Manuscript ID JOU-22-0264

Developing the ‘News Niche’ as an Audience-Level Indicator of Fragmentation: A Theoretical Application of Community Detection Algorithms

**Editor’s Comments**

The reviewers have recommended specific revisions to your manuscript, which I believe will strengthen its argument and presentation overall. Therefore, I invite you to respond to these comments and revise your manuscript in accordance with them.

**Reviewer’s Comments**

Reviewer: 1

**Comment:** I appreciate the ground being covered in this piece but have a number of theoretical-conceptual questions/concerns that point to some existing gaps that need to be addressed.  It is argued that (1) politically motivated selective exposure is at work but that (2) people do not necessarily avoid that which is incongruent (p. 6).  In addition, the actions of others in social media networks influence what content is presented to an individual (the actuarial function) (p. 7).  I believe that what needs to be added to this discussion is a more formal incorporation of the concept of incidental exposure.  Nanz and Matthes have been doing solid work on this concept, and I know Stroud is as well. However, there are two types of incidental exposure to consider in the current work. One, there is casual consumption.  This is where someone comes across a bit of political content while not actively seeking nor actively trying to avoid it. Causal consumption is something noted in the work under review when mentioning the FOX and MSNBC relationship.  However, there is another type of incidental exposure, ineffective avoidance. There needs to be a recognition that audience members may be trying to avoid specific pieces of content but are still having contact with this messaging regardless of their efforts.  This may be due to the algorithmic dynamics being noted in this work, but there is evidence of ineffective avoidance in traditional mass media as well (See Levendusky et al.’s Democracy Amid Crises, Oxford University Press 2023).  Overall, there needs to be an accounting of people being much more effective seekers than effective avoiders of media content across all media.  In general, people can find the content they want (if they really want it).  However, they are much less able to avoid that content which they wish to not be part of their diets.

**Response:** This is an excellent point. Ultimately, other types of studies have extensively addressed the challenge of IE for both thinking about, and measuring, news exposure. Those broader conversations are beyond the scope of the paper, as our goal here was to link network analysis techniques with original theoretical perspectives on reading publics and niche news. Though this is a fascinating avenue for future research, etc.

- Have the INE measure in the dataset. Re-ran the models and it does nothing. Our measure of INE was not designed for this study (as it deals solely with social media), therefore, we left it out. Need to decide if we want to include it. See the alternative models.

-Need to update the conversation around IE to discuss this in the lit review. Be sure to add these citations.

-Update the Discussion section for future work.

**Comment:** On the industry front, there needs to be an acknowledgement that the various social media platforms vary widely in how much they curate content based on individual activity versus a more networked approach.  For example, it is my understanding that TikTok takes much less of a network approach than a platform like Facebook. TikTok looks at the actions of others to dictate curated content for a specific individual, but those others are not directly part of one’s social network. Whereas a place like Facebook does not in terms of accounting for the actions of more direct social connections. In short, how much are your claims about the actuarial function appropriate for the varied Web 2.0 platforms?  It seems to me that it would be variable between platforms.  We know from the work of Russ Neuman and others that we should not generalize empirical findings from one platform (e.g., Twitter) to all of social media, but I believe this would also be true of the field’s theoretical and conceptual claims. How universal are your claims about the actuarial function to the varied digital platforms?

**Response:** This is a thought-provoking idea. There are a couple of notes here. First and foremost, since we are empirically looking at the broad ‘attention network’ as patterns at the audience level, we really cannot make any claims about effects due to specific platforms. However, conceptually, we agree that actuarial mechanisms may not even be the same for each individual, as they would be dictated by one’s idiosyncratic choices (or at least dictated by some interaction between individual and their media environment). Thus, we can only claim that those within a cluster are more likely to share certain media habits relative to those in another cluster. We assume that the most popular news organizations in our sample are most likely being consumed within and across platforms (for example, some may stick to cable news on TV, while others may watch FOX TV and watch FOX on YouTube). Thus, the actuarial metaphor can only explain that audience clusters exist relative to organizations, and based on that assumption, the individual and audience-level analysis can describe general characteristics of the clusters (e.g., the ideological makeup of these clusters), but they cannot tell us exactly how clustering takes place. This is due to as you point out, the ‘black box’ nature of algorithms, namely that we do not know exactly how platforms may differ in their content filtering strategies, nor can we account for everyone’s mix of attention or exposure to various content filtering systems. In short, we are attempting to identify and describe audience clusters, but the actuary metaphor cannot tell us how clustering takes place, nor does it offer more fine-grained qualities of various filtering systems.

