

【互联网政治与法律】

个人信息的边界、敏感度与中心度研究

——基于专家和公众认知的数据分析

吴标兵,许和隆

(南京邮电大学 马克思主义学院,江苏 南京 210023)

摘要:个人信息立法的首要问题是确定个人信息的边界。因为信息与通信技术的进化、大数据聚合分析、隐私悖论、敏感度的个体体验等因素,一般的个人信息和敏感的个人信息的二分法存在不周延性。专业技术人员通过对一般的个人信息进行二次数据分析或自动化处理,可以获得“推测数据”甚至是敏感的个人信息的,因而存在一般的个人信息是潜在的敏感个人信息这种情形。实务界将这些“非敏感”的、有价值的个人信息确定为“核心数据”。研究表明:公众对性取向、固定电话通话记录等敏感度相对较低,对与个人密切相关的身份、生活、工作和政务信息敏感度相对较高。敏感的个人信息和核心的个人信息具有重叠性,核心的个人信息包括部分敏感的个人信息和部分一般的个人信息。

关键词:边界;个人信息;敏感度;中心度

中图分类号:C912 **文章编号:**1673-5420(2018)05-0044-10

人类自产生之时起,就开始试图利用语言等符号工具拟制另一个世界:网络虚拟社会^[1]。互联网技术进一步提升了网络虚拟社会的边界和链接效率。随着物联网社会的到来,现实世界泛在地、全息地转译为数字世界,再通过对数据的加工、赋义行为,创造虚拟现实的信息世界。现实世界与信息世界既有真实的映射关系,也有虚拟的构建关系和实时的互动关系。其精妙之处在于人们假借对信息悄无声息的观察、组织和控制,在幕后神不知鬼不觉地完成现实世界的权力再生产和资源再配置。

自20世纪40年代开始,源起于现代科学技术(特别是信息和通信技术)的双面刃效应,世界各国和地区开始制定个人信息保护的相关法律。以人为本的个人信息保护法的制定是国家和地区法治文明的标志之一。2003年周汉华受国务院信息管理办公室委托起草个人信息保护法(专

收稿日期:2018-05-25 本刊网址:<http://nysk.njupt.edu.cn>

作者简介:吴标兵,副教授,研究方向:智慧城市治理、ICT与产业创新、ICT与隐私安全。
许和隆,教授,研究方向:政治学。

基金项目:国家社会科学基金项目“网络社会公众维权与社会维稳均衡治理研究”(15BZZ089);江苏省高校哲学社会科学重点研究基地项目“智慧城市建设中物联网技术选择模式及发展路径研究”(JDS215002);南京邮电大学引进人才科研启动基金“基于ICT进化的中国特色科技服务业创新规律研究”(NYY215002);江苏高校哲学社会科学重点研究重大项目“大数据时代公民权益保障与社会稳定治理”(2015ZDIXM018)

家建议稿),并于2005年完成^[2]。同年,齐爱民公开发表个人信息保护法专家建议稿^[3]。继2014年吴晓灵在两会上提交关于加快制定《中华人民共和国个人信息保护法》的立法建议之后,杨震在2016年两会上建议启动个人信息保护立法。然而,个人信息保护法始终没有被正式列入立法日程。

个人信息立法的首要问题是个人信息的边界。杨震指出,立法的障碍在于如何界定个人信息。他认为个人信息的概念非常广,什么样的信息需要保护?界限在哪儿?没有定论^[4]。个人信息的界定需要考虑到不同的文化、社会习俗和时代特点。法律不能脱离民情,只有将法律关系和民情结合到一起,才能真正达到我们法律的目的。个人信息的明确化和类型化是国际法律规范确定个人信息边界的通行方法,也是我国未来个人信息立法的可行路径。个人信息类型化,首先需要区分一般的个人信息和敏感的个人信息的。鉴于信息和通信技术的发展等因素,上述区分方式存在不周延性。

