控制与决策
Control and Decision

时序网络构建的理论和方法

李阿明, 侯谷庾, 王龙

引用本文:
李阿明,侯谷庾,王龙. 时序网络构建的理论和方法[J]. 控制与决策, 2023, 38(6): 1473−1490.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.1797

时序网络构建的理论和方法

李阿明[1,2], 侯谷庚[3], 王 龙[1,2†]

(1. 北京大学 工学院，北京 100871；2. 北京大学 人工智能研究院，北京 100871；
3. 北京大学 前沿交叉学科研究院，北京 100871)

摘 要: 20世纪末复杂网络小世界与无标度特性的发现,使多类复杂系统的结构特性、动力学、决策与控制在21世纪初得到了前所未有的关注与发展. 鉴于复杂网络在刻画复杂系统拓扑结构方面的有效性,首先介绍构建具有典型特征静态复杂网络的重要模型与方法,这些模型与方法使复杂网络的构建不再依赖有限且高成本的个体真实交互数据,为多领域研究人员探讨相关科学问题提供了便利条件. 其次,随着高精度海量群体交互数据构建采集能力的不断提升,构建随时间演化的动态时序复杂网络成为可能,作为时序网络的一个典型特征,个体交互的时间间隔往往呈现幂律分布,即具有爆发特性,这种爆发特性可显著改变系统中的信息传播、博弈决策过程,鉴于此,总结对真实个体交互数据进行幂律分布定量检验的参数估计方法,介绍泊松过程与排队系统,给出几类时序网络构建的理论与方法.

关键词: 复杂系统；复杂网络；时序网络；幂律分布；无标度特性；爆发特性

中图分类号: TP393.0 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1797

引用格式: 李阿明,侯谷庚,王龙. 时序网络构建的理论和方法[J]. 控制与决策, 2023, 38(6): 1473-1490.

Theory and method for constructing temporal networks

LI A-ming[1,2], *HOU Gu-yu*[3], *WANG Long*[1,2†]

(1. College of Engineering，Peking University，Beijing 100871，China；2. Institute for Artificial Intelligence，Peking University，Beijing 100871，China；3. Academy for Advanced Interdisciplinary Studies，Peking University，Beijing 100871，China)

Abstract: At the end of the 20th century, the discovery of the characteristics of complex networks, such as small world and scale-freeness, brought unprecedented attention and development in the structural characteristics, dynamics, group games, decision-making, and control of various complex systems in the early 21st century. Given the effectiveness of complex networks in characterizing the underlying topology of complex systems, here we first introduce several representative models and methods for constructing synthetic static complex networks with essential characteristics. Such models and methods change the state of constructing complex networks from individuals' limited and high-cost empirical interaction data, and thus provide an effective solution for constructing static networks for researchers in multiple fields to further explore related scientific issues. In recent years, with the continuous improvement of the ability to collect massive high-precision interaction data of individuals, it is possible to construct temporal networks, which dynamically evolve over time. As an essential feature of temporal networks, the inter-event time of interactions often presents a power-law distribution, namely, the bursty behavior. Existing results have shown that the bursty behavior may significantly change the information dissemination, game, decision-making, and control of temporal networks. We further summarize the parameter estimation method for testing the power-law distribution of empirical individuals' interaction data, and introduce the Poisson process and queuing system, and present typical theories and methods for constructing various temporal networks.

Keywords: complex system；complex network；temporal network；power-law distribution；scale-freeness；bursty behavior

0 引 言

1998年,康奈尔大学数学系非线性动力学家Strogatz及其学生Watts通过分析电影演员合作网络(collaboration network of film actors)、美国西部电力网络(power grid of the western United States)和秀丽隐杆线虫神经网络(neural network of the worm

收稿日期: 2022-10-17；**录用日期:** 2023-03-28.

基金项目: 国家重点研发计划项目(2022YFA1008400)；国家自然科学基金项目(62036002, 62173004)；北京市科技新星项目(Z211100002121105).

†通讯作者. E-mail: longwang@pku.edu.cn.

Caenorhabditis elegans)等真实数据,发现小世界网络(small world network)在真实世界中普遍存在[1]. 以演员网络为例,将电影演员视为网络节点,若任意两位演员同时参演同一部电影,则在节点间建立一条连边. 根据一段时间内影片的数据记录,可构建出相应的小世界网络. 同时定量的研究发现,真实世界中的小世界网络节点间不仅具有高聚类(high clustering)特性,而且任意两节点之间最短路径往往较短. 可以说,网络中高聚类与短特征路径(characteristic path)特性的发现有力推动了传统研究中所关注的规则网络(regular network,每个节点的连边数量(即节点度)相等,且节点间具有高度对称性,此类网络中节点间往往具有高聚类特性)和随机网络(random network,节点间以某一概率相连,以ER随机网络模型[2]为代表,此类网络中节点间往往具有短特征路径特性)的长期探索. 进一步,Strogatz和Watts提出了小世界网络模型,该模型可有效构建具有以上特性的小世界网络.

1999年,美国圣母大学的物理学家Barabási及其学生Albert通过探索万维网(World Wide Web)、学术论文引用网络(citation network of scientific publications)、电影演员合作网络、电力网络等几类真实数据集,发现网络中节点度为k的节点出现概率$P(k)$随着k幂律递减,即$P(k) \sim k^{-\gamma}$,且指数γ往往介于2与3之间,即具有无标度特性(scale-freeness)[3]. 在随机网络模型和小世界网络模型刻画的网络中,节点几乎不会出现度较大的情形,这与真实数据中频繁观测到大度节点存在的现象不符. 对此,Barabási等在已有模型的基础上,提出了无标度网络(scale-free network)模型. 该模型不仅解释了真实网络的无标度特性形成过程,还为构建无标度网络提供了有效方法. 可以说,无标度网络构建模型的提出不仅使具有此类特性网络的产生不再依赖有限且高成本的个体真实交互数据,还为多领域研究人员开展多类复杂系统动力学[4]、群体博弈[5-6]、决策与控制[7]等研究打下坚实的网络拓扑结构的基础.

基于真实数据传统静态网络构建过程如图1所示. 图1(a)表示根据不同系统中对节点和边的定义,以6个节点为例,由真实数据建立节点间(随时间先后)交互的关系,并由所有节点在整个数据采集窗口内的交互数据,通过聚合所有节点间所有连边构建相应的静态网络. (b)展示了当所构建的静态网络拓扑结构呈现出随机连接的随机网络时对应的网络结构. 根据每个节点所连接的边数量(称为节点度,记为k),(c)给出了相应的度分布. (d)展示了当所构建的静态网络拓扑结构为无标度网络时对应的网络结构. (e)给出了双对数坐标下度分布的幂律特性.

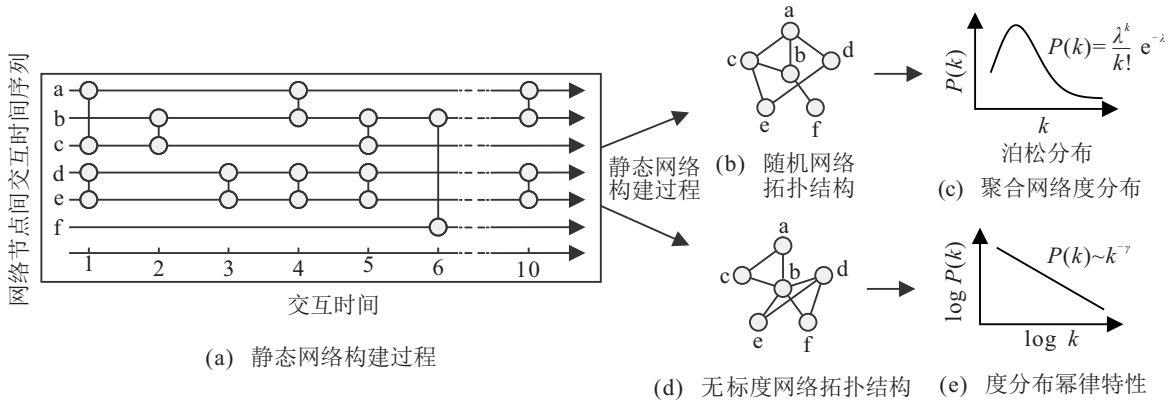

图1 基于真实数据的传统静态网络构建过程

近年来,随着高精度海量节点交互数据采集能力的不断提升,获得带有节点交互时间的交互序列成为可能. 与此同时,在聚合数据采集整个时间窗口内用个体交互数据构建传统静态网络(static network)的基础上(图1),采用较短时间窗口构建时间依赖的动态时序网络(temporal network)成为研究人员关注的焦点问题[5-24](图2).

基于真实数据的时序网络构建过程如图2所示. 图2(a)以6个节点间交互数据为例,不同于传统静态网络构建过程,而是基于较小数据采集窗口Δt,分先后时段分别构建出相应的子网络(snapshot),进而由系列子网络构成拓扑结构随时间演化的时序网络. 当Δt为整个数据采集窗口时,对应的时序网络退化为传统静态网络. (b)表示在时序网络中,个体交互时间的间隔(τ_i)满足一定的分布. 以节点b为例,此分布可以根据个体的交互时间进行确定,如(c)所示. 同时,对于时序网络而言,两节点交互(即节点间的边活跃)时对应的时间间隔(δ_i)往往也满足一定的分布,

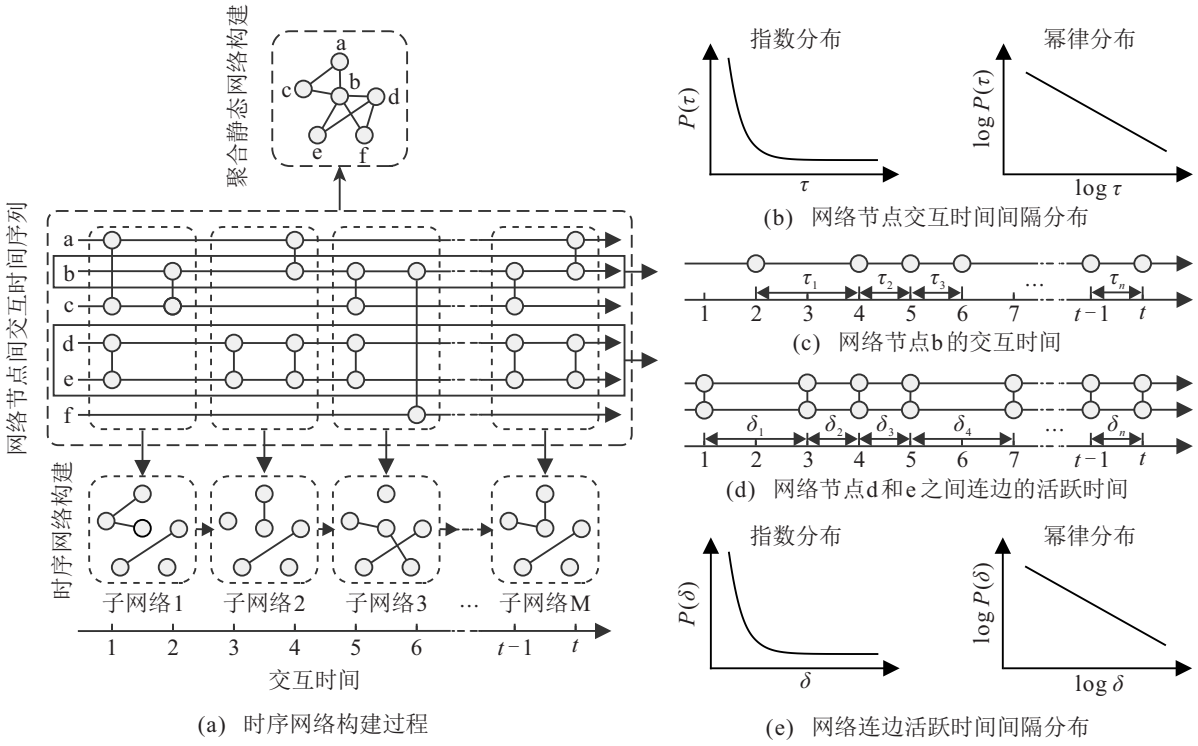

图 2　基于真实数据的时序网络构建过程

如(d)所示. 以节点d和e为例,可将此节点间边活跃的时间点单独取出,进而给出相应的活跃时间的间隔分布如图(e)所示.

