REUTERS INSTITUTE
FOR THE STUDY OF JOURNALISM

Introduction to Social Network Analysis

Workshop for LSE CSS Hackathon

Sílvia Majó-Vázquez, PhD

Oxford, 19th April 2018

UNIVERSITY OF OXFORD

Workshop’s Outline

1. Social Network Analysis (SNA)

2. Levels of SNA

3. Hands-on SNA

OXFORD ACADEMIC

Sign In ▾ Register

JOURNAL OF COMPUTER-MEDIATED COMMUNICATION

Issues · Submit ▾ · Purchase · Alerts · About ▾

All Journal of Comp ▾ · Advanced Search

Volume 22, Issue 5
1 September 2017

Article Contents

Digital News Consumption
Data and Methods
Results
Discussion
Conclusion
References
About the Author

< Previous Next >

Digital News Consumption and Copyright Intervention: Evidence from Spain before and after the 2015 "Link Tax" FREE

Sílvia Majó-Vázquez, Ana S. Cardenal, Sandra González-Bailón

Journal of Computer-Mediated Communication, Volume 22, Issue 5, 1 September 2017, Pages 284–301, https://doi.org/10.1111/jcc4.12196
Published: 16 November 2017

Views ▾ PDF Cite Permissions Share ▾
AddThis

We analyze patterns of digital news consumption before and after a "link tax" was introduced in Spain. This new legislation imposed a copyright fee for showing snippets of content created by newspapers and resulted in the shutdown of Google News Spain. The Spanish copyright law is a precedent to the Copyright Directive currently submitted to the European Parliament, which is planning to impose a similar "link tax." We offer empirical evidence that can help evaluate the impact of that sort of intervention. We analyze data tracking news consumption behavior to assess changes in audience reach and audience fragmentation. We show that the law has no discernible impact on reach, but we identify an increase in the fragmentation of news consumption.

Issue Section: Original Article

View Metrics

Email alerts

New issue alert
Advance article alerts
Article activity alert

Receive exclusive offers and updates from Oxford Academic

Related articles in

Google Scholar

Related articles in PubMed

Relationship between jumping abilities and skeletal muscle architecture of lower limbs in humans: Systematic review and meta-analysis.

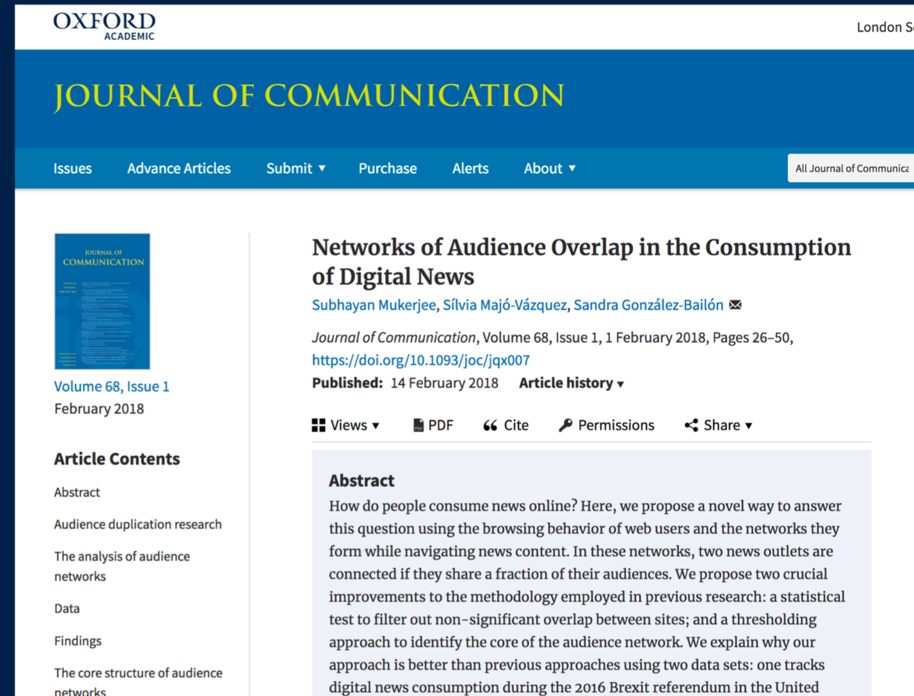
Contextual determinants of psychopathology.

Sílvia Majó-Vázquez

RISJ Post-Doctoral Research Fellow

@silviamajo

Silvia.majo-vazquez@politics.ox.ac.uk



OXFORD ACADEMIC

London Sch

JOURNAL OF COMMUNICATION

Issues | Advance Articles | Submit ▾ | Purchase | Alerts | About ▾

All Journal of Communica ▾

Volume 68, Issue 1
February 2018

Article Contents

- Abstract
- Audience duplication research
- The analysis of audience networks
- Data
- Findings
- The core structure of audience networks

Networks of Audience Overlap in the Consumption of Digital News

Subhayan Mukerjee, Sílvia Majó-Vázquez, Sandra González-Bailón ✉

Journal of Communication, Volume 68, Issue 1, 1 February 2018, Pages 26–50,
https://doi.org/10.1093/joc/jqx007
Published: 14 February 2018 **Article history** ▾

Views ▾ | PDF | Cite | Permissions | Share ▾

Abstract

How do people consume news online? Here, we propose a novel way to answer this question using the browsing behavior of web users and the networks they form while navigating news content. In these networks, two news outlets are connected if they share a fraction of their audiences. We propose two crucial improvements to the methodology employed in previous research: a statistical test to filter out non-significant overlap between sites; and a thresholding approach to identify the core of the audience network. We explain why our approach is better than previous approaches using two data sets: one tracks digital news consumption during the 2016 Brexit referendum in the United

Sílvia Majó-Vázquez

RISJ Post-Doctoral Research Fellow
@silviamajo
Silvia.majo-vazquez@politics.ox.ac.uk

Workshop’s Outline

1. Social Network Analysis (SNA)
2. Levels of SNA
3. Hands-on SNA

Network Science

- Network science is still a young discipline, it just emerged in the twenty-first century (Barabási, 2015)

- But it has seen an **explosive interest** from social researchers

- This popularity is explained by two main factors:
  - The potential of network tools to shed light on **the processes that underpin** social relations;

  - The increasing availability of digital trace data, which provides the opportunity to **map complex human interactions** in an unprecedented scale.

