

News Reading Publics & Audience Fragmentation: Evidence from Online India (2014-2018)



Subhayan Mukerjee | Dissertation Defense | May 27, 2020

Outline

- Motivation
- The Indian Context
- Theoretical Framework
- Data and Methods
- Findings
- Discussion

Motivation

Motivation

- India as an understudied sociopolitical context in the news consumption literature
- India as a case study of the global south – in a field that is largely dominated by US-centric literature
- Need for theory (and methods) informed by a non-western context

Why India?

- India is a useful foil for theory-building
- India isn't unique – a similarly diverse country (ethnically, religiously, culturally, or linguistically) would “work”
- World's largest democracy that is still under-represented in existing news consumption research
- There is very little we know about how Indians consume news

Why India?

- In what ways is India substantively different from a Western country from the perspective of news consumption research?

The Indian Context

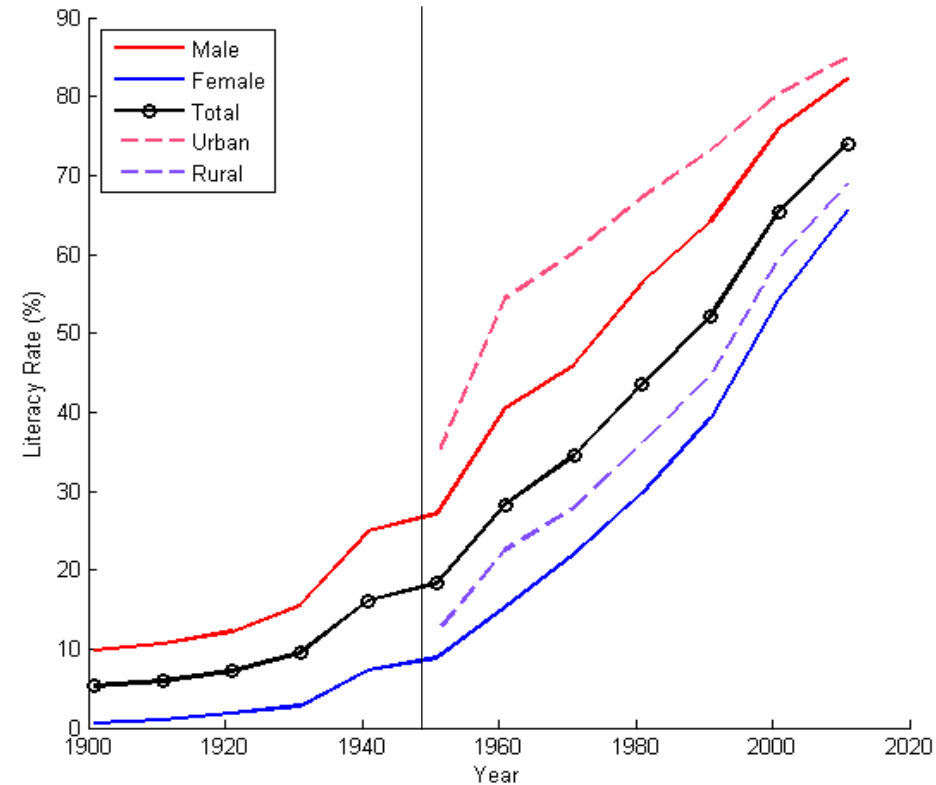
An (embarrassingly) Quick Introduction

- Population of 1.3 billion
- High cultural heterogeneity
 - 780 languages (excluding dialects)
 - 23 languages deemed official in the Constitution (including English)
 - 29 states formed around these cultural identities
- Culture is a big determinant of what drives people to media sources, but is obviously not the only factor

An (embarrassingly) Quick Introduction

- Other useful indicators
 - Literacy
 - Internet penetration

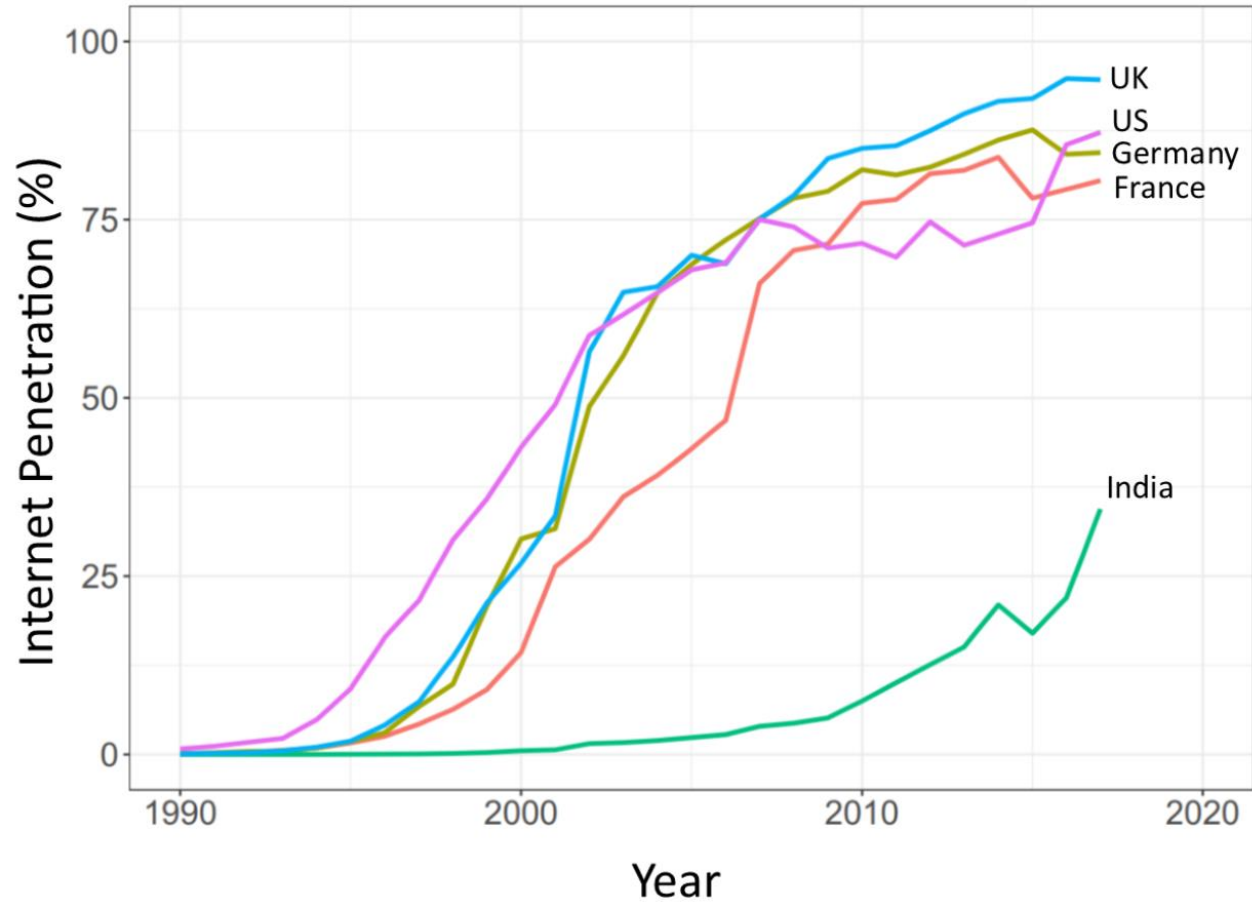
Literacy: Substantial Growth



The vertical line indicates the year of independence.

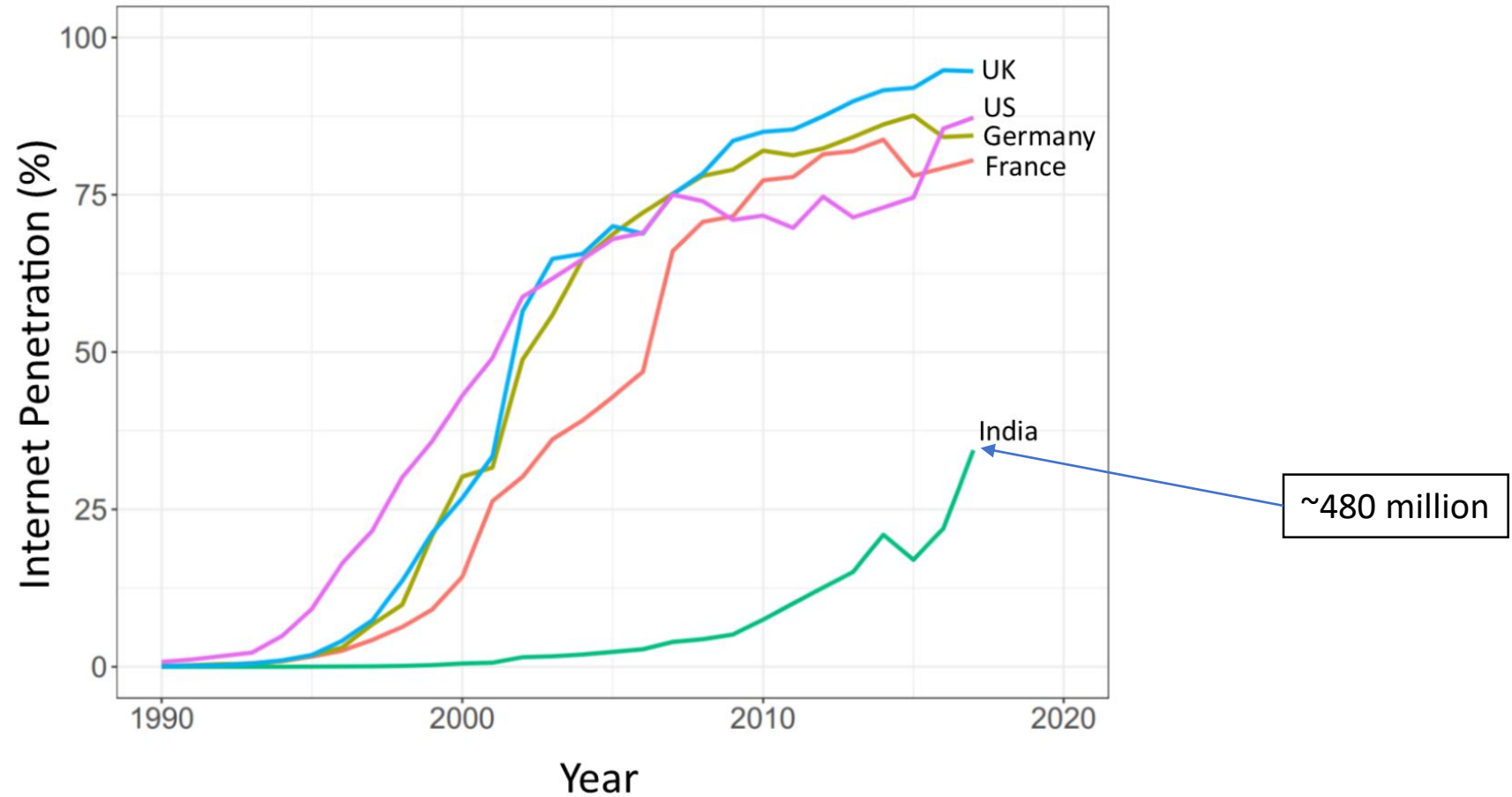
India Literacy trends (Source: Census data)

Internet Penetration: Slow Growth



Source: Internet Telecommunication Union of the United Nations

Internet Penetration: Slow Growth



Source: Internet Telecommunication Union of the United Nations

Indian Culture: Regional vis-à-vis National

- Socially:

- Different festivals, sets of holidays, languages, cuisine

“Scotland is more like Spain than Bengal is like Punjab” – Sir John Strachey, 1888

- Politically:

- Parties and issues vary between states
- Regional parties and issues can be very different from national parties and issues

Indian Culture: Regional vis-à-vis National

- Indians identify with their country and with their region
- This duality potentially shapes news consumption patterns

Political Systems: Indian vis-à-vis Western

Western democracies	India
Pre-ponderance of formal organizations	Formal as well as informal (for eg. Clientelist, quasi legal, or downright illegal) politics
Nationally integrated political systems	Sub-nationally variegated political systems
Relatively consolidated, settled, or established patterns of political cleavage and order	Evolving, flexible, or contingent patterns of political cleavage and order – difficult to capture in terms of traditional “left-right” vectors of power

(Chakravartty & Roy, 2013)

Evolution of Indian Media

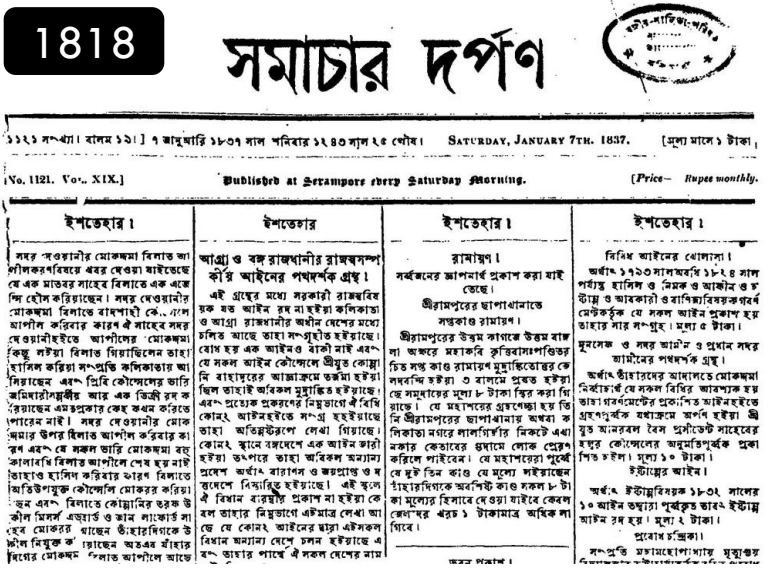
- Media imperialism (Fejes, 1981; Schiller, 1976)
- An Indianizing pushback (Thussu, 1998, 2007; Chadha & Kavoori, 2000)*
- What were these phases?

*see also the cultural proximity literature (Pool, 1977; Straubhaar, 1991)



Print Media in Colonial India

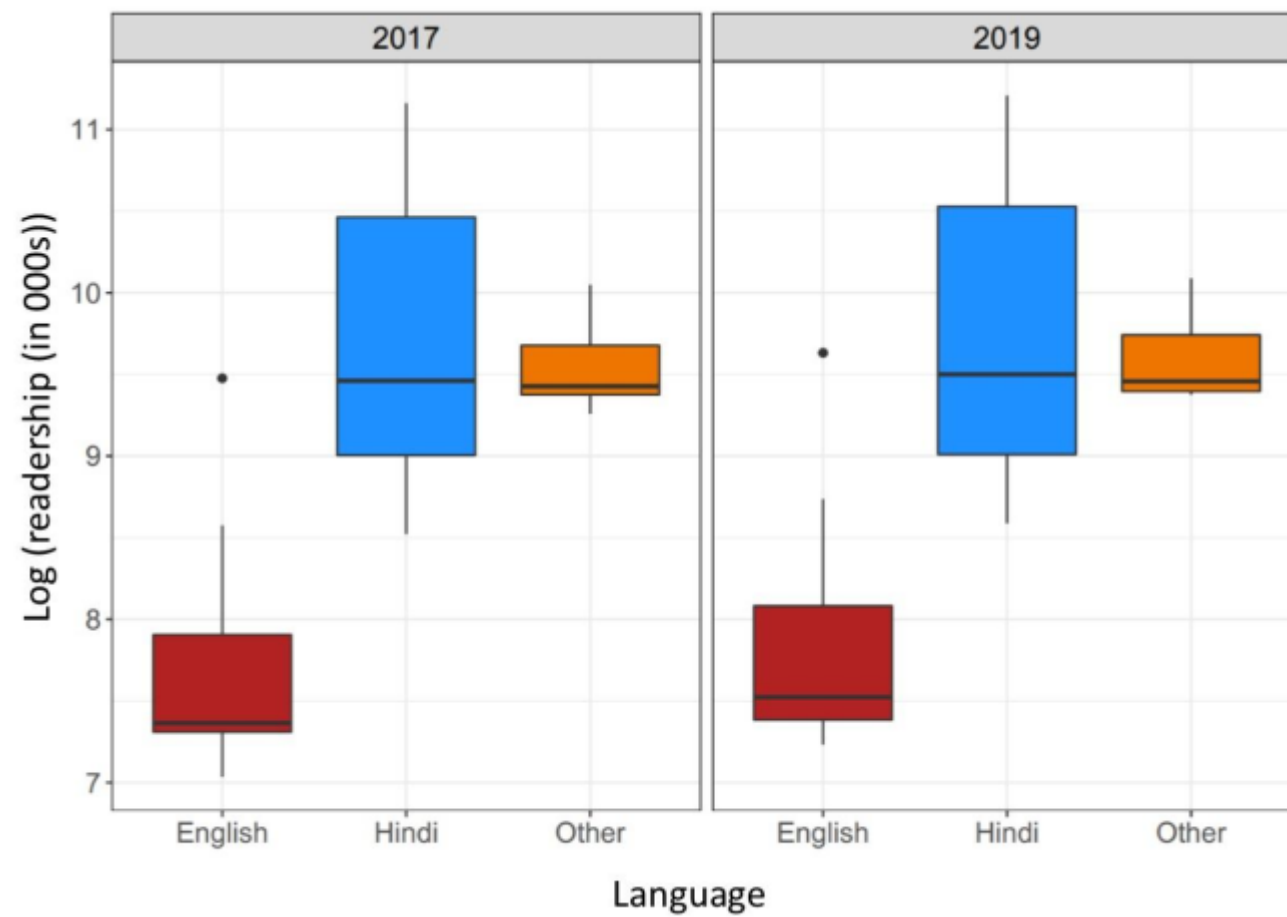
- The printing press as an agent of British colonialism (Athique, 2012)
- The printing press as a counter-agent against British colonialism (Sengoopta, 2016)



- The divide between “English” media and “vernacular” media

Image sources: University of Heidelberg Archives, Banglapedia

Print Media More Popular in Vernacular Than in English



Readership numbers of the top 10 most popular newspapers by language (source: Indian Readership Survey)

