

◎ 汤姆 A.B. 斯奈德斯(Tom A.B. Snijders)

网络分析的动力学进路

许多社会网络分析都贯彻着动态思想。网络动力学对很多领域都很重要,例如,从朋友关系网(如 Pearson and West, 2003; Burk et al., 2007)到组织间关系网(见 Borgatti and Foster, 2003; Brass et al., 2004 的综述性文章)皆如此。但是,诸多网络分析的形式模型(无论在离散数学的传统还是在统计推断的传统中)都长期将注意力放在单类分析方法(即截面方法)上。

一些历史:经验研究

早期的一些重要的纵贯网络研究是由 Nordlie (1958) 和 Newcomb (1961) 作出的,他们基于所收集的经验数据,在一所大学的大学生互助会(fraternity)中研究朋友关系;Coleman (1961) 利用 10 所学校的 9 702 位个体的朋友关系数据,开展了青少年协会(Adolescent Society)的研究;在赞比亚(当时为北罗得西亚^①)劳资纠纷时期,Kapferer (1972) 对一家服装店中发生的互动作了 10 个多月的观察和研究;Sampson (1969) 的博士论文研究了一所寺院内 18 位僧侣的群体关系演变;还有 Hallinan 用七轮数据作的研究(参见 Hallinan, 1974, 1979; Sørensen and Hallinan, 1976)。不过,在大约 1990 年以前,学者的注意力还主要集中在观察单一的网络上。在 UCINET 程序包的数据集当中(Borgatti et al., 1998) 共有 20 个数据,其中只有三个提供了纵向数据:Kapferer 的服装店数据、Newcomb 的大学生互助会数据和 Sampson 的寺院数据。Wasserman 和 Faust (1994) 编写的社会网络分

^① 罗得西亚为前南非的一个地区,包括北罗得西亚(Northern Rhodesia)(今赞比亚)和南罗得西亚(Southern Rhodesia)(今津巴布韦)两部分——译者注。

析教材走在了前列,但是也只有半页纸的篇幅谈及了动态与纵向网络模型。如何才能明确地分析社会网络的历时性,人们并不太关注这个问题,这可以从两个方面来理解。一方面是网络动态数据难以收集,当研究人员想纵向收集它们时,这种困难会加剧;另一方面是社会网络动力学模型难以建立。

从1980年代起,研究者开始更广泛地收集网络面板(panel)数据。面板数据是指研究人员就一个既定的社会行动者群体,在两个或多个连贯的时刻(所谓面板期[panel waves])上收集的数据。Bauman等(1984)研究了五所学校中的朋友关系网,数据是从一项关注吸烟动因(dynamics of cigarette smoking)的研究中收集到的,在两期研究中,共形成了954份完整的问卷;在苏格兰做的青少年朋友与生活方式研究(the Teenage Friends and Lifestyle Study)有三期(West and Sweeting, 1995; Michell and Amos, 1997; Pearson and West, 2003)。目前最有名的研究大概是在美国做的青少年及成人健康研究(the Add Health study),这项研究进行了三期(Harris et al., 2003; Udry, 2003)。有学者(Christakis and Fowler, 2007)在弗兰明汉心脏研究(the Framingham Heart Study)中发现了有趣的网络数据,而这个历时研究最初并没想包含其网络的部分。也可以把官方档案和名录用作历时性网络数据的来源。这种研究中的一些例子有(Gulati and Gargiulo, 2000; Powell et al., 2005)以及一篇述评性文献(Hagedoorn, 2002)。

501

一些历史:统计模型

网络动力学的概率模型要求明确设定 $\{X(t) | t \in T\}$ 的同时概率分布(simultaneous probability distribution),其中 t 是时间参数,它假设了一个时间点集合 T 中的值, $X(t)$ 是时间 t 上的网络。在概率论中,它被称为随机过程,其中的结果空间是一个网络空间。将网络思考为是一个有向图(digraph)会很方便。虽然依照当下的情形,这个网络可能会有不同的结构,例如无向网、多值网等。对于一个有向图来说,网络 $X(t)$ 是由有向关系变量 $X_{ij}(t)$ 构成的,在时间 t 上,如果存在一条弧 $i \rightarrow j$,就用数值1表示,不存在这样的关系,其值就是0。在所有的情况下,我们都假设没有任何自我圈(self-loops),即总有 $X_{ii}(t) = 0$ 。我们会关注点集固定的情形,用 $\{1, \dots, n\}$ 表示。因此,网络是由 n 个行动者组成的。有些行动者是在数据收集开始之后才进入(enter),或在结束之前已离开,如果我们允许表示这类行动者的点具有某些灵活性,对于网络面板数据来说,这通常具有意义。应该注意的是,也有一些针对成长性网络(点进入网络中)的模型,它常常伴有附加的假定,即关系一旦被创建,就不会改变,网络的变化取决于新创建的点所创立的关系。在随机图的数学理论中(如Bollobás, 1985),这是一个经典的研究进路。

动态网模型必须反映反馈过程,这个过程是网络的特征。作为实例,我们可以考虑一些社会网络分析中经典的关系创建过程:互惠性(reciprocation)(Moreno, 1934)、传递性闭合(Rapoport, 1953a, 1953b; Davis, 1970)和马太效应(“因为凡有的,还要加给他,叫他有众。没有的,连他所有的也要夺回来”;

Merton, 1963; de Solla Price, 1965, 1976; Barabási and Albert, 1999 称之为“偏好依附”[preferential attachment])。如果在某个时刻 t , 关系 $i \rightarrow j$ 不存在, 那么如果现在有一个关系 $j \rightarrow i$, 互惠性就有可能在随后的某个时刻被创建; 如果有两个关系被安排在一个 2-路径 $i \rightarrow h \rightarrow j$ 中, 即现在有一个从 i 到 j 的间接关联, 那么传递性闭合就有可能被创建; 如果存在许多其他行动者 h , 对他们来说, 存在关系 $h \rightarrow j$, 即现在从高点入度 (indegree) 的意义上看, 行动者 j 是受欢迎的 (popular), 那么马太效应就有可能被创建。这些例子说明, 网络动力学的统计模型必须表达历时态之间的依赖性和关系之间的依赖性。

历时态之间的依赖性

在对历时态的依赖性建模时, 绝大多数已发表的模型都看似利用了马尔可夫性质中的某种变异。宽泛地讲, 这种性质是针对随机过程定义的, 其意在于, 未来经由现在依赖于过去。一个较形式化的定义 (虽然仍然有些不完整) 是: 对于时点 $t_1 < t_2 < t_3$ 来说, $X(t_3)$ 以 $X(t_2)$ 为条件, 独立于 $X(t_1)$ 。在最早提出的模型中, 假定面板数据是 $X(t_1), X(t_2), \dots, X(t_n)$, 这 n 个连续的观测点就构成了一个马尔可夫过程。例如, 很多学者 (Katz and Proctor, 1959; Wasserman, 1987; Wasserman and Iacobucci, 1988; Robins and Pattison, 2001) 都作了这样的假定。由于观测值在数量上有限, 这被称为离散-时间的马尔可夫过程。

