



**HAL**  
open science

## Follow the guides to the filter niches

Quentin Villermet, Jérémie Poiroux, Manuel Moussallam, Thomas Louail,  
Camille Roth

### ► To cite this version:

Quentin Villermet, Jérémie Poiroux, Manuel Moussallam, Thomas Louail, Camille Roth. Follow the guides to the filter niches. RecSys'21: Fifteenth ACM Conference on Recommender Systems, Sep 2021, Amsterdam, Netherlands. hal-03495269

**HAL Id: hal-03495269**

**<https://hal.science/hal-03495269v1>**

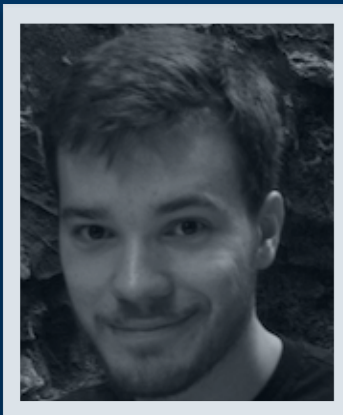
Submitted on 20 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Follow the guides to the filter niches

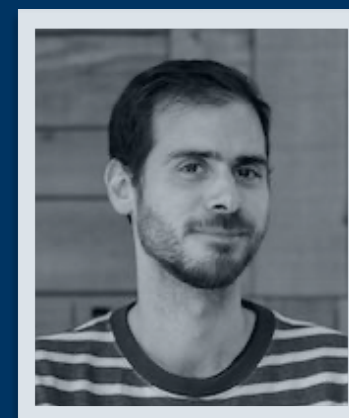
Disentangling human and algorithmic curation  
in online music consumption



