

语言特征到社交影响：极性情感推文可读性与传播关系的实证研究

摘要：在日益情绪化的网络环境中，理解社交媒体信息的传播机制已成为把握在线生态的关键。已有研究基于 X 平台的推文，强调了极性情感（高度积极或消极的情感）对传播的强大驱动力；部分观点还认为，推文可读性越高，其传播能力越强。然而，在极性情感日益普遍的语境下，这一关系亟需重新审视。本研究从实证角度出发，探讨了极性情感推文的可读性特征，并分析可读性在该语境中对传播效果的影响。研究结合情感分类模型与多项可读性、传播指标，对七万余条推文开展了差异性、相关性分析。结果表明，极性情感推文在用词长度、音节数量、词汇难度等方面可读性略高于非极性内容，但二者在实际表达中差异有限。同时，在极性情感语境下，推文可读性与传播效果之间几乎不存在实际关联。该发现揭示了可读性作用可能存在的条件边界，即在强烈情绪内容中，可读性对传播的作用可能受情感倾向、身份立场等因素削弱。本研究进一步提出了相应的理论、实践启示，并讨论了研究的局限性。

关键词：社交媒体；在线传播；文本可读性；极性情感

1. 引言

社交媒体信息的传播机制，是理解当前在线舆论生态的核心议题。以 X（原推特）这一全球性、实时性的社交媒体平台为代表，众多研究探索了其中社交内容“推文”在传播时的驱动因素（Stieglitz & Dang-Xuan, 2013; Brady 等, 2017; Leonhardt & Makienko, 2018; Pancer 等, 2019; Firouzjaei & Özdemir, 2020; Rathje, Van Bavel & van der Linden, 2021）。一方面，高度积极、消极的情绪，或“极性情感”，在网络中逐渐增多，并被证实对推文传播有显著影响（Berger & Milkman, 2012; Stieglitz & Dang-Xuan, 2013; Brady 等, 2017; Rathje, Van Bavel & van der Linden, 2021）。另一方面，“文本可读性”通过降低用户处理推文的认知负荷，营造“认知流畅效应”，同样成为了传播的关键因素（Alter & Oppenheimer, 2009）。研究反映，可读性更高、更易被理解的推文，更容易引发用户点赞、回复、转发（Leonhardt & Makienko, 2018; Pancer 等, 2019; Firouzjaei & Özdemir, 2020）。然而，该发现在强烈情绪主导的内容环境中是否依然成立？其可能存在局限性。极性情感内容下，社区语言习惯（Danescu-Niculescu-Mizil 等, 2013）、情感“认知捷径”（Metzger & Flanagin, 2013）、群体认同动机（Crockett, 2017）均可能模糊可读性对推文传播的影响。因此，在极性情感这一特殊但日益普遍的语境下，推文可读性与传播的关系亟待重新审视。本研究旨在从实证角度考察极性情感推文的可读性特征，并探讨极性情感内容中，可读性是否仍对推文传播存在显著影响。在此基础上，本文有望探讨情感内容下可读性对推文传播的作用边界，为理

解日益情绪化的网络传播生态提供新的视角。研究基于以下问题展开。

1. 极性情感推文与非极性情感推文在可读性上是否存在差异？这能反映前者哪些语言特征？
2. 极性情感内容下，推文可读性与其传播效果是否相关？

2. 文献综述

2.1 极性情感与社媒推文传播

在当前社交媒体环境中，情绪内容逐渐增多。同时，情感对内容信息传播有显著的驱动作用。研究表明，社交媒体中的情绪内容会在用户间相互传染，影响用户的表达倾向（Kramer 等，2014）。且在心理学视角下，当内容引发用户“高唤醒情绪”，即高度积极（敬叹）或消极（愤怒、焦虑）的情感时，其更能引发病毒式传播（Berger & Milkman, 2012, p. 1）。这样高度积极、消极的情感，便可称为社交媒体内容中的极性情感。极性情感内容也因其情绪张力、传播潜力，受到了众多学者的关注。

