北京郵電大學

本 科 毕 业 设 计（论文）

题目：社猜猜看这个毕设题目是什么

姓　　名　猜　猜

学　　院　信息与通信工程学院

专　　业　通信工程

班　　级　2014211199

学　　号　2014210999

班内序号　99

指导教师　猜　猜

2018 年 5 月

北 京 邮 电 大 学

本科毕业设计（论文）诚信声明

本人声明所呈交的毕业设计（论文），题目《社交网络多媒体信息可信度评估》是本人在指导教师的指导下，独立进行研究工作所取得的成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：_____ 日期：_____

社猜猜看这个毕设题目是什么

摘　要

在人类发展过程中，绝大多数媒介都是 2D 媒介，比如壁画、竹简、书本、油画、影片、相片、电脑绘图等等把壮丽的风景、美丽的模型浓缩到一个平面上。但是 VR 本质上就是 3D 媒介，它是有空间感的，你可以在 VR 中看到整个空间，把感受到的环境、物体，在虚拟世界中忠实完整地呈现出来。这会让所有的创意、设计产业有非常大的改变。

基于虚拟现实技术在现实世界中的发展，传统意义上的第一部虚拟现实电影在 80 年代初开始出现，而计算机技术的进步使得人们开始研究更经济更能被普通民众所能接触到的的 VR 技术。90 年代，全球范围内大量涌现的虚拟现实电影对于出现在商场中的科技产生了影响。自从 2012 年的 Oculus Kickstarter 以来，虚拟现实技术在大屏幕上出现了复苏，它逐渐地成为了一个更亲近普通群众和可被大家接收的消费者设备。

区别于以往的视频类型，对观众来说，以往的 2D、3D 电影我们都是在“观看”屏幕当中的人物表演，观众会受到画框的限制而 VR 则是邀请你进入一种 360° 沉浸式的体验状态。因此也意味着视频观看不再受到以往导演想让观众看哪儿，观众就只能看哪儿的局限。简单来说，VR 就是戴上这些可穿戴眼镜或者头盔后，你就是主角，进入到设计师们给你设定好的这些场景，在这里你可以尽情体验真实的过山车、跳伞运动等现实视觉体验。

因此本文致力于研究全景视频微电影在虚拟现实头戴式显示器中的体验方式，在 Unity 游戏引擎中应用了球面投影着色器并通过“时光机”脚本使得全景视频能够在游戏引擎中进行播放控制。同时，基于 Valve 公司的 SteamVR 插件，提出了能够应用在微电影中的交互方式，例如凝视交互、手柄交互等。

关键词　全景拍摄　投影模型　游戏引擎框架　VR 交互

Have a try to guess what the title is

ABSTRACT

Traditionally based off of technological developments happening in the real world, some of the first movies with VR technology started coming out in the early 80’s when advancements in computer technology allowed research to begin on more affordable VR. With the 90’s came a large influx of VR movies on the coattails of the tech appearing in arcades across the globe. Since the Oculus Kickstarter in 2012, VR has seen a revival on the big screen as it becomes a more affordable and accessible consumer device.

In the process of human development, the vast majority of media are 2D media, such as murals, bamboo slips, books, paintings, films, photos, computer graphics, etc. make the magnificent scenery and models condensed to a plane. But VR is essentially a 3D media, it is a sense of space,so you can see the whole space in the VR, the feelings of the environment, objects, in the virtual world faithfully and completely presented.

Different from the previous video type, for the audience, the previous 2D, 3D movies we are in the ”watch” screen where the characters perform. Therefore the audience will be limited by the frame while VR invites you into a 360 ° immersion type of experience. So it also means that the video is no longer subject to the limitations made by directors who want to let the audience see where the audience can only see. In short, VR is when wearing these wearable glasses or helmets, you are the protagonist, stepping into the designers who have already set these scenes, where you can enjoy the real roller coaster, skydiving and other realistic visual experience.

Therefore, this paper is devoted to the study of panoramic micro-film in the virtual reality headset in the experience of the way, applying the equirectangular projection shader and the ”time machine” script that makes the panoramic video can be played in the game engine playback controller. At the same time, based on Valve’s SteamVR plug-in , this paper proposes the interactive ways which can be applied in micro-films, such as staring interaction, handle interaction.

KEY WORDS Panorama Projection Models Game Engine Framework VR Interaction

目　录

第一章　引言

1.1　研究背景及意义

1.1.1　社交媒体发展现状

社交媒体是一种供用户创建在线社群来分享信息、观点、个人信息和其它内容（如视频）的电子化交流平台，社交网络服务（social network service, SNS）和微博客（microblogging）都属于社交媒体的范畴[1]，国外较为知名的有 Facebook[1]、Instagram[2]、Twitter[3]、LinkedIn[4]等，国内较为知名的有新浪微博[5]。社交媒体营销公司 We Are Social 的《2018 数字报告》[2] 显示，截至 2018 年 1 月，全球的活跃社交媒体用户已达到 31.96 亿，同比增长 13%，全人口渗透率达到 42%。其中，知名 SNS 服务商 Facebook 月活跃用户数高达 21.67 亿，微博客服务商 Twitter 月活跃用户数达到 3.3 亿，新浪微博月活跃用户数达到 3.76 亿。可以说，社交媒体已经成为了互联网用户的必需品之一。

在社交媒体的强覆盖下，新闻信息的传播渠道也悄然了发生变化。根据美国皮尤研究中心的 2017 年 9 月发布的调查结果[3]，67% 的美国民众会从社交媒体上获取新闻信息，其中高使用频率用户占 20%。在国内，中国互联网信息中心《2016 年中国互联网新闻市场研究报告》[4] 也显示，社交媒体已逐渐成为新闻获取、评论、转发、跳转的重要渠道，在 2016 年下半年，曾经通过社交媒体获取过新闻资讯的用户比例高达 90.7%，在微信、微博等社交媒体参与新闻评论的比例分别为 62.8% 和 50.2%。社交媒体正在成为网络上热门事件生成并发酵的源头，在形成传播影响力后带动传统媒体跟进报道，最终形成更大规模的舆论浪潮。

然而，社交媒体在改变用户获取新闻途径，加速信息传播分发的同时，也为虚假信息的传播提供了有利环境。2016 年美国大选后，Facebook 爆出“假新闻事件”[6]，其被指控在 Facebook 平台上传播的假新闻严重影响了美国大选结果。2018 年 3 月，《Science》发表了麻省理工学院学者针对真假新闻传播情况的研究[5]。研究发现，在 Twitter 平台上，包含虚假新闻的推文更容易被转发，且更容易形成“病毒式传播”，真实消息传播至 1500 人的时间，比虚假消息长 6 倍。而在国内，新浪微博由于其发布方便、传播迅速、受众广泛且总量大的特点，成为了虚假信息传播的重灾区：《中国新媒体发展报告（2013）》[6] 显示，2012 年的 100 件微博热点舆情案例中，有超过 1/3 出现谣言；《中国新媒体发展报告（2015）》[7] 对 2014 年传播较广、比较典型的 92 条假新闻进行了多维度分析，发现有 59% 的虚假新闻首发于新浪微博。

此等信息的传播严重损害了有关公众人物的名誉权，降低了社交媒体服务商的商业

[1] http://www.facebook.com/

[2] https://www.instagram.com/

[3] http://www.twitter.com/

[4] http://www.linkedin.com/

[5] http://www.weibo.com/

[6] https://www.recode.net/2017/4/28/15476142/facebook-report-trump-clinton-russia-us-presidential-election

美誉度，扰乱了网络空间秩序，冲击着网民的认知，极易对民众造成误导，带来诸多麻烦和经济损失，甚至会导致社会秩序的混乱。针对社交媒体谣言采取行动成为了有关部门、服务提供商和广大民众的共同选择。

1.2 国内外研究现状

本节将与下文有关的关键知识点的研究现状进行概述。

1.2.1 文本的表示方法

传统的文本挖掘通常会使用字符匹配、词典比对、知识库搜索等手段和工具，但它们难以起到学习并挖掘抽象的语义联系的作用，难以满足自然语言处理（Natural Language Processing, NLP）任务的需求。为了把文字内容纳入可计算、可度量的范围中来，学者对文字内容进行了编码，对每个词语进行向量化表示，以便作为机器学习任务的输入。其中最著名的是独热表示（One-Hot Represetation）和一种分布式表示模型——Word2Vec。

