

News Selection by News Aggregators and Incentives for Newspapers to Invest in Journalism: The Case of South Korea

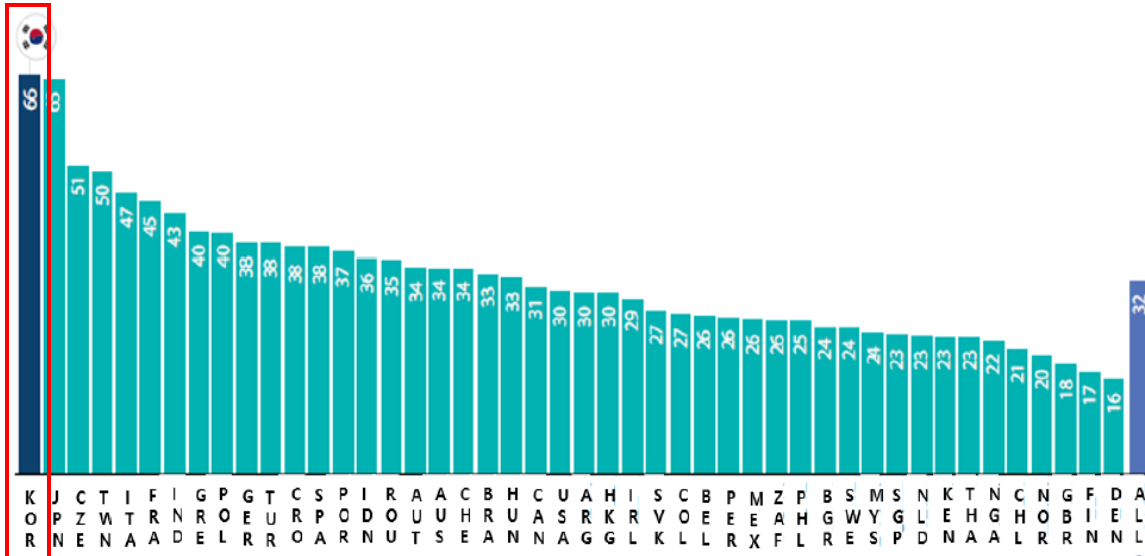
Dong Ook Choi (Sangmyung Univ.)

Doh-Shin Jeon (Toulouse School of Economics)

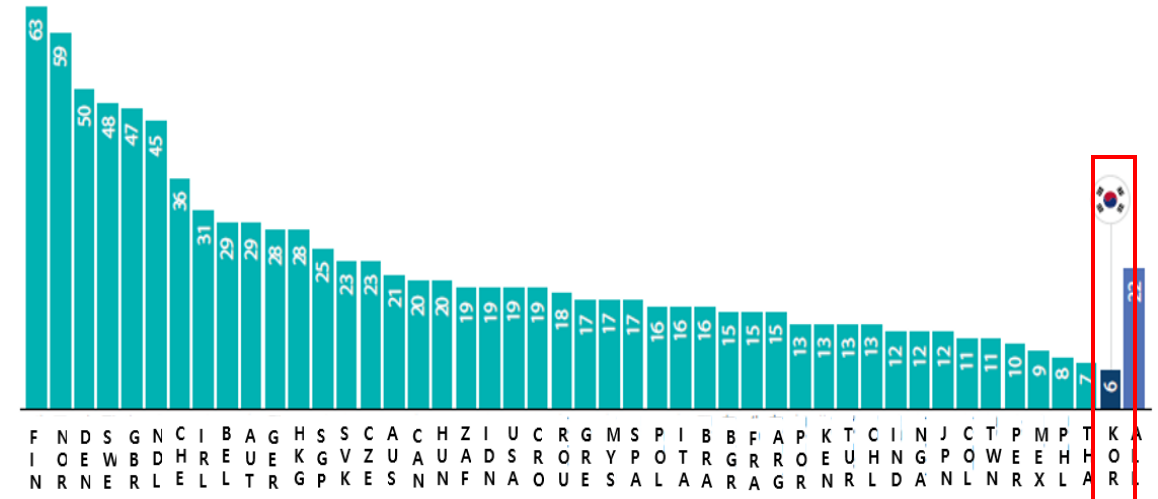
Changmin Oh (Konan Tech.)

