

Unveiling the secret of information rediffusion process on social media from information coupling perspective: a hybrid approach of machine learning and regression model

Presenter : Zhen Yan

Co-authors: Zhen Yan, Xi' an Jiaotong University;

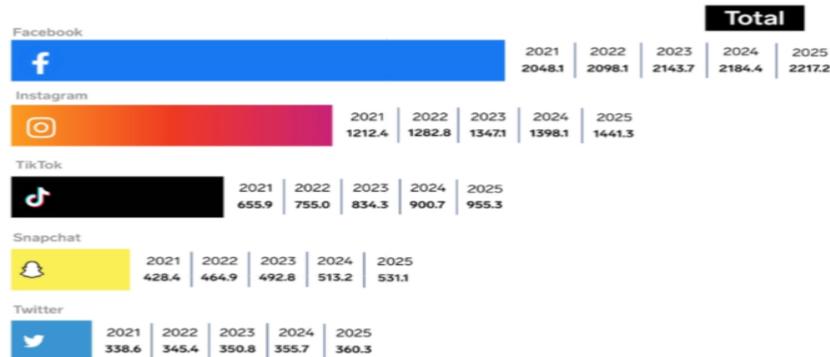
Rong Du, Xidian University;

Hua Wang, Xi' an Jiaotong University

Date: 23rd, April, 2024

Motivation

Social Network Users Worldwide, by Platform, 2021-2025 millions



Notes: internet users of any age who use social networks via any device at least once per month; Snapchat forecast includes mobile phone users who access their Snapchat account via mobile phone app at least once per month
Methodology: Estimates are based on the analysis of survey and traffic data from research firms and regulatory agencies; company releases; historical trends; internet and mobile adoption trends; and country-specific demographic and socioeconomic factors.
Source: eMarketer, Nov 2021

1271750

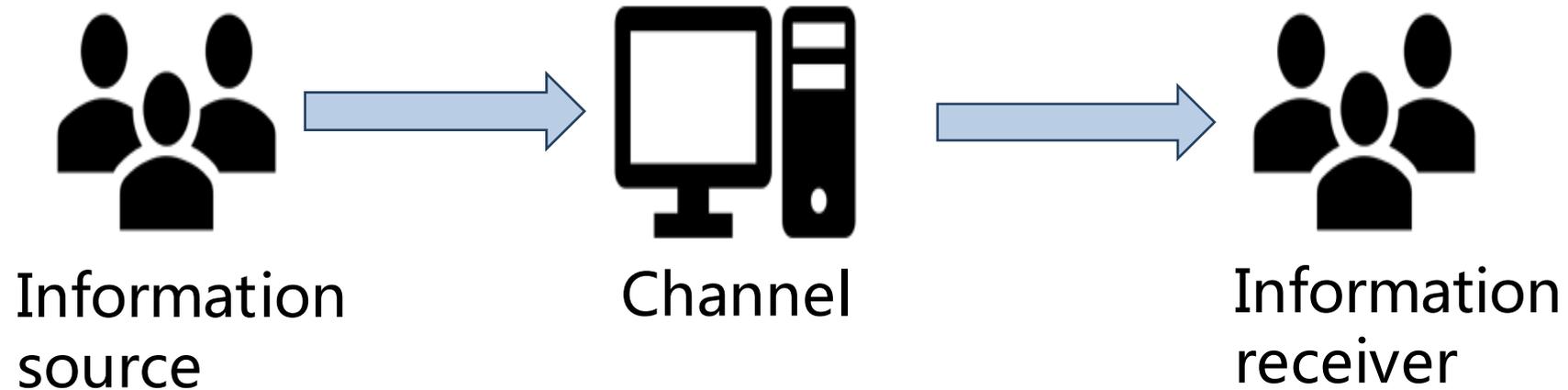
insiderintelligence.com



In the information booming era, the development of social media platforms:

- Provide access for users to share information and interact with others;
- Melt the barriers of communication;
- Motivate information diffusion with a fast speed.