独热表示

该方法首先需要统计表示范围内所有词的数量 N，然后给这 N 个词分别编号为 $1, 2, \ldots, N$，最终使用一个仅第 k 维非 0（通常为 1）的 N 维向量来表示编号为 k 的词。例如，在词语空间 $\Omega = \{\text{中国}, \text{首都}, \text{北京}\}$ 中，“中国”的独热编码为 $[1, 0, 0]$，“首都”的独热编码为 $[0, 1, 0]$，“北京”的独热编码为 $[0, 0, 1]$。从计算机存储的角度上讲，其结构就是一个 Hash 表，再与最大熵、条件随机场（Conditional Random Field，CRF）、支持向量机（Support Vector Machine，SVM）等算法相配合，可以解决大多数自然语言处理的基础任务。

显然，这种表示方式的优势在于操作简单，表示简洁，但其缺陷不容忽视：首先需要表示的所有词数量越多，则表示向量就越长，在实际计算中存在严重的稀疏问题，无法像音频、图像等数据获取稠密表示，形成“维数灾难”（Curse of Dimensionality）[8]；更为关键的是，独热表示仅仅将词语离散符号化，不能表达词与词之间的关系，从而丢失了许多语义信息。

此外，这种表示方法也经常用于其它取值空间不大的非数值数据的表示上。

分布式表示——Word2Vec

词的分布式表示（Distributed Representation）最早由“神经网络之父”Geoffrey Hinton 于 1986 年提出[9]，其基本思想是通过训练将每个词表示为 K 维实值短向量（这里的“短”是相对于独热编码而言的），并通过词嵌入（Word Embedding）在向量空间中的距离来表征词语之间的语义相似度。其之所以被称为“分布式”表示，核心在于一个词的 K 个维度中，每一个都承载着一部分词语的抽象语义信息。然而，其在实际应用上算法复杂度过高，故一直没有被广泛地采用。

直到 2013 年，谷歌提出了著名的词嵌入学习模型 Word2Vec[10]，解决了效率问题。自此词语的分布式表示几乎成为了所有自然语言处理任务的标准预处理方法。为了表达表格的用法，下面插入一个跟这段话没有关系的表格。

表 1-1 基于浏览者行为的特征

特征	描述	形式与理论范围
点赞量	微博的点赞数量	数值，$\mathbb{N}$
评论量	微博的评论数量	数值，$\mathbb{N}$
转发量	微博的转发数量	数值，$\mathbb{N}$

1.3 模型描述

1.3.1 基于主成分分析

在 Weiling Chen[11] 和 Yan Zhang[12] 的工作中，均使用了主成分分析（Principle Component Analysis, PCA）作为基本的数据降维方法。下面对主成分分析进行介绍。

主成分分析是一种简单的机器学习算法，其功能可以从两方面解释：一方面可以认为它提供了一种压缩数据的方式，另一方面也可以认为它是一种学习数据表示的无监督学习算法。[20] 通过 PCA，我们可以得到一个恰当的超平面及一个投影矩阵，通过投影矩阵，样本点将被投影在这一超平面上，且满足最大可分性（投影后样本点的方差最大化），直观上讲，也就是能尽可能分开。

对中心化后的样本点集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_i,\ldots,\boldsymbol{x}_m\}$（有 $\sum_{i=1}^{m}\boldsymbol{x}_i=0$），考虑将其最大可分地投影到新坐标系 $\boldsymbol{W}=\{\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_i,\ldots,\boldsymbol{w}_d\}$，其中 $\boldsymbol{w}_i$ 是标准正交基向量，满足 $\|\boldsymbol{w}_i\|_2=1$，$\boldsymbol{w}_i^T\boldsymbol{w}_j=0$（$i\neq j$）。假设我们需要 d'（$d'<d$）个主成分，那么样本点 $\boldsymbol{x}_i$ 在低维坐标系中的投影是 $\boldsymbol{z}_i=(z_{i1};z_{i2};\ldots;z_{id'})$，其中 $z_{ij}=\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{x}_i$，是 $\boldsymbol{x}_i$ 在低维坐标系下第 j 维的坐标。对整个样本集，投影后样本点的方差是

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{z}_i\\=&\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{W})^{\mathrm{T}}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{W})\\=&\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{W}\\=&\frac{1}{m}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\end{aligned}\qquad\text{式（1-1）}$$

由于我们知道新坐标系 $\boldsymbol{W}$ 的列向量是标准正交基向量，且样本点集 $\boldsymbol{X}$ 已经过中心化，则 PCA 的优化目标可以写为

$$\begin{aligned}\max_{\boldsymbol{W}}\quad&tr(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W})\\\text{s.t.}\quad&\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I}\end{aligned}\qquad\text{式（1-2）}$$

由于 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 是协方差矩阵，那么只需对它做特征值分解，即

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^{\mathrm{T}}\qquad\text{式（1-3）}$$

其中 $\boldsymbol{\Lambda} = diag(\boldsymbol{\lambda})$，$\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$。

具体地，考虑到它是半正定矩阵的二次型，存在最大值，可对式（1-2 ）使用拉格朗日乘数法

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}_i = \lambda_i \boldsymbol{w}_i \qquad \text{式（1-4）}$$

之后将求得的特征值降序排列，取前 d' 个特征值对应的特征向量组成所需的投影矩阵 $\boldsymbol{W}' = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{d'})$，即可得到 PCA 的解。PCA 算法的描述如算法1所示。

算法 1 主成分分析（PCA）

输入： 样本集 $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_m\}$，低维空间维数 d'
输出： 投影矩阵 $\boldsymbol{W}' = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{d'})$

对所有样本中心化 $\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i - \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{x}_i$
计算样本的协方差 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$
对协方差矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 做特征值分解
取最大的 d' 个特征值所对应的特征向量 $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{d'}$

论文[11] 认为，通过 PCA 保留一定的主成分，可以更好地把握历史微博的共性，使历史上的非谣言微博与谣言微博产生可度量的距离。

论文采取了排序的检测方式：如果待判别的微博在特征空间中距离非谣言微博数据的“重心”比任何非谣言微博都要远，即成为了离群点，则认为该微博是一条谣言。但考虑到实验过程中应尽量统一化比较手段，在本节中，我们采用了 Yan Zhang 论文[12] 中的阈值法来进行判别：在特征空间中，如果待判定微博没有阈值范围内的相邻点，则认为该微博是一条谣言。

记待判定微博 $\boldsymbol{w}_0$ 的经典特征向量为 $\boldsymbol{f}_0^c$，它的发布者在 $\boldsymbol{w_0}$ 前发布的 k 条微博为 $\boldsymbol{W} = \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k$，这 k 条微博对应的经典特征向量集为 $\boldsymbol{F}_W^c = \{\boldsymbol{f}_1^c, \boldsymbol{f}_2^c, \ldots, \boldsymbol{f}_k^c\}$。令 $label = 1$ 代表谣言，$label = 0$ 代表非谣言。算法的具体流程如算法2所示。

算法 2 基于 PCA 的信息可信度评估

输入： $\boldsymbol{f}_0^c$，$\boldsymbol{F}_W^c$，保留主成分数 n
输出： 标签 $label \in \{0, 1\}$

对所有特征向量应用 PCA，保留前 n 个主成分 $\boldsymbol{o}_i^c \leftarrow PCA(\boldsymbol{f}_i^c, n)$（$i = 0, 1, \ldots, k$）
计算 $\boldsymbol{F}_W^c$ 中各向量的平均距离 μ 和标准差 σ
计算阈值 $thr = \mu/\sigma$
if $\min_{1<j\le k} \|\boldsymbol{o}_0^c - \boldsymbol{o}_j^c\|_2 > thr$ **then**
    $label \leftarrow 1$
else
    $label \leftarrow 0$
end if

在该工作中，取主成分数 n 为 5，k 为 50。

1.3.2 基于欠完备自编码器

在 Mayu Sakurada 的论文[21] 中，学者使用了自编码器进行异常检测来进行非线性降维。考虑到，PCA 和 TSVD 均停留在线性降维的范畴中，而社交媒体信息的可信度