Quentin Villermet



Jérémie Poiroux

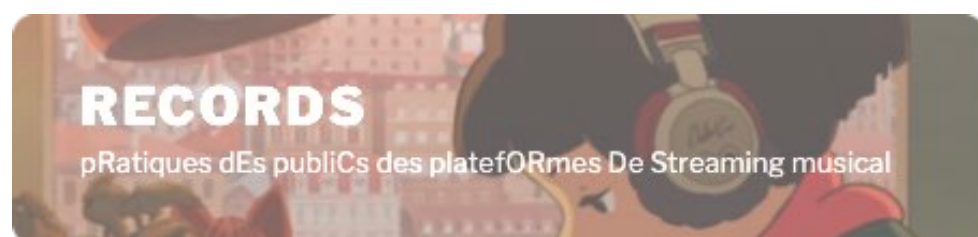


Manuel Moussallam



Thomas Louail

Camille Roth

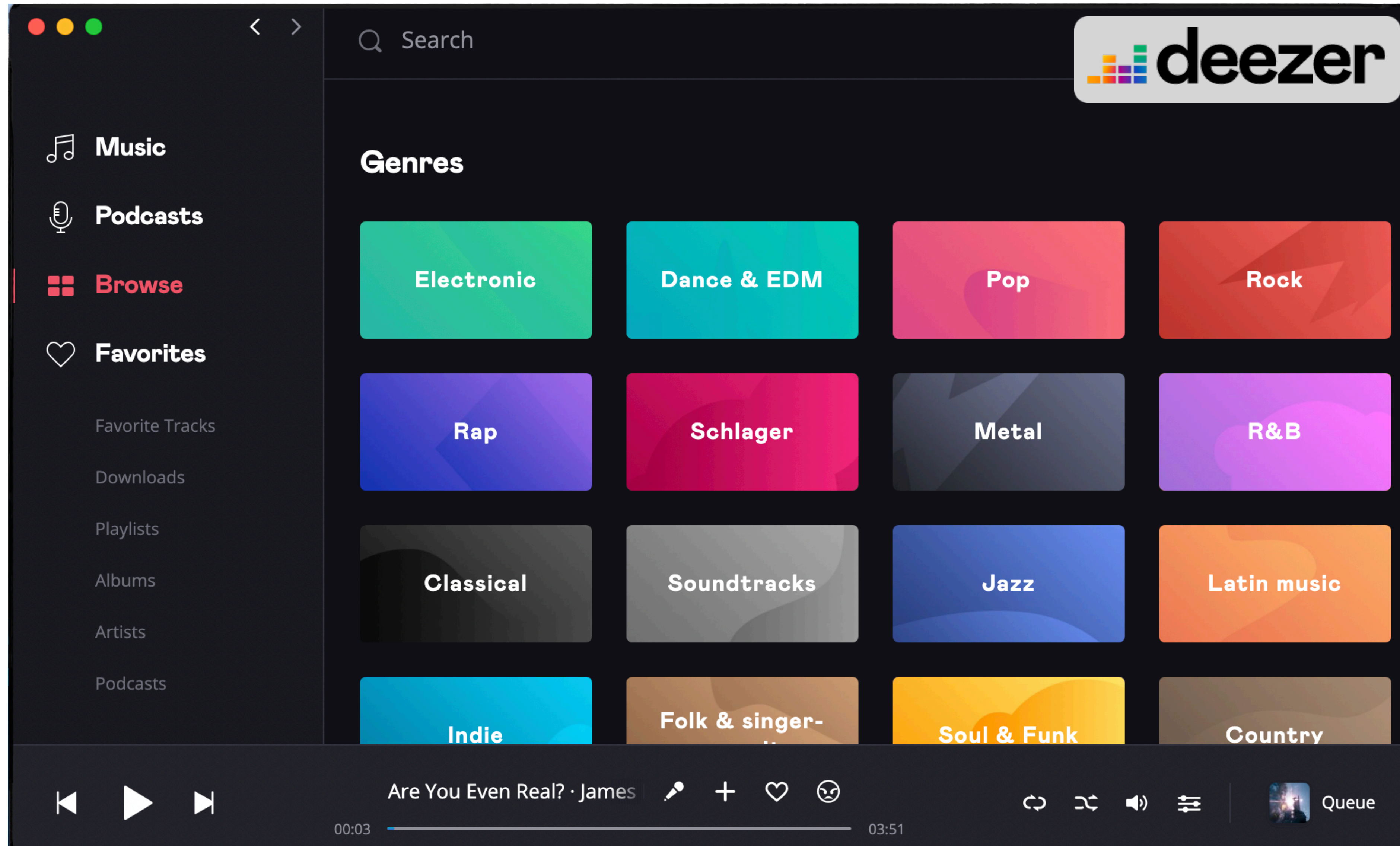


Géographie-cités  
UMR 8504



Centre Zentrum Marc Bloch  
Computational Social Science Team

# Three types of affordances



# Filter bubbles ?

State of the art: not very bubbly...

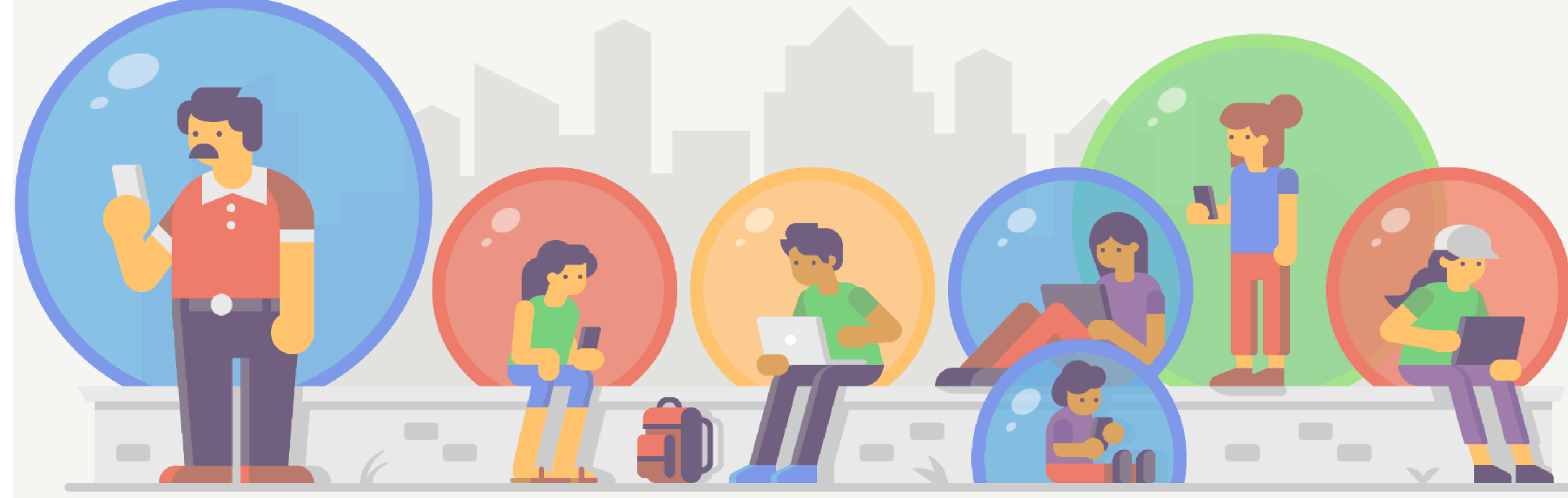


illustration  
courtesy of  
DuckDuckGo

*youtube*

Zhou, Khemmarat,  
Gao, 2010

"The impact of YouTube  
recommendation  
system on video views"