Introduction: Access to Online News

- South Korea is an exception in the world:
 - The proportion of people for whom **search and news aggregators** are the main gateway to online news: 66%
 - The proportion of people for whom **direct access to websites and apps** is the main gateway to online news: 6%



Search engine and news aggregators



Direct access through websites and apps

South Korea is an exception

- A large majority of South Koreans consume news through two news aggregators:
 - Naver.com (called Naver, hereafter)
 - Daum.net (called Daum hereafter)
- In 2015 (the year our data is collected), the share of the two aggregators, Naver and Daum, was 77.8%.
- The way the two aggregators select news has a significant impact on the production of news in South Korea.

News Aggregators in South Korea

NAVER 뉴스 TV 연예 | 스포츠 | 뉴스스탠드 | 날씨

08.04 (화) 오전 30°C 주요뉴스 ▶ 남평 집안일 도울수록 아내 우울증 심해진다? ▶ 상식인뉴스 · 언론사 뉴스 · 라이브러리 · 편집이력

뉴스를 속보 정치 경제 사회 생활/문화 세계 IT/과학 오피니언 포토 TV 행킹뉴스 뉴스검색

신문 헤드라인 · 저녁 방송 뉴스

이 시각 주요뉴스 최종편집 2015.08.04. 15:25

8월 14일 임시공휴일 지정
전국 고속도로 통행료 면제 국무회의, '광복 70주년 국민사기 진작방안' 확정 14~16일, 4대 고공, 국립중앙박물관 등 무료개방외국인...
방과후학교에서 영어 수학 선행학습 가능해진다
수능 D-100, 4당5락은 거짓이다
5선 법무장관 역임 박상천 전 민주당 대표 별세
축구장 3배 면적 초등학교 차기디엔장로켓 실전제치

자전거 탈 때 이런 행동은 삼가세요!

광고 영역입니다.

오늘의 핫뉴스

롯데홈 경영권 분쟁
· '경영권 전횡당 싸움' 롯데, 핵심 사업차질 우려
· 신동빈 롯데 회장 "글로벌 기업으로 성장하기 위한 진통"
· 8인론 '롯데 이진투구' 크게 보도...일본 내 이미지 악화 불가피
· '악화일로' 치달는 롯데...세무조사에 '롯데 불매운동'까지
· '롯데 때문에...' 속앓이 하는 재계

연립뉴스
뉴스1
연립뉴스
노컷뉴스
문화일보

네트즌 외곽외국

"한 컷 사진 욕심에 소나무 베고..."
4당5락은 거짓이다
'적자의 늑' 벗어난 공공부문...7년 만에..
성장을 떨어지면 청년층이 지갑 더 닫는다

TV연예 TV | 포토 | 스타캐스트

스포츠 야구 축구 해외야구 해외축구

TODAY
운세, 8월 4일 화요일
날씨, 전국 폭염 계속...동해 36.6도
코스피 2,027.99 ▲19.50 +0.97%
코스닥 730.03 ▲15.69 +2.20%
종목검색

다음 카페 메일 블로그 쇼핑 지도 사전 지식 뉴스 증권 부동산 영화 아고라 뉴스 웹툰 열정팬서 게임 회화해 더보기

다음 중국 태풍피해 강정호 올린 조윤경 체중감량 이안류 주의

너무 답글한 순간 너무 실례는 여행 아주 알맞은 제안 루프트한자 A380 유럽 50,000원 왕권! 지금 예약

시즌OFF 대방송 엄마를 감동시킬 옷 마담4060

로그인 유지 | IP보안 ON
아이디 or 이메일 비밀번호 로그인
회원가입 | 아이디 비밀번호 찾기 | 비밀번호 로그인

2015년, 내 보험료는 얼마일까? 내 보험료 알아보기

소창호 소호 육류갈조 웨버 롯데닷컴
육선 · GO·C·GS·11번가 · CJ·위메프
롯데홈 · 하프클럽 · 이마트홈 · 신세계 · 롯데닷컴
현대홈 · 패션플러스 · 쇼핑기획전