然而, 可以将上文提到的反馈过程假设为是未被观察到却在观测值之间运行着的过程。例如, 在一个马太效应运行于其中的群体里, 如果某个点 i 在时点 t_1 上有一个低的点入度, 在下一个观测点 t_2 上有一个非常高的点入度, 那么通过逐渐积累那些指向 i 的关系, 如下情况就更容易发生: 这些关系本来就有发生的机会, 只不过一旦点入度相对较高, 它就会变成一个自我强化 (self-reinforcing) 的过程了。这样一个模型预设了在观察点 t_1 和 t_2 之间会发生变化。最为漂亮的和在数学上最易处理的建模方式是假设一个连续-时间的 (continuous-time) 马尔可夫过程 $\{X(t) \mid t_1 \leq t \leq t_m\}$ 。换句话说, 当仍然坚持对观察数据作面板设计时, 令过程 T 的时点集合为整个区间 $[t_1, t_m]$, 因此就假设了网络变化过程是发生在没有被观测到的那些数据收集的時刻之间。这个假设是由 Sørensen 和 Hallinan (1976; Holland and Leinhardt, 1977) 等提出来的。这些作者也提出, 在这个变化过程中, 在任何时间 t 的情况下, 都至多有一个关系变量 $X_{ij}(t)$ 在变化。这就将变化过程分解为它的最小可能成分 (smallest possible constituents), 排除了同时创建一系列关系的协调形式, 正如在一见钟情的恋人或在默契十足的一群朋友中那样。这是一个合理的要求, 它会大大减少建模的复杂性。Sørensen 和 Hallinan (1976) 的模型聚焦在三方谱系 (census) 的动因上 (Holland and Leinhardt, 1975), 这个模型具有一个矢量集合, 该集合将三方谱系的结果定义成了结果空间。然而, 这个模型是不完整的, 因为它不能详尽说明一个网络中三方组之间的依赖性。Hallinan (1979) 提出了一个类似但更为简单的模型, 该模型聚焦于二方谱系 (dyadic census)。Holland 和 Leinhardt (1977) 提出了将网络动态表达为连续-时

间马尔可夫过程的一般模型,这个模型中的关系是逐个变化的。但是,他们并没有详述如何去设定网络中的关系依赖性的方式。

关系之间的依赖性

Katz 和 Proctor(1959)的马尔可夫链模型假设关系变量之间要独立,这种变量在每个后续观测点上都依据马尔可夫链变化。当然,关系的独立性只不过是一个稻草人(straw-man)假设,因为它违反了社会网络分析的基本思想。对这个假设的第一次放宽是假定二方组或 $(X_{ij}(t), X_{ji}(t))$ 之类的一对对关系变量(pairs of tie variables)的独立性。Wasserman(1977,1979,以及其他出版物)针对历时态模型、Hallinan(1979; Leenders,1995 和其他出版物)针对连续-时间的马尔可夫过程、Wasserman 和 Iacobucci(1988)针对离散-时间的马尔可夫过程等都作出了这一假设。

独立二方组假设将随机过程分裂为 $n(n-1)/2$ 个独立的子过程。这是易处理的,但是在上文举例时提到的三个基本的成分过程(component process)(互惠性、可传递性和马太效应)中,它只反映了互惠性。Wasserman(1980)提出了所谓的流行度模型(popularity model),可以认为这个模型反映了马太效应,却没有反映互惠过程。在这个模型中,随机邻接矩阵 $(X_{ij}(t))$ 的各行是独立的,这就再次简化了这个模型,使其易于处理。

针对等级(rankings)形式的数据,有学者提出了允许三方组与更高阶依赖性存在的随机模型,正如 Snijders(1996)在 Newcomb-Nordlie 数据中,Snijders 和 Van Duijn(1997)以及 Snijders(2001)针对有向图数据所做的那样。本章后文将详述后一个模型。

无尺度网络

在 De Solla Price(1976),Barabási 和 Albert(1999),Dorogovtsev 等(2000)提出的模型中,将新的点加入现有的网络中,每个新点都以某种概率与 m 个现有的点相联系,这个概率线性地依赖于现有点的度数。这就导致了所谓的无尺度网络(scale-free networks),该网络中的度数呈现出一种幂分布(power distribution)。从人类个体之间的大多数网络类型看,这好像并不现实,因为各种约束条件会限制极高度数发生的频率。

网络动力学的随机模型

在 1990 年代之前,网络动力学随机模型并没有得以快速发展,原因之一是,描述网络特征的依赖性结构(dependence structures)太复杂了,那些貌似合理的网络动力学模型只能作为计算机模拟模型才能实现(至少在现有的知识状态下似乎是如此)。正如在基于能动者的模型(agent-based models)中所提到的那样,

在前计算机时代,我们是无力在数据分析中作精确计算的。

在这一节中,我们首先介绍基于关系的(tie-based)动态模型,然后介绍基于行动者的模型(actor-based models)。前者比较简单,后者更接近于社会科学中的大多数理论。应该将两者都看作是可以通概率规则来定义的过程模型,这个概率规则能反映网络怎样从一个观测值发展到下一个观测值。从技术上讲,上述所有模型都是在有向图空间上的马尔可夫过程。它们都是连续-时间模型,这意味着时间是以一种无限小的方式递增的,某种变化会在任意时刻发生。为了使模型相对简单,就要给出一个假定,这一点最先由 Holland 和 Leinhardt (1977) 提出。该假设认为,在任何给定的时刻(在任何瞬间),只有一个关系可能发生。这就在最小可能的步骤(steps)中将网络动态作了分解。它假设行动者不具有同时协调(coordination)的可能性:行动者是相互依赖的,因为行动者之间相互反应(参见 Zeggelink, 1994),而不是相互协调。

基于关系的模型

要建构非常具有一般性的依赖性结构的动态网络模型,最简单的进路就是构想这样一个模型,选定一个随机对(i, j),确定以多大概率来改变关系变量 X_{ij} 的值:是创建一个新关系(将值 0 改变为 1),还是终止一个现有的关系(将值 1 变为 0)。这一改变的概率可以由这个网络的各种函数决定,因而反映的是几种“机制”、理论与限制等的联合体。在技术上,它基于的是指数随机图模型的马尔可夫过程思想与 Gibbs 的抽样思想结合。我们还是先考虑这样一个例子,它有四种驱动网络运动的理论或机制成分(component),这四种成分是:趋于指定的度数均值、互惠性、传递性以及马太效应的倾向。在这里,马太效应被解释为自我加强(self-reinforcing)的流行过程,它对点入度的分散性有贡献。所有这些都理解是随机且非确定性的倾向。可以通过下列网络统计量反映这四种成分:

$$L(X) = \sum_{ij} X_{ij} \quad \text{关系数} \quad (33.1)$$

$$M(X) = \sum_{ij} X_{ij} X_{ji} \quad \text{互惠二方组数} \quad (33.2)$$

$$T(X) = (1/6) \sum_{ij} X_{ij} X_{jh} X_{ih} \quad \text{传递性三方组数} \quad (33.3)$$

$$V_{in}(X) = (1/n) \sum_i (X_{+i} - X_{+})^2 \quad \text{点入度的方差} \quad (33.4)$$

其中

$$X_{+i} = \sum_j X_{ji} \quad i \text{ 的点入度} \quad (33.5)$$

$$X_{+} = (1/n) \sum_i X_{+i} \quad \text{平均度数} \quad (33.6)$$

如果这些网络动力的趋向是分别增加这四个统计量的值,那么这种趋势就会将这个网络过程分别导入到一个密度更高、互惠更多、传递性更强或点入度(流行性)差异性更大的方向上去。通过下列方式,用一个模型就可以实现这一点。首先,让我们重新将点入度的方差 $V_{in}(X)$ 改写如下:

$$V_{in}(X) = (1/n) \sum_i X_{+i}^2 - \bar{X}_{+}^2$$

$$\begin{aligned}
&= (1/n) \sum_i X_{+i} (X_{+i} - 1) + \bar{X}_+ - \bar{X}_+^2 \\
&= (1/n) S_2(X) - \bar{X}_+ (\bar{X}_+ - 1)
\end{aligned}$$

其中, $S_2(X)$ 是有向图 X 中的 2-入星 (two-in-stars) 数, 即满足 $j \rightarrow i; k \rightarrow i$ 以及 $j \neq k$ 的 i, j, k 的构型数。这表明, 对于一个固定的平均度数 X_+ 来说, 有一个大点入度的方差 $V_m(X)$ 完全等同于有大量 2-入星数的 $S_2(X)$ 。下文就一直用 2-入星, 而不再用表达马太效应的点入度方差了。

