

社交媒介上谣言的传播机制

——基于新冠疫情的案例分析

学生姓名：杨益

学号：PB21000308

专业：少年班学院统计学

传播学研究方法

授课教师：林爱兵

中国科学技术大学

May 30, 2023

Contents

1	选题背景，意义和目的	2
1.1	研究背景	2
1.2	研究目的	2
1.3	研究意义	3
2	选题论述	4
2.1	选题介绍	4
2.2	名词规范	4
3	研究方案	4
3.1	数据收集和数据预处理	4
3.2	利用内容分析法分析谣言的情感特征	4
3.3	建模分析谣言的传播机制	5
3.3.1	假定得到结论后的验证	5
3.4	可能的不足和改进措施	6
4	运用调查法评估结论的合理性	6
4.1	问卷设计的考虑	6
4.2	验证结论的方案设计	6
4.3	设计的问卷地址	7
4.4	问卷数据处理的建议和补充	7
4.5	问卷数据处理的建议和补充	7
5	研究总结和感想	8
6	问卷表示	11

社交媒体上谣言的传播机制：基于新冠疫情的案例分析

1 选题背景，意义和目的

1.1 研究背景

谣言，是凭空想象或根据主观意愿刻意编造的传言¹，往往激发人们产生恐惧、厌恶和惊慌的负面情绪。社交媒体的出现便捷了人类的生活，但也为谣言的快速传播提供了温床[12]。新冠疫情期间社交媒体更是不乏谣言的传播，这些谣言往往假借官方名义，例如“虚假政府消息”“西安市确诊人群超一半是在核酸检测过程中被感染”，“公寓 800 多人被拉到沙漠隔离”²，“虚假药物消息”“钟南山院士建议盐水漱口防病毒？”³，造成了很大社会恐慌。如何发掘谣言的传播机制，并在谣言散布的早期及时甄别[19]成为了一个非常重要的社会问题。虽然一些研究[2, 14, 5]表明谣言在选举[2]、战争[5]、健康[14]等领域有着相应的传播机制与特点，但是对于新冠疫情期间谣言传播的研究依然还是一个尚未被完全开发的领域。

研究[17, 6]通过对网络数据进行建模，分析社交媒体上谣言消息的传播模式，通过文本的感情色彩发现谣言更为吸引人[17]，相比正常消息传播速度更快、范围更广、影响程度更深，而且不同种类的谣言影响程度不同，例如政治类谣言传播速度往往快于生活类，给出了鉴别社交媒体上谣言的一些可行方式。不可忽视的是，社交媒体上存在着相当一部分机器人[10]散布了大量的恶意谣言，往往加剧了谣言的传播[6]，在本次研究设计中，我们也将充分考虑这一因素在社交媒体谣言传播上的影响。

以往传播学领域对于社交网络信息的研究，往往使用调查[4]和传统统计分析方法[1]推导出结论，近些年，伴随着大数据学科特别是机器学习方法的快速发展，越来越多的研究[19, 14, 17]使用机器学习方法来分析海量的社交数据，特别是新兴的计算传播学也表明了数据驱动代替演绎推理科学的发展，我们也将主要使用机器学习领域最新的方法如图神经网络[11]、Transformer[16]，加上传播学的基本理论假设进行研究分析。

1.2 研究目的

- 本文利用微博 2020-2022 年期间与此相关的所有话题，旨在运用大数据方法了解新冠疫情期间谣言的传播机制，分析重大安全危机期间谣言传播的特点和规律和谣言在不同人群和不同话题上的差别。
- 研究谣言传播的速度、范围和影响程度：通过分析大规模的社交媒体数据，探讨谣言传播的速度、范围和影响程度，揭示其与正常消息传播的差异，进一步了解谣言传播的特点。
- 分析谣言传播的路径和关键节点：通过网络分析和图论方法，揭示谣言传播的路径和关键节点，找出影响谣言传播的因素和机制，为准确预测谣言传播趋势和采取干预措施提供依据。
- 研究谣言传播的动力机制：通过文本情感分析和用户行为模式研究，探讨谣言传播的动力机制，分析人们在传播和接收谣言时的心理和行为特征，深入了解谣言在社交媒