최근 클릭상품 베스트100 1/11

GO·C·C
드문트 고계 7500원급
리넨블라우스 50%1만원대
블랙헤드락-슬릭스틱 30%
레이스-주어 인 기 복 주
레니본-신상 단품10%할인
퍼다정지한택 초트기53%↓
엔틱고급가구 무라워서너리
1등 엄마와 최대 70% ↓
일기만해도 매력이남쳐

실시간 이슈 부산 29.8°C

1 현마
2 목한지리
3 김태희
4 소비즈코리아 홈스타
5 활화드라마
6 남장고를 부탁해
7 아육대
8 박기춘
9 임금피크제
10 서울시교육청

KBO리그
0 : 0
18:30
넥센 양승준계 삼성

08.10 날씨 증권 한올 TV 운세

브랜즈 스페셜 4/4

FUN 동영상 공감 부동산 마켓넷

오버마도 먹은 미국식 수제버거
잡지만 매우 하트 워터슬라이드
말이 안 나오는 프리스타일 기술
착색크림 이 정도 효과 일 줄

News in Aggregators

- In-link service:
 - Selected news are shown in the aggregators' own pages.
- Short length of news text:
 - Usually less than 30 seconds reading.

檢, 조지호-김봉식 구속기소...국회 봉쇄하고 체포조 편성한 혐의

최미송 기자 · 2025. 1. 8. 19:29

2

🔍 🔊 🔄 ⌂ 📄

12·3 불법 비상계엄 선포 사건을 수사 중인 검찰 비상계엄 특별수사본부(본부장 박세현 서울고검장)가 8일 조지호 경찰청장과 김봉식 전 서울경찰청장을 내란 등의 혐의로 구속기소했다.

검찰 비상계엄 특별수사본부(본부장 박세현 서울고검장)는 이날 조 청장과 김 전 청장을 내란 중요인물 중사 및 직권남용권리행사방해 혐의로 재판에 넘기며 이들이 "비상계엄 선포에 대비해 국회를 봉쇄하고 체포조 편성에 가담하였으며 선거관리위원회(선관위)의 서버 반출을 시도했다"고 밝혔다.

검찰에 따르면 이들은 계엄 선포 직전인 지난달 3일 오후 7시 20분경 서울 종로구 삼청동 안가에서 윤 대통령과 만나 "중복 좌파 세력, 반국가 세력들이 사회 곳곳에서 나라를 혼란스럽게 하고 있다. 내가 오늘 밤 10시에 비상계엄을 선포할 것인데 계엄군이 국회에 갈 때 경찰이 나가서 국회 통제를 잘 해달라"는 지시를 받았다. 조 청장과 김 전 청장은 이 자리에서 김용현 전 국방부 장관으로부터 '2200 국회', '2230 더불어민주당', '여론조사꽃' 등 계엄군이 출동할 시간과 장소가 기재된 A4용지 문서 1장씩을 건네받은 것으로 조사됐다.



신사스토어
추위 박멸 목둘라
고품격 남성리류 전문 쇼핑몰
구매하기

검찰은 해당 지시를 받은 조 청장 등이 이후 계엄군을 돕기 위해 국회 출입을 통제하고 6개 기동대를 국회 인근으로 미리 이동 대기시켰다고 밝혔다. 특히 대통령의 국회 통제 지시에 따라 국회로 들어가는 사람을 전면 차단하고, 일부 선별적 출입을 허용한 뒤에도 조 청장은 "포고령을 따르지 않으면 우리들이 다 체포된다, 지시대로 하라"며 국회의원 등의 국회 출입을 다시 차단한 것으로 나타났다. 김 청장 역시 "서울청장이 지시한다. 포고령에 근거해 국회 출입을 통제하라"며 총 1740명의 경력을 배치해 국회 출입을 차단하려고 시도한 것으로 파악됐다.

