

基于谱约束的随机化社会网络多点扰动方法

强小强, 何小卫, 韩建民, 李 静

(浙江师范大学数理与信息工程学院, 浙江 金华 321004)

摘 要: 现有基于谱约束的随机化社会网络扰动方法只采用 4 个点的扰动, 扰动后社会网络的隐私保护程度不强。为此, 基于邻接矩阵及无符号拉普拉斯矩阵, 提出一种多点扰动方法, 在随机化过程中将社会网络的谱半径控制在一定约束范围内, 能在保证扰动后社会网络可用性的同时提高其隐私保护程度。实验结果表明, 该方法可以更好地保护社会网络结构。

关键词: 社会网络; 匿名性; 谱半径; 邻接矩阵; 无符号拉普拉斯矩阵

Multi-point Disturbance Method of Randomization Society Network Based on Spectrum Constraint

JIANG Xiao-qiang, HE Xiao-wei, HAN Jian-min, LI Jing

(College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321004, China)

【Abstract】 There are only four points perturbation method for the randomization society network which based on spectrum constraint, the privacy protection degree is not well after perturbation social network. To solve the problem, this paper proposes a multi-point disturbance method of randomization society network. In random process, this method is based on the adjacency matrix and signless Laplace matrix. It controls the social network spectral radius in certain within constraints, and ensures the usability and improves the privacy protection degree of the social network. Disturbance algorithm and experimental result proves that this method can better protect the social network structure.

【Key words】 social network; anonymous; spectral radius; adjacency matrix; signless Laplace matrix

DOI: 10.3969/j.issn.1000-3428.2011.09.033

1 概述

社会网络的隐私保护问题越来越受到关注, 简单地匿名化节点信息已经不能保证社会网络的安全性。文献[1-3]提出随机化扰动社会网络的方法, 该方法通过扰动社会网络的边或节点的方法保护社会网络的隐私。典型方法有: (1)随机添加/删除边^[2-4]; (2)随机交换边^[2-3], 随机化社会网络方法扰动后网络可用性较差。

文献[1]提出基于谱约束的随机化社会网络方法, 该方法分为谱约束的随机添加/删除边方法和谱约束的随机交换边的方法。谱约束的随机添加/删除边方法的思想是: 在满足一定谱约束的情况下, 随机添加 2 条边, 再删除 2 条边, 重复 K 次。类似地, 谱约束随机交换边方法的思想是: 在满足一定谱约束的情况下, 随机选取 2 条边 (t, w) 、 (u, v) , 且满足不存在边 (t, u) 、 (v, w) , 则将边 (t, w) 、 (u, v) 交换为 (t, u) 、 (v, w) , 重复 K 次。基于谱约束的随机化社会网络方法随机化后社会网络的可用性较好, 但是隐私保护程度不高。该方法增加/删除边或交换边的方法固定, 本文增加了 3 个点边之间边的搅动与谱半径的关系, 从而能够更好地保护社会网络的隐私。

2 相关知识

2.1 符号说明

社会网络可用图 $G(V, E)$ 表示, 节点集 $V = \{1, 2, \dots, n\}$, 边集 $E = \{(i, j) | a_{ij} = 1\}$, 其中, $|E| = m$ 。

为描述方便, 本文引入以下符号: 矩阵 A 表示图 G 的邻接矩阵。当节点 i 与 j 有边时, $a_{ij} = a_{ji} = 1$, 否则 $a_{ij} = a_{ji} = 0$ 。

对角矩阵 D 表示图的度序列, 其中, d_{ii} 表示第 i 个节点的度数, 当 $i \neq j$ 时, $d_{ij} = 0$ 。

谱半径是图的一个重要性质, 反映了图的谱性质, 并与图的结构性质密切相关。

设 A 为图 G 的邻接矩阵, λ_i 为矩阵 A 的特征值, e_i 是对应于 λ_i 的特征向量, 且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 则有 $A = \sum_{i=1}^n \lambda_i e_i e_i^T$ 、 $A e_1 = \lambda_1 e_1$ 。集合 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 是图的谱。其中, λ_1 是矩阵的最大特征值, 是图的谱半径; e_1 为 λ_1 对应的主特征向量, 且 $e_1 = (x_1, x_2, \dots, x_n)^T$ 。