评估问题被认为是极为复杂的非线性问题，Yan Zhang 在另一篇工作[13] 中，将历史信息中特征的提取方法由主成分分析变为了欠完备自编码器（undercomplete autoencoder, UAE），得到了基于欠完备自编码器的可信度评估模型。下面介绍自编码器的相关知识。

自编码器（autoencoder, AE）的概念最早源自 Rumelhart 等人于 1986 年发表在《Nature》上的文章《Learning representations by back-propagating errors》[22]。自编码器是神经网络的一种，经过训练后能尝试将输入复制到输出。[20] 自编码器的结构如图 1-1 所示。

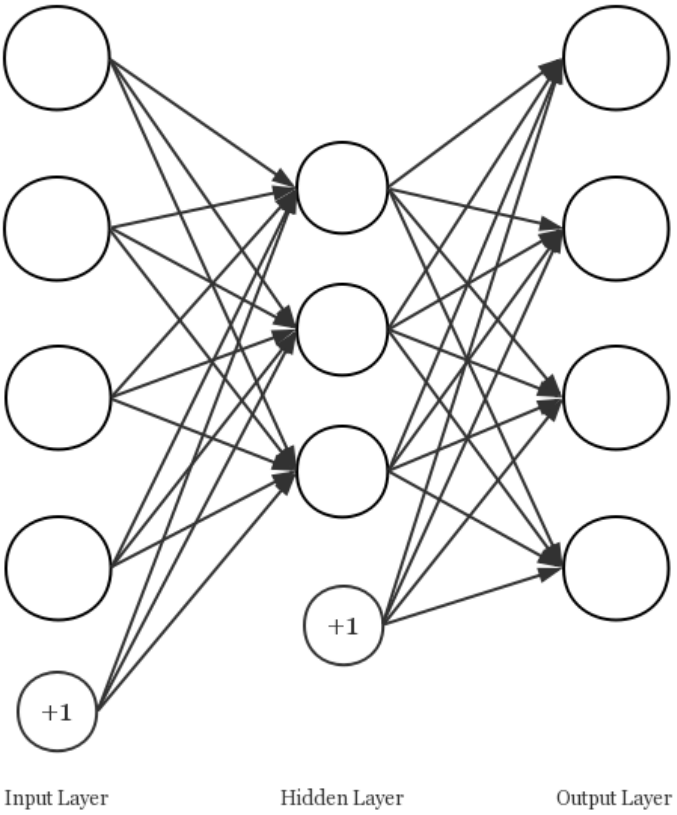


图 1-1 自编码器结构

自编码器内部有一个隐藏层（hidden layer）$\boldsymbol{h}$，可以产生用于表示输入数据的编码（code）。如果输入层（input layer）的输出数据为 $\boldsymbol{x}$，输出层 (output layer）的输出结果为 $\hat{\boldsymbol{x}}$，那么输入层、隐藏层和输出层有如下函数关系：

$$\boldsymbol{h} = g_1(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}_1) \qquad \text{式（1-5）}$$

$$\hat{\boldsymbol{x}} = g_2(\boldsymbol{V}\boldsymbol{h} + \boldsymbol{b}_2) \qquad \text{式（1-6）}$$

其中，$\boldsymbol{b}_1$ 和 $\boldsymbol{b}_2$ 是偏置项，g_1 和 g_2 分别是输入层到隐藏层和隐藏层到输出层间的激活函数（activation function），正是由于激活函数 f 和 g 的存在，层与层之间的映射才是非线

性的。

由上可知，图 1-1 表示的是一个由含有 4 个神经元的输入层、含有 3 个神经元的隐藏层和含有 4 个神经元的输出层组成的自编码器，+1 代表偏置项。

自编码器分为多种，其中最经典的结构正如 1-1 所示，其特点是隐藏层单元数小于输入输出层，Ian Goodfellow 在《深度学习》[20] 一书中将这种自编码器称为欠完备自编码器（undercomplete autoencoder）。欠完备自编码器的特点，使其可以**迫使**隐藏层用小于原始数据的维数来尽可能表示原数据，以期在输出层尽可能将原始数据重构。那么，隐藏层的表示实际上就是一种**有损压缩编码**的结果，那么由输入层到隐藏层的部分，就可以被看做一个有损的编码器（encoder），而隐藏层到输出层的部分，自然就是解码器（decoder）。

进一步，对于压缩表示这种任务，自编码器的损失函数理应表征输入与输出之间的差别，容易想到使用均方误差（Mean Square Error，MSE）：

$$\begin{aligned} Loss &= MSE(\boldsymbol{X}, \hat{\boldsymbol{X}}) \\ &= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij}-\hat{x}_{ij})^2 \end{aligned} \qquad \text{式（1-7）}$$

其中，$\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$ 是输入数据，$\hat{\boldsymbol{X}}=\{\hat{\boldsymbol{x}}_1,\hat{\boldsymbol{x}}_2,\ldots,\hat{\boldsymbol{x}}_n\}$ 是输出数据，假设输入向量和输出向量都有 m 维，即 $\boldsymbol{x}_i=(x_{i1},x_{i2},\ldots,x_{im})$，$\hat{\boldsymbol{x}}_i=(\hat{x}_{i1},\hat{x}_{i2},\ldots,\hat{x}_{im})$，$i=1,2,\ldots,n$。

如果欠完备自编码器只针对一条原始数据进行编码表示，其应当尽可能拟合该数据的特征；而如果针对一批数据，其训练学习的结果应当是拟合所有数据中最有共性的部分，以期降低损失。基于这种认识，仿照基于 PCA 和基于 TSVD 方法的思路，我们就可以基于 UAE 的社交媒体信息可信度评估模型。其算法描述如3所示（沿用基于 PCA 的评估算法中的符号）。

算法 3 基于 UAE 的信息可信度评估

输入： $\boldsymbol{f}_0^c$，$\boldsymbol{F}_W^c$

输出： 标签 $label \in \{0,1\}$

用 $\boldsymbol{F}_W^c$ 中的经典特征向量，基于反向传播，训练自编码器网络
使 $\boldsymbol{f}_0^c$ 通过训练好的网络，得到输入输出之间的方差损失 l_0
使 $\boldsymbol{F}_W^c$ 中的经典特征向量，通过训练好的网络，得到各自的损失，并求得其均值 μ 和标准差 σ
计算阈值 $thr=\mu+\sigma$
if $l_0 > thr$ **then**
    $label \leftarrow 1$
else
    $label \leftarrow 0$
end if

参考文献

[1] Merriam-Webster. Social Media [EB/OL]. 2018 [2018-04-15]. http://www.merriam-webster.com/dictionary/socialmedia.

[2] We Are Social. Digital in 2018: World's Internet Users Pass the 4 Billion Mark [EB/OL]. 2018 [2018-04-15]. https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018.

[3] Pew Research Center. News Use Across Social Media Platforms 2017 [EB/OL]. 2017 [2018-04-15]. http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017.

[4] 中国互联网络信息中心. 2016 年中国互联网新闻市场研究报告 [EB/OL]. 2017 [2018-04-15]. http://www.cnnic.cn/hlwfzyj/hlwxzbg/mtbg/201701/P020170112309068736023.pdf.

[5] Vosoughi S, Roy D, Aral S. The spread of true and false news online [J]. Science, 2018, 359 (6380): 1146–1151.

[6] 唐绪军，吴信训，黄楚新 等. 中国新媒体发展报告 No.4(2013) [M]. 社会科学文献出版社，2013.

[7] 唐绪军，吴信训，黄楚新 等. 中国新媒体发展报告 No.6(2015) [M]. 社会科学文献出版社，2015.

[8] Bengio Y, Ducharme R, Vincent P et al. A neural probabilistic language model. [J]. Journal of Machine Learning Research, 2006, 3 (6): 1137–1155.

[9] Hinton G E. Learning distributed representations of concepts. [C]. In Eighth Conference of the Cognitive Science Society, 1986.

[10] Mikolov T, Sutskever I, Chen K et al. Distributed representations of words and phrases and their compositionality [C]. In Advances in neural information processing systems, 2013: 3111–3119.

[11] Chen W, Chai K Y, Lau C T et al. Behavior deviation: An anomaly detection view of rumor preemption [C]. In Information Technology, Electronics and Mobile Communication Conference, 2016: 1–7.

[12] Zhang Y, Chen W, Chai K Y et al. A distance-based outlier detection method for rumor detection exploiting user behaviorial differences [C]. In International Conference on Data and Software Engineering, 2017: 1–6.