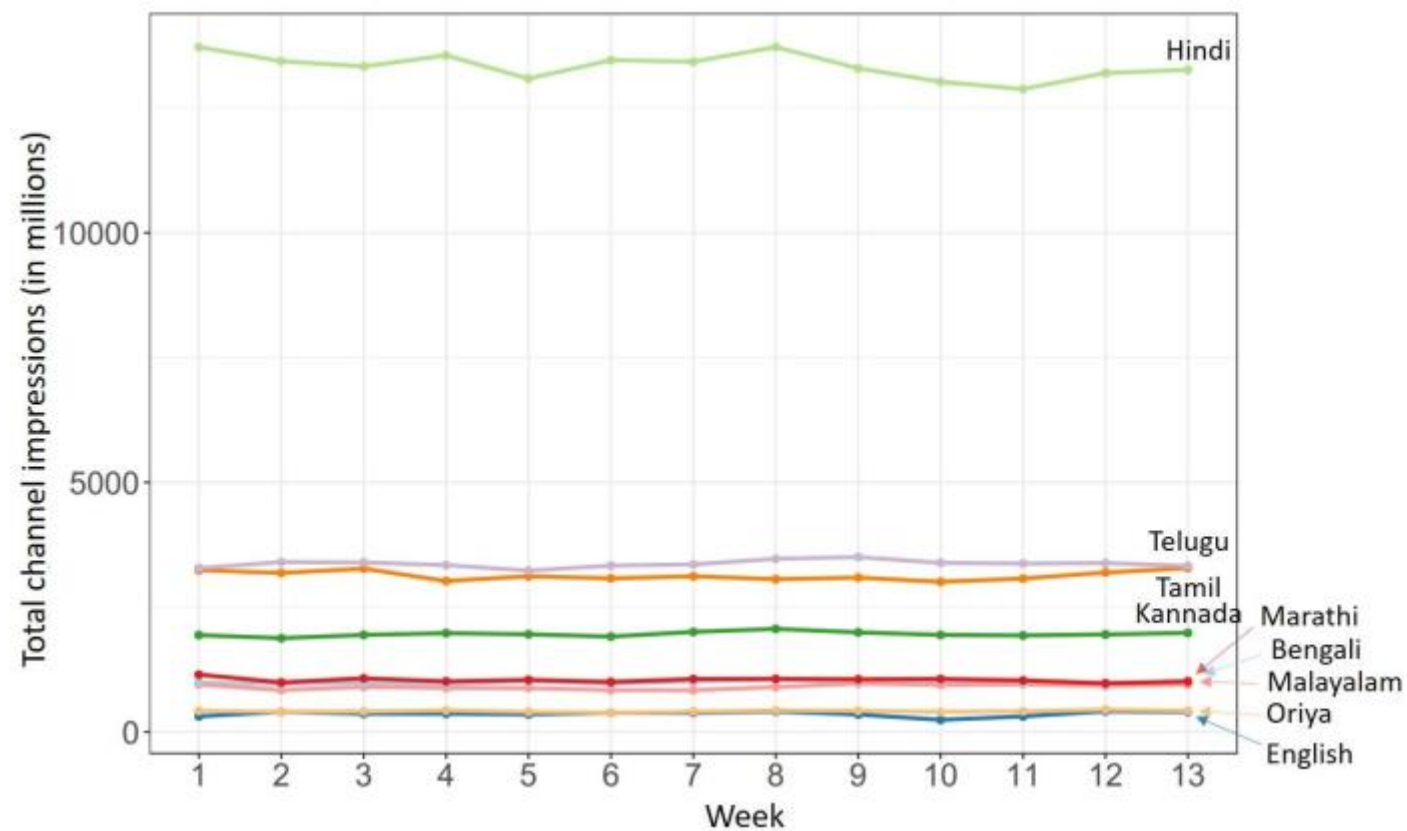


Television: Statism and Privatization

- 1965-1990: Statism (DD)
- 1988: Private programming began (NDTV)
- 1991: Economic reforms and *corporate media imperialism*
 - STAR TV (1991) and Rupert Murdoch (1993)
- Similar Indianization (Thussu, 2008)
 - Zee TV in North India, Sun TV in South India

Image sources: Livemint, NDTV, indiantelevision.com

TV Media More Popular in Vernacular Than in English



TV channel impressions over 13 consecutive weeks in 2018 (source: Broadcast Audience Research Council)

1995



2000s



Internet in India

- 1995: Opened to the public
- 1998-2004: Telecom sector deregulated
 - Increased influx of western content
 - A similar “Indianization”: viral websites, hyper-partisan outlets, streaming services
- Online Indian audiences are more urban, affluent, educated, and English literate

Image sources: Indian Express, Airtel, Financial Times, mouthshut.com

Media Structure: Indian vis-à-vis Western

	Western democracies	India
Media ownership	<p>Consolidation and convergence</p> <p>Dominance of transnational corporate capital</p>	<p>Initial signs of vertical integration, along with regional fragmentation</p> <p>Variegated forms of capital (transnational, domestic, non-corporate)</p>
Media structure	Nested within an apex national media	Polycentric: multiple media systems

(Chakravartty & Roy, 2013)

Theoretical Framework

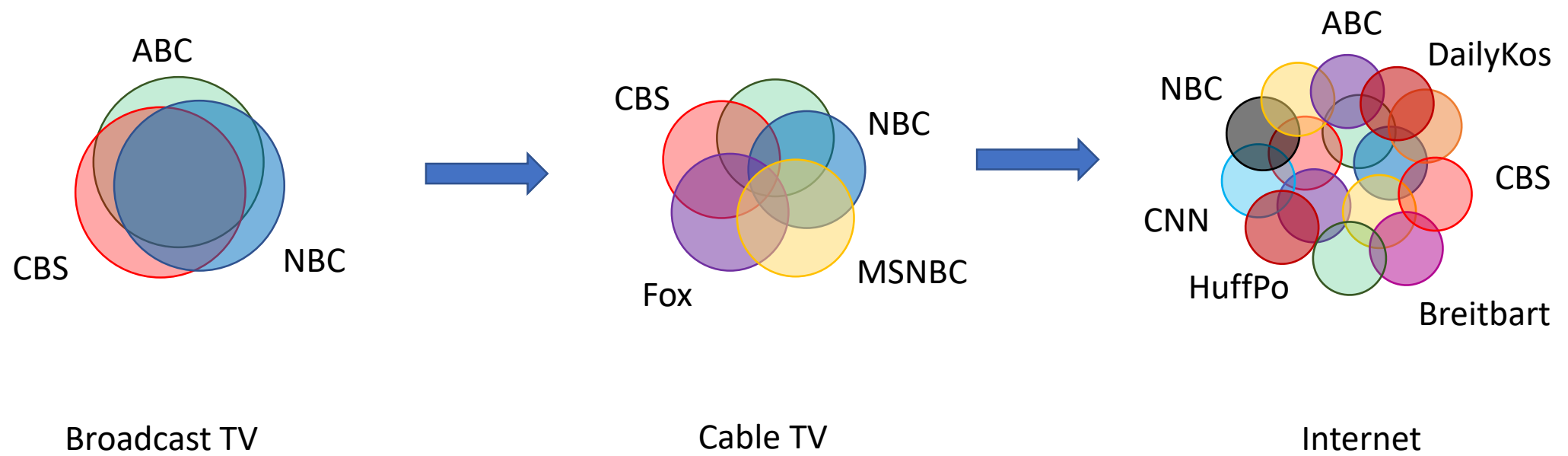
Theoretical Framework

Given the characteristics of the Indian media landscape, what theoretical framework is best positioned to understand patterns of news consumption?

And how can that make us rethink notions of audience behavior?

Audience Fragmentation in the West

- Audiences were distributed over a wider larger set of outlets



Audience Fragmentation in India

- TV
 - 2 channels till 1991, serving only big cities
 - < 70% penetration in 2018
- Newspapers
 - Historically low literacy
- Internet
 - < 40% today

Audience Fragmentation in India

- Historically, Indian audiences
 - Comprised a small section of the population
 - Were fragmented geographically
 - Were segregated linguistically/culturally

Audience Fragmentation in India

- Historically, Indian audiences
 - Comprised a small section of the population
 - Were fragmented geographically
 - Were segregated linguistically/culturally

- Enter News Reading Publics

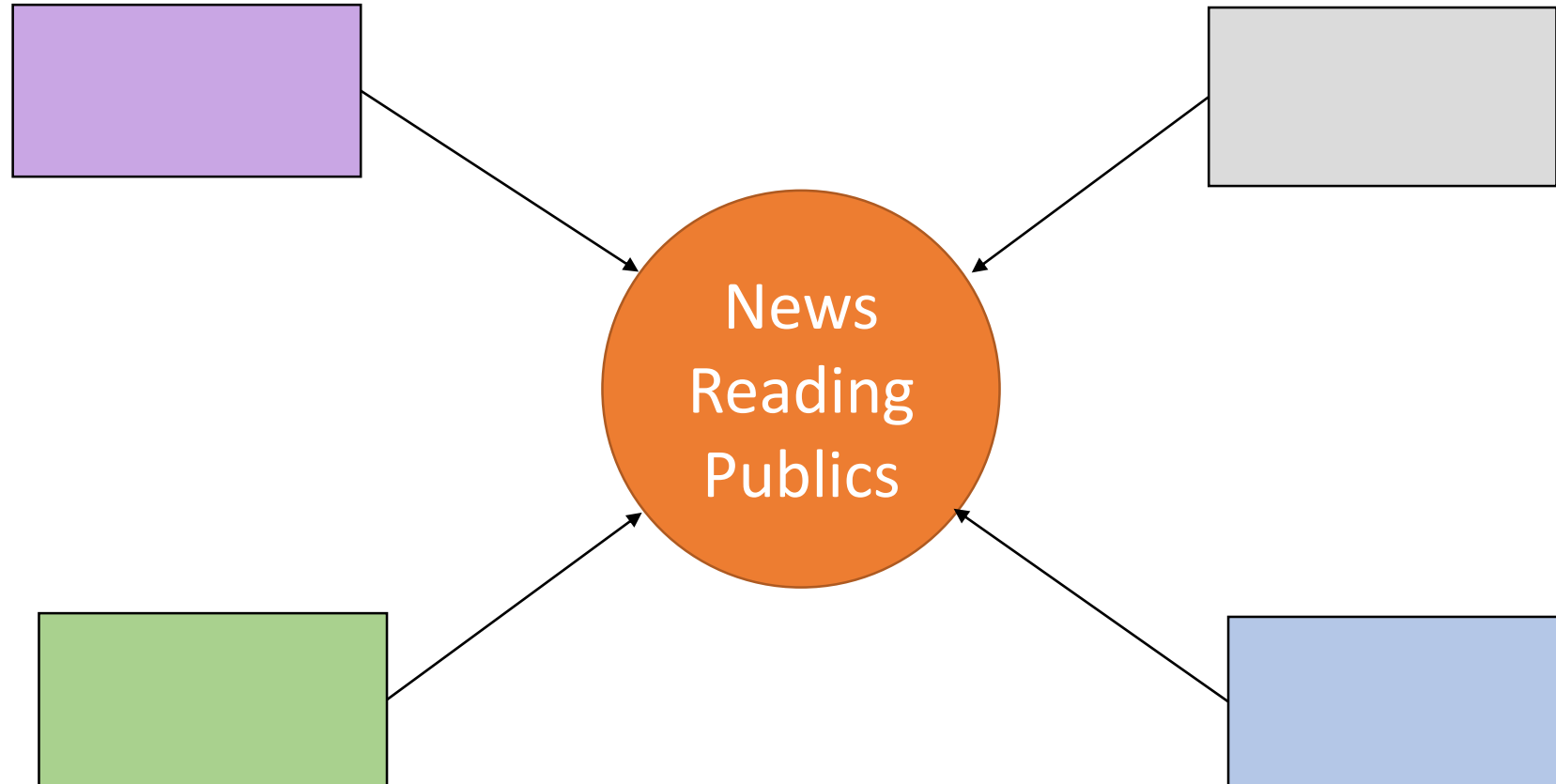
What is a News Reading Public?

- A group of news consumers who share access to the same set of media sources
- Could be due to:
 - Shared cultural markers like language
 - Shared issues they are interested in
 - Shared expectations and gratifications
 - Share identities

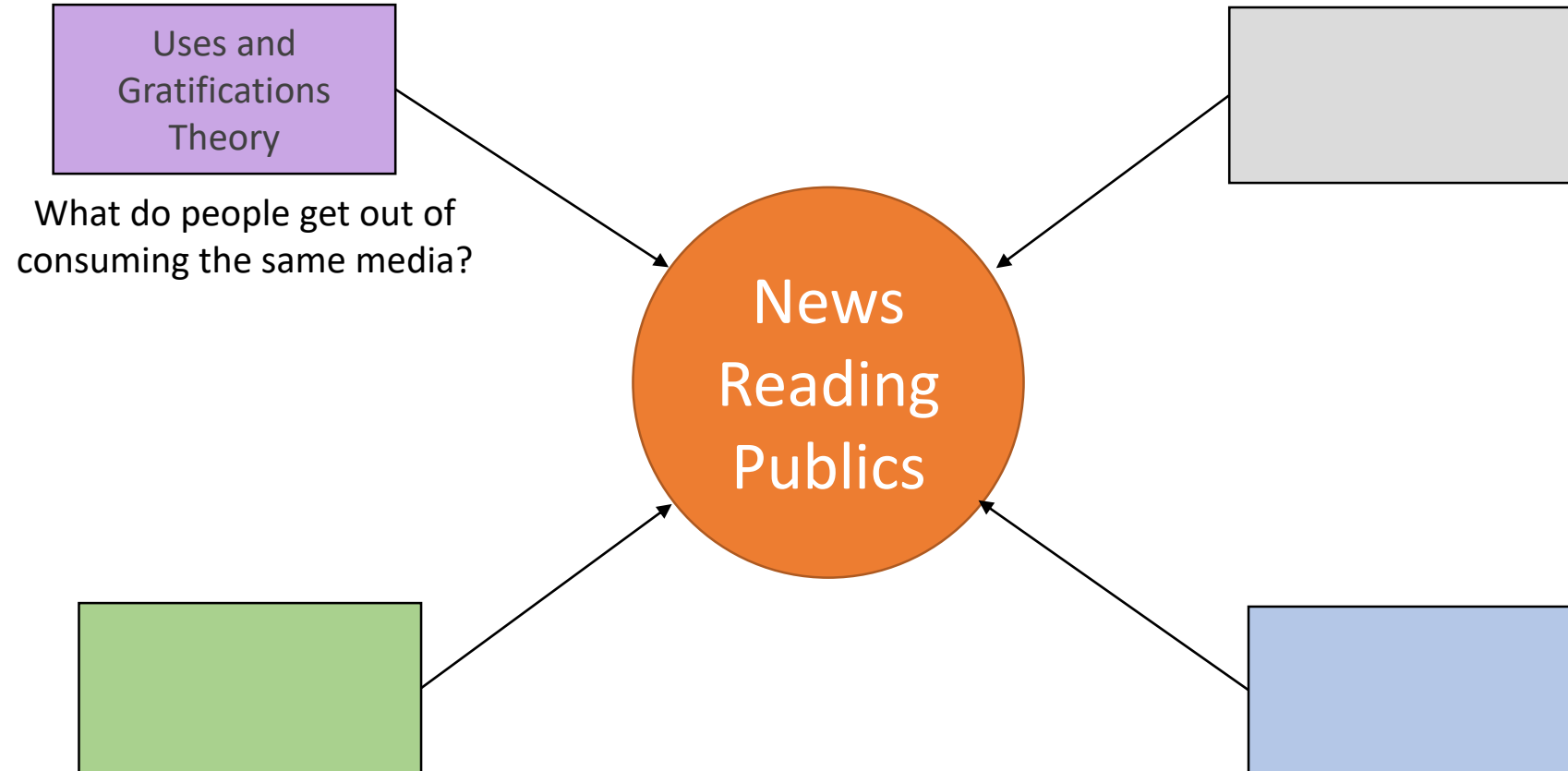
News Reading Publics in India

- Useful to start with the national / regional media divide
- The dual role of the average news consumer of India
 - Consumer of national media
 - Consumer of regional media
- Belong to different “news reading publics”

