

基于预测误差修正的时序链路预测方法

邓志宏* 老松杨 白亮

(国防科学技术大学信息系统与管理学院 长沙 410073)

摘要: 论文主要针对时序链路预测方法进行研究。分析了静态链路预测方法的弊端,认为忽视网络演化趋势信息会对链路预测产生负面影响;还提出了链路预测误差的概念用于描述网络趋势信息,并以此为基础提出一种基于预测误差修正的时序链路预测方法。该方法首先对待预测时刻之前一个时间窗口内的多幅网络图分别采用静态链路预测方法进行预测,记录每次的预测误差并计算其修正值,最后对待测时刻静态预测结果进行修正得到最终预测结果。通过在两个真实网络数据集上进行系列实验表明,该方法较大提升了静态链路预测方法的预测精确度,与另一种典型的时序链路预测方法相比其精度也有所提升,且算法时间复杂度较低。另外,实验中还发现链路预测误差序列与网络链路总数序列存在“镜面对称”关系,分析其内在原因证明了所提方法的普适性。

关键词: 复杂网络; 链路预测; 预测误差修正

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2014)02-0325-07

DOI: 10.3724/SP.J.1146.2013.00657

A Temporal Link Prediction Method Based on Link Prediction Error Correction

Deng Zhi-hong Lao Song-yang Bai Liang

(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: The temporal link prediction method is investigated in this paper. The disadvantages of the static link prediction methods are analyzed, considering that ignoring the evolving information of networks will lead to a negative impact on link predicting. The concept of link prediction error is proposed to describe the evolving information of networks, and a temporal link prediction method is proposed based on the prediction error correction. Firstly, several static link prediction are carried out using each graph in the previous periods window, and then the prediction errors are recorded and used for calculating the modification value. At last, the final prediction result is acquired through refining the static prediction result with the modification value. Several experiments are conducted using two real network datasets. The results show that the proposed method achieves better performance than the static link prediction methods and a typical temporal link prediction method. In addition, it can be found that a relation of ‘mirror symmetry’ exists between prediction error series and total link number series, which demonstrates the universality of the proposed method.

Key words: Complex network; Link prediction; Prediction error correction

1 引言

网络中的链路预测(link prediction)问题是关于复杂网络的演化特性研究中一个重要内容,是指如何通过已知的网络结构等信息预测网络中尚未产生连边的两个节点之间产生连接的可能性。网络中的链路预测既包含了对未知链接(existent yet unknown links)的预测,也包含了对未来链接(future links)的预测。预测已经存在但尚未被发现连接实际上是一种数据挖掘的过程,而对于未来可能产生

的连边的预测则是对网络演化规律的把握。

链路预测具有重要的应用价值,如解决推荐系统中数据稀少的问题^[1];用于电子商务中向用户推荐可能感兴趣的商品,在线社交网络好友推荐^[2];在揭示蛋白质作用网络中隐而未知的链接的实验中,事先通过链路预测缩小实验检测范围,可以大大减少实验所耗费的成本^[3,4];在科学家合作网中利用链路预测方法识别潜在合作可能^[5];此外链路预测方法在识别犯罪网络结构,检测和控制网络攻击等应用中也能发挥重要作用。

链路预测方法研究近年来受到了广泛关注。文献[6,7]是最近的关于链路预测的综述文献,其中文

2013-05-10 收到, 2013-10-14 改回

国家自然科学基金(60902094)资助课题

*通信作者: 邓志宏 lingyu207@gmail.com

献[7]将链路预测方法归纳为 3 大类: (1) 基于最大似然估计的方法^[8,9]; (2) 概率模型方法^[10,11]; (3) 基于结构相似性的方法^[12-14], 这种分类方法体现了对链路预测问题建模思想的不同, 其中基于结构相似性的方法由于方法简单、计算复杂度相对较低、使用范围广等特点最近受到广泛关注。文献[13]提出了基于网络拓扑结构的相似性定义方法, 并分析了若干指标在社会合作网络中链路预测的效果。文献[14]在多种不同网络中比较了 9 种已知的基于局部信息的相似性指标在链路预测中的效果, 并提出了两种新指标: 资源分配指标和局部路径指标。

