群组交流系统新功能探究

——基于QQ群组聊天记录分析

郝建锋

中央财经大学

狗熊会人才计划第二期学员

CONTENTS

背景介绍

SOLOMO 浪潮

John Doerr在2011年首次提出了“SoLoMo” 的概念。

如今，基于SoLoMo的发展模式已被公认为是互联网行业的发展趋势。

Social

社交化——社交类网站和应用，Social毫无疑问是当下乃至未来的潮流

Local

本地化——基于用户当时位置的互联网服务，如大众点评等

Mobile

移动化 ——智能手机所支持的各类移动互联网应用及软件

据艾瑞咨询发布数据，截止2016年12月底，中国移动社交网民超过6亿，占总体移动网民的比例接近 90%，同比增长率高于全球水平。超过一半为30岁以下的新生代用户，互联网的发展伴随着他们的成长，因此他们对互联网更加熟悉，也更愿意尝试移动社交的新玩法与新功能。新生代的大批加入以及同业竞争的不断增强，使得如何设计社交软件的新功能来吸引用户、提升用户粘性成了软件供应商面临的重要问题。

背景介绍

现状

如右图所示，群组交流系统一般包含成员、聊天记录和群组功能三个要素，典型应用QQ在这三个方面进行了许多功能拓展。

缺陷

通讯交流中极为重要的一环——“聊天记录”的功能开发还十分浅显。目前仅停留时间及发言统计上，这种纯频率统计的图表分析的趣味性与互动性都十分低下，并不能满足网络新生代们的好奇心，难以吸引用户。

本文

本文基于上述情况对QQ群组聊天记录进行分析来开发群组新功能，从而提升软件的竞争力，争取到更多有网络社交需求的用户。

QQ群组功能一览

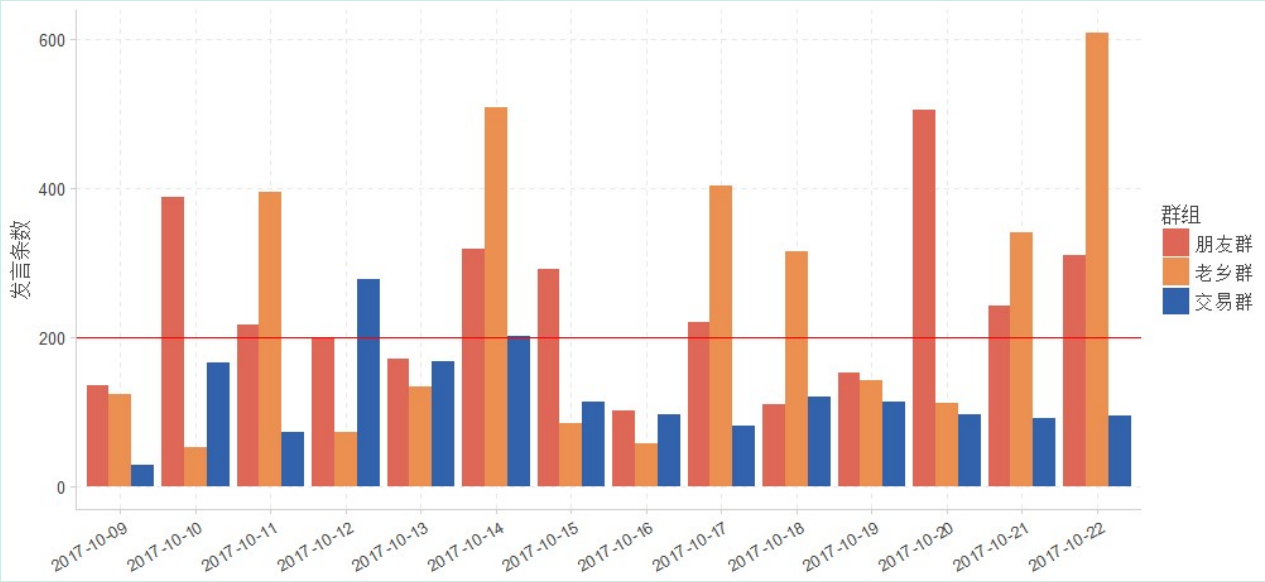
变量说明

本文对来自三个不同类型群组（朋友群、大学同学群和校园交易群）的有效文本数据18798行提取变量如表1.

