

网络科学中的互联网加

——理念、评述和展望

史定华

上海大学数学系, 上海 200444

摘要: 本文从互联网加的角度回顾网络科学的某些进展。众所周知网络科学的第一波浪潮, 互联网加大数据, 是由外国学者在千禧年前夕掀起的。紧接着的网络科学的第二波浪潮, 互联网加动力学, 则是由中国学者发动的。我们预测, 网络科学的第三波浪潮, 互联网加博弈论和纯数学研究或许即将来临。衷心希望中国年轻学者对网络科学做出更大的贡献。

关键词: 网络科学, 互联网加, 大数据, 动力学, 博弈论, 纯数学

互联网+大数据

众所周知, 网络科学第一波浪潮是由外国学者掀起的。他们利用互联网获取大量实际网络的数据, 提出了与传统随机图论不同的网络模型, 通过实证统计和模型分析发现了大多数网络具有小世界和无标度特性。小世界特性由网络平均路径和群集系数刻画; 无标度特性由网络度分布描述。主要的结果是: 网络(节点之间的)平均路径 $\bar{l} \sim \ln N$, 网络(节点)度分布 $P(k) \sim k^{-\gamma}$, 其中 N 为网络节点数, $1 < \gamma$ 为网络自身决定的常数, 而网络群集系数大仅只是定性说法。基本的网络模型有: 随机网络 ER 模型^[1]; 小世界网络 WS 模型^[2]; 无标度网络 BA 模型^[3]。小世界特性经过发展在信息检索中有着重要的应用, 这应归功于可导航网络模型^[4]的提出。这类网络不仅易于形成短的路径, 而且人们还可利用网络的局部信息找到这条短路径。可导航网络赋予节点坐标信息, 然而实际信息往往是节点连接数目或其它。而无标度特性研究虽报道了不少观察结果: 如对蓄意攻击 hub 节点脆弱而对节点的随机失效稳健, 及当 $\gamma \leq 3$ 时, 疾病容易在网上传播等, 但都存有争议, 见 NSR 上文章^[5]的相关介绍。

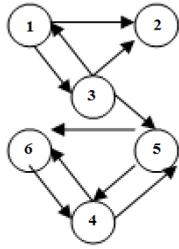
网络马氏链

模型网络的生成规则, 都是在上一步完成后再重复加点和连线, 因此可通过马氏链来刻画。如文[6]用非齐次马氏链刻画了 BA 模型, 文[7]提出了增长网络马氏链模型。

网络马氏链另一个重要应用是研究网页排序。对网络而言, 众多节点在其中的地位并不一样。如何识别哪些节点是最重要的? 借助节点大数据信息, 可通过它在网络中的得分来描述。设节点得分 x_i , 它依赖于邻居节点的数量和重要性。于是得到 $x_i = c \sum_j a_{ij} x_j$, 其中 c 是比例常数, a_{ij} 是网络邻接矩阵的元素。写成矩阵形式有 $x = cAx$ 或 $Ax = c^{-1}x$, 即 x 是邻接矩阵 A 特征值 c^{-1} 对应的特征向量。 c^{-1} 是矩阵 A 的哪个特征值呢? 理论分析应该是模最大的特征值 λ_1 。于是有: $x = \lambda_1^{-1}Ax$, 这就是计算特征向量中心性^[8]的公式。

对大型网络而言, 计算特征值是件困难的事情。能够找到模最大的特征值为 1 的矩阵就可省去计算特征值。根据矩阵理论, 不可约随机矩阵模最大的特征值是 1。随机矩阵是行和为 1 的非负矩阵。若对应的有向图(网络)强连通, 则它不可约。

对有向网络, 节点 i 有连线指向节点 j 时 a_{ij} 为 1, 这时矩阵 A 一般不对称, 且矩阵行和不一定为 1, 其中还可能存在全为 0 的行。如以下面有向网络为例,



$$\bar{P} = \begin{bmatrix} \frac{1}{60} & \frac{28}{60} & \frac{28}{60} & \frac{1}{60} & \frac{1}{60} & \frac{1}{60} \\ \frac{10}{60} & \frac{10}{60} & \frac{10}{60} & \frac{10}{60} & \frac{10}{60} & \frac{10}{60} \\ \frac{19}{60} & \frac{19}{60} & \frac{1}{60} & \frac{1}{60} & \frac{19}{60} & \frac{1}{60} \\ \frac{1}{60} & \frac{1}{60} & \frac{1}{60} & \frac{1}{60} & \frac{28}{60} & \frac{28}{60} \\ \frac{1}{60} & \frac{1}{60} & \frac{1}{60} & \frac{28}{60} & \frac{1}{60} & \frac{28}{60} \\ \frac{1}{60} & \frac{1}{60} & \frac{1}{60} & \frac{55}{60} & \frac{1}{60} & \frac{1}{60} \end{bmatrix}$$

$$\hat{P} = \begin{bmatrix} & \frac{1}{3} & \frac{1}{3} & & & \frac{1}{3} \\ & & & & & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & & \frac{1}{4} & & \frac{1}{4} \\ & & & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ & & & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ & & \frac{1}{2} & & & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

PageRank 算法: 为了得到不可约随机矩阵, 先采用随机修正 $\bar{a}_{ij} = \alpha a_{ij} / k_i^{out} + (1 - \alpha) / N$, 其中 N 为节点数, k_i^{out} 是节点出度, α 为凸组合常数。例如取 $\alpha = 9/10$, 再归一化得 \bar{P} 。问题是 α 取什么值最合适? 根据经验 PageRank 算法^[9]取 $\alpha = 0.85$ 。