已有的链路预测方法大多是以待预测时刻点之前的静态网络图作为预测依据, 没有考虑网络的历史演化信息, 这在许多具体应用中是不合理的, 因为对复杂网络所描述的现实系统而言, 其未来时刻的结构状态不是单纯由前一时刻结构状态决定, 而是在时间轴上统一于系统的整体变化趋势, 因此链路预测方法中选取多幅连续网络图作为预测依据更为合理。Huang 等人^[15]基于这种思想, 依据是否考虑网络时序演化信息将链路预测方法分为静态链路预测方法(SLPM)和时序链路预测方法(TLPM)两大类, 关于时序链路预测方法的研究近年来逐渐被重视, 文献[15]将多幅连续网络图转化为链路的出现序列数据, 然后采用时间序列分析方法预测下一时刻指定链路是否会出现; 文献[16]首先用三阶张量来描述时序网络图数据, 然后采用矩阵和张量分解的方法进行预测; 文献[17]提出一种频繁子图概念, 将时序网络图转换为频繁子图集合序列, 然后对频繁子图的出现概率进行估计, 从而实现了对链路的预测。这些方法的一个共同点在于放弃了时序网络图本身蕴含的丰富的结构演化信息, 通过一些其它数学模型来描述时序网络图, 进而将链路预测问题转化为对应数学模型求解问题, 这种方式虽然有可能获得较高精度解, 但都存在算法复杂度较高的问题, 针对规模比较大的网络难以满足算法效率要求。

本文以基于结构相似性的静态链路预测方法为基础, 针对时序链路预测问题, 提出一种基于预测误差修正的时序链路预测方法, 主要思想是改变以往只依据某一时刻网络拓扑结构信息进行预测的方式, 通过观察待预测点前一个时间段内的预测结果与真实网络的结构差异, 记录这些差异作为下一时刻预测结果的修正值, 从而使得最终预测结果向真实网络拓扑结构逼近。

2 问题描述与符号注记

定义 $G = \{g_1, g_2, \dots, g_n\}$ 表示一个由 n 个时间片

组成的演化网络, g_t 表示在时刻 t 时演化网络的快照, $V(g_t)$ 表示 g_t 的节点集, $E(g_t)$ 表示 g_t 的边集。给定一种链路预测方法, 就是要对任意节点对 (v_x, v_y) 在下一时刻产生新的连边的可能性给出一种度量方法, 记这种产生新的连边的可能性为 s_{xy} , 即一种 s_{xy} 的表示方法就对应一种链路预测方法, 时序链路预测方法可以用一个映射表示:

$$\text{TLPM}: G' \rightarrow \mathbf{S} \quad (1)$$

其中 $G' \subset G$ 是已观测到的历史网络拓扑信息, 包含多幅连续网络历史拓扑图, 是链路预测的数据依据, \mathbf{S} 表示连边可能性矩阵, 反映了对未来时刻网络的拓扑结构的猜测, 是对任意节点对 (v_x, v_y) 在下一时刻产生新的连边的可能性的一种度量。本文采用基于结构相似性 $s_{xy} \in \mathbf{S}$ 的方法对这种可能性进行度量, 这类方法的一个重要前提假设就是两个节点之间的相似性越大, 它们之间产生连边的可能性就越大, 这里所说的节点相似性完全基于网络的拓扑结构信息, 而不是基于节点的属性信息。特别地, 当 G' 只包含 1 幅网络图时, 对应的就是 SLPM, 基于结构相似性的 SLPM 是本文方法的基础。Liben-Nowell 等人^[12]最早提出了基于网络拓扑结构的相似性定义方法, 并分析了若干指标在社会合作网络中链路预测的效果, 本文选择其中 4 种作为参考, 分别是共同邻居(CN)指标, Adamic-Adar(AA)指标, 优先链接(PA)指标和 Katz 指标。

3 方法

SLPM 仅仅依据前一个时刻网络拓扑快照进行预测, 这就类似于对移动目标进行位置预测时, 单纯依靠前一时刻目标快照信息对目标下一时刻位置进行预测。这类方法忽视了预测对象在过去一段时间内的变化趋势, 必然会很大程度上受网络演化在单一时刻表现出的随机性以及噪声的影响, 因而采用 SLPM 方法预测得到的结果必然会和网络真实结构存在差异。这种结构差异的现实含义可以很丰富, 在不同类型的网络中, 产生这些结果差异的深层原因也是多种多样的。