时序网络较传统网络不仅拓扑结构随时间动态演化,而且带有个体交互时间的交互序列还显示出个体交互时间间隔分布的幂律特性,也称为个体交互的爆发特性(bursty behavior),即个体在长时间等待后的交互往往更为频繁[25]. 在时间维度上,爆发特性是时序网络区别于传统静态网络的一个典型特征. 研究发现,相较于传统静态网络,时序网络上的信息扩散[8-9]、传染病传播[10]、博弈策略演化[22-23]以及系统能控性[7,24,26-29]都将产生本质的变化. 如同利用小世界模型、无标度模型构建具有真实网络拓扑结构特性的小世界、无标度网络一样,对拓扑结构为时序网络的复杂系统的探索在根本上依赖于此类随时间动态演化的拓扑结构的构建. 需要注意的是,时序网络的构建必须满足两类本质特性:针对时序网络的静态聚合网络满足传统静态网络的典型特征;时序网络自身所特有的个体交互爆发特性等.

本文首先介绍构建具有传统静态网络典型特性的重要模型与方法;然后系统阐述对真实数据幂律分布参数估计与检验的有效方法;最后通过引入泊松过程和排队系统给出针对动态时序网络典型特征的几类时序网络构建理论和方法.

1　构建无标度网络的方法

自发现真实复杂网络中存在无标度特性以来,多种构建具有此类特性的无标度网络方法被相继提出,其中具有代表性的有:优先链接模型(preferential attachment model)[3]、静态模型(static model)[30]、配置模型(configuration model)[31]、Bianconi-Barabási模型[32-33]. 本节主要对这4种典型方法进行简要介绍.

1.1　优先链接模型

优先链接模型于1999年提出[3],该模型假设现实世界的网络结构具有生长和偏好链接两个基本性质,即网络规模通过添加新节点不断增长且新加入节点倾向于与度较高的已有节点相连. 以学术论文引用网络为例,该网络中学术论文作为网络节点,若两篇论文存在引用关系则此节点间建立一条连边;随着新的论文不断发表,网络规模随着新节点的加入不断增大,同时新论文更倾向于引用影响力较大的论文(此处指节点度较高的节点). 基于优先链接模型构造无标度网络的具体步骤如下(假设初始时刻网络中有m_0个节点随机连接,且不存在孤立节点).

step 1: 生长. 在每个时间步t,当前网络中新加入一个节点,且该节点选择与m个已有节点相连,其中$m < m_0$.

step 2: 偏好链接. 在新节点选择已有节点建立连

边时,选中节点i的概率为

$$p_i=\frac{k_i}{\sum\limits_j k_j}.$$

其中:k_i为节点i的度,$\sum\limits_j k_j$为网络中已有节点的度之和. 由上式可知,度越大的节点与新节点相连的概率越大,越有可能产生新链接.

经过T个时间步后,可得节点数$N=m_0+T$与链接数$V=l_1+mT$的新网络,其中l_1为初始时刻网络中的链接数. Barabási等从数学上证明了基于优先链接模型生成的网络中节点的度服从幂律分布,且幂指数$\gamma=3$,即生成的网络为无标度网络.

优先链接模型是一种较为简单的无标度网络生成模型,目前已在数学、物理学、计算机科学等多领域得到了广泛应用. 从优先链接模型中可以看出,度越大的已有节点与新加入节点建立连边的可能性越大,此举无疑进一步使其自身的度增加. 这很好地解释了现实世界中“富者越富”的马太效应. 然而,优先链接模型中生长和偏好链接假设具有一定的局限性,例如节点数目固定的网络不具备生长性质,此时优先链接模型将不再适用.

1.2 静态模型

静态模型由Goh等[30]于2001年提出,该模型给出了节点数N固定时生成无标度网络的方法,具体步骤如下.

step 1: 对网络中的N个孤立节点依次赋予编号$i\,(i=1,2,\dots,N)$,并基于节点编号同时赋予节点i一个权重

$$p_i=i^{-\alpha},$$

其中$\alpha\in[0,1)$可用于调整最终生成无标度网络的幂指数.

step 2: 分别以概率$p_i\Big/\sum\limits_k p_k$和$p_j\Big/\sum\limits_k p_k$选中节点$i$和$j$, 若此节点对$(i,j)$间不存在连边,则在该节点对间建立一条连边.

step 3: 重复step 2直到网络中有mN条连边,此时网络的平均度为$2m$.

下面分析由静态模型生成的网络中节点的度分布. 由构建网络的具体步骤可以看出,节点i的度k_i满足

$$\frac{k_i}{\sum\limits_j k_j}=\frac{\sum\limits_j p_ip_j}{\sum\limits_w\sum\limits_z p_wp_z}=\frac{p_i}{\sum\limits_j p_j}.$$

根据级数公式$\sum\limits_j p_j\approx\dfrac{N^{1-\alpha}}{1-\alpha}$,可得

$$\frac{k_i}{\sum\limits_j k_j}\approx\frac{1-\alpha}{N^{1-\alpha}i^{\alpha}}. \tag{1}$$

由于网络的平均度为$2m$,即$\sum\limits_j k_j=2mN$,将其代入式(1)可得

$$k_i\approx 2m(1-\alpha)N^{\alpha}i^{-\alpha}. \tag{2}$$

进一步,基于式(2)可得网络中节点的度分布为

$$P(k_i\leqslant k)\approx P\Big(i^{\alpha}\geqslant\frac{2m(1-\alpha)N^{\alpha}}{k}\Big).$$

求解上式可看作求解满足不等式

$$i^{\alpha}\geqslant\frac{2m(1-\alpha)N^{\alpha}}{k}$$

的变量i的数目占总数N的比例,因此上式可求解为

$$P\Big(i^{\alpha}\geqslant\frac{2m(1-\alpha)N^{\alpha}}{k}\Big)=\frac{N-k^{-\frac{1}{\alpha}}(2m(1-\alpha))^{\frac{1}{\alpha}}N}{N}=1-k^{-\frac{1}{\alpha}}(2m(1-\alpha))^{\frac{1}{\alpha}}. \tag{3}$$

最后通过式(3)可得网络中节点的度服从幂律分布,其概率密度函数(probability density function, 记为PDF)为

$$P(k)\sim k^{-\gamma},$$

其中幂指数为$\gamma=(1+\alpha)/\alpha$. 由于参数$\alpha\in[0,1)$导致幂指数满足$2<\gamma<\infty$,可以通过调节参数α得到不同的幂指数γ.

相比于优先链接模型,静态模型不要求网络具有生长的性质,可以在给定网络节点数的条件下生成无标度网络. 对于现实世界中的大量网络,如电力网络、公路网络等,一旦建成便不会有大规模改变,此时若想研究网络上的负载、流量等相关问题,采用静态模型建立网络结构比优先链接模型更简洁实用. 基于静态模型生成无标度网络时,仅能调节幂指数这一参数,由式(2)可以看出,所得到的度序列较为单一. 然而度服从同一幂律分布的网络度序列可能是不同的,因此静态模型具有一定的局限性.

1.3 配置模型

无标度网络与小世界网络、随机网络最大的区别之一是无标度网络中节点的度服从幂律分布,而小世界网络、随机网络中节点的度服从泊松分布. 对于任意给定的度序列,如何生成对应网络具有重要的理论意义和应用价值. 配置模型是一种基于度序列生成网络结构的普适性方法[31]. 假设网络中有N个节点及其度序列$k_1,k_2,\dots,k_N$,基于配置模型构建具有

目标度序列网络的具体步骤如下.

step 1: 对网络中N个节点依次赋予编号$i(i=1,2,\ldots,N)$及相应的度k_i.

step 2: 从每个节点i引出k_i条连边,此时$\sum_i k_i = 2m$,其中m为总连边数.

step 3: 随机选择两个节点,将它们引出的且未连的边连在一起建立网络中一条连边.

step 4: 重复step 3,直至所有节点引出的连边均形成网络中的连边.

对于任意的度序列$k_1,k_2,\ldots,k_N$,利用配置模型可以生成相应的网络. 当度序列服从幂律分布时,生成的网络为无标度网络. 配置模型是一种具有普适性的网络生成模型,既可用于生成无标度网络,也可用于生成随机网络、小世界网络等任意结构的网络,在不同领域有着广泛应用. 配置模型作为一种抽象普适的数学模型,凭借其理论简单、效果显著而深受欢迎. 真实世界中有限的网络数据往往不具有代表性,因此仅仅基于真实数据难以得到具有一般意义的结论. 在同一组度序列下,通过配置模型生成大量具有相同度序列的网络同时进行研究可以得到具有普适性的结论. 然而,配置模型的不足之处在于现实意义不充分、缺乏足够的直观解释、更多被用于理论研究.

1.4 Bianconi-Barabási模型

在真实网络系统中,不同节点间的品质、特性等往往存在较大的差异,可能影响网络的结构、度分布等性质. 针对此情况,基于网络节点的适应度(fitness), Bianconi和Barabási[32-33]提出了一种更具现实意义的网络增长模型,称为Bianconi-Barabási模型. 具体地,假设初始时刻网络中有m_0个节点,基于该模型构建网络的步骤如下.

step 1: 在每个时间步t,当前网络中新加入一个节点,且该节点选择与m个已有节点相连,其中$m<m_0$.

step 2: 每个节点i生成时(包括初始时刻),赋予其一个适应度η_i,所有节点的适应度均从分布$\rho(\eta)$中抽取,给定后不再发生改变.

step 3: 新节点添加时,与已有节点连接的概率正比于函数$a_k(\eta)$,其中$a_k(\eta)$与已有节点的度k和适应度η相关.

下面分析由Bianconi-Barabási模型生成的网络中节点的度分布. 特别强调,函数$a_k(\eta)$可以灵活地取自不同的数学形式[32-33]. 这里以$a_k(\eta)=\eta k$为例进行说明,当$a_k(\eta)=\eta k$时,已有节点i与新节点相连的概率定义为

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}.$$

由每个新节点加入时网络新增m条边可知,已有节点i的度k_i的变化满足

$$\frac{\partial k_i}{\partial t} \approx E\Big(\frac{k_i(t+1)-k_i(t)}{t+1-t}\Big) = m\Pi_i = m\frac{\eta_i k_i}{\sum_j k_j\eta_j}. \tag{4}$$

为了进一步分析节点度k_i的分布,不妨先假设其近似服从幂律分布

$$k_{\eta_i}(t,t_0) = m\Big(\frac{t}{t_0}\Big)^{\beta(\eta_i)}, \tag{5}$$

其中t_0为节点i产生的时刻,且$0<\beta(\eta)<1$. 为了求解式(5)对应$\partial k_i/\partial t$的解析式,首先需要计算$\sum_j \eta_j k_j$的期望

$$\Big\langle \sum_j \eta_j k_j \Big\rangle = \int \rho(\eta)\eta \mathrm{d}\eta \int_1^t k_\eta(t,t_0)\mathrm{d}t_0 = \int \rho(\eta)\eta\mathrm{d}\eta\, m\frac{(t-t^{\beta(\eta)})}{1-\beta(\eta)}.$$

当$t\to\infty$时,有

$$\Big\langle \sum_j \eta_j k_j \Big\rangle \overset{t\to\infty}{=} Cmt(1+O(t^{-\epsilon})), \tag{6}$$

其中参数ϵ、C分别为

$$\epsilon = (1-\max_\eta \beta(\eta)) > 0,$$

$$C = \int \rho(\eta)\frac{\eta}{1-\beta(\eta)}\mathrm{d}\eta.$$

记$k_\eta = k_{\eta_i}(t,t_0)$,将式(6)代入(4),有

$$\frac{\partial k_\eta}{\partial t} = \frac{\eta k_\eta}{Ct}.$$

由此可得$\beta(\eta)=\eta/C$. 对于某一特定的节点$k_\eta(t)$,有

$$P(k_\eta(t)>k) = P\Big(t_0 < t\Big(\frac{m}{k}\Big)^{C/\eta}\Big) \propto t\Big(\frac{m}{k}\Big)^{\frac{C}{\eta}}.$$

因此,网络中节点度的概率密度函数为

$$P(k) = \int_0^{\eta_{\max}} \frac{\partial P(k_\eta(t)>k)}{\partial t}\mathrm{d}\eta \propto \int \rho(\eta)\frac{C}{\eta}\Big(\frac{m}{k}\Big)^{\frac{C}{\eta}+1}\mathrm{d}\eta. \tag{7}$$

由式(7)可知,网络中节点的度是否服从幂律分布取决于分布$\rho(\eta)$的具体形式. 若$\rho(\eta)$为常数,则相应的度分布为幂律分布. 然而,由于η可以取自多种形式的分布,基于Bianconi-Barabási模型生成网络的度可能不服从幂律分布.