LeaderRank 算法: 吕琳媛等人采用增加一个领导节点来修正, 让它与所有节点双向连线保证强连通。新的邻接矩阵归一化后, 得不可约随机矩阵 \hat{P} , 命名为 LeaderRank 算法^[10]。

BrowseRank 算法: 马志明院士等人则利用用户浏览网页的大数据, 得到 $Q = [q_{ij}]$ 矩阵, 见图 1。然后将 $Q = [q_{ij}]$ 转化为不可约随机矩阵 $P = [q_{ij} / q_i]$, 命名为 BrowseRank 算法^[11]。

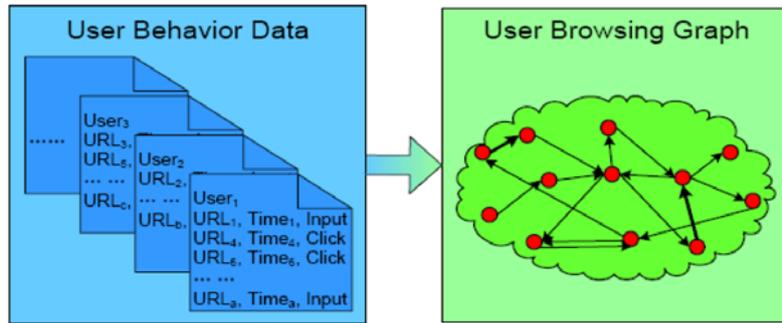


图 1 用户浏览网页的大数据和网页之间转移

计算平稳分布: 有了不可约随机矩阵, 基于马氏链, 就可以通过计算 $\pi \bar{P} = \pi$, $\pi \hat{P} = \pi$, 或 $\pi P = \pi$ 得到平稳分布 π , 然后按分量大小排序。因为随机矩阵模最大的特征值是 1, 这样就省去了计算邻接矩阵特征值的中间步骤。因此特征向量中心性才是谷歌搜索算法的源头! 而矩阵 \hat{P} 比较稀疏更易计算, 去掉领导节点可得归一化的节点价值向量。

互联网+动力学

网络科学第二波浪潮则是由中国学者发动的。主要思想是对每个节点加上动力学, 研究网络拓扑和动力学的相互影响, 具有**互联网+**的显著特征。开创性工作由汪小帆和陈关荣首先研究复杂网络同步问题, 随后带出了多智能体的一致性研究。他们^[12]考虑网络同步数学模型 $\dot{x}^i(t) = f(x^i(t)) - c \sum_{j \in V} l_{ij} H(x^j(t)), i \in V$, 其中 $f(\bullet)$ 和 $x^i(t)$ 描述了节点的非线性动力学, c 是耦合常数, V 是节点集合, $H(x^j(t))$ 是内部耦合机制, l_{ij} 是网络拓扑拉普拉斯矩阵元素。对无向连通网络完全同步, 他们得到条件 $\alpha < c \lambda_2$, 其中 α 与网络结构和节点动力学有关, λ_2 是网络拉普拉斯矩阵 L 的最小非零特征值, 这是无界区域情况。对有界区域情况, 外国学者^[13]得到特征比 λ_2 / λ_N 的条件 $\alpha_1 < c \lambda_2$ 和 $c \lambda_N < \alpha_2$ 。若网络自身不能实现同步, 他们又引入牵引控制使得网络能够被牵引到同步状态^[14]。对牵引节点数, 在自适应控制情况下吕金虎等人^[15]有近似估计, 李翔等人^[16]进一步得到公式 $N_D^P = N - \text{rank}(L)$ 。尔

后刘洋或等人^[17]考虑网络线性控制模型 $\dot{x}(t) = Ax(t) + Bu(t)$ ，其中 $x(t)$ 是 N 维状态向量， A 是耦合矩阵，描述网络拓扑和耦合强度， $u(t)$ 是 $M \leq N$ 维外部输入控制向量， B 是 $N \times M$ 维控制输入矩阵。采用经典的状态可控概念，即网络从任意初始状态能在有限时间被控制到希望的任意目标状态，他们利用图匹配方法，得到结构可控的驱动节点数公式 $N_D^S = \max\{1, N - |M^*|\}$ ，其中 $N - |M^*|$ 表示未匹配节点数。进一步，利用空穴(cavity)方法他们推导出无标度网络驱动节点比公式 $n_D^S \approx \exp\{-1/2[1 - 1/(\gamma - 1)]\langle k \rangle\}$ 。据此他们认为当 $\gamma = 2$ 时几乎所有节点都需要被控制。接着，王文旭等人^[18]利用矩阵特征值理论得到一般有向加权网络精确可控的驱动节点数公式 $N_D^E = \max\{\mu(\lambda_i)\}$ ，其中 $\mu(\lambda_i)$ 表示耦合矩阵 A 特征值的几何重数。三个控制节点数之间有何关系？李翔等人得到 $N_D^P \leq N_D^S$ ，并给出了不等的例子(文章待发)。我们猜测：对一般有向加权网络，应有关系 $N_D^S \leq N_D^E$ 。

网络最优化

网络科学首先是对大量存在的实际网络和模型网络进行实证研究，统计它们的拓扑性质；然后再研究其上的动力学行为并与功能相结合，这属于认识世界。在此基础上，下一步自然就是去改造世界，即寻找最好的网络结构，以便设计网络能够实现某些完美的功能。