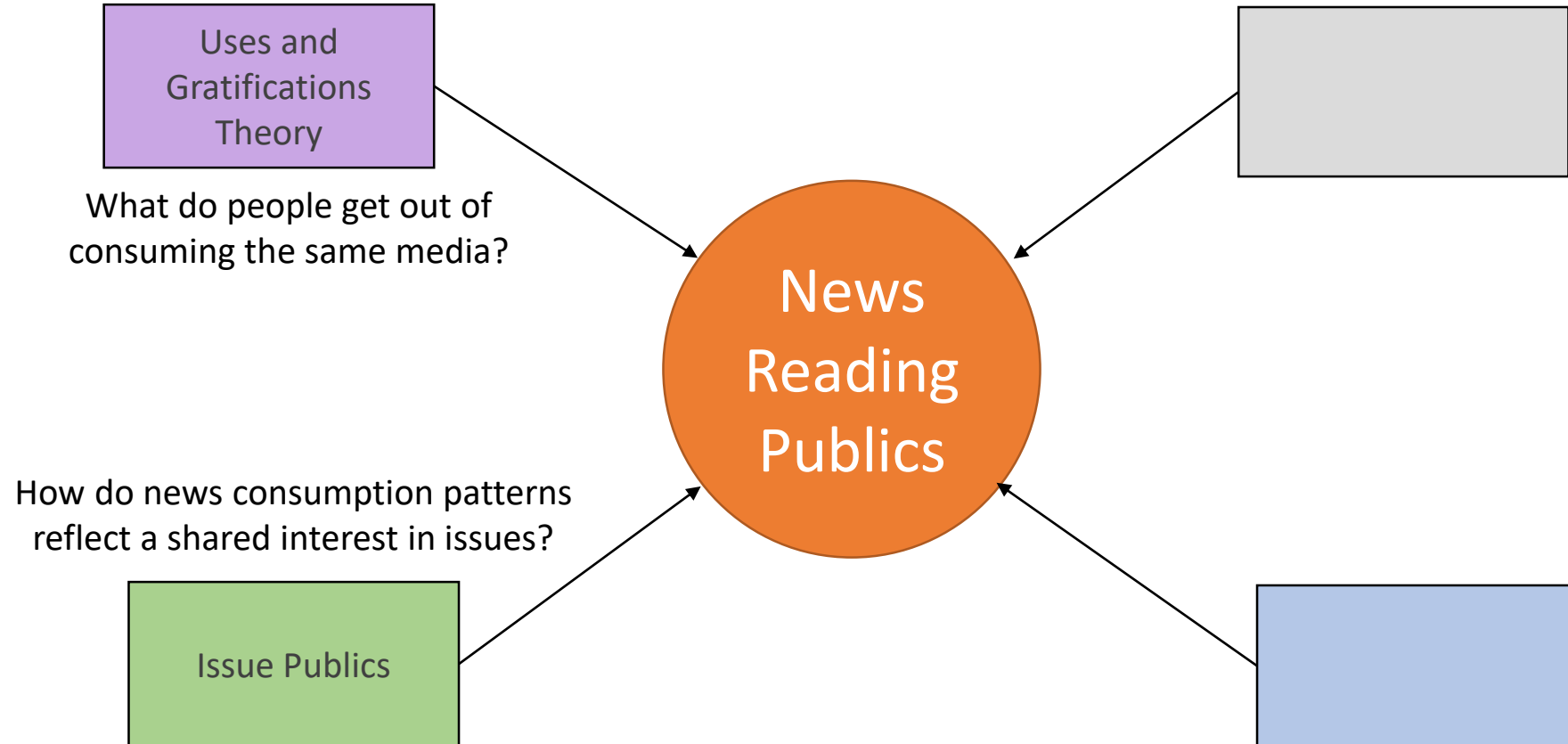
Theoretical Framework



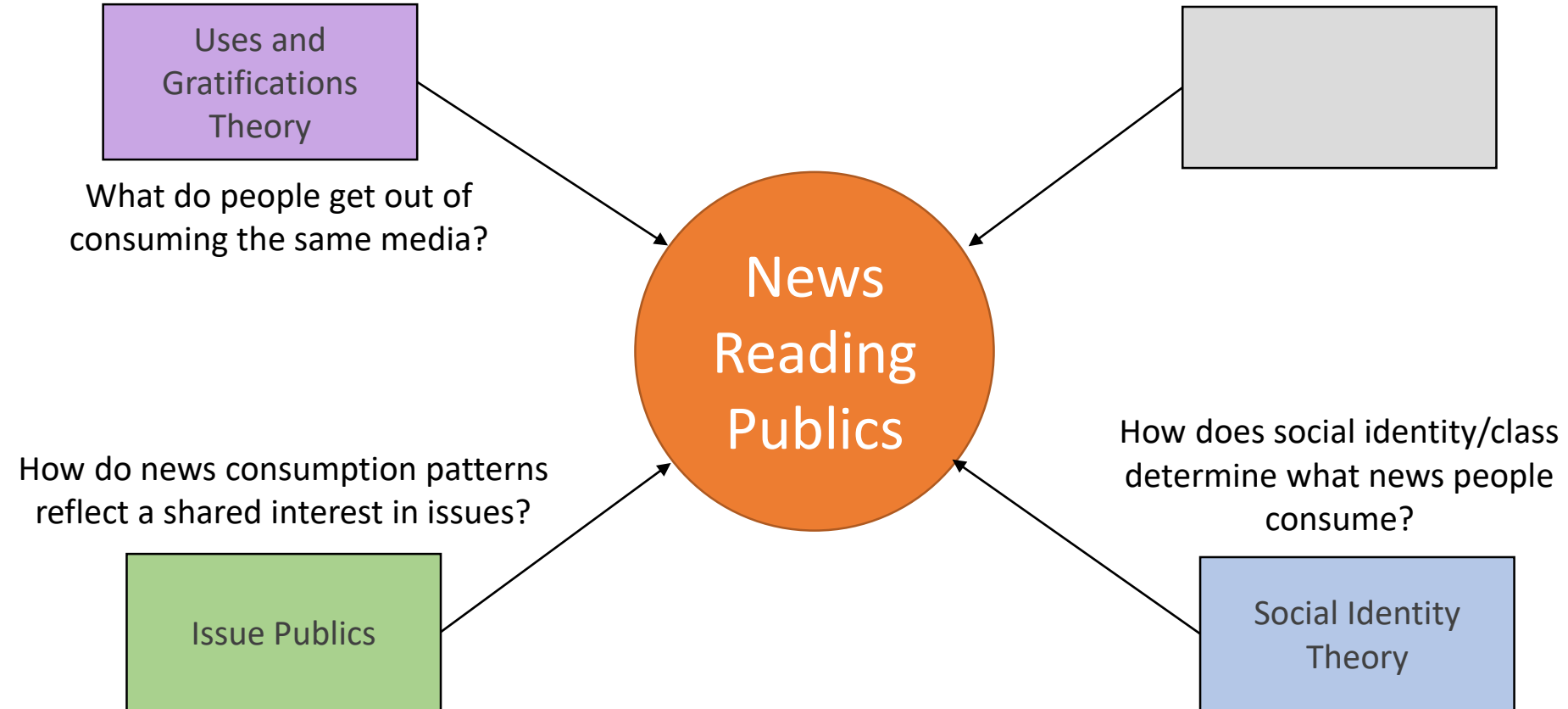
Theoretical Framework



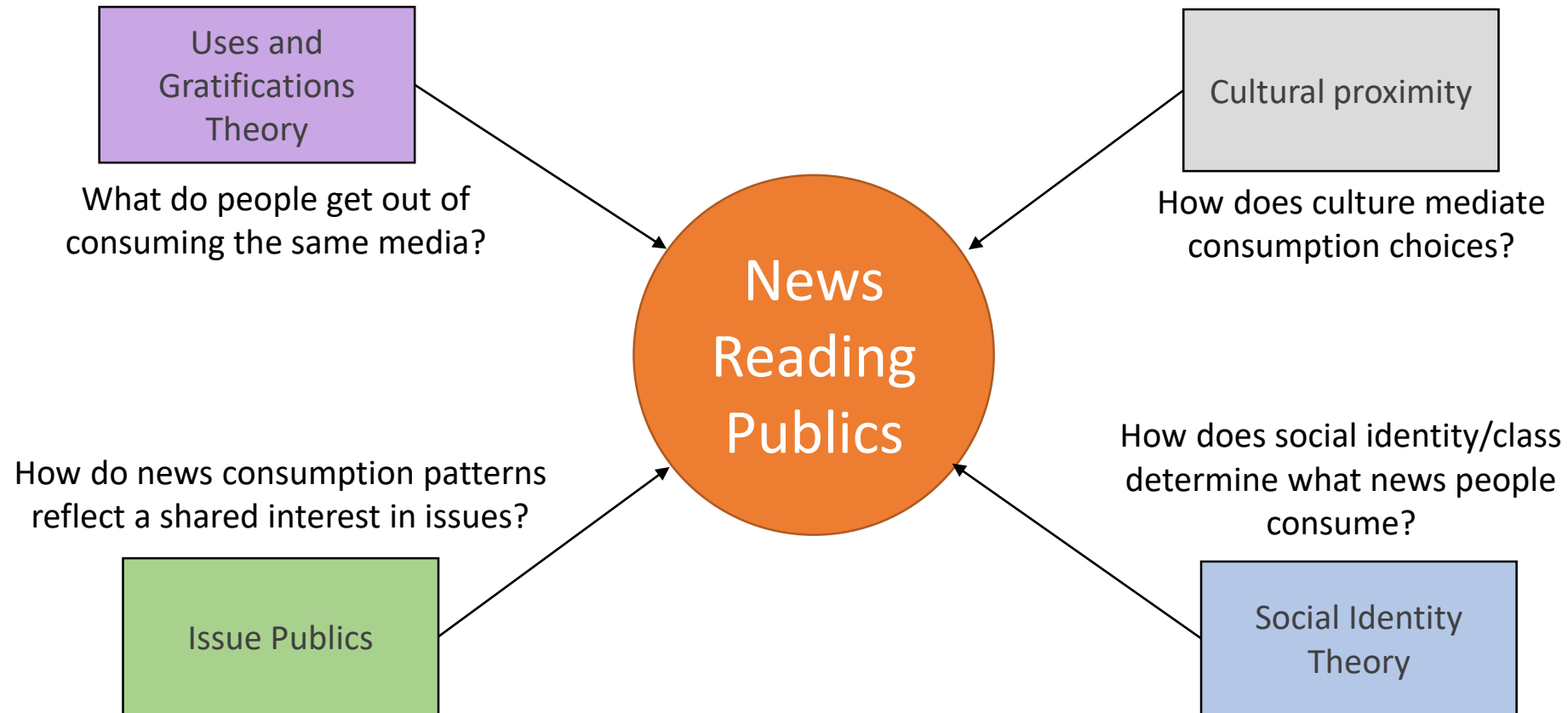
Theoretical Framework



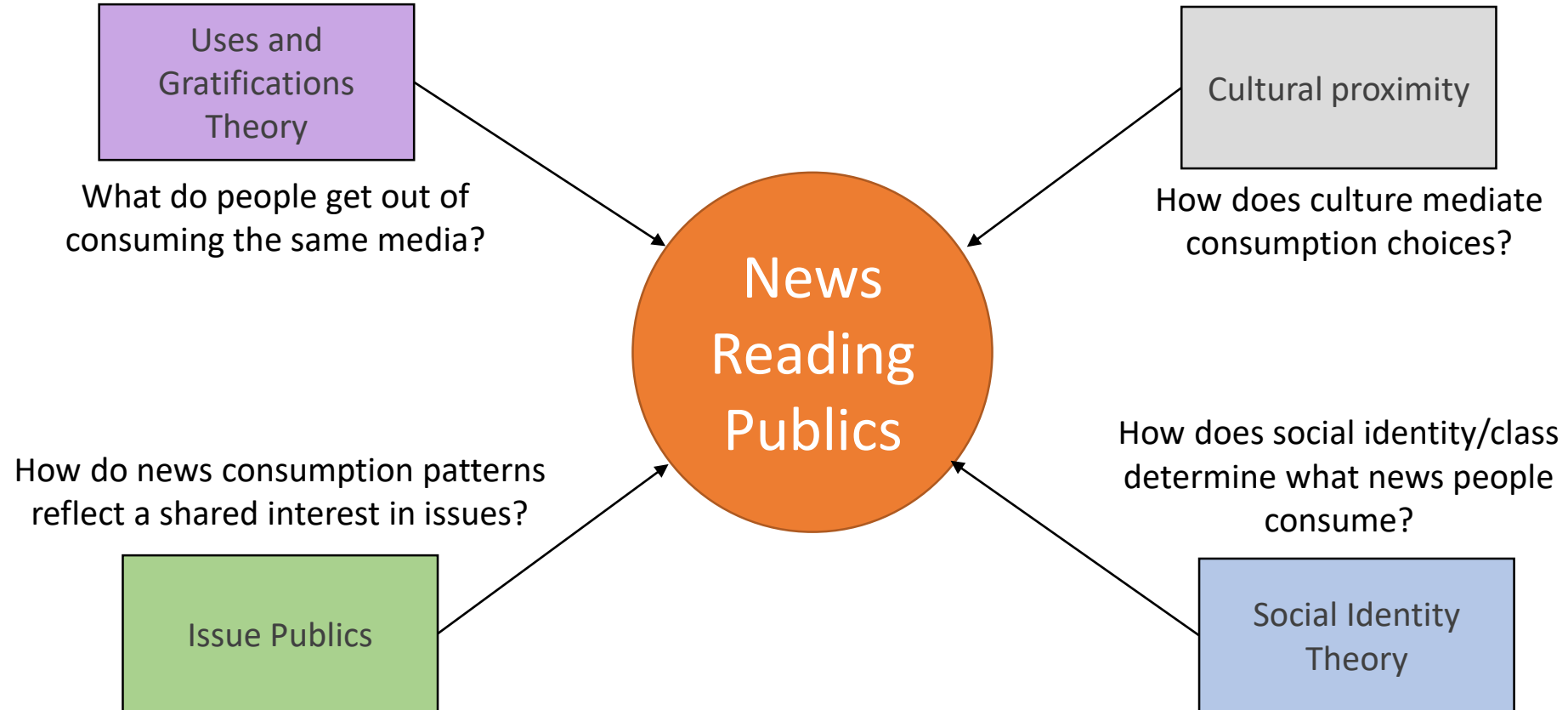
Theoretical Framework



Theoretical Framework



Theoretical Framework



Hypotheses

H1: The media consumption landscape in India is segregated along linguistic lines

Hypotheses

H1: The media consumption landscape in India is segregated along linguistic lines

H2: Vernacular news reading publics will have smaller overlap with each other than with national news reading publics

Hypotheses

H1: The media consumption landscape in India is segregated along linguistic lines

H2: Vernacular news reading publics will have smaller overlap with each other than with national news reading publics

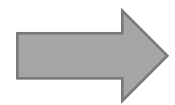
H3: The presence of national English news reading publics reduces fragmentation in the online Indian space

Hypotheses

H1: The media consumption landscape in India is segregated along linguistic lines

H2: Vernacular news reading publics will have smaller overlap with each other than with national news reading publics

H3: The presence of national English news reading publics reduces fragmentation in the online Indian space



Rethink our normative (western) understanding of news consumption

Data and Methods

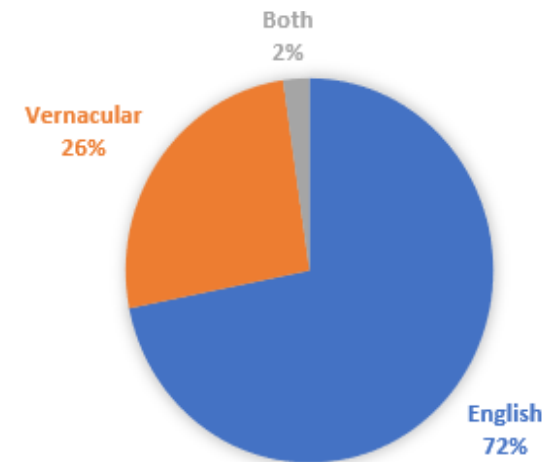
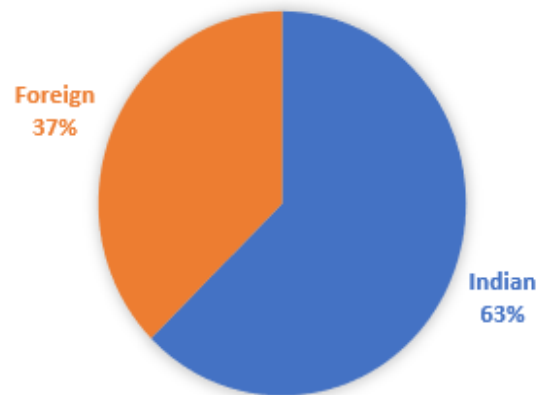
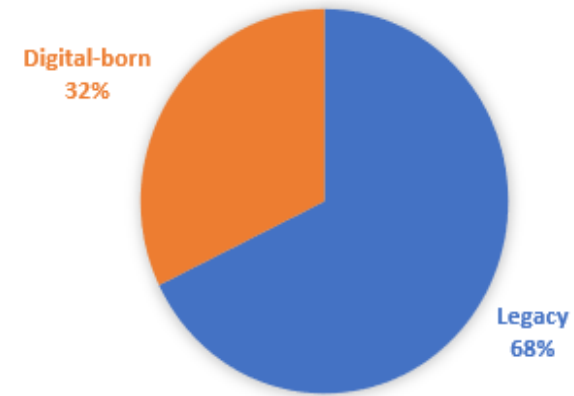
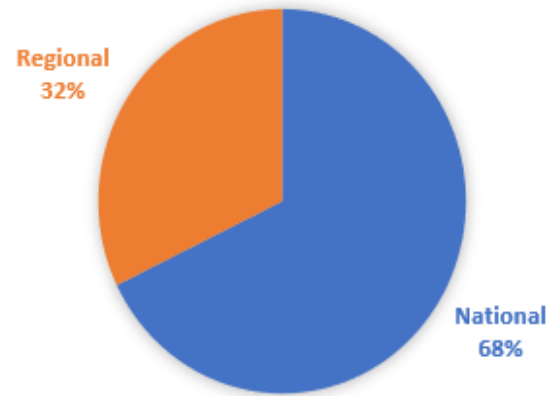
Data

- Obtained from ComScore
 - Browsing patterns over a period of 45 months (Oct 2014 – June 2018)
 - Only news websites that have a minimum reach of 0.1% of the month's audience
- Desktop browsing data, **not mobile**

Data

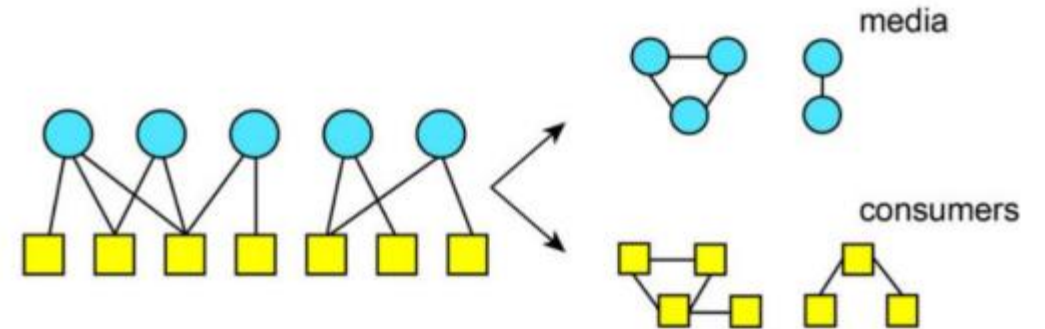
- Three monthly metrics (45 months)
 - Audience reach - number of unique visitors to an outlet
 - Cross-visiting - number of unique visitors to every pair of outlets
 - Average time per visitor
- 352 media outlets in total, 174 appear every month

Data: Media Outlets by Type



Methods: Networks of Audience Overlap

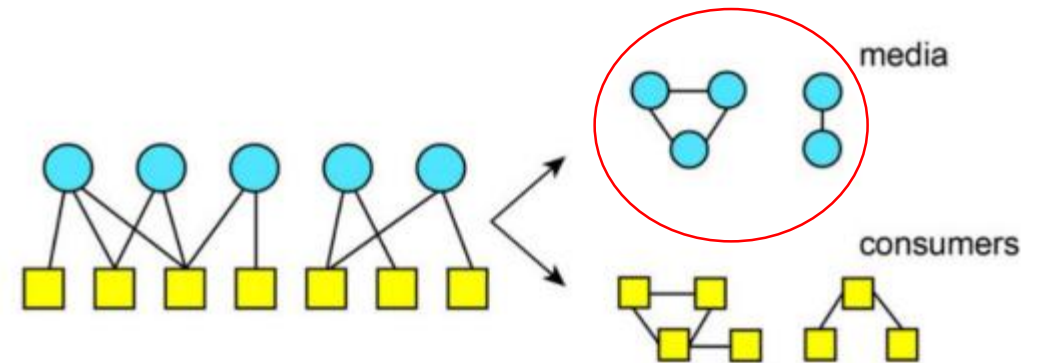
- Audience overlap networks
 - Each node is a news outlet
 - Edge between nodes denotes audience overlap
 - The weight of the edge is the actual strength of overlap



Source: Mukerjee et al. 2018

Methods: Networks of Audience Overlap

- Audience overlap networks
 - Each node is a news outlet
 - Edge between nodes denotes audience overlap
 - The weight of the edge is the actual strength of overlap