遵循这四种理论成分的倾向在强度上是各不相同的, 为了容纳这些不同的强度, 可以定义线性组合如下:

$$F(x; \beta) = \beta_1 L(x) + \beta_2 M(x) + \beta_3 T(x) + \beta_4 S_2(x) \quad (33.7)$$

其中, 参数 β_k 的值决定了这四种倾向的强度, x 是任意一个有向图。现在就可以定义网络的变化过程了, 它是通过改变 (“切换”) 单个关系变量 $X_{ij}(t)$ 来操作的, 这个操作根据系数 b_k 的值促成了统计量 L, M, T 和 S_2 的改变。这要通过下列算法来实现, 这些算法会表明, 当关系发生改变时, 现有的图 $X(t)$ 是怎样转换为下一幅图的。

算法 1: 基于关系的网络动力学

对于有向图 x 来说, 除了下面的关系变量外, 可以将 $x^{(ij+)}$ 和 $x^{(ij-)}$ 定义为等同于 x 的两个图, 即对于有序对 (i, j) 关系来说, 如果 $x^{(ij+)}$ 确有 $i \rightarrow j$ 这个关系, 那么 $x^{(ij-)}$ 就没有这个关系。换句话说, $x_{ij}^{(ij+)} = 1$ 和 $x_{ij}^{(ij-)} = 0$ 。

1. 选择一个有同样概率的随机点对 (i, j) , 假定 $i \neq j$ 。
2. 定义 $x = X(t)$ 。
3. 定义下式

$$P_{ij} = \frac{\exp(f(X^{(ij+)}) ; \beta)}{\exp(f(X^{(ij+)}) ; \beta) + \exp(f(X^{(ij-)}) ; \beta)} \quad (33.8)$$

的概率为 π_{ij} , 将下一个网络选定为 $x^{(ij+)}$; 定义概率为 $1 - \pi_{ij}$, 将下一个网络选定为 $x^{(ij-)}$ 。

4. 时间变量 t 的增量为 Δt , 它是一个随机变量, 服从参数为 ρ 的指数分布。

这是一个网络动力学模型, 它与 Frank 和 Strauss (1986; Frank, 1991; Wasserman and Pattison, 1996) 所提出的指数随机图模型关系密切。为了阐明这个网络动力学模型与指数随机图模型的关系, 一个基本的问题是, 上述指数随机图分布的条件下, 假定除了 $i \rightarrow j$ 这个特殊关系是否存在以外, 假定我们是知道整个网络 x 的, 那么 (33.8) 就是关系 $i \rightarrow j$ 存在的条件概率。而指数随机图分布是由下面的概率函数定义的:

$$P(X = x) = \frac{\exp(f(x; \beta))}{C} \quad (33.9)$$

其中, C 是标准化常数:

$$C = \sum_x \exp(f(x; \beta))$$

504 这是在全部有向图 x 上的总和。因此,在对关系 $i \rightarrow j$ 是否存在进行选择时,上面的动态算法利用了模型(33.9)下该关系的条件概率,这个条件就是该关系之外的整个网络构型。从马尔可夫过程的一般原理,或具体而言从吉布斯的抽样原理出发(Geman and Geman, 1983),就可以认为,当这个算法被无限反复使用时, $X(t)$ 的分布(其中,无限反复意味着 t 趋于无穷大)就趋于有概率函数(33.9)的分布。从这个模型中可以得到随机抽取(random draws)的多种标准算法,这个动力学算法就是其中之一(见 Snijders, 2002; Robins et al., 2005)。

通过选取(33.7)中的参数 β_k ,就可以用不同强度的密度、互惠性、传递性与自我强化的流行性等倾向去选择不同的模型。例如,当 $\beta_2 = \beta_3 = \beta_4 = 0$ 时,就得到一个随机图,“鄂尔多斯-仁义图”(Erdős-Rényi),“伯努利图”(Bernoulli)。当 $\beta_3 = \beta_4 = 0$ 时,这是 Wasserman(1977, 1979)互惠性模型的一个特例,它有着独立的二方组。当 $\beta_3 \neq 0$ 或 $\beta_4 \neq 0$ 时,这种二方组之间的独立性就被破坏了。当 $\beta_2 = \beta_3 = 0$ 时,就得到了 Wasserman(1977, 1980)的名气模型(popularity model)。 β_3 及 β_4 的值可能为正,这样的模型可表达传递性与马太效应等倾向。

基于行动者的模型

网络分析的挑战之一是如何将能动性(agency)融入网络模型。Emirbayer 和 Goodwin(1994)令人信服地阐述了这一点,他们同样强调了文化的重要性,不过本章必须将这一点放置一边。在一个统计模型中,要将能动性与结构结合在一起,一种自然的方式就是采用网络动力学模型。在这个模型中,关系的改变是由行动者发起的。这样的模型可能是一个表达与检验社会科学理论的良好工具,行动者在这种理论中发挥着重要的作用(参见 Udehn, 2002; Hedström, 2005)。Snijders(1996)针对等级网数据、Snijders 和 van Duijn(1997)针对二值网数据都提出了基于行动者的模型。接下来,我们会介绍 Snijders(2001)的模型。他们(Snijders et al., 2010)对这些模型做了一个入门教程,其中包括怎样使用与设定它们的实际建议。

在连续-时间的表现(representation)中,关系一次只能改变一个,而且改变的概率要顾及当前的全部网络构型,在这些限制条件下的模型就具有了基于行动者的性质,这意味着该模型中的行动者好像已经控制了他们发送的关系似的。这个模型设定用了两个函数。一个是所谓的比率函数(rate function) $\lambda_i(x; \alpha)$,它由行动者 i 和目前的网络状态 x 所决定,表明了每个单位时间的频次,行动者 i 就是以这个频次获得了改变一个发出关系的机会;另一个是目标函数(objective function) $f_i(x; \beta)$,可以将它解释为是一种关于网络状态 x 对行动者 i 有多大吸引力的测量。比较中立地讲,目标函数是这样的:当作出一种改变时,行动者会以较高的概率移向网络 x , x 的目标函数 $f_i(x; \beta)$ 也较高。可以用统计参数 α 和 β 来反映包含在比率函数与目标函数中的各种不同成分的强度。(创建一个新关系与终止一个现有关系之间有反对称性,至于在不含反对称性时对该模型的诸多

扩展,参见上文提及的满意 (gratification) 或捐赠 (endowment) 函数文献中的讨论)。

这个算法只是根据概率分布来构想的,但是也可以这样来解释,即它反映了嵌入于一个网络中的行动者。网络是他们彼此之间正在改变的环境(参见 Zeggelink, 1994),在其发出的每个关系中,他们都以某种比率 (rate) $\lambda_i(x; \alpha)$ 在改变(它可能是常数,但如果比率函数是关于 x 的一个非恒量函数,它就是变化的),这样就能在假定随机扰动 (disturbance) 被加入目标函数的条件下,优化他们在作出改变之后获得的目标函数值了。可以称该算法为目标函数的短视 (myopic) 随机优化,在网络形成的博弈论模型中常常会用到它(如 Bala and Goyal, 2000)。

算法 2: 基于行动者的网络动力学

对于有向图 x 来说,除了有序对 (i, j) 这个关系变量和 $x_{ij}^{(ij\pm)}$ 中恰好是 x 中关系变量的反变量(即在 $x_{ij}^{(ij\pm)} = 1 - x_{ij}$ 的意义上)之外,针对所有其他关系变量,定义 $x_{ij}^{(ij\pm)}$ 为等于 x 的图。

定义 $x_{ij}^{(ij\pm)} = x$ (作为一个方便的没有未指明意义的形式定义)。

1. 定义 $x = X(t)$ 。
2. 对于 $i \in \{1, \dots, n\}$, 定义

$$\tau_i = \frac{\lambda_i(x; \alpha)}{\sum_{h=1}^n \lambda_h(x; \alpha)} \quad (33.10)$$