Networks

- Networks are basic representations of processes and capture the **basic connections** among those who take part of them

- The discipline of network science is devoted to understanding the **underlying forces** that drive those processes

- Since the landmark paper by Granovetter (Granovetter, 1973), social scientist mainly refers to this discipline as **social network analysis** and to the structures they analyze as **social networks**

References

The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

Analysis of social networks is suggested as a tool for linking micro and macro levels of sociological theory. The procedure is illustrated by elaboration of the macro implications of one aspect of small-scale interaction: the strength of dyadic ties. It is argued that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. The impact of this principle on diffusion of influence and information, mobility opportunity, and community organization is explored. Stress is laid on the cohesive power of weak ties. Most network models deal, implicitly, with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups.

A fundamental weakness of current sociological theory is that it does not relate micro-level interactions to macro-level patterns in any convincing way. Large-scale statistical, as well as qualitative, studies offer a good deal of insight into such macro phenomena as social mobility, community organization, and political structure. At the micro level, a large and increasing body of data and theory offers useful and illuminating ideas about what transpires within the confines of the small group. But how interaction in small groups aggregates to form large-scale patterns eludes us in most cases.

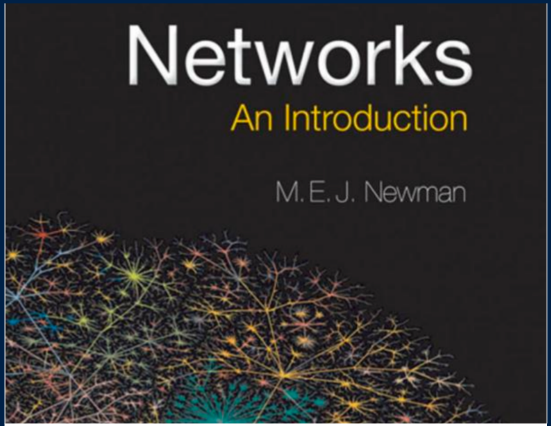
I will argue, in this paper, that the analysis of processes in interpersonal networks provides the most fruitful micro-macro bridge. In one way or another, it is through these networks that small-scale interaction becomes translated into large-scale patterns, and that these, in turn, feed back into small groups.

Granovetter, M. S. (1973). The strength of weak ties. American Journal of Sociology, 78(6), 1360–1380.

References

Barabasi, A-L. (2012) Network Science.

González-Bailón, S. (2017). Decoding the Social World. Data science and the Unintended Consequences of Communication. MIT Press.



Newman, M.E.J Networks and Introduction. Oxford University Press

Key Terms in Network Analysis

- **Node**/Actor/Vertex: The "individual"
- **Tie**/Line/Edge/Arc: The relationship between them
- **Matrix**: An array of elements, rows and columns, that represent the relationships between nodes in the network
- **Edgelist**: common use of representation of a graph by listing each of its dyads and their type of relation row by row.
- **Attribute:** characteristics of a node (e.g., race, gender, age, location, number of followers, network centrality scores….)

Ann McCrannie, 2013

Key Network Dimensions

- **Size**: Typically refers to the number of nodes (and edges) a network has.

- **Directionality**:
  - **Directed (or Asymmetric)**: a relationship can flow in both directions
  - **Undirected (or Symmetric)**: the relationship by definition must be the same for both parties.

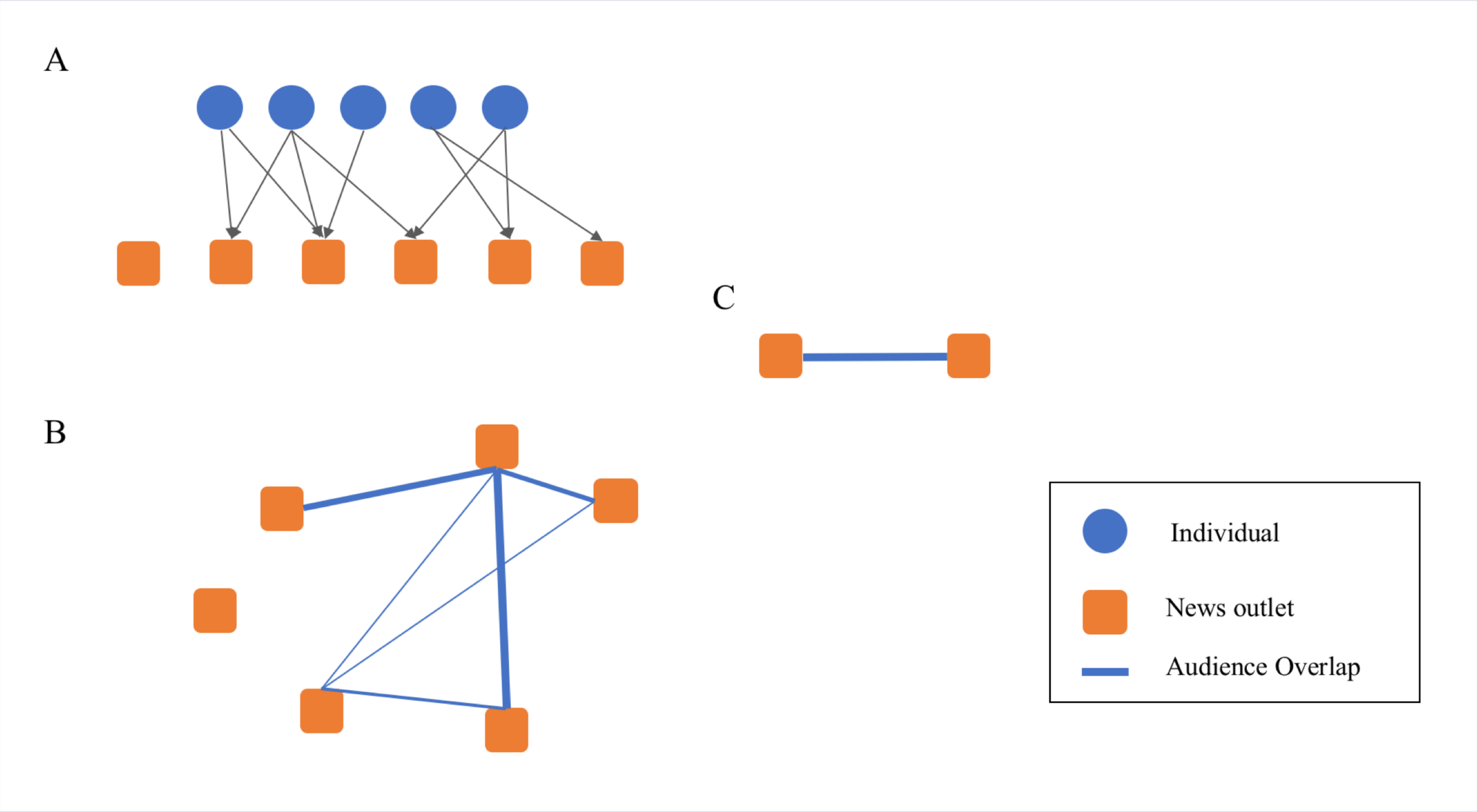
- **Value**:
  - **Unweighted** (binary): The tie is simply present or not (unvalued, unweighted)
  - **Weighted**: There is a value attached to the ties