社媒研究中，X（原推特）因其全球用户基础、实时信息传播等特征，被视为重要研究平台，其社交内容“推文”也作为关键研究对象。因此，部分研究（Stieglitz & Dang-Xuan, 2013; Brady 等, 2017; Rathje, Van Bavel & van der Linden, 2021）以推文为基础，深入探讨了极性情感内容的传播特性。其中，Stieglitz 与 Dang-Xuan（2013）在政治语境下发现，情绪化的推文相比中立推文往往会被更频繁、更迅速地转发（p. 241）。与之类似，Rathje 等（2021）则发现，推文若包含对政治异己的“敌意”，尤其是“愤怒”情感，其互动量将显著提升（p. 7）。Brady 等（2017）则进一步提出“道德传染”概念，指出推文中每增加一个“道德情感词”，其传播量平均增加 20%（p.7316）。这些研究共同表明，带有极性情感内容的推文，同样更容易引发用户互动与病毒式传播。此类推文在社交网络的扩散，也长期作为社媒研究的焦点之一。

2.2 文本可读性与社媒推文传播

除了情感内容，推文传播与文本可读性也存在紧密关联。研究发现，当文本由熟悉词汇和较短句子构成时，读者更容易处理这些信息，产生“认知流畅效应”（Alter & Oppenheimer, 2009）。易阅读的内容也因此在社交媒体上容易获得更多的点赞、评论、分享（Pancer 等, 2019）。具体来说，可读性关注的是“表达方式”上词、句法结构对读者文本理解的影响（Dale & Chall, 1948, p. 3）。其被 Klare（1963）定义为“写作风格影响下理解文本的简易程度”。通常文本越容易理解，其可读性也就越高。在可读性的衡量计算上，相关研究涉及音节数、单词数、句子数、复杂词（音节数 ≥ 3 ）（Gunning, 1952）、长词（字母数 ≥ 7 ）（Björnsson, 1968）、难词（四年级 3000 词表外的单词）（Dale & Chall, 1948）等语言特征。基于这些特征，还提出了 FRE（Flesch, 1948）、SMOG（McLaughlin, 1969）、

Dale-Chall (Chall & Dale, 1995) 等可读性计算指标。基于上述计算方法, 部分研究 (Davenport & DeLine, 2014; Davidson 等, 2017; Leonhardt & Makienko, 2018; Jacob & Uitdenbogerd, 2019; Pancer 等, 2019; Firouzjaei & Özdemir, 2020) 也从可读性角度出发, 分析了推文相关文本特征。

这些研究中, Leonhardt & Makienko (2018)、Pancer 等 (2019)、Firouzjaei & Özdemir (2020) 的发现均对推文可读性及其传播的关系有所启示。首先, Leonhardt 与 Makienko (2017) 结合 FRE 指标分析了大量品牌推文, 发现推文可读性与点赞、转发、回复数量均呈显著正相关。类似地, Pancer 等 (2019) 参考 Dale-Chall 指标, 同样发现易于阅读的社交媒体内容会促进用户的评论、分享行为。同时, Firouzjaei 与 Özdemir (2020) 使用了词、句、语篇层面的多项可读性指标, 发现政治语境下易读推文能吸引用户积极参与相关讨论。这些研究都反映出, 可读性更高、更易被理解的推文, 更容易引发用户互动与社交媒体传播。这也再次强调了可读性对推文传播的重要性。