*movielens*

Nguyen, Hui, Harper,  
Terveen, Konstan, 2014

"Exploring the filter  
bubble: the effect of using  
recommender systems  
on content diversity"

*facebook*

Bakshy, Messing,  
Adamic, 2015

"Exposure to ideologically  
diverse news and opinion  
on Facebook"

*tumblr*

Aiello, Barbieri,  
2017

"Evolution of ego-  
networks in social  
media with link  
recommendations"

*spotify*

Datta, Knox,  
Bronnenberg, 2018

"Changing Their Tune: How  
Consumers' Adoption of  
Online Streaming Affects  
Music Consumption and  
Discovery"

*google news*

Haim, Graefe,  
Brosius, 2018

"Burst of the Filter  
Bubble? Effects of  
personalization on  
the diversity of  
Google News"

# Filter bubbles ?

State of the art: not very bubbly...



illustration  
courtesy of  
DuckDuckGo

*youtube*

Zhou, Khemmarat,  
Gao, 2010

"The impact of YouTube  
recommendation  
system on video views"

*movielens*

Nguyen, Hui, Harper,  
Terveen, Konstan, 2014

"Exploring the filter  
bubble: the effect of using  
recommender systems  
on content diversity"

*facebook*

Bakshy, Messing,  
Adamic, 2015

"Exposure to ideologically  
diverse news and opinion  
on Facebook"

*tumblr*

Aiello, Barbieri,  
2017

"Evolution of ego-  
networks in social  
media with link  
recommendations"

*spotify*

Datta, Knox,  
Bronnenberg, 2018

"Changing Their Tune: How  
Consumers' Adoption of  
Online Streaming Affects  
Music Consumption and  
Discovery"

*google news*

Haim, Graefe,  
Brosius, 2018

"Burst of the Filter  
Bubble? Effects of  
personalization on  
the diversity of  
Google News"

**except (admittedly...) in the case of explicit preferences**

Thurman,  
Schiffers, 2012

"The Future of  
Personalization at News  
Websites"

Zuiderveen Borgesius, Trilling, Möller,  
Bodó, de Vreese, Helberger, 2016

"Should we worry about filter bubbles?"

Dylko, Dolgov, Hoffman,  
Eckhart, Molina, Aaziz, 2017

"The dark side of technology: An  
experimental investigation of the influence  
of customizability technology on online  
political selective exposure"

# Filter bubbles ?

State of the art: not very bubbly...



illustration  
courtesy of  
DuckDuckGo

*youtube*

Zhou, Khemmarat,  
Gao, 2010

"The impact of YouTube  
recommendation  
system on video views"

*movielens*

Nguyen, Hui, Harper,  
Terveen, Konstan, 2014

"Exploring the filter  
bubble: the effect of using  
recommender systems  
on content diversity"

*facebook*

Bakshy, Messing,  
Adamic, 2015

"Exposure to ideologically  
diverse news and opinion  
on Facebook"

*tumblr*

Aiello, Barbieri,  
2017

"Evolution of ego-  
networks in social  
media with link  
recommendations"

*spotify*

Datta, Knox,  
Bronnenberg, 2018

"Changing Their Tune: How  
Consumers' Adoption of  
Online Streaming Affects  
Music Consumption and  
Discovery"

*google news*

Haim, Graefe,  
Brosius, 2018

"Burst of the Filter  
Bubble? Effects of  
personalization on  
the diversity of  
Google News"

**except (admittedly...) in the case of explicit preferences**

Thurman,  
Schiffes, 2012

"The Future of  
Personalization at News  
Websites"

Zuiderveen Borgesius, Trilling, Möller,  
Bodó, de Vreese, Helberger, 2016

"Should we worry about filter bubbles?"

Dylko, Dolgov, Hoffman,  
Eckhart, Molina, Aaziz, 2017

"The dark side of technology: An  
experimental investigation of the influence  
of customizability technology on online  
political selective exposure"

**...and some cases**

*youtube*

Roth, Mazières,  
Menezes, 2020

"Tubes and bubbles:  
topological confinement of  
YouTube recommendations"

*spotify*

Anderson, Maystre, Anderson,  
Mehrotra, Lalmas, 2020

"Algorithmic effects on  
the diversity of  
consumption on spotify"