¹Wikipedia:<https://en.wikipedia.org/wiki/Rumor>

²谣言:https://m.thepaper.cn/baijiahao_19430102

³谣言:<https://zhuanlan.zhihu.com/p/105891081>

体环境中的传播机理。

- 对比新冠疫情期间谣言与选举、战争等领域的相似点和不同点，观察一般谣言传播机制的共同点与差别性，并通过文本分析法和调查法交叉验证研究结论的合理性。

1.3 研究意义

- 通过研究新冠疫情这一近期案例谣言传播的机制，填补这一方向研究的空白。在研究设计的过程中引入最新大数据分析的方法，为后续研究提供方法上的新思路。
- 理论意义：通过对谣言传播机制的研究，可以丰富传播学领域对于信息传播的理论体系。谣言作为一种特殊的信息传播现象，其传播机制的深入探索将为传播学理论的发展提供新的视角和思路，丰富传播学研究的内容和范畴。
- 社会意义：谣言的传播给社会带来了许多负面影响，如社会恐慌、误导决策等。通过深入研究谣言传播机制，可以为政府、媒体和社会组织提供科学依据和有效策略，帮助他们更好地应对谣言传播，减少谣言的负面影响，维护社会的稳定和秩序。探讨新冠疫情这种突发重大公众安全危机期间谣言传播的根本逻辑和路径，便于政府在谣言早期进行甄别、删除和辟谣，做好防范，维护社会稳定。通过数据抓取和问卷设计调查不同谣言种类、人群年龄、教育程度等因素对谣言反映的影响，便于政府实施精准的应对策略和建议。
- 实践意义：本研究的结果可以为社交媒体平台提供参考和指导，帮助他们改进谣言检测和防控机制，提高信息质量和用户体验。同时，研究结果还可以为公众提供科学的谣言辨别方法和策略，提高公众对谣言的抵抗能力和信息素养。
- 方法意义：本研究将运用大数据分析、网络分析和文本挖掘等方法，为谣言传播研究提供新的研究思路和技术手段。研究中所采用的方法和技术对于其他相关领域的研究具有借鉴意义，推动跨学科研究和学科交叉融合的发展。

综上所述，本研究旨在深入探索谣言在社交媒体平台上的传播机制，通过揭示谣言传播的特点和规律，为准确预测和干预谣言传播提供理论支持和实践指导，从而促进社会的健康发展和信息传播的良性循环。

2 选题论述

2.1 选题介绍

在本次研究设计中，我们主要采用**内容分析法**，辅以**调查法**。我们通过抓取数据，首先研究谣言的传播机制，并从数据出发，分析这样的传播机制如何得到了结论，说明其因果性。我们还设计了一套问卷方案辅助证明我们结论的合理性，从不同的角度进行相互验证。

2.2 名词规范

我们沿用学术界的通用定义 [15]，把社交媒体中任何虚假的消息都认定为谣言，并不区分传播者的主观意图 [17]。对于谣言的判别，我们借助多个独立的事实检查组织⁴，通常重合度可以达到 95% [17] 以上，说明了其准确性。如果样本量不太大，也可以雇佣少量人员进行独立少次重复的检验以保证严谨性。这里的社交媒介，特指互联网上基于用户关系的内容生产与交换平台。

3 研究方案

3.1 数据收集和数据预处理

本研究的第一步是收集与疫情相关的微博数据，并进行预处理以便后续分析。我们将使用爬虫技术 [9] 来获取微博平台上 2020 年至 2022 年期间与疫情相关的所有话题的数据。考虑到数据量庞大，我们将选择一部分母话题，并追溯所有评论和转发的子话题。初步估计数据量在一亿条左右。在进行数据爬取时，我们将严格遵守微博的安全协议⁵，保证数据收集的合法性和用户隐私的保护。

3.2 利用内容分析法分析谣言的情感特征

在这一步中，我们将采用内容分析法来分析谣言的情感特征。首先，我们将以文本数据的形式将收集到的微博消息导入到程序中。然后，利用文本情感分析的方法 [7]，结合机器学习方法如 Transformer [16]、RNN [13] 和 Bert 等已经训练好的大模型，对消息的情感色彩进行分析。我们将绘制情感分析结果的图表，直观地展示虚假消息与正常消息、不同类型的虚假消息的情感特征差异。具体分析结果的样式可以参考图 1(b)。为了确保结果的稳定性和可靠性，我们可以采用不同的大模型进行数据处理。

我们预期的结论将主要描述谣言传播的基本逻辑和危害性，具体如下：

- 谣言会更新奇，刺激人们传播。
- 谣言往往激发人们产生恐惧、厌恶和惊慌的负面情绪。

⁴几个比较权威事实检查组织的网址, snopes.com, politifact.com, factcheck.org, truthorfiction.com, hoax-slayer.com, and urbanlegends.about.com