如图 1 所示在线社会网络中采用 CN 算法进行朋友推荐, 如果仅观察 t_i 时刻网络拓扑信息, 可以发现 A, B 两节点之间存在较多共同邻居节点。依据

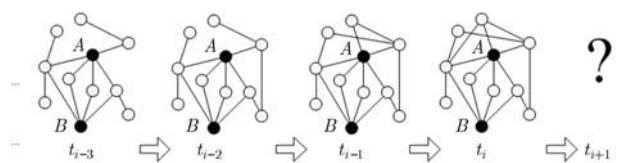


图 1 示例演化网络

CN 算法, 这两个节点之间将会产生连边(成为朋友)的可能性很大, 但如果观察前一段时间的网络拓扑信息, 将会发现节点 B 在较长一段时间内连边未发生任何变化, 这可能是由于节点 B 已退出该网络社区。这种情况下, 采用静态 CN 算法得出的结果必然和实际情况存在较大误差, 究其原因就是 SLPM 方法忽略了类似“节点 B 可能已退出该网络社区”这种蕴含在网络演化过程中的趋势信息, 而这种信息往往会对链路预测结果产生较大影响。对于不同类型的网络, 这种趋势信息的表现不尽相同, 难以用统一的指标去度量, 但不论这种趋势信息是什么含义, 对于链路预测而言, 都会带来相同的结果, 即预测误差。这种差异存在于每次独立预测过程中, 通过观察每次预测误差以及相邻间隔预测误差变化可以在一定程度上反应出网络的演化趋势信息, 充分利用这种演化趋势信息能够对链路预测产生积极影响。本文基于上述思想提出基于预测误差修正的时序链路预测方法, 方法框架如图 2 所示, 其主要思想是通过观察待预测时刻之前一个时间窗口内的多幅网络拓扑图, 采用 SLPM 方法分别进行多次静态预测, 测量每次的链路预测误差, 以这些误差值作为下一次预测的修正依据, 通过这种方式尽量减小下一次预测的预测误差。例如如图 1 所示示例网络, 在过去 3 个时刻内, 采用 CN 方法得到的节点 A, B 间链路的生成概率均较高, 但在实际网络演化中, A, B 间一直未出现连边, 因而在下一次预测时, 应当适当减小该节点对间的链路生成概率值。

链路预测误差是本文方法的一个核心概念。

定义 链路预测误差(LPE)是指应用 SLPM 方法得到的预想网络拓扑结构和真实网络拓扑结构之间的差异, 可以用式(2)进行度量。

$$LPE = S - M \quad (2)$$

其中 M 为下一时刻真实网络的邻接矩阵, 由于本文研究的是无向无权网, M 中的元素非 1 即 0; S 为 SLPM 方法得到的链路生成概率矩阵, S 中的元素需要经过如式(3)的归一化处理:

$$s_{ij} = \frac{s_{ij}}{\max S} \quad (3)$$

如何依据记录的预测误差序列计算预测误差修正值是本文方法的关键。最简单的方法可以对所有预测误差序列值求平均值, 但这种方式没有考虑误差序列的时间特性以及误差值的变化趋势, 本文采用式(4)计算链路预测误差修正值。

$$LPEC_T = \frac{\sum_{T-T_\Delta-1 \leq t < T} \alpha^{T-t} LPE_t}{T_\Delta - 1} + \frac{\sum_{T-T_\Delta-1 < t < T} \alpha^{T-t} (LPE_t - LPE_{t-1})}{T_\Delta - 2} \quad (4)$$

其中 LPE_t 为时刻 t 的预测误差, T 为待预测时刻, T_Δ 为观测时间窗口大小, α 为误差影响衰减因子, 反映了预测误差对最终预测结果的影响是随着与当前时刻的距离增大而不断衰减的。式中两部分分别对应链路预测误差修正值 $LPEC_T$ 的两个组成部分: 反应时间窗口内误差整体水平的误差加权平均值和反应时间窗口内误差变化趋势的误差变化值的加权平均值。

基于以上分析, 采用式(5)计算最终链路生成概率矩阵。

$$S_TLP_T = S_SLP_T - LPEC_T \quad (5)$$

其中 S_TLP_T 为时刻 T 链路生成概率矩阵最终值, S_SLP_T 为采用 SLPM 计算得到的时刻 T 链路生成概率矩阵的静态预测值。

4 实验

4.1 数据集

本文在以下两个不同类型的真实网络数据集上测试所提方法的有效性。

(1) 安然公司邮件数据集^[18](Enron email dataset);

(2) 高能量子物理学家合作数据集^[19] (high-energy particle physics coauthorship dataset)。

本文使用的安然邮件数据集(Enron)是由 USC 大学 Jitesh Shetty 等人整理的一个完整性、一致性均较好的数据集版本, 数据集总共包含了 252,759 封邮件。本文实验中抽取了 1999 年 5 月至 2002 年 6 月中 151 个员工之间的总计 21,254 封邮件构建时序网络图, 151 个员工的邮件地址作为网络的节点, 每一条邮件表示发件人和收件人代表的两个节点间存在连边, 邮件的发送时间作为演化网络的时间戳。根据实验需要, 本文依据划分时序网络图的时间段长度的不同分别生成两组时序网络图, 一组是将邮