-Need to mention this in the lit review

-Need to clearly state that we can ID clusters and describe their characteristics but cannot reasonably generalize about exactly how the sorting of people or content takes place. In other words, future work could compare platforms and clustering characteristics.

-Can also stress that algorithms are one sorting mechanism for some audiences, but not all content is delivered this way. Thus, algorithms are one example of the actuary process, but we are developing a theoretical framework of the news niches that attempts to account for a complex interaction of people, technology, and institutions.

**Comment:** As for the competition and symbiosis distinction being made on p. 8, I believe there is a need to separate and flesh out the role of political interest (e.g., Prior’s work).  Heightened political interest creates shared audience for FOX and MSNBC.  Heightened political interest creates shared audience between The New York Times and The Washington Post.  With this being stated, shared audience may also come through a correspondingly low levels of political interest. For example, someone with low political interest may be television channel surfing and this leads to coming across both MSNBC and FOX content. Or, folks with low political interest may go to a media news aggregator on their smartphones (perhaps for entertainment news) and come across content from both The New York Times and The Washington Post.  Overall, I believe political interest plays a role in the communication dynamics being explored in this work, but this important concept has not been given its proper due in the work’s arguments.  The hypotheses and research questions are all about ideology, but what about the role of political interest?

**Response:** Good point. We have this variable, so should flesh it out a bit in the lit review and add a model that predicts niche membership. Made a correlation table to show these relationships.   
  
**Comment:** Some methodological assumptions to address as well – for example, the concept of “selection valence” (p. 13)– does the “selection” here refer to what outlets they offered in the open-ended responses or does “selection” mean that these are the outlets they actively selected for consumption?  If the former, that is fine (but may create some confusion). If the latter, then I think it is a big assumption that the three outlets people list are necessarily sought out by the audience members.  I would have liked to have seen follow-up questions for each asking (1) whether the respondents actively sought out the listed outlets and (2) whether the respondents actively sought to avoid these outlets.  Given these folks are only being asked to list the three outlets, perhaps the avoidance items are not needed.  However, I would think it would be wise to account for the active seeking component.

**Response:** OK. Selection only refers to what they selected. We assume there is more intent here, as the response rely on free recall. A note on why free recall might better reflect one’s choices. Next, can we add the basic IE measure as a control? That answers this question (see models).   
  
**Comment:** I see that you do have a measure of political interest but that it is treated as exogenous.  Overall, I am making an argument in this review that political interest needs to be made more of an endogenous component for the study.  Ideology is one player, but interest is another. Given the ground being covered, I find it challenging to study one as endogenous while the other sits in the wings (as an exogenous element). The findings being offered represent viewing the mediated information world through a purely ideological prism.  However, what is being overlooked is the fact that people’s contact with political information is determined in no small measure by political interest as well.  The dynamics of ideology and interest need to be studied in combination, not in relative isolation of one another.  I would like to see what this work can do to make interest more of an endogenous element.   Such an undertaking would involve not just using the individual-level audience measure.  It would also involve coding the outlets for something like information complexity or original information gathering.  Places like The New York Times or The Washington Post or The Wall Street Journal are sought after by those high in political interest because these outlets are the primary information gatherers – they offer the new reporting that everyone else talks about. Is there a way to begin to look at the media system from a political interest lens and to then look at how this all shakes out? I could see the ideological-based analyses being offered and then looking at the same dynamics through a political interest lens, but then what would be interesting would be to look at a true combination of ideology and interest.

**Response:** OK. Could add an RQ (RQ1b: with b asking about interest of individuals within each niche). Made a correlation table to address this. Also test a model with Interest as an interaction term. Looks good if we let ideology float but not if we randomize Interest across groups.   
  
  
Reviewer: 2  
  
**Comment:** This was a very interesting article that made an argument for the use of “news niches” to evaluate media partisanship and audience news consumption habits instead of the more traditional hyperlink tracing that shows heightened polarization. The news niche perspective helps account for fluctuation, variation, and overlap in audience habits. It can also address the use of algorithmic recommendations for news content.

**Response:**

**Comment:** On one hand, the methodology is sound - the Qualtrics distributed survey and statistical significance are clear. Yet, there were still many questions when it came to claims that could be inferred from this data:

**Response:**  
  
**Comment:** 1. There is no clear definition of "mainstream" or "media system" in the literature review.

**Response:** Thanks for catching this. We clearly define our use of the term on P.X

**Comment:** 2. p.11 states that "survey respondents were asked three times to 'write the name of a news outlet' yet on p. 23 under limitations it stated that "multiple responses" from each participant were not included. Does that mean if they wrote Fox News twice they didn't count it twice or does that mean that only one source from each participant was used? this needs to be clarified.

