

基于网络同步的链路预测连边机理分析研究

潘永昊¹ 于洪涛¹

摘 要 链路预测是研究复杂网络结构演化趋势的重要组成部分, 用于预测网络丢失的连边和未来可能出现的连边, 具有极大的理论和应用价值. 当前链路预测研究成果主要基于网络结构特征对连边进行预测, 具体分析其连边机理的研究较少. 网络同步的研究能够深刻反映节点的动力学演化行为与网络结构之间的内在机理. 本文针对链路预测考虑的静态网络引入节点动力学模型构成动态网络, 通过分析链路预测连边与动态网络模型同步之间的关系, 对链路预测连边机理进行分析研究. 通过实验与理论分析总结发现了链路预测连边具有同步能力稳定性的规律. 进一步讨论了链路预测连边的动力学机理, 并揭示了链路预测连边机理与真实网络演化的差别.

关键词 复杂网络, 链路预测, 同步, 主稳定函数方法

引用格式 潘永昊, 于洪涛. 基于网络同步的链路预测连边机理分析研究. 自动化学报, 2020, 46(12): 2607–2616

DOI 10.16383/j.aas.c180469

Analysis of Linkage Mechanism of Link Prediction Based on Network Synchronization

PAN Yong-Hao¹ YU Hong-Tao¹

Abstract Link prediction is an important part of the study of the evolutionary trend of complex network structure. It is used to predict the missing links and future links of network, and has great theoretical and practical value. The current link prediction researches are mainly based on the characteristics of the network structure. The study of network synchronization can profoundly reflect the internal mechanism between the node's dynamic evolution behavior and network structure. In this paper, we introduce a node dynamic model for link prediction research. By analyzing the relationship between the link prediction and the synchronization of the dynamic network model, we analyze the link prediction mechanism. Furthermore, we discuss the dynamic mechanism of link prediction, and reveal the difference between link prediction mechanism and real network evolution.

Key words Complex network, link prediction, synchronization, master stability function

Citation Pan Yong-Hao, Yu Hong-Tao. Analysis of linkage mechanism of link prediction based on network synchronization. *Acta Automatica Sinica*, 2020, 46(12): 2607–2616

近年来随着科学技术的不断发展, 人们的生活、工业生产越来越多地依托各类复杂网络系统, 如社交网络、电力网络、交通运输网络、工控网络等, 这些复杂网络系统在方便人们生活、提高生产效率等方面发挥了极大的作用. 在复杂网络系统的应用中, 常常需要对网络的结构连边信息进行链路预测^[1], 这种预测既包括对已经存在但未被观测到的连边的预测, 也包括了未来可能出现的连边的预测. 近年来, 关于链路预测的研究已经成为复杂网络理

论研究的一个重要部分, 已有的研究成果主要基于网络拓扑结构, 而基于网络同步的链路预测研究成果相对较少, 复杂网络中的链路预测问题仍然具有很大的研究空间.

唯物辩证法中, 对立统一规律揭示了“事物发展变化的源泉和动力在于自身的矛盾运动”, 统一性和斗争性是矛盾双方的固有属性, 对立面之间既相互依存、相互渗透、相互贯通, 同时也相互排斥、相互否定, 矛盾双方既统一又斗争, 从而推动事物发展变化. 网络的演化发展过程中, 节点之间通过连边关系相互影响、对立统一, 是网络演化过程中的基本动力. 链路预测与网络同步正是这种动力在不同角度的体现, 链路预测研究网络静态结构演化, 网络同步则是研究节点动力学演化过程与网络静态结构的相互关系, 因此, 把链路预测与网络同步联系在一起, 能够进一步接近并且深入研究网络演化发展运动的动力和规律.

现有链路预测研究主要从网络拓扑结构出发,

收稿日期 2018-07-07 录用日期 2018-11-05

Manuscript received July 7, 2018; accepted November 5, 2018
国家自然科学基金创新研究群体科学基金 (61521003), 国家自然科学基金 (61803384) 资助

Supported by Science Fund for Creative Research Groups of National Natural Science Foundation of China (61521003) and National Natural Science Foundation of China (61803384)

本文责任编辑 陈积明

Recommended by Associate Editor CHEN Ji-Ming

1. 国家数字交换系统工程技术研究中心 郑州 450002

1. National Digital Switching System Engineering and Technological Research and Development Center, Zhengzhou 450002

通过充分挖掘拓扑结构特征中包含的信息,对网络的连边情况进行预测. 现有链路预测方法多为基于相似性的方法,基于局部信息相似性的指标如共同邻居 (Common neighbours, CN)^[2]、Adamic-Adar (AA)^[3]、资源分配 (Resource allocation, RA)^[4]、偏好受限 (Preferential attachment, PA)^[5]、拓展资源分配 (Extended resource allocation)^[6] 等,基于路径相似性的指标如 Local path (LP)、Katz 等,以及基于随机游走相似性的指标如基于随机游走的余弦相似性 (Cos+)^[7]、有重启的随机游走 (Random walk with restart, RWR)^[8] 等,另外还有其它相似性指标^[1]. 在链路预测中还有一种复杂的框架,基于似然分析的链路预测方法,如层次结构模型^[9]、随机分块模型^[10]、闭路模型^[11] 等. 近年来有学者利用不同链路预测方法的互补性,提出了对多种链路预测方法的叠加优化方法,如文献 [12] 中采用三种不同的 OWA (Ordered weighted averaging) 算子对九种基于局部信息的结构相似性指标进行了融合,文献 [13] 将链路预测问题定义为二分类问题,并提出一种基于 Adaboost 的链路预测优化算法,利用 Boosting 方法通过错误反馈提升弱学习算法,从而得到强学习算法的思想,并利用链路预测计算结果的互补性原则选取若干链路预测指标作为弱分类器,提出一种基于 AdaBoost 算法的链路预测优化算法. 文献 [14] 则提出一种基于 Choquet 模糊积分的链路预测算法,引入模糊数学中模糊测度和模糊积分的概念,在考虑不同链路预测指标的重要性和交互作用的基础上,提出对不同链路预测指标进行叠加优化的算法. 从以上研究中可以看出,仅仅从网络拓扑结构出发的链路预测研究,并没有考虑到每个节点自身的动力学特征.

对网络中的节点行为和特性进行动力学建模,即为复杂网络动力学模型. 同步正是复杂网络动力学模型中,节点动力学演化行为的重要现象. 网络同步主要就是研究节点动力学行为与网络拓扑结构的关系^[15]. 在 Pecora 等^[16] 提出的主稳定函数方法中推导出了网络达到完全同步的必要条件,网络的 Laplacian 矩阵的特征值就是其中一个主要变量,网络同步的研究主要包括对不同网络的同步研究以及网络同步的方法,如单层网络^[17-28] 和多层网络^[29-32] 的同步、Lyapunov 方法^[33]、主稳定函数方法^[16] 以及连接图方法^[34-37] 等. 在 Zhou 等^[38] 的研究中表明,节点的同步过程能够反演网络拓扑结构,进行网络拓扑结构识别以及参数识别. 可以看到,网络同步研究包含了节点动力学行为与网络拓扑结构两个部分,互为支撑且互为补充.