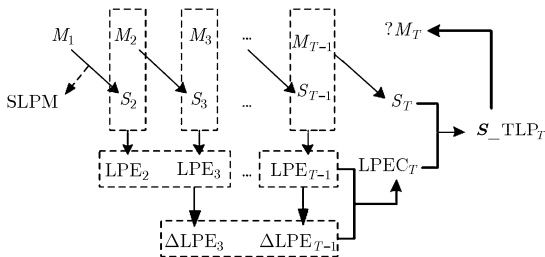


图 2 本文方法框架

件数据按月划分得到的总共 38 个月的演化网络, 记为 GE_M , 另一组是将邮件数据按照周划分得到的总共 158 个周的演化网络 GE_W 。

高量子物理科学家合作数据集(Hep-th)包含了 1992 年至 2003 年期间总共 9200 位作者, 29555 篇论文和 87794 个合作关系的数据, 本文选取其中最高产并且与其他作者至少合作过一次的 96 位作者之间的总共 1796 条合作记录用于实验, 每条论

文合作记录表示两个作者代表的节点间存在连边, 以论文最终提交时间作为链路的时间戳, 以季度为时间步长构建时序网络图, 记为 G_H 。

4.2 实验设计

为了全面验证本文方法的可行性与有效性, 本节设计了 4 个实验, 如表 1 所示。实验结果与 4 种 SLPM 方法以及文献[15]提出的一种时序链路预测方法(TSCP)进行对比。

表 1 实验设计

序号	实验目的	主要参数设置		
		Enron(Monthly)	Enron(Weekly)	Hep-th
1	算法总体性能验证	$T_\Delta = [5, 20], \alpha = 0.8$	$T_\Delta = [5, 20], \alpha = 0.8$	$T_\Delta = [5, 20], \alpha = 0.8$
2	时间窗口 T 参数分析	$T_\Delta = [3, 20], \alpha = 0.8$	$T_\Delta = [3, 40], \alpha = 0.8$	$T_\Delta = [3, 30], \alpha = 0.8$
3	误差影响衰减因子参数分析	$T_\Delta = 8, \alpha = [0.1, 1.0]$	$T_\Delta = 13, \alpha = [0.1, 1.0]$	$T_\Delta = 13, \alpha = [0.1, 1.0]$
4	预测误差分析	$T_\Delta = 8, \alpha = 0.8$	$T_\Delta = 8, \alpha = 0.8$	-

本文提出的时序链路预测方法是以基于结构相似性的 SLPM 方法为基础, 对应 4 种 SLPM 方法分别命名为 TLP-CN, TLP-AA, TLP-PA 和 TLP-KZ。实验中采用 AUC(Area Under ROC Curve)值作为链路预测方法的评价指标, 在计算时序链路预测方法 AUC 值时采用图 3 所示方法选取训练集和测试集, 针对每个时间窗口 T_Δ 的取值, 首先选取时序网络图中前 T_Δ 个时刻网络邻接矩阵作为训练集, 以 $T_\Delta + 1$ 时刻网络邻接矩阵为测试集计算 1 次 AUC 值, 然后将时间窗口往后移动一个步长, 进行下一次 AUC 值计算, 最后对所有 AUC 值取平均作为对应时间窗口 T_Δ 取值时链路预测方法的 AUC 值。

4.3 实验结果与分析

实验 1 算法总体性能验证实验

图 4 所示为针对 3 个数据集, 选取 $T_\Delta = [5, 20]$ 和 $\alpha = 0.8$ 时本文方法和各种 SLPM 方法预测结果对比。从图 4 中可以看出, 4 种经典链路预测方法中, CN, AA 两种方法效果相近, PA 算法最差,