[13] Goodfellow I, Bengio Y, Courville A. Deep Learning [M]. The MIT Press, 2016.

[14] Sakurada M, Yairi T. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction [C]. In Mlsda 2014 Workshop on Machine Learning for Sensory Data Analysis, 2014: 4.

[15] Zhang Y, Chen W, Chai K Y et al. Detecting rumors on Online Social Networks using multi-layer autoencoder [C]. In 2017 IEEE Technology Engineering Management Conference (TEMSCON), 2017: 437–441.

[16] Rumerlhar D E. Learning representation by back-propagating errors [J]. Nature, 1986, 323 (3): 533–536.

致　　谢

此处请写致谢的内容。它可以有多段。

附　　录

附录 1　缩略语表

英文缩写	英文名称	中文
AE	autoencoder	自编码器
CRF	conditional random field	条件随机场
LR	logistic regression	逻辑回归
LSTM	Long Short Term Memory	长短时记忆单元

附录 2　数学符号

数和数组

a	标量（整数或实数）
$\boldsymbol{a}$	向量
$dim()$	向量的维数
$\boldsymbol{A}$	矩阵
$\boldsymbol{A}^{\mathrm{T}}$	矩阵 **A** 的转置
$\boldsymbol{I}$	单位矩阵（维度依据上下文而定）
$diag(\boldsymbol{a})$	对角方阵，其中对角元素由向量 $\boldsymbol{a}$ 确定

外 文 译 文

真假新闻的在线传播

Soroush Vosoughi, Deb Roy, Sinan Aral
麻省理工学院

决策、合作、通信和市场领域的基础理论全都将对真实或准确度的概念化作为几乎一切人类努力的核心。然而，不论是真实信息还是虚假信息都会于在线媒体上迅速传播。定义什么是真、什么是假成了一种常见的政治策略，而不是基于一些各方同意的事实的争论。我们的经济也难免遭受虚假信息传播的影响。虚假流言会影响股价和大规模投资的动向，例如，在一条声称巴拉克·奥巴马在爆炸中受伤的推文发布后，股市市值蒸发了 1300 亿美元。的确，从自然灾害到恐怖袭击，我们对一切事情的反应都受到了扰乱。新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

SOCIAL SCIENCE

The spread of true and false news online

Soroush Vosoughi,[1] Deb Roy,[1] Sinan Aral[2]*

We investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprise ~126,000 stories tweeted by ~3 million people more than 4.5 million times. We classified news as true or false using information from six independent fact-checking organizations that exhibited 95 to 98% agreement on the classifications. Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. We found that false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust. Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it.

Foundational theories of decision-making (*1*–*3*), cooperation (*4*), communication (*5*), and markets (*6*) all view some conceptualization of truth or accuracy as central to the functioning of nearly every human endeavor. Yet, both true and false information spreads rapidly through online media. Defining what is true and false has become a common political strategy, replacing debates based on a mutually agreed on set of facts. Our economies are not immune to the spread of falsity either. False rumors have affected stock prices and the motivation for large-scale investments, for example, wiping out $130 billion in stock value after a false tweet claimed that Barack Obama was injured in an explosion (*7*). Indeed, our responses to everything from natural disasters (*8*, *9*) to terrorist attacks (*10*) have been disrupted by the spread of false news online.

New social technologies, which facilitate rapid information sharing and large-scale information cascades, can enable the spread of misinformation (i.e., information that is inaccurate or misleading). But although more and more of our access to information and news is guided by these new technologies (*11*), we know little about their contribution to the spread of falsity online. Though considerable attention has been paid to anecdotal analyses of the spread of false news by the media (*12*), there are few large-scale empirical investigations of the diffusion of misinformation or its social origins. Studies of the spread of misinformation are currently limited to analyses of small, ad hoc samples that ignore two of the most important scientific questions: How do truth and falsity diffuse differently, and what factors of human judgment explain these differences?

Current work analyzes the spread of single rumors, like the discovery of the Higgs boson (*13*) or the Haitian earthquake of 2010 (*14*), and multiple rumors from a single disaster event, like the Boston Marathon bombing of 2013 (*10*), or it develops theoretical models of rumor diffusion (*15*), methods for rumor detection (*16*), credibility evaluation (*17*, *18*), or interventions to curtail the spread of rumors (*19*). But almost no studies comprehensively evaluate differences in the spread of truth and falsity across topics or examine why false news may spread differently than the truth. For example, although Del Vicario *et al.* (*20*) and Bessi *et al.* (*21*) studied the spread of scientific and conspiracy-theory stories, they did not evaluate their veracity. Scientific and conspiracy-theory stories can both be either true or false, and they differ on stylistic dimensions that are important to their spread but orthogonal to their veracity. To understand the spread of false news, it is necessary to examine diffusion after differentiating true and false scientific stories and true and false conspiracy-theory stories and controlling for the topical and stylistic differences between the categories themselves. The only study to date that segments rumors by veracity is that of Friggeri *et al.* (*19*), who analyzed ~4000 rumors spreading on Facebook and focused more on how fact checking affects rumor propagation than on how falsity diffuses differently than the truth (*22*).

In our current political climate and in the academic literature, a fluid terminology has arisen around "fake news," foreign interventions in U.S. politics through social media, and our understanding of what constitutes news, fake news, false news, rumors, rumor cascades, and other related terms. Although, at one time, it may have been appropriate to think of fake news as referring to the veracity of a news story, we now believe that this phrase has been irredeemably polarized in our current political and media climate. As politicians have implemented a political strategy of labeling news sources that do not support their positions as unreliable or fake news, whereas sources that support their positions are labeled reliable or not fake, the term has lost all connection to the actual veracity of the information presented, rendering it meaningless for use in academic classification. We have therefore explicitly avoided the term fake news throughout this paper and instead use the more objectively verifiable terms "true" or "false" news. Although the terms fake news and misinformation also imply a willful distortion of the truth, we do not make any claims about the intent of the purveyors of the information in our analyses. We instead focus our attention on veracity and stories that have been verified as true or false.

We also purposefully adopt a broad definition of the term news. Rather than defining what constitutes news on the basis of the institutional source of the assertions in a story, we refer to any asserted claim made on Twitter as news (we defend this decision in the supplementary materials section on "reliable sources," section S1.2). We define news as any story or claim with an assertion in it and a rumor as the social phenomena of a news story or claim spreading or diffusing through the Twitter network. That is, rumors are inherently social and involve the sharing of claims between people. News, on the other hand, is an assertion with claims, whether it is shared or not.

A rumor cascade begins on Twitter when a user makes an assertion about a topic in a tweet, which could include written text, photos, or links to articles online. Others then propagate the rumor by retweeting it. A rumor's diffusion process can be characterized as having one or more cascades, which we define as instances of a rumor-spreading pattern that exhibit an unbroken retweet chain with a common, singular origin. For example, an individual could start a rumor cascade by tweeting a story or claim with an assertion in it, and another individual could independently start a second cascade of the same rumor (pertaining to the same story or claim) that is completely independent of the first cascade, except that it pertains to the same story or claim. If they remain independent, they represent two cascades of the same rumor. Cascades can be as small as size one (meaning no one retweeted the original tweet). The number of cascades that make up a rumor is equal to the number of times the story or claim was independently tweeted by a user (not retweeted). So, if a rumor "A" is tweeted by 10 people separately, but not retweeted, it would have 10 cascades, each of size one. Conversely, if a second rumor "B" is independently tweeted by two people and each of those two tweets is retweeted 100 times, the rumor would consist of two cascades, each of size 100.