Information diffusion process



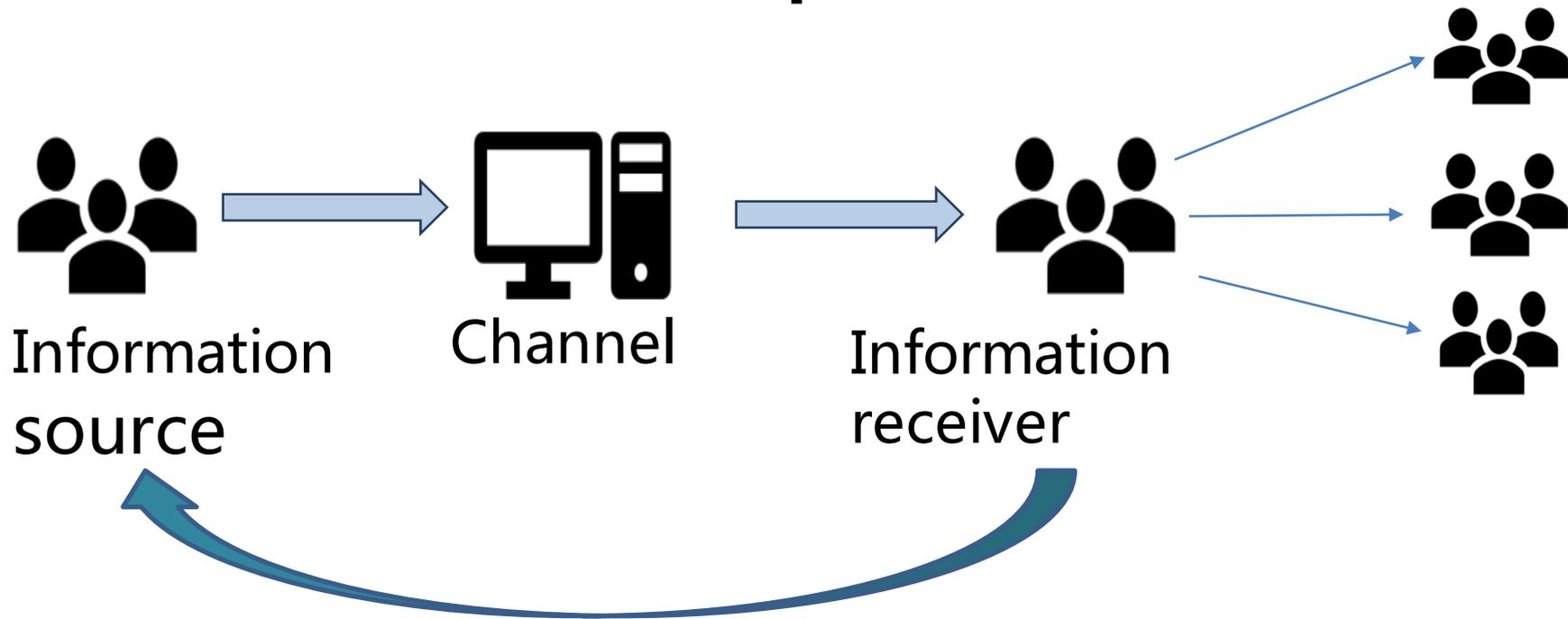
Key factors in information diffusion process:

- Emotional changes
- Semantic relations

Information rediffusion process

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
楼主ID	评论者ID	话题内容	评论内容	评论者昵称	评论者性别	是否认证	认证类型	是否认证	发博数量	关注人数	粉丝数	微博等级	发布日期	评论日期	评论时间	转发	回复	
1699432410	1746320005	【正在直播: 解封不解防	一棵有梦想f			TRUE		0	0	528	684	726	37	2022/5/25	2022/5/25	10:05:56	11	36
1699432410	2337854044	【正在直播: 个人防护	dancing Vf			FALSE	220	0	1671	634	1289	43	2022/5/25	2022/5/25	10:07:10	19	33	
1699432410	1935435353	【正在直播: 已经管控并	司号_Z	m		TRUE		0	0	4630	661	224	31	2022/5/25	2022/5/25	10:03:14	1	25
1699432410	5289596304	【正在直播: 少出们, 严	imee是女f			TRUE		0	0	292	171	116	4	2022/5/25	2022/5/25	10:24:00	8	24
1699432410	5019967163	【正在直播: 消杀很重要	lielen月 f			TRUE		0	0	63	398	155	0	2022/5/25	2022/5/25	10:01:33	3	25
1699432410	3847762200	【正在直播: 第一时间复	卡福先生 m			FALSE	-1	0	183	217	74	14	2022/5/25	2022/5/25	10:07:39	2	13	
1699432410	2140696327	【正在直播: 走到如今不	请把鸡蛋做f			FALSE	-1	0	7270	572	201	40	2022/5/25	2022/5/25	10:05:55	1	12	
1699432410	7017481729	【正在直播: 还好还好,	么毅nyth m			FALSE	-1	0	269	588	33	0	2022/5/25	2022/5/25	10:31:17	0	6	
1699432410	5565257462	【正在直播: 我爱的上海	灵雨花海 f			FALSE	-1	0	206	705	35	8	2022/5/25	2022/5/25	10:17:08	0	5	
1699432410	1742664423	【正在直播: 管控区的保	老阿姨在注f			FALSE	-1	0	296	285	1361	35	2022/5/25	2022/5/25	10:08:25	1	6	
1699432410	5282924179	【正在直播: 转发	今日余山 m			TRUE	1	0	84225	167	1871	28	2022/5/25	2022/5/25	20:52:09	0	0	
1699432410	6009036327	【正在直播: 转发微博	扎兰屯网_m			TRUE	1	0	2585	444	154	14	2022/5/25	2022/5/25	10:35:58	0	1	
1699432410	6600832433	【正在直播: 虽然疫情开	风若羽羽f			FALSE	-1	0	623	226	46	4	2022/5/25	2022/5/25	10:51:56	0	1	
1699432410	1779124422	【正在直播: 所有密接者	大杯全糖f			FALSE	220	0	70	170	308	29	2022/5/25	2022/5/25	10:03:58	0	3	
1699432410	2631817152	【正在直播: 当路上人开	虹桥-招商m			FALSE	-1	0	1891	447	12000	4	2022/5/25	2022/5/25	10:02:25	0	2	
1699432410	1746320005	【正在直播: 解封不解防	一棵有梦想f			TRUE		0	0	528	684	726	37	2022/5/25	2022/5/25	10:05:56	11	36
1699432410	2337854044	【正在直播: 个人防护	dancing Vf			FALSE	220	0	1671	634	1289	43	2022/5/25	2022/5/25	10:07:10	19	33	
1699432410	1935435353	【正在直播: 已经管控并	司号_Z	m		TRUE		0	0	4630	661	224	31	2022/5/25	2022/5/25	10:03:14	1	25
1699432410	5289596304	【正在直播: 少出们, 严	imee是女f			TRUE		0	0	292	171	116	4	2022/5/25	2022/5/25	10:24:00	8	24
1699432410	5019967163	【正在直播: 消杀很重要	lielen月 f			TRUE		0	0	63	398	155	0	2022/5/25	2022/5/25	10:01:33	3	25
1699432410	3847762200	【正在直播: 第一时间复	卡福先生 m			FALSE	-1	0	183	217	74	14	2022/5/25	2022/5/25	10:07:39	2	13	
1699432410	2140696327	【正在直播: 走到如今不	请把鸡蛋做f			FALSE	-1	0	7270	572	201	40	2022/5/25	2022/5/25	10:05:55	1	12	
1699432410	7017481729	【正在直播: 还好还好,	么毅nyth m			FALSE	-1	0	269	588	33	0	2022/5/25	2022/5/25	10:31:17	0	6	
1699432410	5565257462	【正在直播: 我爱的上海	灵雨花海 f			FALSE	-1	0	206	705	35	8	2022/5/25	2022/5/25	10:17:08	0	5	
1699432410	1742664423	【正在直播: 管控区的保	老阿姨在注f			FALSE	-1	0	296	285	1361	35	2022/5/25	2022/5/25	10:08:25	1	6	
1699432410	5282924179	【正在直播: 转发	今日余山 m			TRUE	1	0	84225	167	1871	28	2022/5/25	2022/5/25	20:52:09	0	0	
1699432410	6009036327	【正在直播: 转发微博	扎兰屯网_m			TRUE	1	0	2585	444	154	14	2022/5/25	2022/5/25	10:35:58	0	1	
1699432410	6600832433	【正在直播: 虽然疫情开	风若羽羽f			FALSE	-1	0	623	226	46	4	2022/5/25	2022/5/25	10:51:56	0	1	
1699432410	1779124422	【正在直播: 所有密接者	大杯全糖f			FALSE	220	0	70	170	308	29	2022/5/25	2022/5/25	10:03:58	0	3	
1699432410	2631817152	【正在直播: 当路上人开	虹桥-招商m			FALSE	-1	0	1891	447	12000	4	2022/5/25	2022/5/25	10:02:25	0	2	
1699432410	1746320005	【正在直播: 解封不解防	一棵有梦想f			TRUE		0	0	528	684	726	37	2022/5/25	2022/5/25	10:05:56	11	36
1699432410	2337854044	【正在直播: 个人防护	dancing Vf			FALSE	220	0	1671	634	1289	43	2022/5/25	2022/5/25	10:07:10	19	33	
1699432410	1935435353	【正在直播: 已经管控并	司号_Z	m		TRUE		0	0	4630	661	224	31	2022/5/25	2022/5/25	10:03:14	1	25
1699432410	5289596304	【正在直播: 少出们, 严	imee是女f			TRUE		0	0	292	171	116	4	2022/5/25	2022/5/25	10:24:00	8	24
1699432410	5019967163	【正在直播: 消杀很重要	lielen月 f			TRUE		0	0	63	398	155	0	2022/5/25	2022/5/25	10:01:33	3	25
1699432410	3847762200	【正在直播: 第一时间复	卡福先生 m			FALSE	-1	0	183	217	74	14	2022/5/25	2022/5/25	10:07:39	2	13	
1699432410	2140696327	【正在直播: 走到如今不	请把鸡蛋做f			FALSE	-1	0	7270	572	201	40	2022/5/25	2022/5/25	10:05:55	1	12	
1699432410	7017481729	【正在直播: 还好还好,	么毅nyth m			FALSE	-1	0	269	588	33	0	2022/5/25	2022/5/25	10:31:17	0	6	
1699432410	5565257462	【正在直播: 我爱的上海	灵雨花海 f			FALSE	-1	0	206	705	35	8	2022/5/25	2022/5/25	10:17:08	0	5	
1699432410	1742664423	【正在直播: 管控区的保	老阿姨在注f			FALSE	-1	0	296	285	1361	35	2022/5/25	2022/5/25	10:08:25	1	6	
1699432410	5282924179	【正在直播: 转发	今日余山 m			TRUE	1	0	84225	167	1871	28	2022/5/25	2022/5/25	20:52:09	0	0	
1699432410	6009036327	【正在直播: 转发微博	扎兰屯网_m			TRUE	1	0	2585	444	154	14	2022/5/25	2022/5/25	10:35:58	0	1	