矩阵 $Q = D + A$ 表示图 G 的无符号拉普拉斯矩阵, q_i 是矩阵 Q 的特征值, 且 $q_1 \geq q_2 \geq \dots \geq q_n$, 集合 $\{q_1, q_2, \dots, q_n\}$ 是图的无符号拉普拉斯谱(也称拟拉普拉斯谱), 其中, q_1 是无符号拉普拉斯矩阵的最大特征值, $2 \min d_i \leq q_1 \leq 2 \max d_i$ ^[4-5]; d_i 表示节点 i 的度数; μ_1 是 q_1 对应的特征向量, 且有 $\mu_1 = (y_1, y_2, \dots, y_n)$ 。

2.2 图的结构特性

本文介绍 2 个图的特性:

(1)图的调和平均最短距离。图的调和平均最短距离反映的是图中任意 2 个节点之间最短距离的调和平均值。定义^[1]为:

作者简介: 强小强(1985—), 男, 硕士研究生, 主研方向: 数据挖掘, 信息安全; 何小卫(通讯作者)、韩建民, 副教授; 李 静, 硕士研究生

收稿日期: 2010-10-21 **E-mail:** jhxxw@zjnu.cn

$$h = \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \right\}^{-1}$$

其中, d_{ij} 表示节点 i 与节点 j 的最短路径。

(2)图的传递性系数。图的传递性系数(这是一种聚类系数^[6])指少量节点循环的概率, 在大的社会网络中趋于 0, 而这些小的循环更能反映出社会网络的真实结构。定义为:

$C = 3N_{\Delta}/N_3$ ^[1], 其中, N_{Δ} 表示图中三角形的数目; N_3 表示图中连通三元组的数量。

2.3 基于图谱半径大小的扰动

谱扰动主要是通过改变特征值修改图, 如删除或增加一条边。 \bar{A} 表示扰动后图的邻接矩阵, $\bar{\lambda}_1$ 表示 \bar{A} 的最大特征值, 且 $\bar{\lambda}_1 \geq \lambda_1$ 。为了更好地描述图谱的扰动分析, 在矩阵的扰动分析中给出以下定理。

定理 1 设 $\lambda_1, \bar{\lambda}_1$ 分别是图 G 和图 \bar{G} 的最大特征值, 从 G 中删除一条边变为 \bar{G} , 则有 $\bar{\lambda}_1 \leq \lambda_1$; 同理从 G 中增加一条边变为 \bar{G} , 有 $\bar{\lambda}_1 \geq \lambda_1$ ^[5]。

本文用到盖尔圆盘定律及其推论并给出图扰动后特征值的变化范围。

3 谱约束随机化社会网络

3.1 社会网络的扰动分析

一般在社会网络中, 节点代表个体, 边表示 2 个体之间的关系。这样社会网络可以用一个图表示, 社会网络扰动问题可转化为图的扰动问题。随着扰动程度 k 的增加, 谱约束的随机化社会网络图的结构性质会不断变化, 最后逐渐趋于稳定。

下面给出带有谱约束的无符号拉普拉斯 Spectrum Switch 算法的约束关系式。

定理 2 设 q_1, \bar{q}_1 分别为图 G, \bar{G} 的无符号拉普拉斯矩阵的最大特征值, $\lambda_1, \bar{\lambda}_1$ 分别是图 G 和图 \bar{G} 的最大特征值, μ_1 为对应于 q_1 的特征向量, ε_1 为对应于 λ_1 的特征向量, 将图 G 的 2 条边 $(t, w), (u, v)$, 交换为 $(t, u), (v, w)$, $\mu_1 = (y_1, y_2, \dots, y_n)^T$, $\varepsilon_1 = (x_1, x_2, \dots, x_n)^T$ 。若有 $(y_t - y_u)(y_v - y_w) < 0$, 且:

$$q_1 - q_2 > \frac{(y_t - y_u)}{(y_w - y_v)} + \frac{(y_w - y_v)}{(y_t - y_u)}$$