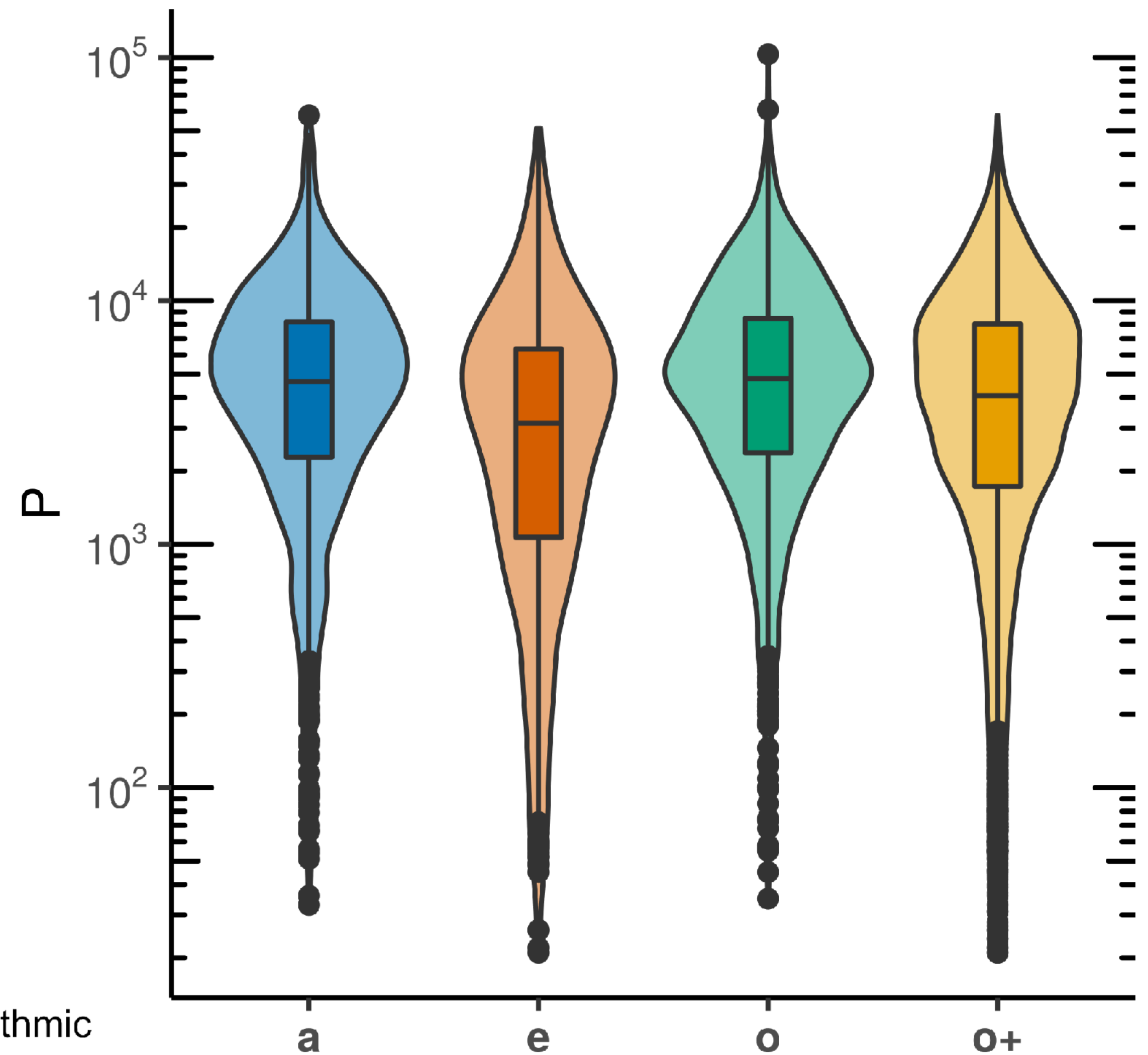