Key Network Dimensions

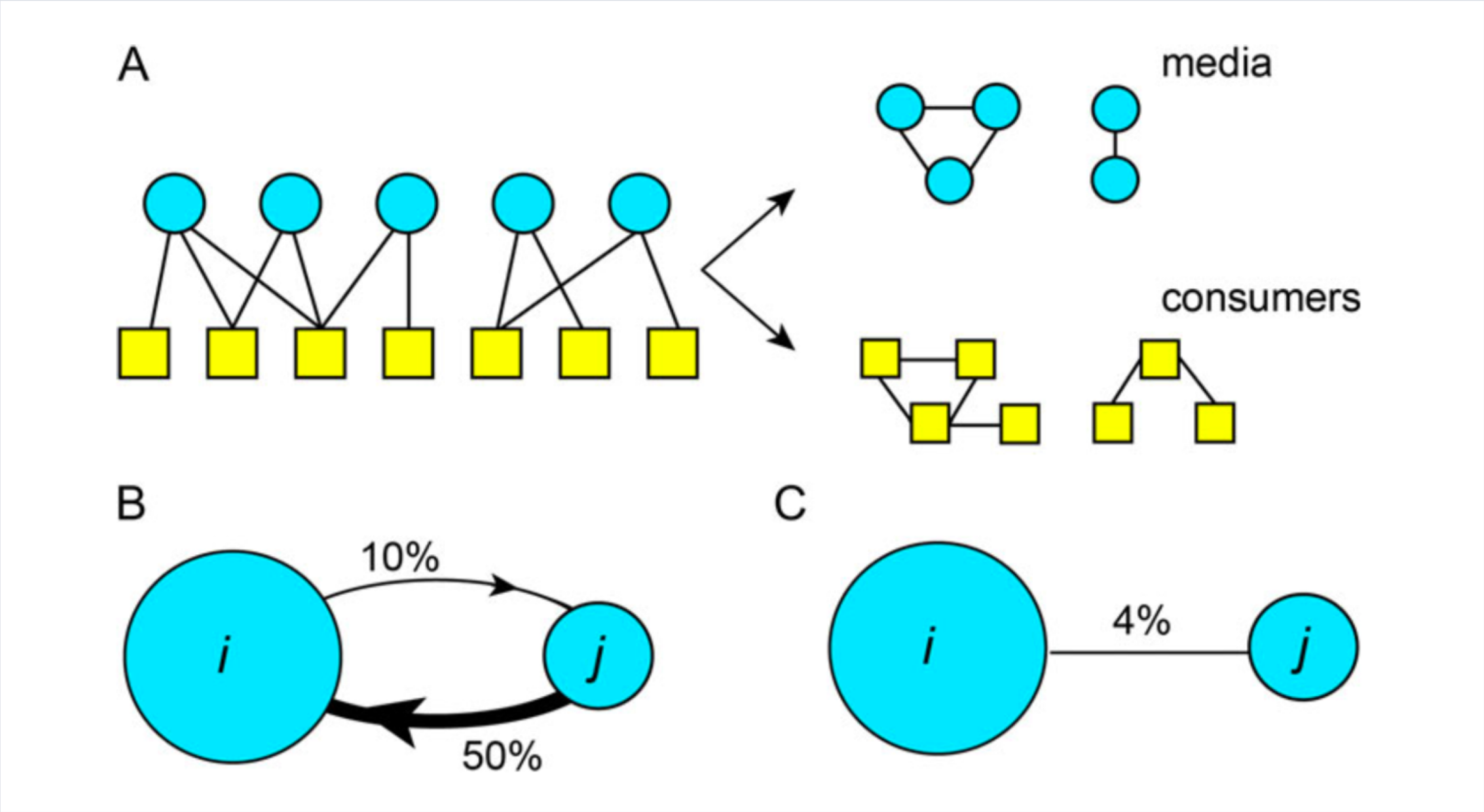
- **Mode:**
  - **One-mode** : relationships between one type of actor
  - **Two-mode**: two types of nodes and relationship only exist between nodes of different type (individual and event; actor and film; employers and employees; minister and governments)
- **Multiplexity**: Whether the relationship is of a single type, or whether there are multiple relations between the same actors
- **Time**: Whether the data is cross sectional or time-dependent, ”dynamic”