则有 $\bar{q}_1 < q_1$ 。

证明: 令 $t=1, w=2, u=3, v=4$, 有扰动矩阵:

$$E = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix} \oplus O_{n-4}$$

若将矩阵对角化得到 $U^T E U = \text{diag}(2, -2, 0, 0, \dots, 0)$, 则很容易求出 E 的最大特征值 $\varepsilon_1 = 2$ 。本文构造 $\bar{E} = (\delta + 2)I - E$ 。其中, I 为单位矩阵。则有:

$$\bar{E}^{-1} = (2 + \delta)^{-1} + U \begin{pmatrix} \delta^{-1} - (2 + \delta)^{-1} \\ \\ \\ 0 \end{pmatrix} U^T$$

$$\delta(2 + \delta)(4 + \delta)\bar{E}^{-1} = \delta(4 + \delta)I +$$

$$\begin{pmatrix} 2 & 2 + \delta & -2 & -2 - \delta \\ 2 + \delta & 2 & -2 - \delta & -2 \\ -2 & -2 - \delta & 2 & 2 + \delta \\ -2 - \delta & -2 & 2 + \delta & 2 \end{pmatrix} \oplus O_{n-4}$$

又因为 $\bar{\lambda}_i = \lambda_i(-Q - (2 + \delta)I)$ ($i=1, 2, \dots, n$), 则有:

$$\bar{\lambda}_1 = -q_1 - 2 - \delta, \bar{\lambda}_2 = -q_2 - 2 - \delta$$

其中, Q 为无符号拉普拉斯矩阵。又有矩阵 $\bar{Q} + \bar{E}R$ 与矩阵 $\text{diag}(\bar{\lambda}_1 + \gamma, \bar{\lambda}_2, \bar{\lambda}_3, \dots, \bar{\lambda}_n)$, 且满足:

$$\gamma^{-1} = \gamma_1^{-1} = [v_1, v_2] = [\bar{E}^{-1}u_1, \bar{E}^{-1}u_1] = y^T \bar{E}^{-1}y$$

则有 $\gamma = q_1 - q_2 > 2 + \frac{a}{b}$, 其中:

$$a = (y_1 + y_2 - y_3 - y_4)^2, b = (y_1 - y_3)(y_2 - y_4)$$

又:

$$2 + \frac{a}{b} = \frac{(y_1 - y_3)}{(y_4 - y_2)} + \frac{(y_4 - y_2)}{(y_1 - y_3)}$$

则有:

$$q_1 - q_2 > \frac{(y_1 - y_3)}{(y_4 - y_2)} + \frac{(y_4 - y_2)}{(y_1 - y_3)}$$

所以:

$$\bar{q}_1 < q_1$$

证毕。

定理 3 设 q_1, \bar{q}_1 分别为图 G, \bar{G} 的无符号拉普拉斯矩阵的最大特征值, $\lambda_1, \bar{\lambda}_1$ 是图 G 和图 \bar{G} 的最大特征值, μ_1 为对应于 q_1 的特征向量, ε_1 为对应于 λ_1 的特征向量, $\mu_1 = (y_1, y_2, \dots, y_n)^T$, $\varepsilon_1 = (x_1, x_2, \dots, x_n)^T$, 图 G 围绕节点 r 将边 (r, s) 扰动到 (r, t) , 则得出以下结论:

(1)若 $x_t < x_s$, 则 $\bar{\lambda}_1 < \lambda_1$;

(2)若 $y_t < y_s$, 则 $\bar{q}_1 < q_1$ 。

证明过程与文献[7]证明 $\bar{\lambda}_1 > \lambda_1$ 和 $\bar{q}_1 > q_1$ 相似, 证明略。

无符号拉普拉斯 Spectrum Switch 方法扰动用到文献[1,4,7]中给出的约束关系式:

(1)若 $(x_t - x_u)(x_v - x_w) > 0$, 则 $\bar{\lambda}_1 > \lambda_1$;

(2)若 $(x_t - x_u)(x_v - x_w) < 0$, 且:

$$\lambda_1 - \lambda_2 > \frac{(x_t - x_u)}{(x_w - x_v)} + \frac{(x_w - x_v)}{(x_t - x_u)}$$