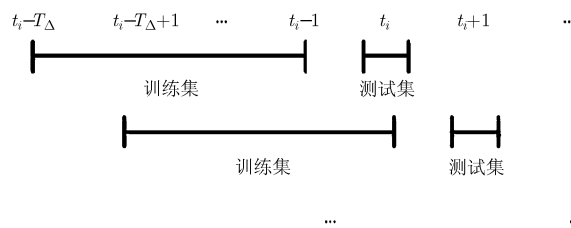


图 3 AUC 值计算训练集与测试集的选取方法

KATZ 算法表现最好, 而本文方法对 4 种经典链路预测方法均有较大提升, 且不论对应哪种经典算法, 改进后的算法性能均较为稳定; 另外, 对比图 4(a) 和图 4(b) 可以发现, 演化网络 G_W 中各方法表现均优于演化网络 G_M , 这是因为演化网络 G_W 中时序网络图的时间间隔小于 G_M , 数据更加精细。因此在实际应用中, 如果条件允许, 应尽可能获得更精细的演化网络数据, 这对提升链路预测方法性能具有很大帮助。

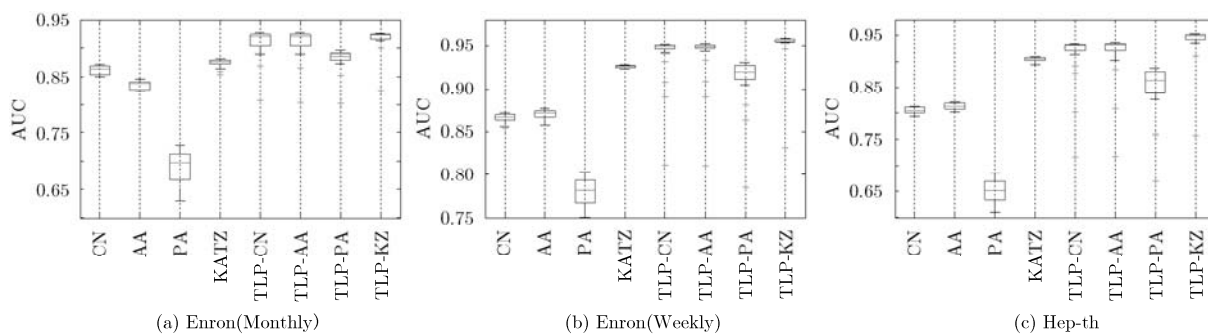


图 4 本文方法与 SLPM 方法对比

图 5 给出了针对 3 个数据集，选取 $T_{\Delta} = 13$ 和 $\alpha = 0.9$ 时本文方法与文献[15]的时序链路预测方法对比。从图 5 中可以看出，本文方法预测精度要普遍优于 TSCP 方法，除了在 Hep-th 数据集中 TLP-PA 方法要略差，这是由于 Hep-th 数据集过于稀疏，而 PA 算法本身预测精度又较低，两个因素综合导致静态 PA 算法的 AUC 值只有 0.6632，最终影响到 TLP-PA 的预测精度。另外，本文算法时间复杂度为 $nO(\text{SLPM})$ ，也优于 TSCP 算法的时间复杂度 $n^2O(\text{SLPM})$ 。

实验 2 时间窗口参数影响实验

时间窗口是预测方法中一个重要参数，为了验证其对预测方法的影响，固定误差影响衰减因子设为 $\alpha = 0.8$ ，时间窗口 T_{Δ} 取不同值对 3 个数据集进行实验，实验结果如图 6 所示。从图 6(a)~图 6(c)中均可以看出，随着时间窗口的增大，时序链路预测方法的性能先是逐渐增大，之后会有一个平稳阶段，当时间窗口继续增大，部分时序链路预测方法的性能会有较小程度的下降。因此可以得出结论，时序链路预测方法中，时间滑窗取值大小对预测结果具有重要影响，取值过小会导致时间滑窗内的网络拓扑序列无法满足反映网络演化规律的要求，取值过大会导致时间滑窗内存在一些过于久远的网络拓扑数据，这些数据由于距离待预测时刻较远，对链路预测会产生负面影响。但由于本文方法在计算预测误差修正值时不仅考虑了时间窗口内误差加权平均值，还考虑了时间窗口误差变化值的加权平均值，时间窗口过大对预测结果的负面影响得到了有效控制。为了验证这一点，通过改变误差修正值的计算公式，去除其中误差变化值的加权平均值项，在不改变参数设置的情况下对 Enron(Monthly)数据集进行实验，实验结果如图 6(d)所示。对比图 6(a)可以看出，改变误差修正值的计算公式后，预测结果在时间窗口较大时性能下降更明显，由此可以证明