基于以上研究结论,本文的研究中考虑节点动力学行为与网络拓扑结构两个方面对链路预测问题

进行研究分析,具体为链路预测计算的连边对于网络同步过程的影响,通过对节点的同步过程以及网络的同步能力的讨论,对链路预测连边的动力学特性以及动力学机理进行研究,揭示链路预测的连边在网络同步中的基本规律. 通过数值仿真实验总结发现,链路预测计算的连边对于网络的同步过程影响极小,即网络拓扑结构按照链路预测的方式演化 (后文中称为网络的链路预测演化),网络的同步能力与演化前近似相等,具有稳定性 (后文中称为链路预测的同步能力稳定性),并通过理论分析,给出同步能力稳定性的理论解释,同时也从网络同步的角度,对链路预测进行了讨论,分析了链路预测连边的动力学机理. 本文创新点如下: 1) 从网络同步的角度出发研究链路预测连边与节点动力学演化的相互作用关系; 2) 结合节点动力学研究讨论了链路预测的连边机理; 3) 以同步能力作为网络的一种结构特征,提出链路预测的同步能力稳定性.

1 预备知识

1.1 四种链路预测方法定义

在链路预测计算中,通过计算两个节点之间的相似性,得出这两个节点之间连边的可能性. 下面给出文中选择的四种具有代表性的链路预测相似性指标的定义.

1) CN 指标^[2]. 计算两个节点的共同邻居节点集合的势,对于网络中的节点 v_x , 定义其邻居节点集合为 $\Gamma(x)$, 则两个节点 v_x, v_y 的相似性定义为共同邻居的个数, 定义为

$$s_{xy} = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

2) AA 指标^[3]. 考虑节点共同邻居度的信息,认为度小的共同邻居节点的贡献大于度大的共同邻居,节点的相似性指标定义为

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k_z} \quad (2)$$

该指标计算了共同邻居集合中所有节点的度的对数的倒数和.

3) PA 指标^[5]. 直接把两个节点的相似性定义为了节点度的乘积,认为节点的连接概率正比与节点的度, 定义为

$$s_{xy} = k_x k_y \quad (3)$$

4) Katz 指标^[1]. 考虑了节点的全局路径相似性,即两个节点的所有路径数,通过路径权重的加权, 定义为

$$s_{xy} = \sum_{l=1}^{\infty} \alpha^l \times \left| \text{path}_{x,y}^{(l)} \right| = \alpha A_{xy} + \alpha^2 (A^2)_{xy} + \alpha^3 (A^3)_{xy} + \cdots \quad (4)$$

1.2 网络同步的主稳定函数方法

考虑 N 个相同节点构成的连续时间的复杂动态网络

$$\dot{x}_i = f(x_i) - c \sum_{j=1}^N l_{ij} H(x_j), \quad i = 1, 2, \dots, N \quad (5)$$

其中, 节点的动力学方程为 $\dot{x} = f(x)$, $x_i \in \mathbf{R}^n$ 为第 i 节点的状态变量, 常数 $c > 0$ 为网络的耦合强度, 耦合矩阵 $L = (l_{ij})_{N \times N}$ 为 Laplacian 矩阵, $H: \mathbf{R}^n \rightarrow \mathbf{R}^n$ 为各个节点状态变量之间的内连函数. 设孤立节点动力学方程为 $\dot{s} = f(s)$, 其解为 $s(t)$.

在主稳定函数方法中, 假设动态网络模型满足如下条件: 1) 所有节点的动力学完全相同; 2) 各个节点之间的耦合函数完全相同; 3) 同步流形是不变流形; 4) 在同步流形附近可以做线性变化. 令 $\xi_i(t) = x_i(t) - s(t)$, 将网络方程在 s 上做变分, 得到

$$\dot{\xi}_i = Df(s) \xi_i - c \sum_{j=1}^N l_{ij} DH(s) \xi_j, \quad i = 1, 2, \dots, N \quad (6)$$

其中, $Df(s)$, $DH(s)$ 分别是 $f(x)$ 和 $H(x)$ 在 s 处的 Jacobian 矩阵, 令 $\xi = [\xi_1, \xi_2, \dots, \xi_N]$, 则方程可重写为

$$\dot{\xi} = Df(s) \xi - c DH(s) \xi L^T \quad (7)$$

设存在可逆矩阵 P , 使得 $L^T = P \Lambda P^{-1}$, $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\}$, 可知 $\lambda_k (k = 1, 2, \dots, N)$ 是耦合矩阵 L 的特征值. 令 $\eta = [\eta_1, \eta_2, \dots, \eta_N] = \xi P$, 则有

$$\dot{\eta}_k = [Df(s) - c \lambda_k DH(s)] \eta_k, \quad k = 2, 3, \dots, N \quad (8)$$

本文只讨论无权无向网络, 因此耦合矩阵 L 为实对称非负定矩阵, 其特征值有且仅有一个重数为 1 的零特征值, 其余均为正实数, 即 $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_N$. 主稳定方程可以写为

$$\dot{y} = [Df(s) - \alpha DH(s)] y \quad (9)$$

在给定节点动力学函数 f 和内连矩阵 H 时, 主稳定方程的最大 Lyapunov 指数 $L_{\max}$ 是变量 α 的函数, 记为 $L_{\max} = L_{\max}(\alpha)$, 主稳定函数为负值的区

域称为主稳定区域, 即为同步化区域, 记为 $SR = \{\alpha | L_{\max}(\alpha) < 0\}$. 耦合强度 c 与耦合矩阵 L 的所有特征值的乘积属于 SR , 即

$$c \lambda_k \in SR, \quad k = 2, 3, \dots, N \quad (10)$$

是网络达到局部完全同步的必要条件.

根据网络的同步化区域的不同情形, 可分为四种类型: 1) 类型 I, 同步化区域为有界区域 $SR = (\alpha_1, \alpha_2)$; 2) 类型 II, 同步化区域为无界区域 $SR = (\alpha_1, +\infty)$; 3) 类型 III, 同步化区域为不连通多区域 $SR = (\alpha_1, \alpha_2) \cup (\alpha_3, \alpha_4) \cup \dots \cup (\alpha_{2k-1}, +\infty)$; 4) 类型 IV, 同步化区域为空集 $SR = \emptyset$.

假设 1. 在本文中, 假设网络的同步化区域为无界情况, 即类型 II, 使用 Laplacian 矩阵的最小非零特征值, 即 λ_2 作为刻画网络同步能力 (Synchronizability) 的指标. λ_2 越大, 网络的同步能力越强.