2.3 当前研究

一方面, 已有研究强调了极性情感内容在推文传播中的关键作用 (Stieglitz & Dang-Xuan, 2013; Brady 等, 2017; Rathje, Van Bavel & van der Linden, 2021); 另一方面, 推文可读性也被认为与讨论度、传播度呈显著正相关 (Leonhardt & Makienko, 2018; Pancer 等, 2019; Firouzjaei & Özdemir, 2020)。然而, 已有研究就可读性与推文传播的发现, 可能在“极性情感”的特殊语境下存在局限。首先, 用户为适应极性情感内容的相关社区讨论, 可能改变语言习惯 (Danescu-Niculescu-Mizil 等, 2013)。这可能模糊可读性差异与真实影响。其次, 用户的传播行为可能依赖于发布者身份、情绪倾向等“认知捷径” (Metzger & Flanagin, 2013)。直观的情绪表达可能降低用户的内容敏感性, 削弱可读性作用; 最后, 传播极性情感内容的行为, 有时并非基于信息交换, 而是表达立场、寻求群体认同的社会信号 (Crockett, 2017)。站队行为可能压倒内容本身的传播优势。可见, 可读性在极性情感语境下对推文的传播作用尚不确定, 仍需实证检验。因此, 本文旨在从实证角度考察极性情感推文的可读性特征, 并探讨极性情感内容中, 可读性是否仍对推文传播存在显著影响。

3. 研究方法

3.1 研究数据

本研究以 X (原推特) 为目标平台爬取语料。为确保语料包含充分极性情感内容, 我们分别爬取了“性少数群体”、“公共健康”、“战争与地缘政治”、“女性权利”、“生态环境”、“种族问题”六个话题下的推文。推文检索时, 我们设置了三项条件。第一, 推文检索词为话题相关#Hashtag。#Hashtag 具有将社交行为结构化, 通过“可搜索对话”形成公共讨论空间的功能 (Zappavigna & Ross, 2024)。

因此, #Hashtag 便于我们聚焦目标议题和社群; 第二, 推文检索按照“最新”机制排序。这可以避免“热门”排序机制下平台优先推荐高传播内容, 导致数据偏向; 第三, 推文时间无范围限制。由于大多数推文的传播基本在一天内结束 (Pfeffer 等, 2023), 我们假设时间变量对多数推文传播力影响较小。针对每篇推文, 我们还爬取了语言、来源链接、点赞数、回复数、转发数、引用数等信息。

按上述条件爬取推文后, 我们进行了数据清洗。具体流程包括: 删除非英文推文; 删来源相同的重复推文; 使用正则表达式删除推文中的表情、#Hashtag、@Mention、链接; 内容长度筛选。此处, 控制推文长度有助于保证后续推文可处理、可比较性。一方面, 为保证推文基本内容量, 我们要求推文最低应包含 10 个单词; 另一方面, 为控制推文长度, 我们根据四分位距 (IQR) 确定了推文最大词数。IQR 为上四分位数 (Q3) 与下四分位数 (Q1) 的差。通常来说, 我们将 $Q3 + 1.5 * IQR$ 的结果视作数据的上界, 用于划分潜在离群值。在本研究中, 推文长度的上界为 58。因此, 我们过滤了长度高于 58 的推文。最终, 研究得到有效推文共 70472 篇, 如表一所示。

表一：话题、检索词与推文数量

话题	检索词	推文数量
性少数	#LGBTRights #LGBTQRights #TransRights #TransRightsAreHumanRights #LGBTQPolicy #StopTransHate #SexualMinority	5,504
公共健康	#HealthEquity #HealthJustice #VaccinesWork #HealthcareIsAHumanRight #VaccineMandates #AntiMandate #HealthFreedom #MedicalFreedom	13,962
战争与地缘政治	#RussiaUkraineWar #UkraineRussiaWar #PalestineIsraelWar #IsraelPalestineWar #FreePalestine #Gaza #PakistanIndiaWar #IndiaPakistanWar #GazaGenocide #WarCrimes	8,414
女性权利	#WomansRights #WomensRights #WomanRights #WomenRights #GenderEquality #Feminism #ReproductiveRights #AbortionRights #EndViolenceAgainstWomen #BelieveWomen	17,322
生态环境	#ClimateCrisis #ActOnClimate #EcoJustice #StopEcocide #ClimateJustice #ClimateChange #ExtinctionRebellion	18,426
种族问题	#RacialJustice #BlackLivesMatter #StopAsianHate #AntiRacism #Racism #Racist #WhitePrivilege #EndRacism #SystemicRacism	6,844
总计		70,472