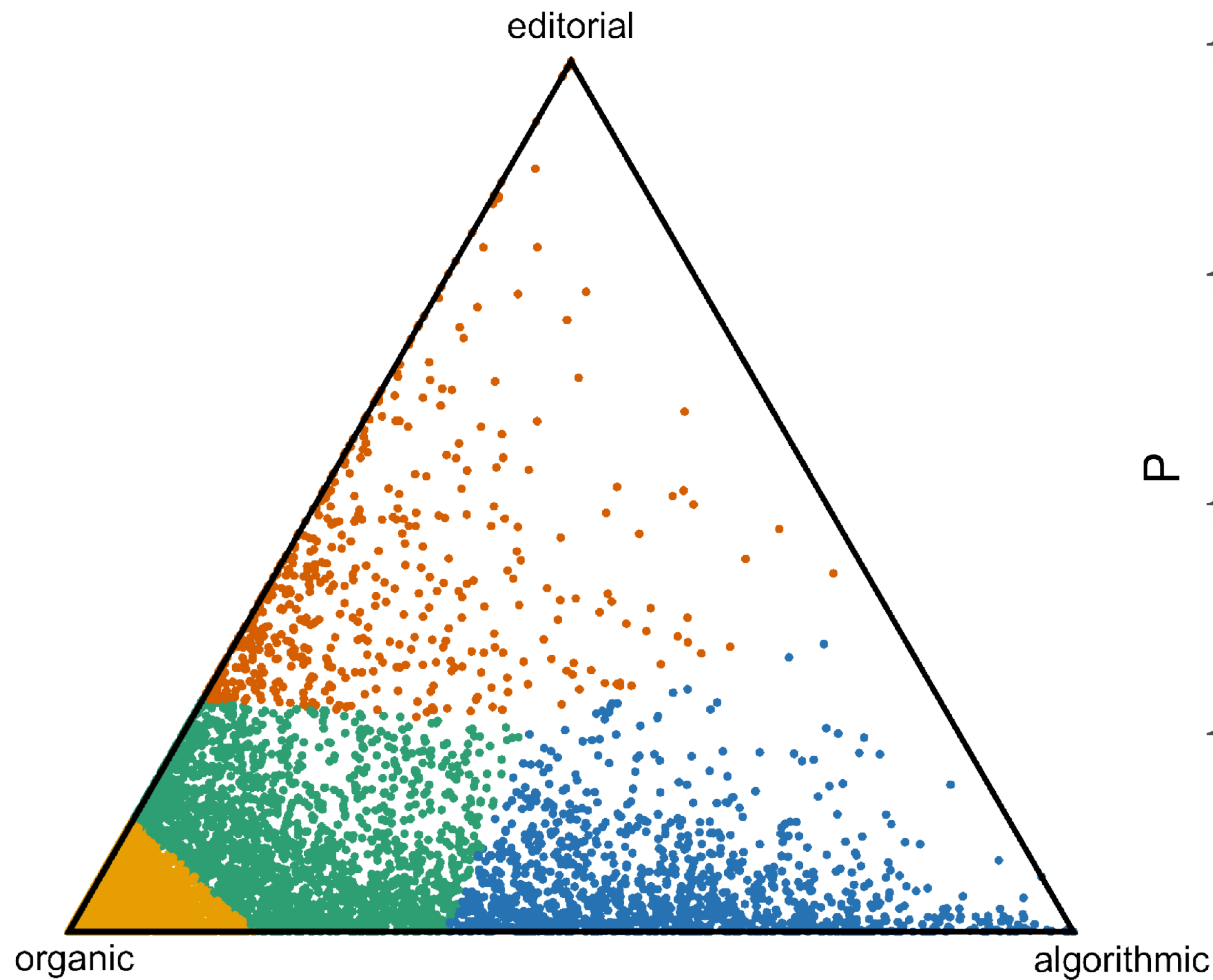