Information rediffusion process



Information coupling: the degree of **difference** between **information source** and the **User-generated-content** (UGC), the content that is created by members of the general public and distributed over the internet (Daugherty et al. 2008, Krumm et al. 2008)

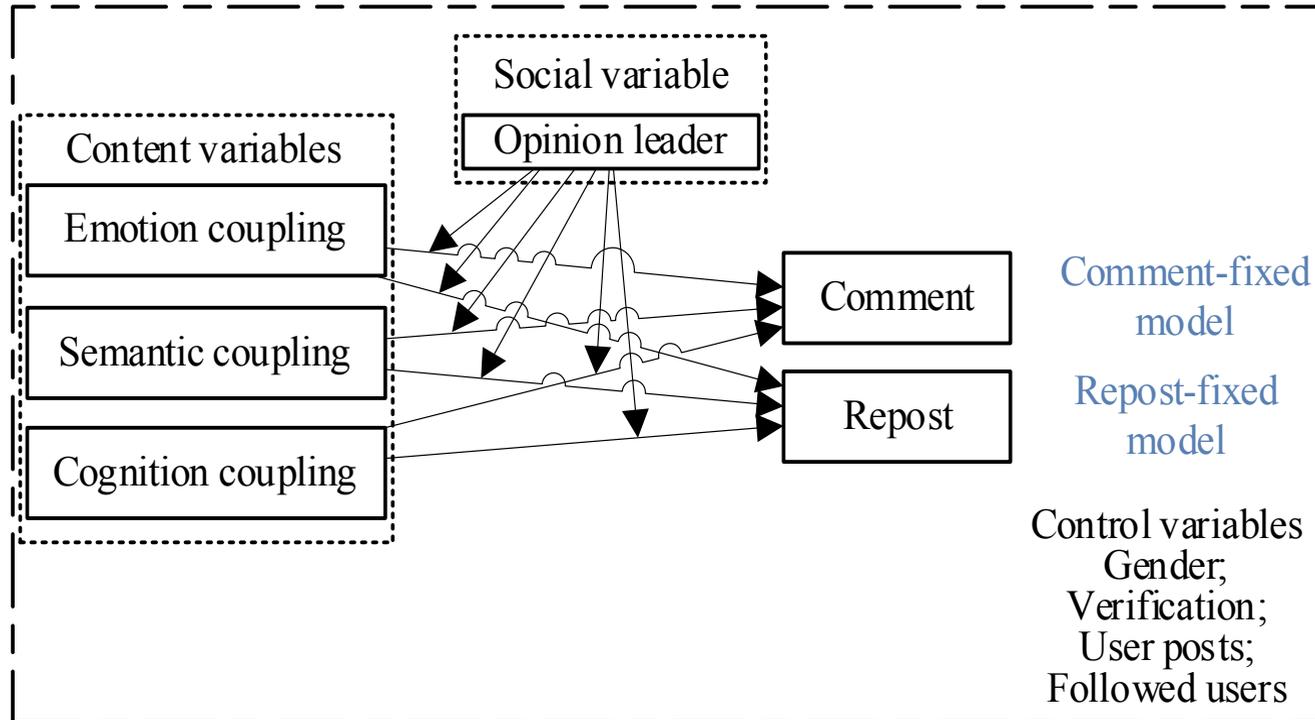
Research Questions

RQs 1: How does information coupling influence information rediffusion in terms of commenting?