3.2 数据处理与分析

为回答当前研究问题, 本研究基于所得推文分别进行了情感分析、可读性计算、差异性检验与相关性检验。

3.2.1 情感分析

为了识别极性情感推文, 我们使用预训练语言模型 robust-sentiment-analysis (Tabularisai, 2024) 进行了情感分析。使用该模型基于两点考量。首先, 该模型“尤其适合社交媒体分析”(Tabularisai, 2024), 在验证集得到 95% 的准确率;

同时，该模型输出含 Very Positive、Positive、Neutral、Negative、Very Negative 的五分类结果，便于细分和判别极性情感。我们认为带有极性情感倾向的文本更加趋向两级。因此，我们将预测结果为 Very Positive 和 Very Negative 的推文归为极性情感推文，其余三类视作非极性情感推文，构建情感变量。

3.2.2 可读性计算

为衡量推文可读性，研究首先需要确定可用指标。然而，选择适合推文的可读性指标存在两大问题。一方面，大多可读性公式需要文本包含至少 100 个单词，才能确保结果可信度 (Lenzner, 2013)。而社交媒体推文以短文本为特征，难以满足可读性公式计算的文本长度要求；同时，推文常存在标点符号不严谨、难以分句的问题 (Davenport & DeLine, 2014)，难以提供公式计算的句子信息。先前部分研究 (Leonhardt & Makienko, 2018; Jacob & Uitdenbogerd, 2019; Pancer 等, 2019; Firouzjaei & Özdemir, 2020) 直接使用过 FRE、Dale-Chall 等可读性指标。而为应对推文难以分句的情况，Davenport & DeLine (2014)、Davidson 等 (2017) 则对 FRE 指标进行了修改，将文本句数固定为一。总体来看，这些方法或受分句影响，或难以避免可读性公式在短文本中长度敏感的问题。因此，本研究主要关注可读性计算中词汇层面的语言特征，来避免推文分句与文本长度限制。在指标选择上，本研究从相关研究、常用指标、可用工具 (Dale & Chall, 1948; Gunning, 1952; Björnsson, 1968; Firouzjaei & Özdemir, 2020; Leileibama, 2023; Andreasvc, 2025) 中选取了“每词字符数”、“每词音节数”、“复杂词占比”、“长词占比”、“难词占比”等可读性指标（见表二）。通常来说，这些指标的值越低，推文的可读性越高。在指标计算时，我们使用了 readability 0.3.2 (Andreasvc, 2025) 与 syntok 1.4.4 (Leitner, 2022) 等工具。其中，readability 为可读性计算工具，能够基于文本特征得到研究所需可读性指标；syntok 为 readability 推荐使用的分词工具，用于推文文本分词。

表二：词汇层面可读性指标

词汇层面可读性指标	含义	与可读性的关系/指标方向性
每词音节数	平均每个词的音节数	越高表示词越复杂，越难读
每词字符数	平均每个词的字符数	越高表示词越长，越难读
复杂词占比	音节≥3 的词占比	复杂词越多，越难读
长词占比	字符≥7 的词占比	长词越多，越难读
难词占比	Dale-Chall 四年级 3000 词表外的词占比	难词越多，越难读

3.2.3 差异性检验

为探讨极性情感推文与非极性情感推文是否存在可读性差异，研究对两类推文在各项可读性指标上进行了差异性检验。首先，我们使用 pingouin 0.5.5

(Vallat, 2018) 在两组推文各项可读性指标上进行了 Shapiro-Wilk 正态性、Levene 方差齐性检验。检验发现，各项可读性指标均不服从正态分布($p > 0.05$)、且不满足方差齐性 ($p > 0.05$)。因此，我们选用更为稳健的 Mann-Whitney U 非参数检验方法，比较两组推文在可读性指标上的差异。相应地，我们以中位数代表指标值， p 值报告差异显著性，rank-biserial r 值报告差异数量。