Here we investigate the differential diffusion of true, false, and mixed (partially true, partially false) news stories using a comprehensive data set of all of the fact-checked rumor cascades that spread on Twitter from its inception in 2006 to 2017. The data include ~126,000 rumor cascades spread by ~3 million people more than 4.5 million times. We sampled all rumor cascades investigated by six independent fact-checking organizations

[1]Massachusetts Institute of Technology (MIT), the Media Lab, E14-526, 75 Amherst Street, Cambridge, MA 02142, USA. [2]MIT, E62-364, 100 Main Street, Cambridge, MA 02142, USA.
*Corresponding author. Email: sinan@mit.edu

(snopes.com, politifact.com, factcheck.org, truthorfiction.com, hoax-slayer.com, and urbanlegends.about.com) by parsing the title, body, and verdict (true, false, or mixed) of each rumor investigation reported on their websites and automatically collecting the cascades corresponding to those rumors on Twitter. The result was a sample of rumor cascades whose veracity had been agreed on by these organizations between 95 and 98% of the time. We cataloged the diffusion of the rumor cascades by collecting all English-language replies to tweets that contained a link to any of the aforementioned websites from 2006 to 2017 and used optical character recognition to extract text from images where needed. For each reply tweet, we extracted the original tweet being replied to and all the retweets of the original tweet. Each retweet cascade represents a rumor propagating on Twitter that has been verified as true or false by the fact-checking organizations (see the supplementary materials for more details on cascade construction). We then quantified the cascades' depth (the number of retweet hops from the origin tweet over time, where a hop is a retweet by a new unique user), size (the number of users involved in the cascade over time), maximum breadth (the maximum number of users involved in the cascade at any depth), and structural virality (*23*) (a measure that interpolates between content spread through a single, large broadcast and that which spreads through multiple generations, with any one individual directly responsible for only a fraction of the total spread) (see the supplementary materials for more detail on the measurement of rumor diffusion).

As a rumor is retweeted, the depth, size, maximum breadth, and structural virality of the cascade increase (Fig. 1A). A greater fraction of false rumors experienced between 1 and 1000 cascades, whereas a greater fraction of true rumors experienced more than 1000 cascades (Fig. 1B); this was also true for rumors based on political news (Fig. 1D). The total number of false rumors peaked at the end of both 2013 and 2015 and again at the end of 2016, corresponding to the last U.S. presidential election (Fig. 1C). The data also show clear increases in the total number of false political rumors during the 2012 and 2016 U.S. presidential elections (Fig. 1E) and a spike in rumors that contained partially true and partially false information during the Russian annexation of Crimea in 2014 (Fig. 1E). Politics was the largest rumor category in our data, with ~45,000 cascades, followed by urban legends, business, terrorism, science, entertainment, and natural disasters (Fig. 1F).

When we analyzed the diffusion dynamics of true and false rumors, we found that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information [Kolmogorov-Smirnov (K-S) tests are reported in tables S3 to S10]. A significantly greater fraction of false cascades than true cascades exceeded a depth of 10, and the top 0.01% of false cascades diffused eight hops deeper into the Twittersphere than the truth, diffusing to depths

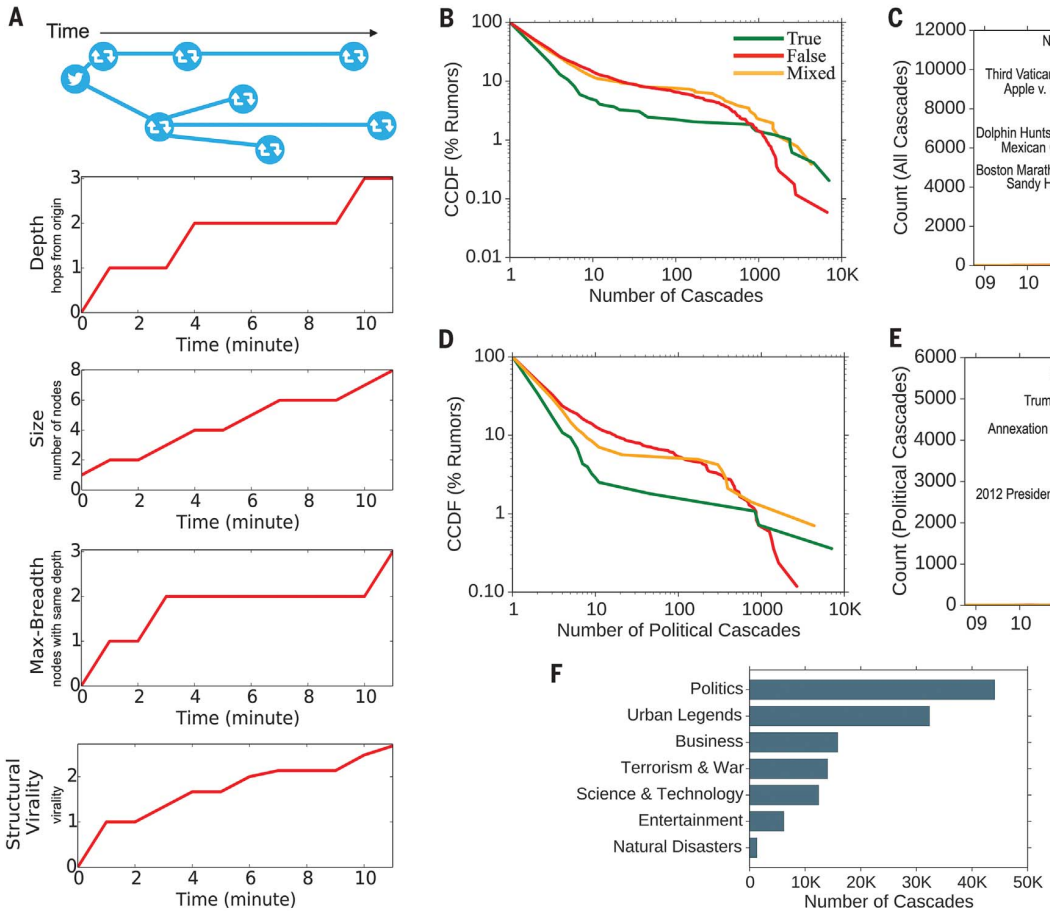


Fig. 1. Rumor cascades. (**A**) An example rumor cascade collected by our method as well as its depth, size, maximum breadth, and structural virality over time. "Nodes" are users. (**B**) The complementary cumulative distribution functions (CCDFs) of true, false, and mixed (partially true and partially false) cascades, measuring the fraction of rumors that exhibit a given number of cascades. (**C**) Quarterly counts of all true, false, and mixed rumor cascades that diffused on Twitter between 2006 and 2017, annotated with example rumors in each category. (**D**) The CCDFs of true, false, and mixed political cascades. (**E**) Quarterly counts of all true, false, and mixed political rumor cascades that diffused on Twitter between 2006 and 2017, annotated with example rumors in each category. (**F**) A histogram of the total number of rumor cascades in our data across the seven most frequent topical categories.

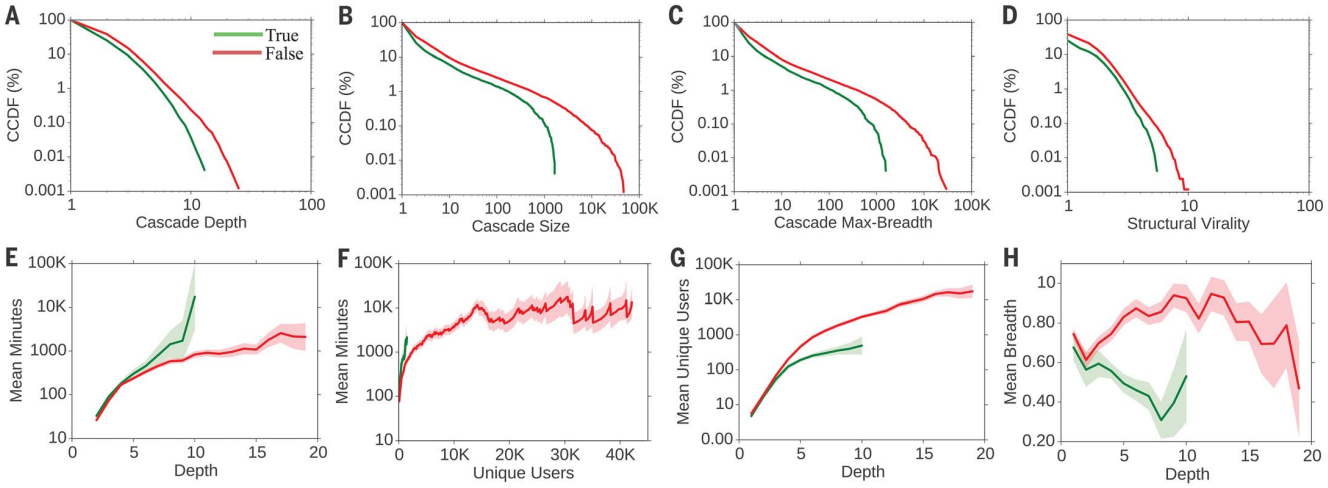


Fig. 2. Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades. (**A**) Depth. (**B**) Size. (**C**) Maximum breadth. (**D**) Structural virality. (**E** and **F**) The number of minutes it takes for true and false rumor cascades to reach any (E) depth and (F) number of unique Twitter users. (**G**) The number of unique Twitter users reached at every depth and (**H**) the mean breadth of true and false rumor cascades at every depth. In (H), plot is lognormal. Standard errors were clustered at the rumor level (i.e., cascades belonging to the same rumor were clustered together; see supplementary materials for additional details).