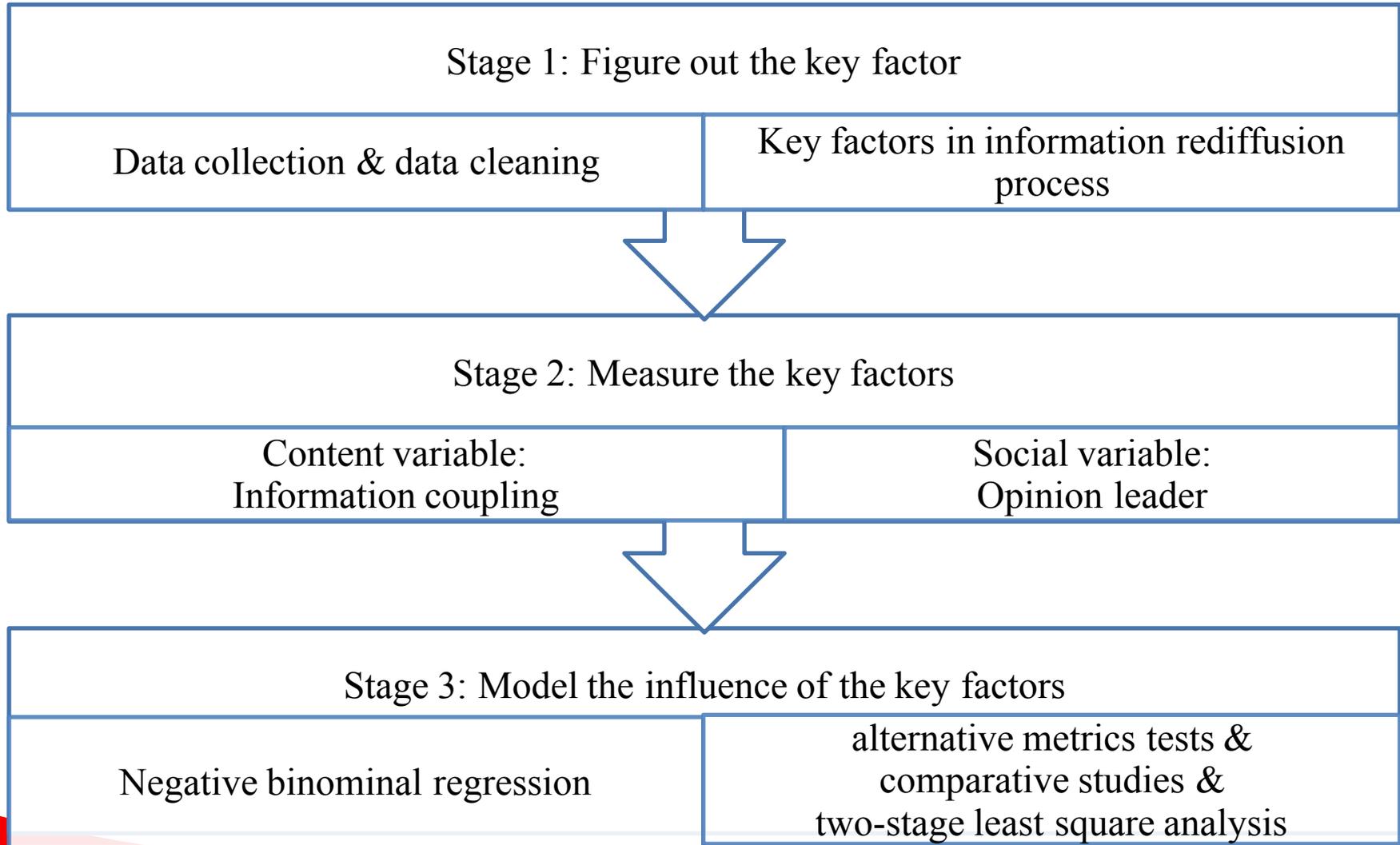
RQs 2: How does information coupling influence information rediffusion in terms of reposting?

RQs 3: How do opinion leaders affect information rediffusion?

