

现或消失的节点.对以上 3 个网络,本文均建模为无向加权网络,权重的计算使用按时间衰减的思想,即:离当前时刻越远的网络快照,对当前时刻节点权重影响越小.权重计算公式如下:

$$w_{a,b}(t) = \sum_i w_i e^{-\lambda(t-t_i)} \quad (4)$$

其中, w_i 为边 ab 在 t_i 时刻的权重,相应的邻接矩阵序列为 $\langle A_1, A_2, \dots, A_r \rangle$.

3 个数据集的详细信息见表 2.

Table 2 Datasets details

表 2 数据集详细信息

数据集	节点数	*边数	*特征数	角色数	快照数	快照长度	时间区间
Enron	2 114	16 413	70	4	12	1 month	2001.1~2001.12
Facebook	5 111	14 438	144	4	12	1 month	2008.1~2008.12
DBLP	29 747	96 874	159	4	10	1 year	2002~2011

*表示平均值

4.2 对比方法

本文使用 3 种方法作为对比.

- (1) PRE(Baseline):直接用当前时刻的节点角色取值作为下一时刻的预测值,即用 t 时刻的节点角色矩阵作为 $t+1$ 时刻的预测结果,即, $G'_{t+1} = G_t$;
- (2) TM:TM 方法是前面提到的转移矩阵方法^[11],该方法利用 $t-1$ 和 t 时刻的角色矩阵 G_{t-1} 和 G_t ,根据非负矩阵分解得到角色转移矩阵 $T:G_{s(t-1)}T \approx G_{s(t)}$,由 G_t 和 T 相乘得到 $t+1$ 时刻的目标角色矩阵.

$$G'_{t+1} : G'_{t+1} = G_t T;$$

- (3) STR-RP:以基于单目标回归算法得到的一阶模型为最终模型, G'_{t+1} 表示节点 n 在 $t+1$ 时刻的预测向量.

$$G'_{t+1} = [h_1(G_t^n), \dots, h_4(G_t^n)].$$

4.3 评估方法

给定预测值 G'_{t+1} 和真实值 G_{t+1} ,本文使用 F 范数来度量两个矩阵的差异.

$$Frobenious Loss = \|G'_{t+1} - G_{t+1}\|_F \quad (5)$$

$\|\cdot\|_F$ 为矩阵的 F 范,对矩阵 A 计算 F 范数,显然, F 范数值越小,预测效果越好.

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2} \quad (6)$$

4.4 实验效果

在本文模型中,参数 k 是用于提取属性集合的时刻个数,直接决定用于预测的数据,因而首先验证参数 k 对预测结果的影响.下面以 Facebook 数据集 2008 年 12 月的角色矩阵为预测目标, k 取值从 2 到 7,每次分别计算预测矩阵 G'_{t+1} 和真实矩阵 G_{t+1} 的 F 范数值.图 9 为对每次实验重复 10 遍后取平均的结果,可以看出, k 的值从 2 变化到 7,预测效果并没有随之变好,这说明用作预测的属性个数并非越多越好.事实上这也是合理的:一方面, k 取值过大带来太多的预测属性,很可能造成数据过拟合;另一方面,可以认为只有离当前时刻较近的历史数据才能有效帮助预测下一时刻网络状态,太远的历史数据可能反而不利于预测效果.当 k 取 3 时,预测效果相对理想.

下面分别在 Enron,Facebook 以及 DBLP 这 3 个数据集上验证 MTR-RP 模型的有效性.此处参数 k 均设定为 3,因而预测目标从时刻 5 开始,分别表示 Enron 数据集中 2001 年 5 月~12 月(如图 10 所示)、Facebook 数据集中 2008 年 5 月~12 月(如图 11 所示)、DBLP 数据集中 2006 年~2011 年(如图 12 所示).

对图 10~图 12 的分析包括以下几点.

- (1) 总体分析.