⁵微博安全协议: <https://m.weibo.cn/c/regagreement>

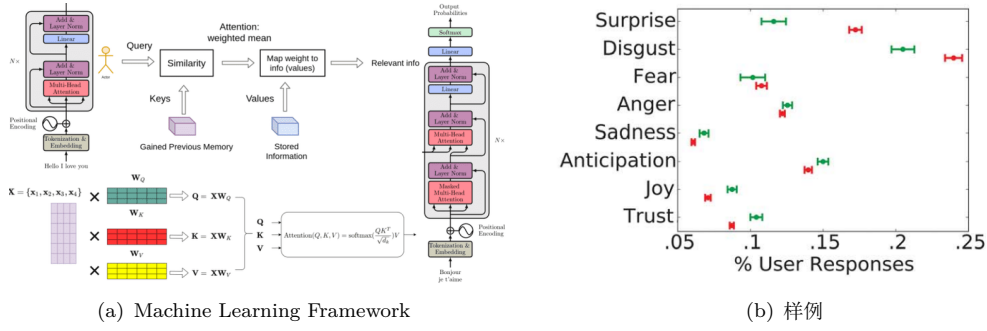


Figure 1: 文本情感分析

3.3 建模分析谣言的传播机制

在这一步中，我们将建立模型来分析谣言的传播机制。我们将从每个谣言的源头开始，每一次谣言消息的转发和评论将形成一条有向边，从而构建一张有向图。同时，我们还将考虑到意见领袖的存在 [3]，因为在消息传递过程中，不同节点的影响力可能不同。因此，我们将以每个用户的关注量作为权重，构建一张加权有向图（如图 2(b)所示）。值得注意的是，图 2(a)是从以往对谣言传播研究的通用范式 [17] 中获取的样例，我们特别考虑到意见领袖可能导致影响力不均衡的问题，因此在我们的研究中将纳入这一因素。

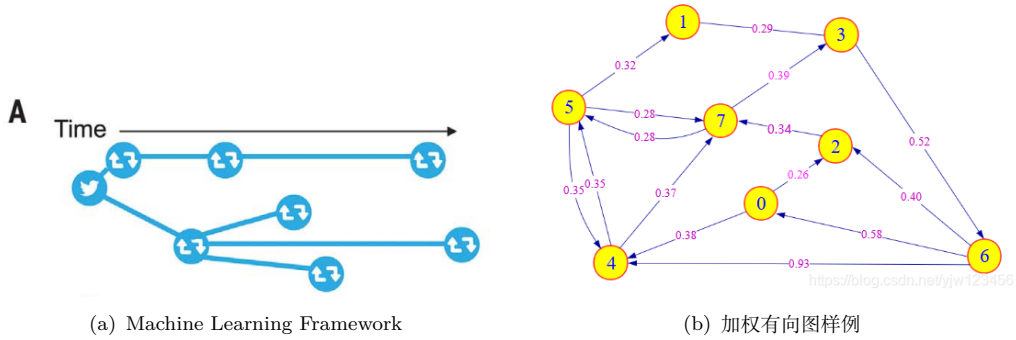


Figure 2: 谣言消息建模

我们将采用统计方法 [17] 来统计疫情期间不同谣言消息传播的深度、广度、传播速度和消息总量等特征。同时，我们还将结合抓取数据中用户的年龄、学历等有效信息，结合公开数据（如图 3(b)、3(a)所示），对以下方面进行建模和研究：

- 不同类别谣言如新冠药物、政府封控等在传播特征上的差别
- 不同学历和年龄的用户在谣言传播中的差别
- 谣言相比正常消息在传播的深度、广度、传播速度的差别

3.3.1 假定得到结论后的验证

假使我们已经得到了相应的结论，由于为研究设计，不妨先假定我们得到的结论为如下三条。下面我们将使用调查法，验证我们结论的合理性。

- 疫情期间政府封控相关的谣言总体传播的最快
- 青年人和老年人更容易转发谣言，低学历者更容易转发谣言
- 谣言相比正常消息传播的更快，更广，更深。

3.4 可能的不足和改进措施

在当前社交媒体环境中，存在大量机器人 [8] 散布虚假消息的问题。因此，在进行调查信任度时，我们可以采用先进的机器人检测方法 [18] 等来排除机器人生成的数据，可以先把数据划分为多个独立的小块进行机器人检测测试，以确定最优的检测方法，以确保分析的准确性和可靠性。