一、个人信息类型化问题与争议

以“可识别性”为依据对个人信息进行界定,并将敏感的个人信息公开具体化,是国际立法的趋势。如由于地域文化差异,匈牙利将国籍规定为敏感的个人信息的,意大利与荷兰将商会身份规定为敏感的个人信息的。即使文化传统相似的欧盟成员国之间也存在差异,一些欧盟国家将照片视为敏感数据,而欧盟数据保护工作小组并不将网络照片归为敏感的个人信息的^[5]。同一国家或地区,对敏感的个人信息的界定也不完全一致,例如美国《金融隐私法》和联邦贸易委员会将财务信息列为敏感的个人信息的,但是HIPAA法案却并未将该信息列为敏感的个人信息的。对于刑事犯罪等有关司法纪录,英国、爱尔兰均将其看作敏感信息。美国则允许将民事诉讼、民事裁定和逮捕记录作为消费者报告的内容,但要求犯罪等判决和未判决的公共纪录完整、准确。德国则将行政违法记录也视为敏感信息。

在我国个人信息立法中是否区分一般的个人信息和敏感的个人信息的,学界存在分歧。张新宝等人主张采用敏感的个人信息的概念,以调和个人信息保护与利用的需求冲突,实现利益平衡^[6]。周汉华等人则认为在个人信息保护法中不宜采用敏感的个人信息的概念,原因在于“敏感的个人信息的所包含的范围非常广泛,但其中很多不适合我国国情”,“会导致个人信息保护法与我国宪法及根本政治制度的冲突;如果另起炉灶,采用含义狭窄的敏感信息概念,还可能在国际人权对话中授人以柄、陷于被动”^{[2]79}。

根据领域理论,个人生活领域在保护程度上存在等级体系。个人信息的保护一般采取信息自主权模式和隐私权模式。在隐私权模式下,个人信息分为隐私性个人信息和非隐私性个人信息。敏感的个人信息的即是隐私性个人信息。隐私性为个人信息保护的前提条件。个人信息保护的力度取决于隐私的边界。我国主流学说将隐私的边界扩展至“生活安宁不受他人非法干扰”,具体化为“私人信息、私人生活和私人领域”。尽管如此,因隐私权保护模式存在缺陷,有学者建议将某些个人一般信息纳入保护范围^[7]。在司法实践中,上海市浦东新区人民法院将个人的姓名、性别、职业、学历、联系方式、家庭住址、婚姻状况等与个人及其家庭密切相关的信息纳入隐私的范围。

敏感的个人信息和一般的个人信息的界限日益模糊,隐私和个人信息呈现同质性特征^[1]。就共性而言,欧盟将敏感的个人信息举报为以下几种:种族或民族起源;宗教或哲学信仰;政治权利、政治观点;工会组织成员;刑事、行政违法纪录;个人基因识别数据、生物数据,涉及健康、性生活或性取向的数据等。

如何科学地对个人信息进行类型化?这是本文要研究的问题。通过头脑风暴法、德尔菲法及对大样本数据的挖掘,对个人信息的边界、敏感度和中心度进行实证研究:其一,明确个人信息分类;其二,将个人信息类型化为一般的个人信息、敏感的个人信息和核心的个人信息。

二、数据来源与方法

(一)头脑风暴法

由于政治、经济、历史、文化背景等方面的差异,世界各国或地区对个人信息的规定不一。随着信息和通信技术的发展,个人信息的外延不断地扩大。为了对个人信息有更为明确、科学的划界,本研究采用头脑风暴法,激发专家的创造性,以达到预期效果。

选取5名计算机专业的专家和5名人文社科类专业的专家,专家职称为副教授及以上,以我国台湾地区“个人资料类别”为蓝本,采用网络会议讨论个人信息分类。会议前后进行3次,每隔1个星期1次,每次45分钟。最终,以讨论形成的个人信息分类方案来设计调查问卷。