Ann McCrannie, 2013

Two-Mode to One-Mode

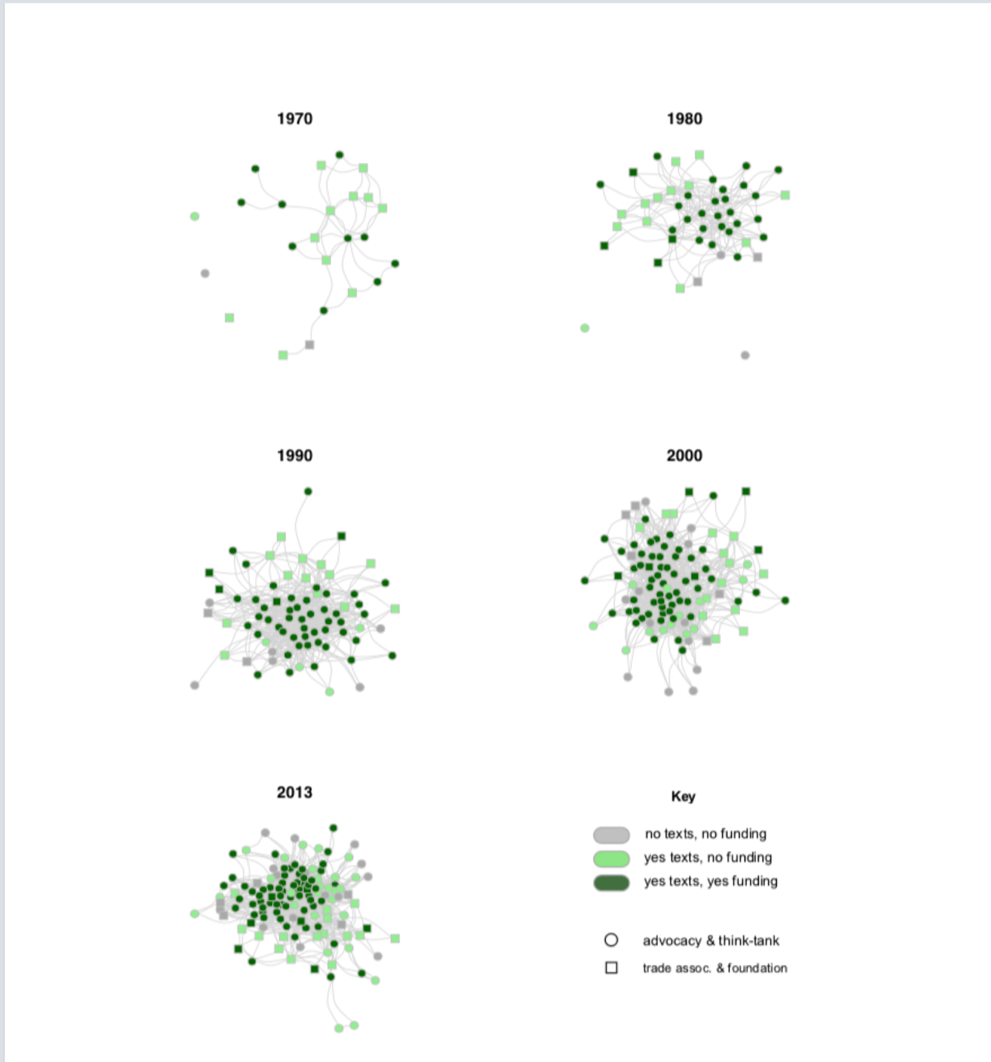


Weighted & Unweighted

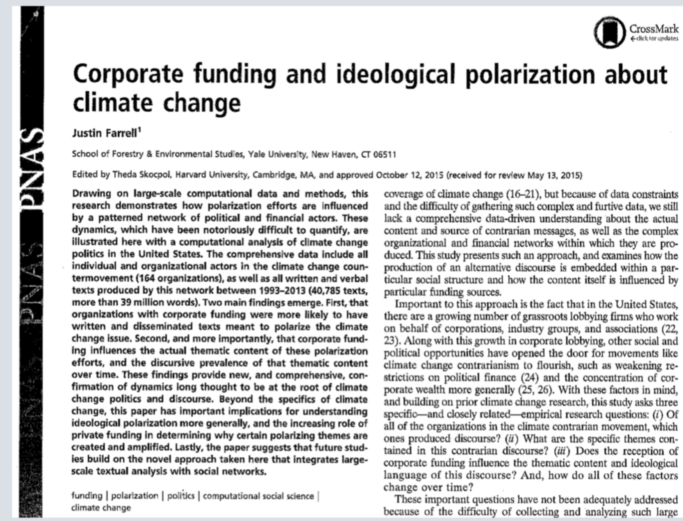


Mukerjee, S., Majo-Vazquez, S., & González-Bailón, S. (2018).

Organizational Networks



- One-mode network
- Undirected (symmetric)
- Unweighted (binary)



PNAS

CrossMark

Corporate funding and ideological polarization about climate change

Justin Farrell[1]

School of Forestry & Environmental Studies, Yale University, New Haven, CT 06511

Edited by Theda Skocpol, Harvard University, Cambridge, MA, and approved October 12, 2015 (received for review May 13, 2015)

Drawing on large-scale computational data and methods, this research demonstrates how polarization efforts are influenced by a patterned network of political and financial actors. These dynamics, which have been notoriously difficult to quantify, are illustrated here with a computational analysis of climate change politics in the United States. The comprehensive data include all individual and organizational actors in the climate change countermovement (164 organizations), as well as all written and verbal texts produced by this network between 1993–2013 (40,785 texts, more than 39 million words). Two main findings emerge. First, that organizations with corporate funding were more likely to have written and disseminated texts meant to polarize the climate change issue. Second, and more importantly, that corporate funding influences the actual thematic content of these polarization efforts, and the discursive prevalence of that thematic content over time. These findings provide new, and comprehensive, confirmation of dynamics long thought to be at the root of climate change politics and discourse. Beyond the specifics of climate change, this paper has important implications for understanding ideological polarization more generally, and the increasing role of private funding in determining why certain polarizing themes are created and amplified. Lastly, the paper suggests that future studies build on the novel approach taken here that integrates large-scale textual analysis with social networks.