Figure 1: Use profiles and classes, where each dot on the ternary plot corresponds to a user of barycentric coordinates  $(p_a, p_e, p_o)$ , and each color refers to one of the four categories **a** (blue), **e** (red), **o** (green), **o+** (yellow).

8639 users in total, of which:

- 989 **a**
- 655 **e**
- 1614 **o**
- 5381 **o+**

# Two dimensions of diversity

## Dispersion, i.e. S/P

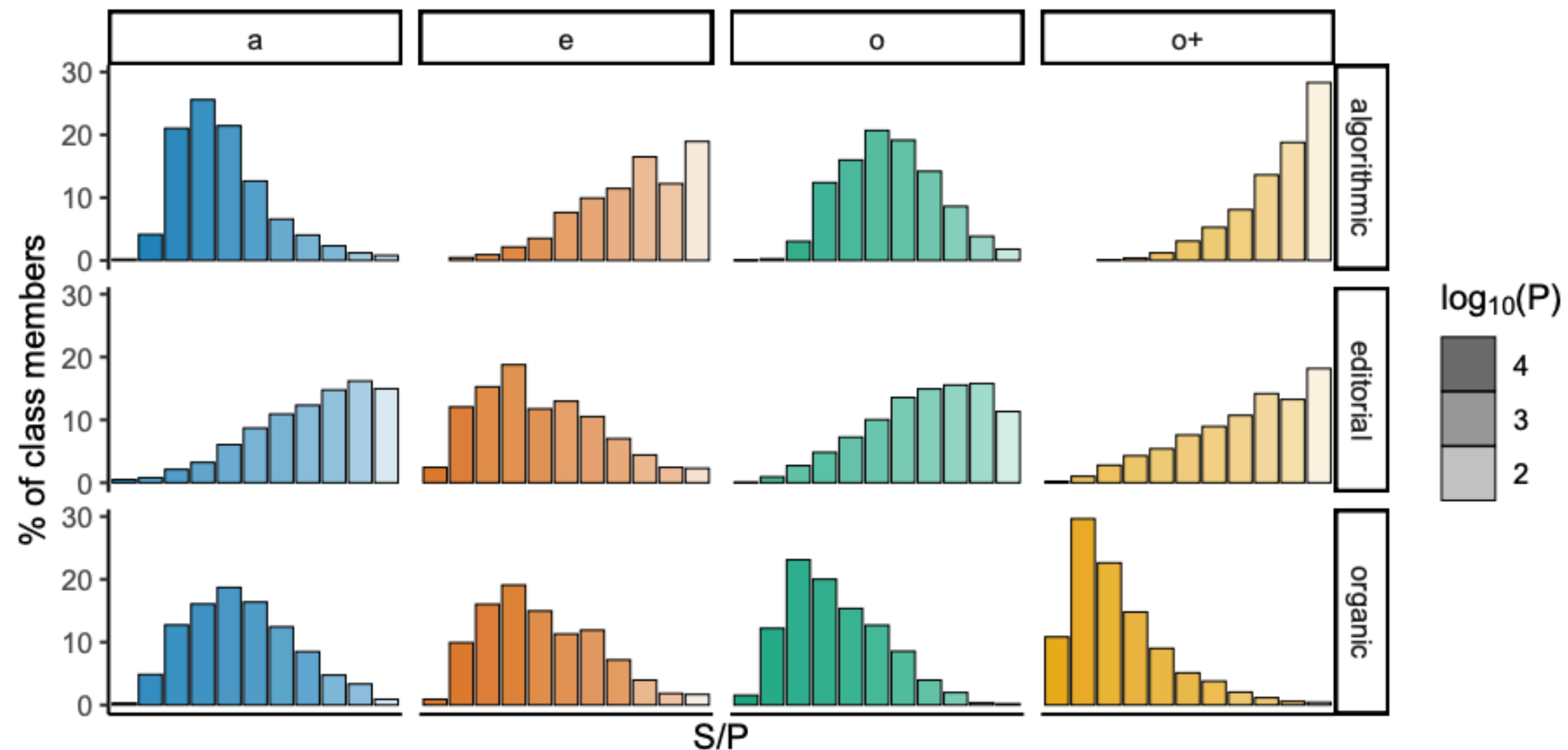


Fig. 3. Breakdown of dispersion values for each user class and for each access mode. Histograms are further binned by deciles of increasing  $S/P$  values (from 0 to 1 from left to right) and indicate how many users of each class (**a**, **e**, **o** and **o+**) exhibit which dispersion value for a certain access mode (algorithmic, editorial or organic). Average activity values for users of each decile bar are further indicated by a grayscale, where darkest shades correspond to highest  $P$  values.

- Dispersion generally lower as a function of activity
- Dispersion lower for the main access mode, especially as activity increases
- Generally lower for organic access, especially for **o+** users
- **o** users still appear to have lower dispersion in the algorithmic access mode



# Two dimensions of diversity

## Artist popularity

bin	# artists	<i>access mode</i>			
		algorithmic	editorial	organic	
$\nu_1$	73	9%	<b>8%</b>	<b>83%</b>	100%
$\nu_2$	319	16%	<b>8%</b>	76%	100%
$\nu_3$	1462	<b>18%</b>	5%	77%	100%
$\nu_4$	166869	15%	5%	80%	100%
all	164955	14%	7%	79%	100%

Table 1: Proportion of access modes for each nicheness bin (preferred bins for each access mode are marked in bold).