Donetti等人^[19]提出保持连线数不变使得网络同步最优的问题，并引入了使特征比最大的缠绕网络(entangled networks)概念。然而缠绕网络是一个数学定义，其实当时并不识庐山真面目，见图2。他们设计了模拟退火算法去寻找缠绕网络，见图3，并且认为谱隙 λ_2 最大往往就是最优，据此对大规模网络提议了一个代数图论算法。宣琦等人^[20]则提出对称网络概念，即以结点度相同的网络作为近似，并提出平均路径最短算法去寻找。

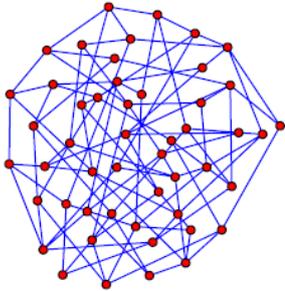


图2 缠绕网络的示意图

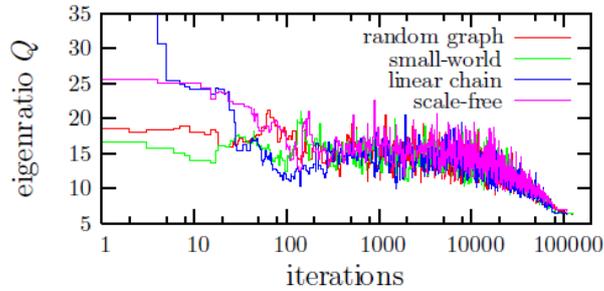


图3 模拟退火算法运行结果, 这里 $Q = \lambda_N/\lambda_2$

随后史定华和陈关荣等人^[21]首先研究了无向连通网络完全同步能力最强的网络结构形态，发现圈结构，即周长最大才是关键。由于同步能力最强等价于特征比 λ_2/λ_N 最大，这个特征比最大网络就是几乎全齐性网络。全齐性网络指网络中每个节点的度数、路和、周长都相同。对给定的网络规模，全齐性网络可能有许多，为了使得特征比最大还要求平均路径最短，周长最大，这时就有可能得牺牲点齐性，所以最优拓扑是几乎全齐性网络。在完成了同步能力最强网络结构形态研究之后，史定华等人又开始研究可控性好的网络结构形态。刘洋或等人^[17]认为无标度网络难以控制，但史定华和阎小勇等人^[22]发现自然数同余网络是 $\gamma = 2$ 的无标度网络，因拥有完美的长链结构反而具有最佳的强可控性，这里强可控指与连线权重无关。自然数同余网络以自然数作为节点，若每对自然数相除同为给定余数时则在它们之间连线，连线方向指定从小到大(当然也可以从大到小)。由于带余除法公式 $a_n^i = nr + i$ ，

只要余数 $r > 0$ ，就存在 r 条长链，使得只需控制 r 个最大度的根节点，而且面对蓄意攻击根节点稳健而对节点的随机失效脆弱。这些都与刘洋或等人^[17]的结论恰好相反。

我们发现的两类网络：几乎全齐性网络(对称性)和自然数同余网络(标度律)代表图形如图 4 所示。图 4b 中的自然数同余网络中虚线表示特殊情形 $r = 0$ ，即整除网络，这个退化情形没有了带余除法，因长链结构的缺失，它的可控性能差，需要控制网络半数节点。

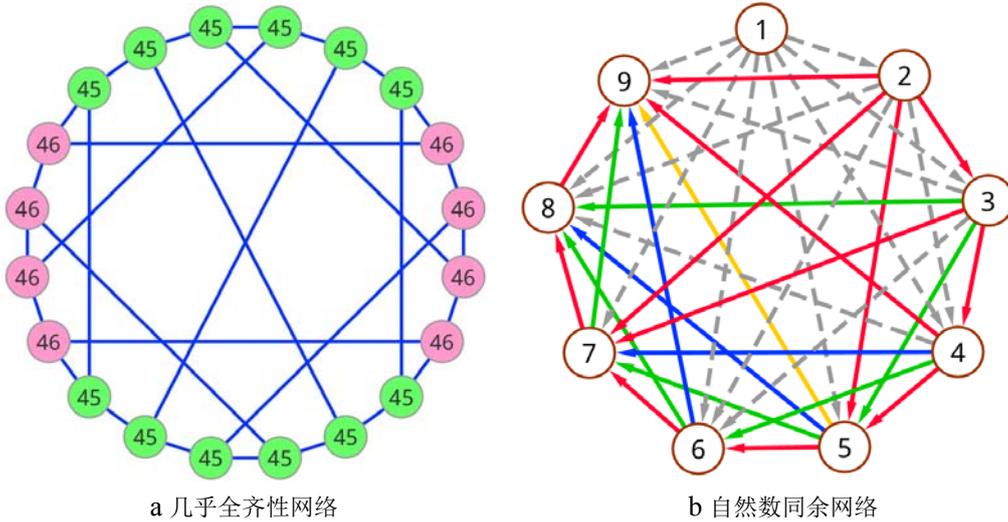


图 4 两类网络的例子

这两类网络从政治哲学上讲，对社会组织也极具参考价值。同步是要求群体步调一致协调运作，因此最好的结构自然应该是个体之间尽量整齐划一。可控是希望群体按照领导集体意图迈向指定目标，这时环环相扣的链式结构才能保证上级意图得以顺利贯彻。

互联网+博弈论

继达尔文基因突变和自然选择之后，Nowak^[23]提出进化的第三大原则——自然合作(合作互惠)。合作者为了帮助他人会付出一定的代价，而背叛者通常因不劳而获，其短期收益会高于合作者。在一次性囚徒困境博弈中，“纳什均衡”告诉我们背叛是最好的选择。但只有选择合作才能得到社会效益最优，那么这个社会是怎样合作起来的呢？