4 运用调查法评估结论的合理性

4.1 问卷设计的考虑

本研究旨在评估社交媒体上谣言传播的特征，并验证之前提出的假设。因此，在设计问卷时，我们需考虑以下因素：首先，调查对象必须是社交媒体上活跃的用户，因此问卷将通过社交媒体渠道进行发布。此外，为了方便数据处理，我们选择在线问卷工具进行设计。在问卷的筛选问题中，如果参与者不是活跃的社交媒体用户，我们将感谢其参与并结束调查。

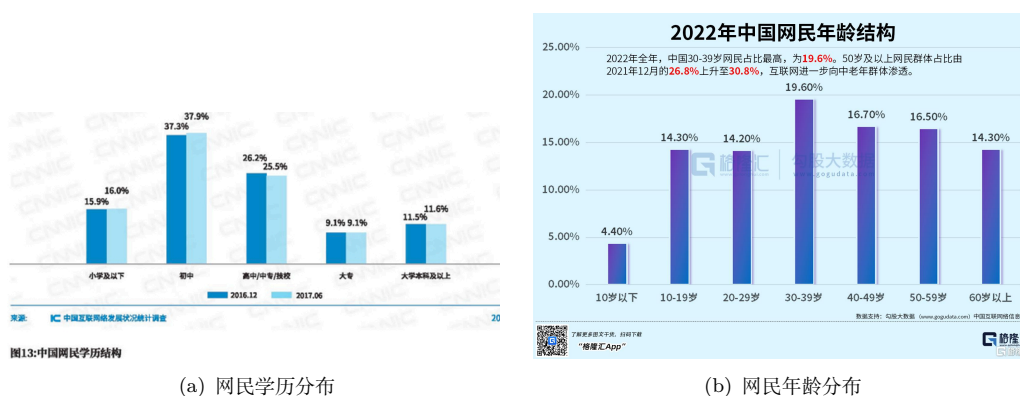
4.2 验证结论的方案设计

根据之前提出的假设，我们希望验证以下结论：

- 谣言会因其新奇性而刺激人们进行传播。
- 谣言往往引发人们产生负面情绪，如恐惧、厌恶和惊慌。
- 在疫情期间，与政府封控相关的谣言总体传播速度最快。
- 青少年和老年人更容易转发谣言，而学历较低者也更容易转发谣言。
- 谣言相比正常消息传播的速度、范围和深度更大。

为了验证这些结论，我们将通过社交媒体发布问卷，并考虑网民的性别、年龄（参见图 3(b)）、学历等因素（参见图 3(a)）。这些基本问题将放在问卷的最前面，以确保获取参与者的基本信息。

在问卷中，我们设计了一个用于判断消息是否为谣言的多选题，以调查不同背景群体对谣言的判别能力。同时，为了了解谣言对情感的影响，我们设置了多个单选题，每个题目都包含相等数量的谣言和正常消息，并要求参与者选择最符合对应情感色彩的选项。此外，为了探究谣言的种类，我们设置了多个单选题，每个题目包含不同的谣言种类。最后，为了研究参与者更愿意转发哪些消息，我们提供了一些关于谣言和正常消息的处理方式的问题。



(a) 网民学历分布

(b) 网民年龄分布

Figure 3: 社交网络基本数据

4.3 设计的问卷地址

我使用问卷星⁶进行问卷设计,并生成了问卷的链接:<https://www.wjx.cn/vm/hxBYJee.aspx>。你可以在研究设计的最后部分(第6节)找到问卷选项。

需要注意的是,在实际的问卷设计中,对于谣言的选择应充分考虑随机性,即应在抓取到的数据总体中进行随机选择,以避免主观因素的影响。同样的问题可以提供多个单选题选项,以确保准确性。

4.4 问卷数据处理的建议和补充

对于不同群体的统计分析,我们可以利用参与者的统计背景信息进行分类,以得出结论。如果一个参与者选择了谣言或正常消息,我们将相应的结果记为+1,对于未选择的部分,记为0。通过将所有结果相加并除以有效填写数,我们可以得到各自的比例,并以可视化的方式展示结果。

另外,考虑到问卷可能存在中途未填完而退出的情况,我们注意将参与者的基本信息放在问卷的最开始部分进行提问。因此,如果一个参与者填写了基本信息,那么他们已回答的部分应纳入统计,而未回答的部分应予以丢弃,以确保数据的准确性。