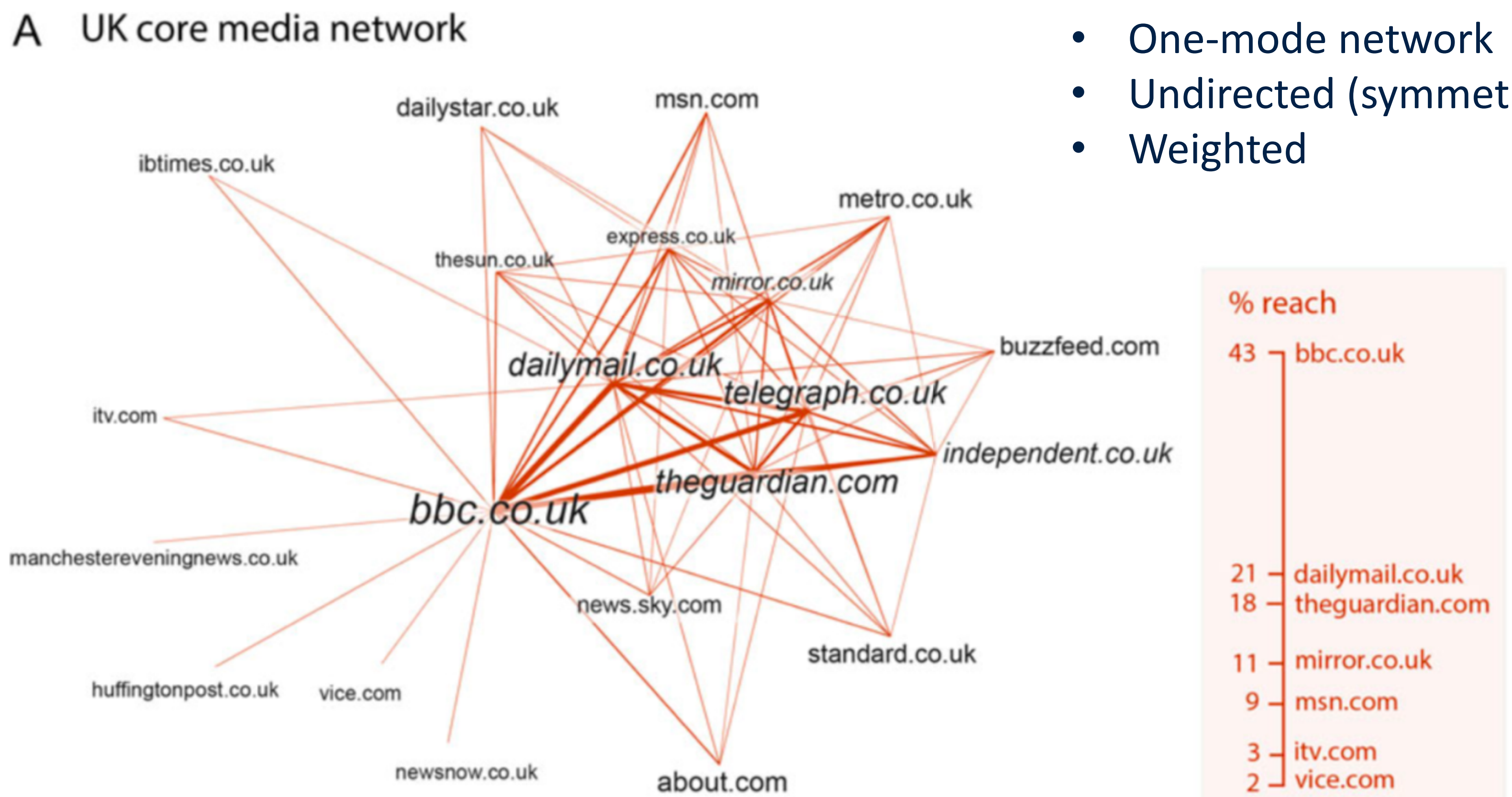
funding | polarization | politics | computational social science | climate change

coverage of climate change (16–21), but because of data constraints and the difficulty of gathering such complex and furtive data, we still lack a comprehensive data-driven understanding about the actual content and source of contrarian messages, as well as the complex organizational and financial networks within which they are produced. This study presents such an approach, and examines how the production of an alternative discourse is embedded within a particular social structure and how the content itself is influenced by particular funding sources.

Important to this approach is the fact that in the United States, there are a growing number of grassroots lobbying firms who work on behalf of corporations, industry groups, and associations (22, 23). Along with this growth in corporate lobbying, other social and political opportunities have opened the door for movements like climate change contrarianism to flourish, such as weakening restrictions on political finance (24) and the concentration of corporate wealth more generally (25, 26). With these factors in mind, and building on prior climate change research, this study asks three specific—and closely related—empirical research questions: (*i*) Of all of the organizations in the climate contrarian movement, which ones produced discourse? (*ii*) What are the specific themes contained in this contrarian discourse? (*iii*) Does the reception of corporate funding influence the thematic content and ideological language of this discourse? And, how do all of these factors change over time?

These important questions have not been adequately addressed because of the difficulty of collecting and analyzing such large

News Audience Networks



- One-mode network
- Undirected (symmetric)
- Weighted

Mukerjee, S., Majo-Vazquez, S., & González-Bailón, S. (2018).

News Audience Networks

Weighted & Unweighted

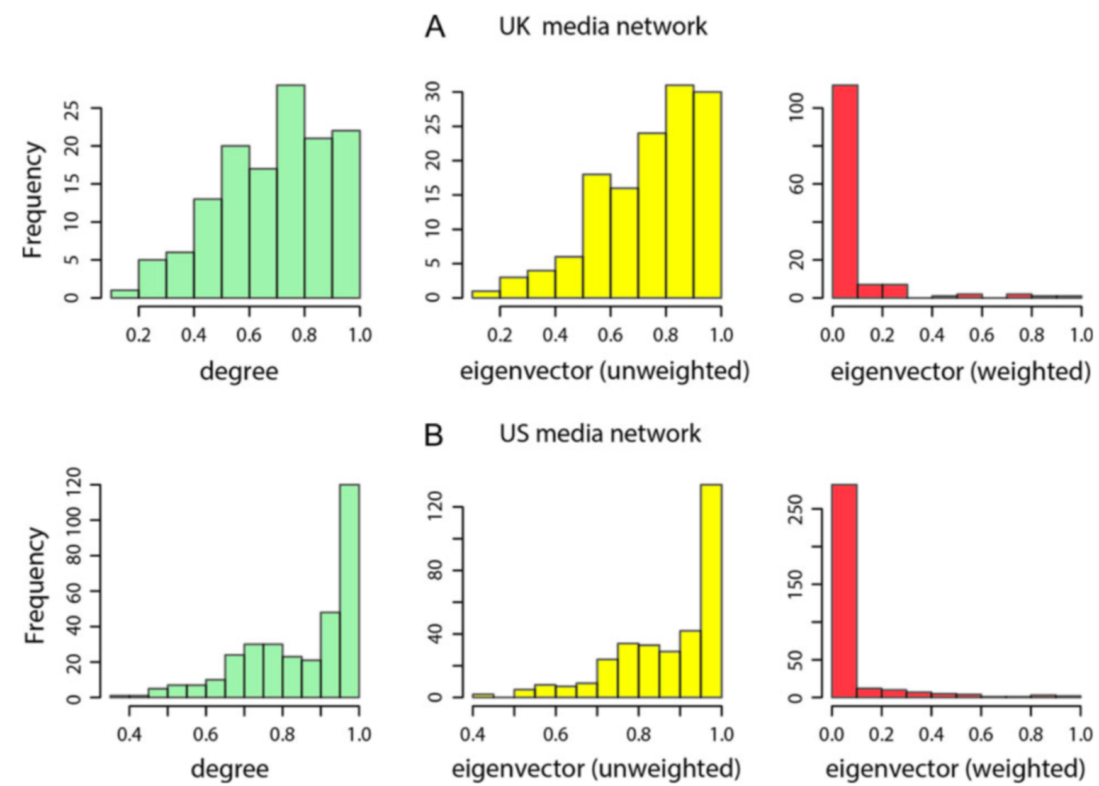


Figure 4 Centrality distributions in networks with unweighted and weighted ties.