最近，Press和Dyson^[24]针对迭代囚徒困境博弈提出了一个新的研究博弈理论的框架。他们指出，在无限次重复博弈中长程记忆策略并不比一步记忆策略占优。由于一步记忆策略中个体的本轮行动只取决于上一轮双方的行为，因此可通过马氏链来揭示博弈双方策略与期望收益之间的关系。该理论为重新认识各种策略的性质及相互关系提供了全新的思想，有可能推动对囚徒困境乃至合作博弈理论研究产生革命性的转变。

用 X 和 Y 代表参与囚徒困境博弈的双方，以 C 表示合作， D 表示背叛。博弈状态有 4 种： CC, CD, DC, DD 。设收益矩阵元素为 R, S, T, P ，假定它们满足两个不等式： $T > R > P > S$ 和 $2R > T + S$ 。令 $\mathbf{p}^X = (p_1^X, p_2^X, p_3^X, p_4^X)^T$ 是 X 的策略向量， $\mathbf{p}^Y = (p_1^Y, p_3^Y, p_2^Y, p_4^Y)^T$ 是 Y 的策略向量。用 \mathbf{M} 表达各状态间的转移概率，有

$$\mathbf{M} = \begin{bmatrix} p_1^X p_1^Y & p_1^X (1-p_1^Y) & (1-p_1^X) p_1^Y & (1-p_1^X)(1-p_1^Y) \\ p_2^X p_3^Y & p_2^X (1-p_3^Y) & (1-p_2^X) p_3^Y & (1-p_2^X)(1-p_3^Y) \\ p_3^X p_2^Y & p_3^X (1-p_2^Y) & (1-p_3^X) p_2^Y & (1-p_3^X)(1-p_2^Y) \\ p_4^X p_4^Y & p_4^X (1-p_4^Y) & (1-p_4^X) p_4^Y & (1-p_4^X)(1-p_4^Y) \end{bmatrix}$$

而其平稳分布向量 \mathbf{v} 与矩阵 $\mathbf{M}' = \mathbf{M} - \mathbf{I}$ 的伴随矩阵(由原矩阵元素的代数余子式组成的转置矩阵) $\text{adj}(\mathbf{M}')$ 的每一行呈正比关系, 因为 $\text{adj}(\mathbf{M}')\mathbf{M}' = \det(\mathbf{M}')\mathbf{I} = 0$ 。

任意向量 \mathbf{u} 与 \mathbf{v} 的点积等于将 \mathbf{M}' 的第四列用 \mathbf{u} 替代后矩阵的行列式 $D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{u})$, 即

$$\mathbf{v}^T \cdot \mathbf{u} \equiv D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{u}) = \det \begin{bmatrix} -1 + p_1^X p_1^Y & -1 + p_1^X & -1 + p_1^Y & u_1 \\ p_2^X p_3^Y & -1 + p_2^X & p_3^Y & u_2 \\ p_3^X p_2^Y & p_3^X & -1 + p_2^Y & u_3 \\ p_4^X p_4^Y & p_4^X & p_4^Y & u_4 \end{bmatrix}, \quad (1)$$

式中的第 2 和第 3 列单独分别由 X 和 Y 控制, 称为控制策略。

$$\text{双方期望收益为 } E^X = \frac{\mathbf{v}^T \cdot \mathbf{u}^X}{\mathbf{v}^T \cdot \mathbf{1}} = \frac{D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{u}^X)}{D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{1})} \text{ 和 } E^Y = \frac{\mathbf{v}^T \cdot \mathbf{u}^Y}{\mathbf{v}^T \cdot \mathbf{1}} = \frac{D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{u}^Y)}{D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{1})}, \text{ 其}$$

中 $\mathbf{u}^X = (R, S, T, P)^T$ 和 $\mathbf{u}^Y = (R, T, S, P)^T$ 由囚徒困境博弈中收益矩阵元素组成。

若 X 选择 $\tilde{\mathbf{p}}^X = \alpha \mathbf{u}^X + \beta \mathbf{u}^Y + \gamma \mathbf{1}$ 或者 Y 选择 $\tilde{\mathbf{p}}^Y = \alpha \mathbf{u}^X + \beta \mathbf{u}^Y + \gamma \mathbf{1}$, 都会使 X 和 Y 的收益由于(1)式的行列式为零而满足如下的线性关系:

$$\alpha E^X + \beta E^Y + \gamma = \frac{D(\mathbf{p}^X, \mathbf{p}^Y, \alpha \mathbf{u}^X + \beta \mathbf{u}^Y + \gamma \mathbf{1})}{D(\mathbf{p}^X, \mathbf{p}^Y, \mathbf{1})} = 0. \quad (2)$$

它暗示着一个聪明的个体可以控制对手的长期收益并迫使其合作。

他们发现使(2)式成立的策略是零行列式策略(Zero-determinant strategy, 简称ZD策略)。不论对手采取何种策略, 使用ZD策略的个体都可以单方面保证双方收益满足线性关系, 从而可以控制不知情对手的收益。进一步, 他们详细研究了两大类重要ZD策略的子集——设定策略(Equalizer strategy)和剥削策略(Extortion strategy)的性质。前者指可以单方面让对手长期收益限定在某个预先设定的固定值; 而后者则可以保证自身收益是对手收益的 x 倍。当 $x=1$ 意味着二人长期收益相同, 著名的Tit-for-Tat(TFT, 针锋相对)策略是其一个特例; 而 $x > 1$ 时意味着剥削行为。之后的学者开始重点关注剥削策略的性质, 发现由于剥削策略与背叛策略之间收益相同, 它容易被背叛行为入侵而可能导致策略的不稳定。

