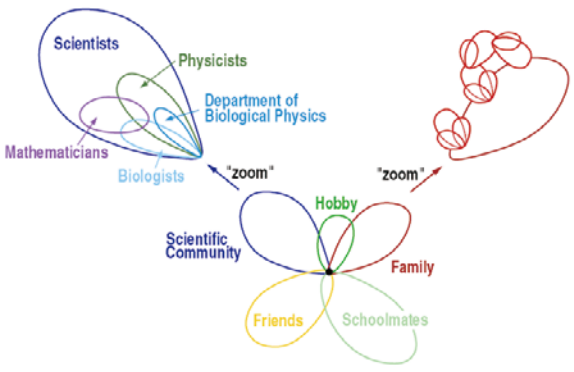

图2 层次重叠社区结构示例[15]

能覆盖网络中少数节点，大多数节点不属于任何一个社区；其次，该方法无法像传统层次聚类社区发现方法那样找到社区的层次结构。更重要的是，该方法的参数选择对于不同的网络差异较大，至今没有合适的参数选择方法。

我们通过对真实世界复杂网络的实际观察发现，很多网络社区，包括社会信息网络，既相互重叠、又具有层次结构。因此，提出一种能够同时揭示网络层次重叠社区结构的社区发现方法对于认识网络的结构具有重要意义。我们从分析重叠社区结构的成因入手，指出使用单个节点作为社区的基本构成单元是阻碍模块度方法不能处理社区重叠现象的根本原因；提出了使用极大完全子图来代替单个节点作为社区的基本构成单元[16]，定义了一种新的网络模块度，能够度量网络的重叠社区结构，通过构造极大完全子图网络，我们证明了以新模块度为目标函数，任何基于模块度优化的非重叠社区发现方法都可以不加修改地直接应用于发现重叠社区结构[17]；采用聚合式层次聚类的方式，我们提出了一种能够同时揭示网络层次重叠社区结构的社区发现方法，并成功应用于词共现网络、学术合作网络以及蛋白质交互网络等多个领域的复杂网络中。

社区结构和网络动力学

模块度优化方法仅考虑了网络的静态拓扑结构，并不能很好地反映网络结构与网络上发生的诸如信息扩散、同步等动态过程的关系。研究网络动力学的目的在于揭示网络拓扑结构对发生在其上的动态过程的影响，以及这些动态过程是否能够反映其“承载网络”的拓扑结构特征。在自然界和人类社会中，大量存在着这样的动态过程，包括疾病传播、流言散布、计算机病毒扩散、新产品推广（像口口相传等）、交通流、级联失效、同步、渗流、随机行走等。因此，探索能够和网络动力学具有天然关联关系的社区结构成为重要的研究问题。

针对这一问题，发表在***Phys. Rev. Lett.***上的文献[18]通过研究网络同步过程和社区结构的关系，给出了一种能够揭示网络同步过程中阶段性特征的社区发

现方法。文献[19]通过考察网络上随机行走路径的最短描述长度来研究网络的社区结构，提出了著名的InfoMap方法，该方法是目前为止最好的基于网络划分的社区发现方法。

研究网络动力学和社区结构关系的核心在于找到网络上发生的动态过程和网络结构的关联机理。针对信息扩散这一在社会信息网络中最为常见的动态过程，我们使用微分方程刻画该动态过程，采用自由基表达方式给出了该微分方程的闭式解。该闭式解清晰地揭示了信息扩散过程中涌现的局部均衡态和网络归一化的拉普拉斯矩阵谱之间的关系[20]。我们发现了一个有趣的现象，如图3所示，网络所固有的社区结构和扩散过程的两个稳定均衡态具有一一对应关系，还发现这种对应关系在自由基表达上对应着网络归一化拉普拉斯矩阵的较大谱间距，

