

社会网络的中心性分析在即时通信监控中的应用^{*}

李建强^{1,2,**}, 吕世超¹, 范杰¹, 文红^{1,2}, 刘国超¹, 陈霄¹

(1. 电子科技大学 通信抗干扰技术国家级重点实验室, 成都 611731;

2. 东南大学 移动通信国家重点实验室, 南京 210096)

摘要: 将社会网络分析方法引入即时通信系统监控的分析, 用社会网络分析方法对即时通信系统的联系参与者之间的关系进行深入分析, 从中心性的角度分析在即时通信监控系统中各参与者的联系紧密度指标。通过对目标群体的分析, 找出在这个群体中属于领导地位或者意见领袖或信息传递的关键桥点目标, 指导监控系统重点检测这些目标节点, 从而实现对即时通信系统网络的有效监控和不法信息的及时阻断。

关键词: 社会网络; 中心性分析; 即时通信; 监控系统

中图分类号: TN915; TP311 **文献标志码:** A **文章编号:** 1001-893X(2013)02-0214-05

Application of Central Analysis of Social Network in Instant Messages Monitor

LI Jian-qiang^{1,2}, LV Shi-chao¹, FAN Jie¹, WEN Hong^{1,2}, LIU Guo-chao¹, CHEN Xiao¹

(1. National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. National Mobile Communication Research Laboratory, Southeast University, Nanjing 210096, China)

Abstract: This paper applies social network analysis to Instant Messages(IM) monitoring system, by which the relationship between participants of IM system is deeply analysed. It is the index of contact compactness of every participant in IM monitoring system that is investigated based on analysis of centrality. With analysing the target group, this paper finds targets among the group which belongs to the leaders or key peers of message passing, which guides the monitor system to focus on these target nodes. Therefore, those target nodes are put into monitoring. As a result the monitoring system can effectively monitor IM system and block illegal information.

Key words: social network; central analysis; IM messages; monitor system

1 引言

即时通信(Instant Message, IM)满足了人们对于信息交换中快捷方便、高效实时的要求, 深受互联网用户的喜爱, 因此得到了广泛的应用。信息技术的不断发展, 也给非法信息的传递提供了更多的途

径^[1]。例如, 非法组织利用 IM 工具快速传播有害的社会舆论、邪教思想、不健康的色情内容、非法的诈骗信息等。因此, 对即时通信软件的监控对网络的健康和安全使用越来越重要。

社会网络分析(Social Network Analysis, SNA)是研究社会结构和社会关系的一种方法, 是在社会计

^{*} 收稿日期: 2012-05-09; 修回日期: 2012-09-21 Received date: 2012-05-09; Revised date: 2012-09-21

基金项目: 东南大学移动通信国家重点实验室开放研究基金项目(2010D05); 四川省科技支撑计划(2011GZ0183); 四川省应用基础研究项目(2012JY0001)

Foundation Item: The Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (2010D05); The Sichuan Science and Technology Support Plan(2011GZ0183); The Sichuan Application and Basic Research Program(2012JY0001)

^{**} 通讯作者: jianqianglee@163.com **Corresponding author:** jianqianglee@163.com

量法基础上发展起来的分析方法,是数据挖掘的一个重要分支。通过研究社会网络,理解网络中各种参与者之间的关系。目前关于即时通信系统内容的直接监控的研究已有文献讨论,但关于即时通信系统联系网络的有效分析和控制工具还很少。本文将 SNA 分析方法引入即时通信系统监控的分析,用社会网络分析方法对即时通信系统联系参与者之间的关系进行深入的分析,利用社会网络分析中的中心度分析方法对其进行量化研究,检测网络中传输不良信息的核心节点,通过阻断核心节点实现不良信息传播的有效阻断。

2 社会网络分析基本概念

社会网络指的是社会行动者 (Actor) 及其间关系的集合。一个社会网络是由多个点 (社会行动者) 和各点之间的连线 (行动者之间的关系) 组成的集合。社会网络中所说的“点”是各个社会行动者,而社会网络中的“边”指的是行动者之间的各种社会关系。在图论中,社会网络关系可以形象地描述为社会网络图谱和社会网络关系矩阵^[2]。

(1) 社会网络的图示法

社会网络图 G 定义为一个有序对 (V, E) , 记为 $G = (V, E)$, 其中非空集合 $V = \{v_1, v_2, \dots, v_N\}$ 表示社会网络中所有参与者的集合,也可以称为点集,其一个元素对应到一个具体的个人或者其他的社会单位, $|V|$ 表示参与者的个数 N , E 是由 V 中的点组成的无序点对构成的集合, $E = \{E_1, E_2, \dots, E_N\}$, 称为边集,其中每一条边表示的是社会参与者之间的关系^[3]。