2 链路预测网络同步能力数值仿真实验

选择一些网络进行同步能力的数值仿真实验, 对比网络进行链路预测演化与随机演化前后节点的动力学同步过程的差异与网络的同步能力的变化, 研究网络的链路预测演化与同步能力的关系.

2.1 网络节点的同步过程对比情形

对 NW 小世界网络和 Jazz (爵士音乐家网络), 以及按照方法 1 产生的链路预测演化网络和随机演化网络进行数值仿真实验, 假设网络节点采用式 (11) 所示的 Rossler 混沌系统, 其参数为 $\omega = 0.1$, $\alpha = 0.165$, $\beta = 0.2$, $\gamma = 10$. 耦合方式采用线性耗散耦合, 内耦合矩阵 H 选为单位矩阵, 耦合系数为 $c = 1.0$. 每个节点的三维初始状态在 $(1, -1)$ 内随机取值.

$$\begin{cases} \dot{x} = -\omega y - z \\ \dot{y} = \omega x + \alpha y \\ \dot{z} = \beta + z(x - y) \end{cases} \quad (9)$$

经过计算得到上述 Rossler 混沌系统的最大 Lyapunov 指数为 $\gamma_{\max} \approx 0.11$, 两个网络的最小非零特征值分别为 3.8409 和 0.5720, 满足 $|c \lambda_2| > \gamma_{\max}$, 网络能够达到完全同步.

方法 1. 对于网络 $G(V, E)$, 其中 V 表示节点集合, E 表示连边集合, $|E|$ 表示边集合的势, 即网络中连边的数量. 分别按链路预测 CN, AA, PA, Katz 指标为网络增加 $0.1|E|$ 的连边, 节点集合 V 保持不变, 即得到四种链路预测演化网络, 分别记为 $G_{\text{CN}}(V, E_{\text{CN}})$, $G_{\text{AA}}(V, E_{\text{AA}})$, $G_{\text{PA}}(V, E_{\text{PA}})$, $G_{\text{Katz}}(V, E_{\text{Katz}})$. 作为对比, 为原始网络 $G(V, E)$

随机增加 $0.1|E|$ 的连边, 得到随机演化网络, 记为 $G_{rd}(V, E_{rd})$.

NW 小世界网络与 Jazz 网络中节点的 x 分量

的动力轨迹如图 1 和图 2 所示. 图中 X_1 表示原始网络中的节点的 x 分量集合, X_2 表示链路预测演化网络中节点的 x 分量集合, X_3 表示随机演化网络中