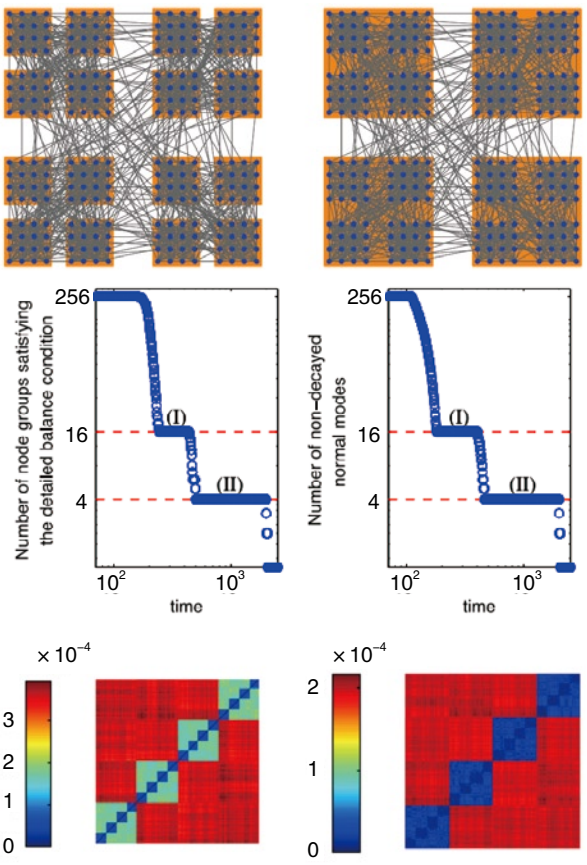

图3 社区结构与扩散过程的稳定局部均衡态的对应关系[20]

为此提出了使用网络谱分析方法揭示这种和扩散过程紧密关联的网络社区结构。

异质网络多尺度社区发现

模块度优化方法通过寻找模块度最大的网络划分来刻画网络的社区结构。由于该方法仅使用一个网络划分来表达社区，因此无法揭示网络的多尺度社区结构。实际上，很多网络都具有多个尺度的社区结构。如图4所示，该网络天然具有两个尺度的社区结构，即分别包含4个社区和16个社区。尽管直接优化模块度无法识别网络的多尺度社区结构，纽曼等人通过分析模块度矩阵的谱，给出了一种能够揭示网络多尺度社区结构的方法。

我们发现，基于模块度矩阵的谱分析方法在揭示网络多尺度社区结构时，其准确率受到网络节点度的异质性严重影响。众所周知，真实世界复杂网络中节点的度服从幂律分布，这便是节点度的异质性。基于模块度矩阵的多尺度社区发现方法之所以受到节点度异质性影响，原因在于模块度的定义对节点度的处理不恰当。我们前面介绍模块度时指出，模块度使用的参照网络考虑了节点度的异质分布。具体做法是：实际网络中节点i和节点j之间的连边数为A_{ij}，参照网络中，节点i和节点j之间的连边数为$k_i k_j / 2m$，使用同一个社区内部的两个节点i和j的连边数的差值$A_{ij} - k_i k_j / 2m$作为模块度定义的基本要素。我们从线性降维的角度分析了社区发现，指出模块度矩阵是网络的一种协方差矩阵，基于模块度矩阵谱分析的多尺度社区发现方法仅考虑了平移和旋转两种变换，因此受到节点度异质性的严重影响。为此，我们引入尺度伸缩变换[14]，将之与平移、旋转结合，很好地解决了异质网络的多尺度社区发现问题。通过在国际公认的标准测试集上进行实验发现，我们的多尺度社区发现的准确率显著好于基于模块度矩阵的方法，当网络中社区的模糊程度（mixing ratio）较高时，我们的方法能够准确地识别出网络固有的多尺度社区结构。