Mukerjee, S., Majo-Vazquez, S., & González-Bailón, S. (2018).

Workshop's Outline

1. Social Network Analysis (SNA)

2. <u>Levels of SNA</u>

3. Hands-on SNA

Levels of SNA

- **Micro Level (role identification) :** We look at the individual properties of the nodes.

  The focus lies on single nodes and their specific positions within the overall structure; (node degree, betweenness, amongst other metrics).

(for more information on levels of analysis see Borge-Holthoefer & Gonzalez-Bailon, 2015).

Levels of SNA

- **Micro Level (role identification) :** We look at the individual properties of the nodes. The focus lies on single nodes and their specific positions within the overall structure; (node degree, betweenness, amongst other metrics).

- **Macro-level**: the focus is on the aggregation of node-level measures and their distribution (average degrees $\langle k \rangle$, the range of the degree distribution P(k), the average path length *L).*

(for more information on levels of analysis see Borge-Holthoefer & Gonzalez-Bailon, 2015).

Levels of SNA

- **Micro Level (role identification) :** We look at the individual properties of the nodes. The focus lies on single nodes and their specific positions within the overall structure; (node degree, betweenness, amongst other metrics).

- **Macro-level**: the focus is on the aggregation of node-level measures and their distribution (average degrees 〈k〉 , the range of the degree distribution P(k), the average path length *L)*.

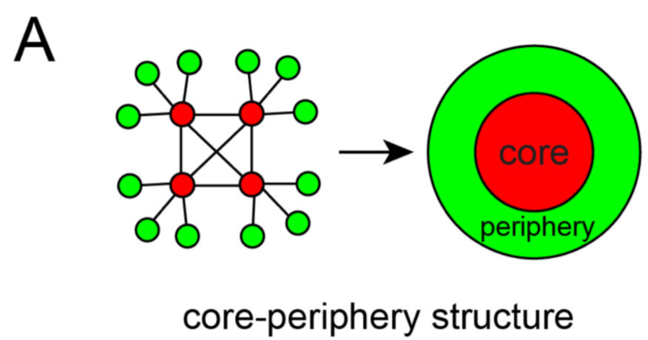
- **Meso-level:** one can account for the complexity of networks between the position of individual nodes and **the relational properties of the groups** where nodes are embedded (community detection and reduction techniques operate at this level)

(for more information on levels of analysis see Borge-Holthoefer & Gonzalez-Bailon, 2015).

Network Reduction Techniques

- **Core vs Periphery :**
  - A network has a core-periphery structure when there is a subset of nodes that are very well connected to each other and to peripheral nodes (this would be the core);
  - and another set of nodes that are well connected to the core, but not well connected to each other (these would be the periphery)

Figure 2. Core-Periphery Structures and Role Identification



Borge-Holthoefer, González-Bailón, 2014

Significant Ties

Filtered & Unfiltered (tie based filter)