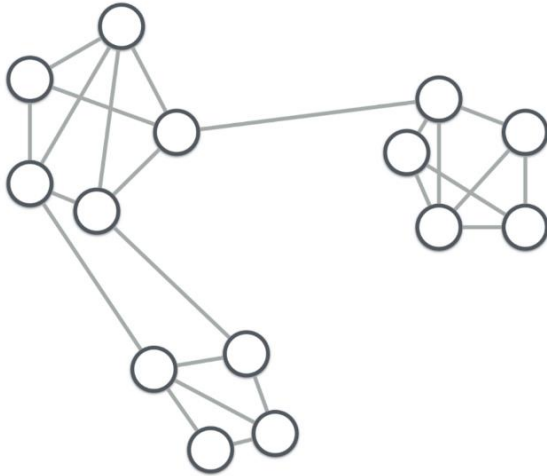


Source: Mukerjee et al. 2018

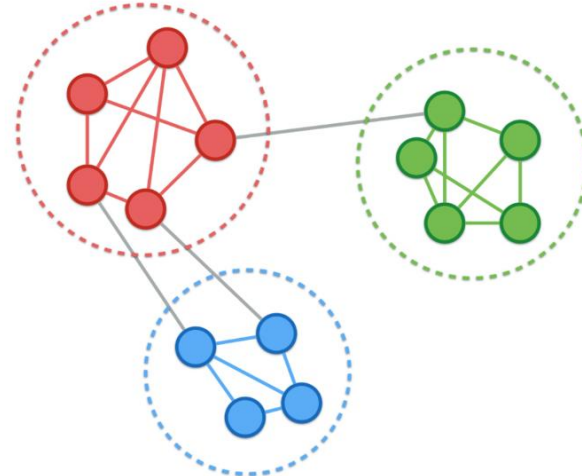
Methods: Network Analysis

- Identifying news reading publics using network analysis
 - Community detection (Pons & Latapy, 2006)

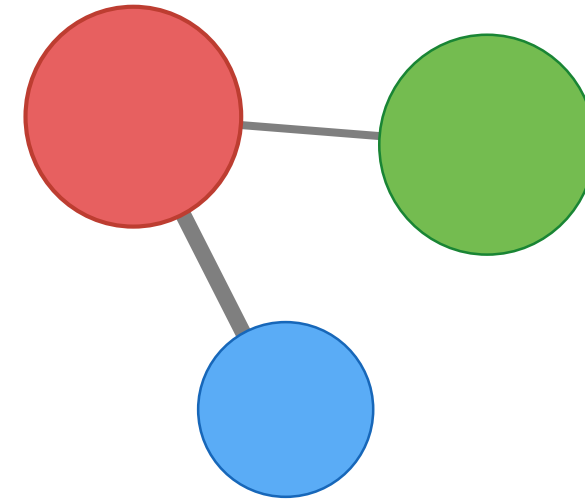
Methods: Network Analysis



Raw Network



Communities



Community Network

Methods: Network Analysis

- Identifying news reading publics using network analysis
 - Community detection (Pons & Latapy, 2006) with a methodological improvement
 - Evaluation of the “goodness” of community structure

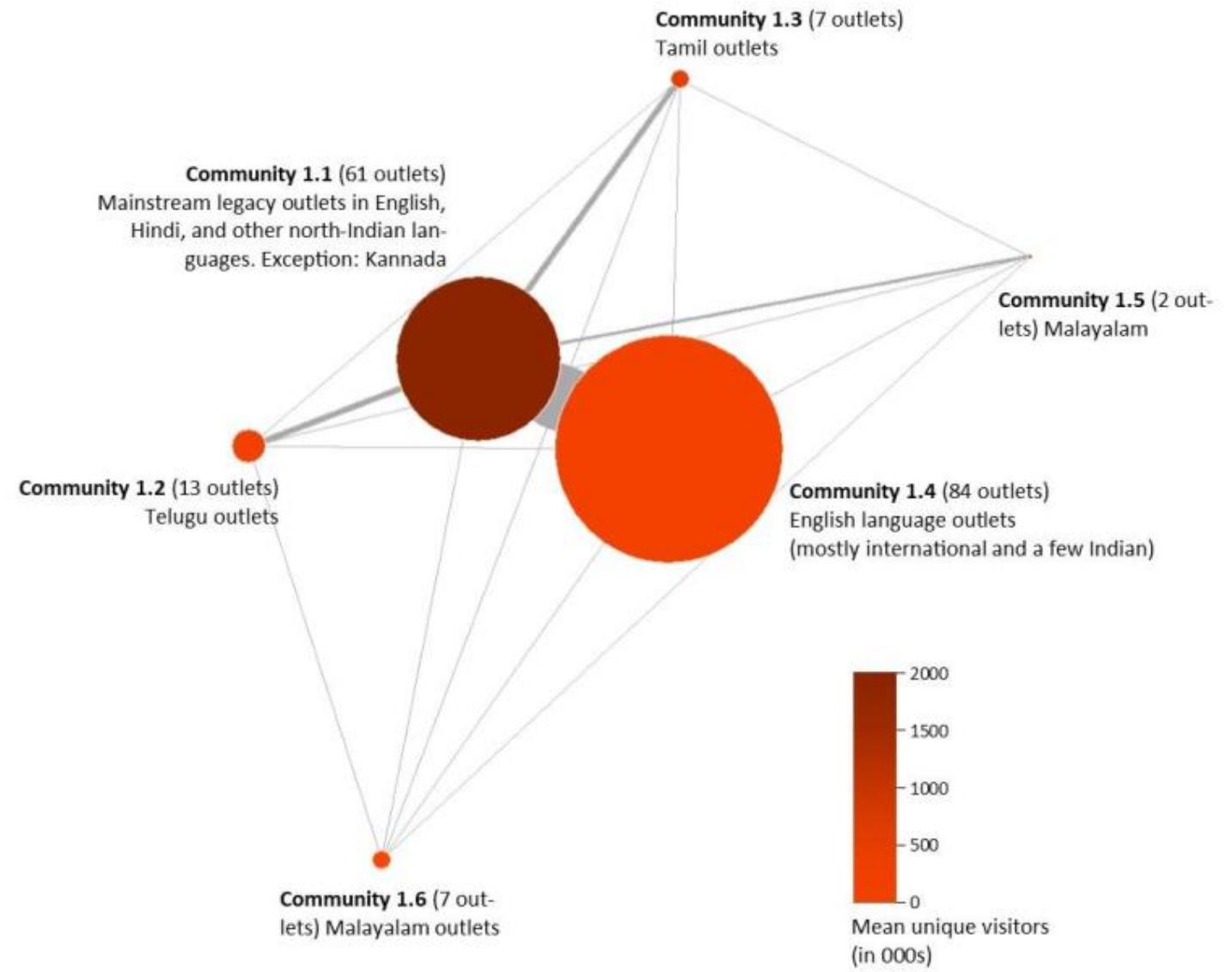
Methods: Network Analysis

- Identifying news reading publics using network analysis
 - Community detection (Pons & Latapy, 2006) with a methodological improvement
 - Evaluation of the “goodness” of community structure
- Assessing audience fragmentation using Network Thresholding with Community Extraction

Findings

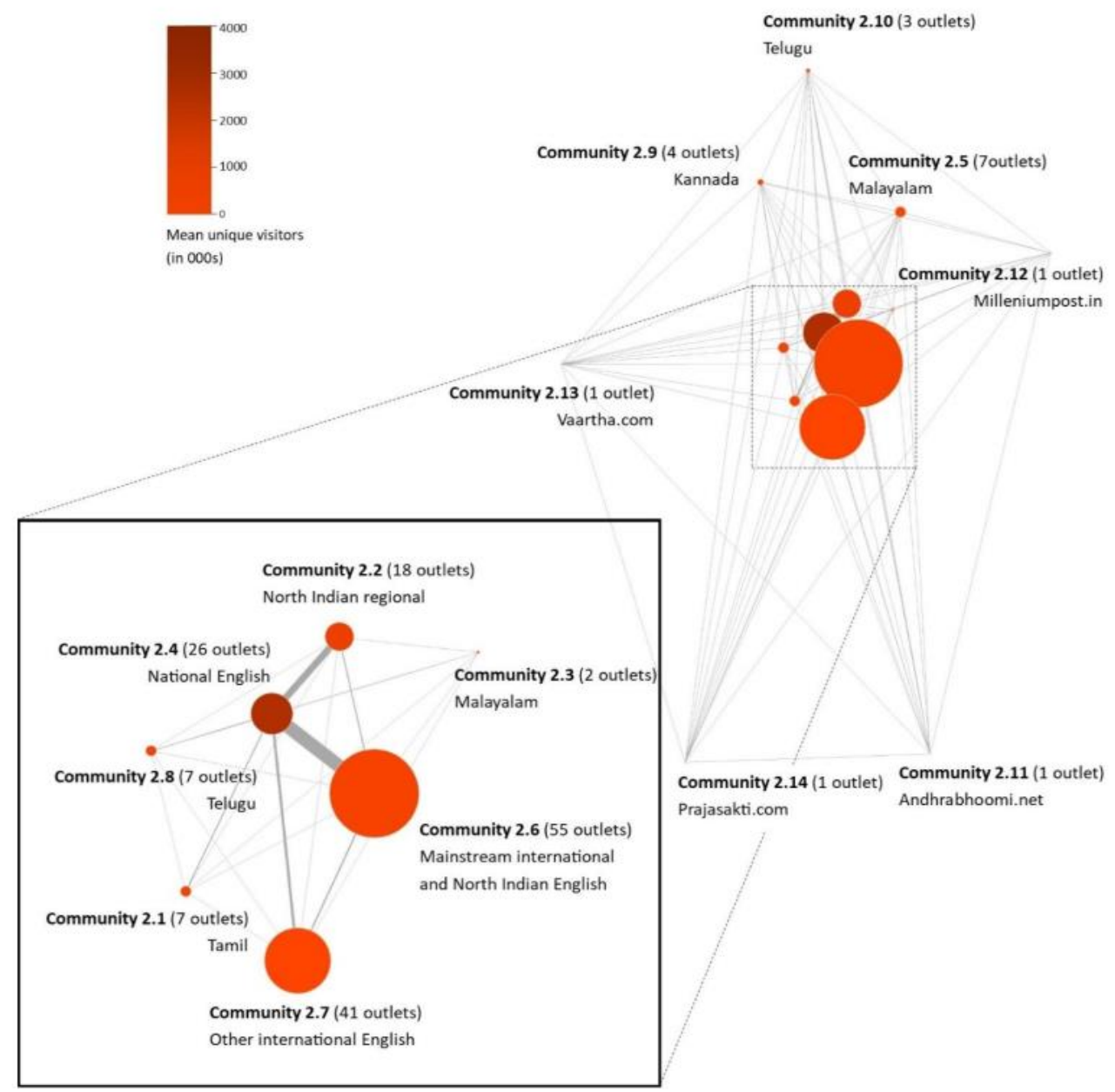
New Reading Publics: Linguistic Segregation

Original Algorithm

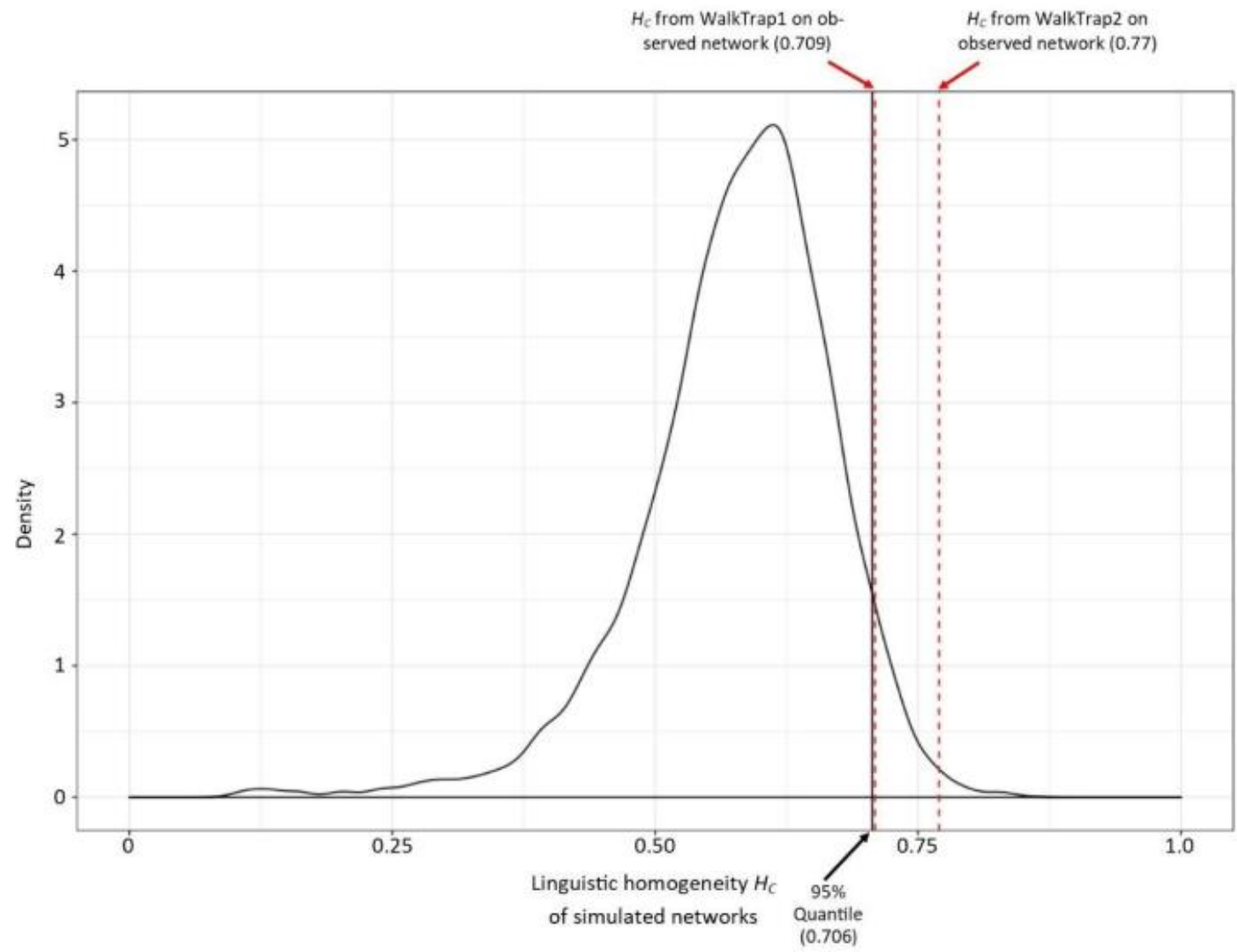


New Reading Publics: Linguistic Segregation

Refined Algorithm



Linguistic Segregation: Statistical Validation

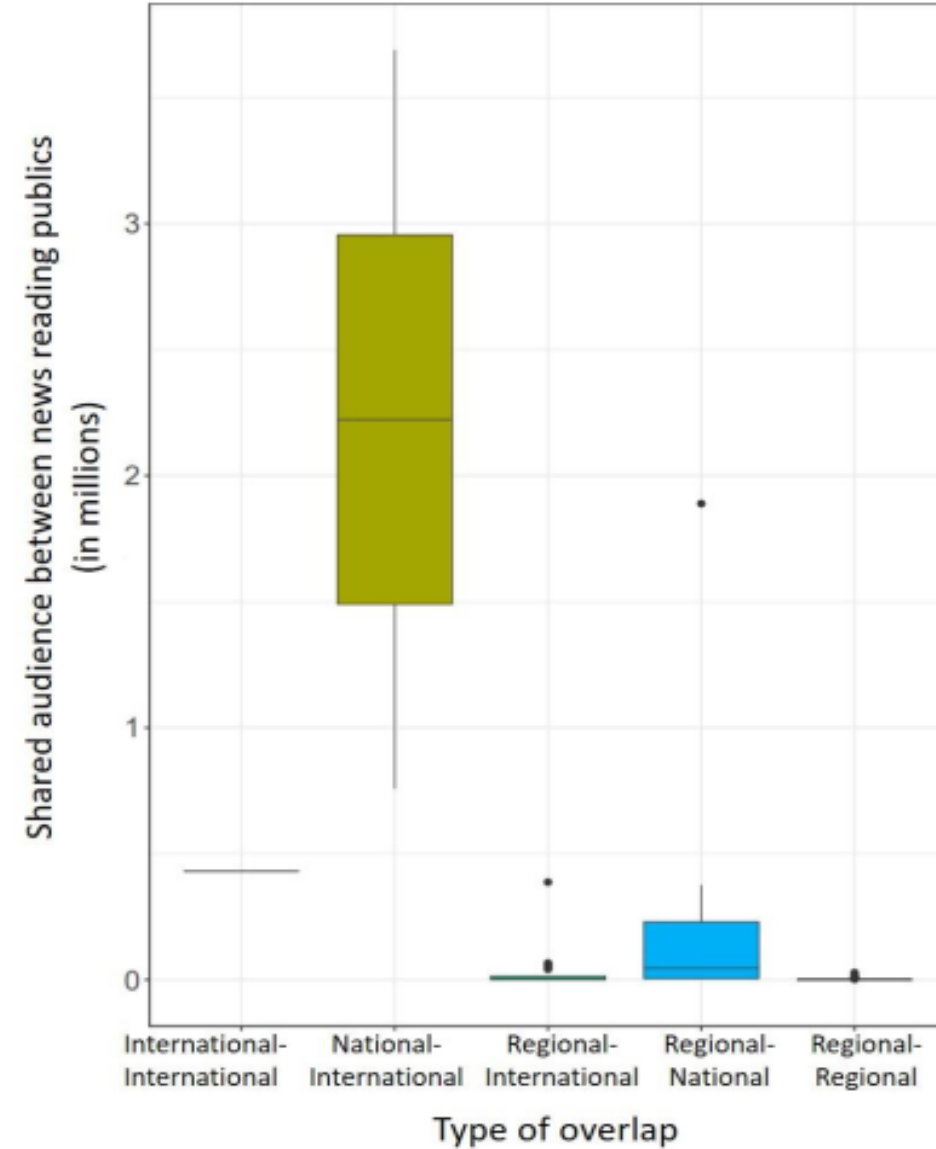


H1: The media consumption landscape in India is segregated along linguistic lines



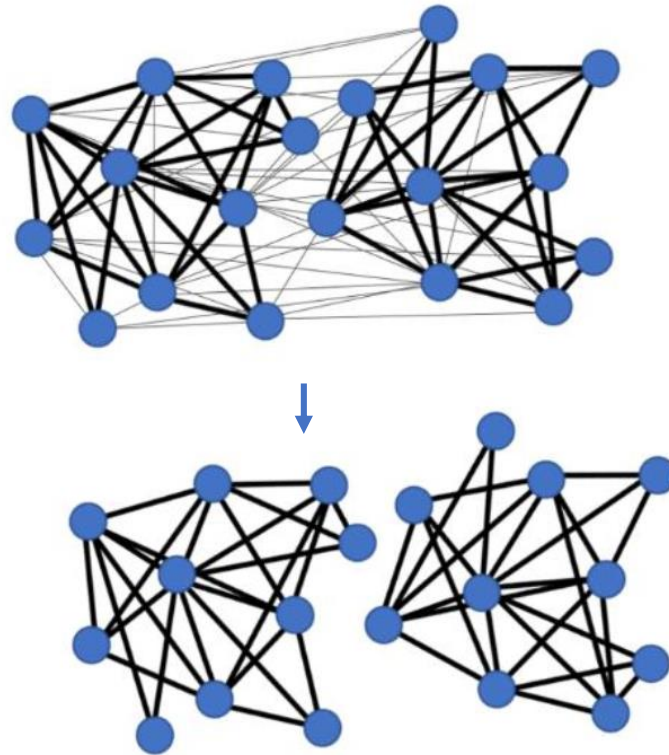
New Reading Publics: Linguistic Segregation