社区发现的发展趋势

社区演化

社区演化是社区结构研究中的一个重要问题。演化是真实网络所具有的基本特性，社区演化是网络自身结构和在其上频繁发生的交互过程相互作用的结果。社区演化分析主要研究社区随时间变化的情况，并分析导致这些变化的机制和原因。社区演化主要包括社区形成、社区生长、社区缩减、社区合并、社区分裂、社区消亡等。与社区发现相比，社区演化的研究仍然处于“婴儿”期，造成这一结果的主要原因有：（1）针对静态网络拓扑进行社区发现仍然是一个存在较大争议、尚未解决好的问题，吸引了社区结构研究的主要注意力；（2）缺乏或很少具有时间标签的数据用于支持社区演化分析的研究。近几年，在社区发现取得长足进步的基础上，社区演化得到了越来越多的关注。文献[21]基于他们提出的完全子图渗流社区发现方法研究社区演化，得到一个有趣结论：小社区的稳定性是保证其存在的前提，而大社区的动态性是它存在的基础。随着含时间数据的积累，关于社区演化的研究将会成为一个热点。

局部社区发现

现有社区发现方法大多要求知道网络的全部结构，然而在很多真实情况下，这种要求是无法满足的，主要包括两方面的原因：（1）网络结构高度动态变化，难以实时获取全部结构；（2）网络规模大，获取全部结构困难。另外，一方面基于网络全部结构的社区发现方法计算复杂度高，难以满足实时性的应用需求；另一方面在很多应用场景中，人们并不需要知道网络的全部社区结构，只需关注特定节点所在的社区结构。这些都亟待局部社区发现方法的提出。

实际上，局部社区发现方法的研究由来已久，也出现了很多局部社区发现方法[22]。然而，局部社区发现方法面临性能不稳定等诸多问题，这就大大制约了局部社区发现方法的应用。尽管如此，在社会信息网络规模快速增长的今天，此方法的研究仍然是社区结构研究的发展主流。

异质关系网络的社区发现

现有社区结构的研究面临一个共同问题，即所研究的网络均为具有单一关系的同质网络，且所发现的社区多为单一尺度的社区。社会信息网络的繁荣产生了众多的大规模异质关系网络，这类网络大多具有多尺度的社区结构。因此，最近社区结构的研究主要关注大规模异质关系网络的多尺度社区结构之上，2010年的代表性论文包括：文献[23]通过扩展模块度研究异质关系网络的多尺度社区结构，并将之用于分析美国东北部某大学1640名在校学生因四种社会关系构成的异质关系网络；文献[24]通过使用节点间的连边代替节点自身作为社区的组成单元，提出一种基于层次聚类的多尺度社区发现方法，并将其用于分析手机用户通话关系网络的多尺度社区结构；文献[25]从社区随时间变化的稳定性角度研究多尺度社区结构，应用于分析蛋白质的原子级结构特征；文献[26]以一个大型多用户在线网络游戏的用户社会关系网络为对象，研究了异质关系网络中多种关系并存对网络结构（包括社区结构）的影响，对游戏用户的行为进行了分析。这些研究表明，研究异质关系网络的多尺度社区结构是网络结构分析的基本问题，对于分析社会信息网络中个体、群体乃至整个网络的行为具有重要意义。

综上所述，关于网络社区结构的动态研究及社会信息网络的快速发展均表明，关于大规模异质关系网络的多尺度社区结构的研究是社区结构研究的发展方向，并在用户行为分析、人类动力学、社会搜索、网络舆情、信息推荐等方面具有广阔的应用前景。

基于社区结构的其他相关应用

与搜索技术互补，推荐技术是解决信息过载问题的关键技术。在过去20年间，基于协同过滤的推

荐技术得到了长足的发展，出现了包括user/item-based的协同过滤方法、基于矩阵分解的方法等等。这些推荐算法的背后都蕴藏着共同的基本假设：行为（评论、打分、购买、浏览等）相似的个体在未来仍然具有相似的行为。该假设恰恰体现了社区的思想，这里的社区是由用户行为结成的隐式社区，同一个社区的用户之间并没有直接关系。