本文方法中计算预测误差修正值的方法是合理的，能够有效控制时间窗口过大时对预测方法产生的负面影响。

实验 3 误差影响衰减因子参数影响实验

图 7 所示为针对 3 个数据集分别固定时间窗口 $T_{\Delta EM} = 8$ ， $T_{\Delta EW} = 13$ 和 $T_{\Delta H} = 13$ ，误差影响衰减因子 α 取不同值时，本文方法预测结果变化情况。从图 7 中可以看出，误差影响衰减因子 α 取值对预测结果具有一定影响， α 取值小于 0.5 时，预测结果较差，这是因为 α 取值小时相当于缩小了计算预测误差的时间窗口的有效范围， $\alpha = 0$ 时即退化为 SLPM 方法。取值跨越范围 $[0.5, 0.6]$ 时，预测结果会有较大提升， α 取值接近 1 时，算法精度反而下降，这是因为 α 取值接近 1 会导致较早时刻的网络历史演化信息和近期的网络演化信息同等重要，这是不合理的。另外从图 7 中可以看出，预测算法精度越低的方法对 α 取值变化越敏感。综合来看， α 取值在范围 $[0.8, 0.9]$ 内，预测方法效果最好。

实验 4 链路预测误差分析实验

图 8 所示下方曲线为选取时间窗口大小 $T_{\Delta} = 8$ ，误差影响衰减因子 $\alpha = 0.8$ 时，KATZ 方法的预测误差序列，上方曲线为各时刻链路总数的归一化序列值。可以发现，链路预测误差曲线与链路总数曲线呈现出“镜面对称”的状态，预测误差的波动行为和链路总数变化的波动相吻合，链路总数发生激增时，一般都对应了预测误差的一个峰值，而这些峰值往往对应了现实世界某些突发事件和重大事件，从图 8(a)中可观察到一个明显峰值，对应的是 2001 年 10 月期间 Enron 由于巨额资金流失被美国联邦能源规划委员会(SCE)调查，图 8(b)中可观察到 3 个明显峰值区域，分别对应了 2000 年 10 月和 2001 年 8 月两次更换 CEO 的事件，以及 2001 年 10 月的 SCE 调查事件。