Bianconi-Barabási模型具有较强的现实意义和应用价值,传统的优先链接模型暗含了新加入节点的

度无法高于已加入节点的度这一假设. 事实上,这一假设在现实世界中往往很难成立,如新发表的创新性更高的论文其被引次数很快会超过多数已存在的创新性不够高的论文、社交媒体中新入驻名人的粉丝数也可能很快超过多数老用户等. Bianconi-Barabási模型开创性地引入了节点适应度这一概念,从而使得模型能够更好地刻画这种真实世界的网络情况,即允许节点度“后来居上”这种现象出现. 同时Bianconi-Barabási模型在特定情况下可以退化为优先链接模型,由此可见Bianconi-Barabási模型更具普适性.

优先链接模型通过网络生长和偏好链接机制保证了节点的先到者优势,即先加入网络中的节点越容易获取更多的连边. 该模型形式简单,但其生成网络的幂律度分布指数理论上仅为-3. Bianconi-Barabási模型在优先链接模型的基础上引入了节点适应度这一特征,使网络中可以出现节点度“后来居上”的现象,且所得网络幂律度分布的指数可变. 静态模型给出了节点数固定时生成无标度网络的方法,不需要网络不断生长且生成网络度分布的幂指数可变. 然而,该模型生成网络的度序列是固定的,且连通性往往无法保证. 配置模型可以生成固定节点数的具有任意度序列的网络,是一种具有普适性的网络生成模型,既可用于生成无标度网络,也可用于生成随机网络. 然而,该模型的现实意义不充分,缺乏足够的直观解释,更多被用于理论研究.

2 幂律分布检验

自复杂网络的无标度特性被发现以来,无标度网络在现实世界中普遍存在已是公认的事实. 然而,无标度网络和幂律分布的关系至今仍处于争论之中[34-38]. 以Barabási为代表的专家学者将网络中节点度满足幂律分布的网络称为无标度网络(幂指数介于2与3之间),而Brodio等[34]在分析大量真实世界的经验数据后得出了无标度网络在真实世界中并不常见这一结论,对无标度网络理论产生一定的质疑. 目前接受度最高的说法是真实世界中的网络数据往往存在大量噪声与偏差,进而影响对网络中节点度的幂律分布在严格数学定义上的精确检验,正如无法在现实世界中精确验证万有引力定律一样[36]. 文献[38]认为,无标度特性检验的重点在于度分布是否为重尾(heavy-tailed)分布.

无标度网络和幂律分布的关系有待进一步验证,而Clauset等[39]在研究中使用的幂律分布检验方法在数学上值得借鉴学习. 对于数据幂律分布的判断往往依赖于对数据的定性评价,例如基于可视化的评价. 但定性研究缺少定量的实证检验,得到的结论可能具有欺骗性并导致对幂律行为的判断与实际相背离,定量的数据检验结果在数学上具有更高的可信度. 本部分将对该方法进行详细介绍.

在实践中,经验数据并非严格服从幂律分布,通常情况下,幂律分布只适用于大于某个最小值$x_{\min}$的观测数据,即只有经验数据的尾部服从幂律分布. 数学上,幂律分布通常有两种形式: 1) 连续分布,自变量为连续实数; 2) 离散分布,自变量取一组离散值. 对于连续的幂律分布,其概率密度函数$P(x)$满足如下形式:

$$P(x)=\frac{\alpha-1}{x_{\min}}\Big(\frac{x}{x_{\min}}\Big)^{-\alpha}.$$

其中: $\alpha>1, x\geqslant x_{\min}>0$. 对于离散情况,当仅考虑变量取整数值时,其概率函数为

$$P(x)=\frac{x^{-\alpha}}{\zeta(\alpha,x_{\min})}.$$

其中: $\zeta(\alpha,x_{\min})=\sum\limits_{n=0}^{\infty}(n+x_{\min})^{-\alpha}$, $\alpha>1$, $x\geqslant x_{\min}>0$. 对于幂律分布而言,互补累积分布函数(complementary cumulative distribution function, CCDF)即$F(x)=P(X\geqslant x)$相比于PDF对有限样本量造成的波动更加稳健,特别是在分布的尾部. 连续情况下幂律分布的CCDF为

$$F(x)=\Big(\frac{x}{x_{\min}}\Big)^{-\alpha+1}.$$

离散情况下幂律分布的CCDF为

$$F(x)=\frac{\zeta(\alpha,x)}{\zeta(\alpha,x_{\min})}.$$

Clauset等[39]提出了一个对幂律分布定量检验的基本方法.

1) 利用幂律分布拟合经验数据: 估计幂律分布中的参数α和最小值$x_{\min}$.

2) 进行拟合优度检验: 计算经验数据与幂律分布之间的拟合优度,若计算得到的拟合优度值$p>0.1$,则幂律分布是经验数据的一个合理假设,否则拒绝该假设.

3) 备择分布似然比检验: 通过似然比检验将幂律分布与备择分布进行比较,对于任一备选分布,如果计算得到的对数似然比明显偏离0,则可通过对数似然比的符号判断该分布是否优于幂律分布.

相比于离散分布,连续分布的数学表达更加简单,因此实际检验中通常用连续幂律分布近似离散幂律分布. 但近似方法的选择至关重要,一种可靠的方法是将离散幂律分布看作是由连续幂律分布生成的,即将连续幂律分布中的值四舍五入到整数. 下面

就上述幂律分布定量检验的基本方法展开详细介绍.

2.1 数据拟合

首先考虑对尺度参数α的估计. 由于经验数据中的幂律行为存在下界,即前文提到的$x_{\min}$,这里假设该下界已知(后文给出$x_{\min}$的估计方法). 拟合幂律分布到经验数据的方法是极大似然法,该方法在大样本容量的极限下可以给出准确的参数估计. 假设经验数据对于$x \geqslant x_{\min}$严格遵循幂律分布,基于此可以分别得到离散和连续情况下尺度参数α的极大似然估计$\hat{\alpha}$. 在连续情况下,利用极大似然估计理论可得参数α的估计值为

$$\hat{\alpha} = 1 + n\Big(\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\Big)^{-1},$$

其中$x_i\,(i = 1, 2, \ldots, n)$为经验数据集中大于$x_{\min}$的值. 此时估计的标准差为

$$\sigma = \frac{\hat{\alpha} - 1}{\sqrt{n}} + O\Big(\frac{1}{n}\Big).$$

在离散情况下对参数α的极大似然估计为如下方程的解:

$$\frac{\zeta'(\hat{\alpha}, x_{\min})}{\zeta(\hat{\alpha}, x_{\min})} = -\frac{1}{n}\sum_{i=1}^{n} \ln x_i.$$

通过连续幂律分布近似离散分布,可得离散情况下α的极大似然估计的近似值为

$$\hat{\alpha} \cong 1 + n\Big(\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min} - \frac{1}{2}}\Big)^{-1}.$$

下面给出$x_{\min}$的估计方法,$x_{\min}$估计的准确性对幂律分布的检验至关重要. 如果估计值$\hat{x}_{\min}$偏低,则将引入较多非幂律噪声数据,使得检验结果出现较大偏差;如果估计值$\hat{x}_{\min}$偏高,则会损失较多的有效信息,从而增加参数α估计的统计误差. 对此,Handcock等[40]提出了一种离散情形下估计$x_{\min}$的方法,该方法使用一个广义模型刻画所有的经验数据. 对于一个估计值$\hat{x}_{\min}$,所有大于$\hat{x}_{\min}$的数据分布由一个标准离散幂律分布拟合;每个小于$\hat{x}_{\min}$的数据分布由一个单独的概率p_x刻画,其中$1 \leqslant x < \hat{x}_{\min}$,$p_x$为观测数据中$x$出现的频率. 因此每个不同的$\hat{x}_{\min}$对应一个参数个数为$\hat{x}_{\min}$的广义模型. 在参数个数不定的情况下,增大参数个数可以实现更高的似然值,但是会产生过拟合的风险,因此采用贝叶斯信息(BIC)准则对参数个数加以惩罚,即

$$\ln P(x|x_{\min}) \simeq \mathcal{L} - \frac{1}{2}x_{\min} \ln n,$$

其中$\mathcal{L}$为传统的极大对数似然值. 最后通过最大化上述BIC准则求得$\hat{x}_{\min}$. 该方法在一些情况下可以获得较好的效果,但也可能出现困境. 例如,该方法需要$x_{\min} - 1$个参数对$x_{\min}$以下的数据建模,往往使得BIC低估$x_{\min}$的值,进而导致随后估计的尺度参数值α出现偏差. 更重要的是,该方法无法推广到连续数据的情况.

随后,Clauset等[41]提出了一种在离散和连续情形下均适用的方法,其基本思想为:选择$\hat{x}_{\min}$使得经验数据中大于$\hat{x}_{\min}$的数据尽可能服从幂律分布,即经验数据与拟合得到的幂律分布间的距离尽可能小. 通过采用Kolmogorov-Smirnov(记为KS)检验方法可以量化两个非正态分布之间的距离,此时经验数据和拟合模型的互补累计分布函数的最大距离D可表示为

$$D = \max_{x \geqslant x_{\min}} |S(x) - Q(x)|.$$

其中:$S(x)$为经验数据中大于等于$x_{\min}$的值的互补累积分布函数,$Q(x)$为拟合得到的幂律分布的互补累积分布函数. 此时最优的估计值$\hat{x}_{\min}$为使得距离D最小的$x_{\min}$值.

2.2 拟合优度检验

对于任意数据集,均可以采用上述方法将其拟合到幂律分布,并给出参数α和下界$x_{\min}$的具体估计,然而此方法无法判断该数据集是否真正服从幂律分布. Clauset等[39]提出了一种检验经验数据是否服从幂律分布的方法,具体步骤如下.

step 1: 使用数据拟合方法将经验数据拟合得到幂律分布,并计算经验数据和拟合分布对应的KS统计量.

step 2: 根据拟合的幂律分布,通过采样得到大量合成数据集,将每个合成数据集分别拟合到对应的幂律分布,并计算每个合成数据集对应的KS统计量.

step 3: 计算KS统计量大于经验数据KS统计量的合成数据集数量ψ与合成数据集总数ϕ之比$p = \psi/\phi$. 若p值接近1,则接受经验数据服从幂律分布假设;若p值接近0,则拒绝该假设.

对于每个合成数据集,分别拟合其对应的幂律分布以计算KS统计量,而非根据经验数据集的原始分布进行计算. 通过这种方式,可以确保对每个合成数据集执行的计算与对真实数据集执行的计算相同,从而得到p值的无偏估计. 另外,合成数据集的采样准确度对检验经验数据是否服从幂律分布至关重要,因此合成数据集应满足如下要求:合成数据小于$\hat{x}_{\min}$时应与经验数据的分布接近,大于$\hat{x}_{\min}$时应服从幂律分布. 利用半参数法可以生成符合要求的数据,假设经验数据集中共有n个观测值,其中有n_{tail}个大于$\hat{x}_{\min}$的值,以n_{tail}/n的概率从参数为$\hat{\alpha}$的幂律分布中

选择大于$\hat{x}_{\min}$的值,以$1-n_{\text{tail}}/n$的概率随机选择一个小于$\hat{x}_{\min}$的值,将此过程重复n次得到一个合成数据集. 对于合成数据集的数量,根据对测试情况分析,如果希望p值的精度在ϵ以内,则至少生成$\epsilon^{-2}/4$个数据集. 通过计算得到的p值决定是否拒绝幂律假设,一般而言,如果p值小于等于0.1则拒绝幂律假设.