- *algorithmic*: less popular
- *editorial*: more popular
- *organic*: U-curve

# Two dimensions of diversity

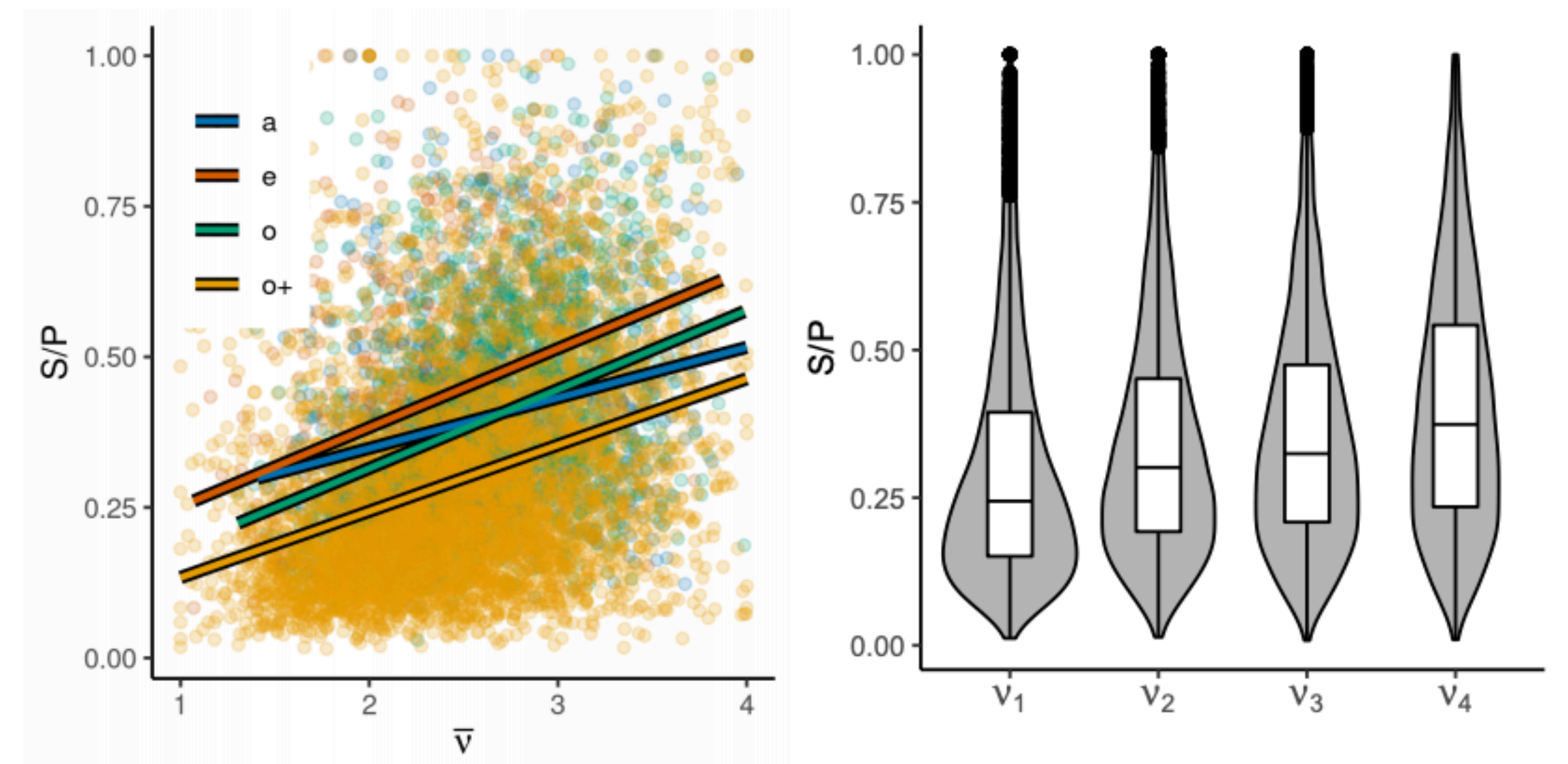
## Artist popularity and dispersion

- dispersion increases with less popular content on average, for all user types

bin	# artists	access mode			
		algorithmic	editorial	organic	
$\nu_1$	73	9%	<b>8%</b>	<b>83%</b>	100%
$\nu_2$	319	16%	<b>8%</b>	76%	100%
$\nu_3$	1462	<b>18%</b>	5%	77%	100%
$\nu_4$	166869	15%	5%	80%	100%
all	164955	14%	7%	79%	100%

Table 1: Proportion of access modes for each nicheness bin (preferred bins for each access mode are marked in bold).