以概率 τ_i 选择行动者 i 。

3. 对于 $j \in \{1, \dots, n\}$, 定义

$$\pi_{ij} = \frac{\exp(f_i(x^{(ij\pm)}; \beta))}{\sum_{h=1}^n \exp(f_i(x^{(ij\pm)}; \beta))} \quad (33.11)$$

以概率 π_{ij} , 选择下一个网络为 $x^{(ij\pm)}$ 。

505

4. 时间变量 t 的增量为数量 Δt , 它是一个随机变量, 服从参数为 $\sum_{h=1}^n \lambda_h(x; \alpha)$ 的指数分布。

这一指数函数的性质意味着方程式(12)可以改写为:

$$\pi_{ij} = \frac{\exp(f_i(x^{(ij\pm)}; \beta) - f_i(x; \beta))}{\sum_{h=1}^n \exp(f_i(x^{(ij\pm)}; \beta) - f_i(x; \beta))} \quad (33.12)$$

也就是说, 给定变化的概率是由目标函数的增加而单调决定的, 而目标函数又是由这一改变引起的。这表明, 对于行动者 i 来说, 如果他目前的网络状态 x 接近于目标函数 $f_i(x; \beta)$ 的最优值, 他就很有可能不改变, 因为选择维持现状至下一个网络 $x^{(ij\pm)} = x$ 的概率 π_{ii} 较高。

模型设定

在基于关系与基于行动者的模型中, 为了设定模型, 研究者必须分别设定

$f(x; \beta)$ 或 $f_i(x; \beta)$ 函数 (在基于行动者的模型中, 还要设定改变比率 $\lambda_i(x; \alpha)$)。这个选择应该建立在要研究的主题、理论思考与假设的基础之上。这里仅讨论基于行动者的情况。

如在广义线性建模中一样, 如下线性组合会提供一类方便的函数:

$$f_i = \sum_k \beta_k s_{ki}(x) \quad (33.13)$$

其中, $s_{ki}(x)$ 是网络函数, 从行动者 i 的视角看, 在许多情况下, 它就是 i 的个体网函数。该函数与 (33.7) 相类似, 但是, 现在它被定义为是基于行动者的模型, 即:

$$f_i(x; \beta) = \beta_1 \sum_j x_{ij} + \beta_2 \sum_j x_{ij} x_{ji} + \beta_3 \sum_{j,h} x_{ij} x_{jh} x_{ih} + \beta_4 \sum x_{ij} x_{hj} \quad (33.14)$$

该模型很像 (33.7) 中的那四项, 不过现在是从行动者 i 的视角上看待的, 这四个统计量分别代表关系数、互惠关系数、传递性三方组 $\{i \rightarrow j \rightarrow h, i \rightarrow h\}$ 数以及补加的行动者 j 的点入度 $\sum_h x_{hj}$, 行动者 i 向 j 发出了一个关系。基于关系的模型具有设定 (33.7), 基于行动者的模型具有设定 (33.14), 两个模型都定义了网络动力学的概率分布, 这些定义的方差相似, 却仍然有差别; 如果想识别出任何拟合上的差异, 就要在基于关系的与基于行动者的设定之间作抉择, 就必须建立在理论偏好或经验拟合的基础之上。

在这里, 这个模型设定只是一个例子, 用这个例子可以说明, 这些模型是怎样被用来通过四个参数 β_1 至 β_4 来表达一些趋向的, 即趋于一个给定度数均值、互惠性、传递性闭合以及对已受欢迎的行动者的偏好的趋势。值得注意的是, 这四个统计量高度相关, 这意味着虽然参数 β_2 、 β_3 和 β_4 可以被用来检验各自的趋势, 但是, 一些统计量的概率分布是可以从这个网络中计算出来的, 从这些参数对于这些统计量概率分布的意义来讲, 这些参数是互相勾连的 (collaborate)。在所有实际的情况中, 我们都希望能控制平均度数, 因此, 总体上, 检验关于 β_1 的假说似乎不是非常地有意义。

在目标函数的表达式 (33.13) 中, 行动者 i 的个体网的许多其他统计量都可以被用作 $s_{ki}(x)$ 。这样的统计量被称为效应 (effects)。由于行动者只能控制发出的关系变量, 所以, 这里重要的是, 效应是怎样依赖于发出的关系变量 x_{ij} 的; 效应只依赖于进入的 (incoming) 关系变量, 它们对条件概率 (33.11) 没有任何影响。Snijders 等 (2010) 充分地讨论了可以加入的许多统计量, 它们能反映有各种理论指向的网络倾向, 也会有助于充分表达关系变量之间的依赖性。下面是一个不完全的提纲。

1. 两个基本的统计量

(a) 点出度 $\sum_j x_{ij}$, 它的参数——如例子 (33.14) 中的 β_1 ——可用来拟合平均度数的水平与倾向; 大多数其他的统计量会与平均度数相关, 这意味着这个参数的精确值会高度依赖于其他参数。

(b) 互惠度定义为 $\sum_j x_{ij} x_{ji}$ (即行动者 i 涉及的互惠关系数)

也包括在(33.14)中;在几乎所有的有向社会网络中,互惠性都是一个基本的倾向,将这个效应包含在内,就可以更好地表达趋于互惠的倾向。 506

2.网络的局部结构是由三方组决定的,即三个点上的子图(Holland and Leinhardt,1975)。三方组中关系之间的主依赖性(main dependencies)是由下列两个量把握的:

(a)传递性:“朋友的朋友成为朋友,或继续为友”的倾向,用个体网中的传递性三方组数 $\sum_{j,h} x_{ij}x_{jh}x_{ih}$ 来表达,如(33.14)中包含的第三项所示。

(b)3-回路(three-cycles):形成回路 $i \rightarrow j \rightarrow h \rightarrow i$ 的倾向,用 $\sum_{j,h} x_{ij}x_{jh}x_{hi}$ 来测量。这可以反映一般化的交换(generalized exchange)(Bearman,1997);但是常见的情况是这个效应有个负号,这意味着3-回路往往会被避免(Davis,1970),这是一种局部等级性(local hierarchy)的符号。

3.点入度与点出度是个体网位置的两个基本面向,关系的创建与终止或多或少取决于所涉及的行动者的度数。这可以用与度数相关的效应来表达。基本的度数效应有:

(a)点入度知名度(in-degree popularity),它表示的是,当前点入度高的行动者作为新关系的接收者在多大程度上更知名,这就是前文提到的马太效应和(33.14)中的第四项。

(b)点出度活跃度(out-degree activity),它表示的是,目前点出度高的行动者是否有更大的创建而非终止关系的倾向。

(c)点出度知名度。

(d)点入度活跃度。

同样,高阶的度数效应(higher-order degree effects),如基于度数的配套性 assortativity)也有可能被纳入进来,它取决于两者在度数上的结合,表达了一种形成并保持关系的或强或弱的倾向。

4.除了这些基于网络结构本身的效应之外,把那些由行动者属性决定的统计量包含进来也很重要,如他们的人口统计学特征、资源的指标等。可以将一个给定的行动者变量作为一个自我效应纳入进来,它反映了该变量对发送关系倾向的影响,并且作为一个他者效应,它也反映了对接收关系倾向的影响。此外,发送者与接收者的结合通常是重要的,如他们在突出属性上的相似性反映了趋于同质性的倾向(McPherson et al.,2001)。

5.也可能把行动者对(pairs of actors)的属性包括在内,它们可能是其在一个不同网络中的相关性。例如,这样的二方组协变量(dyadic covariates)可能表达了二方组关系的见面机会、成本或收益等。