2.3 备择分布似然比检验

通过拟合优度检验可以检验经验数据是否服从幂律分布,当检验通过时,则接受幂律假设,然而是否存在其他分布能够更好地拟合经验数据是进一步必须考虑的问题. 为解决该问题,Clauset等[39]提出了似然比检验方法,其基本思想是计算数据集在两种分布下的似然度,似然度更高的分布与经验数据的拟合程度更高. 等价地,计算两种分布的似然比并取其对数,然后通过该对数值的符号对比两种分布拟合的准确性. 对数似然比具体计算方法如下:对于概率密度分别为$p_1(x)$和$p_2(x)$的两种不同分布,给定的经验数据集在两种分布上的似然度L_1和L_2分别为

$$L_1=\prod_{i=1}^{n}p_1(x_i),\ L_2=\prod_{i=1}^{n}p_2(x_i).$$

相应地两分布的似然比R为

$$R=\frac{L_1}{L_2}=\prod_{i=1}^{n}\frac{p_1(x_i)}{p_2(x_i)}. \tag{8}$$

对式(8)两端同时取对数得到对数似然比为

$$\ln R=\sum_{i=1}^{n}(\ln p_1(x_i)-\ln p_2(x_i)).$$

通过对数似然比判断备选分布与幂律分布的优劣性时,由于不可避免的统计误差影响,不能仅凭对数似然比的符号下结论. 为定量判断对数似然比是否可信,需要计算对数似然比的标准差以反映数据的波动程度. 记

$$l_i=\ln p_1(x_i)-\ln p_2(x_i).$$

根据中心极限定理,当样本量n足够大时,$\ln R$服从正态分布,故对数似然比的标准差σ满足

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(l_i-\bar{l})^2,$$

其中$\bar{l}$为所有l_i的均值. 为判断$\ln R$的符号是否受数据波动影响,计算对数似然比的分布中绝对值大于等于$|\ln R|$的概率p'为

$$p'=\frac{1}{\sqrt{2\pi n\sigma^2}}\Big(\int_{-\infty}^{-|\ln R|}\mathrm{e}^{-t^2/2n\sigma^2}\mathrm{d}t+\int_{|\ln R|}^{+\infty}\mathrm{e}^{-t^2/2n\sigma^2}\mathrm{d}t\Big).$$

当$p'<0.1$时,一般认为检验结果是可信的,可以利用备择分布似然比检验方法判断哪种分布与经验数据的拟合效果更好;当$p'\geqslant 0.1$时,常常认为该似然比检验不足以区分两种分布的优劣.

3 爆发特性

3.1 真实数据的典型特征

当事件的发生随着长时间的静默与短暂的频繁活跃交替进行时,该事件发生的统计特性称为爆发特性. 在21世纪初期之前,许多研究都是基于泊松分布近似分析网络的度分布、网络上的动力学、人类行为等[42-43],自Barabási等[3]发现真实世界中无标度网络普遍存在后,幂律分布逐渐进入人们的视野,自然界中的事件发生规律更多地服从幂律分布还是泊松分布?这一问题引起了学术界持续广泛的关注[44].

Vázquez等[45]对以下5个数据集进行了深入的研究:

1) 网页浏览数据集[46-47]. 该数据集对匈牙利最大的新闻和娱乐门户网站中大约25万独立访客的浏览历史进行了搜集,主要整理了2002年8月11日~2002年8月12日期间每个访问者的访问记录,并以s为单位记录各访问者的每次下载时间. 在该数据集中,事件间隔时间(inter-event time)定义为同一用户下载或者点击网页的时间间隔.

2) 邮件收发数据集[48-49]. 该数据集对某大学个人电子邮件的收发记录进行了搜集,主要整理了3 188和9 665个用户在3个月和6个月内的发件人、收件人以及每封电子邮件的发送时间,以s为单位共记录了129 135封电子邮件. 在该数据集中,事件间隔时间定义为同一用户发送电子邮件的时间间隔.

3) 图书借阅数据集. 该数据集对圣母大学中教师从图书馆借阅书籍的记录进行了搜集,主要整理了2 247名教师在3年内共进行的48 409次借阅活动,并以min为单位进行记录. 在该数据集中,事件间隔时间定义为同一读者借阅图书或期刊的时间间隔.

4) 贸易交易数据集. 该数据集对1999年6月~2003年5月之间一家中欧银行股票经纪人发起的所有“买/卖”交易记录进行了搜集,主要整理了每天从7:00到19:00的交易记录,共54 374笔交易,并忽略当天结束时最后一笔交易与下一个交易日开始时第一笔交易的时间间隔. 在该数据集中,事件间隔时间定义为经纪人交易的时间间隔.

5) 爱因斯坦、达尔文和弗洛伊德各自的信件往来数据集[50-52]. 该数据集对3位科学家一生中所有信件的发件人、收件人、日期记录进行了搜集,主要整理了:达尔文发出的7 591封信件和收到的6 530封信件,共计14 121封;爱因斯坦发出的14 512封信件

和收到的16 289封信件,共计30 801封;弗洛伊德发出的3 183封信件和收到的2 675封信件,共计5 858封. 在该数据集中,事件间隔时间定义为3位科学家收到某人来信与回复该人信件的时间间隔.

通过对上述5个数据集的分析发现,Vázquez等[45]发现事件间隔时间τ呈现出一种爆发特性,即其服从幂律分布

$$P(\tau) \sim \tau^{-\alpha}.$$

对于网页浏览、邮件收发、图书借阅3个数据集,对应的$\alpha \approx 1$;对于爱因斯坦、达尔文和弗洛伊德各自的信件往来数据集,对应的$\alpha \approx 3/2$;而对于贸易交易数据集,对应的α位于两者之间. 由于贸易交易数据集包含的数据较少,具有偶然性,Vázquez等主要对前两种情况进行分析,并基于排队论的思想给出如下数学结论:

1) 当任务的队列长度有限,即同一时刻任务数量有限时,单个任务的等待时间服从参数为1的幂律分布,即$\alpha = 1$.

2) 当任务的队列长度没有限制,即个体在任意时刻的任务数量没有限制时,单个任务的等待时间服从参数为$3/2$的幂律分布,即$\alpha = 3/2$.

真实世界中的事件间隔时间往往呈现爆发特性,用幂律分布来刻画更为准确,因此之前基于泊松分布得到的结果有待进一步验证.

3.2 爆发特性生成机制

3.2.1 爆发特性与泊松过程

将事件到达速率为λ的泊松过程中相邻事件间隔时间记为τ,则τ服从参数为λ的指数分布,即

$$P(\tau) = \lambda \mathrm{e}^{-\lambda\tau},\ \tau > 0.$$

随着间隔时间τ的增大,其出现的概率呈指数下降,导致泊松过程中的事件间隔时间彼此差异较小,几乎不会出现差异较大的间隔时间,这实际上是无记忆性的体现. 若一个随机变量X是无记忆的,则有

$$\forall s,\ t \geqslant 0,\ P(X > s + t | X > t) = P(X > s).$$

记此随机变量的累计分布函数为$F(x)$,则等价地有$1 - F(s + t) = (1 - F(s))(1 - F(t))$. 而对于定义在非负实数域上的$f(x)$,若$f(s + t) = f(s)f(t)$,则该方程的右连续解只有$f(x) = \mathrm{e}^{-\lambda x}$,其中$\lambda$为任意常数,这使得满足无记忆性的非负连续随机变量只能服从指数分布.

对于具有爆发特性的过程,事件间隔时间具有明显的差异性. 在一段时间内事件频繁发生,间隔时间较短;此后可能长时间内无事件发生,该现象与泊松过程事件产生的模式有明显区别. 泊松过程的事件间隔时间服从指数分布,而通常具有爆发特性的过程中事件间隔时间服从幂律分布,即$P(\tau) \sim \tau^{-\gamma}, \gamma > 0$. 幂律分布在长时间间隔处的衰减速度远小于指数分布,方差也会较大. 若分布在非负实轴上取非零值且$\gamma < 3$,则此时分布方差的理论值为无穷,这表明不仅长间隔时间出现的概率增高,并且间隔时间之间的差异性也会增大.

3.2.2 间隔时间与排队系统

在描述同一种事件发生规律的情况下,由于事件间隔时间和等待时间(waiting time)描述对象相同,一般对二者不作区分,但对于描述多种事件情形,需对二者进行区分对待[53]. 在某人收发电子邮件的例子中,以此人回复一封邮件为事件,他回复一封邮件到回复下一封邮件的时间差称为事件(回复邮件)间隔时间. 而对于一封具体的邮件,它被接收到被回复之间的时间差称为此邮件的等待时间. 一般而言,事件间隔时间多用于描述两个具体动态的事件,等待时间描述维持某一状态的时长,很多情况下二者可用于描述同一对象.

探究爆发特性的生成机制本质上是探索事件间隔时间的规律. 为此,对排队系统及等待时间相关的模型和结论进行简要介绍,首先介绍排队系统的基本组成部分如下:

1) 服务器: 指维护队列并执行任务的个体. 在排队论中,可以让多台服务器同时执行任务,而人的行为建模中一般假定为单一服务器,即个体只负责执行自己列表上的任务.

2) 任务到达规律: 排队论中通常假设任务的到达是一个参数为λ的泊松过程,即各任务彼此独立且以速率λ到达队列.

3) 服务过程: 指定执行单个任务所需时间,并假设为参数μ的泊松过程,从而各任务的时间间隔服从指数分布. 另外,假设服务时间与任务到达过程或优先列表上的任务数量无关.

4) 选择协议/队列规则: 指定队列中任务选择执行的方式,刻画人类行为中常见的选择协议为先到先得(first in first out, FIFO),而后进先出(last in first out, LIFO)在存储系统中较为常见.

5) 队列长度: 一般排队模型假设队列容量是无限的,且队列长度可以根据单个任务的到达和执行速度动态变化. 在某些排队模型中,队列长度受物理限制,例如,人类面对太多任务时会放弃部分或不再接受其他任务,或者由于即时记忆能力有限导致任务列

表长度有限.

3.2.3 优先级排队模型

优先级排队模型是一类与等待时间相关的系统模型,其基本思想为:每个顾客到达队列时赋予其一个优先级,系统按照当前队列中顾客的优先级从高到低的顺序对顾客进行服务. 根据优先级的分布形式可将模型分为离散优先级模型和连续优先级模型.

1) 离散优先级模型.

Cobham[54] 首先研究了 $M/M/1$ 系统在优先级 $p\,(p=1,2,\ldots,r)$ 取有限离散值模式下队列中优先级为 p 的顾客的平均等待时间. 基于总是优先级最高的顾客被系统服务,相同优先级的顾客按照到达的先后顺序服务的假定,Cobham 得到了平均等待时长 τ 与到达率 λ 和处理率 μ 之间的理论关系式. 特别地,Abate 等[55] 得到了 $r=2$ 时低优先级顾客等待时间的累积分布函数 $\tilde{P}(\tau)$ 的近似表达式

$$1-\tilde{P}(\tau)\sim\tau^{-\alpha}\mathrm{e}^{-\beta\tau},$$

其中 $\alpha=3/2$. 在一些特殊参数选择下,也可以得到 α 为0或0.5,β 为 λ 和 μ 的关系式.

2) 连续优先级模型.

Cohen[56] 给出了 $M/G/1$ 系统在优先级从连续分布 $\rho(x)=U[0,1]$ 中随机采样的模式下,等待时间 τ 与优先级 p 的联合分布密度 $W(\tau,p)$ 的积分表达式.

对于连续优先级模型,Vázquez 等[45] 给出了不同条件下的数值分析,记任务到达速率和任务处理速率之比 $\rho=\lambda/\mu$ 为流量强度,有:

① 亚临界状态 $\rho=\lambda/\mu<1$:若任务到达速率 λ 小于处理速率 μ,即任务到达队列后都能很快被执行,则多数时候队列中不存在任务. 仿真结果表明,当 $\rho<1$ 时等待时间分布呈现指数衰减,当 $\rho\to1$ 时等待时间呈现尺度参数 $\alpha=3/2$ 的具有指数截断的幂律形式.

② 临界状态 $\rho=\lambda/\mu=1$:若任务到达速率 λ 等于处理速率 μ,则队列长度 l 以0为界随机波动,优先级最低的任务会等到队列长度 $l=0$ 时才会执行,此时随机波动的返回时间 τ 服从 $P(\tau)\sim\tau^{-3/2}$.