H2: Vernacular news reading publics will have smaller overlap with each other than with national news reading publics

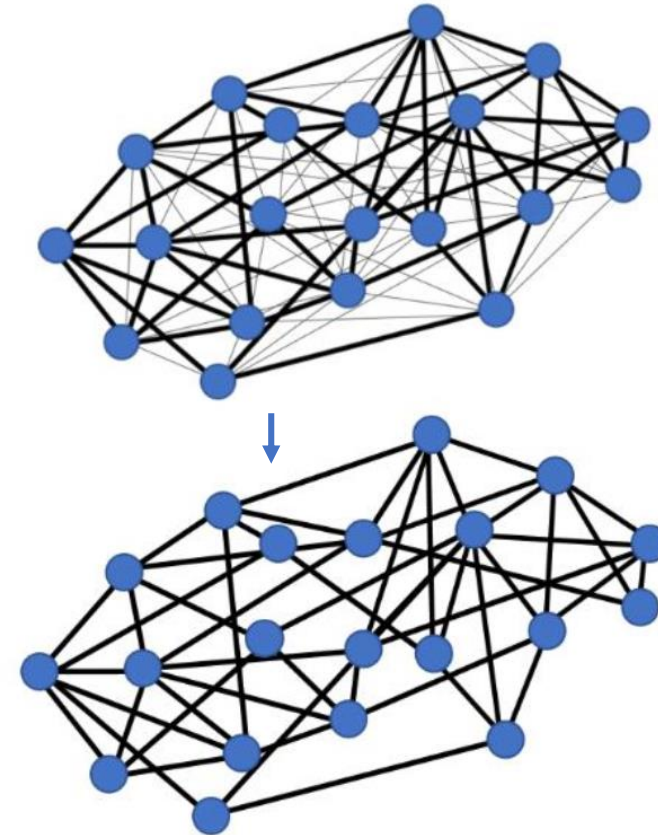


Fragmentation: Theoretical Expectation

A

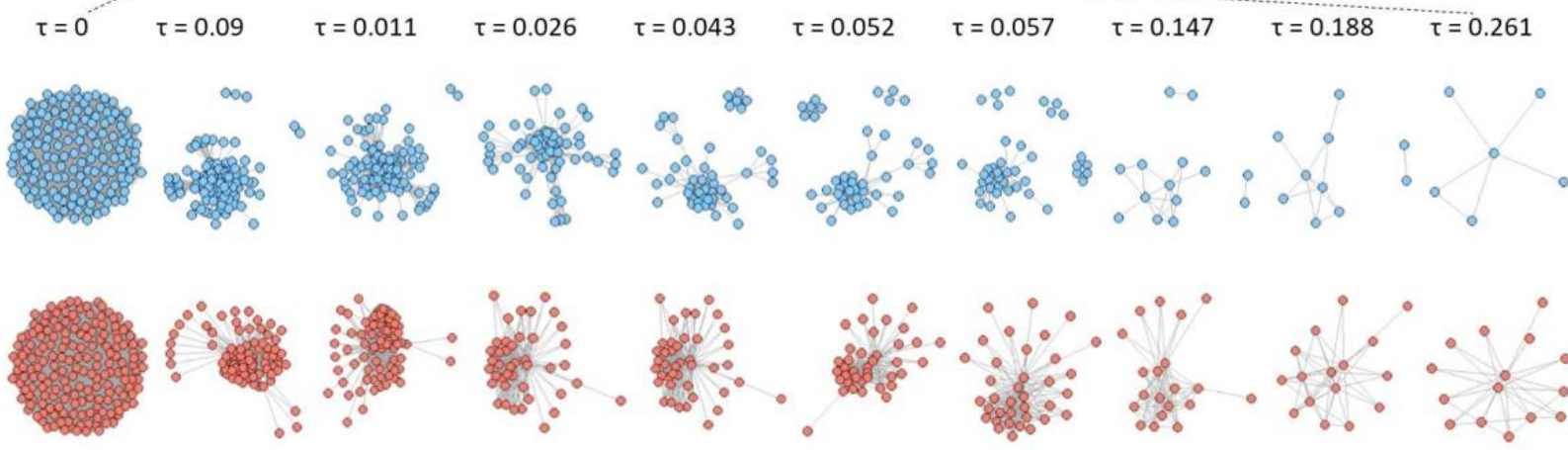
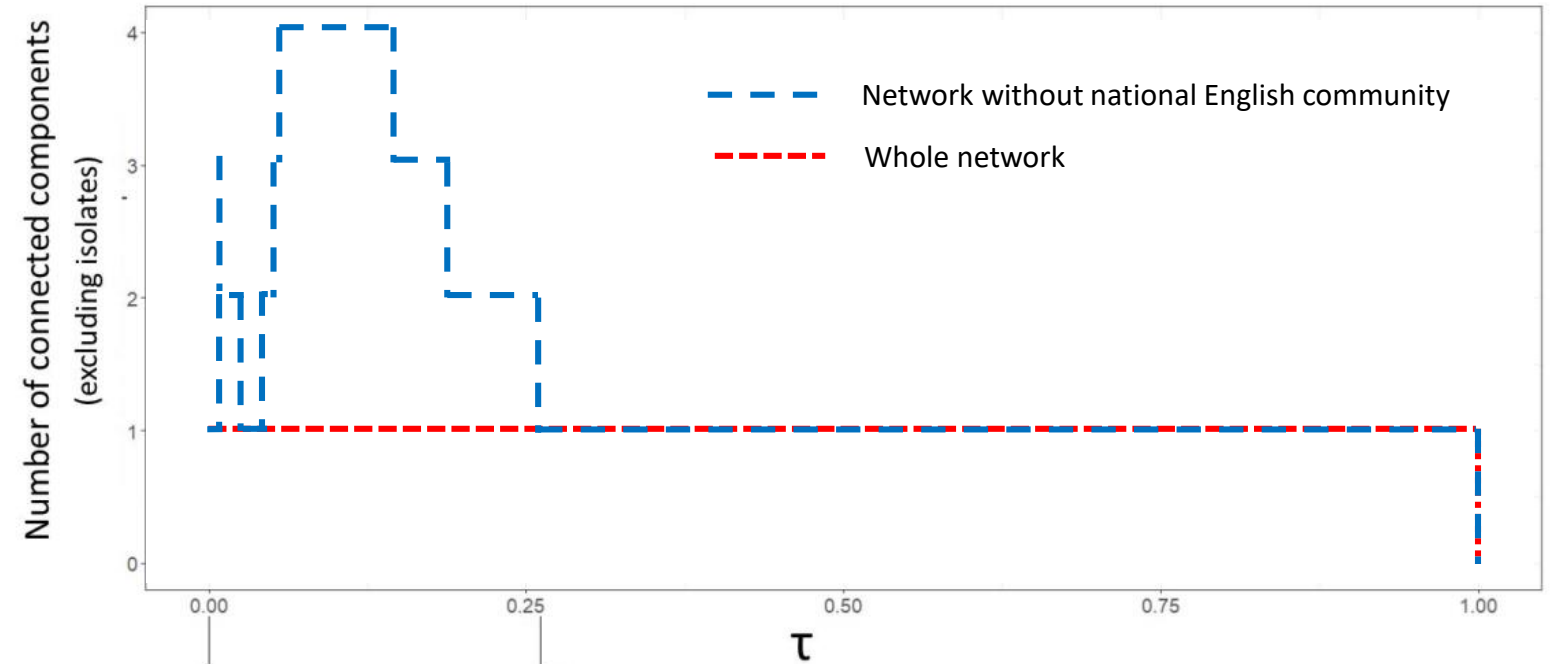


B

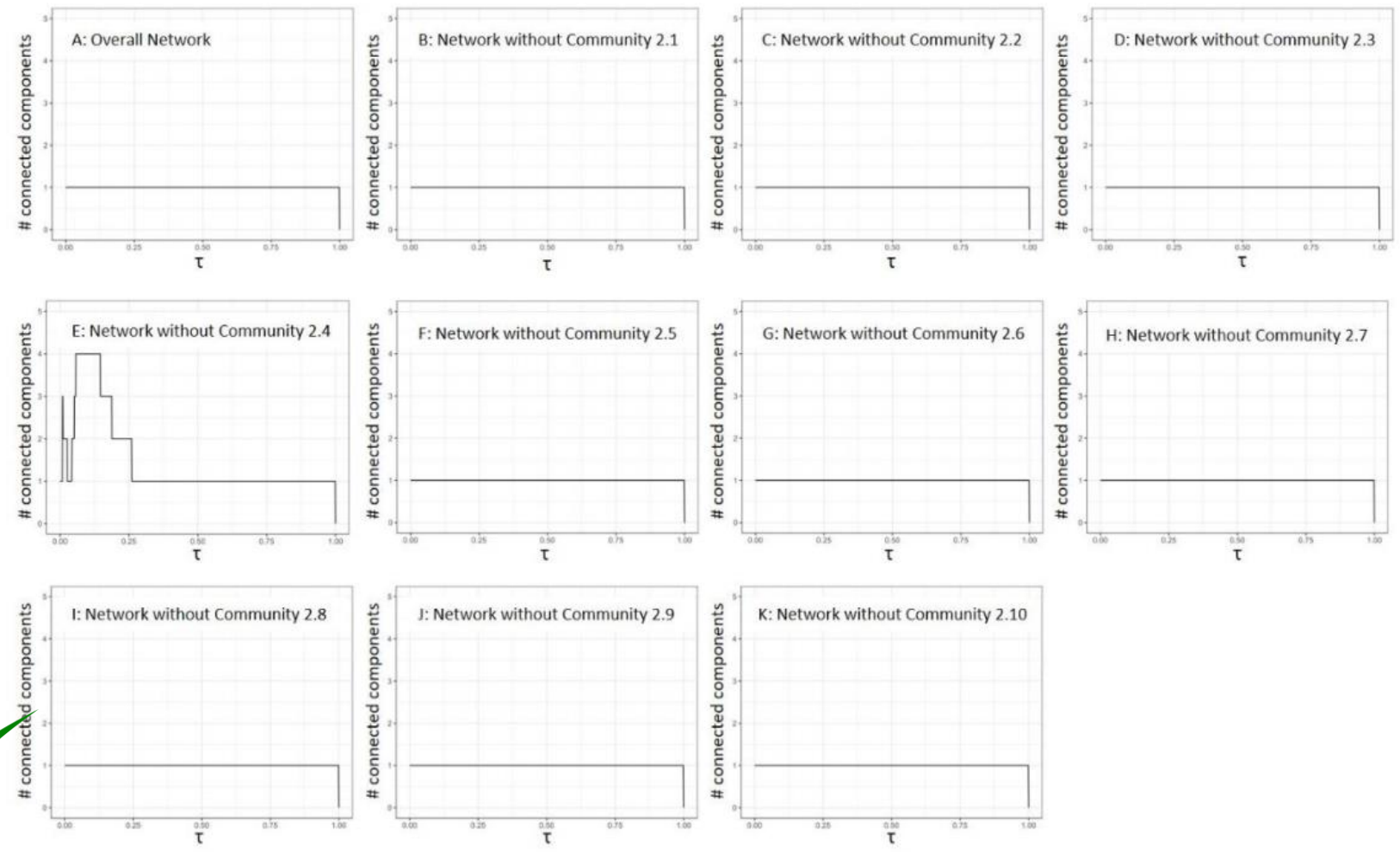


Intuition behind Thresholding and Network Fragmentation

The Unifying Role of National English Media



The Unifying Role of National English Media



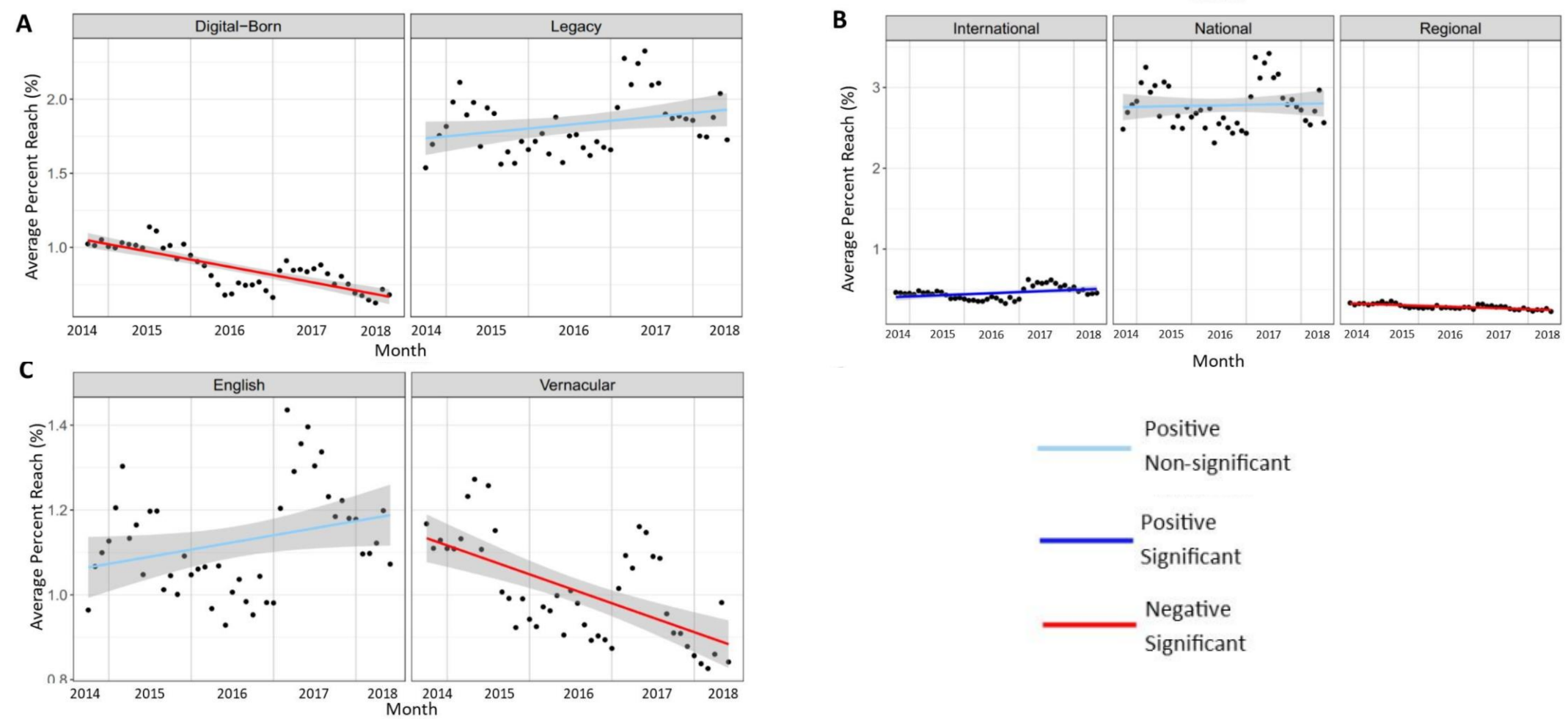
H3: The presence of national English news reading publics reduces fragmentation in the online Indian space

Audience Mobility

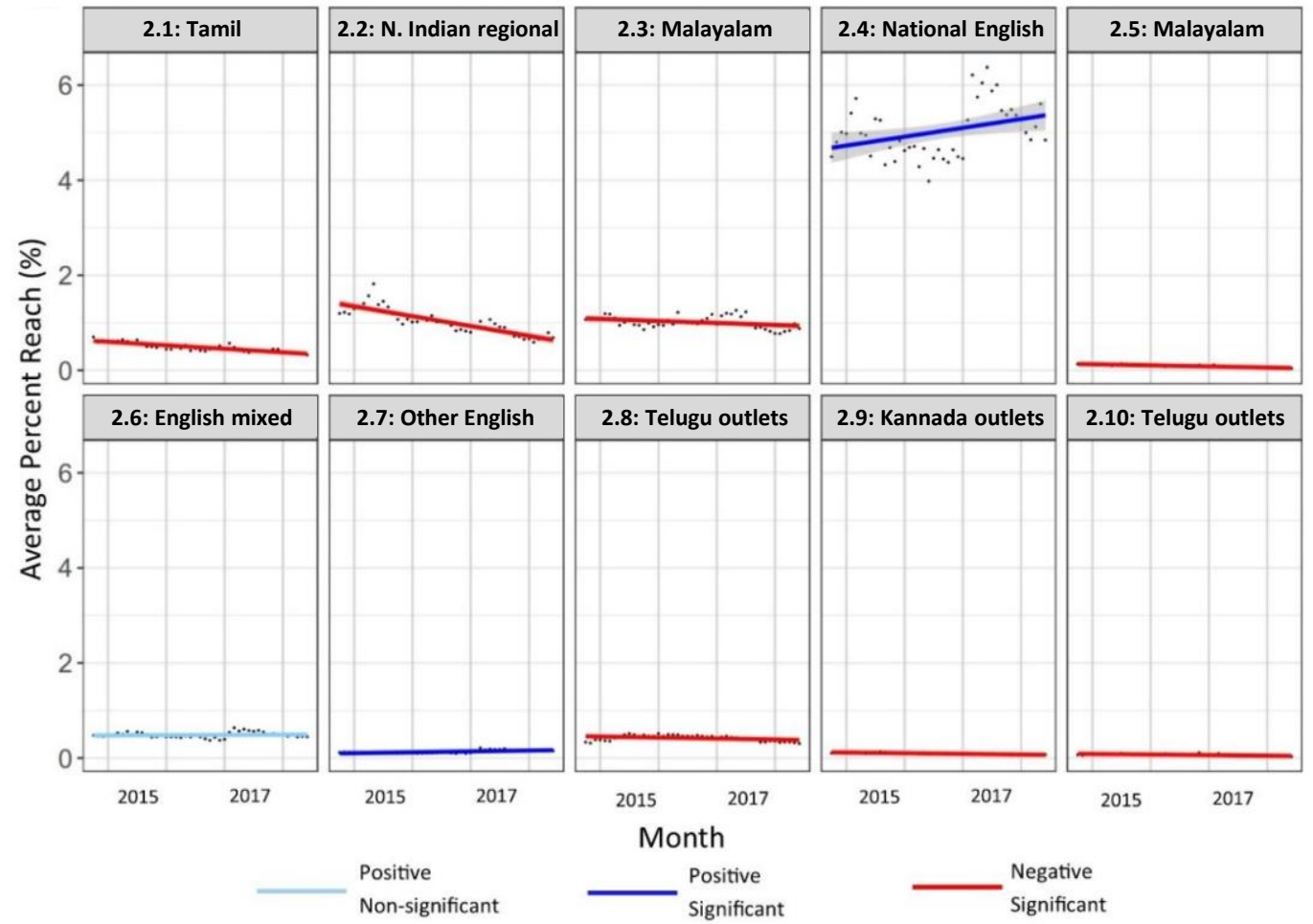
The migration of audience(s) from some media types/formats to others