现实生活中个体之间一般具有特定的网络组织结构, 它与合作演化行为存在着密切的联系, 因此网络互惠是一类重要的合作演化机制, 也是互联网加博弈的产物。随着群体网络结构特征的引入, 根本改变了均匀混合群体中博弈演化的结果。Szolnoki和Perc^[25]基于囚徒困境博弈发现, 剥削策略可以在大规模规则网络和无标度网络上稳定存在。

荣智海、吴枝喜和周涛等人^[26]探讨了策略演化时间尺度的多样性对网络上剥削策略的演化作用机理。通过比较规则格子、随机网络和无标度网络中的剥削策略演化过程, 他们发现策略演化时间尺度因素的引入会促使剥削策略在网络环境中的稳定存在, 并进一步导致合作行为的涌现。由于个体收益与时间尺度之间的反馈作用, 无标度网络中度大的节点更倾向于采取剥削策略, 这似乎预示合作行为在异质的无标度网络中更容易涌现。

衡量网络博弈合作涌现的指标有合作频率, 即趋于稳态时的合作者的比例, 以及背叛者出现的阈值和合作者湮灭的阈值。Szabó和Fath^[27]报道了囚徒困境博弈在5种规则格子上的合作行为, 见图5。个体X随机选择一个邻居Y的策略作为自己策略, 演化采用费米规则

$$P = \frac{1}{1 + \exp[(W_x - W_y)/K]}, \quad (3)$$

其中 W_x 是个体X本轮收益, 低于邻居Y收益时, 容易接受邻居策略; K 刻画选择理性。

从图中可以看出，当 $K > 0.5$ 时，合作行为在贝特格(d)上最不容易湮灭。一个未决的问题是到底何种网络结构——标度律抑或对称性——更容易促使合作保持或涌现？基于“物以类聚，人以群分”的古训，由于贝特格与由群生成的凯莱网络有关，我们猜测具有良好群结构的层次网络可能更易促使合作保持或涌现。

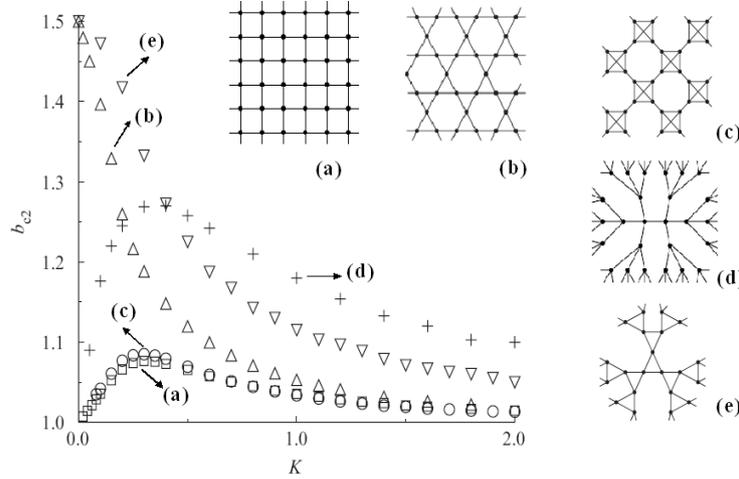


图 5 合作者湮灭阈值 b_{c2} 随 K 的变化情况

从宏观上看，网络传播过程节点的动态演化与网络博弈类似，它们都是0(易感、合作)，1(染病、背叛)交替直至稳定，只是微观机制网络传播不如网络博弈丰富。Kitsak等人^[28]发现对于单个传播源情形，hub节点或者高介数节点不一定是具有影响力的节点，而通过K-shell分解确定的网络核数(coreness)节点(即K-shell值大的节点)才是具有影响力的节点。最近，吕琳媛和周涛等人^[29]给出了节点度、H指数、核数三者之间一个漂亮的数学关系。注意到每层K-shell中的节点都具有很好的置换性，见图6，应该可能会与自同构群有关。

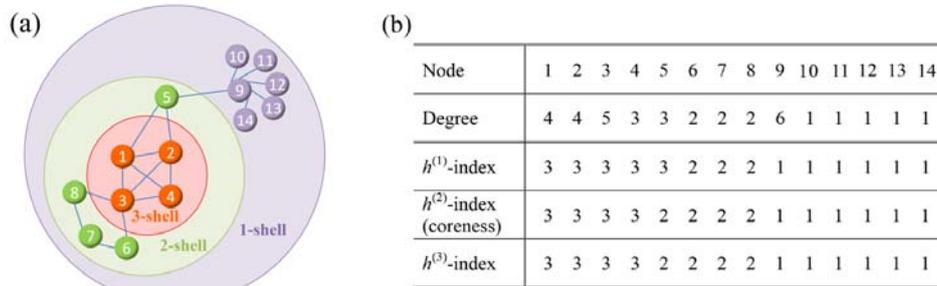


图 6 网络的 K-shell 分解和节点度、H 指数、核数

网络反问题

顾名思义，反问题是相对于正问题而言的。如何区分某个问题的“正”与“反”？大致可以这样认为：所谓正问题，一般是由因推果；而反问题则是倒果求因。反问题几乎都是非线性的和不适定的，因此反问题研究的难度一般比相应的正问题要大。

如前所述，网络科学主要是已知网络去研究网络的拓扑性质及其上的动力学特性，这是正问题。但许多实际情况，我们只能得到系统的部分信息，却要求重构网络，这是反问题。例如，前面介绍的网络最优化，只知道网络规模去设计同步最优网络。网络科学中研究热点的反问题有链路预测，网络重构等等，在这些领域中国学者做出了重要贡献。