与图论中的有向图和无向图一样,社会网络图 G 也是根据参与者之间是否需要表示关系的方向性来划分为有向和无向社会网络图的。这两类图如图 1 和图 2 所示。一般情况下,社会网络中参与者之间的关系是具有方向性的,该方向可通过参与者之间连线的箭头来表示。

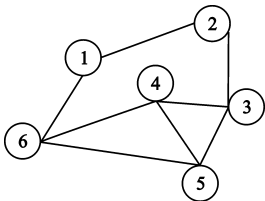


图 1 无向图
Fig.1 Undirected graph

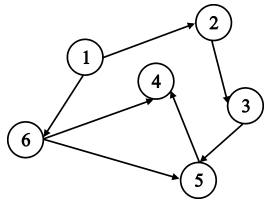


图 2 有向图
Fig.2 Directed graph

(2) 社会网络的邻接矩阵表示法

邻接矩阵中的行和列都对应着集合 V 中的参与者,行和列对应的矩阵元素则表示相应参与者之间的关系,例如有 N 个点的社会网络图 G 的邻接矩阵 $A = (a_{ij})$ 是一个 $N \times N$ 的矩阵,其中,如果参与者 v_i 与 v_j 有联系则 $a_{ij} = 1$, 否则 $a_{ij} = 0$, 于是在有 N 个参与者的社会网络图 G 与对角线为零的 N 阶对称二元矩阵 (其中元素取值为 0 和 1) 之间是一一对应的关系 (若 a_{ij} 取为连接参与者 v_i 与 v_j 的实际边数, 则称 A 为推广的邻接矩阵)。与图示法一样,社会网络的矩阵邻接也分为有向和无向的社会网络关系矩阵,如图 3 和图 4 所示。

节点	1	2	3	4	5	6
1	0	1	0	0	0	1
2	1	0	1	1	0	0
3	0	1	0	1	1	0
4	0	1	1	0	1	0
5	0	0	1	1	0	1
6	1	0	0	0	1	0

图 3 无向对称邻接矩阵
Fig.3 Non-symmetrical adjacency matrix

节点	1	2	3	4	5	6
1	0	0	1	0	0	1
2	1	0	1	0	1	0
3	1	1	0	1	1	0
4	0	0	0	0	0	1
5	1	0	0	1	0	1
6	0	1	1	1	0	0

图 4 有向非对称邻接矩阵
Fig.4 The asymmetric adjacency matrix

3 社会网络的中心性分析方法

中心性是社会网络分析的一个重点,它衡量的是一个节点或参与者在这个网络中居于怎样的中心

位置,具有什么样的“权利”。社会网络的中心性分析方法有度数的中心度和中间中心度。

3.1 点度中心度

社会网络的点度中心度分析定义网络中参与者的数目和某一参与其余参与者之间的联系紧密程度。在一个具体的社会网络中,如一个参与者与大量的其他参与者之间存在着直接的关系,则可以判定该参与者处于这个网络的中心部位,并且很有可能是一个权力比较大的参与者或者是一个领导者。

点度中心度分为绝对中心度和相对中心度。绝对中心度就是指社会网络图 G 中某一参与者节点的度数,也就是与该参与者直接相连的参与者节点的个数。相对中心度是将参与者节点的绝对中心度除以网络中节点最大可能度数所得到的结果,如在一个 N 点图中,任一节点的最大可能度数一定是 $N - 1$,因此在一个有 N 个点的网络图中,任意节点的相对中心度为:绝对中心度/ $(N - 1)$ 。相对中心度用来比较多个参与者节点的中心度程度,与绝对中心度都反应了一个参与者节点在网络中所处的核心程度。

3.2 中间中心度

中间中心度定义为某一参与者控制其他两个参与者之间联系的能力。所谓某一参与者控制其他两个参与者指的是该参与者位于其他两个参与者联系的最短途径上。如果某一参与者的中间中心度高于平均水平,即他处于大量参与者节点对的捷径上,那么就称该参与者起到了沟通其他参与者的桥梁作用,由于他可以控制其他参与者之间的联系,所以也可以认为此参与者居于网络中的重要地位。

简单示例如图 5 所示,该图中的节点 1 和 2 都要通过节点 3 才能与节点 4 建立起连接关系,因此节点 3 的中间中心度就比较高,它在该网络中就充当了一个中介的角色,为 2 和 3 节点起到了信息传递的作用。