- e.g. audiences “moving” to print media as they become literate
- audiences “moving” to cable TV as it becomes available

Trends: Sharp Decline for Regional, Vernacular, Digital-born Media



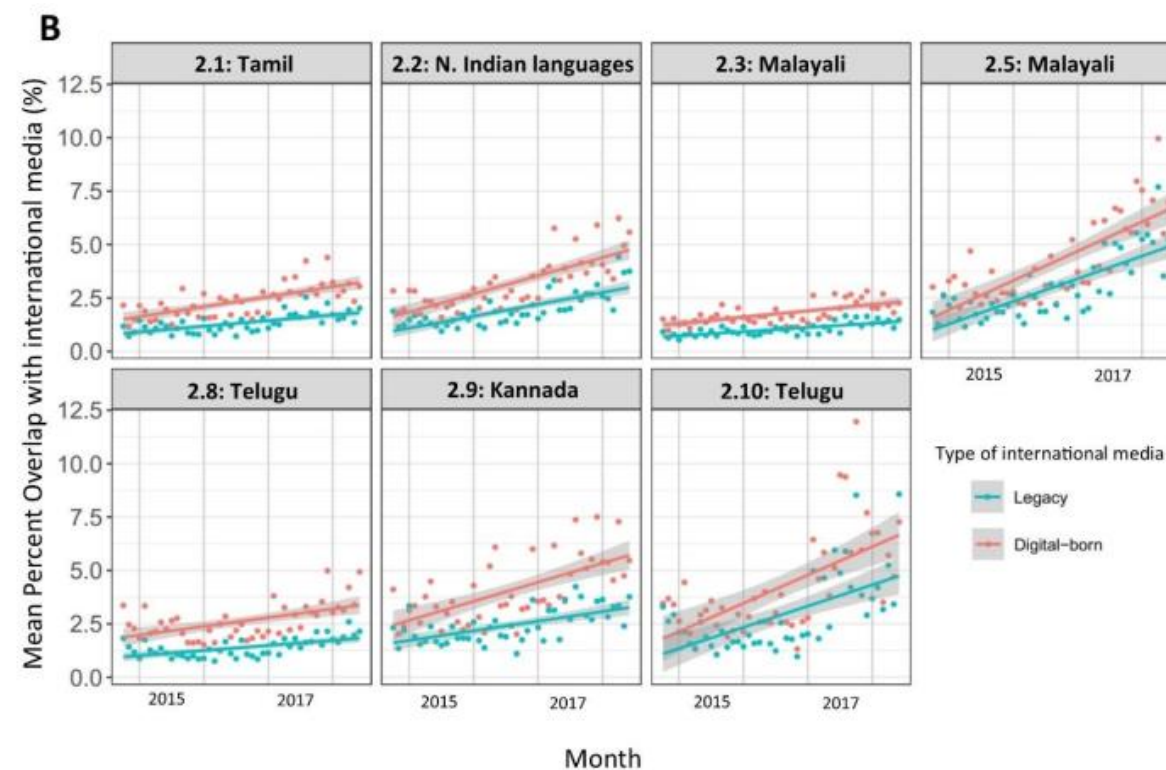
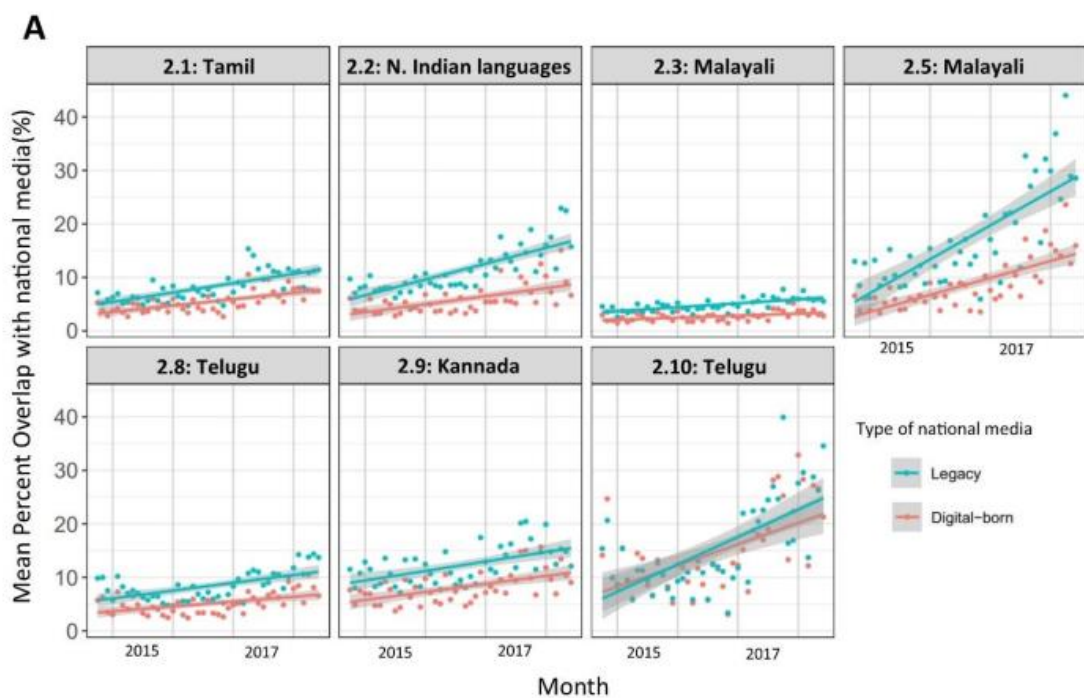
Trends: Sharp Decline for Regional, Vernacular Media



Trends: Migration From Vernacular to National and International

Vernacular audiences moving to **national** media prefer legacy brands to digital-born brands

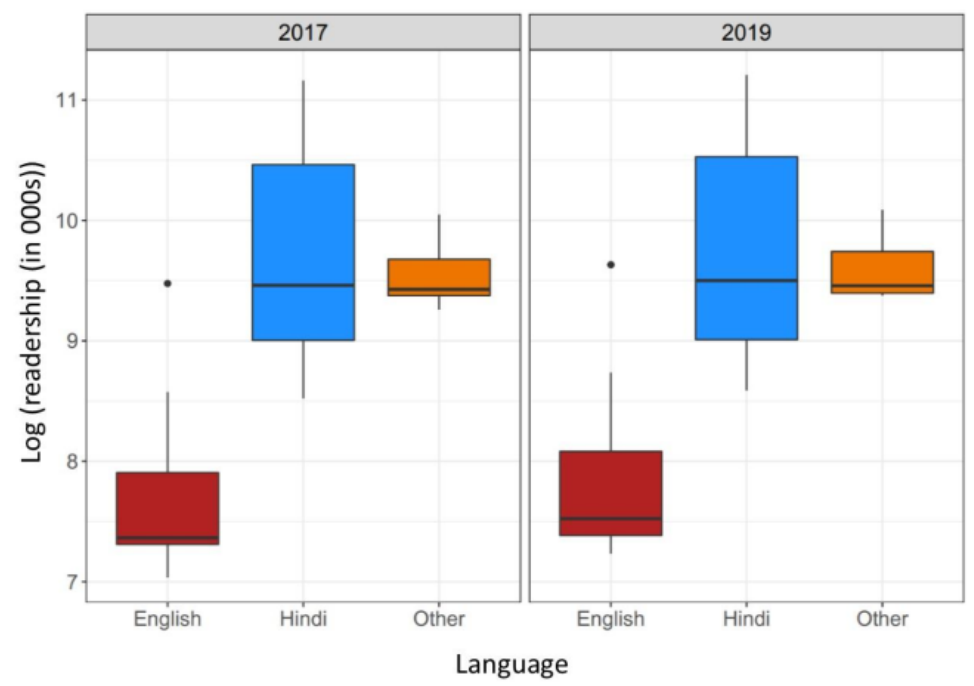
Vernacular audiences moving to **international** media show no such preference



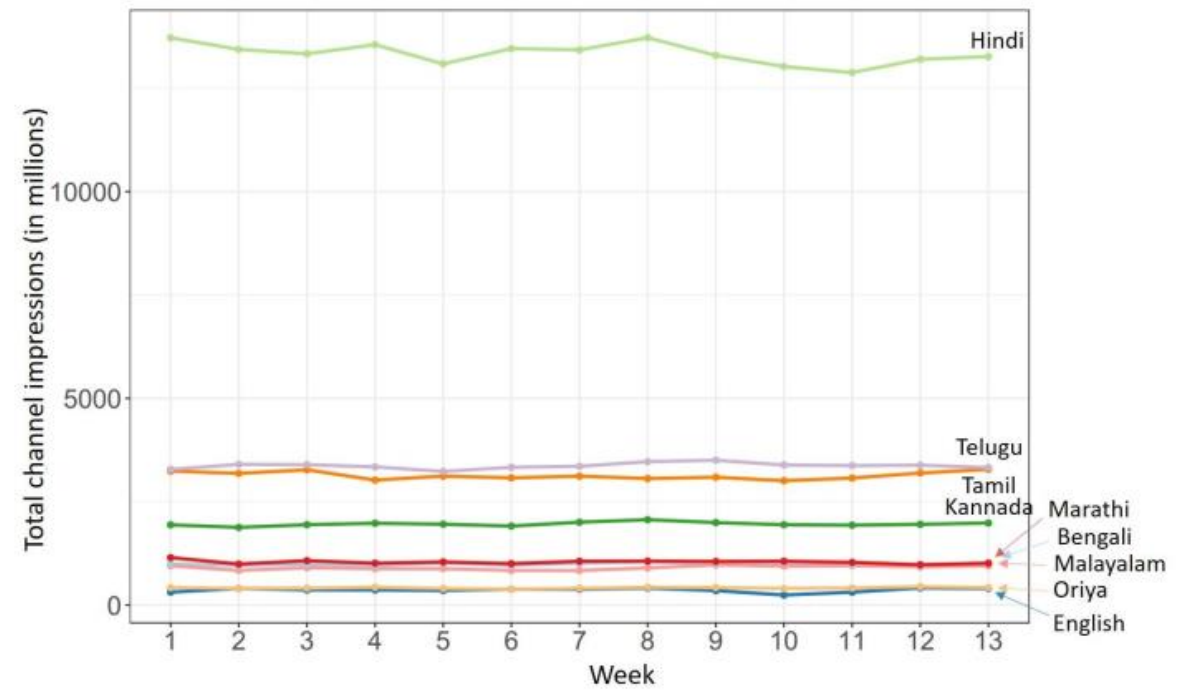
Summary: Structure of News Reading Publics

- The Indian online landscape is segregated along linguistic lines (H1)
- Vernacular-National duality of news reading behavior (H2)
- National, English news reading publics prevent audience fragmentation online (H3)

Summary: Online vis-à-vis offline



Readership numbers of the top 10 most popular newspapers by language (source: Indian Readership Survey)



TV channel impressions over 13 consecutive weeks in 2018 (source: Broadcast Audience Research Council)

Summary: Longitudinal Trends

- Online audiences **increasingly** consuming international, and legacy national media
- Online audiences **decreasingly** consuming vernacular, regional, digital-born media
- Vernacular news readers increasingly **prefer legacy national media** to digital-born media
- Vernacular news readers have **no significant brand preference** with international media

Discussion

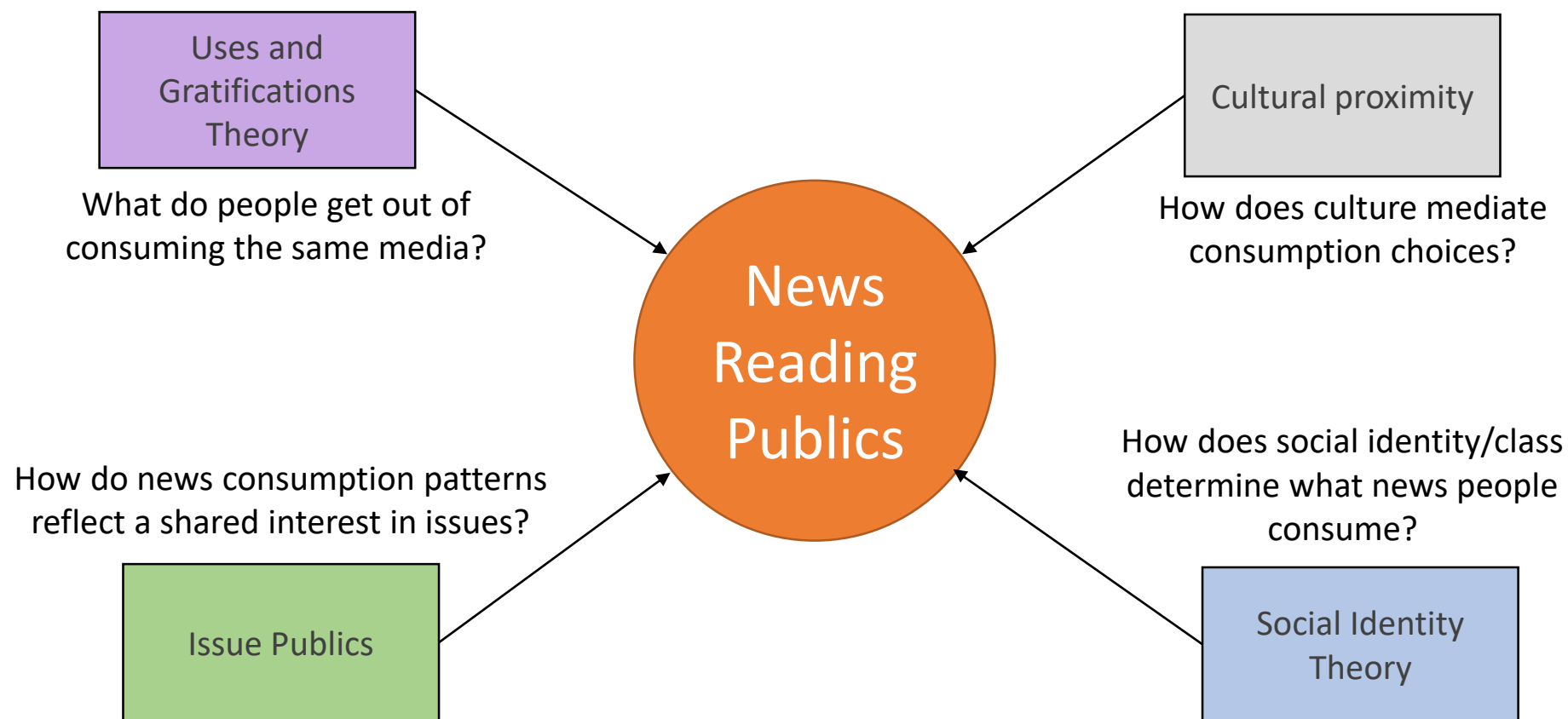
Online vis-à-vis Offline

- Online news consumption is more centralized, less fragmented
 - Potentially owing to the demographic differences in online versus offline
- Likely to increase more – literacy, internet penetration, English education
- Implications for regional media industries?

Future of Regional Media?

- Dire future of regional media
 - Still profitable in print and TV, but not for long
 - Need to invest in digital
- Decline in local news around the world

Revisiting the Theoretical Framework



Limitations

- ComScore's data collection / integration methods are proprietary
 - ComScore's estimates are likely the best available for India
 - ComScore's US estimates correlate highly with Nielsen's
- Desktop data only
 - Including mobile data when available did not change qualitative findings
 - English-vernacular power dynamics were similar

Contributions

- **Main contribution:** Novel evidence of news reading behavior of the second largest online population
- Analytical framework with a context-agnostic methodology
- Instrument for comparative research to understand structural differences in audience organization in different countries
- News Reading Publics as an umbrella theory – echo-chambers, partisan selective exposure, and demographic segmentation are *special cases*