链路预测是通过已知的网络节点和网络结构去预测节点之间未知的连接或未来的连接。已知的网络节点和网络结构往往只是冰山一角，如何窥见深藏海水中的冰山全貌自然不易。

链路预测本质上属于计算机数据挖掘方向之一，其方法主要基于马氏链和机器学习。链路预测对需要大量或昂贵实验才能推断节点之间是否存在连接的网络具有重要应用价值。另外，通过链路预测还能揭示网络演化的关键机制等等。随着互联网加理念的不断扩展，电子商务网站如雨后春笋出现，链路预测在推荐系统设计上也有重要应用。因为推荐系统设计可归结为用户-产品二分网络上的链路预测问题。吕琳媛和周涛等人在链路预测^[30]和推荐系统^[31]方面做了许多好的研究工作。他们发现链路预测的准确程度往往依赖网络的可预测性，于是最近又提出了结构一致性概念和结构微扰法^[32]。这不仅提供了一种链路预测的新方法，还提供了刻画网络结构的一个重要新参数。

然而，对节点间的连接无法被直接探测的复杂网络，从可观测数据和时间序列信号间接重构网络则更具有挑战性。王文旭和狄增如等人^[33]将信息领域中的压缩感知(Compressed sensing)理论^[34]与复杂网络理论相结合，提出了针对二元时间序列的网络重构方法。他们利用复杂网络的天然稀疏性，如图7中向量X的非零元素，相对网络节点数N要小得多，将复杂网络重构问题转化为稀疏信号重构问题，极大地降低了重构网络的数据需求。借助压缩感知理论解决网络重构问题的思想新颖，效果很好。

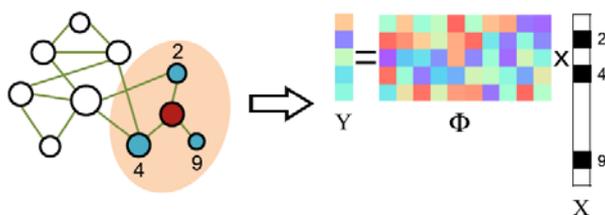


图7 重构一个节点局部结构的示意图

另外，随着 Y 染色体检测技术的成熟，为人类寻根开辟了科学新途径。自从有了姓氏之后，人们开始记录纸质家谱，但由于种种原因会出现中断和错误。而世代相传的血缘家谱就隐藏在 Y 染色体中，史定华^[35]则研究了如何用网络科学与方法重构家族血缘树的问题。

互联网+纯数学

超网络是网络科学发展的新方向，椭圆曲线是证明费马大定理的钥匙，BSD(贝赫和施维讷通-戴尔)猜想是7个“千禧年大奖难题”之一，它们之间会有互联网加的故事吗？

20世纪60年代，朗兰兹^[36]提出了一个宏伟的计划，被称为朗兰兹纲领(见附录)，其主要思想是认为数学领域的许多猜想组成一个错综复杂的网络。这是一个伟大的想法！怀尔斯就是通过攻克 TSW(谷山丰-志村五郎-韦伊)猜想最后推出了费马大定理。

我们以千年数论难题之一的同余数问题来讲述有关故事。费马不定方程 $x^2 + y^2 = z^2$ 显然有整数解。我国古代周髀算经中高高定理就有“勾三股四弦五”之说，即(3, 4, 5)是一组整数解，国外也称为“毕达哥拉斯三元组”。边为(3, 4, 5)的直角三角形有面积6，推而广之，若一个自然数为某个由有理数边组成的直角三角形的面积，则称为同余数。所谓同余数问题是说：寻求比较简单的判别法则来决定一个自然数 n 是否为一个同余数。

利用计算机计算不定方程 $x^2 + y^2 = z^2$ 的整数解并不难。求方程 $x^2 + y^2 = z^2$ 的整数解，得 $x = u^2 - v^2$ ； $y = 2uv$ ； $z = u^2 + v^2$ ， $u > v > 0$ 。取 $u = 2, \dots, 6$ ； $v = 1, 2$ ，可算出：(3, 4, 5)；(8, 6, 10)；(15, 8, 17)；(24, 10, 26)；(35, 12, 37)；(5, 12, 13)；(12, 16, 20)；(21, 20, 29)；(32, 24, 40)，再算面积可得同余数。现在我们把上述解中出现的自然数看成节点，每组解看成一条超边(因现在一条边有三个节点)，则可以画出超网络图形如图8a。

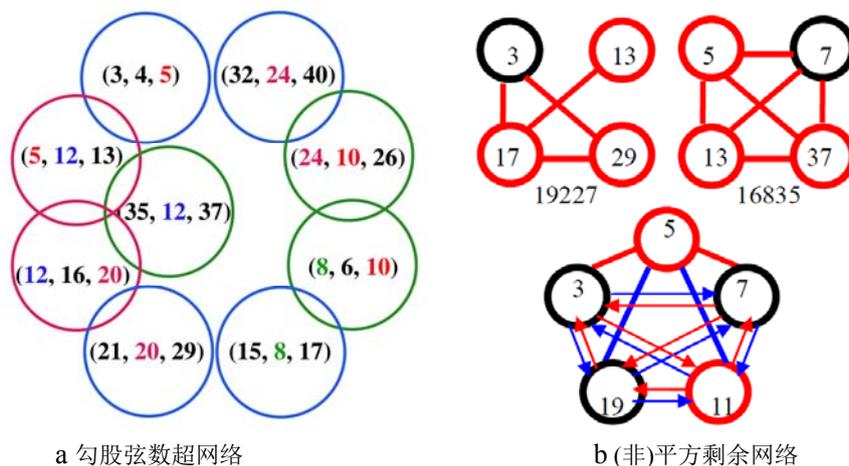


图 8 与同余数有关的两个网络

为了研究同余数，数学家还构造了一条特殊的椭圆曲线。从代数上讲， n 为一个同余数的条件是下面方程组及其单价形式有非零有理数解：

$$\begin{cases} x^2 + y^2 = z^2 \\ xy/2 = n \end{cases} \Leftrightarrow (x \pm y)^2 = z^2 \pm 4n \Leftrightarrow \left(\frac{x^2 - y^2}{4} \right)^2 = \left(\frac{z}{2} \right)^4 - n^2 \Leftrightarrow u^2 v^2 = u^6 - n^2 u^2$$