③ 超临界状态 $\rho=\lambda/\mu>1$:若任务到达速率 λ 大于处理速率 μ,则队列长度与时间存在线性关系,$\langle l(t)\rangle=(\lambda-\mu)t$ 成立. 超临界状态存在永远不会被解决的任务,使得大部分任务会永远处在队列中,等待时间为无穷长,并且由于一些等待时长无法统计,导致统计到的等待时间分布集中于较短的时间间隔.

假设每个个体有一个长度为 L 的任务清单,每个任务被赋予一个优先级 $x_i\,(i=1,2,\ldots,L)$,且这些优先级均从同一个分布 $\rho(x)$ 中独立随机采样得到. 基于此,Barabási 提出了优先级队列模型(称为Barabási优先级模型)[25,27]:

1) 在每个离散时间步,个体都选择其任务清单上优先级最高的任务并在该时间步立即完成,然后将此任务移出清单;

2) 随即添加一项新的任务进入清单,保证清单长度 L 不变,新添加任务的优先级仍从分布 $\rho(x)$ 中随机采样获得;

3) 进入下一时间步并重复上述步骤.

为了研究模型属性,首先将该模型进行拓展,选择任务的规则更改为:以固定概率 p 选择优先级最高的任务执行;以概率 $1-p$ 从 L 个任务中均匀随机选择一个任务执行. 对于该拓展模型,当 $p\to1$ 时任务选择准则为最高优先级执行准则,当 $p\to0$ 时为均匀随机执行准则. 下面将上述拓展模型视为采用概率不均匀的随机任务执行准则进而分析该模型的任务等待时长分布:每个时间步选择优先级为 x 的任务执行概率为 $\Pi(x)=x^{\gamma}\Big/\sum_{i}^{L}x_i^{\gamma}$,其中 γ 是一个可调整常数. 当 $\gamma=0$ 时,所有 L 个任务被选择执行的概率相同,对应上述 $p\to0$ 的情况;当 $\gamma=\infty$ 时,优先级最高的任务被选择执行的概率为1,其他任务为0,对应上述 $p\to1$ 的情况. 注意到,采取这种方式只能近似刻画 $p\to1$ 和 $p\to0$ 的情况,不适用其余 $0<p<1$ 的情况.

在此近似模型下,任务等待时间分布为

$$P(\tau)=\frac{1}{\gamma}\frac{\rho(\tau^{-1/\gamma})}{\tau^{1+1/\gamma}}.$$

当 $\gamma\to\infty$ 时,有 $P(\tau)\approx\tau^{-1}$,对应模型中最高优先级执行准则的情况. 当 $\gamma=0$ 时,τ 与 x 无关,有

$$P(\tau)=(1-\Pi)^{\tau-1}\Pi=\frac{1}{L}\Big(1-\frac{1}{L}\Big)^{\tau-1},$$

此时对于较大的 L 和 τ,该分布可近似为参数 $1/L$ 的指数分布.

特别地,Vázquez[57] 给出了队列长度 $L=2$ 时任务等待时间分布的数学推导,得到

$$P(\tau)=\begin{cases}1-\dfrac{1-p^2}{4p}\ln\dfrac{1+p}{1-p},\ \tau=1;\\[2ex]\dfrac{1-p^2}{4p}\Big[\Big(\dfrac{1+p}{2}\Big)^{\tau-1}-\dfrac{1-p}{2}^{\tau-1}\Big]\dfrac{1}{\tau-1},\\ \quad\tau>1.\end{cases}$$

对于上式,当 $p\to0$ 时,有

$$\lim_{p\to0}P(\tau)=\Big(\frac{1}{2}\Big)^{\tau};$$

当$p \to 1$时,有

$$\lim_{p\to 1} P(\tau) = \begin{cases} 1 + O\Big(\dfrac{1-p}{2}\ln(1-p)\Big), & \tau = 1; \\ O\Big(\dfrac{1-p}{2}\Big)\dfrac{1}{\tau-1}, & \tau > 1. \end{cases}$$

此时,几乎所有任务的等待时间均为$\tau = 1$,而没有被一步执行掉的任务等待时长服从幂律分布,且幂指数为-1.

4 典型的构建理论与方法

4.1 活跃度驱动模型

不同于静态网络生成模型,如优先链接模型、配置模型等,Perra 等[58] 提出的活跃度驱动模型(activity-driven model, AD模型)是一种简单的时序网络生成模型,即网络结构随时间动态变化,使网络节点间的交互存在时序信息. 活跃度驱动模型通过如下方式构建离散的时序网络序列: 假设网络规模固定为N,且每个节点$i\,(1 \leqslant i \leqslant N)$被赋予一个激活概率$a_i = \eta x_i$,此概率表示在单位时间内该节点与其他节点产生交互的概率. 其中η为归一化常数,确保每个节点在固定时间间隔Δt内的激活概率$a_i\Delta t \in [0,1]$,同时也可控制网络中单位时间内激活节点的平均数量; x_i从一个概率密度函数为$\rho(x)$的分布中抽取,且$\epsilon \leqslant x_i \leqslant 1$. 为了避免$\rho(x)$在$x = 0$处出现发散的情况,约束$x_i \geqslant \epsilon$. 每个节点的激活概率$a_i$分配后不再改变,具体的网络生成过程如下.

step 1: 在每个时间步t,网络以N个孤立节点为初始状态.

step 2: 每个节点i以相应激活概率$a_i\Delta t$变为活跃状态,以概率$1 - a_i\Delta t$处于非活跃状态,其中Δt为相邻两个时间步的固定时间间隔. 各个节点的激活过程相互独立,且每个处在激活状态的节点产生m条无向边,并随机选择m个其他节点相连.

step 3: 到下一个时间步$t + \Delta t$,网络仍以N个孤立节点为初始状态,并重复step 2.

处于非活跃状态的节点虽不能主动产生连边,但可以被其他活跃节点选择连接. 每个离散时间步形成的网络称为t时间步的网络切片(snapshot),合并网络切片序列形成的网络为静态聚合网络(图2(a)最上面部分). 即聚合网络中节点i和j存在连边,当且仅当存在某个时间步t,使得该时间步的网络切片中节点i与j间存在连边.

对于每个时间步t的网络切片,其平均度为

$$\langle k\rangle_t = 2m\langle a\rangle = 2m\eta\langle x\rangle,$$

其中$\langle a\rangle = \int \eta x\rho(x)\mathrm{d}x = \eta\langle x\rangle$为激活概率的均值. 记截止到$T$时刻形成的聚合网络为$G_T$,节点$i$在聚合网络中的度为$k_i(T)$,则当网络规模$N$充分大且$T/N$充分小时,可近似得

$$k_i(T) = N(1 - \mathrm{e}^{-Tm\eta x_i/N}). \tag{9}$$

由式(9)可以看出,聚合网络中节点的度与其激活概率为近似的线性关系. 直观地,一个节点的激活概率越高,其激活的平均次数与其链接的总边数也呈比例增高,故节点的度呈现近似线性增长的趋势. 根据节点的度及其激活概率之间的单调关系可知, $P_T(k)\mathrm{d}k \sim \rho(x)\mathrm{d}x$,从而得到

$$P_T(k) \sim \frac{1}{Tm\eta}\rho\Big(\frac{k}{Tm\eta}\Big).$$

因此,聚合网络的度分布与激活概率的分布基本相同,相差一个缩放系数. 为了得到具有幂律度分布的聚合网络,应选择激活概率分布$\rho(x) \propto x^{-\gamma}$,其中$\epsilon \leqslant x \leqslant 1, \gamma > 2$. 此时活跃度驱动模型生成的聚合网络度分布近似服从幂律分布,且幂指数为γ. 活跃度驱动模型生成网络结构的更多特性参见文献[59].

从活跃度驱动模型的网络生成机制可以看出,各个时间步生成的网络切片是相互独立的,导致真实世界网络切片间的相关性无法在该模型中体现. 另外,在每个离散时间步,网络中各节点i均以相同的激活概率$a_i\Delta t$变为活跃状态,与其他节点产生交互. 此时节点i在相邻活跃状态间的时间间隔服从几何分布,特别地,当$\Delta t \to 0$时近似服从指数分布,从而每个节点的激活可看作一个泊松过程,不具有爆发特性. 活跃度驱动模型中假设节点间交互的持续时长为固定间隔时间Δt,具有较强的约束性. 此外,聚合网络呈现幂律度分布的结果是建立在N充分大以及T/N充分小的基础上,若网络持续演化,即T不断增大,则最终聚合网络将收敛至平凡的N阶完全图. 因此,活跃度驱动模型虽然生成机制简单,但其具有较大的约束性,缺乏一般的时序网络交互机制与特征(如爆发特性).

4.2 记忆性活跃度驱动模型

活跃度驱动模型生成的时序网络中网络切片间是相互独立的,与真实数据集中反映出网络序列间的相关性不相符. 例如,在人类的行为活动中,人们出行更倾向于曾经到访过的地方[60],人们的社交活动更倾向于在自己的朋友间展开. 因此为了刻画网络序列间的相关性,后续一些研究对活跃度驱动模型进行拓展,使当前时间步形成的网络切片依赖于历史网络中的交互信息,特别是邻居节点信息. 这些网络生成模型称为记忆性活跃度驱动模型(activity-driven

model with memory).

在活跃度驱动模型的基础上,Karsai等[61]提出了记忆性扩展模型,具体的模型实现如下:将活跃度驱动模型的step 2改为

step 2': 在当前时间步t,记节点i在聚合网络G_{t-1}中的度为k_i,若以其激活概率$a_i\Delta t$变为活跃状态,则将连接m条边至其他节点,其中每条边以概率$c/(k_i+c)$随机连接非邻居节点集中的一个,以互补概率$k_i/(k_i+c)$随机连接一个曾经的邻居节点,其中$c>0$为模型参数,可根据具体数据而定. 对于该模型,在N充分大以及T/N充分小的条件下,聚合网络G_T的度分布形式并未改变,但网络序列间相关性使网络上的动力学行为发生显著改变. 例如,在含记忆性的网络中,网络最大连通分支规模的增长速度会显著慢于传统无记忆性活跃度驱动模型,网络上的谣言传播速度相比于活跃度驱动模型有明显下降. 该结论与大众普遍认知一致,因为网络节点更倾向与邻居节点交互,从而使得节点向外扩张的速度减慢,所以更多的交互会保持在邻居内部,而不倾向于向外扩散.

对于记忆性活跃度驱动模型,Kim等[62-63]进一步拓展,具体实现如下:将活跃度驱动模型的step 2改为

step 2": 在时间步t,若节点i以其激活概率$a_i\Delta t$激活且为第1次激活时,将与网络中的其他节点随机连接一条边. 若节点i并非第1次激活,则节点i以概率$As_i(t)^{-\beta}$随机连接一个未曾连接过的节点;以互补概率$1-As_i(t)^{-\beta}$选择一个曾经的邻居节点j连接,且选择节点j的概率为$\Pi_{ij}(t)=w_{ij}(t)\Big/\sum\limits_{l\in\mathcal{N}_i(t)}w_{il}(t)$. 其中:$\mathcal{N}_i(t)$为截止到时间$t$节点$i$曾经的邻居节点集,$w_{ij}(t)$为节点$i$与$j$从初始时刻到当前时刻$t$的交互次数. 概率$As_i(t)^{-\beta}$中:$A>0;\beta\in[0,1]$为模型参数;$s_i(t)$为节点$i$在时刻$t$的交互总强度,即到时刻$t$节点$i$产生链接交互的总次数. 具体地,$\mathcal{A}_{ij}(t)$为节点$i$与$j$在$t$时间步发生链接交互的指示变量,取值为1表示产生交互链接,反之取值为0,则有

$$s_i(t)=\sum_{t'=0}^{t}\sum_{j=1}^{N}\mathcal{A}_{ij}(t'),\ w_{ij}(t)=\sum_{t'=0}^{t}\mathcal{A}_{ij}(t').$$

相比于Karsai等[61]提出的记忆性活跃度驱动模型,此模型给出一个可调控的记忆程度参数β,$\beta=0$对应无记忆情形,$\beta>0$表示记忆的影响程度. 其次,此模型表明节点在邻居间的选择存在差异,交互越频繁的邻居在未来被链接的可能性越大. 当网络规模充分大时,此网络生成模型对应静态聚合网络中节点的度近似服从幂律分布$P(k)\sim k^{-\gamma_k}$,$\gamma_k=(\gamma-\beta)/(1-\beta)$,其中$\gamma$为激活概率分布的幂指数. 此外,节点交互强度$s$和连边权重$w$也有近似的幂律分布形式$P(s)\sim s^{-\gamma_s}$,$P(w)\sim w^{-\gamma_w}$,$\gamma_s=\gamma_w=\gamma$.