可以看到,MTR-RP 模型在规模不同的 3 个真实数据集中都取得了很好的效果.与直接用一阶模型方法

STR-RP 相比,MTR-RP 得到更准确的预测值.这说明节点在 4 种角色上分别对应的取值是具有一定联系的.本文的节点角色取值在归一化处理,节点在各个角色上的取值存在求和为 1 的关系.PRE 和 TM 都只使用了前一个时刻的数据做预测,在 3 个数据集上的预测效果都不如 MTR-RP,STR-RP;并且从曲线形状来看,PRE 和 TM 模型的曲线都比较波动.相比之下,MTR-RP 和 STR-RP 能够得到相对稳定的预测结果.

(2) DBLP 网络的有趣现象.

观察 PRE 曲线与其他 3 条曲线的距离可以发现一个有趣的现象:相比 Enron 和 Facebook 网络,在 DBLP 数据集中,PRE 的预测效果与其余 3 种方法更为接近.PRE 直接使用当前时刻节点的角色作为下一时刻的预测值,这是基于相邻时刻网络结构不会发生剧烈变化的假设.继续分析可知,在 Enron 和 Facebook 网络中 PRE 表现很差.这说明 Enron 和 Facebook 网络结构并不稳定.事实上,2001 年 7 月~10 月间,Enron 公司发生了巨大的人事调动,年底公司破产,Enron 网络结构理应是不稳定的;而 2008 年的 Facebook 网络正处于超速发展中,6 月正式成为全球最大、增长最快的社交网络,Facebook 的网络结构也不会是稳定的;但在 DBLP 网络中,学者一般具有稳定的研究兴趣和合作学者,网络结构自然相对稳定.

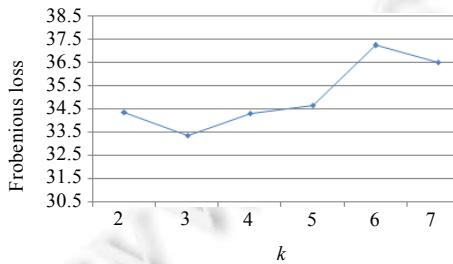


Fig.9 Influence of k to the prediction result—Facebook dataset

图 9 参数 k 对预测结果的影响——Facebook 数据集

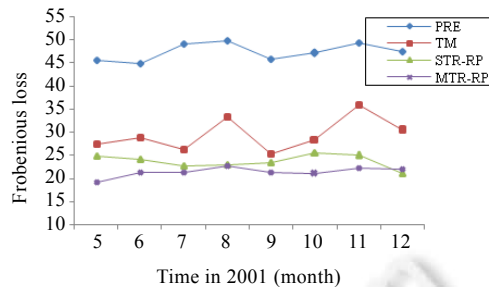


Fig.10 Prediction in Enron dataset

图 10 Enron 数据集预测效果

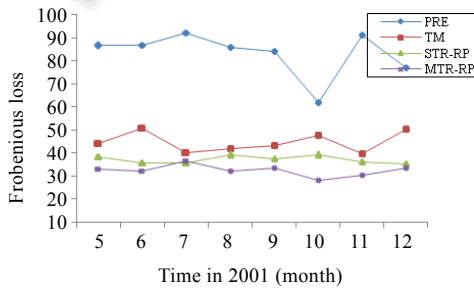


Fig.11 Prediction in Facebook dataset

图 11 Facebook 数据集预测效果

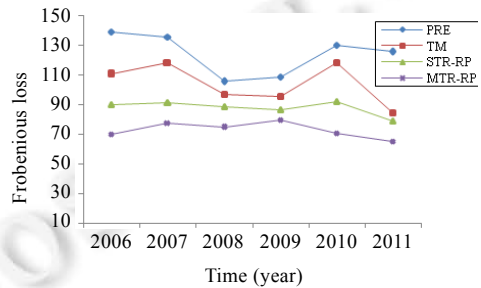


Fig.12 Prediction in DBLP dataset

图 12 DBLP 数据集预测效果

5 相关工作

目前,动态网络的结构演化分析的相关研究可以分为动态网络结构预测、动态网络的演化模式挖掘以及动态社团发现等.动态网络结构预测的相关研究主要包括网络的链接预测以及节点中心性预测两部分.链接预测主要有:Sarkar 等人^[13]提出了一种基于隐藏空间模型(latent space model)的方法对动态网络的边进行预测,该方法基于马尔可夫假设,即,下一时刻节点的边至于当前时刻有关系;Liben 等人^[14]研究动态网络的链接预测问题,提出了基于分类思想的网络链接预测方法;Huang 等人^[15]将动态网络的链接数据视为时间序列,通过定义新的时间序列相似性度量进行链接预测.动态网络的模式挖掘主要集中于网络频繁子图模式挖掘问题的研究^[16].Borgwardt 等人^[17]提出了动态频繁子图的概念,提出了 Dynamic GREW 算法用于挖掘动态频繁子图,通过应用