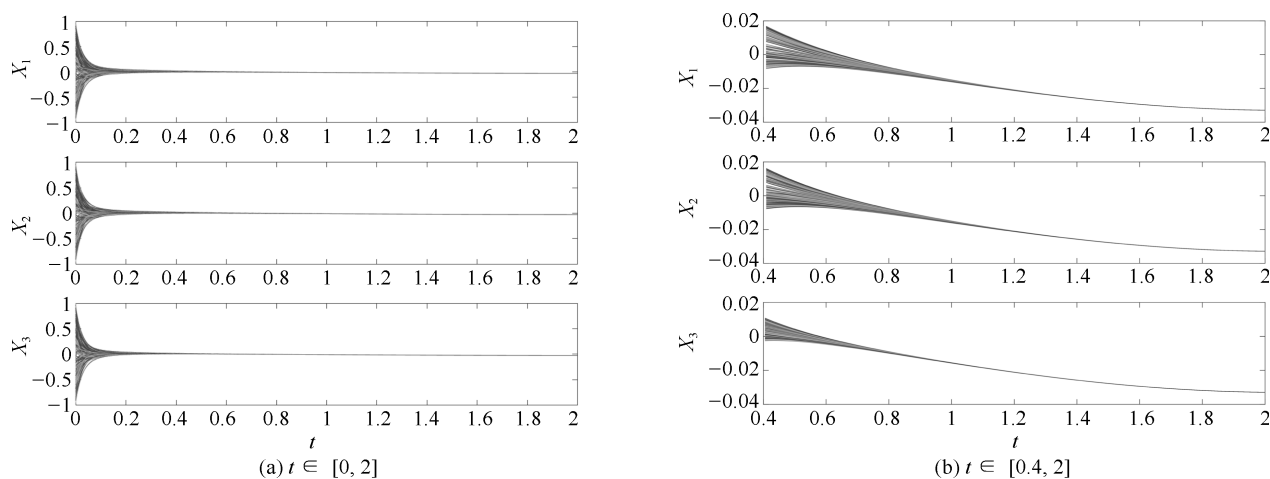


图 1 NW 小世界网络的 x 分量运动轨迹图

Fig. 1 x component trajectories of NW small-world network

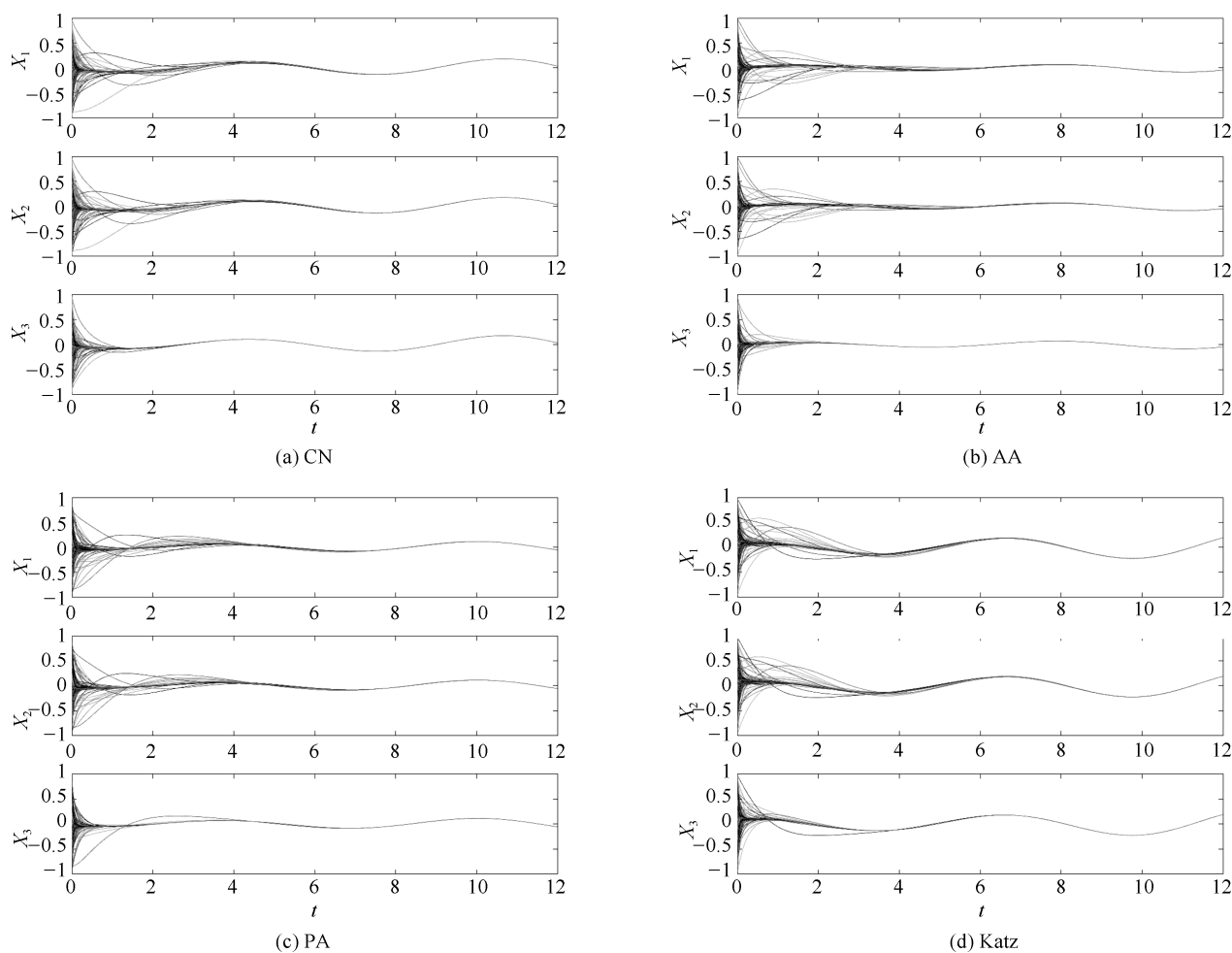


图 2 Jazz 网络的 x 分量运动轨迹图

Fig. 2 x component trajectories of Jazz network

节点的 x 分量集合.

图 1 中 NW 小世界网络具有 200 个节点, 平均路径长度为 2.406, 聚类系数为 0.6113, 节点平均度为 21.855. 图 1(a) 为从 $t = 0$ 开始, 描述节点运动同步化过程. 图 1(b) 为 $t = 0.4$ 开始, 是对于图 1(a) 的局部放大. 图中 X_1 与 X_2 在 $t = 1.2$ 时达到完全同步, X_3 在 $t = 0.8$ 时达到完全同步. 可以观察得到, 原始网络与链路预测演化网络 (注: 使用链路预测的 CN 方法) 达到完全同步所需时间相同, 随机演化网络所需时间明显少于原始网络与链路预测演化网络.

图 2 中 Jazz 网络, 具有 198 个节点, 平均路径长度为 2.235, 聚类系数为 0.6175, 节点平均度为 27.697. 四个分图分别对比了四种链路预测方法演化网络的同步化过程. 从图中可以明显观察到, X_2 表示的链路预测演化网络的同步化过程与 X_1 表示的原始网络的同步化过程完全一致, X_3 表示的随机演化网络达到完全同步所需的时间明显少于前两者, 同时值得注意的是, 四个分图中随机演化网络的同步化过程各有不同, 图 2(a) 和图 2(b) 中所需的时间明显少于图 2(c) 和图 2(d) 所需的时间.

通过图 1 和图 2 的数值仿真实验对比, 可以明显观察得到, 链路预测演化网络与原始网络的节点动力演化过程具有很高的一致性, 各个节点由随机赋予的初始状态到完全同步所需的时间相同, 而随机演化网络达到完全同步的时间明显小于前两者.

2.2 同步能力的对比情形

由假设 1 中的条件, 对 NW 小世界网络, Jazz 网络, USair (美国航空网络), PB (政治博客网络), Yeast (蛋白质相互作用网络), FWFB (佛罗里达海湾雨季的食物链网络), 以及按照方法 2 产生的训练集网络、链路预测演化网络和随机演化网络的同步能力进行计算对比. NW 小世界网络参数与上一节中一致, 其余网络的具体参数见表 1 中网络的特征参数统计, N 为节点总数, M 为边总数, k 为平均度, L 为平均路径长度, C 为聚类系数.

表 1 网络的拓扑特征参数
Table 1 Topological characteristic parameters of the network

网络	N	M	k	L	C
Jazz	198	2 742	27.697	2.235	0.6175
USair	332	2 126	12.807	2.738	0.6252
PB	1 222	16 714	27.355	2.738	0.3203
Yeast	2 375	11 693	9.848	5.096	0.3057
FWFB	128	2 075	32.422	1.776	0.3346

方法 2. 对于网络 $G(V, E)$, 其中 V 表示节点集合, E 表示连边集合, $|E|$ 表示边集合的势, 即网络中连边的数量. 在保证网络连通的条件下, 随机划分 90% 的边组成训练集网络, 记为 $G_t(V, E_t)$, 分别按链路预测 CN, AA, PA, Katz 方法为训练集网络增加 $0.1|E|$ 条连边, 节点集合 V 保持不变, 即得到四种链路预测演化网络, 分别记为 $G_{CN}(V, E_{CN})$, $G_{AA}(V, E_{AA})$, $G_{PA}(V, E_{PA})$, $G_{Katz}(V, E_{Katz})$, 作为对比, 为训练集网络 $G_t(V, E_t)$ 随机增加 $0.1|E|$ 条连边, 得到随机演化网络, 记为 $G_{rd}(V, E_{rd})$.

注: $|E| = |E_{CN}| = |E_{AA}| = |E_{PA}| = |E_{Katz}| = |E_{rd}|$.

图 3 中分别对 6 种网络以及其演化以后的网络的同步能力, 纵轴为同步能力数值, 横轴为网络连边数, 五角星为原始网络 $G(V, E)$ 对应的同步能力与连边数, 左下角的点为训练集网络 $G_t(V, E_t)$ 的同步能力与连边数, 右下角分别为四种链路预测方法演化后的网络 $G_{CN}(V, E_{CN})$, $G_{AA}(V, E_{AA})$, $G_{PA}(V, E_{PA})$, $G_{Katz}(V, E_{Katz})$ 的同步能力, 右上角为随机演化网络 $G_{rd}(V, E_{rd})$ 的同步能力 (注: 图中虚线用于辅助表示数值的相对关系, 无实际物理含义).

表 2 中计算了演化网络与训练集网络的同步能力的增长率, 对图 3 中描述的数据增长关系进行量化计算.