真实的消息数据格式如屏幕所示，数据虽有固定格式但需进一步处理。

表1 “聊天记录”数据变量说明表

变量名		变量类型	取值范围	备注
时间		日期型	2017-10-09 19:21:05 - 2017/10/24 18:45:35	包含日期、时间 分析时可用函数计算时间差
用户名		文本型	包含昵称、QQ号	由于涉及用户隐私，在本文分析中将昵称与QQ号换为游戏或漫画中的人物名称
聊天内容	图片/表情	文本型	三种格式 【详见备注】	图片、表情在导出的聊天数据中表现格式有以下三种： [图片] [表情] /托腮
	应用外链接	链接型	如： http://s.kugou.com/song.html?id=5DXAmdarAV2	包括网页分享和其他应用数据分享（如通过音乐软件分享的音乐）
	群应用	文本型	有固定格式 【详见备注】	如： [群签到]请使用手机QQ进行查看。 [QQ红包]请使用新版手机QQ进行查看。
	内容 文本	文本型	—— ——	除图片/表情、应用外链接和 群应用消息之外的群成员聊天文本



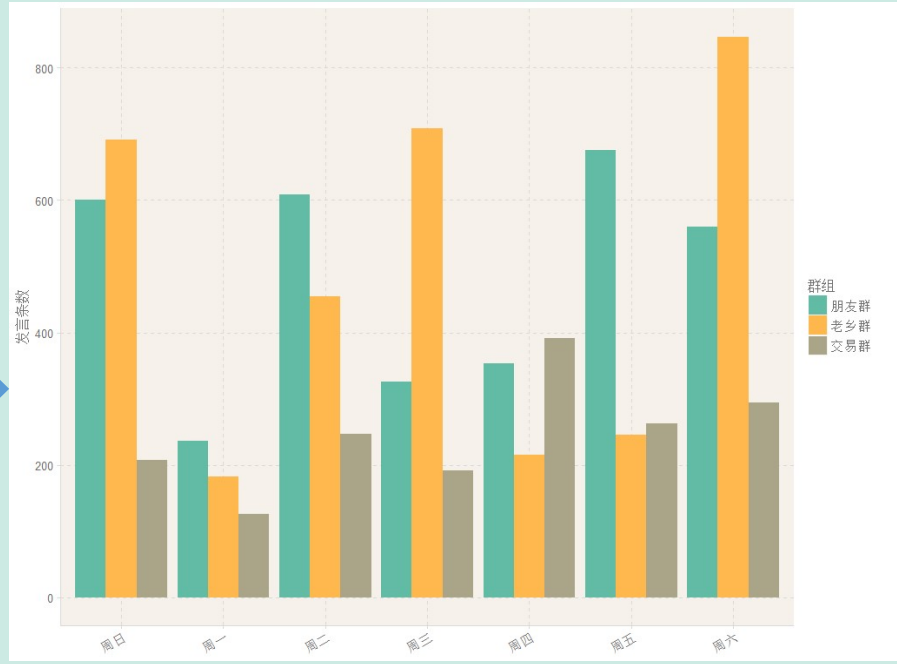
描述性分析

发言时间分布

- 由上图可以看出，在分析时段内交易群较其他两个群组交易群发言数量较少。
- 老乡群在10月22日发言条数最多，超过了600条，看来那一天有某个话题引起了群成员的热烈讨论。
- 朋友群和老乡群都在某几天发言条数很高，这是否涉及到星期内的分布呢？