4.5 问卷数据处理的建议和补充

对于不同群体的统计分析,我们可以借助参与者的背景信息进行分类,并得出相应的结论。当一个用户选择了谣言或正常消息时,我们将相应的结果记为+1,对于未选择的部分,记为0。将所有结果相加并除以有效填写数,可以得到各自的比例,并可以通过可视化方式展示结果。

考虑到问卷可能存在中途未填完而退出的情况,我们需要注意将参与者的基本信息放在问卷的最开始部分进行提问。因此,如果一个用户填写了基本信息,那么其已回答的部分都应该纳入统计,而其他未回答的部分应予以丢弃,以确保数据的准确性。

另外,在进行数据处理时,我们可以使用统计软件来分析和计算结果,以提高数据处理的效率和准确性。统计软件如SPSS、Excel等可以帮助我们进行统计分析、生成图表,并支持进一步的数据解读和可视化展示。

⁶问卷星网址: <https://www.wjx.cn/>

5 研究总结和感想

《传播学研究方法》是我辅修人文学院网络与新媒体专业选择的第一门专业课。在这门课程的学习过程中，我阅读了《社会研究方法》、《传播学研究方法》和《计算传播学》等专业书籍，学习了调查法（问卷设计）、抽样、统计软件等具体知识，进行了多个案例分析，深入了解了社会科学思考和研究的范式。

在撰写研究设计的过程中，我意识到初步的研究设计在实施过程中往往会遇到各种意外情况。因此，在再次修改设计时，我努力考虑研究设计的鲁棒性和严谨性。例如，引入多个独立且权威的方法进行相互验证，考虑可能发生问题的解决方案，并在给出结论和方法时尽量引用已经被广泛接受的方法。同时，我也意识到英文文献往往比中文文献更容易查询和获取，因此在研究设计中基本上都引用了英文文献。

最后，我衷心感谢林老师在本学期的精彩授课和对本次研究设计的耐心修改意见。经过多次修改和完善，本次研究设计得以不断提升，并得到了更加全面的考虑和改进。通过本门课程的学习，我不仅掌握了传播学研究的基本方法和技巧，还深入了解了科学研究的思维方式和逻辑。这将对我的未来学习和研究产生积极的影响，为我打下了坚实的基础。

References

- [1] Alan Agresti, Barbara Finlay, et al. *Statistical methods for the social sciences*. Vol. 207. Pearson Prentice Hall Upper Saddle River, NJ, 2009.
- [2] Hunt Allcott and Matthew Gentzkow. “Social media and fake news in the 2016 election”. In: *Journal of economic perspectives* 31.2 (2017), pp. 211–236.
- [3] Leisa Reinecke Flynn, Ronald E Goldsmith, and Jacqueline K Eastman. “Opinion leaders and opinion seekers: Two new measurement scales”. In: *Journal of the academy of marketing science* 24 (1996), pp. 137–147.
- [4] Lawrence Frey, Carl H Botan, and Gary Kreps. “Investigating communication”. In: *NY: Allyn & Bacon* (2000).
- [5] Irina Khaldarova and Mervi Pantti. “Fake news: The narrative battle over the Ukrainian conflict”. In: *The Future of Journalism: Risks, Threats and Opportunities*. Routledge, 2020, pp. 228–238.
- [6] David MJ Lazer et al. “The science of fake news”. In: *Science* 359.6380 (2018), pp. 1094–1096.
- [7] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. “Sentiment analysis algorithms and applications: A survey”. In: *Ain Shams engineering journal* 5.4 (2014), pp. 1093–1113.
- [8] Claudio Tinoco Mesquita et al. *Infodemia, fake news and medicine: science and the quest for truth*. 2020.
- [9] Ryan Mitchell. *Web scraping with Python: Collecting more data from the modern web*. ” O’Reilly Media, Inc.”, 2018.
- [10] Mariam Orabi et al. “Detection of bots in social media: a systematic review”. In: *Information Processing & Management* 57.4 (2020), p. 102250.
- [11] Ladislav Rampášek et al. “Recipe for a general, powerful, scalable graph transformer”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 14501–14515.
- [12] Yung-Cheng Shen et al. “Why people spread rumors on social media: developing and validating a multi-attribute model of online rumor dissemination”. In: *Online Information Review* 45.7 (2021), pp. 1227–1246.
- [13] Alex Sherstinsky. “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network”. In: *Physica D: Nonlinear Phenomena* 404 (2020), p. 132306.
- [14] Silvia Sommariva et al. “Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study”. In: *American journal of health education* 49.4 (2018), pp. 246–255.
- [15] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. “Defining “fake news” A typology of scholarly definitions”. In: *Digital journalism* 6.2 (2018), pp. 137–153.