또 이들은 여인형 국군방첩사령관으로부터 '이재명, 우원식, 한동훈 등' 주요인사를 체포하기 위해 위치를 파악해달라는 이야기를 듣고, 우중수 국가수사본부장의 도움을 받아 필요한 인력을 대기시키는 등 체포조에도 가담한 것으로 나타났다.

검찰은 이들이 중앙선관위 전산실 서버를 탈취하기 위해 K1 소총 5정 등으로 무장한 경력 114명을 과천청사로 출동시킨 사실도 확인, 비상계엄 선포 당일 헌법에 위배되는 국회 통제 등을 시도하며 내란에 가담한 것으로 보고 있다.

최미송 기자 cms@donga.com

Copyright © 동아일보. 무단전재 및 재배포 금지.

이 기사에 대해 어떻게 생각하시나요?



추위 박멸 양모 울 바지
고품격 남성리류 전문 쇼핑몰

신사스토어

함께 찾는 검색어

국회 인근 국회 출입
비상계엄 ... 비상계엄 ...

이 시각 추천뉴스

유인태 "한덕수, 문제 생기면 내 이름 빼달라고 켜... 뉴스1 · 5시간 전



[단독] 대법원 "내란특검 무제한 압색 우려"... 민... 한국일보 · 5시간 전



광동제약의 아심작! 멈추지 않는 기침, 가래엔... 광동제약



'국회의원 출입 차단 맞나' 문의에도...경찰청장 "지... SBS · 21시간 전



尹 지지율 42.4%...국민의 힘 41% 민주당 38.9% ... 한국경제 · 6시간 전



'현금 68억' 훔친 관리인 "비번 누르고 들어가"..... 뉴스1 · 1시간 전



응답률 22% 여론조사인 데...민주 36%-국힘 32... 이데일리 · 1시간 전



輿, 송영길 징역 2년 판결에 "민주당, 손절 말고 사... 연합뉴스 · 3시간 전



'관저 밖 윤석열'도 체포할 수 있나요? 전직 판사가 ... 하경례 · 1시간 전



Changes in Korean News Market

- Transition to online news from early in 2000s.
- Competition has shifted towards producing a larger volume of articles at a faster pace.
 - Quality issues: short length of articles and copying behaviors
- A shortage of journalists compared to the number of articles published.
 - The number of media professionals increased by 11% from 2000 to 2012, but the number of published articles surged by 498.6% (Song, 2020).
- In South Korea, despite the dominant role of the aggregators, these problems do not appear to be effectively addressed.

Objectives

- This paper aims to highlight two factors in the news selection process of news aggregators:
 - The presence of incomplete information in selecting issues
 - Being myopic in selecting news articles within an issue
- We also point out the problem of the aggregator's objective function being focused on maximizing clicks.
- Empirical questions:
 - Which issues are selected by the aggregators?
 - Which news articles are selected by the aggregators?
 - Do consumers prefer original news?

Online News Market (Cage et al., 2019)

- Studies online news production in France:
 - Collects all online news content from France in 2013 to measure the extent of copying.
 - Cluster news articles and measure originality of each article.
- However, the authors **do not consider the behavior of news aggregators** in the market.
- Our study employs similar methodology but considers the effect of news aggregators in the market.

Effect of News Aggregators

- Google news increases traffic to newspapers (Athey et al., 2017; Calzada and Gil, 2020; Chiou and Tucker, 2017)
 - Google News is a complement to overall news reading.
 - Newspapers have an incentive to opt-in.
- Theoretical Study: Jeon and Nasr (2016)
 - If readership expansion effect is positive, the aggregator increases quality chosen by each newspaper.
- No studies address **the news selection of aggregators** yet.