表 2 演化网络同步能力增长率 (%)

Table 2 Growth rate of evolutionary network synchronizability (%)

	G_{CN}	G_{AA}	G_{PA}	G_{Katz}	G_{rd}
NW	2.74	2.98	1.92	2.51	19.95
Jazz	0.09	0.10	0.51	0.07	402.73
USair	0.04	0.04	0.45	0.04	599.47
PB	0	0	0	0	183.85
Yeast	0	0	0.01	0	1 277.52
FWFB	0	0	0.27	0.08	491.64

由图 3 和表 2 中的数据可以看出, 链路预测演化网络的同步能力与训练集网络的同步能力数值非常相近, 在 PB, Yeast, FWFB 网络中有完全相同的情况出现. 作为对比, 随机演化网络的同步能力较训练集网络有大幅的增加, 而原始网络的同步能力介于链路预测演化网络与随机演化网络之间. 同时注意到, NW 小世界网络的链路预测演化网络的同步能力增长大于其他网络, 而随机演化网络的同步能力增长小于其他网络.

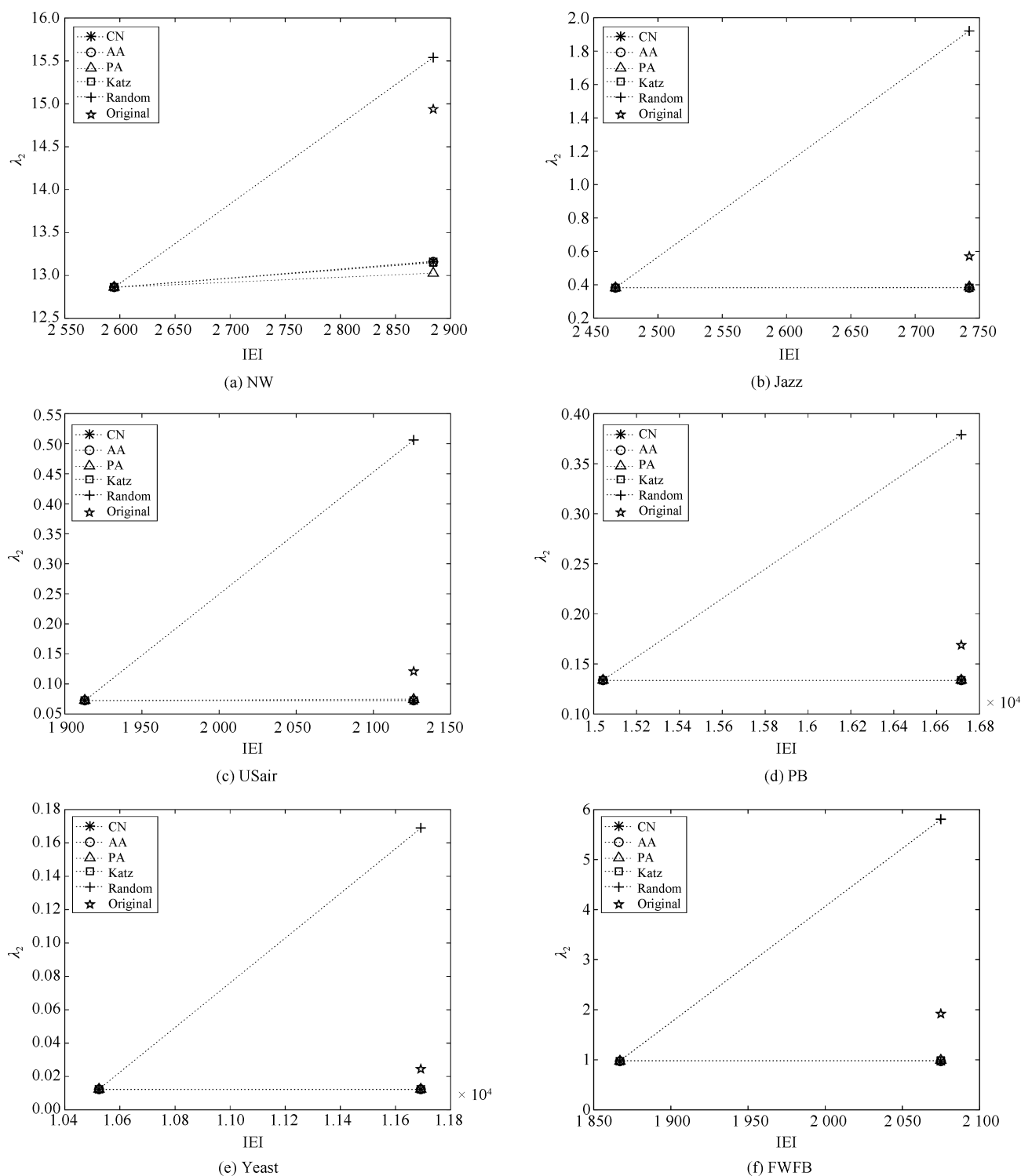


图3 网络的同步能力

Fig. 3 Synchronizability of network

2.3 链路预测网络规模与同步能力的关系

链路预测在不同的应用场景中, 根据实际需要选择不同的预测规模对网络的连边进行补充, 同样在假设 1 的条件下, 对 NW 小世界网络、Jazz

网络、USair (美国航空网络)、PB (政治博客网络)、Yeast (蛋白质相互作用网络)、FWFB (佛罗里达海湾雨季的食物链网络), 分别按照网络中现有连边数量 1%, 2%, 3%, 5%, 7%, 10% 的规模进行链路预测增加连边, 链路预测方法仍然选择 CN, AA,

PA, Katz 方法, 分别计算不同规模的链路预测演化网络与训练集网络的同步能力的增长率, 结果如表 3~6 所示.

表 3 CN 演化网络同步能力增长率 (%)

Table 3 Growth rate of CN evolutionary network synchronizability (%)

	1	2	3	5	7	10
NW	0.11	0.22	0.31	0.82	1.03	2.74
Jazz	0	0	0.02	0.04	0.07	0.09
USair	0	0	0	0.01	0.02	0.04
PB	0	0	0	0	0	0
Yeast	0	0	0	0	0	0
FWEB	0	0	0	0	0	0

表 4 AA 演化网络同步能力增长率 (%)

Table 4 Growth rate of AA evolutionary network synchronizability (%)

	1	2	3	5	7	10
NW	0.10	0.23	0.30	0.93	1.34	2.98
Jazz	0	0	0.02	0.05	0.08	0.10
USair	0	0	0	0	0.02	0.04
PB	0	0	0	0	0	0
Yeast	0	0	0	0	0	0
FWEB	0	0	0	0	0	0

表 5 PA 演化网络同步能力增长率 (%)

Table 5 Growth rate of PA evolutionary network synchronizability (%)

	1	2	3	5	7	10
NW	0.07	0.29	0.31	0.72	1.06	1.92
Jazz	0.09	0.10	0.13	0.21	0.42	0.51
USair	0.03	0.04	0.06	0.20	0.32	0.45
PB	0	0	0	0	0	0
Yeast	0	0	0	0	0	0.01
FWEB	0	0	0	0.05	0.12	0.27

由表中数据可以看出, 链路预测增加连边的规模与同步能力增长有关系, 从整体上来看, 预测规模越小, 则同步能力增长越小, 从单个网络的数据看, 其增长率多数为零, 或近似为零.