(二)问卷调查法和德尔菲法

1. 问卷设计

调查问卷包括个人信息项和4项非敏感的人口统计学信息(性别、年龄、学历、婚姻状况)。考虑到问卷的体量,采用常用的李克特5级量表形式,要求被调查者对个人信息敏感度从非常不敏感到非常敏感进行评价,赋值分别为1~5分。问卷的个人信息项全部被随机打乱。为了便于被调查者对评价信息项的理解,对较为专业的信息项进行了必要的解释和说明。

在正式调查之前,选取不同性别、年龄、学历的公众对问卷进行预调查,根据反馈的意见修改问卷,对较为专业的个人信息给予扩充解释。调查方式为滚雪球非概率抽样,通过社交网络(电子邮件、微信、QQ等即时通信工具)发放电子样本,每一台设备和IP只限作答1次。问卷调查时间为2015年12月—2017年12月。一共发放3700份问卷,回收3199份,回收率为86.46%。剔除187份无效问卷,有效问卷率为81.41%。判断无效问卷的标准如下:首先是草率作答,用时少于200秒(预调查测试最低值);其次是恶意作答,在200秒内全部赋予同一值;最后是赋值逻辑存在问题,撒谎或欺瞒,相似信息选项赋值相差悬殊等。样本的区域构成方面,江苏省占比45.1%,其他地区占比54.9%。样本的人口特征方面,男性占比49.9%,女性占比50.1%;18~33岁占比76.1%,34~49岁占比21.8%,50岁及以上占比2.1%;大专以下学历(含高中)占比35.3%,本科占比57.4%,研究生占比7.2%;已婚占比42.9%,未婚占比55.6%,离异占比1.5%。

当柯龙巴赫系数 α 在0.8以上时,则判定该调查量表信度较好。总量表的柯龙巴赫系数 α 为0.972,说明量表具有较高的可靠性。KMO值为0.980,且Bartlett球体检验结果均小于0.01,说明本量表的结构效度较好。

2. 德尔菲法

鉴于公众对个人信息认知的非专业性所产生的偏差,问卷同时采用德尔菲法征询专家意见。17名专家对所有的个人信息按照敏感度从非常不敏感到非常敏感进行1~5分的赋值,并就可能遗漏的、有单独评价必要的个人信息予以列出和赋值。经过反复讨论,在权利资质中将“学历学位证书、房屋产权证和金融权证”单独予以赋值。最终得到108项个人信息列表。所有均值采取公众赋值50%和专家赋值50%进行加权(对于缺省值采用专家赋值)。

三、数据分析

(一)个人信息分类与编码

根据头脑风暴法和德尔菲法,将个人信息分为10类,共含108项。

第1类:身份信息。

个人识别:1 姓名;2 昵称/小名/曾用名;3 年龄;4 性别;5 出生日期;6 星座;7 籍贯;8 学历;9 职业;10 政治面貌;11 婚姻状况;12 宗教信仰;13 民族或种族;14 肖像(证件照或工作照);15 个人履历(学习、工作经历)等。

身份证件:16 出生证;17 户口簿;18 身份证、工作证、退休证、护照等证明身份的文书。

权利或资质文书:19 学历学位证书、驾驶证、行驶证;20 专业技术资格证、许可证或营业执照;21 房屋产权证;22 金融权证、专利权证书等以权利或资质证明身份的文书。

个人特征:23 DNA;24 指纹、掌纹、声纹、视网膜、3D 脸部特征等;25 身体特征(视力、身高、体重、三围、肩宽、腿长、脚长、血型等);26 个人特长;27 个性/性格;28 兴趣/爱好;29 个人笔迹等。

联络方式:30 手机号码;31 电子邮箱;32 QQ/微信号等网络虚拟账号;33 居住地址等。

第2类:家庭信息。

34 住家设施:住宅性质(所有还是承租)、类型、面积、装潢、设施、租金、租期;35 家庭经济状况:家庭收入等;36 家庭成员:配偶、父母、子女、兄弟姐妹等;37 家庭生活:私生活、家庭生活、家居生活信息;38 家庭住址;39 家庭电话号码等。