Thank You



Subhayan Mukerjee | Dissertation Defense | May 27, 2020

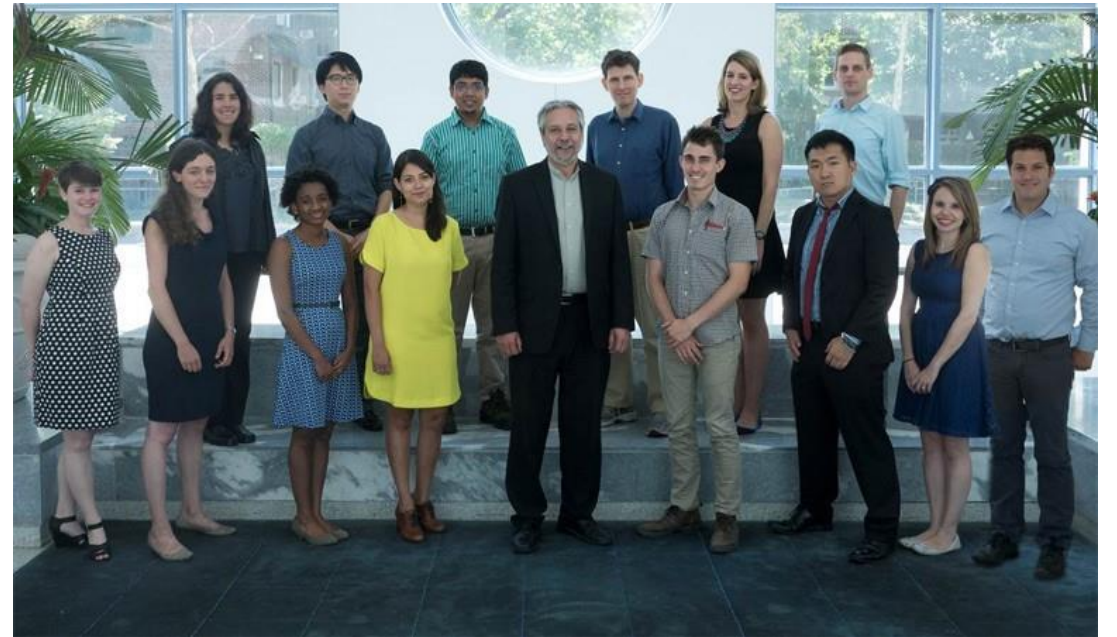


github.com/wrahoool/news-reading-publics

Acknowledgements: Committee



Acknowledgements: Annenberg



Acknowledgements: DiMeNet



Acknowledgements: Family



Thank You

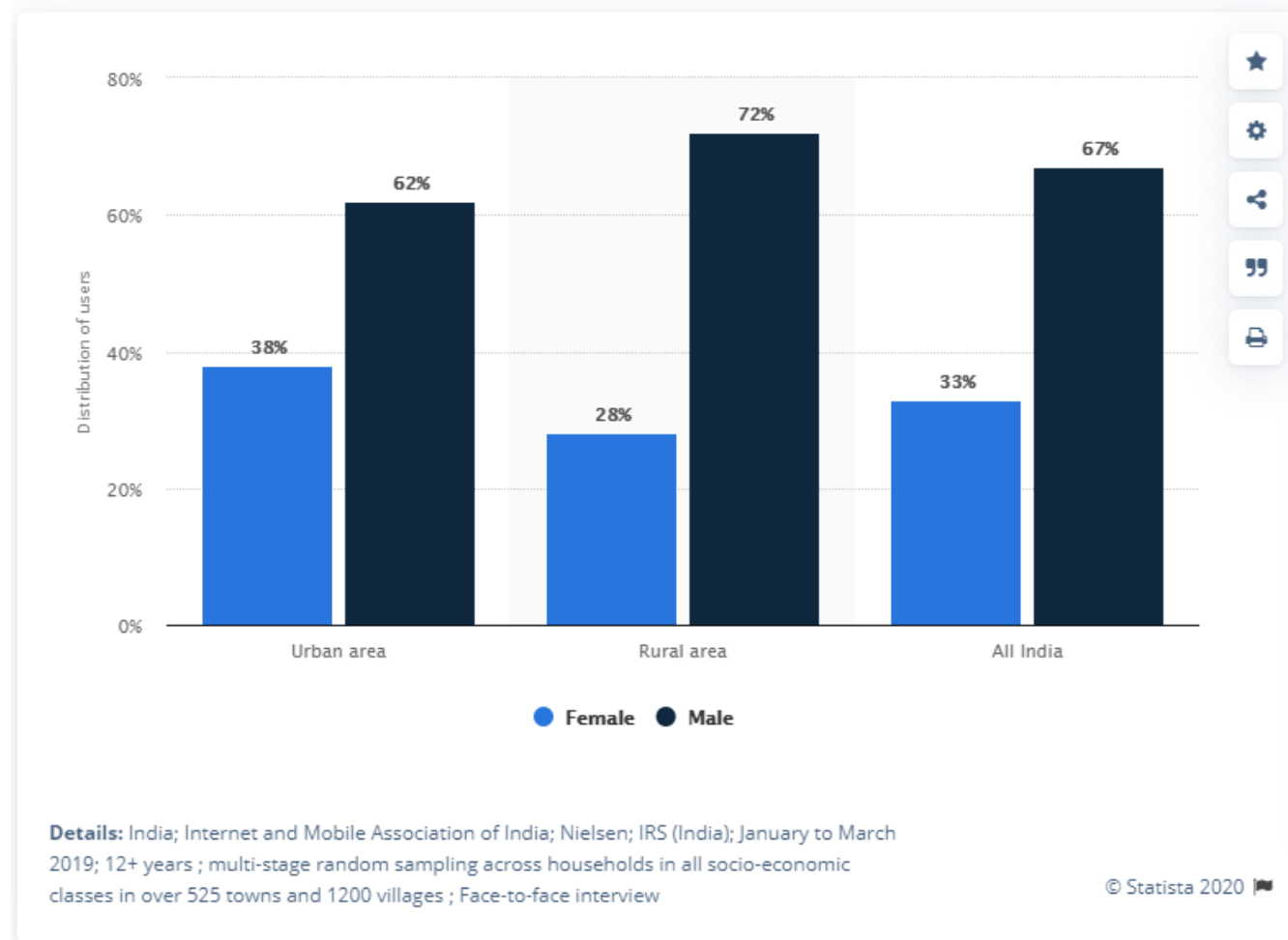


Subhayan Mukerjee | Dissertation Defense | May 27, 2020

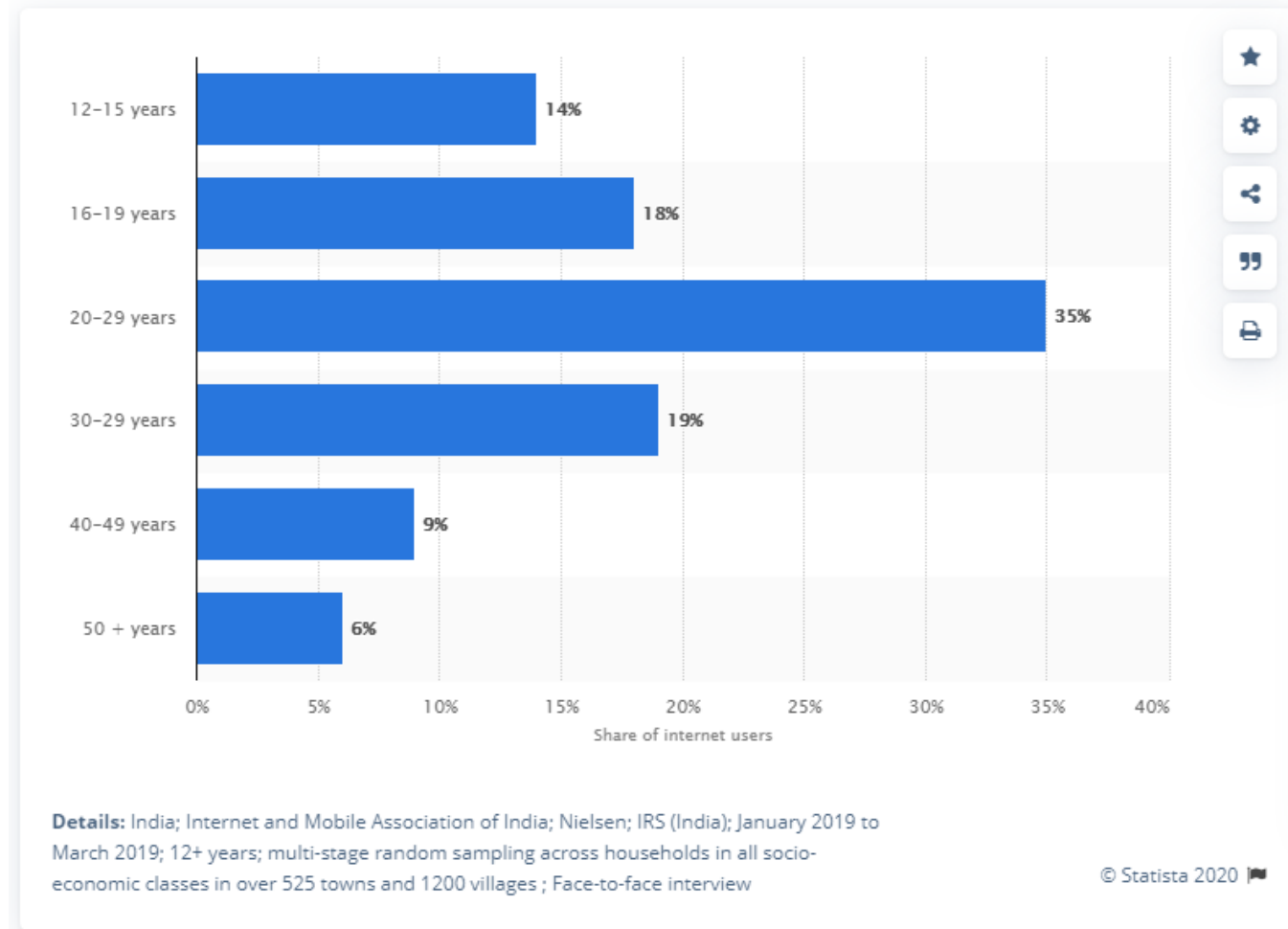


github.com/wrahoool/news-reading-publics

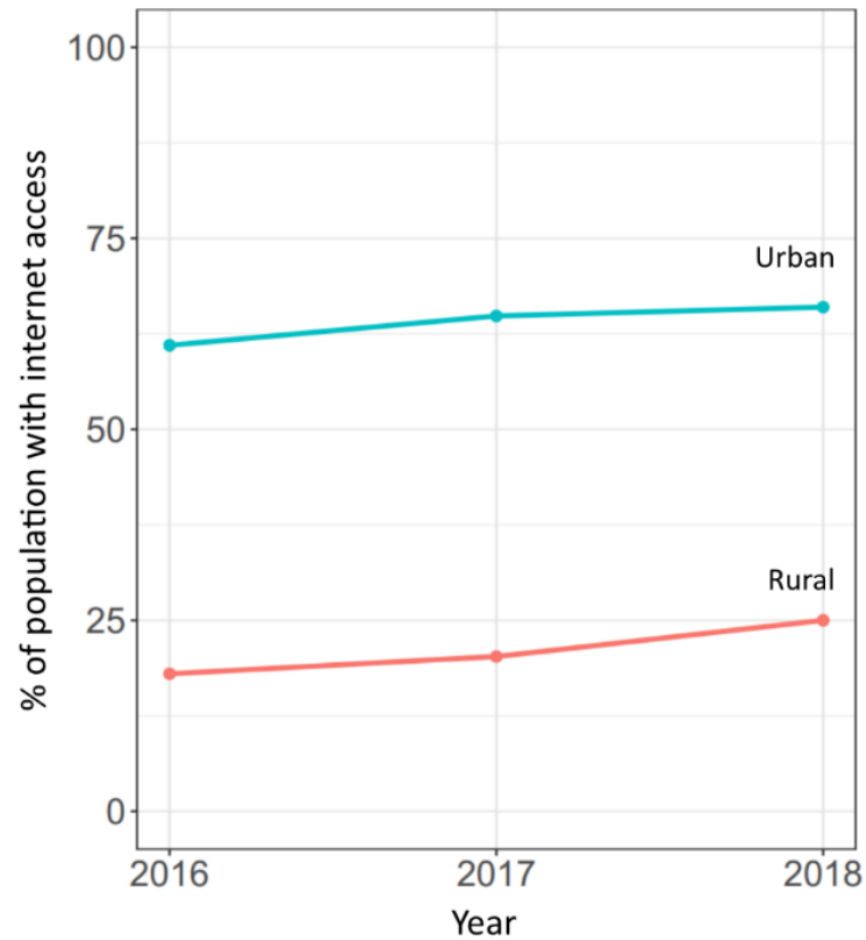
Supplementary: Online Population is More Male



Supplementary: Online Population is (slightly) Younger



Supplementary: Online Population is More Urban



Supplementary: Growth of Mobile in India

- High growth in recent years

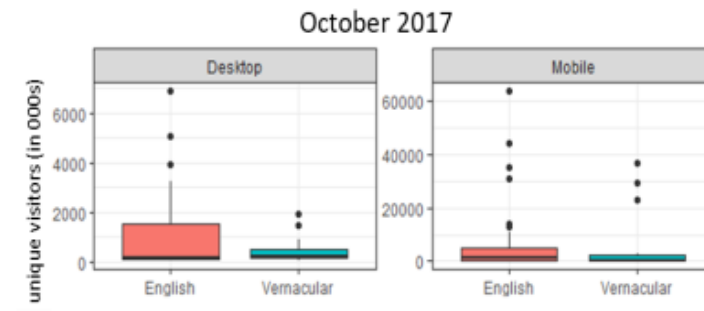
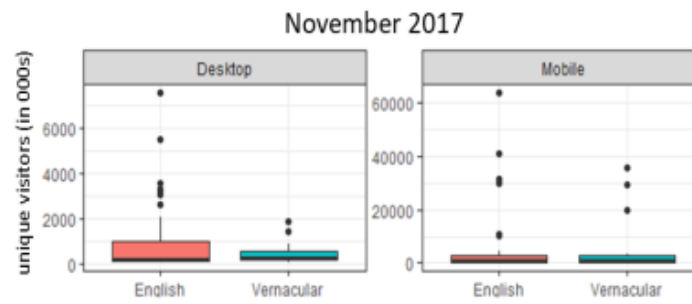
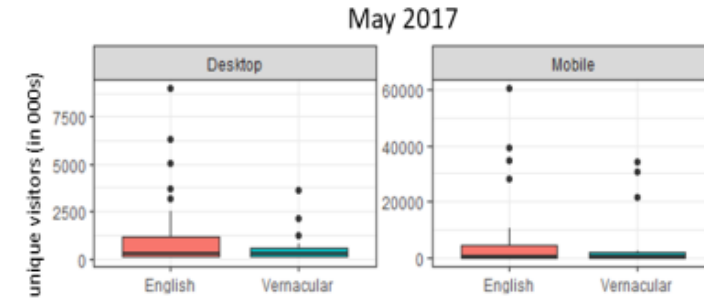
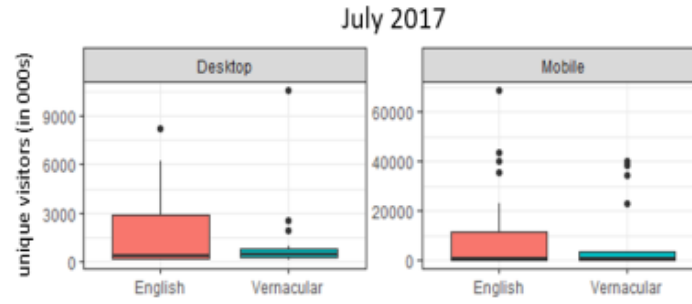
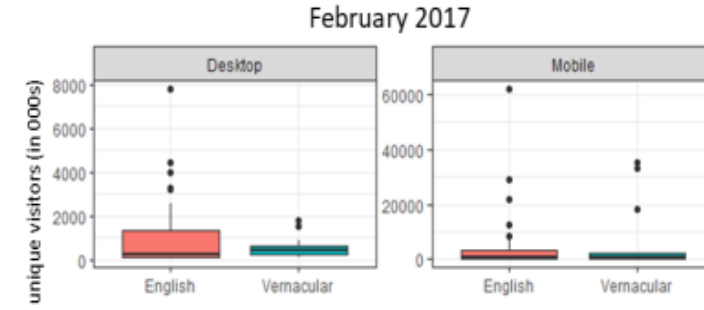
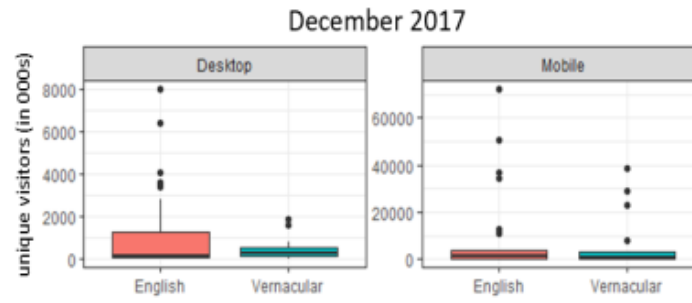
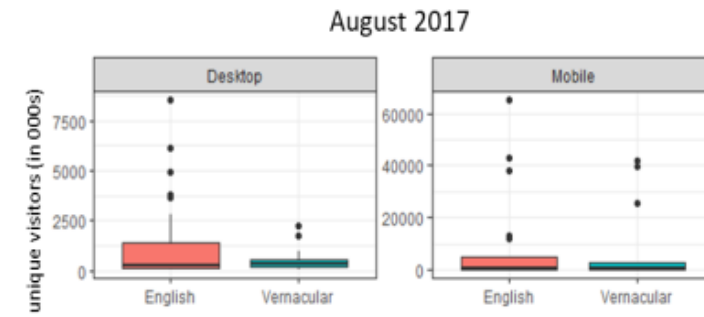
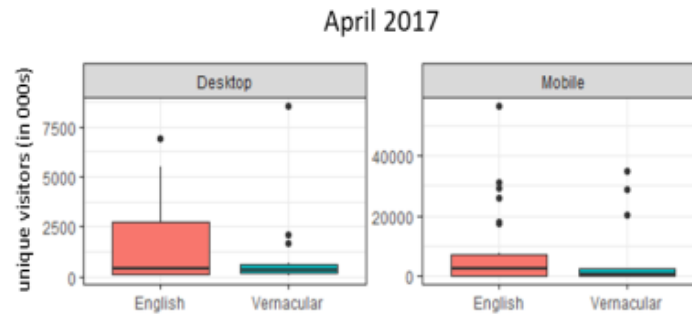
But

- Growth driven by feature phones
- Vast majority of people in rural areas don't use the internet

From a survey administered in rural Karnataka:

“the majority (85%) were unfamiliar with internet communication channels including email and Skype, while only 11% were familiar with Facebook, WhatsApp and YouTube, 4% with gaming, and less than 1% with online shopping” (Vaijayanti, 2018)