基于行动者模型的统计推断

当参数 α 和 β 取不同值时,就会得到不同的网络动力学。对于一个给定的历时网络数据集来说,问题是如何确定这些参数值,以达到模型与数据之间的充分拟合(good fit)。这就是常见的统计推断问题。这里存在一个技术上的困难,即对于模型与数据之间的拟合来说,还不存在任何一种便利的、可计算的指标,如方差分析中的平方和那样。实际上,对模型性质的评估只能通过计算机模拟。基于行动者的模型的确可以被看作一个基于能动者的计算模型(参见 Macy and Willer, 2002),这意味着要去模拟网络演化的方式。

估计

文献中基于行动者的模型提出了三种参数估计方法。第一种是积矩法(the Method of Moments)(Snijders and van Duijn, 1997; Snijders, 2001),它适当选择一系列历时态网络数据集的统计量,每个统计量对应于一个待估计的参数,确定这些参数,使得对于这些统计量来说,在观测值与来自该模型的全部模拟总体的期望值之间能完美拟合(perfect fit):期望值应该等于观测值。实际上,这只能通过一种随机逼近算法近似地达到,由于实际做的模拟次数有限,结果中就会有某些随机性。

第二种是 Koskinen 和 Snijders(2007),Schweinberger(2007)提出的贝叶斯程序(Bayesian procedures)。贝叶斯方法假定参数的概率分布反映了某些先验的信念(prior beliefs)或先验的无视(prior ignorance),然后计算或粗略估计所谓的参数后验分布。给定观察数据后,后者就是参数的条件分布,它反映了先验的信念是怎样通过经验观察转换的。第三种是 Snijders 等(2010)提出的一种逼近最大似然估计量的算法。这个算法需要模拟可能的连续-时间过程,该过程可能从一期观测引向下一期观测,用一种适当的平均法就能近似地估计这些参数。

对于不那么小的数据集来说,如果这个模型能做到非常近似,那么这三种方法就会产生相似的估计结果。

检验

积矩法和最大似然法都是估计方法,与这些方法相关联,遵循建构统计检验的一般原则(如参见 Cox and Hinkley, 1974),就有一些检验参数的统计假设的程序了。最直接的方式常常会用参数估计值及其标准误。例如,为了检验

$$H_0: \beta_k = 0$$

这个原假设,那么检验所用的统计量就是这个估计值与标准误之比

$$t = \frac{\hat{\beta}_k}{s.e.(\hat{\beta}_k)} \quad (33.15)$$

这可以用一个标准正态参考分布来检验,也可称之为 t -检验,因为它建立在 t -比率的基础之上。用一种类似的方式可推导出多参数检验。用积矩法得到估计值的检验,可以被称为是沃尔德式检验(Wald-type tests),用最大似然法得到估计值的检验则被称为是沃尔德检验。

还有一种不同的假设检验方式,它不需要对有待检验的参数进行估计。这就是 Rao 的效率分检验(Rao's efficient score test)一般原则。对于积矩法来说,它需要作一种特殊的适应性改变,才能产生 Schweinberger(2008)提出的得分式检验(score-type test)。对于这些模型来说,得分或得分式检验都有一种特殊实用的优点。这是因为在模型相对复杂的情况下,给定数据信息量,参数估计的蒙特卡罗算法可能会不收敛;因此,即使没有作参数估计,得分法也可以提供检验。

与最大似然估计相关的是似然率检验。Snijders 等(2010)对此提供了一种算法。

与最大似然估计与检验算法相比,目前可用的积矩法算法更节省时间。不过,这是一个发展迅速的领域,现有算法的计算效率可能会改变。

网络与行为的动力学

使网络变得重要起来的因素常常是个体的行为与其他个体的结果,它们以某种方式与行动者的网络嵌入性相关(如可参见 Granovetter, 1973; Burt, 1992; Lin et al., 2001)。然而,这样的个体特征在解释网络动力学时也发挥着作用。因此,我们遭遇到了一种可以将网络与行为都看作是因变量的情形,它们的变化相互依赖,这里用了“行为”这一术语,它是行动者的相关可变特征的简略表达,也可以指态度、操行等。这里的假定是,行为变量是定序的离散变量,值为 1、2 等,直至某个最大值,如一个从左至右的量表上的多级酒精消费或多级政治态度;二值变量则是一种特殊的情形。网络动力对网络与行为的依赖性被统称为社会选择过程(social selection process),行为动力对网络与行为的依赖性被称为社会影响过程(social influence process)(An, 本书)。

在有关系的行动者之间,社会影响与社会选择都能引起相似性,可以描述性地称之为网络自相关(network autocorrelation)(Doreian, 1989; Leenders, 1997)。这个网络自相关主要由影响引起,还是由选择引起,这是一个重要问题。Ennett 和 Bauman(1994)针对吸烟、Haynie(2001)和 Carrington(本书)针对少年犯罪行为说明了这一点。

基于行动者的模型

要回答这样的问题,用过程模型可能会有帮助,该模型反映了关系变量及行动者行为变量之间的互依性演化(interdependent evolution)。这里,很自然地会想到基于行动者的模型;在 Snijders 等(2007; Steglich et al., 2010)的研究中,这种

模型就被设定了。他们假设,一个行动者发出的关系及其行为是在该行动者的控制之下的,它会受到各种限制。

过程模型假定,在任意时刻,可能改变的要么是一个网络关系,要么是一个行动变量。行动者分别有网络与行为的比率函数与目标函数。网络与行为是由不同的过程潜在地控制着的,这一点可以被证明,例如,有学者将网络选择与行为选择看作是由不同的决策框架确定的(Lindenberg, 2001)。

将这些变化分解为最小可能的步骤,在这里,它意味着,对于一个给定的(“无限小”)瞬间,一个行动者改变他或她行为变量的可能性只限于在定序尺度上左右移动一个类别。

在向量 $Z(t)$ 中可以采集到行动者 i 在时间 t 上的行为,我们标记为 $Z_i(t)$ 。现在假设,网络变化的概率取决于该网络的目前状态及行为;行为变化的概率取决于该行为的目前状态与网络。用 $f_i^X(x, z; \beta)$ 表示行动者 i 关于网络的目标函数,用 $f_i^Z(x, z; \beta)$ 表示关于行为的目标函数。与网络的目标函数相类似,行为的目标函数也如此,这样的话,趋于更高目标函数值的改变就比趋于更低目标函数值的改变更有可能发生。用 $\lambda_i^X(x, z; \alpha)$ 表示行动者 i 对于网络改变的比率函数,用 $\lambda_i^Z(x, z; \alpha)$ 表示对于行为改变的比率函数。

算法 3: 基于行动者的“网络与行为动力学”

对于网络来说,其采用的算法定义等同于基于行动者的网络动力学的算法。对于行为来说,就任何行动者 i 和潜在的增量 d 来说,我们将 $z^{(i+d)}$ 定义为行为向量,除了将 d 加到第 i 个坐标上之外,它都等于 z 。 $z_i^{(i+d)} = z_i + d$ 。

1. 定义 $x = X(t)$, $z = Z(t)$ 。
2. 以概率 ϕ_X 计算如下比率

$$\phi_X = \frac{\sum_{h=1}^n \lambda_h^X(x, z; \alpha)}{\sum_{h=1}^n (\lambda_h^X(x, z; \alpha) + \lambda_h^Z(x, z; \alpha))} \quad (33.16)$$

转到第 3 条,迈一个网络步(make a network step); 否则(以概率 $1 - \phi_X$) 转向第 5 条,迈一个行为步(make a behavior step)。

3. 对于 $i \in \{1, \dots, n\}$ 来说,定义

$$\tau_i^X = \frac{\lambda_i^X(x, z; \alpha)}{\sum_{h=1}^n (\lambda_h^X(x, z; \alpha))} \quad (33.17)$$

以概率 τ_i^X 选择行动者 i 。

4. 对于 $j \in \{1, \dots, n\}$ 来说,定义

$$\pi_{ij}^X = \frac{\exp(f_i^X(x^{(ij\pm)}, z; \beta))}{\sum_{h=1}^n \exp(f_i^X(x^{(ij\pm)}, z; \beta))} \quad (33.18)$$