Conceptual model



Research Framework



Data collection & cleaning

Source: Sina Microblog 

Time span: 1 Dec 2021- 1 Jun 2022

Data volume: 4,017 pieces of news and 416,358 pieces of UGC

Topic: public health emergencies

Resaved data volume after cleaning: 4,017 pieces of news and resaved 415,473 pieces of UGC (i.e. remove repeated data and symbol-only data and Jieba word split).

Variable description and calculation

Variables	Definition	Description	Measurement
Independent variables			
comment	Sum of comments	Number of comments	Frequency
Repost	Sum of reposts	Number of repost	Frequency
Dependent variables			
Emotional coupling	Emotional intensity	Degree of sentiment polarity difference between the news and UGC	BERT & Difference
Semantic coupling	Content similarity in semantic level	Degree of topic similarity between the news and UGC	LDA & K-means & Cosine similarity
Cognitive coupling	Cognition similarity in metaphorical level	Degree of metaphorical expressions similarity between the news and UGC	WordNet & Cosine similarity
Moderate variable			
Social variables	Opinion leader	Followers of user more than 10,000 and with Big V badge	Followers more than 10,000 \cap Big V badge
Control variables			
Gender	The gender of the user	The gender	Male=1; Female=0
Verification	Microblog level of the user	The status of the user	Status= (0,1,2...50)
Number of user posts	Sum of user posts	Number of users posts in his account	Number of posts
Number of followed users	Sum of user followed accounts	Number of followed users in microblog	Number of followed accounts

Variable description and calculation

Variables	Definition	Description	Measurement
Independent variables			
comment	Sum of comments	Number of comments	Frequency
Repost	Sum of reposts	Number of repost	Frequency
Dependent variables			
Emotional coupling	Emotional intensity	Degree of sentiment polarity difference between the news and UGC	BERT & Difference
Semantic coupling	Content similarity in semantic level	Degree of topic similarity between the news and UGC	LDA & K-means & Cosine similarity
Cognitive coupling	Cognition similarity in metaphorical level	Degree of metaphorical expressions similarity between the news and UGC	WordNet & Cosine similarity
Moderate variable			
Social variables	Opinion leader	Followers of user more than 10,000 and with Big V badge	Followers more than 10,000 \cap Big V badge
Control variables			
Gender	The gender of the user	The gender	Male=1; Female=0
Verification	Microblog level of the user	The status of the user	Status= (0,1,2...50)
Number of user posts	Sum of user posts	Number of users posts in his account	Number of posts
Number of followed users	Sum of user followed accounts	Number of followed users in microblog	Number of followed accounts

$$\begin{aligned}
 & \textit{Sentiment polarity} = \textit{positive} + \textit{negative} \\
 & \textit{Emotional coupling} = SP_{UGC} - SP_{news}
 \end{aligned}$$

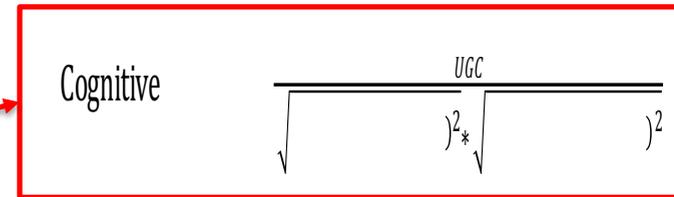
Variable description and calculation

Variables	Definition	Description	Measurement
Independent variables			
comment	Sum of comments	Number of comments	Frequency
Repost	Sum of reposts	Number of repost	Frequency
Dependent variables			
Emotional coupling	Emotional intensity	Degree of sentiment polarity difference between the news and UGC	BERT & Difference
Semantic coupling	Content similarity in semantic level	Degree of topic similarity between the news and UGC	LDA & K-means & Cosine similarity
Cognitive coupling	Cognition similarity in metaphorical level	Degree of metaphorical expressions similarity between the news and UGC	WordNet & Cosine similarity
Moderate variable			
Social variables	Opinion leader	Followers of user more than 10,000 and with Big V badge	Followers more than 10,000 \cap Big V badge
Control variables			
Gender	The gender of the user	The gender	Male=1; Female=0
Verification	Microblog level of the user	The status of the user	Status= (0,1,2...50)
Number of user posts	Sum of user posts	Number of users posts in his account	Number of posts
Number of followed users	Sum of user followed accounts	Number of followed users in microblog	Number of followed accounts