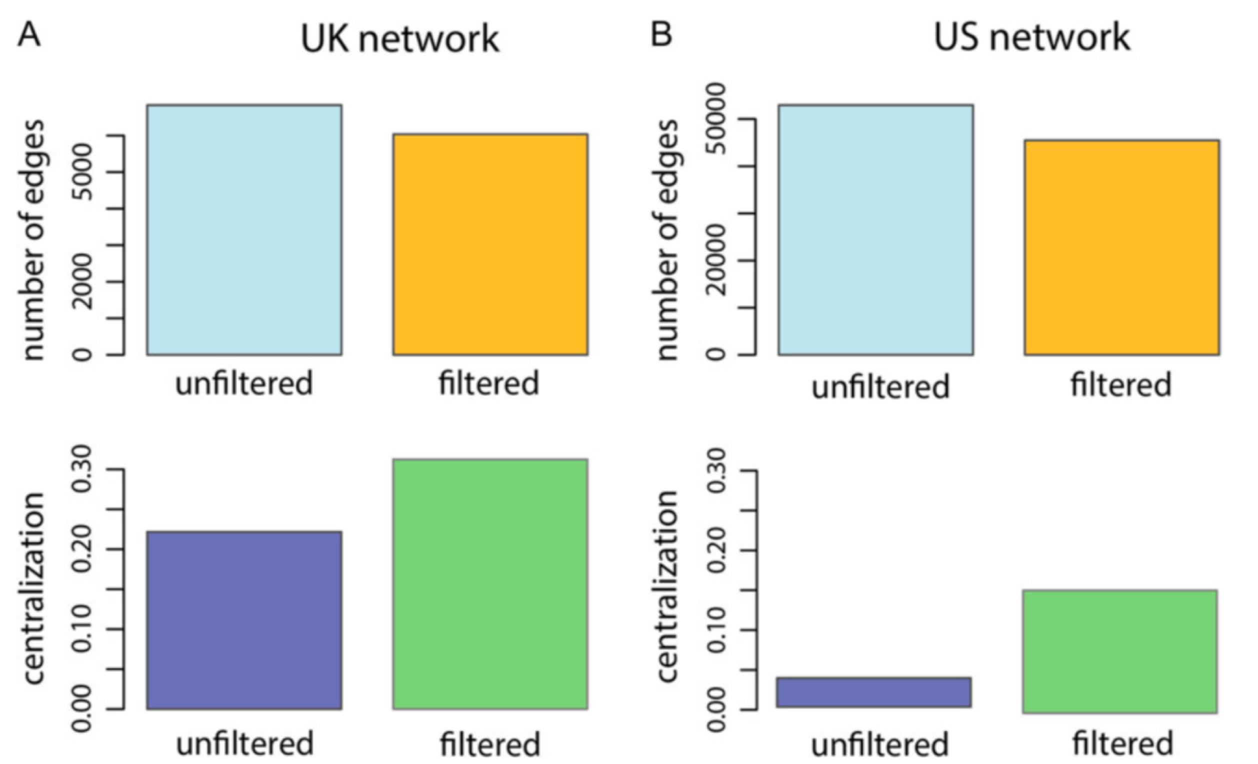
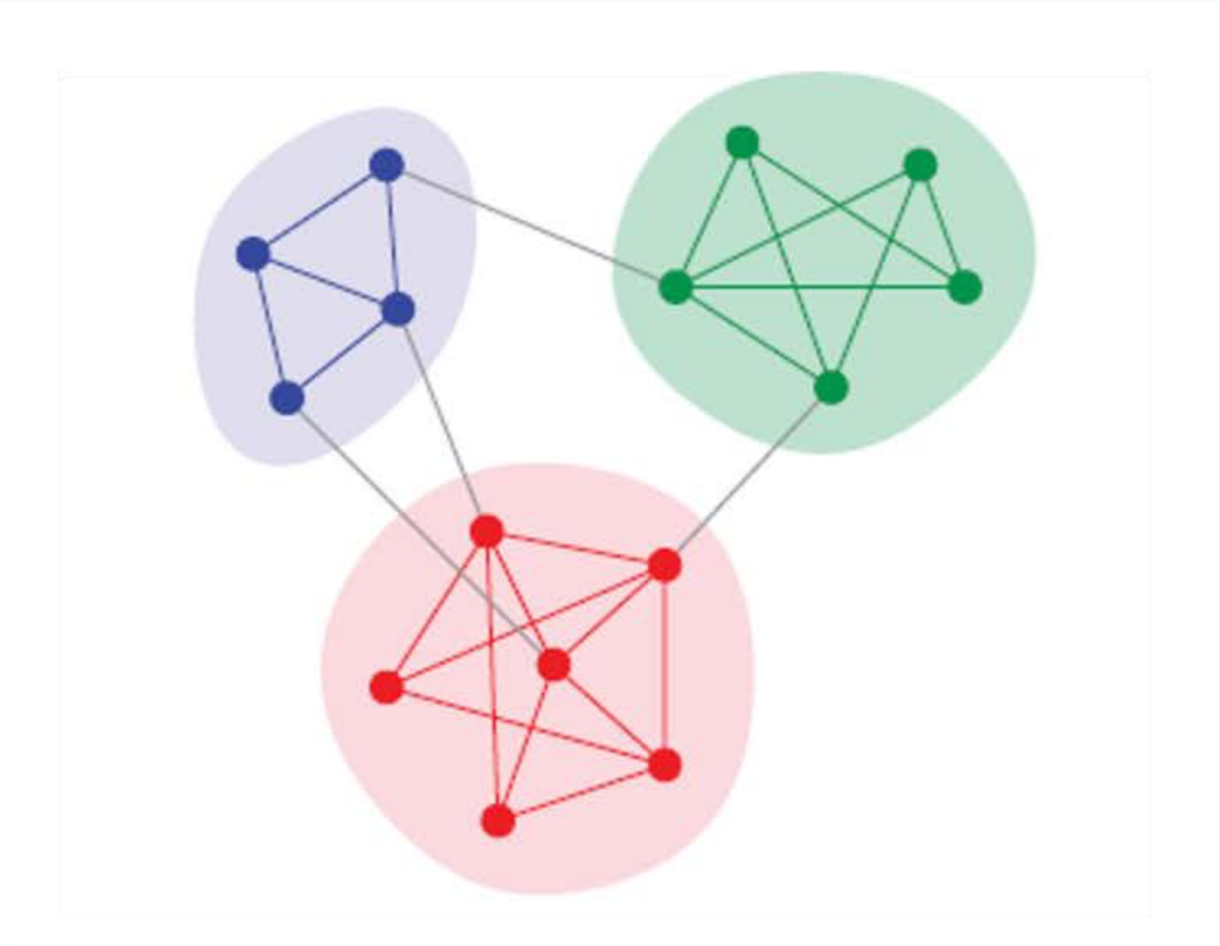


Figure 3 Audience networks before and after significance test.

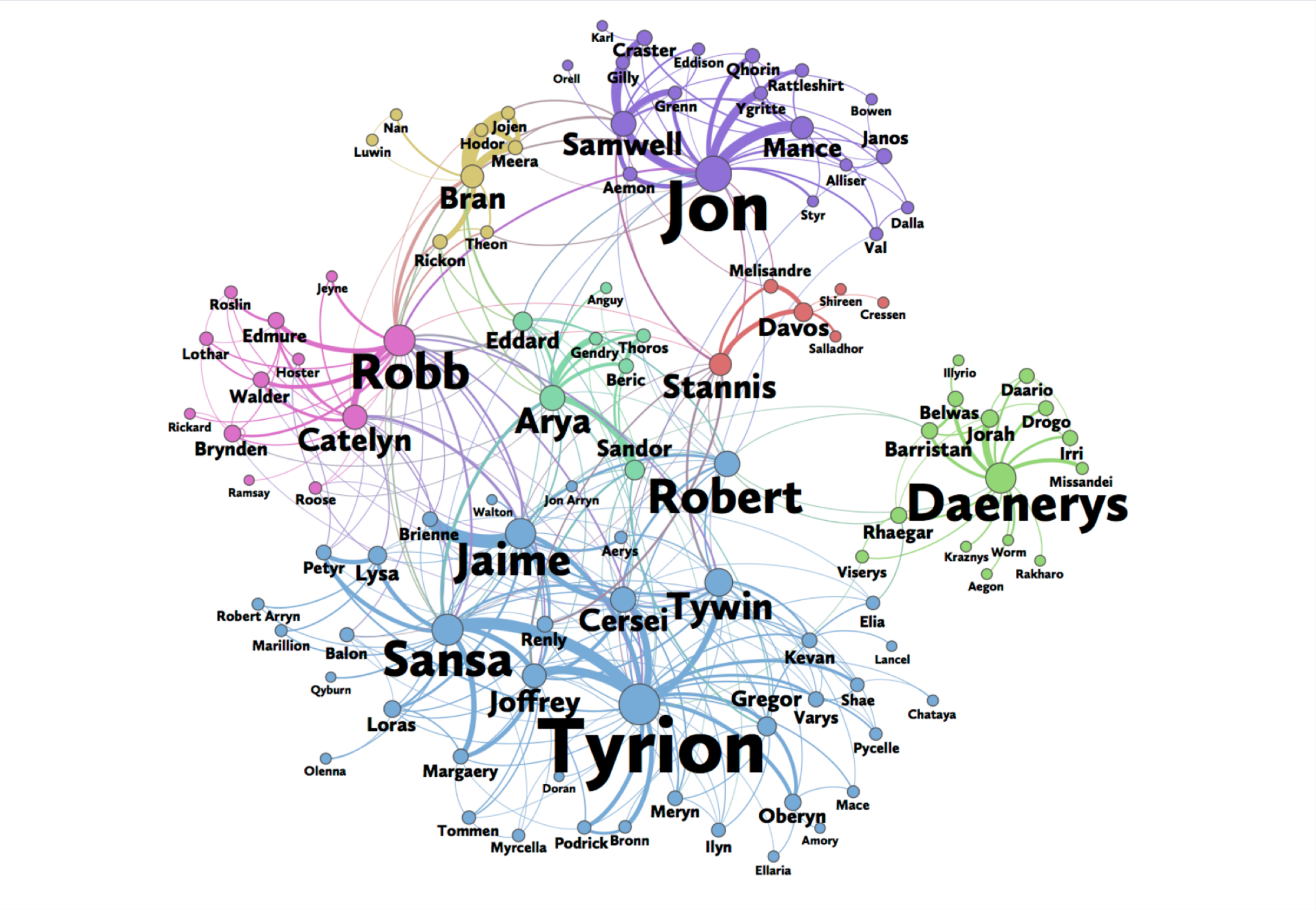
Mukerjee, S., Majo-Vazquez, S., & González-Bailón, S. (2018).