对活跃度驱动模型的记忆性扩展使得生成网络序列具有非马尔可夫性,具有更强的现实意义. 然而,具有记忆性的模型在其他方面仍然与活跃度驱动模型存在相同的问题,例如网络中节点的活跃方式仍然近似为泊松过程,且当网络演化时间充分长时,聚合网络仍将收敛到全链接结构.

4.3 吸引力活跃度驱动模型

在活跃度驱动模型中加入对历史交互信息的记忆可以使生成的网络序列具有相关性,但是这种记忆性的机制是针对网络局部实施的. 在真实数据中,很多时序的交互信息无法通过局部作用机制来解释和编码. 例如,在社交网络中,人们通常会以更高的概率与知名用户进行交互,这也是其影响力和流行度的体现,这种行为是一种全局作用影响的结果,而非局部的作用机制. 在活跃度驱动模型中,每个节点$i\,(1\leqslant i\leqslant N)$被赋予一个激活概率$a_i$,用于刻画该节点产生交互的倾向或意愿,而交互对象的选择方式是随机的. 基于此,Alessandretti等[64]基于活跃度驱动模型对每个节点引入一个吸引力特征b_i,通过全局作用的影响改变激活节点选择交互对象的方式. 具体的吸引力活跃度驱动模型实现(activity-driven model with attractiveness)如下.

在活跃度驱动模型的基础上,每个节点i不仅被赋予一个激活概率a_i,而且被赋予一个吸引力b_i,(a_i,b_i)从联合分布$H(a,b)$中对不同的节点独立抽取. 一般而言,激活概率a与吸引力b是相关的,将活跃度驱动模型的step 2改为

step 2"': 在每个离散时间步,节点i以概率$a_i\Delta t$变为活跃状态,并连接m条边到其他节点,其他任一节点j被一条边选择链接的概率为$b_j\Big/\sum\limits_{i=1}^{N}b_i$.

这里吸引力的作用方式实际上与优先链接机制相同,节点获得连边的概率与其吸引力呈线性关系. 一般而言,吸引力的幂律分布会导致聚合网络度的幂律分布.

4.4 基于静态网络属性的时序网络构建

活跃度驱动模型及其扩展模型通过给网络中的每个节点赋予相应的激活概率来促使节点间交互,生成网络切片序列. 由于模型中每个节点所赋予的激

活概率始终保持不变,在$\Delta t \to 0$时每个节点的活跃间隔时间服从指数分布,整体的活跃模式可视为泊松过程. 通过对大量真实数据的研究表明,时序网络中发生在节点上的事件和边上的事件间隔时间均服从爆发特性(图2(b)~(e)). 因此活跃度驱动模型虽然在一定程度上通过节点间参与交互倾向的异质性刻画了网络拓扑结构的异质性,但是网络中每个节点的活跃模式与真实数据所反应的现象有着显著区别. 近年来,许多研究关注能同时产生网络拓扑结构的无标度特性和网络节点活动模式爆发特性的时序网络模型. 然而,通过动态方式生成同时具有这两种特性的网络仍是一个困难且悬而未决的问题,许多研究通常假设对应的静态网络已具有无标度特性,并在此基础上探究节点和边产生爆发特性的动力学模式.

4.4.1 DLM模型

Fonseca等[65]提出了基于已有的静态网络结构可以同时产生节点活动和边活动间隔时间服从重尾分布的Dos-Li-Masuda(DLM)交互模型. 需要说明的是,节点和边的活动同时呈现出爆发特性这一现象不是平凡的,因为节点上的活动是其所有连边上活动的叠加. 若节点i每条连边的交互均是独立的且具有爆发特性,即每条边上的交互间隔时间均为幂律分布,则叠加形成的节点i的交互模式将是近似泊松过程而非幂律的,这表明节点和边同时存在的爆发特性可能源于边上交互事件的相关性. 基于此,Dos等假设每个节点都存在高活跃度和低活跃度两种状态,当处于高活跃状态时,节点的每条连边上都更有可能产生交互,而处于低活跃状态时,每条连边上的交互以较低的概率发生,节点在两个状态间的转换由一个连续时间两状态的马尔可夫链控制,这样可以使得节点i每条边上的交互与节点所处的状态产生相关性,具体的模型实现如下.

记节点的高活跃度状态为h,低活跃度状态为l,每个节点由状态h转换为l的转移速率为$r_{h\to l}$. 相对地,由状态l转换为h的转移速率为$r_{l\to h}$,且各个节点的状态转换相互独立. 如果两相邻节点i和j都处在状态h,则此时两节点连边(i,j)上交互的发生服从速率为λ_h的泊松过程;如果至少有一个节点处在l状态,则连边上交互的发生服从速率为λ_l的泊松过程,其中$\lambda_l < \lambda_h$. 由构造的连续时间马尔可夫链可知在平稳状态下,网络中一节点处于h状态的概率为

$$p_h^\star = \frac{r_{l\to h}}{r_{l\to h} + r_{h\to l}}.$$

对于网络中的任一条边而言,其上发生交互的间隔时间分布取决于该边两端节点的状态组合. 当两端节点的状态组合为(h,h)时,边上交互的时间间隔服从参数为λ_h的指数分布,其他情形下服从参数为λ_l的指数分布,所以边上交互间隔时间τ的分布为混合指数分布,其概率密度函数可表示为

$$P(\tau) = \frac{\lambda_h p_h^{\star\,2}}{\Omega_1}\lambda_h \mathrm{e}^{-\lambda_h \tau} + \frac{\lambda_l (1-p_h^{\star\,2})}{\Omega_1}\lambda_l \mathrm{e}^{-\lambda_l \tau},$$

其中$\Omega_1 = \lambda_h p_h^{\star\,2} + \lambda_l(1-p_h^{\star\,2})$. 类似地节点的交互间隔时间也为多个指数分布的混合. 注意到,虽然边与节点的交互模式最终用混合指数分布刻画而非幂律分布,但从变异系数(coefficient of variation, CV, $\mathrm{CV} = \sqrt{\langle\tau^2\rangle/\langle\tau\rangle^2 - 1}$)的角度看,这样的混合分布呈现出重尾分布的特点,即变异系数皆大于1. 其分布的异质程度比指数分布高,并且λ_l/λ_h越小,该模型得到的边与节点的交互间隔时间分布的CV值越高.

模型给出了一个使得时序网络中节点与边上的交互间隔时间均服从重尾分布的潜在机制,但其不足之处在于基于静态网络结构的前提,并没有给出时序网络结构演化的机制,而且通过该模型最终得到的节点和边的交互模式服从混合指数分布,并非幂律分布.

4.4.2 HMLJ模型

Hiraoka等[66]基于静态网络结构提出了一种使节点和边的活动均具有爆发特性的Hiraoka-Masuda-Li-Jo(HMLJ)交互模型. DLM模型[65]首先确定了边上的交互模式,通过边上交互的叠加得到节点的交互模式,而HMLJ模型[66]首先确定节点的交互模式,并通过各个节点的状态推出各边的交互模式,该模型实际上是对活跃度驱动模型节点交互方式的扩展. 由于活跃度驱动模型中每个节点的激活概率始终保持不变,从而导致节点的活跃交互模式服从泊松过程. 为了使节点的活跃打破泊松过程的形式,最直接的想法是使节点的激活概率随时间变化. 基于此,具体的模型实现如下.

对网络中的每个节点$i\,(1 \leqslant i \leqslant N)$都赋予一个激活概率$\lambda_i \in [0,1]$,$\lambda_i$从某个分布$F(\lambda)$中独立随机抽取,其中$F(\lambda) \sim \lambda^{\alpha-2}, \alpha > 2$. 在每个离散时间步,节点$i$以概率$\lambda_i$变为活跃状态,以概率$1-\lambda_i$变为非活跃状态. 当节点为活跃状态时,将从分布$F(\lambda)$中重新抽取相应的激活概率$\lambda_i$,并在下一时间步以新的概率$\lambda_i$激活. 对于边上的交互模式,HMLJ模型提供了两种具体方法:

1) 当某一节点激活时,它将与其邻居中所有同时激活的节点交互;

2) 当某一节点激活时,它将随机挑选一个其邻居节点中同时激活的节点进行交互.

两种方法对边上的交互提出了不同的模式,但最终节点和边上的交互间隔时间均服从近似幂律分布,且有概率密度函数

$$P_{\text{vertice}}(\tau)\sim\tau^{-\alpha},\ P_{\text{edge}}(\tau)\sim\tau^{-\alpha},$$

其中$\alpha>2$,与激活概率分布$F(\lambda)$中的α相同.

实际上,该模型中节点的活跃方式可以看作分段的泊松过程,每一段具有不同的速率,而速率则从分布$F(\lambda)$中随机抽取,$F(\lambda)\sim\lambda^{\alpha-2}$的形式源自文献[67]. 相比于DLM模型[65],该模型得到的节点和边上共同的爆发行为服从幂律,而非混合指数分布.

4.5 吸引力动力学模型

针对面对面网络中的人类互动现象,Starnini等[68]建立模型解释了人类活动的爆发特性,即个体或群体之间的持续互动时间和同一个体连续两次互动的时间间隔服从重尾分布. 现有的模型大多试图阐明人类选择任务机制与人类行为中爆发现象的联系,Starnini等基于空间临近性影响个体间的相互作用进行建模. 基于社会网络的异质性,首先假设不同个体对邻居的吸引力不同而导致他们与邻居的互动程度不同,并根据真实数据进一步假设并非所有个体都同时存在于系统中,它们可以选择进入或退出一个活跃状态,活跃个体可以选择随机游走或者与邻居互动. 基于上述假设,Starnini等提出的模型如下.

step 1: 在边长为L的正方形区域中放入N个个体,此时密度$\rho=N/L^2$,且所有节点间不存在交互.

step 2: 对每个个体i,从分布$\eta(r)$和$\zeta(a)$中随机选择r_i和a_i表示该个体对邻居的吸引力和个体自身的活跃度,其中$r_i,a_i\in[0,1)$. 每个个体的邻居定义为与其欧氏距离小于等于d的所有个体.

step 3: 在每个时间步t,对于个体i,若当前时间步其不存在交互行为,则该个体以概率a_i变为活跃个体,以互补概率$1-a_i$为不活跃个体;若当前时间步存在交互行为,则默认其为活跃个体.

step 4: 活跃个体i以概率$p_i(t)$朝随机选择的方向$\xi\in[0,2\pi)$移动距离v,以概率$1-p_i(t)$保持当前位置,其中$p_i(t)=1-\max_{j\in\mathcal{N}_i(t)}\{r_j\}$,$\mathcal{N}_i(t)$为$t$时刻个体$i$的邻居集. 每个个体基于当前位置更新其邻居集合,活跃个体i与其邻居j以概率r_j产生交互. 已产生的交互当且仅当交互双方的距离大于d时自动断开.

step 5: 重复step 3和step 4.

在此框架下,个体在二维空间随机游动的同时与其他个体进行不同时长的互动. 该模型因具有无后效性等特点被视为马尔可夫过程,整个动力学系统主要取决于碰撞概率$p_c=\rho\pi d^2$、活跃度分布$\zeta(a)$、吸引力分布$\eta(r)$. 通过数值模拟并与真实数据比较,Starnini等[68]发现两个体互动持续时间Δt和同一个体两次连续互动间隔时间τ的分布与真实数据相吻合,均服从重尾分布.

该研究框架进一步帮助人们认识了面对面互动网络中的人类行为特性,与之前的工作相比,研究模型不受内部参数的影响,表现出良好的稳定性以及与经验数据的高度一致性. 未来的研究可以基于该模型进行更深入地探讨,将其扩展到更一般的情况进一步讨论面对面互动网络中的人类行为特性.