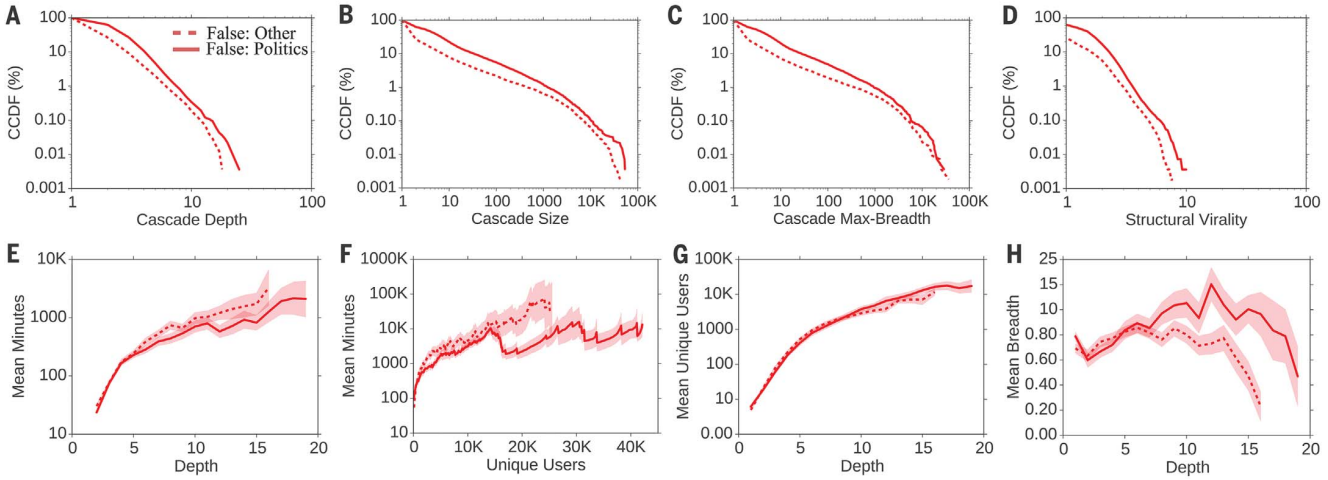


Fig. 3. Complementary cumulative distribution functions (CCDFs) of false political and other types of rumor cascades. (**A**) Depth. (**B**) Size. (**C**) Maximum breadth. (**D**) Structural virality. (**E** and **F**) The number of minutes it takes for false political and other false news cascades to reach any (E) depth and (F) number of unique Twitter users. (**G**) The number of unique Twitter users reached at every depth and (**H**) the mean breadth of these false rumor cascades at every depth. In (H), plot is lognormal. Standard errors were clustered at the rumor level.

greater than 19 hops from the origin tweet (Fig. 2A). Falsehood also reached far more people than the truth. Whereas the truth rarely diffused to more than 1000 people, the top 1% of false-news cascades routinely diffused to between 1000 and 100,000 people (Fig. 2B). Falsehood reached more people at every depth of a cascade than the truth, meaning that many more people retweeted falsehood than they did the truth (Fig. 2C). The spread of falsehood was aided by its virality, meaning that falsehood did not simply spread through broadcast dynamics but rather through peer-to-peer diffusion characterized by a viral branching process (Fig. 2D).

It took the truth about six times as long as falsehood to reach 1500 people (Fig. 2F) and 20 times as long as falsehood to reach a cascade depth of 10 (Fig. 2E). As the truth never diffused beyond a depth of 10, we saw that falsehood reached a depth of 19 nearly 10 times faster than the truth reached a depth of 10 (Fig. 2E). Falsehood also diffused significantly more broadly (Fig. 2H) and was retweeted by more unique users than the truth at every cascade depth (Fig. 2G).

False political news (Fig. 1D) traveled deeper (Fig. 3A) and more broadly (Fig. 3C), reached more people (Fig. 3B), and was more viral than any other category of false information (Fig. 3D). False political news also diffused deeper more quickly (Fig. 3E) and reached more than 20,000 people nearly three times faster than all other types of false news reached 10,000 people (Fig. 3F). Although the other categories of false news reached about the same number of unique users at depths between 1 and 10, false political news routinely reached the most unique users at depths greater than 10 (Fig. 3G). Although all other categories of false news traveled slightly more broadly at shallower depths, false political news traveled more broadly at greater depths, indicating that more-popular false political news items exhibited broader and more-accelerated diffusion dynamics

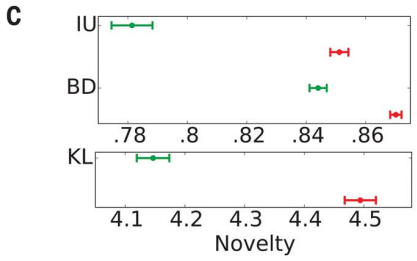
A

	median		mean		mean (log)		stdv (log)		ks-test
	false	true	false	true	false	true	false	true	
followers	410	466	2234	5240	2.62	2.68	0.69	0.88	D=0.104, p∼0.0
followees	383	509	1002	1707	2.59	2.72	0.85	0.96	D=0.136, p∼0.0
verified	0	0	0.002	0.006	nd	nd	nd	nd	D=0.005, p<0.001
engagement	9.52	9.54	19.70	24.65	0.91	0.90	0.65	0.76	D=0.054, p∼0.0
account age	982	1214	1072	1269	2.90	2.97	0.39	0.42	D=0.125, p∼0.0

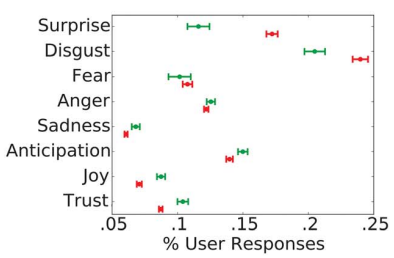
B

	coef	odds ratio	std err	z	P>|z|	[95.0% Conf. Int.]	
account age	0.0002	1.000160	2.07e-05	7.759	0.000	0.000	0.000
engagement	0.0066	1.006648	0.000	18.019	0.000	0.006	0.007
falsehood	0.5350	1.707489	0.084	6.366	0.000	0.370	0.700
followees	-1.639e-05	0.999984	8.73e-06	-1.877	0.060	-3.35e-05	7.22e-07
followers	5.192e-05	1.000052	7.77e-06	6.682	0.000	3.67e-05	6.72e-05
intercept	-2.3941	0.091257	0.072	-33.437	0.000	-2.534	-2.254
verified	1.4261	4.162467	0.090	15.915	0.000	1.250	1.602

C



D



E

	mean		variance		ks-test
	false	true	false	true	
IU	0.85	0.78	0.0052	0.0072	D=0.457, p∼0.0
KL	4.49	4.15	0.1618	0.0948	D=0.433, p∼0.0
BD	0.87	0.84	0.0008	0.0008	D=0.415, p∼0.0

F

	mean		variance		ks-test
	false	true	false	true	
surprise	0.172	0.116	0.0167	0.0072	D=0.205, p∼0.0
disgust	0.240	0.205	0.0260	0.0227	D=0.102, p∼0.0
fear	0.108	0.102	0.0120	0.0095	D=0.021, p∼0.164
anger	0.122	0.126	0.0074	0.0111	D=0.023, p∼0.078
sadness	0.061	0.068	0.0038	0.0065	D=0.037, p∼0.0
anticipation	0.140	0.150	0.0093	0.0154	D=0.038, p∼0.0
joy	0.071	0.087	0.0054	0.0104	D=0.061, p∼0.0
trust	0.087	0.104	0.0058	0.0119	D=0.060, p∼0.0