Community Detection



Newman, M.E.J (2012) Communities, modules and large—scale structure in networks . Nature Physics, 8

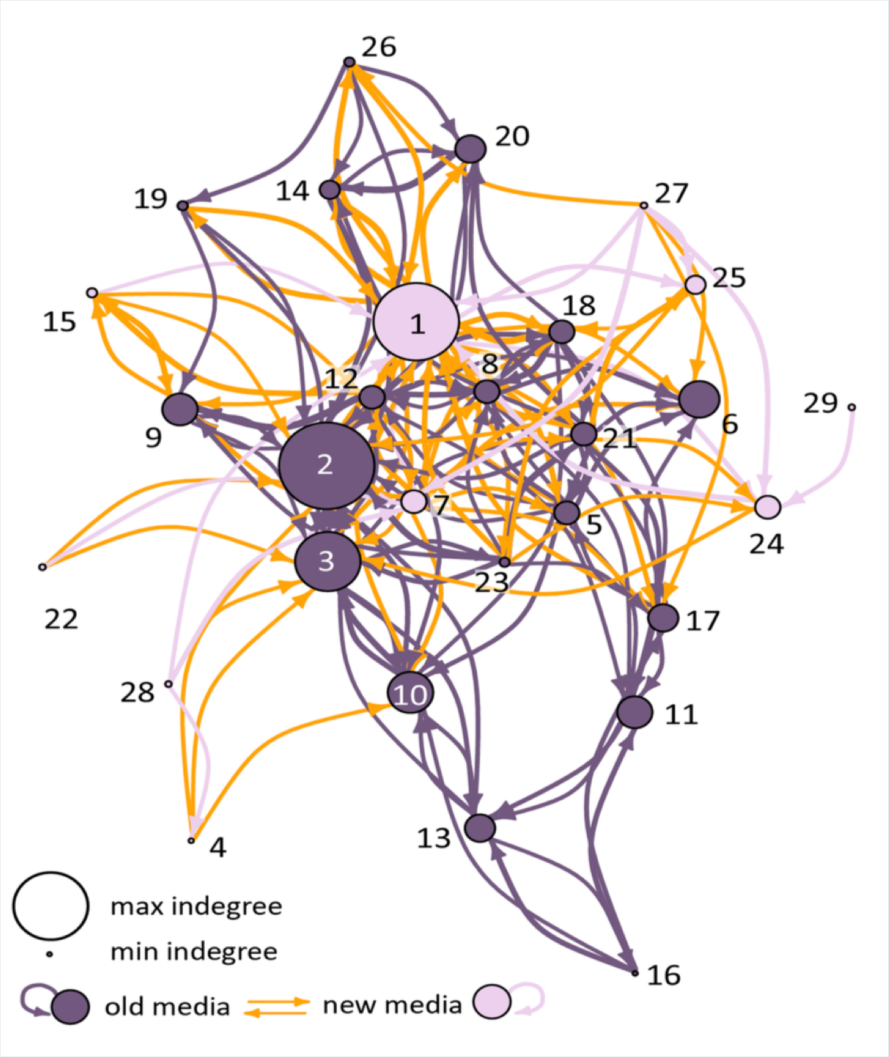
Network of Thrones



https://www.macalester.edu/~abeverid/thrones.html

Fragmentation in the News Production

Random Walk Algorithm
Directed
& Weighted Network



Majó-Vázquez, S., Cardenal, S.A., González-Bailón, S.., 2017

Characteristics of digital trace data

Generally helpful for research: big, always-on, and non-reactive

Generally problematic for research:

- incomplete,
- inaccessible,
- nonrepresentative,
- drifting,
- algorithmically confounded,
- sensitive,
- and dirty

(Salganik, 2017)

Characteristics of digital trace data

- Generally helpful for research:
  - Big
  - always-on: real time events, unexpected events, emergency events. (problems with longitudinal studies)
  - and non-reactive : social desirability bias → tendency of people to present themselves in the best possible way

(Salganik, 2017)

Characteristics of digital trace data

Generally problematic for research:

- **Incomplete**: missing information on demographic characteristics; behavior on other platforms;
- **Inaccessible**: data held by companies and governments are difficult to access
- **Nonrepresentative**: data useful for within-sample comparisons; bad for out-of-sample comparisons
- **Drifting**: population drift (who uses it), usage drift (who they use it) and system drift (change in the system) make it hard to use bid data sources to study long-term trends.
- **Algorithmically confounded**: systems highly engineered to induce specific behaviors
- **Sensitive**: includes personal data; can be used to trace personal identities
- **Dirty:** sophisticated spam or bots can make some political causes intentionally appear more popular

(Salganik, 2017)

References

Salganik, M. J. (2017) Bit by Bit. Princeton Press

“ Let’s Practice ! ”

REUTERS INSTITUTE

FOR THE STUDY OF JOURNALISM

Thank you !

Workshop for LSE CSS Hackathon

Sílvia Majó-Vázquez, PhD

Oxford, 19th April 2018

UNIVERSITY OF OXFORD