活跃度驱动模型通过赋予每个节点不同的活跃度从而控制节点产生交互的时间和次数,当活跃度分布为幂律时,生成的聚合网络度分布也服从幂律,即生成的聚合网络为无标度网络. 该模型因其机制简单受到广泛关注,然而其具有较大的约束性,缺乏一般的时序网络交互机制与特征(如个体交互的爆发特性). 记忆性活跃度驱动模型在活跃度驱动模型的基础上引入对曾经交互的记忆,用来指导节点的下一次交互,使交互对象的选择具有非马尔可夫性,更符合现实特征. 吸引力活跃度驱动模型则是在活跃度驱动模型的基础上,通过赋予节点不同的吸引力指导时序网络中节点的交互对象选择. 这里吸引力的作用方式实际上与优先链接机制相同. 以上模型都旨在生成具有聚合无标度特性的时序网络,缺乏对节点交互的时序爆发模式. 基于静态网络属性的时序网络模型在给定聚合网络无标度结构的基础上,刻画节点间交互的爆发模式. 从而使得生成的时序网络序列既满足聚合静态网络的无标度特性又符合个体交互的爆发特性. 然而,如何通过动态方式生成同时具有这两种特性的网络仍是一个公开的难题. 吸引力动力学模型则是通过规定每个个体随机游走的机制直接刻画人类社会面对面交流网络中的互动现象. 该研究模型不受内部参数的影响,表现出了良好的稳定性以及与经验数据的高度一致性.

5 时序网络上系统动力学

时序网络近年来成为研究热点的原因不仅仅在于其网络拓扑结构相较于传统静态网络而言随着时间动态演化,更在于时序网络上系统动力学行为显著有别于其对应的静态聚合网络上系统动力学行为. 以下仅通过时序网络上信息扩散、博弈策略演化为例作简要介绍,有关传染病传播或系统控制可参见相应文献[7,10].

Li等[22]基于群体间的真实个体动态交互数据构建了随时间演化的时序网络,并基于经典囚徒困境模

型研究了时序网络上群体策略的演化. 通过与对应的聚合静态网络上群体博弈行为进行对比发现,时序网络较相应的静态网络能够促进群体合作行为的涌现. 根据传统研究结果,群体合作行为的涌现往往依赖于合作策略形成相应的团簇(cluster)从而抵御个体最优策略,即纳什均衡策略的入侵,而连通的静态网络为团簇的形成提供了天然的有利条件. 但该文发现,随着时间动态演化的时序网络虽然网络连边在动态变化,但不妨碍合作行为的涌现,结论较为反直观,提供了研究真实动态演化系统中群体行为涌现的新的视角. 进一步,作者还指出时序网络系统中个体交互的爆发特性并非是时序网络较传统静态网络促进合作行为涌现的本质原因;相反,爆发特性反而能够促进纳什均衡策略(即囚徒困境中的纯策略——背叛策略)的演化,这也充分体现了时序网络上系统动力学行为的复杂性. 根据所定义的网络瞬时性,即网络切换速度的快慢,该文进一步指出,中等时间尺度的网络演化速度能够使得时序网络最为促进群体合作行为的涌现. 一般而言,时序网络上群体博弈、信息传播、传染病传播等系统动力学行为,可建模为网络节点与边状态动态演化动力学. 根据传染病传播动力学理论,成功预测了时序网络上群体合作行为的演化趋势,并给出纳什均衡策略在时序网络上得以演化的数学条件,进一步揭示了群体最优策略在动态博弈场景中得以演化的科学机理.

针对时序网络上信息扩散, Scholtes 等[9] 发现时序网络中交互序列间的因果关系,即交互产生的先后顺序与相关性,可以影响时序网络上扩散过程的传播速度. 实际上,这是由于交互序列间的因果关系可能使得时序网络的连通性增强或削弱,从而导致其上的扩散过程加速或减慢. 此外,提出一种基于高阶交互的马尔可夫模型预测时序网络中因果关系驱动的扩散速度变化. 与静态网络不同,时序网络中的连边具有时间尺度,每一条边的出现与消失都有相应的时间序列,因此在聚合网络中存在的连边通路并不一定在时序网路中存在. 例如在聚合网络中存在 (a,b) 与 (b,c) 连边,则导致有一条从 a 到 c 的路径 $a\to b\to c$. 然而在时序网络中,这条路径的存在进一步要求 (a,b) 出现在 (b,c) 之前,且相隔时间在一定的阈值内. 这样的因果逻辑使得时序网络结构随时间演变,且这样的变化通常是非马尔可夫的. Scholtes 等利用线图(line graph),将原有向网络中的连边视为新的节点,新节点间的连边为原网络连边间存在的先后因果关系,构造一个二阶聚合网络. 利用在该高阶网络上的马尔可夫过程,近似在原网络上进行非马尔可夫过程. 设在高阶聚合网络上随机游走的转移矩阵为 T,在原聚合网络上的随机游走转移矩阵为 $\tilde{T}$. 由于对一个具有特征值 $1=\lambda_1>|\lambda_2|>|\lambda_3|\geqslant\ldots\geqslant|\lambda_n|$ 的转移矩阵,可以证明使得其定义的随机游走在 k 步后的访问概率分布 π_k 与平稳分布 π 的全变差距离 $\Delta(\pi_k,\pi):=\dfrac{1}{2}\sum|(\pi)_v-(\pi_k)_v|<\epsilon$ 的步数 k 与 $1/\ln(|\lambda_2|)$ 成正比,定义由因果关系驱动的收敛速度变化为

$$\mathcal{S}^*(T):=\ln(|\tilde{\lambda}_2|)/\ln(|\lambda_2|),$$

其中 λ_2、$\tilde{\lambda}_2$ 分别为 T 和 $\tilde{T}$ 模度量意义下的第2大特征值. 当 $\mathcal{S}^*(T)>1$ 时,时序网络因果关系加速扩散过程,反之则减缓.

进一步, Masuda 等[8] 发现,与所有链接均同时存在的静态网络相比,流行病的传播或扩散等动态过程会因为时序网络拓扑结构的持续切换而减慢. 直观上,将静态聚合网络解构为一系列瞬时网络切片会使得网络的连通性减弱,从而降低扩散速度. 事实上,通过研究时序网络上扩散过程的拉普拉斯频谱,并将其与相应的聚合网络进行比较,发现两者具有相同的特征空间,但前者的特征值小于后者. 设网络节点集为 $V=\{1,2,\ldots,N\}$,边集为 $E\subseteq\{\{i,j\}:i,j\in V,i\neq j\}$,节点 i 的邻居集为 $N_i=\{j\in V:\{i,j\}\in E\}$. 于是静态聚合网络上的扩散过程描述为 $\dot{x}_i=|E|^{-1}\sum\limits_{j\in N_i}(x_j-x_i)$. 其中: x_i 为节点 i 的状态, $|E|$ 为边集的元素个数即聚合网络的连边总数. 写作向量形式为

$$\dot{x}=M^*x,$$

其中为 M^* 非负,为重缩放后的拉普拉斯矩阵

$$M^*_{ij}=|E|^{-1}\begin{cases}1,\ \{i,j\}\in E;\\ -|N_i|,\ i=j;\\ 0,\ \text{otherwise.}\end{cases}$$

时序网络中,假设每时刻扩散过程仅发生在一条边 $e\in E$ 上,并持续一段时间 τ,则有发生扩散的边序列 $e_0,e_1,e_2,\ldots$,此时扩散过程可描述为

$$\dot{x}(t)=M^{e_{r(t)}}x(t).$$

其中

$$r(t)=\lfloor t/\tau\rfloor.$$

$$M^{(v,w)}_{ij}=\begin{cases}1,\ \{i,j\}=\{v,w\};\\ -1,\ i=j,\ i\in\{v,w\};\\ 0,\ \text{otherwise.}\end{cases}$$

当每次都选择从边集E中随机抽取一条边发生扩散时,扩散的平均状态可以表达为

$$\langle x(t+\tau)\rangle = |E|^{-1}\sum_{e\in E}\exp(\tau M^{(e)})\langle x(t)\rangle.$$

记$\hat{T} = |E|^{-1}\sum_{e\in E}\exp(\tau M^{(e)})$,则此时等效的交互矩阵可写为$\hat{M} = \tau^{-1}\ln\hat{T}$. 又因$(M^{(e)})^2 = 2M^{(e)}$,故$\exp(\tau M^{(e)}) = I + \alpha(\tau)M^{(e)}$,其中$\alpha(\tau) = \dfrac{1-\exp(-2\tau)}{2}$. 于是有

$$\hat{T} = |E|^{-1} : \sum_{e\in E}\exp(\tau M^{(e)}) =$$

$$I + \alpha(\tau)|E|^{-1} : \sum_{e\in E} M^{(e)}.$$

进一步得

$$\hat{M} = \tau^{-1}\ln[I + \alpha(\tau)M^*].$$

可以看到,时序网络上扩散过程的拉普拉斯频谱与相应的聚合网络具有相同的特征空间,其特征值有如下关系:

$$\hat{\mu} = f(\mu^*, \tau) = \tau^{-1}\ln[1 + \alpha(\tau)\mu^*].$$

于是,时序网络上扩散过程的拉普拉斯频谱的特征值随τ的增大而减小,扩散速率整体下降.

6 总结与展望

因随着时间演化的时序网络对其上各类系统动力学分析及系统控制的影响,越来越多的研究人员开始关注时序网络相较于传统静态网络所独有的特性. 在取代传统上基于有限的个体真实交互数据来构建具有典型特征的时序网络的同时,已有研究往往先假定构建模型中已经带有某种典型特性. 此类模型所带来的系统性误差使得时序网络的构建无法像经典静态网络构建模型那样仅仅依赖于个体交互的基本规则. 后续研究应重点关注如何基于基本的规则生成同时具备典型特性的时序网络,且定量探究刻画网络特性关键参数间的数学关系.

2021年诺贝尔物理学奖授予了意大利科学家Giorgio Parisi等人,以表彰他们对理解复杂系统理论的开创性贡献. 如何从过去二十多年来网络科学的研究成果出发,将复杂系统的拓扑结构加入时间演化的维度,将是复杂系统研究中不可回避的重要科学问题. 当然,时间维度的引入在丰富传统研究结果的同时,复杂网络化系统的结构分析以及相关的群体智能理论、分布式决策与控制理论等前沿基础科学问题也面临重要挑战. 例如,当多个智能体在复杂时序网络化场景中彼此对抗博弈时,如何分析系统的网络结构特征、建立系统演化动力学、设计智能体的决策交互准则、控制系统沿着期望轨迹演化等问题均具有极大的挑战性. 同时,这也是我国《新一代人工智能发展规划》中“群体智能理论”“自主协同控制与优化决策理论”等基础理论研究中的核心科学问题.

参考文献(References)

[1] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks[J]. Nature, 1998, 393(6684): 440-442.

[2] Erdös P, Rényi A. On the evolution of random graphs[J]. Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, 5(1): 17-60.

[3] Barabási A L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.

[4] 钱学森, 于景元, 戴汝为. 一个科学新领域——开放的复杂巨系统及其方法论[J]. 自然杂志, 1990, 12(1): 3-10.
(Qian X S, Yu J Y, Dai R W. A new field of science — Open complex giant system and its methodology[J]. Nature Magazine, 1990, 12(1): 3-10.)

[5] 王龙, 伏锋, 陈小杰, 等. 演化博弈与自组织合作[J]. 系统科学与数学, 2007, 27(3): 330-343.
(Wang L, Fu F, Chen X J, et al. Evolutionary games and self-organizing cooperation[J]. Journal of Systems Science and Mathematical Sciences, 2007, 27(3): 330-343.)

[6] 王龙, 伏锋, 陈小杰, 等. 复杂网络上的演化博弈[J]. 智能系统学报, 2007, 2(2): 1-10.
(Wang L, Fu F, Chen X J, et al. Evolutionary games on complex networks[J]. CAAI Transactions on Intelligent Systems, 2007, 2(2): 1-10.)

[7] 李阿明, 王龙. 时序网络控制[J]. 系统科学与数学, 2019, 39(2): 184-202.
(Li A M, Wang L. Controlling temporal networks[J]. Journal of Systems Science and Mathematical Sciences, 2019, 39(2): 184-202.)