第3类:社会生活信息。

40 生活格调:消费品的种类、品质和模式及所使用的交通工具等;41 日常生活安排:生活方式/习惯等;42 社交活动记录:聚餐、聚会等社会活动的时间、地点和对象等信息;43 旅行及迁徙信息:车票、飞机票、酒店入住登记信息等;44 地理位置和行踪记录:GPS 定位信息、IP 信息、MAC 地址等;45 私人生活空间:人与人之间应有的距离、独处的空间;46 情感经历:女朋友、情人、情史等婚前恋爱和婚外感情生活;47 难堪的往事或不愉快的事情;48 亲戚/朋友信息:其社会地位、政治面貌和家庭住址等;49 社团或社会组织资格;50 个人生活照;51 个人生活视频和音频。

第4类:教育信息。

52 毕业院校:本科毕业院校等;53 学历/职称专业;54 修习的课程;55 考试成绩等学习记录;56 师生关系。

第5类:工作信息。

57 工作单位名称;58 工作单位地址;59 工作单位电话号码;60 领导/上司;61 同事和同事关系;62 职称;63 行政职务;64 培训记录;65 工作记录:在职记录、请假、未上班的理由等;66 工作经验:以前的工作、失业时间等;67 工作表现评价;68 离职经过:离职日期、原因等;69 选任、受聘、雇佣、委任经过。

第6类:健康信息。

70 身体健康状态;71 心理或精神状态;72 性取向;73 性生活记录;74 医疗记录;75 病史:患病的原因、症状等及历次所患疾病与治疗情况。

第7类:财务信息。

76 商业活动:签订的商业契约、提供或使用的财货或服务、商业种类等;77 财务交易:转账汇款、缴费、支付、个人理财(购买基金、账户金、理财产品、外汇买卖)等;78 银行卡账户或信用卡号码;79 个人负债与支出信息;80 保险信息:保险种类、范围、金额、账号等;81 个人收入、所得、资产与投资信息;82 贷款、抵押、质押、担保等信息;83 票据信用:支票存款、基本资料、退票资料、拒绝往来资料等;84 信息结汇记录;85 津贴、福利、赠款。

第8类:行政司法信息。

86 赔偿申请记录;87 行政或刑事处罚、警察备案记录;88 养老保险信息:生效日期、付出与收入金额、受益人;89 行政奖励记录;90 困难救助或社会救济申请记录;91 负有履行法院文书义务的相关信息;92 行政许可申请记录;93 信用评级;94 驾驶记录。

第9类:表达性信息。

95 政治观点、政治意见;96 非正式场合言论;97 私人文件:日记、记事本、电子邮件等;98 安全密码信息;99 固定电话通话记录;100 手机通讯录、通话记录;101 网络聊天、即时通讯记录;102 网络浏览记录;103 网络购物记录;104 网络下载记录;105 个人博客、网站等信息;106 网络搜索关键词;107 微博、QQ空间、朋友圈等社交网络信息;108 个人著作/出版物等知识信息。

第10类:其他个人信息。比如辐射剂量资料等。

(二)个人信息敏感度

1. 敏感度判断依据

Tosun认为,5级李克特量表得分均值在1~2.4之间表示反对,2.5~3.4之间表示中立,3.5~5之间表示赞同^[8]。公众对个人信息的敏感度均值为3.41,专家对个人信息的敏感度均值为3.65,加权评价值为3.53,单项个人信息最低值为2.4,说明公众的个人信息敏感度总体较高。本研究将加权平均值小于3.41列为一般的个人信息,介于3.41~3.53之间列为中等敏感的个人信息,大于3.53列为敏感的个人信息。

2. 敏感度分析

表1为个人信息敏感度的调查得分,我们发现:公众和专家对于个人信息敏感度的认知基本一致;生活格调、生活方式、社交活动、难堪的往事或不愉快的事情等生活信息敏感度居中,这些信息被认为不是高度隐秘。个人识别信息、个人网络信息敏感度普遍偏低,这与之前的研究结论不同^[9]。我国公众的性观念发生变化,性取向的敏感度下降。随着固定电话退出日常生活,固定电话通话记录也淡出人们关注的视线。与个人密切相关的学习(成绩)、生活(照片、视频/音频、