后缀树挖掘单边动态频繁子图,迭代地合并小的动态频繁子图得到规模更大的动态频繁子图;以 Borgwardt 等人的工作为基础,Wachersreuther 等人^[18]提出了 DFS 算法挖掘动态图中的动态频繁子图,采用字符串方式将静态频繁子图编码,应用字符串的公共子串发现方法得到这些静态图共有的边出现模式,从而获得动态图中的动态频繁子图;Berlingerio 等人^[19]提出一种基于频繁模式挖掘的演化规则挖掘算法 GREM,解决了动态网络的局部演化规律挖掘问题.动态网络社团挖掘方法主要分增量聚类 and 演化聚类两类.增量聚类是分别对每个时刻的网络快照进行聚类;演化聚类多采用时间平滑假设,在多个时刻上对聚类进行连续分析.Toyoda 等人^[20]给出了挖掘社团演化的 6 种类型(出现、消失、增长、收缩、分裂和合并),为每种类型分别定义相应的度量,这是社团演化问题较早的研究之一;Hopcroft 等人^[21]使用演化聚类的方法研究 NEC 的文献引文网络中社团结构的演化情况,发现网络中可供跟踪的稳定存在社团.

6 总 结

本文解决了动态网络的结构表示、结构演化及角色预测等问题,通过将静态网络的角色发现扩展至动态网络,提出动态网络的角色模型,并以此为基础,将结构预测问题角色预测,提出基于多目标回归思想的动态网络角色预测算法 MTR-RP.动态网络是目前复杂网络研究领域中最具活力的新兴研究方向,相比于静态网络的研究成果,目前动态网络的研究还处于起步阶段.本文只针对其中的演化和预测问题进行研究,传统静态网络中许多问题都需要在动态网络中得到进一步研究与扩展,未来的研究工作将继续关注动态网络的演化问题.

References:

- [1] Albert R, Barabasi AL. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002,74(1):47–97. [doi:10.1103/RevModPhys.74.47]
- [2] Wang XF, Li X, Chen GR. *The Theory and Application of Complex Networks*. Beijing: Tsinghua University Press, 2006 (in Chinese).
- [3] Biggs N, Lloyd EK, Wilson RJ. *Graph Theory 1736~1936*. Oxford University Press, 1976.
- [4] Fang BX. *Online social network Analysis*. Beijing: Electronic Industry Press, 2014 (in Chinese).
- [5] Kim J, Wilhelm T. What is a complex graph? *Physica A: Statistical Mechanics and its Applications*, 2008,387(11):2637–2652. [doi: 10.1016/j.physa.2008.01.015]
- [6] Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*, 2001,411:41–42. [doi: 10.1038/35075138]
- [7] Fang JQ, Wang XF, Zheng ZG, Bi Q, Di ZR, Li X. A new cross science: Network science (I). *Progress in Physics*, 2007,21(3): 239–337 (in Chinese with English abstract).
- [8] Henderson K, Gallagher B, Eliassi-Rad T, Tong HH. Rolx: Structural role extraction & mining in large graphs. In: *Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. ACM Press, 2012. 1231–1239. [doi: 10.1145/2339530.2339723]
- [9] Rossi R, Gallagher B, Neville J, Henderson K. Dynamic behavioral mixed-membership model for large evolving networks. *arXiv preprint arXiv:1205.2056*, 2012.
- [10] Rossi R, Gallagher B, Neville J, Henderson K. Role-Dynamics: Fast mining of large dynamic networks. In: *Proc. of the 21st Int'l Conf. on Companion on World Wide Web*. ACM Press, 2012. 997–1006. [doi: 10.1145/2187980.2188234]
- [11] Rossi RA, Gallagher B, Neville J, Henderson K. Modeling dynamic behavior in large evolving graphs. In: *Proc. of the 6th ACM Int'l Conf. on Web Search and Data Mining*. ACM Press, 2013. 667–676. [doi: 10.1145/2433396.2433479]
- [12] Henderson K, Gallagher B, Li L, Akoglu L, Eliassi-Rad T, Tong HH, Faloutsos C. It's who you know: Graph mining using recursive structural features. In: *Proc. of the 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. ACM Press, 2011. 663–671. [doi: 10.1145/2020408.2020512]
- [13] Sarkar P, Moore AW. Dynamic social network analysis using latent space models. *ACM Special Interest Group on Knowledge Discovery and Data Mining*, 2005,7(2):2330.
- [14] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 2007,58(7):1019–1031. [doi: 10.1002/asi.20591]