- [16] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).
- [17] Soroush Vosoughi, Deb Roy, and Sinan Aral. “The spread of true and false news online”. In: *science* 359.6380 (2018), pp. 1146–1151.
- [18] Soroush Vosoughi, Prashanth Vijayaraghavan, and Deb Roy. “Tweet2Vec: Learning Tweet Embeddings Using Character-level CNN-LSTM Encoder-Decoder”. In: *SIGIR*. ACM, 2016, pp. 1041–1044.
- [19] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. “Enquiring minds: Early detection of rumors in social media from enquiry posts”. In: *Proceedings of the 24th international conference on world wide web*. 2015, pp. 1395–1405.

6 问卷表示

新冠疫情期间社交媒介上的谣言传播 调查问卷

第1题：您的性别： [单选题]

选项 #	小计 #	比例
男	1	<div><div></div></div> 100%
女	0	<div><div></div></div> 0%
本题有效填写人次	1	

第2题：您的年龄是？ [单选题]

选项 #	小计 #	比例
20岁以下	1	<div><div></div></div> 100%
20-30岁	0	<div><div></div></div> 0%
30-50岁	0	<div><div></div></div> 0%
50-60岁	0	<div><div></div></div> 0%
60岁以上	0	<div><div></div></div> 0%
本题有效填写人次	1	

第3题：您的学历是？
[单选题]

选项 #	小计 #	比例
高中文凭及以下	0	<div><div></div></div> 0%
中专，职高或大专	0	<div><div></div></div> 0%
本科	1	<div><div></div></div> 100%
研究生（硕士或博士）	0	<div><div></div></div> 0%
本题有效填写人次	1	

第4题：您觉得以下哪些消息是谣言 [多选题]

选项 #	小计 #	比例
新冠病毒可以通过空气传播	1	<div><div></div></div> 100%
服用感冒药能预防新冠病毒	0	<div><div></div></div> 0%
盐水漱口可以防病毒	0	<div><div></div></div> 0%
上海曾经疫情大爆发，新增30000例	0	<div><div></div></div> 0%
以上都不是	0	<div><div></div></div> 0%
本题有效填写人次	1	

第5题：您觉得以下哪条消息更为新奇？ [单选题]

选项 #	小计 #	比例
疫情期间家畜家禽海鲜都不能吃了	1	<div><div></div></div> 100%
疫苗应当及时打加强针，不然效果很差	0	<div><div></div></div> 0%
应当戴多层口罩放病毒感染	0	<div><div></div></div> 0%
得新冠后一个月内不宜剧烈运动	0	<div><div></div></div> 0%
本题有效填写人次	1	

第6题：对于第五题所有的选项，您最愿意倾向于转发以下哪条消息？ [单选题]

选项 #	小计 #	比例
疫情期间家畜家禽海鲜都不能吃了	1	<div><div></div></div> 100%
疫苗应当及时打加强针，不然效果很差	0	<div><div></div></div> 0%
应当戴多层口罩放病毒感染	0	<div><div></div></div> 0%
得新冠后一个月内不宜剧烈运动	0	<div><div></div></div> 0%
本题有效填写人次	1	

第7题：您觉得以下哪条消息最能引起您的恐惧？ [单选题]

选项 #	小计 #	比例
俄罗斯专家确定新冠病毒为人工合成	1	<div><div></div></div> 100%
每周 4 2021年中国因新冠死亡3634人	0	<div><div></div></div> 0%
每周 1 次北京协和医院新冠患者出逃	0	<div><div></div></div> 0%
每月新冠变异毒株具有更强的传染性	0	<div><div></div></div> 0%
每月 1 次	0	<div><div></div></div> 0%
少于每月一次	0	<div><div></div></div> 0%
本题有效填写人次	1	

第8题：您更倾向于转发以下哪条消息？ [单选题]

选项 #	小计 #	比例
每天一解放军将全面管控疫情	1	<div><div></div></div> 100%
每周 4 2021年国家卫生局发布多吃草莓对防新冠肺炎有重大效果	0	<div><div></div></div> 0%
每周 1 次朝鲜7人确诊新冠肺炎	0	<div><div></div></div> 0%
每新冠肺炎确诊病人挖洞逃出医院	0	<div><div></div></div> 0%
本题有效填写人次	1	