旅行、地理位置、空间)、工作(工作记录、选任/受聘经过)、联系方式(手机号码、居住地址)、身份信息、家庭信息、通讯记录、生物特征识别、财务信息、行政司法记录、私人文件和密码等个人信息的敏感度相对较高(见表1灰色字体部分)。

表1 个人信息敏感度调查得分

类别	信息项编号	公众	专家	类别	信息项编号	公众	专家	类别	信息项编号	公众	专家
个人识别	1	2.94	3.47	社会生活	37	3.78	4.00	财务信息	73	3.83	4.06
	2	3.04	3.12		38	3.62	4.24		74	3.45	3.71
	3	2.86	2.88		39	3.62	3.94		75	3.65	3.71
	4	2.57	2.59		40	3.17	3.47		76	3.62	4.12
	5	3.03	3.18		41	3.37	3.47		77	3.83	4.47
	6	2.40	2.53		42	3.34	3.47		78	4.19	4.47
	7	2.72	2.59		43	3.16	4.18		79	4.01	4.41
	8	3.13	3.18		44	3.62	3.59		80	3.67	4.00
	9	3.02	3.29		45	4.17	4.06		81	3.84	4.24
	10	2.91	2.94		46	3.55	3.94		82	3.82	4.29
	11	3.16	3.82		47	3.47	3.35		83	3.77	4.35
	12	2.57	2.94		48	3.58	3.65		84	3.75	3.88
	13	2.60	2.59		49	2.88	2.94		85	3.57	3.71
14	3.73	4.12	50	3.98	3.53	行政司法	86	3.65	4.35		
15	3.32	3.29	51	4.10	4.29		87	3.90	4.35		
身份证明	16	3.84	4.18	教育信息	52		2.94	3.47	88	3.76	4.00
	17	4.00	4.35		53		2.87	3.71	89	3.12	3.18
	18	4.10	4.82		54		2.90	3.35	90	3.35	3.29
权利资质	19	缺省	3.76	55	3.54		3.82	91	3.55	3.65	
	20	3.27	3.35	56	3.05	3.06	92	3.50	3.71		
个人特征	21	缺省	4.59	工作信息	57	3.10	3.47	93	3.87	4.35	
	22	缺省	4.00		58	3.07	3.24	94	3.59	4.24	
	23	3.81	3.65		59	3.16	3.35	表达性信息	95	3.13	3.53
	24	4.12	4.47		60	3.14	3.41		96	3.31	3.59
25	3.39	3.12	61	3.10	3.06	97	4.12		4.41		
26	2.90	3.12	62	3.05	3.47	98	4.35		4.88		
27	3.21	3.41	63	3.08	3.18	99	3.40		3.53		
联系方式	28	3.21	3.65	64	2.96	2.59	100	3.63	4.00		
	29	3.17	3.24	65	3.34	4.00	101	3.69	3.65		
	30	3.69	4.24	66	3.24	3.65	102	3.40	3.24		
	31	3.07	3.53	67	3.40	3.53	103	3.29	3.29		
	32	3.46	3.24	68	3.19	3.29	104	3.23	3.06		
	33	3.42	3.88	69	3.51	3.65	105	3.29	2.71		
	家庭信息	34	3.85	4.18	健康信息	70	3.60	3.88	106	3.09	3.41
35		3.59	4.12	71		3.46	3.82	107	3.56	2.94	
36		3.79	4.29	72		3.17	3.1	108	3.56	3.35	