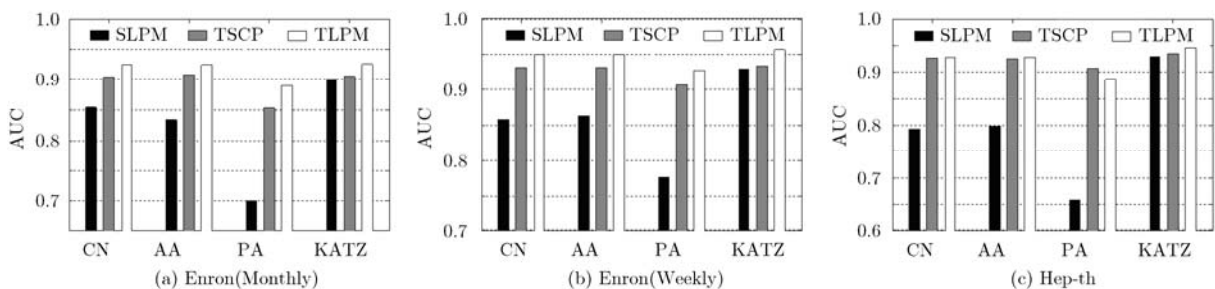


图 5 本文方法与 TSCP 方法对比

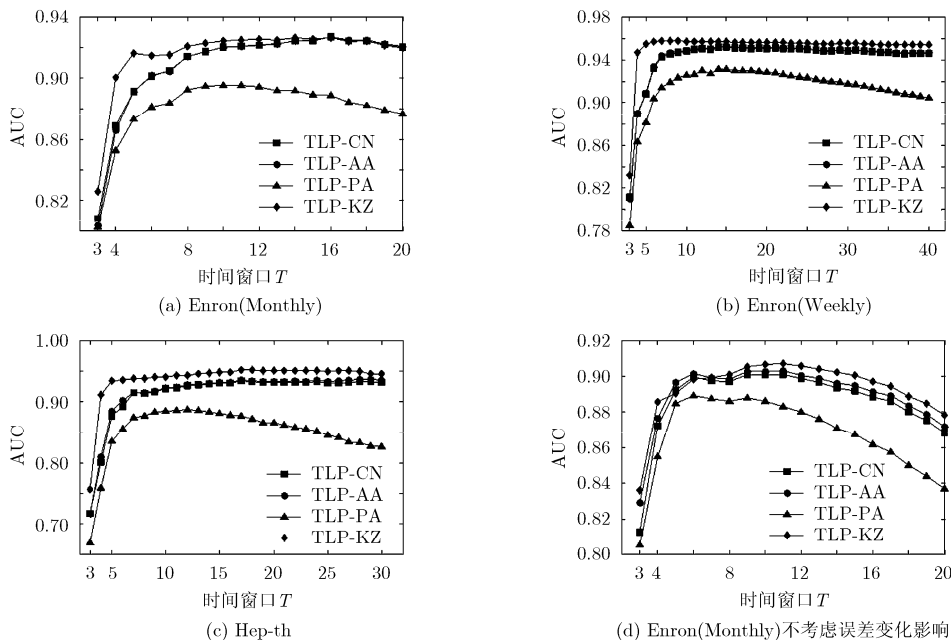


图 6 时间滑窗大小对预测算法的影响

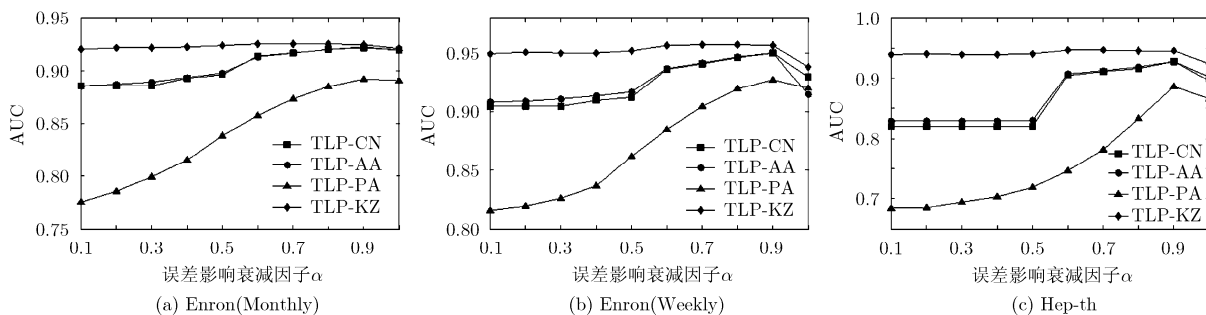


图 7 误差影响衰减因子对预测算法的影响

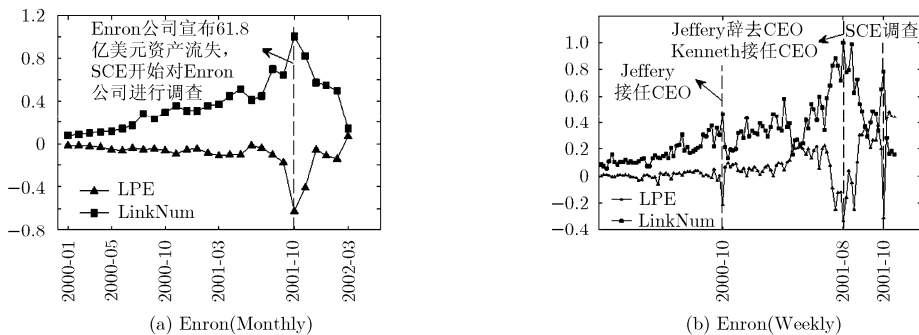


图 8 预测误差序列与各时刻链路总数序列

这是一个很有趣的现象，但不是一个巧合现象。前文分析过，采用链路预测方法预测得到的结果必然会和网络真实结构存在差异，这种结构差异的现实含义可以很丰富，在不同类型的网络中，产生这些结果差异的深层原因也是多种多样的。之前关于在线社会网络朋友推荐的例子中，是由于节点本身某些属性影响造成链路预测产生较大偏差，而在对

Enron 邮件数据预测中，预测误差的波动是由于某些突发事件导致的。这两个例子中，前者表明预测误差与节点度指标存在一定关系，后者表明预测误差与网络链路总数存在一定关系，这种关联关系正好证明了预测误差确实能够反映出网络演化趋势信息，对链路预测起到积极影响。并且与节点度变化、链路总数变化这类指标不同，后者用来描述网络演

化趋势信息会受具体网络类型的制约, 而预测误差由于不是对网络本身属性描述, 因而具有普适性。由此可以得出结论本文方法能够适用于不同类型现实网络的链路预测。

5 结束语

链路预测最初作为数据挖掘领域的研究方向之一, 如今已在多个领域具有重要应用, 链路预测方法近年来也受到广泛关注。本文针对静态链路预测方法忽视网络演化趋势信息的弊端, 提出了一种基于预测误差修正的时序链路预测方法, 采用预测误差来描述网络演化趋势信息, 进而辅助链路预测。以两个真实数据集——安然邮件数据集和高能量子物理学家合作数据集进行了一系列实验, 实验结果表明本文方法能够较大程度地提升 SLPM 方法预测精度, 与已有的一种典型的时序链路预测方法相比, 在预测精度和算法时间复杂度上都具有一定优势。另外还通过实验分析了算法中两个关键参数对预测结果的影响, 最后通过分析预测误差与网络链路总数关系, 证明了预测误差是一种可以用来描述网络演化趋势信息的普适指标, 进而证明了本文方法能够适用于不同类型现实网络的链路预测。