3.2.4 相关性检验

为分析极性情感推文的可读性与传播的关系，研究对极性情感推文的各项可读性指标与传播能力进行了相关性检验。为衡量推文传播力，我们依据 Xiao & Chen (2025) 的方法，基于推文点赞数、回复数、转发数、引用数等指标计算了综合传播得分。公式如下。

$$S_{\text{传播得分}} = w_1 * \text{点赞数} + w_2 * \text{回复数} + w_3 * \text{转发数} + w_4 * \text{引用数}$$

其中， w 为各指标的权重。该权重基于指标信息熵计算得出。其核心思想是，当指标数值分布越不均匀，其包含有效信息就越多，在评价体系中就越重要。具体来说，我们先将每项指标数据归一化到[0.02, 1] (p. 191)。此处设定 0.02 的下限是为确保信息熵对数计算的有效性；基于归一化的数据，我们计算了每项指标的信息熵值；通过信息熵，便可计算指标的信息冗余度 (1-信息熵)，即指标的有效信息程度；每项指标信息冗余度占指标冗余度总和的比例，便为该指标的权重。这种方法确保了赋权的客观性，能够更有意义地反映各项互动指标对传播的实际贡献 (p. 198)。接下来，通过对各指标加权求和，便能得到推文的传播得分。该得分越高，表明推文的传播能力越强。

为确定可读性指标与传播得分的相关性检验方法，我们使用 Shapiro-Wilk 方法分析了数据正态性。由于可读性指标与传播得分均不满足正态分布 ($p > 0.05$)，我们使用 Spearman 方法对两者进行相关性检验。同时，我们以 p 值报告显著性， ρ 值报告相关程度。

4. 研究结果

4.1 极性情感推文与非极性情感推文在可读性上是否存在差异？这能反映前者哪些语言特征？

研究通过 Mann-Whitney U 方法，检验了极性情感推文和非极性情感推文“每词字符数”、“每词音节数”、“复杂词占比”、“长词占比”、“难词占比”五项可读性指标的差异。检验结果以中位数代表指标值， p 值报告显著性，Rank-biserial r (r) 值报告效应量，如表三所示。其中，各项指标值越低，可读性越高。同时， $|r| < 0.1$ 代表极小效应， $0.1 \leq |r| \leq 0.3$ 代表小效应， $0.3 \leq |r| \leq 0.5$ 代表中效应， $|r| \geq 0.5$ 代表大效应。由表三可见，相较非极性情感推文，极性情感推文的各项可读性指标均显著更低 ($p < 0.001$)，说明其可读性更高，具体表现为词汇更短、音节

更少、难度更低；然而，各项可读性指标整体效应量较小 ($0.0 < |r| < 0.3$)，说明极性情感推文与非极性情感推文在现实表达中可读性区别不大。具体指标特征如下。

表三：极性、非极性情感推文的可读性指标差异

可读性指标	效应量 r	中位数（极性）	中位数（非极性）	效应方向
每词字符数***	-0.1069 ⁻	4.8182	4.9524	数值越小，可读性越高
每词音节数***	-0.0942 ⁻⁻	1.5143	1.5625	数值越小，可读性越高
复杂词占比***	-0.0856 ⁻⁻	0.1538	0.1667	数值越小，可读性越高
长词占比***	-0.1128 ⁻	0.2353	0.2632	数值越小，可读性越高
难词占比***	-0.1422 ⁻	0.3636	0.4000	数值越小，可读性越高

注： *** $p < 0.001$ ；⁻ $0.1 \leq |r| \leq 0.3$ （小效应），⁻⁻ $|r| < 0.1$ （极小效应）。

每词字符数。极性情感推文相较非极性情感推文，每词包含的字符数显著更低 ($4.8182 < 4.9524$, $p < 0.001$)。其效应量为-0.107，属于小效应。该结果表明极性情感推文倾向于使用更短的词汇，用词更加简洁。