令 $x = u^2 = (z/2)^2$ ， $y = uv = (x^2 - y^2)z/8$ ，则得椭圆曲线 $E_n: y^2 = x^3 - n^2 x$ 。可以证明 n 为同余数则存在非零有理数点，而全体有理数点定义加法形成阿贝尔群。1920 年代，莫德尔^[37]提出一个重要定理：椭圆曲线上每个有理数点可表示为： $P_0 + n_1 P_1 + \dots + n_r P_r$ ， $r \geq 0$ ，其中 P_0 属于有限阶点生成的子群；而 r 称为椭圆曲线的秩。由此可证明 n 为同余数的充分必要条件是 $r > 0$ 。也可考虑椭圆曲线 $y^2 = x^3 - n^2 x \pmod{p}$ 在有限域 F_p 上的加法点群，其中 p 是除 n 的素因子外的任意素数。这个加法点群是有限阿贝尔群，阶为 N_p 。

这些阿贝尔群能为研究同余数提供什么有用的信息呢？贝赫和斯维讷通-戴尔^[38]考虑了如下的狄利克雷函数： $L(E, s) = \prod_p \frac{1}{1 - a_p p^{-s} + p^{1-2s}}$ ，其中 $a_p = 1 + p - N_p$ ， E 表示一般椭圆曲线。并且提出了著名 BSD 猜想： $L(E, s)$ 能解析开拓到整个复平面，并且满足函数方程 $L(E, s) = f(s)L(E, 2-s)$ ；函数值 $L(E, 1) = 0$ 的阶等于椭圆曲线的秩。在 BSD 猜想成立下，可以证明 $n \equiv 5, 6, 7 \pmod{8}$ 都是同余数，这是很了不起的结果。

如何判断一个自然数是非同余数呢？冯克勤^[39]引入非平方剩余网络，见图 8 b 中红色线条组成的网络(若 q 不是 p 的平方剩余，则 $p \rightarrow q$ 连线)，得到了一系列重要结果。例如，对无平方因子自然数 $n = p_1 \dots p_k \equiv 3 \pmod{8}$ ，其中 $p_1 \equiv 3 \pmod{4}$ ， $p_i \equiv 1 \pmod{4}$ 的非平方剩余网络，若网络奇性(验证 $r = \text{rank}_{F_2} L(G) = k - 1$)，则 $n = 19227$ 为非同余数，而 $n = 16835$ 是同余数。更奇妙的是任何一个无向网络都能够对应于一个这样的非平方剩余网络。

如果最终能够证明任何网络(无向、有向、二分、超网络、网络的网络等)都可与某个自然数(特别是素数)网络对应，就可以充分利用数论知识推进网络科学的发展。又若网络科学能为破解某些数论难题做出贡献，也必将诞生一个网络数论新方向。

参考文献

- [1]Erdős, P., Rényi, A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17 (1960).
- [2]Watts, D. J., Strogatz, S. H. Collective dynamics of small-world networks. *Nature* **393**, 167 (1998).
- [3]Barabasi, A. L., Albert, R. Emergence of scaling in random networks. *Science* **286**, 509 (1999).
- [4]Kleinberg, J. M. Navigation in a small world. *Nature* **406**, 845 (2000).
- [5]Shi, D. H. Critical thinking of scale-free networks: similarities and differences in power-law random graphs. *National Science Review* **1**, 337 (2014).
- [6]Shi, D. H., Chen, Q. H., Liu, L. M. Markov chain-based numerical method for degree distributions of growing networks. *Physical Review E* **71**, 036140 (2005).
- [7]Shi, D. H., Zhou, H. J., Liu, L. M. Markovian iterative method for degree distributions of growing networks. *Physical Review E* **82**, 031105 (2010).
- [8]Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**, 113 (1972).
- [9]Brin, S, Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Sys.* **30**, 107 (1998).
- [10]Lü, L. Y., Zhang, Y. C., Yeung, C. H., Zhou, T. Leaders in social networks, the delicious case. *PLoS One* **6**, e21202 (2011).
- [11]Liu, Y. T., Ma, Z. M., et al. BrowseRank: letting web users vote for page importance. *Internat. Acm Sigr Conference on Research Development in Information Retrieval*, 451 (2008).
- [12]Wang, X. F., Chen, G. R. Synchronization in scale-free dynamical networks: Robustness and fragility. *IEEE Trans. Circ. Syst.* **49**, 54 (2002).
- [13]Barahona, M., Pecora, L. M. Synchronization in small-world systems. *Phys. Rev. Lett.* **89**, 054101 (2002).
- [14]Wang, X. F., Chen, G. R. Pinning control of scale-free dynamical networks. *Physica A* **310**, 521 (2002).
- [15]Zhou, j., Lu, J. A., Lü, J. H. Pinning adaptive synchronization of a general complex dynamical network. *Automatica* **44**, 996 (2008).
- [16]Lu, W. L., Li, X., Rong, Z. H. Global stabilization of complex networks with digraph topologies via a local pinning algorithm. *Automatica* **46**, 116 (2010).
- [17]Liu, Y. Y., Slotine, J. J., Barabasi, A. L. Controllability of complex networks. *Nature* **473**, 167 (2011).
- [18]Yuan, Z. Z., Zhao, C., Di, Z. R., Wang, W. X., Lai Y. C. Exact controllability of complex networks. *Nature Communications* **4**, 2447 (2013).
- [19]Donetti, L. Hurtado, P. I., Muñoz, M. A. Entangled networks, super-homogeneity, and the optimal network topology. *Phys. Rev. Lett.* **95**, 188701 (2005).
- [20]Xuan, Q., Li, Y. J., Wu, T. J. Optimal symmetric networks in terms of minimizing average shortest path length and their sub-optimal growth model. *Physica A*, **388**, 1257 (2009).
- [21]Shi, D. H., Chen, G. R., Thong, W. W. K., Yan, X. Y. Searching for optimal network topology with best possible synchronizability. *IEEE Circ. and Syst. magazine* **10**, 66 (2013).
- [22]Yan, X. Y., Wang, W. X., Chen, G. R., Shi, D. H. Multiplex congruence network of natural numbers. *arXiv*: 1512.01615 (2015).
- [23]Nowak, M. A., Highfield, R. Super Cooperators: Altruism, Evolution, and Why We Need Each Other to Succeed. **Free Press**, New York, 2011.
- [24]Press, W., Dyson F. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences* **109**, 10409 (2012).