$$S = \frac{UGC}{\sqrt{)2 * \sqrt{)2}}$$

Variable description and calculation

Variables	Definition	Description	Measurement
Independent variables			
comment	Sum of comments	Number of comments	Frequency
Repost	Sum of reposts	Number of repost	Frequency
Dependent variables			
Emotional coupling	Emotional intensity	Degree of sentiment polarity difference between the news and UGC	BERT & Difference
Semantic coupling	Content similarity in semantic level	Degree of topic similarity between the news and UGC	LDA & K-means & Cosine similarity
Cognitive coupling	Cognition similarity in metaphorical level	Degree of metaphorical expressions similarity between the news and UGC	WordNet & Cosine similarity
Moderate variable			
Social variables	Opinion leader	Followers of user more than 10,000 and with Big V badge	Followers more than 10,000 \cap Big V badge
Control variables			
Gender	The gender of the user	The gender	Male=1; Female=0
Verification	Microblog level of the user	The status of the user	Status= (0,1,2...50)
Number of user posts	Sum of user posts	Number of users posts in his account	Number of posts
Number of followed users	Sum of user followed accounts	Number of followed users in microblog	Number of followed accounts



Model regression-comment fixed model

logF

logF

logF

logF

Model regression-repost fixed model

$\log F$

$\log F$

$\log F$

$\log F$

Model regression-moderate effect model

$\log F$

$\log F$

Findings and Discussions

Mean, standard error and correlation variables in comment-fixed model

variables	M	SD	Comment-fixed models				
			Model 1	Model 2	Model 3	Model 4	Model 5
Emotional coupling	1.084	0.557		1.007**	1.210**	1.014**	1.001**
Semantic coupling	1.033	0.034			0.667***	0.698***	0.699***
Cognitive coupling	1.401	0.505				0.637***	0.658***
Opinion leader	3.706	0.007					0.227*
Gender	0.800	0.201	0.450**	0.417**	0.415**	0.454**	0.421**
Verification	1.462	0.211	0.599***	0.554***	0.534***	0.522***	0.535***
User posts	-9.895	1.105	-1.122***	-1.145***	-1.146***	-1.136***	-1.131***
Followed users	-2.566	0.001	0.487***	0.424***	0.402***	0.467***	0.435***
R2			0.645	0.786	0.782	0.784	0.788

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

The three types of information coupling are all significantly influencing commenting information redifussion process.

Mean, standard error and correlation variables in repost-fixed effect model

variables	M	SD	Repost-fixed models				
			Model 6	Model 7	Model 8	Model 9	Model 10
Emotional coupling	1.084	0.557		0.946**	0.958**	0.954**	0.967**
Semantic coupling	1.033	0.034			0.417***	0.535***	0.447***
Cognitive coupling	1.401	0.505				0.668***	0.674***
Opinion leader	3.706	0.007					3.388**
Gender	0.800	0.201	0.857**	0.842**	0.756**	0.631**	0.817**
Verification	1.462	0.211	2.345**	2.398**	2.452**	2.354**	2.315**
User posts	-9.895	1.105	-0.475***	-0.425***	-0.397***	-0.545***	-0.465***
Followed users	-2.566	0.001	0.035***	0.041***	0.042***	0.038***	0.048***
R2			0.771	0.782	0.781	0.786	0.788

Note: * p < .05. ** p < .01. *** p < .001.

The three types of information coupling are all significantly influencing reposting in information redifussion process.

The moderated effect of opinion leader on comment and repost

Variables	Model 11 (comment)	Model 12 (repost)
Emotional coupling × opinion leader	2.317*** (0.381)	0.389*** (0.535)
Semantic coupling × opinion leader	0.532*** (0.064)	2.359*** (0.086)
Cognitive coupling × opinion leader	2.304*** (0.443)	2.707*** (0.570)
gender	0.454** (0.293)	0.631** (0.293)
Verification	0.522*** (2.050)	2.354** (2.050)
User posts	-1.136*** (1.434)	-0.545*** (0.434)
Followed users	0.467*** (3.634)	0.038*** (3.634)
	R^2 0.527	0.642

The moderating effects of opinion leaders are all significantly influencing commenting and reposting in information redifussion process.

Note: Standard errors in parentheses; * $p < .05$. ** $p < .01$. *** $p < .001$.

Model validation-Alternative metrics test & comparative studies

Variables	Model 13 (comment)	Model 14(repost)
Emotional coupling	0.281**	0.274**
Semantic coupling	0.192***	0.231***
Cognitive coupling	0.823***	0.713***
Emotional coupling × opinion leader	1.102***	0.901***
Semantic coupling × opinion leader	0.910***	1.085***
Cognitive coupling × opinion leader	0.991***	1.642***
gender	0.890***	0.631**
Verification	0.754*	2.354**
User posts	-4.000***	-0.545***
Followed users	-0.001*	0.038***
R^2	0.235	0.286

Note: p < .05. ** p < .01. *** p < .001.