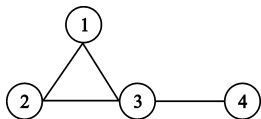


图 5 中间中心度简单示例图
Fig.5 Example of betweenness graph

4 社会网络的中心性分析在即时通信监控控制中的应用

即时通信系统内容的直接监控系统^[4]能够实时地获取通过监控端口的即时通信数据,解析出通信双方的 ID 和内容,检测其中是否含有敏感词汇,同时把这些数据存储在数据库中。通过分析,我们发现很多讨论某一话题的用户通常隶属于某一个群体,而这个群体的核心节点是相对稳定的,因此我们又对其涉及某一话题^[5]的用户提取出来对他们之间的社会网络关系进行分析,找出其中起主导作用的用户。系统结构图如图 6 所示,我们从数据库中读取某一话题中的用户信息,做对称处理后输入社会网络分析工具 UCINET 6.0 中,用 SNA 分析方法对其处理^[6]。

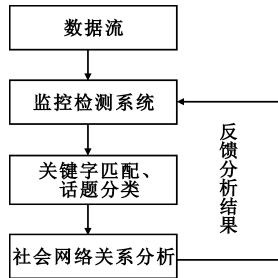


图 6 系统结构框图
Fig.6 Block diagram of system

抽取出具体的联系人,绘制出其整体的网络图谱,为了不泄漏隐私,每一个用户 ID 都用编号代替。图 7 来自社会网络分析软件 UCINET 6.0。

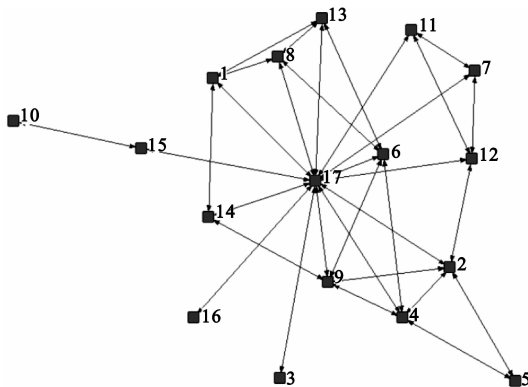


图 7 话题参与者网络结构图
Fig.7 Network structure diagram of topic participants

在图 7 中,我们可以直观地看出这一群体中他们之间的关系,其中节点 17 等占据了比较重要的位

置。下一步我们对其数据进行量化分析,其点度中心度(degree)、中间中心度(betweenness)分别如表 1 和表 2 所示。

表 1 点度中心度详细数据 Tab 1 Detailed data of Degree		
节点	点度中心度 1	相对点度中心度 2
17	14	87.50
2	5	31.25
9	5	31.25
4	5	31.25
6	5	31.25
13	4	25.00
1	4	25.00
8	4	25.00
12	4	25.00
11	3	18.75
7	3	18.75
14	3	18.75

表 2 中间中心度详细数据 Table 2 Detailed data of Betweenness		
节点	绝对中间中心度 1	相对中间中心度 2
17	87.833	73.194
15	15.000	12.500
2	8.000	6.667
4	7.333	6.111
6	2.667	2.222
9	2.333	1.944
12	1.667	1.389
1	1.000	0.833
14	0.500	0.417
13	0.333	0.278
8	0.333	0.278

表 1 中,第一列表示每一个节点与其他节点有联系的节点的个数或者边的条数,也就是绝对中心度。第二列表示节点的相对中心度,其计算公式为:绝对中心度/($N-1$), N 为这个网络的节点个数。表 2 中,第一列表示绝对的中间中心度(简记 C_{ABi}),计算公式为

$$C_{ABi} = \sum_j^n \sum_k^n b_{jk}(i), j \neq k \neq i \text{ 且 } j < k$$

其中, $b_{jk}(i) = g_{jk}(i)/g_{jk}$, g_{jk} 表示节点 j 和 k 之间存

在的捷径数目, $g_{jk}(i)$ 表示节点 j 和 k 之间存在的经过节点 i 的捷径数目, $b_{jk}(i)$ 表示节点 i 控制节点 j 和 k 交往的能力,也就是点 i 处于点 j 和 k 之间的捷径的概率。点 i 相应于网络中所有点对的中间度加在一起就得到了绝对中心度 C_{ABi} ^[7]。表 2 中,第二列表示相对中间中心度,中间中心度才能达到最大值 $C_{\max} = (n^2 - 3n + 2)/2$ 。因此,点的相对中心度

C_{RBi} 为: $C_{RBi} = \frac{2C_{ABi}}{n^2 - 3n + 2}$,其取值范围为 0~1,如表 2 中节点 17 的相对中间中心度为 73.194%,表示控制他人的能力很强,处于这个网络中的核心位置。

由表 1 和表 2 可以很直观的看到:表 1 中占据前几位的分别是节点 17、2、9、4、6,也就表示这几个节点的点度中间度很大,是这个网络中的重要节点;表 2 中占据前几位的是节点 17、15、2、4、6、9 等,表示他们居于重要的中介位置,控制他人或作为桥梁中转不良信息的的能力很强。因此,我们可确定在这个网络中节点 17、9、6、2 等是我们重点监控的对象。如果剔除这几个节点的存在,其网络图由图 7 成为图 8。