- [25]Szolnoki, A., Perc, M. Evolution of extortion in structured populations. *Physical Review E* **89**, 022804 (2014).
- [26]Rong, Z. H., Wu, Z. X., Hao, D., Chen, Michael Z. Q., Zhou, T. Diversity of time scale promotes the maintenance of extortionists in spatial prisoner's dilemma game. *New Journal of Physics* **17**, 033032 (2015).
- [27]Szabó, G., Fath, G. Evolutionary games on graphs. *Physical Reports* **446**, 97 (2007).
- [28]Kitsak, M., et al. Identification of influential spreaders in complex networks. *Nature Physics* **6**, 888 (2010).
- [29]Lü, L. Y., Zhou, T., Zhang, Q. M., Stanley, H. E. The H-index of a network node and its relation to degree and coreness. *Nature Communications* **7**, 10168 (2016).
- [30]Lü, L. Y., Zhou, T. Link prediction in complex networks: A Survey. *Physica A* **390**, 1150 (2011).
- [31]Lü, L. Y., Medo, M., Yeung, C. H., Zhang, Y. C., Zhang, Z. K., Zhou, T. Recommender systems. *Physics Reports* **519**, 1 (2012).
- [32]Lü, L. Y., Pan, L., Zhou, T., Zhang, Y. C., Stanley, H. E. Toward link predictability of complex networks. *Proceedings of the National Academy of Sciences* **112**, 2325 (2015).
- [33]Han, X., Shen, Z. S., Wang, W. X., Di, Z. R. Robust reconstruction of complex networks from sparse data. *Phys. Rev. Lett.* **114**, 028701 (2015).
- [34]Candès, E. J., Romberg, J., Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* **52**, 489 (2006).
- [35]Shi, D. H. Reconstructing the family tree with blood relationship. *Modern Physics* **159**, 4 (2015).
- [36]Langlands, R. Letter to Prof. Weil. 1967.
- [37]Mordell, L. J. On the rational solutions of the indeterminate equation of the third and fourth degrees. *Proc. Cambridge Philos. Soc.* **21**, 179 (1922).
- [38]Birch, B. J., Swinnerton-Dyer, H. P. F. Notes on elliptic curves I and II. *Reine Angew. Math.* **212**; **218**, 7 (1963); **218** (1965)..
- [39]Feng, K. Q. Non-Congruent Numbers and Elliptic Curves with Rank Zero. *University of Science & Technology China press*, 2008.

致谢：互联网+博弈论部分的素材由荣智海提供。陈关荣给予了许多宝贵的修改建议。

附录：罗伯特·朗兰兹发展了一项雄心勃勃的革命性理论，将数学中的两大分支数论和群论之间建立了新的联系。通过一系列的推测和分析，发现了涉及与整数有关的公式有不可思议的对称性。朗兰兹纲领可以追溯到数论中最深刻结果之一的高斯二次互反律，即“ q 是否 p 的平方剩余？”与“ p 是否为 q 的平方剩余？”之间的关系。朗兰兹纲领的一个最初动机，就是要对更一般情形的互反律提供完全的理解。如阿廷互反律，它建立了伽罗瓦群和狄利克雷 L -函数之间的联系。朗兰兹纲领是继克莱因的爱尔兰根纲领和希尔伯特纲领之后的第三个纲领，已经成为现今核心数学的风向标。

The Internet plus in network science——ideals, overviews and perspectives

Shi Dinghua

Dept of Math, Shanghai University, Shanghai 200444

Abstract This paper reviews some advances in network science from the perspective of the Internet plus. In retrospect, the first wave of network science studies, Internet plus big data, was set off by foreign scholars in the eve of the new millennium. Next, the second wave of network science studies, Internet plus dynamics, was launched by Chinese scholars. We predict that the third wave of network science studies, Internet plus game theory and pure math, may be about to come. We hope that the young scholars in China will be able to make greater contributions to the future network science.

Key words: network science, Internet plus, big data, dynamics, game theory, pure math