- [15] Huang Z, Lin DKJ. The time-series link prediction problem with applications in communication surveillance. *INFORMS Journal on Computing*, 2009,21(2):286–303. [doi: 10.1287/ijoc.1080.0292]
- [16] Gao L, Yang JY, Qin GM. Methods for patternmining in dynamic networks and applications. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(9):2042–2061. <http://www.jos.org.cn/1000-9825/4439.htm> [doi: 10.3724/SP.J.1001.2013.04439]
- [17] Borgwardt KM, Kriegel HP, Wackersreuther P. Pattern mining in frequent dynamic subgraphs. In: *Proc. of the 6th IEEE Int'l Conf. on Data Mining*. 2006. 818–822. [doi: 10.1109/ICDM.2006.124]
- [18] Wackersreuther B, Wackersreuther P, Oswald A, Böhm C, Borgwardt KM. Frequent subgraph discovery in dynamic networks. In: *Proc. of the 8th Workshop on Mining and Learning with Graphs*. 2010. 155–162. [doi: 10.1145/1830252.1830272]
- [19] Berlingerio M, Bonchi F, Bringmann B, Gionis A. Mining graph evolution rules. In: *Proc. of the Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer-Verlag, 2009. 115–130. [doi: 10.1007/978-3-642-04180-8_25]
- [20] Toyoda M, Kitsuregawa M. Extracting evolution of Web communities from a series of Web archives. In: *Proc. of the 14th ACM Conf. on Hypertext and Hypermedia*. ACM Press, 2003. 28–37. [doi: 10.1145/900051.900059]
- [21] Hopcroft J, Khan O, Kulis B, Selman B. Tracking evolving communities in large linked networks. *Proc. of the National Academy of Sciences*, 2004,101(Suppl. 1):5249–5253. [doi: 10.1073/pnas.0307750100]

附中文参考文献:

- [2] 汪小帆,李翔,陈关荣.复杂网络理论及其应用.北京:清华大学出版社,2006.
- [4] 方滨兴.在线社交网络分析.北京:电子工业出版社,2014.
- [7] 方锦清,汪小帆,郑志刚,毕桥,狄增如,李翔.一门崭新的交叉科学:网络科学(上).*物理学进展*,2007,21(3):239–337.
- [16] 高琳,杨建业,覃桂敏.动态网络模式挖掘方法及其应用.*软件学报*,2013,24(9):2042–2061. <http://www.jos.org.cn/1000-9825/4439.htm> [doi: 10.3724/SP.J.1001.2013.04439]



李川(1977—),男,河南郑州人,博士,副教授,CCF 专业会员,主要研究领域为数据库,数据挖掘,信息网络数据分析,社会网络分析,生物信息学.



胡绍林(1964—),男,博士,教授,博士生导师,主要研究领域为航天安全与大数据分析技术,过程监控与故障诊断技术,复杂系统建模与仿真.



冯冰清(1990—),女,助理工程师,主要研究领域为数据库,数据挖掘,信息网络.



杨宁(1974—),男,博士,讲师,CCF 专业会员,主要研究领域为时空数据挖掘,时序序列挖掘,异构信息网络挖掘,网络上的信息处理.



李艳梅(1994—),女,硕士,主要研究领域为数据库,数据挖掘,信息网络.



唐常杰(1946—),男,教授,博士生导师,CCF 杰出会员,主要研究领域为数据库系统,数据挖掘和数据仓库,知识工程,计算机安全.