近几年，在线商店与社交网络的融合给社会推荐提供了发展机遇。在社会推荐中，社区的作用日益突出。与传统的推荐相比，社会推荐是利用个体间的直接社会关系结成的显式社区进行推荐的。通过识别出每个个体所属的社区和利用社区中其他成员的行为来为该个体提供推荐服务，从而挖掘社区中个体间的影响力，并提供更为可靠和良好用户体验的推荐服务。

此外，通过识别社区，可以更好地让社会推荐考虑口碑（word-of-mouth）的因素，鼓励用户向其朋友推荐物品，并且在推荐过程中显示用户对该物品的评价，使被推荐的用户更加容易接受，这种推荐更贴近于真实生活中的推荐过程。在这类用户对用户的推

CCF YOCSEF报告会“移动互联网技术趋势”举行

Panel讨论
左起为宋乐永、李琳、刘承涛、陈石、林兴陆、靳岩

移动互联网的发展速度明显快于桌面互联网，其规模将数倍于桌面互联网。CCF YOCSEF“移动互联网技术趋势2012”报告会由此而生。

2011年11月11日，“移动互联网技术趋势2012”学术报告会在北京航空航天大学举办，UC优视副总裁陈石、北京播思通讯技术（北京）有限公司软件经理刘承涛、北京易路联动技术有限公司总经理助理李琳接受邀请作主题报告。

截至2011年6月，网民数量为4.85亿，手机网民已达3.18亿。移动通信与互联网正在通过整合产业资源，形成移动互联网产业链。这个产业链由平台提供商、设备提供商和服务提供商等部门组成，并且向商务、金融、物流、游戏等应用领域快速延伸。

陈石通过对中国移动互联网用户的分析，探讨了移动互联网行业的创新，特别是手机浏览器的创新与开放。刘承涛详细介绍了播思通讯OMS 平台的功能和特点，对OPhone SDN 社区基本情况做了介绍。李琳从移动互联网软件开发和运行支撑等方面论述了Web中间件的重要性，对Web中间件技术及应用做了详细的介绍。

在Panel环节，三位特邀讲者、eoeandroid 社区创始人靳岩、开拓天际App部总经理林兴陆与参会者一起交流了移动互联网产业发展的技术趋势、创业机遇等。

参加报告会的有AC委员陈益强、齐红威等以及YOCSEF大连副主席孙晓鹏等。中国图像图形学学会多媒体专委会副秘书长陈志华博士以及来自高校、研究所、企业界的专业人士共100多人。报告会由AC委员卜佳俊、胡春明和宋乐永主持。

荐方法中，确定用户之间的相互影响力成为研究的关键。与信息传播过程中用户间的关系对信息传播的影响类似，用户间关系对物品的推荐具有很大的影响，而且在不同类型的物品上的影响不相同。如何在真实推荐数据中挖掘社会影响对推荐的作用，并以此来改善推荐系统的性能，仍在探索过程中。

社区是众多社会计算平台普遍具有的核心组件，比如兴趣小组、朋友圈子、群（组）等。社区的存在对于在线社会信息网络的成功与否至关重要，社区可以增强用户的黏着度，提高用户体验。除了用户或系统创建的显式的社区，用户间的交互关系会自发涌现出很多隐式社区，自动发现这些隐式的社区对于精准营销、信息推荐、隐藏组织发现等应用显得十分重要。社会信息社会网络中，好友推荐和社区推荐都是隐式社区发现的重要应用。

结语

作为网络科学研究的热点，网络社区结构的研究得到了很多领域的关注，取得了很多进展，特别是社区发现算法的研究。本文以社区结构的主要研究成果为主线回顾了社区结构的研究历程，并以模块度方法在处理社会信息网络的社区结构时面临的问题为线索，介绍了我们在社区发现方面的工作。总体来看，社区结构的研究方兴未艾，尚存在很多未能解决或者没有很好解决的问题，包括异质多尺度社区发现问题、社区结构重叠现象、网络结构规则性的探索、大规模网络的局部社区发现等。在关注这些研究问题的同时，如何把社区结构的研究成果应用到实际的场景和具体的网络中是一个重要的课题。我们期望看到社区结构的研究在社会信息网络的应用中发挥更大的作用。■