Fig. 4. Models estimating correlates of news diffusion, the novelty of true and false news, and the emotional content of replies to news. (**A**) Descriptive statistics on users who participated in true and false rumor cascades as well as K-S tests of the differences in the distributions of these measures across true and false rumor cascades. (**B**) Results of a logistic regression model estimating users' likelihood of retweeting a rumor as a function of variables shown at the left. coeff, logit coefficient; *z*, *z* score. (**C**) Differences in the information uniqueness (IU), scaled Bhattacharyya distance (BD), and K-L divergence (KL) of true (green) and false (red) rumor tweets compared to the corpus of prior tweets the user was exposed to in the 60 days before retweeting the rumor tweet. (**D**) The emotional content of replies to true (green) and false (red) rumor tweets across seven dimensions categorized by the NRC. (**E**) Mean and variance of the IU, KL, and BD of true and false rumor tweets compared to the corpus of prior tweets the user has seen in the 60 days before seeing the rumor tweet as well as K-S tests of their differences across true and false rumors. (**F**) Mean and variance of the emotional content of replies to true and false rumor tweets across seven dimensions categorized by the NRC as well as K-S tests of their differences across true and false rumors. All standard errors are clustered at the rumor level, and all models are estimated with cluster-robust standard errors at the rumor level.

(Fig. 3H). Analysis of all news categories showed that news about politics, urban legends, and science spread to the most people, whereas news about politics and urban legends spread the fastest and were the most viral in terms of their structural virality (see fig. S11 for detailed comparisons across all topics).

One might suspect that structural elements of the network or individual characteristics of the users involved in the cascades explain why falsity travels with greater velocity than the truth. Perhaps those who spread falsity "followed" more people, had more followers, tweeted more often, were more often "verified" users, or had been on Twitter longer. But when we compared users involved in true and false rumor cascades, we found that the opposite was true in every case. Users who spread false news had significantly fewer followers (K-S test = 0.104, $P \sim 0.0$), followed significantly fewer people (K-S test = 0.136, $P \sim 0.0$), were significantly less active on Twitter (K-S test = 0.054, $P \sim 0.0$), were verified significantly less often (K-S test = 0.004, $P < 0.001$), and had been on Twitter for significantly less time (K-S test = 0.125, $P \sim 0.0$) (Fig. 4A). Falsehood diffused farther and faster than the truth despite these differences, not because of them.

When we estimated a model of the likelihood of retweeting, we found that falsehoods were 70% more likely to be retweeted than the truth (Wald chi-square test, $P \sim 0.0$), even when controlling for the account age, activity level, and number of followers and followees of the original tweeter, as well as whether the original tweeter was a verified user (Fig. 4B). Because user characteristics and network structure could not explain the differential diffusion of truth and falsity, we sought alternative explanations for the differences in their diffusion dynamics.

One alternative explanation emerges from information theory and Bayesian decision theory. Novelty attracts human attention (*24*), contributes to productive decision-making (*25*), and encourages information sharing (*26*) because novelty updates our understanding of the world. When information is novel, it is not only surprising, but also more valuable, both from an information theoretic perspective [in that it provides the greatest aid to decision-making (*25*)] and from a social perspective [in that it conveys social status on one that is "in the know" or has access to unique "inside" information (*26*)]. We therefore tested whether falsity was more novel than the truth and whether Twitter users were more likely to retweet information that was more novel.

To assess novelty, we randomly selected ~5000 users who propagated true and false rumors and extracted a random sample of ~25,000 tweets that they were exposed to in the 60 days prior to their decision to retweet a rumor. We then specified a latent Dirichlet Allocation Topic model (*27*), with 200 topics and trained on 10 million English-language tweets, to calculate the information distance between the rumor tweets and all the prior tweets that users were exposed to before retweeting the rumor tweets. This generated a probability distribution over the 200 topics for each tweet in our data set. We then measured how novel the information in the true and false rumors was by comparing the topic distributions of the rumor tweets with the topic distributions of the tweets to which users were exposed in the 60 days before their retweet. We found that false rumors were significantly more

novel than the truth across all novelty metrics, displaying significantly higher information uniqueness (K-S test = 0.457, $P \sim 0.0$) (*28*), Kullback-Leibler (K-L) divergence (K-S test = 0.433, $P \sim 0.0$) (*29*), and Bhattacharyya distance (K-S test = 0.415, $P \sim 0.0$) (which is similar to the Hellinger distance) (*30*). The last two metrics measure differences between probability distributions representing the topical content of the incoming tweet and the corpus of previous tweets to which users were exposed.

Although false rumors were measurably more novel than true rumors, users may not have perceived them as such. We therefore assessed users' perceptions of the information contained in true and false rumors by comparing the emotional content of replies to true and false rumors. We categorized the emotion in the replies by using the leading lexicon curated by the National Research Council Canada (NRC), which provides a comprehensive list of ~140,000 English words and their associations with eight emotions based on Plutchik's (*31*) work on basic emotion—anger, fear, anticipation, trust, surprise, sadness, joy, and disgust (*32*)—and a list of ~32,000 Twitter hashtags and their weighted associations with the same emotions (*33*). We removed stop words and URLs from the reply tweets and calculated the fraction of words in the tweets that related to each of the eight emotions, creating a vector of emotion weights for each reply that summed to one across the emotions. We found that false rumors inspired replies expressing greater surprise (K-S test = 0.205, $P \sim 0.0$), corroborating the novelty hypothesis, and greater disgust (K-S test = 0.102, $P \sim 0.0$), whereas the truth inspired replies that expressed greater sadness (K-S test = 0.037, $P \sim 0.0$), anticipation (K-S test = 0.038, $P \sim 0.0$), joy (K-S test = 0.061, $P \sim 0.0$), and trust (K-S test = 0.060, $P \sim 0.0$) (Fig. 4, D and F). The emotions expressed in reply to falsehoods may illuminate additional factors, beyond novelty, that inspire people to share false news. Although we cannot claim that novelty causes retweets or that novelty is the only reason why false news is retweeted more often, we do find that false news is more novel and that novel information is more likely to be retweeted.

Numerous diagnostic statistics and manipulation checks validated our results and confirmed their robustness. First, as there were multiple cascades for every true and false rumor, the variance of and error terms associated with cascades corresponding to the same rumor will be correlated. We therefore specified cluster-robust standard errors and calculated all variance statistics clustered at the rumor level. We tested the robustness of our findings to this specification by comparing analyses with and without clustered errors and found that, although clustering reduced the precision of our estimates as expected, the directions, magnitudes, and significance of our results did not change, and chi-square ($P \sim 0.0$) and deviance (d) goodness-of-fit tests ($d = 3.4649 \times 10^{-6}$, $P \sim 1.0$) indicate that the models are well specified (see supplementary materials for more detail).

Second, a selection bias may arise from the restriction of our sample to tweets fact checked by the six organizations we relied on. Fact checking may select certain types of rumors or draw additional attention to them. To validate the robustness of our analysis to this selection and the generalizability of our results to all true and false rumor cascades, we independently verified a second sample of rumor cascades that were not verified by any fact-checking organization. These rumors were fact checked by three undergraduate students at Massachusetts Institute of Technology (MIT) and Wellesley College. We trained the students to detect and investigate rumors with our automated rumor-detection algorithm running on 3 million English-language tweets from 2016 (*34*). The undergraduate annotators investigated the veracity of the detected rumors using simple search queries on the web. We asked them to label the rumors as true, false, or mixed on the basis of their research and to discard all rumors previously investigated by one of the fact-checking organizations. The annotators, who worked independently and were not aware of one another, agreed on the veracity of 90% of the 13,240 rumor cascades that they investigated and achieved a Fleiss' kappa of 0.88. When we compared the diffusion dynamics of the true and false rumors that the annotators agreed on, we found results nearly identical to those estimated with our main data set (see fig. S17). False rumors in the robustness data set had greater depth (K-S test = 0.139, $P \sim 0.0$), size (K-S test = 0.131, $P \sim 0.0$), maximum breadth (K-S test = 0.139, $P \sim 0.0$), structural virality (K-S test = 0.066, $P \sim 0.0$), and speed (fig. S17) and a greater number of unique users at each depth (fig. S17). When we broadened the analysis to include majority-rule labeling, rather than unanimity, we again found the same results (see supplementary materials for results using majority-rule labeling).