2.4 局部网络结构的同步化情形

构造一个含有 8 个节点的网络如图 4(a) 所示,

表 6 Katz 演化网络同步能力增长率 (%)

Table 6 Growth rate of Katz evolutionary network synchronizability (%)

	1	2	3	5	7	10
NW	0.15	0.32	0.39	0.94	1.12	2.51
Jazz	0	0	0.02	0.04	0.06	0.07
USair	0	0	0	0.01	0.02	0.04
PB	0	0	0	0	0	0
Yeast	0	0	0	0	0	0
FWEB	0	0	0	0	0.03	0.08

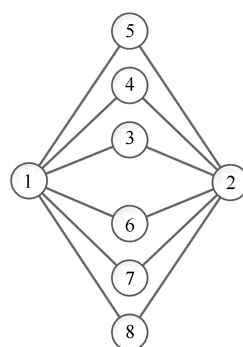
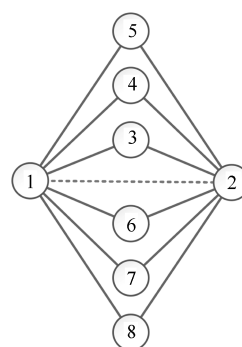
(a) 原始网络
(a) Original network(b) 链路预测演化网络
(b) Link prediction evolution network

图 4 8 节点网络

Fig. 4 8-node network

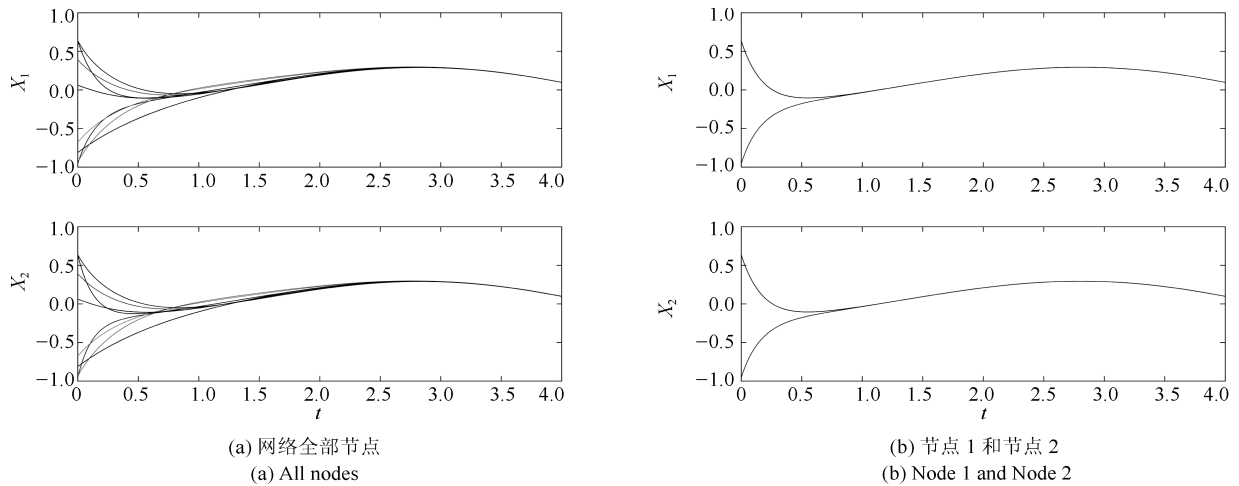
节点 1 和节点 2 共有 6 个共同邻居节点相连. 按链路预测方法对网络进行演化预测, CN、AA、PA 和 Katz 方法都能够计算出节点 1 与节点 2 有很强的相似性, 在这两个节点之间形成连边的可能性很大. 图 4(b) 中所示的网络即为链路预测演化的网络.

对图 4 所示网络的节点运动轨迹进行数值模拟, 假设网络节点采用式 (11) 所示的 Rossler 混沌系统, 其参数为 $\omega = 0.1$, $\alpha = 0.165$, $\beta = 0.2$, $\gamma = 10$. 耦合方式采用线性耗散耦合, 内耦合矩阵 H 选为单位矩阵, 耦合系数为 $c = 1.0$. 每个节点的三维初始状态在 $(1, -1)$ 内随机取值. X_1 为原始网络节点的集合, X_2 为预测演化网络节点的集合.

网络中节点的 x 分量运动轨迹如图 5 所示, 其中图 5(a) 为网络全部节点 x 分量的运动轨迹图, 图 5(b) 为节点 1 和节点 2 的 x 分运动轨迹图, 可以观察到, 原始网络与链路预测演化网络的同步化情况完全一致. 分别计算两个网络的同步能力数值, 两个网络均为 $\lambda_2 = 2$ 其同步能力相等.

3 讨论部分

由以上数值仿真实验可以看出, 链路预测演化

图 5 8 节点网络的 x 分量运动轨迹图Fig. 5 x component trajectories of 8-node network

网络与原始网络的同步能力非常接近,在部分计算中甚至相等,说明网络在链路预测演化中,其同步能力具有稳定性,不会因为新增加的连边而有大幅度的提高.在表 2 中给出了增长幅度的计算结果,除 NW 小世界网络的同步能力增长较大,为 1.92% 到 2.98%,在其余真实数据网络中,同步能力增长幅度都在 1% 以内,变化很小,在部分计算中同步能力没有变化,增长幅度为 0.同时,图 1 和图 2 中的运动轨迹仿真中,各个网络在链路预测演化前后的节点运动轨迹对比看出,链路预测演化网络与原始网络同步过程高度一致,也说明网络的链路预测演化具有同步能力稳定性特点.

链路预测在计算中是对于未连边节点对之间存在连边的可能性进行计算,通常以结构相似性程度为依据,其相似性数值越高,则该节点对之间存在连边的可能性越大.而所有节点对的相似性数值分布,具有近似幂律分布的特点,即数值较大的只占有很少一部分.因此,在考虑链路预测规模的时候,规模越小,则链路预测的连边越有意义,而规模越大时,其中包含的相似性数值较低的节点对数量增大,不具有很大的实际预测意义.通过表 3~6 的数据可以看出,总体上链路预测的规模越小,同步能力增长越少,说明在连边可能性越大的节点对之间增加连边,对网络同步能力的影响越小.从单个网络的数据来看,在预测规模为原始连边数量 1%, 2%, 3%, 5% 的时候,同步能力增长为 0,即链路预测对于网络的同步能力没有影响.

另外构造了 8 个节点的局部网络结构的同步能力分析说明,链路预测计算出的具有很强连边可能性的节点之间,往往已经存在了很多连通的路径,这两个节点之间有没有连边,对于两个节点的动力学同步几乎没有影响.这也说明在整个网络中,链路预

测计算出的连边,对于整个网络的同步化过程没有太大影响.另外,在网络中增加连边,网络的平均度增加,部分节点之间增加连边,使得节点之间相互的动力学作用更直接,因此在网络中增加连边时,网络的同步能力一定不会减小.