以概率 π_{ij}^X 选择的下一个网络为 $x^{(ij\pm)}$ 。

转到第 7 步。

5. 对于 $i \in \{1, \dots, n\}$ 来说,定义

$$\tau_i^Z = \frac{\lambda_i^Z(x, z; \alpha)}{\sum_{h=1}^n (\lambda_h^Z(x, z; \alpha))} \quad (33.19)$$

以概率 τ_i^Z 选择行动者 i 。

6. 对于 $d \in \{-1, 0, 1\}$, 如果 $z_i + d$ 处在 Z 允许的范围之内, 则定义

$$\pi_{id}^Z = \frac{\exp(f_i^Z(x, z^{(i+d)}; \beta))}{\sum_{k=-1}^1 \exp(f_i^Z(x, z^{i+k}; \beta))} \quad (33.20)$$

$z_i + d$ 在允许的范围之外的 d 值不包含在分母之内。

以概率 π_{id}^Z 选择下一个行为矢量为 $z^{(i+d)}$ 。

转向第 7 步。

7. 令时间变量 t 的增量为 Δt , 它是一个随机变量, 服从参数为 $\sum_{h=1}^n (\lambda_h^X(x; \alpha) + \lambda_h^Z(x; \alpha))$ 的指数分布。

这个 $d=0$ 的选择意味着行动者 i 具有改变她或他行为的机会, 但是却克制了自己没有这样去做。对于 $d=-1, +1$ 来说, 与邻域状态的目标函数 $f_i^Z(x, z^{(i+d)}; \beta)$ 相比, 目前状态的目标函数 $f_i^Z(x, z; \beta)$ 取值较高。因此, 行动者克制自己不改变行为的概率较高。

模型设定

对于行为来说, 最方便的目标函数表达式也是一个线性组合

$$f_i^Z(x, z; \beta) = \sum_k \beta_k^Z s_{ki}^Z(x, z) \quad (33.21)$$

其中, $s_{ki}^Z(x, z)$ 是关于行为与行动者 i 其他特征的函数, 但是, 它也可能取决于个体网, 以及与行动者 i 有关系的那些行动者的行为。在关于选择与影响的研究中, 通过设定网络动力学模型, 是可以对依赖于行为 (behavior-dependent) 的选择部分进行建模的, 例如, 可以设定这样一项, 它表示的是与其他行动者建立关系的偏好, 这些行动者在行为变量 Z 上表现类似 (同质性)。

用行为目标函数中的某些适当项是可以对依赖于网络 (network-dependent) 的影响部分进行建模的。该函数设定的一个基本例子是

$$f_i^Z(x, z; \beta) = \beta_1^Z z_i + \beta_2^Z z_i^2 + \beta_3^Z z_i \left(\frac{\sum_j x_{ij} z_j}{\sum_j x_{ij}} \right) \quad (33.22) \quad 509$$

前两项代表了行为 Z 的一个二次偏好函数。如果偏好是单峰的, 那么二次项的系数 β_2^Z 就是负的。然而对于成瘾 (addictive) 行为来说, 这个系数可能为正。第三项表明, 对于行为 z_i 来说, 行动者 i 的“值”取决于 i 向其发出关系的那些行动者的平均行为。

实例

因篇幅有限,本章并没有给出一个详述的经验实例。前文提到的方法论文章进一步解释了网络动力学的基于行动者的模型,其中有一些例子可以查阅。其他发表过的网络动力学实例(以行动者总体的年龄排序)有:学龄前儿童之间的朋友关系动力学,包括互惠性、传递性与流行性效应(Schaefer et al., 2010);人格特征决定青少年朋友关系的动力学方式(Selfhout et al., 2010);大学生之间的那些可察觉的与不可察觉的属性对其朋友关系的动力学(van Duijn et al., 2003);管理人员的流动性如何影响到公司之间的关系(Checkley and Steglich, 2007)。

有关网络与行为的联合动力学(joint dynamics)案例只是在最近才有发表,因为它是一个新近提出的模型。一些例子如下:

Burk 等(2007)研究了青少年友谊与犯罪行为动力学方面的影响与选择过程。Steglich 等(2010)在一所中学的同期群中,研究了友谊、吸烟及喝酒行为的共同演化(co-evolution)。Mercken 等(2009)作了一项大规模的研究,研究涉及6个国家中的70所学校的网络,探讨了青少年中吸烟启动(smoking initiation)的影响与选择过程。还有学者(De Klepper et al., 2010)针对一所海军军官学校,研究了友谊与军纪演变中的互依性。

SIENA 程序

网络动力学的基于行动者的模型,以及针对网络与行为的动力学模型,都可以在“对经验网络分析的模拟探究”(Simulation Investigation for Empirical Network Analysis)即 SIENA 程序中执行。最初,它是一个经由程序 StOCNET 的脱机程序,有用户界面,自 2009 年以来,它就是一个在统计系统 R 内的程序包了(R Development Core Team, 2009),名为 RSiena。R 系统及其程序包是免费软件,在 Windows、Mac 和 Unix/Linux 系统上运行。它有一个全面并频繁更新的手册(Ripley and Snijders, 2010)。这个手册对安装与使用 RSiena 给出了详细的操作指南。

第一项要求是安装 R、程序包 RSiena 和一些辅助性程序包,正如在 RSiena 手册中描述的那样。如果要使用它,显然可以在不了解任何 R 的情况下,利用一个图用户界面(graphical user interface)来操作 RSiena;安装结束后,就不必再操作 R 了。一旦安装结束,RSiena 就可以通过两种方式运行:

1. 运行 R, 装上程序包 RSiena 以及辅助程序包,从 R 内通过命令 `siena01Gui()` 运行 RSiena 图用户界面。它会提供 RSiena 的基本功能,并且有可能将 RSiena 的使用与任何其他 R 程序包的使用整合在一起。它的优点是无需任何 RSiena 命令的知识。

2. 运行 R, 装上程序包 RSiena 和辅助程序包,通过使用 RSiena 的 R 命令

来运行 RSiena。对于熟悉 RSiena 的用户,这是最佳选择。

Snijders 等(2010)的文章是基础性文献,最好将它作为方法论教程。如果要了解数据形式与软件操作,可利用(Ripley and Snijders,2010)(或其近期版本)。

展望与讨论

社会网络分析的统计方法以一种令人满意的方式表达了网络的依赖性,这些方法只在最近才有可能被使用。这里所介绍的分析网络演化及网络与行为共同演化的方法,能够使研究者检验那些关于网络演变的竞争性及互补性的理论。现在,需要从理论及方法论的视角更多地反思如何将统计进路与网络进路相结合。在结构与位置分析上,网络进路成果颇丰。相比之下,统计进路则具有简约性的传统,它常常将假设检验的模型设定限制在对检验变量连同少量控制变量的选择上。统计进路中的大多数研究是完全个体性的(purely individualistic),忽略了区别出多类分析单位的重要性,在“控制 A 后, X 就导致 Y”这样的方案(scheme)下,所构想的假设是唯一的,无须多费周折。对于诸如杂志的审阅者与编辑这样的守门员来说,让他们相信网络进路(其中的理论与统计模型比较复杂)的重要性是有难度的。 510

下面谈谈完全个体性进路的两个主要局限。首先,大多数的网络研究都是观测性的而非实验性的,这意味着分析方法中一定要包含对竞争性假说与理论的适当控制,对观测变量之间统计依赖性的良好设定,这些对于得到可信的结论来说都是必要的。在网络现象中,内生性的(也称为自我参照的、涌现的和自组织的反馈)过程是必要的,这些过程会导致变量之间的依赖性,而不是导致某个测量变量(measured variable) X 对一个因变量 Y 的效应。不能恰当地设定这样的依赖性会导致假设检验不足以控制竞争性的理论。