Contribution

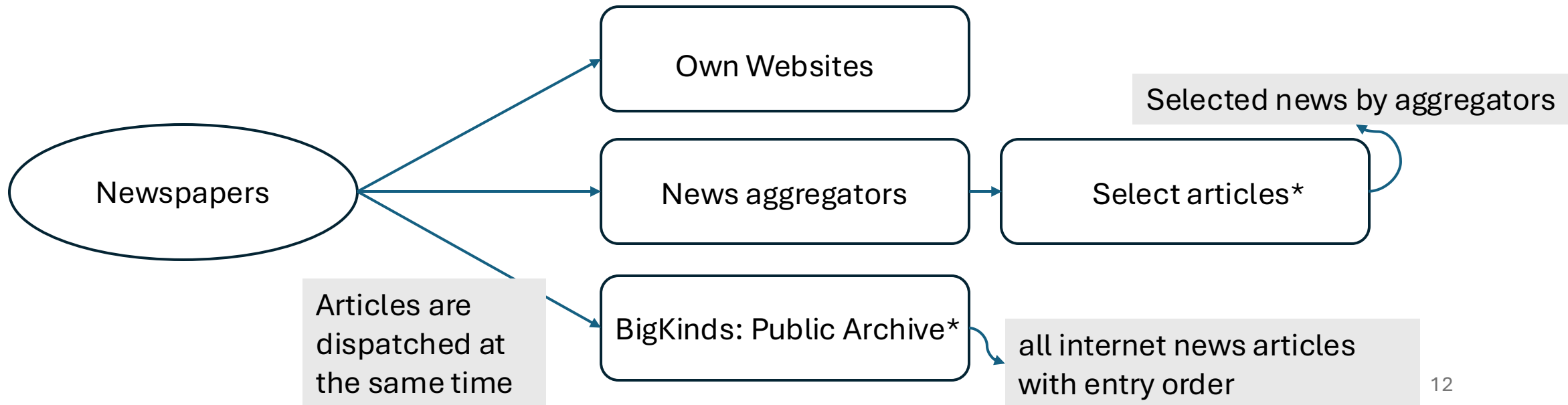
- The first study on news selection by aggregators
- Novel dataset
 - All internet news content and selection info by aggregators.
 - Sequential order of articles, and their depreciation over time.
- Platform's behavior in online news market
 - Incomplete information about importance of issues, and selection of issues.
 - Ad-financed news aggregators that maximize their profits (clicks)
- Unique case from South Korea
 - Dominance of news aggregators in online news market

Data Source

- Study period: 2015
 - No algorithm for personalizing news provision
- Internet news articles from **BigKinds**:
 - The largest news article database in South Korea: all news articles from 52 newspapers for the year 2015, leading to a total of 4,495,009 articles
- News text from the **two aggregators**:
 - All articles posted on the headline pages of the two aggregators during 2015, leading to 147,000 articles
- Consumer data from **Neilsen**:
 - individual/day level data with approx. 1.5 mil. observations per month

Dispatching Articles

- All newspapers in the dataset have content partnership agreements with both aggregators.
- They are **required to share all their articles with the aggregators**.
- **In-link service**: the aggregators show selected news articles on their own pages.
- Newspapers dispatch their articles at the same time to BigKinds and to the aggregators.
 - We observe **the entry order** of articles in BigKinds.



Methods

- **Clustering:**

- Building on Cagé, Hervé and Viaud (RES, 2019)
- Group news articles by topic through a clustering process (DBSCAN algorithm)
- Compute originality rate of each article within each cluster.

- **Identifying news selection:**

- the similarity between a news article in BigKinds and an article in the aggregators is calculated
- Due to some modifications in the aggregators' news text (headers, footers, and copyright notes), articles are considered identical if their Jaccard similarity score is above a certain level.

Description: Clusters

- The number of clusters in the analysis is 45,260.
- The average cluster has 20.7 articles, with duration of 2.28 days. It has 0.3 selected articles with 9.43 newspapers.
- Among clusters in our study, 15% had at least one article selected, and only 2% have the first article in the cluster selected.