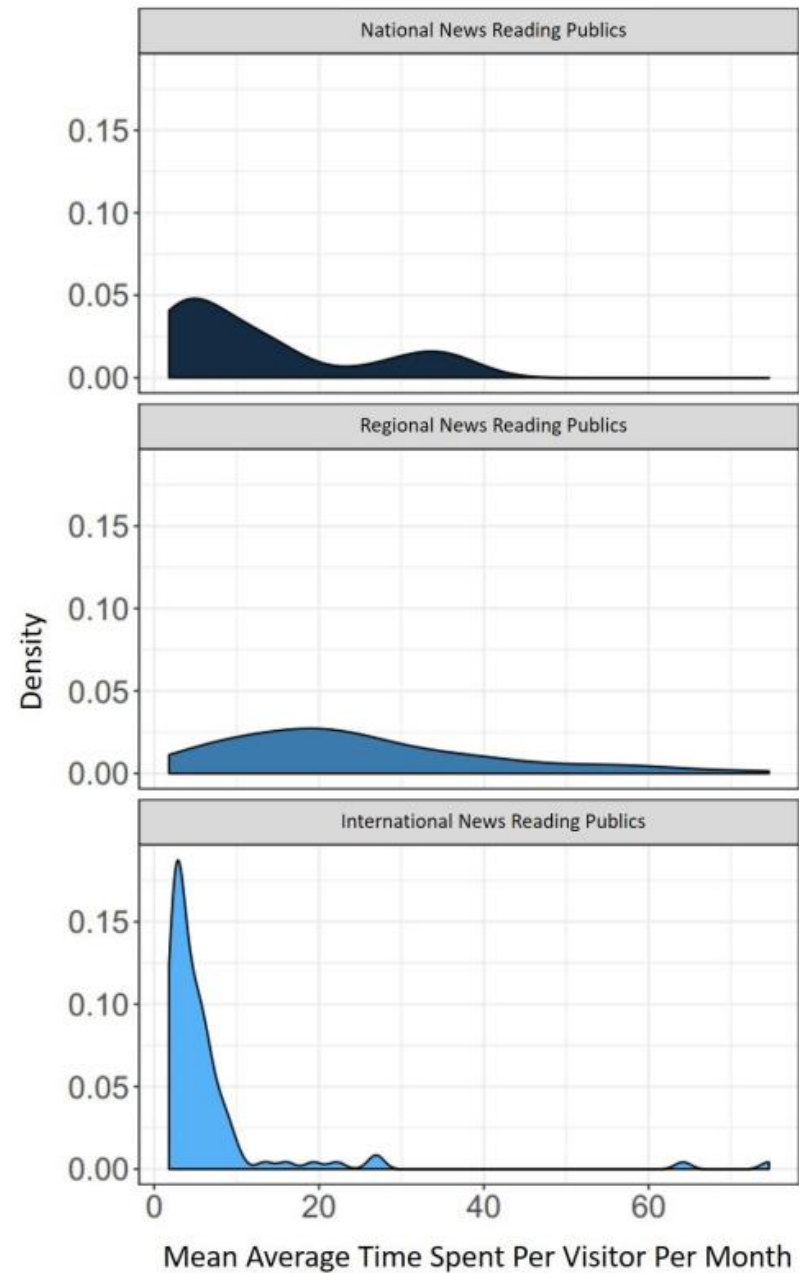
Supplementary: Multi-platform



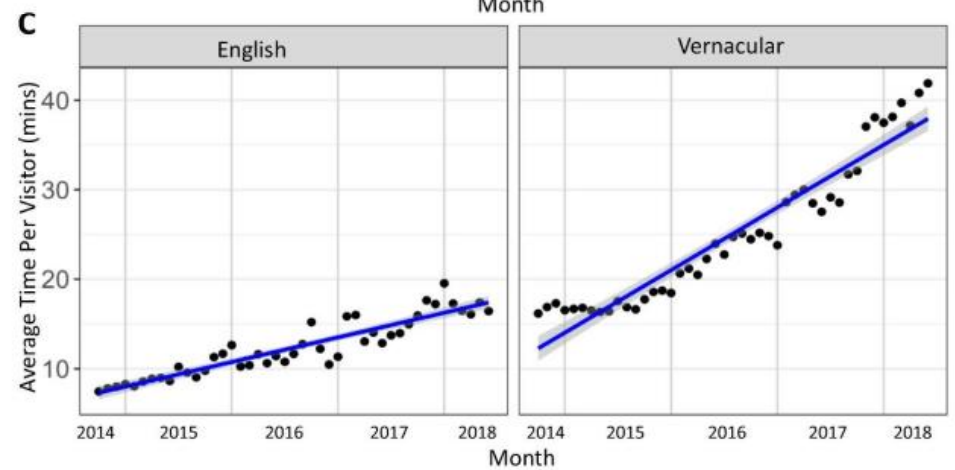
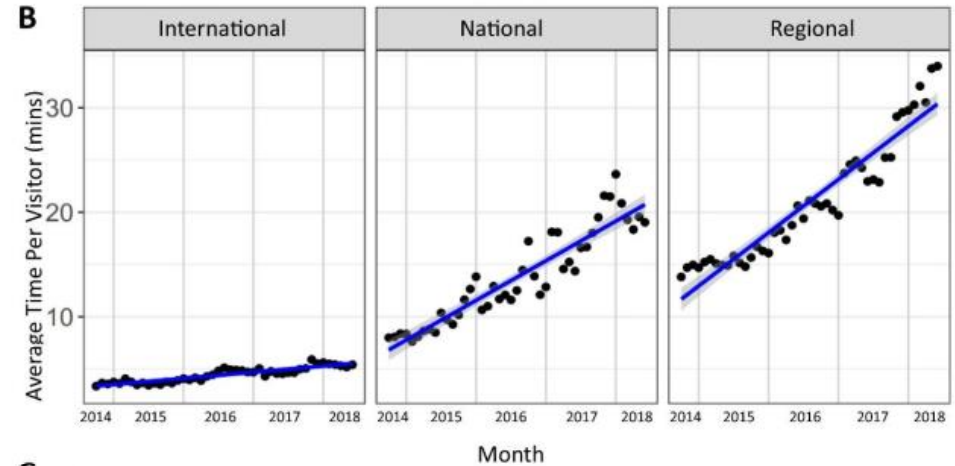
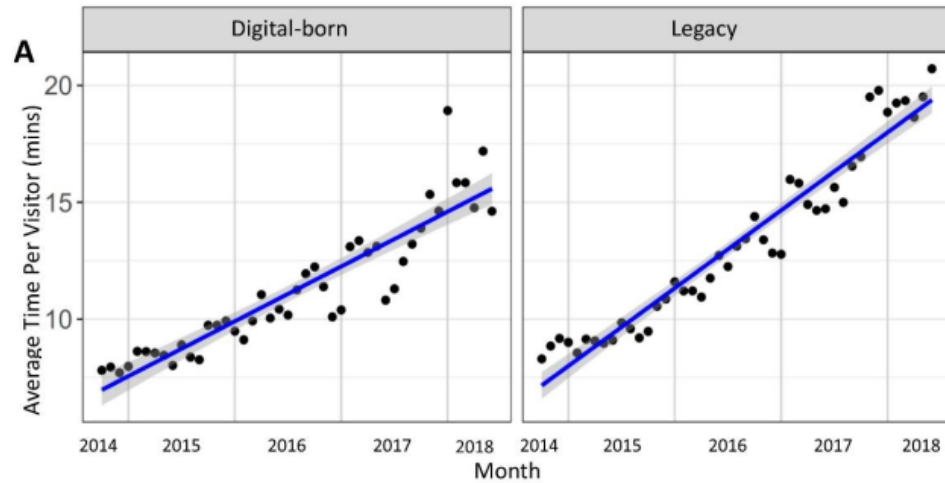
Supplementary: Audience Mobility

- $Percent\ Overlap(PO) = \frac{Shared\ Audience\ Between\ A\ and\ B}{Audience\ Reach\ of\ A} \times 100$
- Trend of (Mean PO / month) for all pairs (A, B) where A is a regional outlet and B is a national Outlet

Supplementary: Audience Engagement



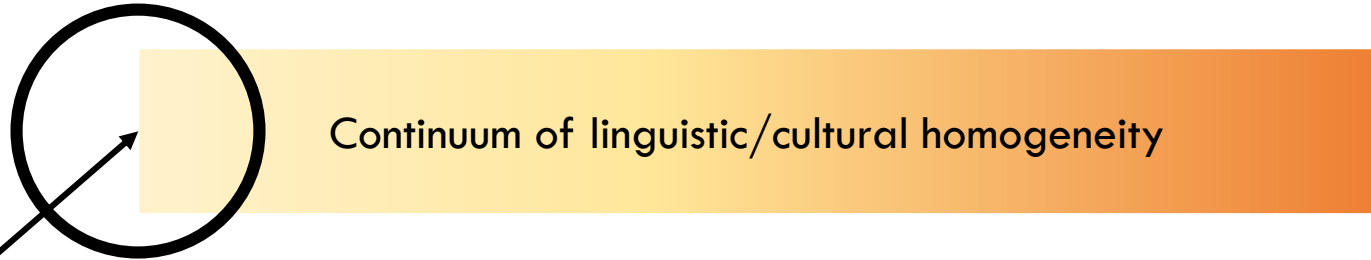
Supplementary: Audience Engagement



Supplementary: Generalizability

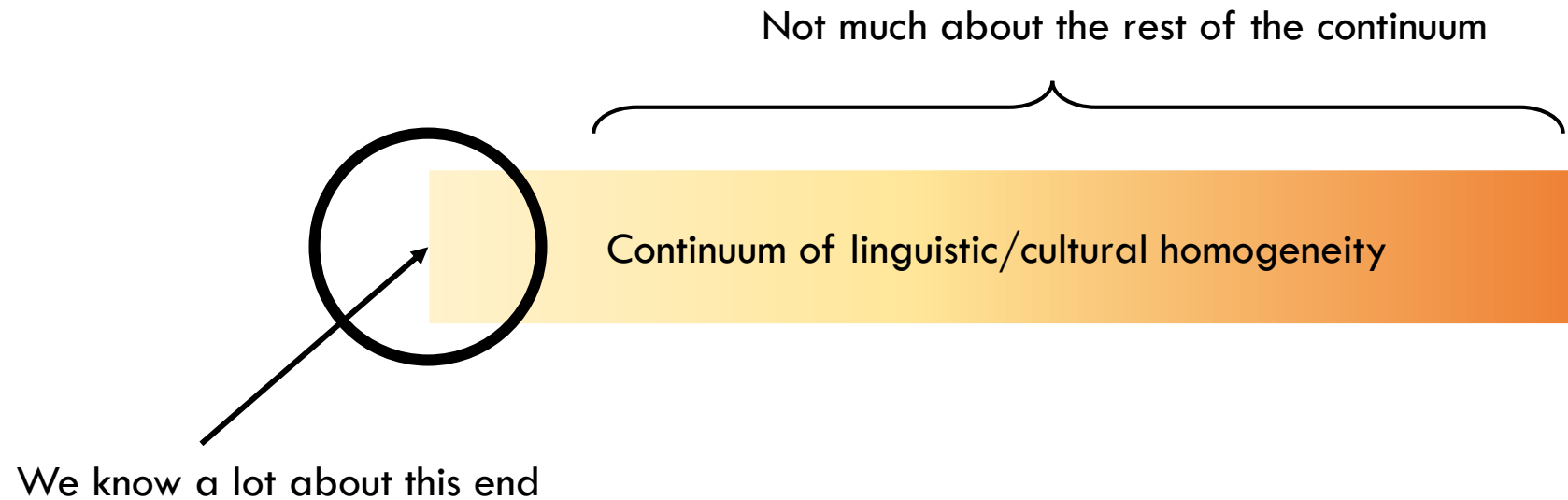
Continuum of linguistic/cultural homogeneity

Supplementary: Generalizability

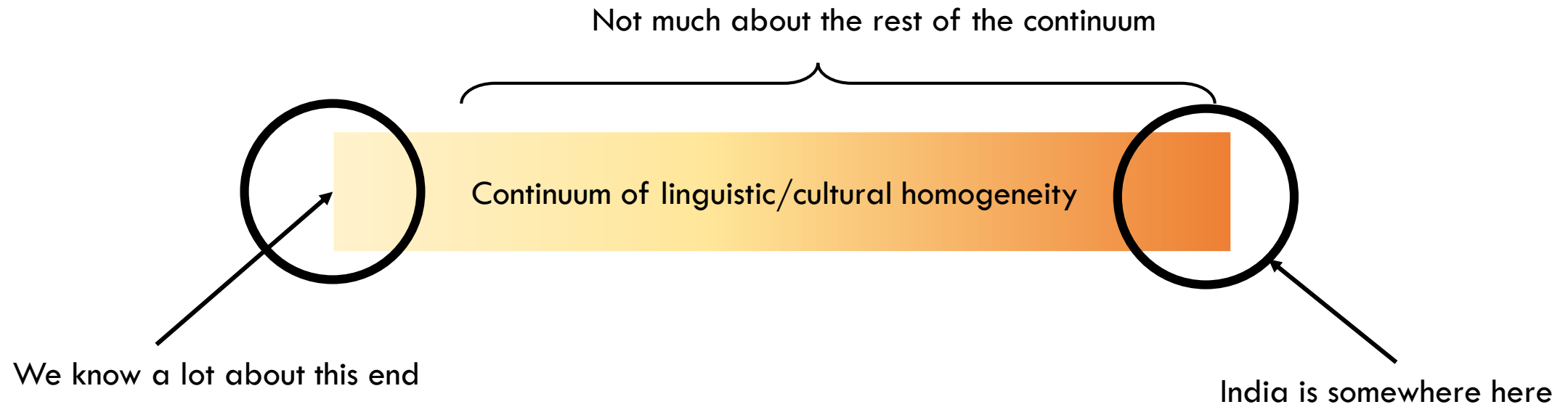


We know a lot about this end

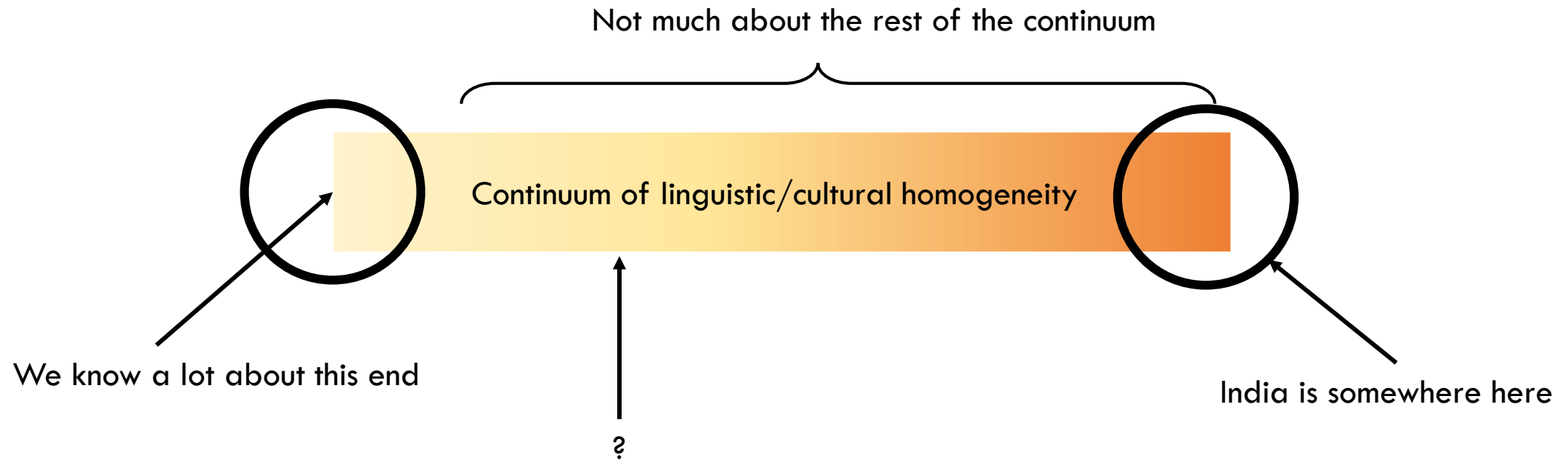
Supplementary: Generalizability



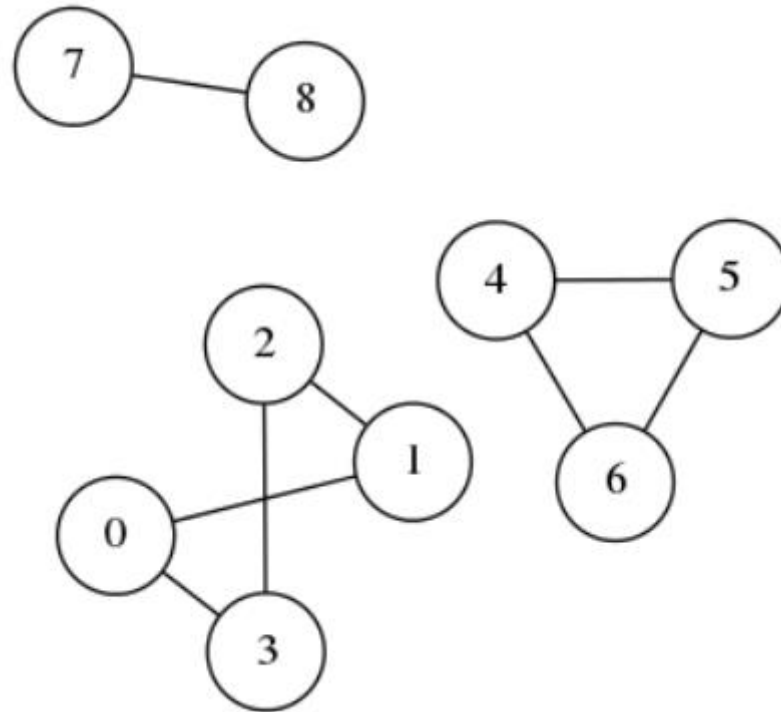
Supplementary: Generalizability



Supplementary: Generalizability



Supplementary: Connected Component



Supplementary: WalkTrap

- Imagine a person walking along the network edges
- At every step, she decides to randomly walk to an adjacent node
- Let her walk for a very long period of time
- The set of nodes within which she gets trapped and spends a lot of time are the “communities” as they have lots of edges between them

Supplementary: WalkTrap Enhancement

- How do you parameterize the WalkTrap algorithm? (Arenas et al. 2008)
 - Add a self loop to each node and increase/decrease weight to control mobility of the walker
 - For audience networks, this weight is the audience of that node

Supplementary: WalkTrap Enhancement

- How do you parameterize the WalkTrap algorithm? (Arenas et al. 2008)
 - Add a self loop to each node and increase/decrease weight to control mobility of the walker
 - For audience networks, this weight is the audience of that node

Media Imperialism

- Media imperialism is a theory based upon the fact that an over-concentration of mass media from larger nations is a significant variable in negatively affecting smaller nations, in which the national identity of smaller nations is lessened or lost due to media homogeneity inherent in mass media from the larger countries
- A vision of Western cultural dominance and imposition, created by a ceaseless flow of cultural products that invaded and overwhelmed the developing world (Chadha & Kavoori, 2000)

Supplementary: Dyadic Thresholding

- Phi-correlation

Association between two binary variables

	Y=1	Y=0	Total
X=1	n_{11}	n_{10}	n_{1*}
X=0	n_{01}	n_{00}	n_{0*}
Total	n_{*1}	n_{*0}	n

$$\Phi_{XY} = \frac{n_{11} n_{00} - n_{10} n_{01}}{\sqrt{n_{1*} n_{*1} n_{0*} n_{*0}}}$$

Supplementary: Dyadic Thresholding

- Phi-correlation

Association between two binary variables

	Y=1	Y=0	Total
X=1	n_{11}	n_{10}	n_{1*}
X=0	n_{01}	n_{00}	n_{0*}
Total	n_{*1}	n_{*0}	n

$$\Phi_{XY} = \frac{n_{11} n_{00} - n_{10} n_{01}}{\sqrt{n_{1*} n_{*1} n_{0*} n_{*0}}}$$

Supplementary: Dyadic Thresholding

- Phi-correlation

Association between two binary variables

	Y=1	Y=0	Total
X=1	n_{11}	n_{10}	n_{1*}
X=0	n_{01}	n_{00}	n_{0*}
Total	n_{*1}	n_{*0}	n

$$\Phi_{XY} = \frac{nn_{11} - n_{1*}n_{*1}}{\sqrt{n_{1*}n_{*1}(n - n_{1*})(n - n_{*1})}}$$

Supplementary: Dyadic Thresholding

- Phi-correlation

Association between two binary variables

	Visits $j = 1$	Visits $j = 0$	Total
Visits $i = 1$	D_{ij}	-	A_i
Visits $i = 0$	-	-	-
Total	A_j	-	N

$$\Phi_{ij} = \frac{D_{ij}N - A_iA_j}{\sqrt{A_iA_j(N - A_i)(N - A_j)}}$$

Supplementary: Dyadic Thresholding

■ Phi-correlation

Association between two binary variables

	Visits j = 1	Visits j = 0	Total
Visits i = 1	D_{ij}	-	A_i
Visits i = 0	-	-	-
Total	A_j	-	N

$$\Phi_{ij} = \frac{D_{ij}N - A_iA_j}{\sqrt{A_iA_j(N - A_i)(N - A_j)}}$$

$$t = \frac{\Phi_{ij} \sqrt{\max(A_i, A_j) - 2}}{\sqrt{1 - \Phi_{ij}^2}}$$

Thank You



Subhayan Mukerjee | Dissertation Defense | May 27, 2020



github.com/wrahoool/news-reading-publics