则:

$$\bar{\lambda}_1 < \lambda_1$$

(3)若 $(y_t - y_u)(y_v - y_w) > 0$, 则 $\bar{q}_1 > q_1$;

(4)若 $x_s > x_t$, 则 $\bar{\lambda}_1 > \lambda_1$;

(5)若 $y_t > y_s$, 则 $\bar{q}_1 > q_1$ 。

同理利用无符号拉普拉斯 Spectrum Add/Del 方法扰动图可得出类似的结论。

3.2 基于谱半径及无符号拉普拉斯谱的扰动算法

该算法思想: 交换(删除/增加)带有谱及无符号拉普拉斯谱约束的社会网络的 2 条边或旋转某条边, 随机化后社会网络的谱半径变化趋势是确定的, 本文基于这一结论将随机化后社会网络的谱半径控制在一定的约束范围内。当随机化后社会网络的谱半径大于它的上界值时, 本文给出带有谱半径和无符号拉普拉斯谱约束的随机化过程, 使下次扰动后随机化社会网络的谱半径减小; 当随机化后社会网络的谱半径小于它的下界值时, 同样给出带有约束的随机化过程, 使得下次扰动后随机化社会网络的谱半径增大; 当随机化后社会网络的谱半径在它的约束范围之内时, 下次扰动则不需要加任何约束随机化边。

下面给出无符号拉普拉斯 Spectrum Switch 保护社会网络的算法:

输入 图的邻接矩阵 A 、度序列矩阵 D 和扰动程度 K

输出 随机化后的图

步骤:

(1) 计算出矩阵 A 的特征值与特征向量 $(\lambda_1, \lambda_2, e_1)$, 矩阵

$Q = D + A$ 的特征值与特征向量 (q_1, q_2, μ_1)

(2) While($i < K$)

{1) 从图 G 中随机选出一条边 (t, w)

2) if ($i \% 4 = 0$ && $|\hat{\lambda}_1 - \lambda_1| \geq \|E\|$)

{① 找出所有满足 $\hat{\lambda}_1 > \lambda_1$ 且 $\hat{q}_1 > q_1$ 的边

② 从满足①的条件中随机选出一条边 (u, v) , 则将边 (t, w) 与 (u, v) 换为 (t, u) 与 (v, w)

}

3) else if ($i \% 4 = 1$ && $|\hat{\lambda}_1 - \lambda_1| \leq \|E\|$)

{① 找出所有满足 $\hat{\lambda}_1 < \lambda_1$ 且 $\hat{q}_1 < q_1$ 的边

② 从满足①的条件中随机选出一条边 (u, v) , 则将边 (t, w) 与 (u, v) 换为 (t, u) 与 (v, w)

}

4) else if ($i \% 4 = 2$ && $|\hat{\lambda}_1 - \lambda_1| \geq \|E\|$)

{① 找出所有满足 $\hat{\lambda}_1 > \lambda_1$ 且 $\hat{q}_1 > q_1$ 的点

② 从满足①的条件中随机选出一边 s 则将边 (t, w) 换为 (s, w)

}

5) else if ($i \% 4 = 3$ && $|\hat{\lambda}_1 - \lambda_1| \leq \|E\|$)

{① 找出所有满足 $\hat{\lambda}_1 < \lambda_1$ 且 $\hat{q}_1 < q_1$ 的点

② 从满足①的条件中随机选出一边 s 则将边 (t, w) 换为 (s, w)

}

6) else 随机交换 (t, w) 、 (u, v) 为 (t, u) 、 (v, w)

7) $k++$

}

算法每次扰动后都要计算矩阵的最大特征值 $\hat{\lambda}_1$ 。因为计算矩阵特征值的算法复杂度为 $O(n^2)$, 扰动程度为 K 和每次扰动后计算满足条件的边的复杂度为 $O(m)$, 所以算法复杂度为 $O(Kn^2)$, 其中, m 为图中边的条数。