每词音节数。该项指标在极性情感推文上的数值显著低于非极性情感推文 ($1.5143 < 1.5625$, $p < 0.001$)，而效应量 (0.0942) 处于极小效应区间。这说明，尽管极性情感推文使用的词汇具有更简单的音节结构，两类推文在这一特征的差异比较微弱。

复杂词占比。极性情感推文使用复杂词的比例相比非极性情感推文显著更低 ($0.1538 < 0.1667$, $p < 0.001$)，其效应量 (-0.0856) 也处于极小效应范围。因此，极性情感推文虽然倾向使用更多简单易读的词汇，两类推文在这一点的差异依然较小。

长词占比。极性情感推文在长词的使用比例上显著低于非极性情感推文 ($0.2353 < 0.2632$, $p < 0.001$)，其效应量 (0.1128) 为小效应。该结果同样说明了极性情感推文的用词相对更短。

难词占比。极性情感推文的难词使用显著少于非极性情感推文 ($0.3636 < 0.4000$, $p < 0.001$)。效应量 (0.1422) 达到小效应区间，在五项可读性指标中最为明显。这表明，极性情感推文中非常见词出现较少。

总体来看，五项可读性指标下，极性情感推文相比非极性情感推文可读性更高展现出统计意义。然而，从效应量上看，上述差异解释力度有限。可以说，尽管极性情感推文相比非极性情感推文在用词长度、音节数量、词汇难度上均展现出更高的可读性，两者在现实表达中可读性并无太大区别。

4.2 极性情感内容下，推文可读性与其传播效果是否相关？

研究通过 Spearman 方法，对极性情感推文“每词字符数”、“每词音节数”、“复杂词占比”、“长词占比”和“难词占比”五项可读性指标与传播得分进行了相关性检验，并用相关系数 ρ 展示相关程度（见表四）。通常来说， $0.0 < |\rho| < 0.1$

表示极弱相关， $0.1 < |\rho| < 0.3$ 表示弱相关， $|\rho| > 0.3$ 对应中等或强相关。由表四可见，“每词字符数”、“每词音节数”、“复杂词占比”、“长词占比”、“难词占比”与传播效果均呈显著相关 ($p < 0.001$)。然而，各指标相关系数 ρ 处于 [0.02, 0.05] 的区间，反映极弱相关性 ($|\rho| < 0.1$)。这说明，极性情感内容下，推文可读性与传播效果之间几乎不存在实际关联。

表四：极性情感推文可读性指标与传播得分的相关性

可读性指标	相关系数 ρ
每词字符数***	0.0297--
每词音节数***	0.0232--
复杂词占比***	0.0406--
长词占比***	0.0457--
难词占比***	0.0214--

注： *** $p < .001$; -- $|\rho| < 0.1$ (极弱相关)。

5. 讨论与总结

本研究从实证角度出发，探讨了极性情感推文的可读性特征，并分析了极性情感内容下可读性对推文传播的影响效果。研究基于七万余条来自 X 平台的推文，使用五分类情感模型区分了极性、非极性情感文本。在此基础上，研究参考“每词字符数”、“每词音节数”、“复杂词占比”、“长词占比”、“难词占比”等词汇可读性指标与基于信息熵加权的传播得分，对目标推文进行了差异性、相关性分析。研究主要发现两点。第一，极性情感推文相比非极性情感推文在用词长度、音节数量、词汇难度上可读性更高，但两者在现实表达中差异不大。第二，极性情感语境下，推文可读性与传播效果之间几乎没有实际关联。