Variable	N	Mean	Median	Min	Max
Cluster size	45260	20.71	15.00	10.00	2015.00
Duration (days)	45260	2.28	2.00	1.00	15.00
N of sel	45260	0.30	0.00	0.00	64.00
Select Dummy	45260	0.15	0.00	0.00	1.00
1st rank selected	45260	0.02	0.00	0.00	1.00
N of newspapers	45260	9.43	9.00	2.00	50.00

Description: Articles in the Clusters

- The total number of observations is 937,549, of which 13,573 articles were selected by the aggregators.
- Articles selected by the aggregators exhibit higher originality and longer text lengths.
 - The average originality of selected articles is 36.6%, while that of non-selected articles is approximately 28%.

	selected by aggregators		
	0	1	Total
N	923,976 (98.6%)	13,573 (1.4%)	937,549 (100.0%)
Originality	27.968 (27.605)	36.600 (29.437)	28.093 (27.652)
length	788.992 (478.252)	1,107.166 (563.088)	793.598 (481.091)
exclusive news dummy	0.001 (0.034)	0.056 (0.231)	0.002 (0.044)
breaking news dummy	0.002 (0.047)	0.005 (0.069)	0.002 (0.048)
byline dummy	0.727 (0.446)	0.908 (0.289)	0.729 (0.444)
soft news dummy	0.225 (0.418)	0.014 (0.117)	0.222 (0.416)
hard news dummy	0.730 (0.444)	0.953 (0.212)	0.733 (0.442)

Description: Selected Articles

- On average, selected articles are the 24.9th article within a cluster and are positioned at the 46.8th percentile.
- The rank of the first selected article is around 10.8, implying that aggregators are not likely to select the first article to break an issue.

Variable	N	Mean	Median	Min	Max
All selected articles					
Rank	13,573	24.88	10	1	1,949
Percentile	13,573	46.84	45.45	0.43	99.7
First selected article dummy	13,573	0.49	0	0	1
Originality	13,573	36.6	35.32	0	100
First-selected articles in clusters					
Rank	6,694	10.82	7	1	338
Originality	6,694	54.21	54.11	0	100

Regression Results: Cluster Level

- Y: selection dummy
- X: cluster characteristics
- The probability for a cluster to be selected increases with
 - The size of the cluster
 - The number of newspapers which contribute articles to the cluster
 - The average originality of articles in the cluster
 - The average length of articles in the cluster

	OLS_1	OLS_2	Probit_1	Probit_2
main				
Cluster size	0.0003** (0.000)	0.0013** (0.000)	-0.0011** (0.000)	0.0125*** (0.001)
Duration (days)	0.0186*** (0.002)	0.0072** (0.002)	0.0775*** (0.007)	0.0094 (0.008)
N of newspapers	0.0179*** (0.000)	0.0099*** (0.001)	0.0757*** (0.002)	0.0331*** (0.003)
Avg orig		0.0019*** (0.000)		0.0165*** (0.001)
Avg length		0.0001*** (0.000)		0.0004*** (0.000)
Byline ratio		0.0022 (0.006)		0.1878*** (0.045)
Excl. dummy		0.3633*** (0.017)		1.1763*** (0.050)
Break dummy		0.1359*** (0.015)		0.3611*** (0.046)
Soft news dummy		-0.1171*** (0.004)		-1.1410*** (0.035)
Constant	-0.0697*** (0.004)	-0.0989*** (0.009)	-1.9862*** (0.020)	-2.5596*** (0.053)
R-squared	0.082	0.143		
N	45,260	45,260	45,260	45,260
ll	-15404.584	-13845.369	-17351.218	-15442.620

* p<0.05, ** p<0.01, *** p<0.001

Regression Results: Article Level

- Probit model
 - Y: selection dummy
 - X: article characteristics
- No significant relationship between the selection and rank
 - No premium for being the first to report on an issue
- Once an issue is selected, the aggregator tends to choose higher quality articles.
 - Originality, article length, and byline dummy are positively correlated.