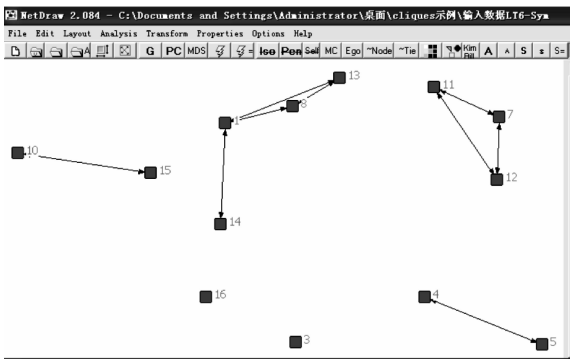


图 8 剔除关键节点后的网络图谱
Fig.8 Network map after removing key node

由图 8 可以清楚地看到整个图被分割成几个不相关的部分,如网络中能有效地剔除或控制这些关键节点的网络传播力,那么网络中不良信息的传播范围和影响将大大减小。因此社会网络分析的结果可以给我们后续的重点监控给予指引。

5 结 论

本文介绍了社会网络分析的含义以及与社会网络分析密切相关的图论和矩阵等数学知识,详细研究了社会网络的中心性分析方法在即时通信系统过滤和监控系统中的应用,分析了即时通信话题团体的社会网络。结果显示:通过社会网络的中心性分

析,可以快速提取出某一话题网络中关于参与者之间的联系以及关键节点、领袖意见等有效信息,通过阻断关键节点,在进行监控时候能有效控制不良信息的扩散。

参考文献:

- [1] Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics[J]. Review of Modern Physics, 2009, 81(2): 591 – 646.
- [2] 刘军. 社会网络分析导论[M]. 北京: 社会科学文献出版社, 2004: 94, 99, 114.
LIU Jun. Introduction to social network analysis[M]. Beijing: Social Science Literature Press, 2004: 94, 99, 114. (in Chinese)
- [3] 张先迪, 李正良. 图论及其应用[M]. 北京: 高等教育出版社, 2005: 1 – 10.
ZHANG Xian-di, LI Zheng-liang. Graph Theory and Its Applications[M]. Beijing: High Education Press, 2005: 1 – 10. (in Chinese)
- [4] 吕世超. 即时通信系统内容过滤和分析研究[D]. 成都: 电子科技大学, 2012.
LV Shi-chao. Research and Design of Instant Message Monitor System[D]. Chengdu: University of Electronic Science and Technology of China, 2012. (in Chinese)
- [5] Dong Haichao, Siu Cheung Hui, He Yulan. Structural analysis of chat messages for topic detection[J]. Online Information Review, 2006, 30(5): 496 – 516.
- [6] 刘军. 整体网分析讲义: UCINET 软件实用指南[M]. 上海: 上海人民出版社, 2009: 119 – 166.
LIU Jun. Overall network analysis handout: UCINET software Practical Guide[M]. Shanghai: Shanghai People's Publishing House, 2009: 119 – 166. (in Chinese)
- [7] Hossain L, Fazio D. The social networks of collaborative process[J]. Journal of High Technology Management Research,

2009, 20(2): 119 – 130.

作者简介:



李建强 (1986—), 男, 四川达州人, 硕士研究生, 主要研究方向为网络安全;

LI Jian-qiang was born in Dazhou, Sichuan Province, in 1986. He is now a graduate student. His research concerns network security.

Email: jianqianglee@163.com

吕世超 (1987—), 男, 河北保定人, 硕士研究生, 主要研究方向为网络安全;

LV Shi-chao was born in Baoding, Hebei Province, in 1987. He is now a graduate student. His research concerns network security.

范杰 (1987—), 男, 山东泰安人, 硕士研究生, 主要研究方向为网络安全;

FAN Jie was born in Tai'an, Shandong Province, in 1987. He is now a graduate student. His research concerns network security.

文红 (1969—), 女, 四川成都人, 博士, 教授、博导, 主要研究方向为无线移动通信。

WEN Hong was born in Chengdu, Sichuan Province, in 1969. She is now a professor with the Ph. D. degree and also the Ph. D. supervisor. Her research concerns wireless communication.

Email: sunlike@uestc.edu.cn

刘国超 (1987—), 男, 湖南澧县人, 硕士研究生, 主要研究方向为信道编码技术;

LIU Guo-chao was born in Lixian, Hunan Province, in 1987. He is now a graduate student. His research concerns channel coding technology.

陈霄 (1987—), 男, 四川资阳人, 硕士研究生, 主要研究方向为信道编码技术。

CHEN Xiao was born in Ziyang, Sichuan Province, in 1987. He is now a graduate student. His research concerns channel coding technology.