[8] Masuda N, Klemm K, Eguíluz V M. Temporal networks: Slowing down diffusion by long lasting interactions[J]. Physical Review Letters, 2013, 111(18): 188701.

[9] Scholtes I, Wider N, Pfitzner R, et al. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks[J]. Nature Communications, 2014, 5(1): 1-9.

[10] Takaguchi T, Masuda N, Holme P. Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics[J]. PLoS One, 2013, 8(7): e68629.

[11] Masuda N, Lambiotte R. A guide to temporal networks[M]. London: World Scientific Publishing Europe Ltd., 2016.

[12] Holme P, Saramäki J. Temporal networks[J]. Physics Reports, 2012, 519(3): 97-125.

[13] 王龙, 丛睿, 李昆. 合作演化中的反馈机制[J]. 中国科学: 信息科学, 2014, 44(12): 1495-1514.
(Wang L, Cong R, Li K. Feedback mechanism in

cooperation evolving[J]. Scientia Sinica Informationis, 2014, 44(12): 1495-1514.)

[14] Cui Y L, He S B, Wu M C, et al. Improving the controllability of complex networks by temporal segmentation[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(4): 2765-2774.

[15] 王龙, 杜金铭. 多智能体协调控制的演化博弈方法[J]. 系统科学与数学, 2016, 36(3): 302-318.
(Wang L, Du J M. Evolutionary game theoretic approach to coordinated control of multi-agent systems[J]. Journal of Systems Science and Mathematical Sciences, 2016, 36(3): 302-318.)

[16] Hou B Y, Li X, Chen G R. Structural controllability of temporally switching networks[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2016, 63(10): 1771-1781.

[17] Lentz H H K, Selhorst T, Sokolov I M. Unfolding accessibility provides a macroscopic approach to temporal networks[J]. Physical Review Letters, 2013, 110(11): 118701.

[18] 关永强, 纪志坚, 张霖, 等. 多智能体系统能控性研究进展[J]. 控制理论与应用, 2015, 32(4): 421-431.
(Guan Y Q, Ji Z J, Zhang L, et al. Recent developments on controllability of multi-agent systems[J]. Control Theory & Applications, 2015, 32(4): 421-431.)

[19] Pedreschi N, Battaglia D, Barrat A. The temporal rich club phenomenon[J]. Nature Physics, 2022, 18(8): 931-938.

[20] Liu X M, Lin H, Chen B M. Structural controllability of switched linear systems[J]. Automatica, 2013, 49(12): 3531-3537.

[21] 王龙, 田野, 杜金铭. 社会网络上的观念动力学[J]. 中国科学: 信息科学, 2018, 48(1): 3-23.
(Wang L, Tian Y, Du J M. Opinion dynamics in social networks[J]. Scientia Sinica: Informationis, 2018, 48(1): 3-23.)

[22] Li A M, Zhou L, Su Q, et al. Evolution of cooperation on temporal networks[J]. Nature Communications, 2020, 11(1): 1-9.

[23] Sheng A Z, Li A M, Wang L. Evolutionary dynamics on sequential temporal networks[J/OL]. 2021, arXiv: 2110.05995.

[24] Li A, Cornelius S P, Liu Y Y, et al. The fundamental advantages of temporal networks[J]. Science, 2017, 358(6366): 1042-1046.

[25] Barabási A L. The origin of bursts and heavy tails in human dynamics[J]. Nature, 2005, 435(7039): 207-211.

[26] Li A M, Liu Y Y. Controlling network dynamics[J]. Advances in Complex Systems, 2019, 22(7n08): 1950021.

[27] Duan G P, Li A M, Meng T, et al. Energy cost for controlling complex networks with linear dynamics[J]. Physical Review E, 2019, 99: 052305.

[28] Duan G P, Li A M, Meng T, et al. Energy cost for target control of complex networks[J/OL]. 2019, arXiv: 1907.06401.

[29] Meng T, Duan G P, Li A M, et al. Control energy scaling for target control of complex networks[J]. Chaos, Solitons & Fractals, 2023, 167: 112986.

[30] Goh K I, Kahng B, Kim D. Universal behavior of load distribution in scale-freenetworks[J]. Physical Review Letters, 2001, 87(27): 278701.

[31] Newman M E, Strogatz S H, Watts D J. Random graphs with arbitrary degree distributions and their applications[J]. Physical Review E, 2001, 64(2): 026118.

[32] Bianconi G, Barabási A L. Competition and multiscaling in evolving networks[J]. Europhysics Letters: EPL, 2001, 54(4): 436-442.

[33] Bianconi G, Barabási A L. Bose-Einstein condensation in complex networks[J]. Physical Review Letters, 2001, 86(24): 5632-5635.

[34] Broido A D, Clauset A. Scale-free networks are rare[J]. Nature Communications, 2019, 10(1): 1-10.

[35] Zhou B, Meng X Y, Stanley H E. Power-law distribution of degree-degree distance: A better representation of the scale-free property of complex networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(26): 14812-14818.

[36] Voitalov I, van der Hoorn P, van der Hofstad R, et al. Scale-free networks well done[J]. Physical Review Research, 2019, 1(3): 033034.

[37] Holme P. Rare and everywhere: Perspectives on scale-free networks[J]. Nature Communications, 2019, 10(1): 1-3.

[38] Erica Klarreich. Scant evidence of power laws found in real-world networks[EB/OL]. (2018-02-15) [2022-09-16]. https://www.quantamagazine.org/scant-evidence- of- power - laws-found-in-real-world-networks-20180215/.

[39] Clauset A, Shalizi C R, Newman M E J. Power-law distributions in empirical data[J]. SIAM Review, 2009, 51(4): 661-703.

[40] Handcock M S, Jones J H. Likelihood-based inference for stochastic models of sexual network formation[J]. Theoretical Population Biology, 2004, 65(4): 413-422.

[41] Clauset A, Young M, Gleditsch K S. On the frequency of severe terrorist events[J]. Journal of Conflict Resolution, 2007, 51(1): 58-87.

[42] Haight F A. Handbook of the Poisson distribution[M]. New York: Wiley, 1967: 16-25.

[43] Anderson H R. Fixed broadband wireless system design[M]. Manhattan: John Wiley & Sons, 2003: 334-342.

[44] Karsai M, Jo H H, Kaski K. Bursty human dynamics[M]. Berlin: Springer, 2018: 31-46.

[45] Vázquez A, Oliveira J G, Dezsö Z, et al. Modeling bursts and heavy tails in human dynamics[J]. Physical Review E, 2006, 73(3): 036127.

[46] Dezsö Z, Almaas E, Lukács A, et al. Dynamics of information access on the web[J]. Physical Review E, 2006, 73(6): 066132.

[47] Rácz B, Lukács A. High density compression of log files[C]. Data Compression Conference. Snowbird, 2004: 557.

[48] Eckmann J P, Moses E, Sergi D. Entropy of dialogues creates coherent structures in e-mail traffic[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(40): 14333-14337.

[49] Ebel H, Mielsch L I, Bornholdt S. Scale-free topology of e-mail networks[J]. Physical Review E, 2002, 66(3): 035103.

[50] Einstein A. The collected papers of albert Einstein — Volume 15[M]. New Jersey: Princeton University Press, 1987: 63-70.

[51] Darwin C, Porter D M, Dean S A, et al. The Correspondence of charles darwin[M]. Cambridge: Cambridge University Press, 2002: 1821-1879.

[52] Oliveira J G, Barabási A L. Human dynamics: Darwin and Einstein correspondence patterns[J]. Nature, 2005, 437(7063): 1251.

[53] Cooper R B, Niu S C, Srinivasan M M. Some reflections on the Renewal-theory paradox in queueing theory[J]. Journal of Applied Mathematics and Stochastic Analysis, 1998, 11(3): 355-368.

[54] Cobham A. Priority assignment in waiting line problems[J]. Journal of the Operations Research Society of America, 1954, 2(1): 70-76.

[55] Abate J, Whitt W. Asymptotics for M/G/1 low-priority waiting-time tail probabilities[J]. Queueing Systems, 1997, 25(1): 173-233.

[56] Cohen J W. The single server queue[M]. Amsterdam: Elsevier Science Ltd, 1982: 237-266.

[57] Vázquez A. Exact results for the Barabási model of human dynamics[J]. Physical Review Letters, 2005, 95(24): 248701.

[58] Perra N, Gonçalves B, Pastor-Satorras R, et al. Activity driven modeling of time varying networks[J]. Scientific Reports, 2012, 2(1): 1-7.

[59] Starnini M, Pastor-Satorras R. Topological properties of a time-integrated activity-driven network[J]. Physical Review E, 2013, 87(6): 062807.

[60] Song C M, Koren T, Wang P, et al. Modelling the scaling properties of human mobility[J]. Nature Physics, 2010, 6(10): 818-823.

[61] Karsai M, Perra N, Vespignani A. Time varying networks and the weakness of strong ties[J]. Scientific Reports, 2014, 4(1): 1-7.

[62] Kim H, Ha M, Jeong H. Scaling properties in time-varying networks with memory[J]. The European Physical Journal B, 2015, 88(12): 315.

[63] Kim H, Ha M, Jeong H. Dynamic topologies of activity-driven temporal networks with memory[J]. Physical Review E, 2018, 97(6): 062148.

[64] Alessandretti L, Sun K Y, Baronchelli A, et al. Random walks on activity-driven networks with attractiveness[J]. Physical Review E, 2017, 95(5): 052318.

[65] Fonseca D R E, Li A M, Masuda N. Generative models of simultaneously heavy-tailed distributions of interevent times on nodes and edges[J]. Physical Review E, 2020, 102(5): 052303.

[66] Hiraoka T, Masuda N, Li A M, et al. Modeling temporal networks with bursty activity patterns of nodes and links[J]. Physical Review Research, 2020, 2(2): 023073.

[67] Masuda N, Rocha L E C. A Gillespie algorithm for non-Markovian stochastic processes[J]. SIAM Review, 2018, 60(1): 95-115.

[68] Starnini M, Baronchelli A, Pastor-Satorras R. Modeling human dynamics of face-to-face interaction networks[J]. Physical Review Letters, 2013, 110(16): 168701.

作者简介

李阿明(1988−), 男, 研究员, 博士生导师, 博士, 从事群体博弈与决策、网络系统控制、群体智能等研究, E-mail: liaming@pku.edu.cn;

侯谷庾(1998−), 男, 硕士生, 从事复杂网络、智能体博弈等研究, E-mail: guyuhou@stu.pku.edu.cn;

王龙(1964−), 男, 教授, 博士生导师, 博士, 从事人工智能、博弈控制理论、演化动力学等研究,E-mail: longwang@pku.edu.cn.

科研团队简介

王龙教授科研团队立足于北京大学系统与控制研究中心,长期专注于人工智能、博弈控制理论的研究,一直倡导将前沿性基础研究成果与国民经济发展和国家重大需求紧密相连. 团队近年来在控制理论与人工智能前沿交叉基础科学领域作出了一系列系统性的创新工作,在国内外著名学术期刊(如Science、PNAS、Nature Communications、IEEE Trans. on Automatic Control等)发表论文100余篇. 先后获得国家自然科学奖二等奖、关肇直奖、《中国科学:信息科学》10年经典论文奖、《控制理论与应用》创刊30周年最具影响力论文奖、教育部自然科学一等奖等多项奖励,相关成果被国内外著名学者大量引用和高度评价,在学术界产生了重要影响.

团队带头人王龙教授连续多年入选Clarivate、Elsevier高被引学者,先后获得国家自然科学奖二等奖(2017年第1获奖人)、三等奖(1999年第2获奖人)以及关肇直奖(1994年)、张嗣瀛奖(2019年)、《中国科学:信息科学》10年经典论文奖(2017年)、《控制理论与应用》创刊30周年最具影响力论文奖(2014)、教育部自然科学一等奖(2015年第1获奖人)、北京市优秀研究生指导教师(2022年)等多项奖励. 王龙教授现任北京人工智能学会副理事长、中国系统仿真学会常务理事、智能物联系统专业委员会主任,《控制与决策》《智能系统学报》《控制理论与应用》编委,国家应用数学中心(四川)学术委员会委员.