✓ 大学朋友群与老乡群聊天主要分布于周六、周日，刚好是学生的放假时间；

✓ 校园交易群的聊天则没有明显的特殊分布，这也符合事实，因为一般来说是否放假并不影响校园内的交易洽谈。



描述性分析

日内分布

学生式分布

由图可以看到在一天中两个群组的聊天峰值出现时间几乎相同，都是中午、傍晚以及晚上十点后。

这不难理解，中午和傍晚刚好是学生们的下课时间，而晚间十点后学生一般是完成学习回到宿舍开始上网闲聊，这样的聊天分布是典型的大学生群组聊天时间分布。

描述性分析

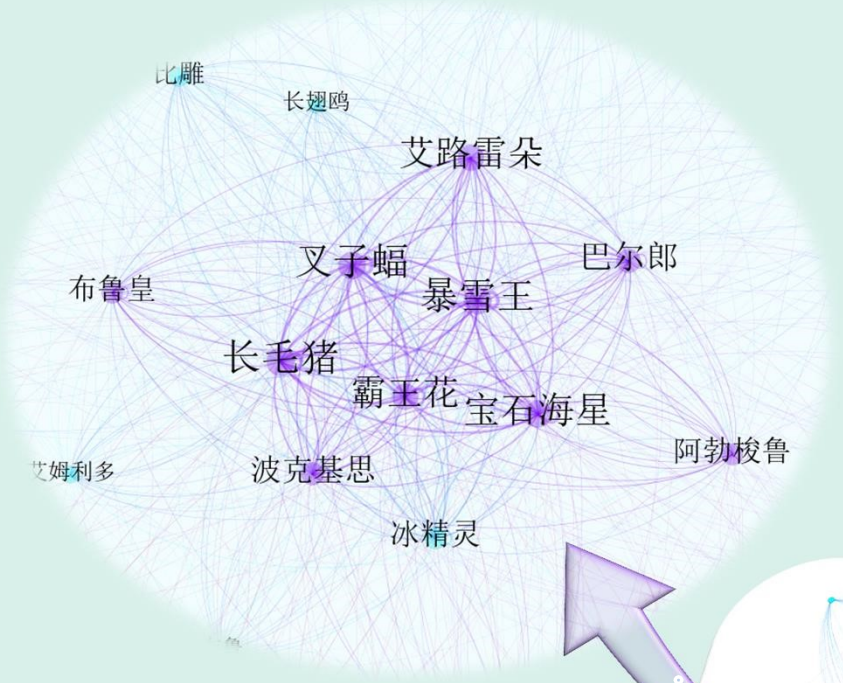
谁是话痨？

从图中可以清晰地看到大学朋友群中平日里发言较多的小伙伴，在该群中，发言用户根据发言次数大致分为三个梯队，其中长毛猪是发言次数最多的那一位——话痨！！！

描述性分析

社交网络图

本文通过对用户连续对话次数的划分以及每次连续对话中群成员两两出现的次数进行统计最终画出了部分群组成员之间的社交网络图。

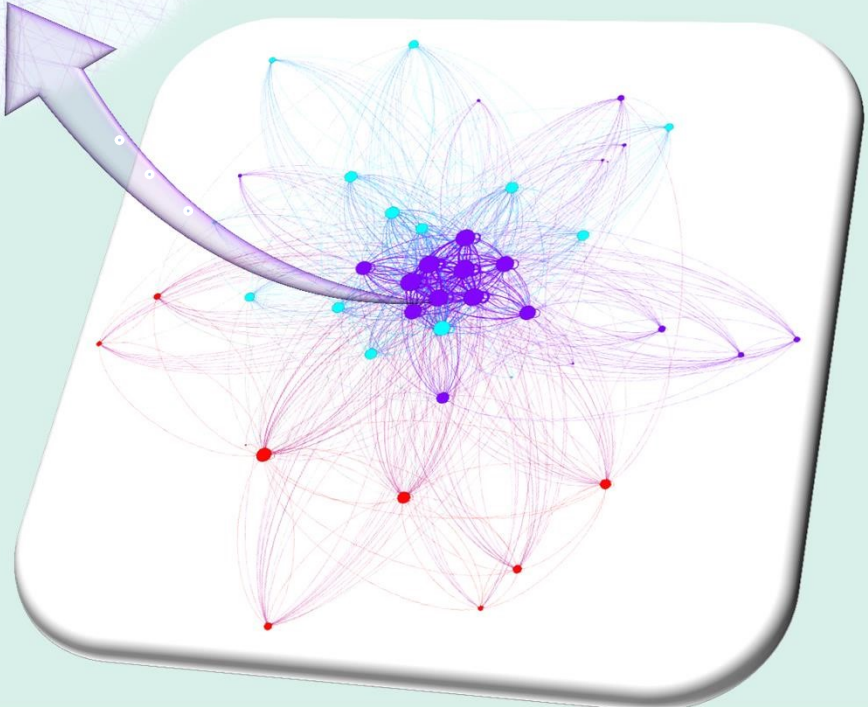


模型解读

右下角图中节点越大代表与他人交流次数越多；
同一颜色的节点代表平时经常一起聊天的小群体。

核心人物

中间紫色的大节点所代表的人员基本与所有人多有交谈，并且发言次数最多，本文对图形进行了局部放大，可以看到平时聊的最开心的要数这几位了，他们称得上是本群的核心人物。



深层分析概述

人员系统

- 建立活跃度变量，查看成员发言天数及每天活跃人数
- 根据群组每周活跃人员的不同来看群组聊天人员的刷新度及退隐度，选取合适图进行可视化

称号系统

- 根据发言人与发言时间确立：

冷场小王子：一段时间无人发言后首个发言并使得其他人加入讨论的人

开聊能手：在他发言后一段时间无人发言

- 根据发言人及发言内容确立：

表情达人：发送表情最多的人

斗图狂魔：发送图片最多的人

- 群组常用功能使用频率[签到、红包、群链接等]可视化

风格系统

通过词典匹配做出群组整体风格雷达图和群组成员风格雷达图[发言次数达一定次数开启]