Third, although the differential diffusion of truth and falsity is interesting with or without robot, or bot, activity, one may worry that our conclusions about human judgment may be biased by the presence of bots in our analysis. We therefore used a sophisticated bot-detection algorithm (*35*) to identify and remove all bots before running the analysis. When we added bot traffic back into the analysis, we found that none of our main conclusions changed—false news still spread farther, faster, deeper, and more broadly than the truth in all categories of information. The results remained the same when we removed all tweet cascades started by bots, including human retweets of original bot tweets (see supplementary materials, section S8.3) and when we used a second, independent bot-detection algorithm (see supplementary materials, section S8.3.5) and varied the algorithm's sensitivity threshold to verify the robustness of our analysis (see supplementary materials, section S8.3.4). Although the inclusion of bots, as measured by the two state-of-the-art bot-detection algorithms we used in our analysis, accelerated the spread of both true and false news, it affected their spread roughly equally. This suggests that false news spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it.

Finally, more research on the behavioral explanations of differences in the diffusion of true and false news is clearly warranted. In particular, more robust identification of the factors of human judgment that drive the spread of true and false news online requires more direct interaction with users through interviews, surveys, lab experiments, and even neuroimaging. We encourage these and other approaches to the investigation of the factors of human judgment that drive the spread of true and false news in future work.

False news can drive the misallocation of resources during terror attacks and natural disasters, the misalignment of business investments, and misinformed elections. Unfortunately, although the amount of false news online is clearly increasing (Fig. 1, C and E), the scientific understanding of how and why false news spreads is currently based on ad hoc rather than large-scale systematic analyses. Our analysis of all the verified true and false rumors that spread on Twitter confirms that false news spreads more pervasively than the truth online. It also overturns conventional wisdom about how false news spreads. Though one might expect network structure and individual characteristics of spreaders to favor and promote false news, the opposite is true. The greater likelihood of people to retweet falsity more than the truth is what drives the spread of false news, despite network and individual factors that favor the truth. Furthermore, although recent testimony before congressional committees on misinformation in the United States has focused on the role of bots in spreading false news (*36*), we conclude that human behavior contributes more to the differential spread of falsity and truth than automated robots do. This implies that misinformation-containment policies should also emphasize behavioral interventions, like labeling and incentives to dissuade the spread of misinformation, rather than focusing exclusively on curtailing bots. Understanding how false news spreads is the first step toward containing it. We hope our work inspires more large-scale research into the causes and consequences of the spread of false news as well as its potential cures.

REFERENCES AND NOTES

1. L. J. Savage, *J. Am. Stat. Assoc.* **46**, 55–67 (1951).
2. H. A. Simon, *The New Science of Management Decision* (Harper & Brothers Publishers, New York, 1960).
3. R. Wedgwood, *Noûs* **36**, 267–297 (2002).
4. E. Fehr, U. Fischbacher, *Nature* **425**, 785–791 (2003).
5. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379–423 (1948).
6. S. Bikhchandani, D. Hirshleifer, I. Welch, *J. Polit. Econ.* **100**, 992–1026 (1992).
7. K. Rapoza, "Can 'fake news' impact the stock market?" *Forbes*, 26 February 2017; www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/.
8. M. Mendoza, B. Poblete, C. Castillo, in *Proceedings of the First Workshop on Social Media Analytics* (Association for Computing Machinery, ACM, 2010), pp. 71–79.
9. A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, in *Proceedings of the 22nd International Conference on World Wide Web* (ACM, 2010), pp. 729–736.

10. K. Starbird, J. Maddock, M. Orand, P. Achterman, R. M. Mason, in *iConference 2014 Proceedings* (iSchools, 2014).
11. J. Gottfried, E. Shearer, "News use across social media platforms," Pew Research Center, 26 May 2016; www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/.
12. C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on Facebook," *BuzzFeed News*, 16 November 2016; www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook/.
13. M. De Domenico, A. Lima, P. Mougel, M. Musolesi, *Sci. Rep.* **3**, 2980 (2013).
14. O. Oh, K. H. Kwon, H. R. Rao, in *Proceedings of the International Conference on Information Systems* (International Conference on Information Systems, ICIS, paper 231, 2010).
15. M. Tambuscio, G. Ruffo, A. Flammini, F. Menczer, in *Proceedings of the 24th International Conference on World Wide Web* (ACM, 2015), pp. 977–982.
16. Z. Zhao, P. Resnick, Q. Mei, in *Proceedings of the 24th International Conference on World Wide Web* (ACM, 2015), pp. 1395–1405.
17. M. Gupta, P. Zhao, J. Han, in *Proceedings of the 2012 Society for Industrial and Applied Mathematics International Conference on Data Mining* (Society for Industrial and Applied Mathematics, SIAM, 2012), pp. 153–164.
18. G. L. Ciampaglia *et al.*, *PLOS ONE* **10**, e0128193 (2015).
19. A. Friggeri, L. A. Adamic, D. Eckles, J. Cheng, in *Proceedings of the International Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, AAAI, 2014)
20. M. Del Vicario *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 554–559 (2016).
21. A. Bessi *et al.*, *PLOS ONE* **10**, e0118093 (2015).
22. Friggeri *et al.* (*19*) do evaluate two metrics of diffusion: depth, which shows little difference between true and false rumors, and shares per rumor, which is higher for true rumors than it is for false rumors. Although these results are important, they are not definitive owing to the smaller sample size of the study; the early timing of the sample, which misses the rise of false news after 2013; and the fact that more shares per rumor do not necessarily equate to deeper, broader, or more rapid diffusion.
23. S. Goel, A. Anderson, J. Hofman, D. J. Watts, *Manage. Sci.* **62**, 180–196 (2015).
24. L. Itti, P. Baldi, *Vision Res.* **49**, 1295–1306 (2009).
25. S. Aral, M. Van Alstyne, *Am. J. Sociol.* **117**, 90–171 (2011).
26. J. Berger, K. L. Milkman, *J. Mark. Res.* **49**, 192–205 (2012).
27. D. M. Blei, A. Y. Ng, M. I. Jordan, *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
28. S. Aral, P. Dhillon, "Unpacking novelty: The anatomy of vision advantages," Working paper, MIT–Sloan School of Management, Cambridge, MA, 22 June 2016; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2388254.
29. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (Wiley, 2012).
30. T. Kailath, *IEEE Trans. Commun. Technol.* **15**, 52–60 (1967).
31. R. Plutchik, *Am. Sci.* **89**, 344–350 (2001).
32. S. M. Mohammad, P. D. Turney, *Comput. Intell.* **29**, 436–465 (2013).
33. S. M. Mohammad, S. Kiritchenko, *Comput. Intell.* **31**, 301–326 (2015).
34. S. Vosoughi, D. Roy, in *Proceedings of the 10th International AAAI Conference on Weblogs and Social Media* (AAAI, 2016), pp. 707–710.
35. C. A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, in *Proceedings of the 25th International Conference Companion on World Wide Web* (ACM, 2016), pp. 273–274.
36. For example, this is an argument made in recent testimony by Clint Watts—Robert A. Fox Fellow at the Foreign Policy Research Institute and Senior Fellow at the Center for Cyber and Homeland Security at George Washington University—given during the U.S. Senate Select Committee on Intelligence hearing on "Disinformation: A Primer in Russian Active Measures and Influence Campaigns" on 30 March 2017; www.intelligence.senate.gov/sites/default/files/documents/os-cwatts-033017.pdf.

ACKNOWLEDGMENTS

We are indebted to Twitter for providing funding and access to the data. We are also grateful to members of the MIT research community for invaluable discussions. The research was approved by the MIT institutional review board. The analysis code is freely available at https://goo.gl/forms/AKIlZujpexhN7fY33. The entire data set is also available, from the same link, upon signing an access agreement stating that (i) you shall only use the data set for the purpose of validating the results of the MIT study and for no other purpose; (ii) you shall not attempt to identify, reidentify, or otherwise deanonymize the data set; and (iii) you shall not further share, distribute, publish, or otherwise disseminate the data set. Those who wish to use the data for any other purposes can contact and make a separate agreement with Twitter.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6380/1146/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S39
References (*37*–*75*)

14 September 2017; accepted 19 January 2018
10.1126/science.aap9559