当扰动程度 K 给定为常数时, 复杂度为 $O(n^2)$ 。实际运行时优化算法可以每隔几次扰动计算一次矩阵的最大特征值 $\hat{\lambda}_1$ 。

3.3 隐私保护分析

根据图 1 结合算法分析当节点集 $\{t, u, v, w\}$ 中每个节点相连错误的边数都大于 1 时, 边 (u, v) 存在的可能性是所有与节点 t 、节点 w 、节点 v 和节点 u 相连的错误边数目的积的倒数, 攻击边 (u, v) 成功的概率等于 $p = 1/c_t c_w c_u c_v$ (c_i 是与 i 节点关联的错误边的个数)。当节点集中与某些节点相连错误边的数目为 0 时, 攻击边 (u, v) 成功的概率为:

$$p = \min \{1/c_u, 1/c_v, 1/c_w, 1/c_t\}$$

当旋转和交换的次数比较大时, 任何一条边错了扰动后的图攻击 (u, v) 都有可能不成功。同理攻击边 (t, u) 成功的概率为:

$$p = \min \{1/c_u, 1/c_v, 1/c_w, 1/c_t, 1/c_u c_w, 1/c_v c_t\}$$

显然比一般的方法要好的多, 在文献[1]中, 在边 (t, u) 假设存在的前提下, 攻击边 (u, v) 成功的概率等于 $p = 1/c_t c_w$ 。说明本文所用的无符号拉普拉斯 Spectrum Switch 方法隐私保护程度要比 Spectrum Switch 方法强。



图 1 扰动分析

4 实验结果与可用性分析

4.1 实验数据来源描述

本文使用的实验数据是社会网络分析经常用到的经典的美国政治书籍数据^[1-4], 它包含 105 个节点, 441 条边。

4.2 实验结果分析

本文用 Rand Switch 方法和无符号 Spectrum Switch 方法(记为 SSpectrum Switch)进行图扰动分析。调和平均最短距离和传递系数的变化情况如图 2(限于篇幅, 选取部分实验数据分析)所示。

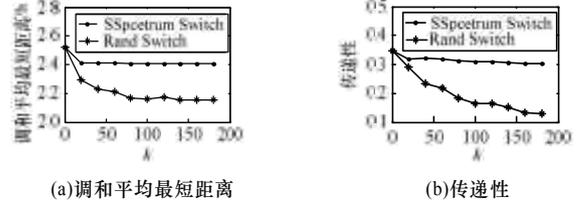


图 2 图的特性随扰动程度的变化 1

从图 2 可以看出, SSpectrum Switch 扰动方法下变化幅度比 Rand Switch 方法下变化幅度小。图 2(a)和图 2(b)随着扰动程度 k 的变化是相似的, 当 k 增加到 100 后, h 和 c 的变化趋于平缓。而且在 SSpectrum Switch 扰动方法下 h 和 c 变化范围远小于 Rand Switch 扰动方法下的变化范围, 说明 SSpectrum Switch 方法比 Random Switch 方法能更好地保护社会网络的结构。

图 3 从传递系数和调和平均最短距离 2 个角度比较了 SSpectrum Switch 方法与 Spectrum Switch 方法的优劣。

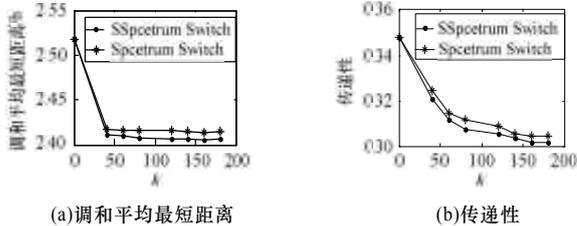


图 3 图的特性随扰动程度的变化 2

从图 3 可以看出, SSpectrum Switch 和 Spectrum Switch 扰动方法随着扰动程度 k 的变化 h 、 c 基本相似, 当 k 增加到 80 后, h 和 c 的变化趋于平缓。这说明 SSpectrum Switch 方法与 Spectrum Switch 方法随机化后社会网络的结构性质相似, 即可用性相近。这是因为拉普拉斯矩阵与无符号拉普拉斯矩阵都能很好地反映图的性质。

5 结束语

本文通过无符号拉普拉斯矩阵对图谱保护随机化社会网络进行改进, 并通过实验验证了本文方法的有效性。下一步(下转第 103 页)