致谢

本文得到国家自然科学基金项目（No. 60873245、No. 60933005）和国家重点基础研究发展计划项目（2012CB316303）资助。

程学旗

CCF 高级会员。中国科学院计算技术研究所研究员。主要研究方向为网络科学、网络与信息安全以及互联网搜索与服务。cxq@ict.ac.cn

沈华伟

CCF会员。中国科学院计算技术研究所助理研究员。主要研究方向包括网络科学、社会网络分析、数据挖掘。shenhuawei@ict.ac.cn

参考文献

[1] 李国杰. 关于网络社会宏观信息学研究的一些思考. 中国计算机学会通讯, 2006, 2(2): 2~7

[2] 程学旗, 陈海强, 韩战钢. 社会信息的网络化分析初探. 中国计算机学会通讯, 2006, 2(2): 18~26

[3] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: homophily in social networks. Annual Review of Sociology. 2001, 27:415~444

[4] F. Heider. Attitudes and cognitive organization. Journal of Psychology. 1946, 21:107~112

[5] M. S. Granovetter. The strength of weak ties. Amer. J. of Sociology. 1973, 78(6): 1360~1380

[6] S. Brin and L. Page. Anatomy of a large-scale hypertextual web search engine. In Proceedings of the 7th international conference on World Wide Web (WWW′98) , 1998, 107~117

[7] J. Kleinberg. Authoritative sources in a hyperlinked environment. J. ACM, , 1999, 46(5):604~632

[8] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. Proc. Natl. Acad. Sci., 2002, 99(12): 7821~7826

[9] S. Fortunato. Community detection in graphs. Phys. Rep., 2010, 486(3-5): 75~174

[10] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Phys. Rev. E, 2004, 69(2): 026113

[11] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner. On modularity clustering. IEEE Trans. Knowl. Data Eng., 2008, 30(2): 172~188

[12] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral. Modularity from fluctuations in random graphs and complex networks. Phys. Rev. E, 2004, 70:025101(R)

[13]. S. Fortunato and M. Barthélemy. Resolution limit in community detection. Proc. Natl. Acad. Sci., 2007, 104(1): 36~41

[14]. H. W. Shen, X. Q. Cheng, and B. X. Fang. Covariance, correlation matrix and the multiscale community structure of networks, Phys. Rev. E, 2010, 82:016114

[15]. G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. Nature, 2005, 435(7043): 814~818

[16]. H. W. Shen, X. Q. Cheng, K. Cai, and M. B. Hu. Detect overlapping and hierarchical community structure in networks, Physica A, 2009, 388(8):1706~1712

[17]. H. W. Shen, X. Q. Cheng, and J. F. Guo. Quantifying and identifying the overlapping community structure in networks, J. Stat. Mech., 2009, P07042

[18]. A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente. Synchronization reveals topological scales in complex networks. Phys. Rev. Lett., 2006, 96(11): 114102

[19]. M. Rosvall, C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci., 2008, 105(4): 1118~1123

[20]. X. Q. Cheng and H. W. Shen. Uncovering the community structure associated with the diffusion dynamics on networks, J. Stat. Mech., 2010, P04024

[21]. G. Palla, A. L. Barabási, and T. Vicsek. Quantifying the social group evolution. Nature, 2007, 446(7136): 664~667

[22]. J. P. Bagrow. Evaluating local community methods in networks. J. Stat. Mech. 2008, P05001

[23]. P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J. P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. Science, 2010, 328(5980): 876~878

[24]. Y. Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. Nature, 2010, 466(7307): 761~764

[25]. J. C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. Proc. Natl. Acad. Sci., 2010, 107(29): 12755~12760

[26]. M. Szell, R. Lambiotte, and S. Thurner. Multirelational organization of large-scale social networks in an online world. Proc. Natl. Acad. Sci., 2010, 107(31): 13636~13641