网络按链路预测方法进行演化,为部分节点之间增加了连边,使得网络的同步能力发生变化,为方便分析,引入如下定理 1,网络特征值谱与节点度序列的相关性定理.

定理 1^[39]. 假设 G 为 N 个节点构成的网络, L 为 G 对应的 Laplacian 矩阵,并且节点度序列向量 $\mathbf{d} = (d_1, d_2, \dots, d_N)^T$ 与特征值向量 $\boldsymbol{\lambda}(L) = (\lambda_1, \lambda_2, \dots, \lambda_N)^T$ 里的元素都按照升序排列的,则

$$\delta = \frac{\|\boldsymbol{\lambda}(L) - \mathbf{d}\|_2}{\|\mathbf{d}\|_2} \leq \frac{\sqrt{\|\mathbf{d}\|_1}}{\|\mathbf{d}\|_2} \leq \sqrt{\frac{N}{\|\mathbf{d}\|_1}} \quad (12)$$

定理 1 表示了网络的拉普拉斯矩阵的特征值向量 $\boldsymbol{\lambda}(L)$ 与节点度序列向量之差极小,从而得出了两个向量之间非常相似,即 Laplacian 矩阵的特征值分布与节点度序列非常相似,具有强相关性.

在网络的链路预测演化中,两个节点之间增加连边,使得两个节点的度分别加 1,从而网络的度序列有所变化.在式 (1) 定义的 CN 指标中,两个节点的度分别为 $k_x = |\Gamma(x)|$, $k_y = |\Gamma(y)|$,所以这两个节点的相似性指标满足如下条件

$$s_{xy} \leq \min(k_x, k_y) \quad (13)$$

即度小的节点与其他节点之间的相似性数值一定低于其节点度,在链路预测计算中与其他节点形成新的连边的可能性极小.在式 (2) 定义的 AA 指标的定义中, $|Z| \leq \min(k_x, k_y)$,即度小的节点,与其他

节点的共同邻居集合的势小于度大的节点的共同邻居集合, 在计算中求和的数目较小, 相似性期望较低, 与其他节点产生新的连边的可能性较低. 在式 (3) 定义的 PA 指标中, 两个节点的相似性数值为节点度的乘积, 显然度小的节点在计算中的相似性明显低于度大的节点, 不可能产生新的连边. 在式 (4) 定义的 Katz 指标中, 主项为 CN 指标的结果, 即表示最短路径对于相似性的贡献最大, 其他路径数由 α 的幂次项依次加权. 很显然, 对于度小的节点, 其与其他节点产生新的连边的可能性较低.

实际的计算中, 网络的度分布情况各异, 链路预测增加连边的节点总会出现度大的节点分布区域, 度大的节点的度增加的可能性远大于度小的节点. 事实上, 上述现象对于其他链路预测方法也成立, 同时, 这也符合网络演化的优先连接规律. 对于定理 1 中的度序列的分布向量 $\mathbf{d}$ 而言, 链路预测不改变向量 $\mathbf{d}$ 中排在前面元素, 由定理 1 的结论可知, 链路预测对于特征值向量 $\boldsymbol{\lambda}(L)$ 中 λ_2 的影响很小, 即为链路预测具有同步能力稳定性.

在同步能力计算部分, 对原始网络划分训练集后, 训练集网络的同步能力明显低于原始网络, 是因为随机减少网络中的连边, 使得度序列整体减小, 特征值普相应减小, 同步能力减小. 同时, 减少网络中的连边, 网络中路径平均长度增加, 节点之间相互的动力学影响减弱, 网络整体同步能力降低. 从这里也可以看出链路预测的演化完全按照网络的结构特征出发, 是在现有结构的基础上按结构特征“生长”出一些连边, 而实际的网络演化中, 还有一部分连边是随机产生的. 对于训练集网络完全随机演化后的同步能力大于原始网络, 说明真实网络演化中, 既有链路预测演化的部分, 也有随机演化的部分.

4 结论

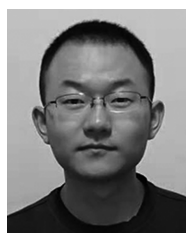
链路预测是复杂网络领域研究网络结构信息与网络演化的重要部分, 网络同步是研究网络结构与节点动力学同步之间的关系, 网络结构的演化会使得网络节点动力学同步发生变化, 而通过节点的动力学同步现象又可以研究网络结构演化的动力学机理. 在本文的研究中发现链路预测演化具有同步能力稳定的特点, 也就是网络通过链路预测演化前后, 网络的同步能力具有稳定性, 并结合链路预测相似性指标的定义, 说明这种稳定性是链路预测的原理方法决定的. 通过对于网络结构演化与节点动力学行为的研究, 进一步分析了结构演化的动力学机理, 清晰的反映出链路预测研究与真实网络演化过程在动力学机理上的区别, 链路预测的演化趋势是完全基于现有网络结构信息的, 而在真实网络演化中既包含了链路预测的部分, 也包含随机演化的部分. 值

得注意的是, 本文的研究把复杂网络领域中链路预测研究与网络同步研究两个方向联系在了一起, 用网络同步的研究方法研究分析了链路预测问题, 同时也提出了网络演化中的同步问题, 是一个新的研究网络结构演化与节点动力学行为的研究思路.

References

- 1 Lv L, Zhou T. Link prediction in complex networks: a survey. *Physica A Statistical Mechanics and Its Applications*, 2011, **390**(6): 1150–1170
- 2 Mitzenmacher M. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 2004, **1**(2): 226–251
- 3 Adamic L A, Adar E. Friends and neighbors on the web. *Social Networks*, 2003, **25**(3): 211–230
- 4 Zhou T, Lv L, Zhang Y C. Predicting missing links via local information. *European Physical Journal B*, 2009, **71**(4): 623–630
- 5 Xie Y B, Zhou T, Wang B H. Scale-free networks without growth. *Physica A Statistical Mechanics and Its Applications*, 2005, **387**(7): 1683–1688
- 6 Liu S, Ji X, Liu C, Bai Y. Extended resource allocation index for link prediction of complex network. *Physica A Statistical Mechanics and Its Applications*, 2017, 479: 174–183
- 7 Fouss F, Pirotte A, Renders J M, Saerens M. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. In: *Proceedings of the 2007 IEEE Transactions on Knowledge and Data Engineering*, 2007, **19**(3): 355–369
- 8 Tong H, Faloutsos C, Pan J Y. Fast random walk with restart and its applications. In: *Proceedings of the 2006 International Conference on Data Mining*. IEEE Computer Society, 2006: 613–622
- 9 Clauset A, Moore C, Newman M E. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008, **453**(7191): 98–98
- 10 Airoldi E M, Blei D M, Fienberg S E, Xing E P. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 2008, **9**(5): 1981–1981
- 11 Pan L, Zhou T, Linyuan L, Hu C K. Predicting missing links and identifying spurious links via likelihood analysis. *Scientific Reports*, 2016, **6**: 22955–22955
- 12 He Y L, Liu J N K, Hu Y X, Wang X Z. OWA operator based link prediction ensemble for social network. *Expert Systems with Applications*, 2015, **42**(1): 21–50
- 13 Wu Zu-Feng, Liang Qi, Liu Qiao, Qin Zhi-Guang. Modified link prediction algorithm based on AdaBoost. *Journal on Communications*, 2014, **35**(3): 116–123
(吴祖峰, 梁棋, 刘峤, 秦志光. 基于 AdaBoost 的链路预测优化算法. *通信学报*, 2014, **35**(3): 116–123)
- 14 Yu H T, Wang S H, Ma Q Q. Link prediction algorithm based on the Choquet fuzzy integral. *Intelligent Data Analysis*, 2016, **20**(4): 809–824