3. 敏感度的不周延性

物联网、大数据时代,隐私已终结。数据挖掘能根据少量的一般个人信息识别个人的身份等敏感信息^[10]。人们甚至能够通过大数据聚合分析,利用一般的个人信息进行具体的个人画像^[11]。对于数据中间商来说,年龄、学历、兴趣、职业、购买记录等一般的个人信息却是“核心数据”^[12]。通过二次数据分析或自动化处理,人们即可获得“推测数据”甚至是敏感的个人敏感信息。一个只知道你的出生日期、邮政编码和性别的研究人员可以在87%的尝试中识别你的名字^{[13]21}。敏感度不高的个人信息并不表示这些个人信息不重要、不敏感。另外,还存在一种“隐私悖论”现象,它通常被解释为:公众对个人信息在态度上较为关注和重视,但实际行为却主动放弃个人信息。所以对于公众来说,个人信息敏感度具有个体体验特征,存在认知上的差异。敏感的个人敏感信息未必真的敏感。基于上述分析,可以得出命题:存在一般的个人信息是潜在的敏感个人信息这种情形。该命题说明“敏感的个人敏感信息”的不周延性。这也是部分学者反对“敏感的个人敏感信息”这一分类的重要原因。如何解决“敏感的个人敏感信息”不周延性这一问题?实务界将“非敏感”的、有价值的个人信息确定为“核心数据”,这为我们提供了新的思路和解决路径。有必要对个人敏感信息网络进行中心度分析,确定哪些个人信息是处于关键节点的核心个人信息。

(三) 个人敏感信息网络中心度

1. 网络节点重要性的评价方法

评价网络节点重要性排序的方法有“节点功能”和“位置属性”两种。“节点功能”方法通过删除节点,考察节点被删除后对网络的破坏性的大小。位置属性方法一般通过网络的拓扑结构来计算得到节点的重要程度,主要分为度(degree)、中间度(betweenness)、接近度(closeness)、K-Shell、PageRank法等。就不同的评价方法,学者之间颇有争议。例如,Daly等人认为中间度和接近度是信息传输最合适的指标^[14]。Holthoefter等人指出,在级联效应的模型下,K-shell指标能更好地识别网络中更有影响力的节点^[15]。陈静等人将接近度和关键度作为评定节点的重要度^[16]。武澎等人则将特征向量中心性的理论方法运用于社交网络信息交互过程中重要节点的评判研究^[17]。节点的重要性不仅由网络结构决定,还与网络的动力学、传播机制和节点本身有关。

2. 中心度分析

本文采用点度、中间度、接近度、特征向量综合法,对基于公众认知的个人敏感信息的网络中心度进行探索性研究。考虑到专家样本不足30份,本部分数据来源于公众问卷调查。首先用SPSS进行相关分析,构建个人敏感信息矩阵。基于大样本的因素,虽然绝大多数信息之间的相关系数、标准系数总体偏小,但是达到显著水平,具有统计学意义。根据主成分分析结果,得到21个主要成分。然后用Netdraw进行网络密度检测,分析各个节点的紧密程度。在个人敏感信息网络中,节点代表“个人敏感信息”项。一个网络的密度越高,越接近1,说明网络节点联系越密切,信息流通的速度越快、效率越高,节点之间的通道就越顺畅。通过网络密度分析,个人敏感信息的网络密度为0.2463。大规模网络的密度一般较小,实际网络中能够发现的最大的密度是0.5。这说明个人敏感信息网络连通度较好,适合做中心度分析。再分析个人敏感信息网络关系中心度(阈值大于0.3),提取排名前21位的核心个人敏感信息网络节点,制定中心度表格(见表2),绘制相应的网络图谱(节点