**Response:** It depends. In general, we used all three responses. However, for the network analysis, and following work on data filtering, we eliminated ‘self-loops’ as to no artificially inflate the audience for any organization by double counting respondents.

**Comment:** 3. did the authors collect demographic information and if so, how did this factor into cable vs. elite?

**Response:** We did. Added a correlation table to address this and respond to R1.

**Comment:** 4. A frequency count of ALL listed sources is needed. On Table B1 it lists "right-leaning sphere" etc. but what was included in those spheres? why is Breitbart separate from this sphere? Does this sphere account for Podcasts, radio, and/or YouTubers?

**Response:**  We chose not to do this for two reasons. First, and again following recent recommendations on network construction and filtering, any organization that did not receive at least two mentions would be deleted. The purpose of this filtering is to minimize noise in the network from inflating the ‘long tail’ of audience dispersion (Citations). We wanted to keep that important information. Second, at the same time, researchers must always balance precision with parsimony. In this case, any outlet that received less than 10 mentions and clearly fell into an ideological category was wrapped into the ‘right’, ‘left’ and ‘neutral’ sphere. This strikes a balance between filtering important descriptive data and the need for concision in network construction. These decisions, as well as others mentioned here, where not fully reported and explained. (There is a footnote in Appendix X). This was due to space considerations. To address this issue, we moved the discussion of the text analysis to an expanded Appendix section that we hope better explains these decisions.

**Comment:** 5. the separation from "elite" and "cable" still needs more work. Many of these sources in Table B1 are not either of those. For example, OAN is classified as "cable/tv" but few people watch through this medium. Also, how did you account for programs like Joe Rogan Experience or other podcasts (were none listed?) or Radio programs?

**Response:** There is unfortunately no perfect label for the clusters, as the nature of the media environment is one of complexity. We understand that platforms and organizations are not completely congruous. This is a reality that we cannot fully address. But as Richard Rorty argued, there is always a risk of failure when scholars attempt to ‘cut nature at its joints’. We chose these labels based on the work around media logic, that is, the set of economic, technological, and institutional incentives and constraints on the production of media content. In general, we feel that the organizations within each category loosely reflect a particular logic, based on the medium and platforms that borne those organizations. Of course, there is considerable overlap, and one could take exception to one or more outlets going in one category over another. This is a limitation that needs to be addressed, both within the current paper, and the field, as clustering based on network analysis is an inherently data-driven process. This fact connects to the general finding of work in this area, that fragmentation of audienes is simply not observed via this methodology. We updated the discussion to address this point and added a fuller explanation of our labels in the revised Appendix.

As for Joe Rogan, no respondent reported relying on that outlet for news. Podcasts were given their own category of ideology was not identifiable. (need to check this)

**Comment:** 6. Although the authors said they used "clustering algorithms" and "Louvain clustering" it is still not clear how the network was built - what were considered nodes and what were considered edges/weight of those edges?

**Response:** Need to clarify this in the text. It is buried in the Methods section but can explitly state this earlier in the paper.

**Comment:** 7. Most organizations in the local niches were proximity based - does that mean they were coded as "neutral" b/c the editorial valence could not be determined or were these still highly partisan?

**Response:** They were coded as such if they were local news markets.

**Comment:** 8. How were the ideological stances of non-traditional sources (like podcasts or tiktok influencers) coded. Were these not ever listed?

**Response:** These were rarely listed, but if we could not ID the ideology (or where there would be no clear bias), they were folded into the ‘neutral’ category. It is a testament to the coding scheme that these outlets cleanly fell into the ‘local/aggregators’ cluster after cluster analysis. Again, point to the Appendix here.

**Comment:** 9. Labels "elite press" and/or "cable" read very dated given the subscriptions to newspapers continue to fall and I believe that less than 50% of people have cable anymore (there is a Pew study on this I think?)

**Response:**  This is similar the comment above. These labels were chosen based on loose adherence to a media logic. See above.   
  
**Comment:** It is unclear what the "three" rhetorical construction advantages are that they state on p.9 - the first one sets off the preliminary advantage but the other two are lost in subsequent paragraphs and need to be clarified.

**Response:** Need to revise these sections.   
  
**Comment:** Overall, it's challenging to identify what the central takeaway from this article is. The data set is impressive, so it's sure to be there but as it reads now is buried and needs to be more clearly identifiable. For example, they mentioned that cable/local niches do not align strongly with ideological slants or might be influenced by organizational symbiosis - could the authors provide some theoretical insights for why this is?

**Response:** Do this.