上述发现拓展了推文在社交媒体传播中极性情感、可读性作用的讨论空间。一方面，在可读性上，极性情感推文与非极性情感推文的实际差异可以忽略不计。这恰好印证了 Danescu-Niculescu-Mizil 等 (2013) 提出的“语言同质化”现象。用户在参与极性情感内容相关话题、社区的讨论时，需要调整语言风格、适应社区规范，才能更好地表达自身观点。因此，用户在同一话题下进行极性、非极性情感表达时，风格的同一便可能削弱可读性差异，并进一步模糊其真实影响。另一方面，就传播作用看，极性情感推文的可读性高低与传播能力无实质联系。该发现为先前研究“推文可读性越高，传播能力越强”(Leonhardt & Makienko, 2018; Pancer 等, 2019; Firouzjaei & Özdemir, 2020) 的通用情境结论提供了重要的补充。具体来说，该发现在极性情感内容下，触及了可读性对推文传播的作用边界。

在强烈情感内容下，推文可读性对传播能力的作用可能失效。其原因可能存在于相关话题传播时的“情感驱动效应”(Berger & Milkman, 2012)、“心理认知捷径”(Metzger & Flanagin, 2013)、“传播动机影响”(Crockett, 2017)。首先，

正如 Berger 与 Milkman (2012) 所揭示的，“高唤醒情绪”本身就是社交媒体内容病毒性传播的强烈驱动因素。作为高唤醒情绪，推文中的极性情感同样更容易吸引用户参与相关讨论与内容传播。此时，相较文本可读性，情感强度可能作为更强的主导因素。其次，“认知捷径”在内容传播时作为用户重要的心理依据 (Metzger & Flanagin, 2013)，可能干扰可读性的作用效果。在网络环境中，由于信息过载，用户很少进行复杂性的评估，而是大量依赖信息来源、观点倾向等直观途径来认识、处理新信息。因此，用户面对极性情感推文时，发布者身份、显性情绪等直观内容更可能成为其主要的判断依据。相反，随着用户对表达方式的敏感性下降，可读性作用便可能被绕过或削弱。最后，在传播动机上，用户对极性情感内容的传播可能并非普通信息分享，而是一种“社会信号”行为 (Crockett, 2017)。通过分享相关话题内容，用户可以低成本、高回报地表明自身立场、提升个人声誉、构建群体认同。在这种情况下，用户的传播动机与信息质量关系较小，可读性对传播的作用也趋于失效。综上可发现，内容可读性在社交媒体传播中的作用存在重要的语境依赖，其影响效果可能受到情感强度、用户心理的调节。也就是说，推文可读性对传播的作用，尤其是在情绪强度显著的内容中，存在条件边界。

本研究对极性情感推文可读性与传播能力的考察，可进一步丰富相关理论认知，并提供面向实际传播策略的启示。理论层面，本文明确提出，推文可读性对传播的驱动能力存在极性情感的条件边界。极性情感内容传播中，情绪倾向、身份立场等非语言因素可能相较文本可读性更占主导。这为先前研究发现做出了重要补充。现实层面，本文发现有利于在当下情绪化的网络生态中，为内容传播策略提供参考。在具有强烈情感色彩的话题讨论中，投入过度资源打磨“易读性”可能事倍功半。关注情感表达的精准性、共鸣性或许更能驱动社交媒体内容传播。

同时，本研究仍存在以下局限。首先，传播效果的度量维度有限。推文的传播效果仅源自点赞、评论、转发、引用等表层互动指标，难以反映更长期或平台范围内信息的传播链条与扩散路径。因此，未来研究可以结合时间动态和社交网络结构，更深入地看待推文传播机制。其次，平台样本与语种单一。本研究对象均为 X 平台中的较短英文推文。这限制了可读性指标的选择，且在语言发现上存在片面性。后续研究可以在不同长度、语言等角度深入探讨极性情感、文本可读性与社交媒体传播的关系。

参考文献

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219–235. <https://doi.org/10.1177/1088868309341564>
- Andreasvc. (2025). Readability (Version 0.3.2). <https://pypi.org/project/readability/>
- Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of*