图标愈大,说明中心度越高)。在表2中,用表1的个人信息项编号表示节点。

表2 个人信息中心度排名前21位的节点

排名	点度		中间度		接近度		特征向量	
	节点	值	节点	值	节点	值	节点	值
1	42	56	58	266	43	565	108	0.165
2	15	55	59	211	28	463	44	0.164
3	59	55	40	209	55	446	92	0.163
4	108	55	15	204	41	426	99	0.161
5	99	54	69	200	25	408	14	0.160
6	92	53	27	197	40	371	100	0.159
7	14	52	85	182	60	368	34	0.158
8	34	52	42	156	72	348	91	0.155
9	100	52	46	139	86	356	42	0.154
10	91	51	9	133	47	346	81	0.153
11	46	51	62	123	6	346	36	0.150
12	58	51	63	122	69	344	48	0.149
13	36	47	57	121	87	342	80	0.147
14	48	47	71	112	53	340	39	0.145
15	88	47	66	107	54	338	23	0.145
16	75	47	28	102	45	336	75	0.144
17	35	46	88	93	3	329	35	0.142
18	39	46	1	87	52	328	38	0.142
19	38	46	65	86	98	326	46	0.141
20	23	46	34	84	4	324	107	0.140
21	82	45	20	82	12	319	38/101	0.139

点度中心度说明一个网络中与某一个点直接相连的其他点的个数。个数越多,说明与其他节点联系越紧密,该点在网络中越处于中心地位。从点度的测度结果来看,社交活动、个人简历、工作单位电话号码和个人著作/出版物等个人信息连接的节点最多,能够有效控制和影响其他个人信息,因而局部中心度较高,可以被看作是个人信息局部网络的中心点。点度中心度的缺陷是仅仅计量连结的数量,而没有考虑其“质量”和控制力。

中间中心度测量的是一个点在多大程度上位于其他点的中间,即中介控制力。从中间度测度结果来看,工作单位地址、工作单位电话号码、生活格调和个人履历等个人信息具有高度间接中心性,在很大程度上是其他节点的中介,起到经纪人的作用。因此,工作单位地址、工作单位电话号码等个人信息在资源占有和信息流通中,处于关键位置,对于个人信息网络具有重要影响。

接近度表明某节点到其他节点所有路径的最小值。如果一个节点较少地依赖于其他节点,其中心度越高。从接近度的测度结果来看,迁徙/旅行、兴趣/爱好、学习考试成绩、生活方式/习惯、身体特征等个人信息虽然在个人信息网络中不接近中心性,但越不处于核心地位,不受其他

节点的控制能力越强,在个人信息网络中具有较好的视野:能够察知个人信息的流通方向和个人信息网络发生的事情。

特征向量是刻画中心度与网络中心势的一种标准化测度。通过特征向量分析,可以在网络整体结构意义上找到核心成员。从特征向量的测度结果来看,个人著作/出版物、地理位置、行政许可申请、固定电话通话记录、肖像、手机通信记录、住宅设施等个人信息在网络中具有重要的影响力。

四、结论和建议

当前无论在立法实践还是理论研究层面,以敏感度作为区分标准的个人信息类型化存在不周延性。个人信息可以区分为一般的个人信息、敏感的个人信息和核心的个人信息。核心的个人信息(见表2)包括敏感的个人信息和一般的个人信息。

建议将个人信息分为10个大类、108项,以明晰个人信息的边界。其中,将出生证、户口簿、房屋产权证,金融权证、DNA、指纹、掌纹、视网膜3D脸部特征、手机号码、居住地址、家庭信息、旅行信息(车票、住宿等)、地理位置和行踪记录、情感经历、亲戚/朋友信息、学习考试成绩、工作记录、选任/受聘经过、个人生活照、个人生活视频和音频、性生活记录、医疗记录、病史、财务信息、行政司法信息、安全密码信息、手机通讯记录、网络即时通讯记录等列为敏感的个人敏感个人信息。对敏感的个人敏感个人信息应采取严格的事先同意原则。

建议将以下信息下列为非敏感的个人敏感个人信息:姓名、年龄、性别、星座、职业、宗教信仰、个人履历、专业技术资格证、身体特征、兴趣/爱好、个性/性格、生活格调、生活方式/习惯、社交活动记录、毕业院校、学历/职称专业、修习的课程、工作单位名称、工作单位地址、工作单位电话号码、领导/上司、职称、行政职务、工作经验、性取向、微博、QQ空间、朋友圈、个人著作/出版物等。对于上述信息,禁止未经允许的以识别个人或推测个人敏感个人信息为目的的二次数据处理。