Alternative metrics test:
 emotional coupling-
 sentiment dictionary;
 semantic coupling-
 Word2vec;
 cognitive coupling-
 metaphor corpus & co-
 occurrences.
Comparative studies :
 OLS.

Model validation- Two stage least square analysis

Variables	Model 15 (comment)	Model 15(repost)
Emotional coupling	0.201**	0.205**
Semantic coupling	0.130***	0.156***
Cognitive coupling	0.432***	0.524***
Emotional coupling × opinion leader	1.502***	1.361***
Semantic coupling × opinion leader	1.263***	1.258***
Cognitive coupling × opinion leader	1.552***	1.559***
gender	0.320***	0.221**
Verification	0.724*	1.364**
User posts	-2.338***	-2.006***
Followed users	-0.238*	0.217***
R ²	0.455	0.512

Note: p < .05. ** p < .01. *** p < .001.

Conclusion

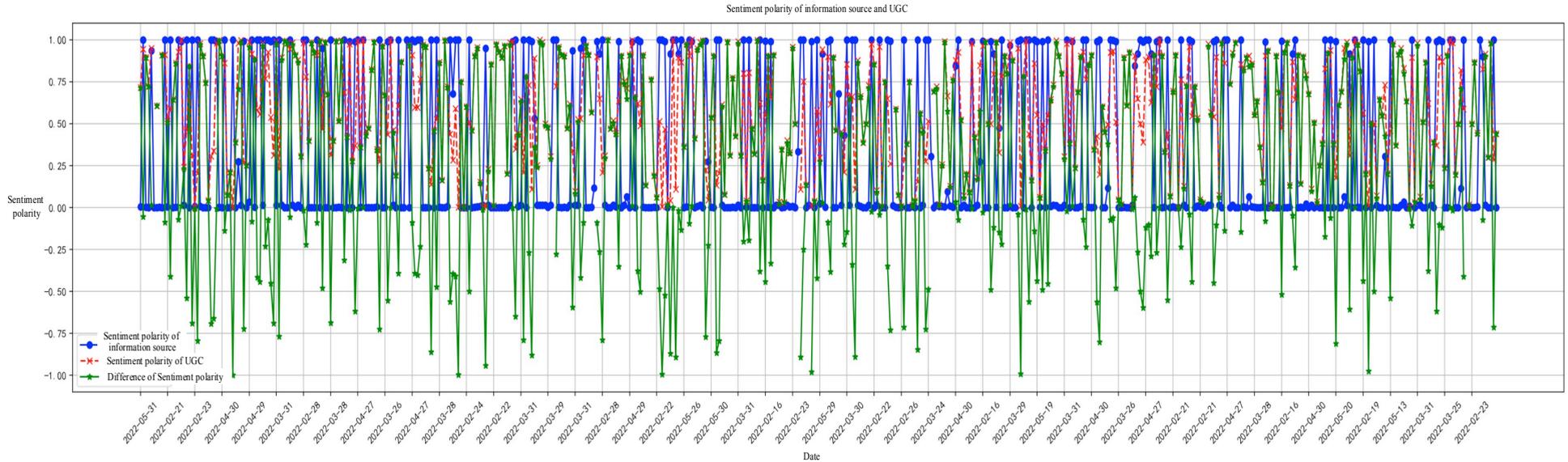


Emotional and semantic coupling prompt information rediffusion through comments, while reposting typically depends on emotional and cognitive coupling.



The moderating effects of opinion leaders in reposting are more significant than commenting, especially the semantic and cognitive coupling.

Conclusion



Emotional fluctuation in time span

Conclusion

Theoretical implications

Difference in emotional intensity between a piece of news and some UGC is a highly significant factor which fluctuates dramatically. This is supported by cognitive dissonance theory.

The effect of opinion leaders in moderation model explicitly displays the “nudge” effect of opinion leaders in social network as two-step flow theory posits.

Conclusion

Practical implications

Social media platforms should be aware that differences in emotion are critical for attracting comments to a post, whereas cognitive differences are the primary reason for the inclination of people to share/repost information with their friends.

Opinion leaders, who would like to enhance their influence, should be aware of the basic human need for trust and visibility by others, especially using metaphorical expressions.

Unveiling the secret of information rediffusion process on social media from information coupling perspective: a hybrid approach of machine learning and regression model

Thanks for your attention!

Q & A!

Presenter : Zhen Yan

Date: 23rd, April, 2024

XI'AN JIAOTONG UNIVERSITY