- Marketing Research*, 49(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>
- Björnsson, C.-H. (1968). *Läsbartet*. Liber.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318.
<https://doi.org/10.1073/pnas.1618923114>
- Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale–Chall readability formula*. Brookline Books.
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1, 769–771. <https://doi.org/10.1038/s41562-017-0213-3>
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. *Educational Research Bulletin*, 27(2), 37–54.
- Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)* (pp. 307–318). Association for Computing Machinery.
<https://doi.org/10.1145/2488388.2488416>
- Davenport, J. R. A., & DeLine, R. (2014). The readability of tweets and their geographic correlation with education. *arXiv*. <http://arxiv.org/abs/1401.6058>
- Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *arXiv*.
<https://doi.org/10.48550/arXiv.1703.04009>
- Firouzjaei, H. A., & Özdemir, S. F. (2020). Effect of readability of political tweets on positive user engagement. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 884–891). IEEE. <https://doi.org/10.1109/ASONAM49781.2020.9381431>
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Fry, E. (1990). A readability formula for short passages. *Journal of Reading*, 33(8), 594–597.
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Gunning, R. (1969). The fog index after twenty years. *Journal of Business Communication*, 6(2), 3–13. <https://doi.org/10.1177/002194366900600202>
- Jacob, P., & Uitdenbogerd, A. (2019). Readability of Twitter tweets for second language learners. In M. Mistica, M. Piccardi, & A. MacKinlay (Eds.), *Proceedings of the 17th Annual Workshop of the Australasian Language Technology Association* (pp. 19–27). Australasian Language Technology

- Association. <https://aclanthology.org/U19-1003/>
- Klare, G. R. (1963). *Measurement of readability*. The Iowa State University Press.
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), 8788–8790. <https://doi.org/10.1073/pnas.1320040111>
- Leileibama. (2023). Alpha Readability Calculator (Version 1.0).
<https://github.com/leileibama/AlphaReadabilityCalculator>
- Leitner, F. (2022). syntok (Version 1.4.4). <https://pypi.org/project/syntok/>
- Lenzner, T. (2013). Are Readability Formulas Valid Tools for Assessing Survey Question Difficulty? *Sociological Methods & Research*, 43(4), 677–698. <https://doi.org/10.1177/0049124113513436>
- Leonhardt, J. M., & Makienko, I. (2018). Keep it simple, readability increases engagement on Twitter: An abstract. In N. Krey & P. Rossi (Eds.), *Back to the future: Using marketing basics to provide customer value. Developments in marketing science: Proceedings of the Academy of Marketing Science* (pp. 333–334). Springer. https://doi.org/10.1007/978-3-319-66023-3_116
- McLaughlin, G. H. (1969). SMOG grading — a new readability formula. *Journal of Reading*, 12(8), 639–646.
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59(Part B), 210–220. <https://doi.org/10.1016/j.pragma.2013.07.012>
- Pancer, E., Chandler, V., Poole, M., & Noseworthy, T. J. (2019). How readability shapes social media engagement. *Journal of Consumer Psychology*, 29(2), 262–270. <https://doi.org/10.1002/jcpy.1073>
- Pfeffer, J., Matter, D., & Sargsyan, A. (2023). The half-life of a tweet (No. arXiv:2302.09654). *arXiv*. <https://doi.org/10.48550/arXiv.2302.09654>
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217–248. <https://doi.org/10.2753/MIS0742-1222290408>
- Tabularisai. (2024). robust-sentiment-analysis.
<https://huggingface.co/tabularisai/robust-sentiment-analysis>
- Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*,

3(31), 1026, <https://doi.org/10.21105/joss.01026>

Xiao, S., & Chen, X. (2025). Measuring social media customer engagement with brands based on information entropy: an application case of luxury brand. *Journal of Brand Management*, 32, 184–202.
<https://doi.org/10.1057/s41262-024-00376-7>

Zappavigna, M., & Ross, A. S. (2024). *Innovations and challenges in social media discourse analysis*. Routledge. <https://doi.org/10.4324/9781003257516>