本文旨在通过多种方法的交叉融合,减少问卷设计、数据采集和数据分析方面的不确定性。由于人力、物力、财力、智力(数据获取和分析方法)等因素的限制,这种不确定性依然存在。

参考文献:

- [1] 吴标兵. 信息哲学视域下的物联网隐私本质及其困境[J]. 理论月刊,2016(2):36-40.
- [2] 周汉华. 中华人民共和国个人信息保护法(专家建议稿)及立法研究报告[M]. 北京:法律出版社,2006.
- [3] 齐爱民. 中华人民共和国个人信息保护法示范法草案学者建议稿[J]. 河北法学,2005(6):2-5.
- [4] 许茜. 个人信息保护法“难产”12年,他们为何还说乐观[N]. 科技日报,2017-03-04(003).
- [5] LIPTON J D. Digital multi-media and the limits of privacy law[J]. Case Western Reserve Journal International Law,2010(16):551-571.
- [6] 张新宝. 从隐私到个人信息:利益再衡量的理论与制度安排[J]. 中国法学,2015(3):38-59.
- [7] 冷传莉,李怡. 司法保护视角下的隐私权类型化[J]. 法律科学,2017(5):79-89.
- [8] TOSUN C. Host perceptions of impacts:a comparative tourism study[J]. Annals of Tourism Research,2002(1):231-253.

- [9] 吴标兵,许和隆. 中国公众隐私敏感度实证研究[J]. 南京邮电大学学报(社会科学版),2015(3):82-90.
- [10] JONATHAN S. Exposed;the erosion of privacy in the internet era[J]. Harvard Magazine,2009(9/10):38-43.
- [11] TENE O,POLONETSKY J. Big data for all:privacy and user control in the age of analytics[J]. Northwestern Journal of Technology and Intellectual Property,2013(5):240-273.
- [12] 城田真琴. 数据中间商[M]. 邓多一,译. 北京:北京联合出版公司,2016.
- [13] 特雷莎·M. 佩顿,西奥多·克莱普尔. 大数据时代的隐私[M]. 郑淑红,译. 上海:上海科学技术出版社,2017.
- [14] DALY E M,HAAHR M. Social network analysis for routing in disconnected delay-tolerant manets[J]. IEEE Transactions on Mobile Computing,2008(5):606-621.
- [15] HOLTHOEFER J B,RIVERO A, MORENO Y. Locating privileged spreaders on an online social network[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics,2012(6):66-123.
- [16] 陈静,孙林夫. 复杂网络中节点重要度评估[J]. 西南交通大学学报,2009(3):426-429.
- [17] 武澎,王恒山. 基于特征向量中心性的社交信息超网络中重要节点的评判[J]. 情报理论与实践,2014(5):107-113.

(责任编辑:范艳芹)

The boundary, sensitivity and centrality of personal information based on data analysis of expert and public perception

WU Biaobing, XU helong

(School of Marxism, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: The primary issue of personal information legislation is to determine the boundary of the information. Due to the factors such as evolution of information and communication technology, aggregation analysis of big data, privacy paradox and sensitivity to individual experiences, the dichotomy of general personal information and sensitive personal information is undistributed. Through secondary data analysis or automated processing of general personal information, professional technicians can obtain “presumed data” and even sensitive personal information. Thus there exists the case that some general personal information that is not defined as sensitive is potentially sensitive. The industry community defines this kind of “non-sensitive” and valuable personal information as “core data”. This research finds that: the public has generally low sensitivity to personal information, with relatively lower sensitivity to such information as sexual orientation and fixed telephone call records, while relatively higher sensitivity to identity, life, work, and government information that is closely related to individuals; sensitive personal information and core personal information are overlapping; and that core personal information includes some sensitive personal information and some general personal information.

Key words: boundary; personal information; sensitivity; centrality