其次,网络依赖性有可能是一个深藏着有趣的理论与假设的宝地,沿着像 Hedström(2005)分析社会学那样的理论路线,将网络进路注入理论思考与统计假说检验之中,就可以更好地解释经验现象,也可以促进诸如公共健康领域中的干预。在更早一些的情境分析中,通过多层建模就已经开始一种类似的发展了。现在,分析性地使用几类单位已经被广为接受,该做法被认为是效果好,甚至有必要,虽然目前还没有被用于实践中,例子可参见(Sampson et al., 2002; O'Campo, 2003)。

当网络动力学的统计建模沿着三条路线进一步发展时,这些理论-方法论上的进步就会更容易一些。这三条路线是:第一,适于更丰富的数据结构的模型、更少限制的模型和更丰富的统计程序。至于数据结构,当还在网络面板设计的范围内时,人们可能思考的是如何将这类数据建模扩展到诸如多值网络、多变量网络与无向图网络这样的数据类型中去。然而,发展不应该仅限于面板设计。在研究两个组织之间的网络时,有时有些观测时点(moments)间隔太紧,这样的话,可以用网络自回归(autoregressive)模型,利用前面的网络状态观测值来预测

下一个观测值,如某些学者(Leenders, 1997; Gulati and Gargiulo, 1999)所做的那样,这种近似估计是合理的;有时候,观测值甚至会提供一个连续的关系创建记录,虽然并不总是关于关系终止的记录,如 Hagedoorn (2002) 所作的研究那样。第二,就模型而言,值得提出非马尔可夫过程的模型,例如有潜变量的模型或更广义的、隐藏的马尔可夫模型(Cappé et al., 2005)。这里介绍的模型显然假设了行动者充分了解这个网络,并以一种貌似合理的方式对更大的网络建模,因此,有必要提出非完备信息假设的模型。第三,必须进一步开发统计程序。应该改进算法,研究它们的数学性质。除此之外,还应该开发估计拟合优度的程序,研究参数估计量的稳健性和对设定错误的检验。这些工作连同软件实现都意味着相当艰巨的方法论工作在等待着我们。

参考文献

- Barabási, A.L. and Albert, R. (1999) 'Emergence of scaling in random networks', *Science*, 286: 509-12.
- Bala, V. and Goyal, S. (2000) 'A noncooperative model of network formation', *Econometrica*, 68: 1181-229.
- Borgatti, S.P. and Foster, P.C. (2003) 'The network paradigm in organizational research: A review and typology', *Journal of Management*, 29: 991-1013.
- Brass, D.J., Galaskiewicz, J., Greve, H.R. and Tsai, W. (2004) 'Taking stock of networks and organizations: a multilevel perspective', *Academy of Management Journal*, 47: 795-817.
- Bauman, K.E., Fisher, L.A., Bryan, E.S. and Chenoweth, R. L. (1984) 'Antecedents, subjective expected utility, and behavior: A study of adolescent cigarette smoking', *Addictive Behaviors*, 9: 121-36.
- Bearman, P.S. (1997) 'Generalized exchange', *American Journal of Sociology*, 102: 1383-415.
- Bollobás, B. (1985) *Random Graphs*. London: Academic Press.
- Borgatti, S., Everett, M.G. and Freeman, L.C. (1998) *UCINET V, Reference Manual*. Columbia, SC: Analytic Technologies.
- Burk, W.J., Steglich, C.E.G. and Snijders, T.A. B. (2007) 'Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors', *International Journal of Behavioral Development*, 31: 397-404.
- Burt, R.S. (1992) *Structural Holes*. Cambridge, MA: Harvard University Press.
- Cappé, O., Moulines, E. and Rydén, T. (2005) *Inference in Hidden Markov Models*. New York: Springer.
- Checkley, M. and Steglich, C. E. G. (2007) 'Partners in power: Job mobility and dynamic deal-making', *European Management Review*, 4: 161-71.
- Christakis, N.A. and Fowler, J.H. (2007) 'The spread of obesity in a large social network over 32 years', *New England Journal of Medicine* 357: 370-379.
- Coleman, J.S. (1961) *The Adolescent Society*. New York: Free Press of Glencoe.
- Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. London: Chapman & Hall.
- Davis, J.A. (1970) 'Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices', *American Sociological Review*, 35: 843-52.
- De Klepper, M., Sleebos, E., van de Bunt, G. and Agneessens, F. (2010) 'Similarity in friendship networks: Selection or influence? The effect of constraining contexts and non-visible individual attributes', *Social Networks*, 32: 82-90.
- de Solla Price, D. (1965) 'Networks of scientific

- papers', *Science*, 149: 510-15.
- de Solla Price, D. (1976) 'A general theory of bibliometric and other advantage processes', *Journal of the American Society for Information Science*, 27: 292-306.
- Doreian, P. (1989) 'Network autocorrelation models: Problems and prospects', in D. A. Griffith (ed.), *Spatial Statistics: Past, Present, Future*. Ann Arbor: Michigan Document Services.
- Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. (2000) 'Structure of growing networks with preferential linking', *Physical Review Letters*, 85: 4633-36.
- Emirbayer, M. and Goodwin, J. (1994) 'Network analysis, culture, and the problem of agency', *American Journal of Sociology*, 99: 1411-54.
- Ennett, S. T. and Bauman, K. E. (1994) 'The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking', *Journal of Personality and Social Psychology*, 67: 653-63.
- Frank, O. (1991) 'Statistical analysis of change in networks', *Statistica Neerlandica*, 45: 283-93.
- Frank, O. and Strauss, D. (1986) 'Markov graphs', *Journal of the American Statistical Association* 81: 832-42.
- Geman, S. and Geman, D. (1983) 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721-41.
- Granovetter, M. S. (1973) 'The strength of weak ties', *American Journal of Sociology*, 78: 1360-80.
- Gulati, R. and Gargiulo, M. (1999) 'Where do interorganizational networks come from?', *American Journal of Sociology*, 104: 1439-93.
- Hagedoorn, J. (2002) 'Inter-firm R&D partnerships: an overview of major trends and patterns since 1960', *Research Policy*, 31: 477-92.
- Hallinan, M. T. (1974) *The Structure of Positive Sentiment*. New York: Elsevier.
- Hallinan, M. T. (1979) 'The process of friendship formation', *Social Networks*, 1: 193-210.
- Harris, K. M., Florey, F., Tabor, J., Bearman, P. S., Jones, J. and Udry, J. R. (2003) 'The national longitudinal study of adolescent health: Research design'. Technical report, University of North Carolina. <http://www.cpc.unc.edu/projects/addhealth/design/>.
- Haynie, D. L. (2001) 'Delinquent peers revisited: Does network structure matter?' *American Journal of Sociology*, 106: 1013-57.
- Hedström, P. (2005) *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
- Holland, P. W. and Leinhardt, S. (1975) 'Local structure in social networks', *Sociological Methodology—1976*, pp. 1-45.
- Holland, P. W. and Leinhardt, S. (1977) 'A dynamic model for social networks', *Journal of Mathematical Sociology*, 5: 5-20.
- Kapferer, B. (1972) *Strategy and Transaction in an African Factory*. Manchester: Manchester University Press.
- Katz, L. and Proctor, C. H. (1959) 'The configuration of interpersonal relations in a group as a time-dependent stochastic process', *Psychometrika*, 24: 317-27.
- Koskinen, J. H. and Snijders, T. A. B. (2007) 'Bayesian inference for dynamic network data', *Journal of Statistical Planning and Inference*, 13: 3930-38.
- Leenders, R. T. A. J. (1995) 'Models for network dynamics: A Markovian framework', *Journal of Mathematical Sociology*, 20: 1-21.
- Leenders, R. T. A. J. (1997) 'Longitudinal behavior of network structure and actor attributes: Modeling interdependence of contagion and selection', in P. Doreian and F. N. Stokman (eds), *Evolution of Social Networks*. New York: Gordon and Breach.
- Lin, N., Cook, K. and Burt, R. S. (eds) (2001) *Social Capital: Theory and Research*. New York: Aldine de Gruyter.
- Lindenberg, S. (2001) 'Social rationality versus rational egoism', in J. Turner (ed.), *Handbook of Sociological Theory*. New York: Kluwer/Plenum. pp. 635-68.
- Macy, M. W. and Willer, R. (2002) 'From factors to actors: Computational sociology and agent-based modelling', *Annual Review of Sociology* 28: 143-66.