- 15 Lu Jun-An, Liu Hui, Chen Juan. *Synchronization in Complex Dynamical Networks*. Beijing: Higher Education Press, 2016. 47–47
(陆君安, 刘慧, 陈娟. 复杂动态网络的同步. 北京: 高等教育出版社 2016. 47–47)
- 16 Pecora L M, Carroll T L. Master stability functions for synchronized chaos in arrays of oscillators. *Physical Review Letters*, 1998, 80: 2109–2112
- 17 Liu H, Chen J, Lu J A, Cao M Generalized synchronization in complex dynamical networks via adaptive couplings. *Physica A*, 389: 1759–1770
- 18 Lu W, Liu B, Chen T. Cluster synchronization in networks of coupled nonidentical dynamical systems. *Chaos an Interdisciplinary Journal of Nonlinear Science*, 2010, 20(1): 175–175
- 19 Chen L, Lu J A, Chi K T. Synchronization: an obstacle to identification of network topology. *IEEE Transactions on Circuits and Systems II Express Briefs*, 2009, 56(4): 310–314
- 20 Suykens J A K, Osipov G V. Introduction to focus issue: synchronization in complex networks. *Chaos an Interdisciplinary Journal of Nonlinear Science*, 2008, 18(3): 268–268
- 21 Arenas A, Díaz-Guilera A, Kurths J, Moreno Y, Zhou C. Synchronization in complex networks. *Physics Reports*, 2008, 469(3): 93–153
- 22 Wu W, Chen T. Global synchronization criteria of linearly coupled neural network systems with time-varying coupling. *IEEE Transactions on Neural Networks*, 2008, 19(2): 319–332
- 23 Han X P, Lu J A, Wu X Q. Synchronization of impulsively coupled systems. *International Journal of Bifurcation and Chaos*, 2008, 18(5): 1539–1549
- 24 Zhou J, Lu J A, Lv J. Pinning adaptive synchronization of a general complex dynamical network. *Automatica*, 2009, 45(2): 598–599
- 25 Liu Q, Fang J Q, Li Y. Synchronization and control of halo-chaos in beam transport network with small world topology. *Communications in Theoretical Physics*, 2007, 47(4): 752–758
- 26 Xin B L, Xiao F W, Jin Q F. Topological transition features and synchronizability of a weighted hybrid preferential network. *Physica A Statistical Mechanics and Its Applications*, 2006, 371(2): 841–850
- 27 Lu W, Chen T, Chen G. Synchronization analysis of linearly coupled systems described by differential equations with a coupling delay. *Physica D Nonlinear Phenomena*, 2006, 221(2): 118–134
- 28 Barahona M, Pecora L M. Synchronization in small-world systems. *Physical Review Letters*, 2002, 89(5): 054101–054101
- 29 Xu Ming-Ming, Lu Jun-An, Zhou Jin. Synchronizability and eigenvalues of two-layer star networks. *Acta Physica Sinica*, 2016, 65(2): 383–395
(徐明明, 陆君安, 周进. 两层星形网络的特征值谱及同步能力. 物理学报, 2016, 65(2): 383–395)
- 30 Aguirre J, Sevilla-Escoboza R, Gutiérrez R, Papo D, Buldú J. Synchronization of interconnected networks: the role of connector nodes. *Physical Review Letters*, 2014, 112(24): 248701–248701
- 31 Um J, Minnhagen P, Kim B J. Synchronization in interdependent networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2011, 21: 5712–5712
- 32 Lu R, Yu W, Lu J, Xue A. Synchronization on complex networks of networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 25(11): 2110–2118
- 33 Lu Jun-An, Liu Hui, Chen Juan. *Synchronization in Complex Dynamical Networks*. Beijing: Higher Education Press, 2016. 81–81
(陆君安, 刘慧, 陈娟. 复杂动态网络的同步. 北京: 高等教育出版社 2016. 81–81)
- 34 Belykh V N, Belykh I V, Hasler M. Connection graph stability method for synchronized coupled chaotic systems. *Physica D Nonlinear Phenomena*, 2004, 195(1): 159–187
- 35 Schaer J. Generalized connection graph method for synchronization in asymmetrical networks. *Physica D Nonlinear Phenomena*, 2006, 224(1): 42–51
- 36 Liu H, Cao M, Wu C W. Coupling strength allocation for synchronization in complex networks using spectral graph theory. *IEEE Transactions on Circuits and Systems I Regular Papers*, 2017, 61(5): 1520–1530
- 37 Liu H, Cao M, Wu C W, Lu J A, Chi K T. Synchronization in directed complex networks using graph comparison tools. *IEEE Transactions on Circuits and Systems I Regular Papers*, 2017, 62(4): 1185–1194
- 38 Zhou J, Lu J A. Topology identification of weighted complex dynamical networks. *Physica A Statistical Mechanics and Its Applications*, 2007, 386(1): 481–491
- 39 Zhan C, Chen G, Yeung L F. On the distributions of laplacian eigenvalues versus node degrees in complex networks. *Physica A Statistical Mechanics and Its Applications*, 2010, 389(8): 1779–1788



潘永昊 国家数字交换系工程技术研究中心硕士研究生. 主要研究方向为复杂网络链路预测.

E-mail: panyonghao2016@163.com

(PAN Yong-Hao Master student at the National Digital Switching System Engineering Technological Research Center. His main research interest is link prediction on complex network.)



于洪涛 国家数字交换系工程技术研究中心研究员, 博士. 主要研究方向为网络大数据分析处理. 本文通信作者.

E-mail: 15937101921@139.com

(YU Hong-Tao Ph.D., professor at the National Digital Switching System Engineering Technological Research Center. His research interest covers network big data analysis and processing. Corresponding author of this paper.)