下一步研究重点包括两个方面: (1) 本文方法对于有向、加权网络的适用性需要进一步研究; (2) 对于存在多种类型链路的网络, 如何预测节点之间会产生何种连边是值得研究的问题。

参考文献

- [1] Chiluka N, Andrade N, and Pouwelse J. A link prediction approach to recommendations in large-scale user-generated content systems[C]. Proceedings of the 33rd European conference on Advances in Information Retrieval, Ireland, 2011: 189-200.
 - [2] Aiello L M, Barrat A, Schifanella R, et al. Friendship prediction and homophily in social media[J]. *Transactions on the Web (TWEB)*, 2012, 6(2): 1-37.
 - [3] Yu H, Braun P, Yildirim M A, et al. High-quality binary protein interaction map of the yeast interaction network[J]. *Science*, 2008, 322(5898): 104-110.
 - [4] Stumpf M P H, Thorne T, Silva E de, et al. Estimating the size of the human interaction[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(19): 6959-6964.
 - [5] Gu Q, Zhou J, and Ding C. Collaborative filtering: weighted nonnegative matrix factorization incorporating user and item graphs[C]. Proceedings of the SIAM International Conference on Data Mining, Columbus, 2010: 199-210.
 - [6] Mohammad Al Hasan and Mohammed J Zaki. A Survey of Link Prediction in Social Networks[M]. New York, Social network Data Analysis, Springer, 2011: 243-275.
 - [7] Lü L and Zhou T. Link prediction in complex networks: a survey[J]. *Physical A: Statistical Mechanics and Its Applications*, 2010, 390(6): 1150-1170.
 - [8] Clauset A, Moore C, and Newman M E J. Hierarchical structure and the prediction of missing links in networks[J]. *Nature*, 2008, 453: 98-101.
 - [9] Guimera R and Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(52): 22073-22078.
 - [10] Jamali M, Huang T, and Ester M. A generalized stochastic block model for recommendation in social rating networks[C]. Proceedings of the fifth ACM Conference on Recommender Systems, New York, 2011: 53-60.
 - [11] Huang S, Chen M, and Luo B, et al. Predicting aggregate social activities using continuous-time stochastic process[C]. Proceedings of the 21st ACM International Conference on Information and Knowledge Management, New York, 2012: 982-991.
 - [12] Liben-Nowell D and Kleinberg J. The link prediction problem for social networks[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(7): 1019-1031.
 - [13] Papadimitriou A, Symeonidis P, and Manolopoulos Y. Scalable link prediction in social networks based on local graph characteristics[C]. ITNG 2012: 9th Int'l Conference on Information Technology- New Generations, Las Vegas, 2012: 738-743.
 - [14] Zhou T, Lü L, and Zhang Y C. Predicting missing links via local information[J]. *The European Physical Journal B*, 2009, 71: 623-630.
 - [15] Huang Z and Lin D K J. The time series link prediction problem with applications in communication surveillance[J]. *Infoms Journal on Computing*, 2009, 21(2): 286-303.
 - [16] Dunlavy D M, Kolda T G, and Acar E. Temporal link prediction using matrix and tensor factorizations[J]. *ACM Transactions on Knowledge Discovery from Data*, 2011, 5(2): 1-27.
 - [17] Lahiri M and Berger-Wolf T Y. Structure prediction in temporal networks using frequent subgraphs[C]. Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining, Hawaii, 2007: 35-42.
 - [18] Adibi J I. Enron email dataset[OL]. http://www.isi.edu/_adibi/Enron/Enron.htm, 2013.
 - [19] The Knowledge Discovery Laboratory (KDL), Hep-th dataset[OL]. <https://kdl.cs.umass.edu/display/public/HEP-Th>, 2013.
- 邓志宏: 男, 1986年生, 博士生, 研究方向为指挥控制与指挥决策分析。
老松杨: 男, 1968年生, 教授, 研究方向为指挥控制与指挥决策分析、视频情报分析。
白亮: 男, 1978年生, 副教授, 研究方向为指挥控制与指挥决策分析、视频情报分析。