- McPherson, M., Lynn, S.-L. and James M. C. (2001) 'Birds of a feather: Homophily in social networks', *Annual Review of Sociology*, 27: 415-44.
- Mercken, L., Snijders, T.A.B., Steglich, C. and de Vries, H. (2009) 'Dynamics of adolescent friendship networks and smoking behavior: Social network analyses in six European countries', *Social Science and Medicine*, 69: 1506-14.
- Merton, R. (1963) 'The Matthew effect in science', *Science*, 159(3810): 56-63.
- Michell, L. and A. Amos, (1997) 'Girls, pecking order and smoking', *Social Science and Medicine*, 44: 1861-69.
- Newcomb, T.M. (1961) *The Acquaintance Process*. New York: Holt, Rinehart and Winston.
- Nordlie, P.G. (1958) 'A longitudinal study of interpersonal attraction in a natural group setting', PhD thesis, University of Michigan.
- O' Campo, P. (2003) 'Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health', *American Journal of Epidemiology*, 157: 9-13.
- Pearson, M. and West, P. (2003) 'Drifting smoke rings: Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking', *Connections*, 25(2): 59-76.
- Powell, Walter W., White, Douglas R., Koput, Kenneth W. and Owen-Smith, Jason. (2005) 'Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences', *American Journal of Sociology*, 110: 1132-205.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rapoport, A. (1953a) 'Spread of information through a population with socio-structural bias: I. Assumption of transitivity', *Bulletin of Mathematical Biophysics*, 15: 523-33.
- Rapoport, A. (1953b) 'Spread of information through a population with socio-structural bias: II. Various models with partial transitivity', *Bulletin of Mathematical Biophysics*, 15: 535-46.
- Ripley, R. and Snijders, T.A.B. (2010) *Manual for SIENA version 4.0*. Oxford: University of Oxford, Department of Statistics, <http://www.stats.ox.ac.uk/siena/>.
- Robins, G. and Pattison, P. (2001) 'Random graph models for temporal processes in social networks', *Journal of Mathematical Sociology*, 25: 5-41.
- Robins, G.L., Woolcock, J., and Pattison, P. (2005) 'Small and other worlds: Global network structures from local processes', *American Journal of Sociology* 110: 894-936.
- Sampson, R. J., Morenoff, J. D. and Gannon-Rowley, T. (2002) 'Assessing "neighborhood effects": Social processes and new directions in research', *Annual Review of Sociology* 28: 443-78.
- Sampson, S.F. (1969) 'Crisis in a cloister', Ph.D. dissertation, Cornell University.
- Schaefer, D. R., Light, J. M., Fabes, R. A., Hanish, L. D. and Martin, C. L. (2010) 'Fundamental principles of network formation among preschool children', *Social Networks* 32: 61-71.
- Schweinberger, M. (2007) 'Statistical methods for studying the evolution of networks and behavior', Ph. D. dissertation, University of Groningen.
- Schweinberger, M. (2008) 'Statistical modeling of network dynamics given panel data: Goodness-of-fit tests'. Submitted for publication.
- Selfhout, M., Burk, W., Branje, S., Denissen, J.J.A., Van Aken, M.A.G. and Meeus, W. (2010) 'Emerging late adolescent friendship networks and big five personality traits: A dynamic social network perspective', *Journal of Personality*, 78: 509-38.
- Snijders, T. A. B. (1996) 'Stochastic actor-oriented dynamic network analysis', *Journal of Mathematical Sociology*, 21: 149-72.
- Snijders, T. A. B. (2001) 'The statistical evaluation of social network dynamics', *Sociological Methodology*—2001, 361-95.
- Snijders, T.A.B. (2002) 'Markov chain Monte

- Carlo estimation of exponential random graph models', *Journal of Social Structure*, 3: 2.
- Snijders, T.A.B. (2005) 'Models for longitudinal network data', in P.J. Carrington, J. Scott, and S. Wasserman (eds), *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.
- Snijders, T. A. B., Koskinen, J. H. and Schweinberger, M. (2010) 'Maximum likelihood estimation for social network dynamics', *Annals of Applied Statistics*, 4: 567-588.
- Snijders, T. A. B., Pattison, P., Robins, G. L., Handcock, M. (2006) 'New specifications for exponential random graph models', *Sociological Methodology*—2006, 99-153.
- Snijders, T. A. B., Steglich, C. E. G. and Schweinberger, M. (2007) 'Modeling the co-evolution of networks and behavior', in Kees van Montfort, Han Oud and Albert Satorra (eds), *Longitudinal Models in the Behavioral and Related Sciences*. Mahwah, NJ: Lawrence Erlbaum. pp. 41-71.
- Snijders, T.A.B., van de Bunt, G.G. and Steglich, C.E.G. (2010) 'Introduction to stochastic actor-based models for network dynamics', *Social Networks*, 32: 44-60.
- Snijders, T. A. B., and van Duijn, M. A. J., 'Simulation for statistical inference in dynamic network models'. In: R. Conte, R. Hegselmann, and P. Terna, (eds) *Simulating Social Phenomena*. Lecture Notes in Economics and Mathematical Systems, 456. Berlin: Springer, 1997, pp. 493-512.
- Sørensen, A. B. and Hallinan, M.T. (1976) 'A stochastic model for change in group structure', *Social Science Research*, 5: 43-61.
- Steglich, C.E.G., Snijders, T.A.B. and Pearson, M. (2010) 'Dynamic networks and behavior: Separating selection from influence', *Sociological Methodology*—2010, pp. 329-393.
- Udehn, L. (2002) 'The changing face of methodological individualism', *Annual Review of Sociology*, 8: 479-507.
- Udry, J. R. (2003) 'The national longitudinal study of adolescent health (add health), Waves I & II, 1994-1996; Wave III, 2001-2002', technical report, Carolina Population Center, University of North Carolina at Chapel Hill.
- van Duijn, M. A. J., Zeggelink, E. P. H., Huisman, M., Stokman, F.N. and Wasseur, F. W. (2003) 'Evolution of sociology freshmen into a friendship network', *Journal of Mathematical Sociology*, 27: 153-91.
- Wasserman, S. (1977) 'Stochastic models for directed graphs', Ph.D. dissertation, University of Harvard, Dept. of Statistics.
- Wasserman, S. (1979) 'A stochastic model for directed graphs with transition rates determined by reciprocity', *Sociological Methodology*—1980, pp. 392-412.
- Wasserman, S. (1980) 'Analyzing social networks as stochastic processes', *Journal of the American Statistical Association*, 75: 280-94.
- Wasserman, S. (1987) 'The conformity of two sociometric relations', *Psychometrika*, 53: 261-82.
- Wasserman, S. and K. Faust. (1994) *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.
- Wasserman, S. and D. Iacobucci. (1988) 'Sequential social network data', *Psychometrika*, 53: 261-82.
- Wasserman, S. and Pattison, P.E. (1996) 'Logit models and logistic regression for social networks; I. An introduction to Markov graphs and p*', *Psychometrika*, 61: 401-25.
- West, P. and Sweeting, H. (1995) 'Background rationale and design of the West of Scotland 11-16 Study', Working Paper No. 52. Glasgow: MRC Medical Sociology Unit Glasgow.
- Zeggelink, E. P. H. (1994) 'Dynamics of structure: An individual oriented approach', *Social Networks*, 16: 295-333.