	All articles		Articles in selected clusters	
	Model_1	Model_2	Model_3	Model_4
Selected				
Rank	0.0000 (0.000)	-0.0001 (0.000)	-0.0022** (0.001)	-0.0020** (0.001)
Originality	0.0031*** (0.000)	0.0014*** (0.000)	0.0038*** (0.000)	0.0024*** (0.000)
length	0.0003*** (0.000)	0.0002*** (0.000)	0.0002*** (0.000)	0.0002*** (0.000)
byline dummy	0.4466*** (0.017)	0.3457*** (0.020)	0.4280*** (0.022)	0.3428*** (0.024)
exclusive news dummy		1.6643*** (0.055)		1.7024*** (0.058)
breaking news dummy		0.1729 (0.092)		-0.0027 (0.103)
soft news dummy		-0.8809*** (0.072)		0.2518* (0.123)
Constant	-2.8881*** (0.026)	-2.7532*** (0.037)	-2.0886*** (0.043)	-2.0961*** (0.052)
Log-Likelihood	-67642.152	-58038.345	-47704.997	-42346.118
N	937,549	812,280	204,070	175,404

* p<0.05, ** p<0.01, *** p<0.001

Marginal effects of ranking on selection probability

- Even within the selected clusters, the effect of ranking on selection prob. is very low.
- When the rank changes from 1 to 20, the percentage reduction in selection probability is 4.7.
- This implies that almost **no premium for breaking the first story** and a **strong incentive to copy**.

<Table 7> Marginal effects estimation results

Rank	<u>dy/dx</u>	Standard Error	P>z
1	-0.00029	0.000106	0.007
20	-0.00027	9.66E-05	0.005
Average rank	-0.00026	8.83E-05	0.003
Observations		175,404	

Note: Marginal effects are estimated using the parameter estimate of the rank variable in column (4) in Table 6.

Consumer preferences

	(1) PV	(2) Duration	(3) PV	(4) Duration
Originality	-0.027* (0.016)	0.207*** (0.038)	-0.016 (0.017)	0.220*** (0.043)
Length	0.009* (0.005)	-0.009 (0.012)	0.009* (0.005)	-0.006 (0.013)
Length^2	-0.000* (0.000)	0.000 (0.000)	-0.000* (0.000)	0.000 (0.000)
Byline	-0.007 (0.012)	0.019 (0.032)	-0.006 (0.012)	0.023 (0.032)
N. of Excl./Break	0.055** (0.017)	0.028 (0.043)	0.054** (0.017)	0.020 (0.044)
N. of articles	0.003 (0.010)	-0.003 (0.028)	0.002 (0.010)	-0.008 (0.028)
N. of newspapers	0.004 (0.016)	0.002 (0.041)	0.004 (0.016)	0.001 (0.041)
N. of news agencies	-0.008* (0.004)	0.019* (0.011)	-0.008* (0.004)	0.019* (0.011)
N. of clusters	0.037** (0.016)	-0.223*** (0.047)	0.041** (0.016)	-0.217*** (0.048)
Diff. political stance			-7.145** (3.524)	-8.375 (8.635)
Avg. cluster size			0.033 (0.025)	0.074 (0.069)
Max cluster size			0.022 (0.055)	-0.156 (0.147)
Observations	380,450	380,450	380,450	380,450
r2	0.003	0.003	0.003	0.003

- Panel FE model results
 - Y: PV or Duration per day of users
 - X: daily average of aggregator's characteristics
- In column (1), a lower originality rate is significantly correlated with the increase in the number of clicks but is not significant in column (3).
- In columns (2) and (4), as the originality increases, duration also increases.
- Duration-based metrics may generate stronger incentives to invest in news quality.

Summary

- On average, the rank of the first selected article within each selected cluster is 10.8th article
- Cluster-level analysis:
 - Higher selection probability for issues with larger sizes (n. of articles) and with higher newspaper participation.
- Article-level analysis:
 - Selection probability has no relationship with article rank, implying that there is no premium for being the first rank in the issue.
 - Low marginal effect of rank implies the strong incentive to copy others' news.
- Consumer preferences:
 - Consumers take time to read and pay more attention to original news articles.

Concluding Remark: Why no premium for breaking a story?

- Incomplete information and belief updating:
 - The aggregators do not know which issues are important.
 - They update their belief about the importance of a issue as additional articles arrive.
- Being myopic:
 - After a good number of articles accumulate on a topic, aggregators recognize its importance but tend to prioritize selecting the most recent articles over the initial ones.
- Furthermore, if the aggregators maximize clicks, they have little incentives to encourage originality.