- *algorithmic*: less popular
- *editorial*: more popular
- *organic*: U-curve



# Two dimensions of diversity

## Artist popularity and access modes

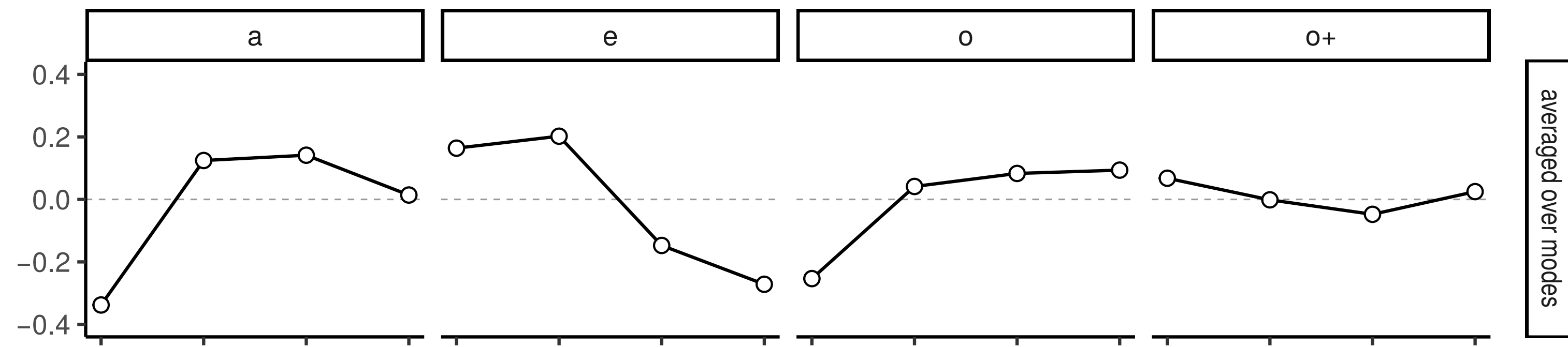
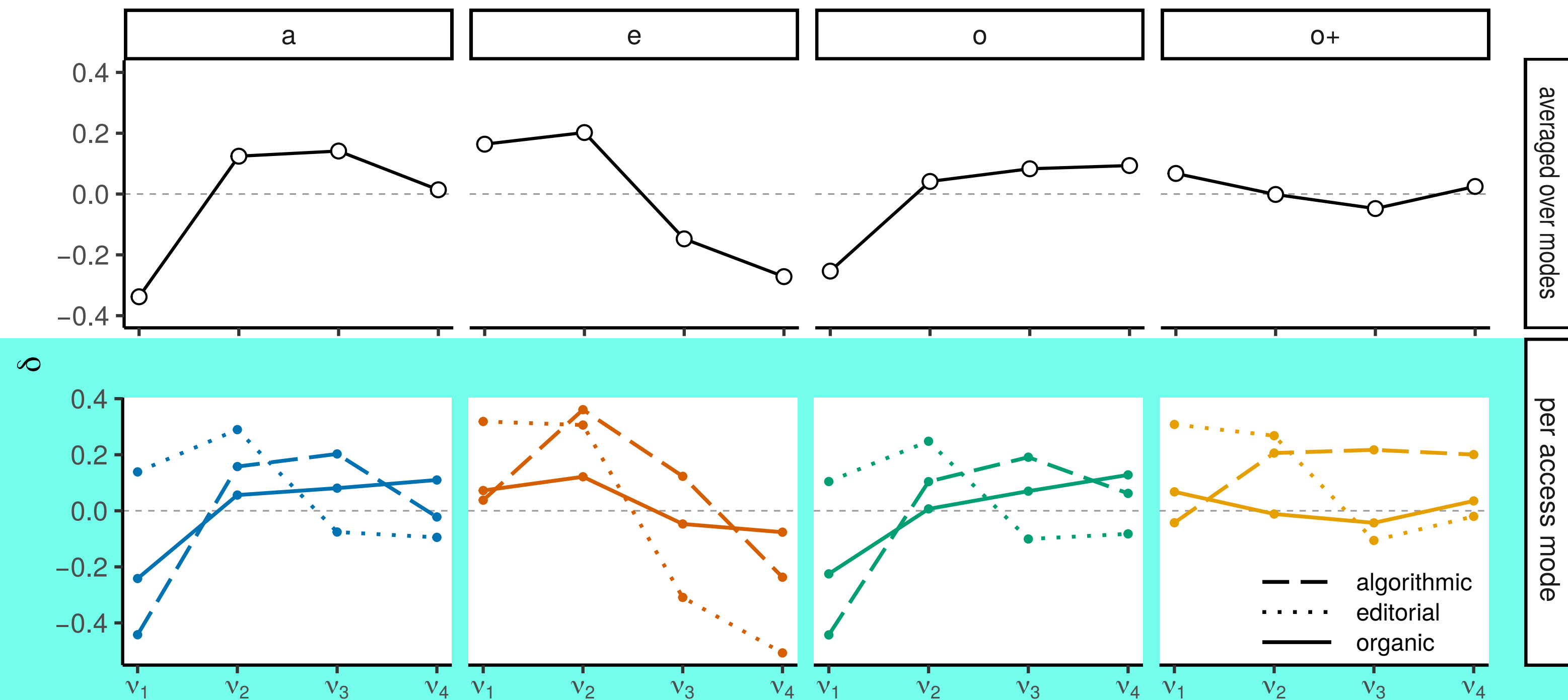


Fig. 5. Relative consumption of content from each popularity bin, average log-ratio with respect to a uniformly random baseline for each bin (0 corresponds to no deviation, the x-axis is ordered from  $v_1$  to  $v_4$  i.e., for musical content from more to less popular artists). *Top*: average over all plays. *Bottom*: breakdown by access mode.

# Two dimensions of diversity

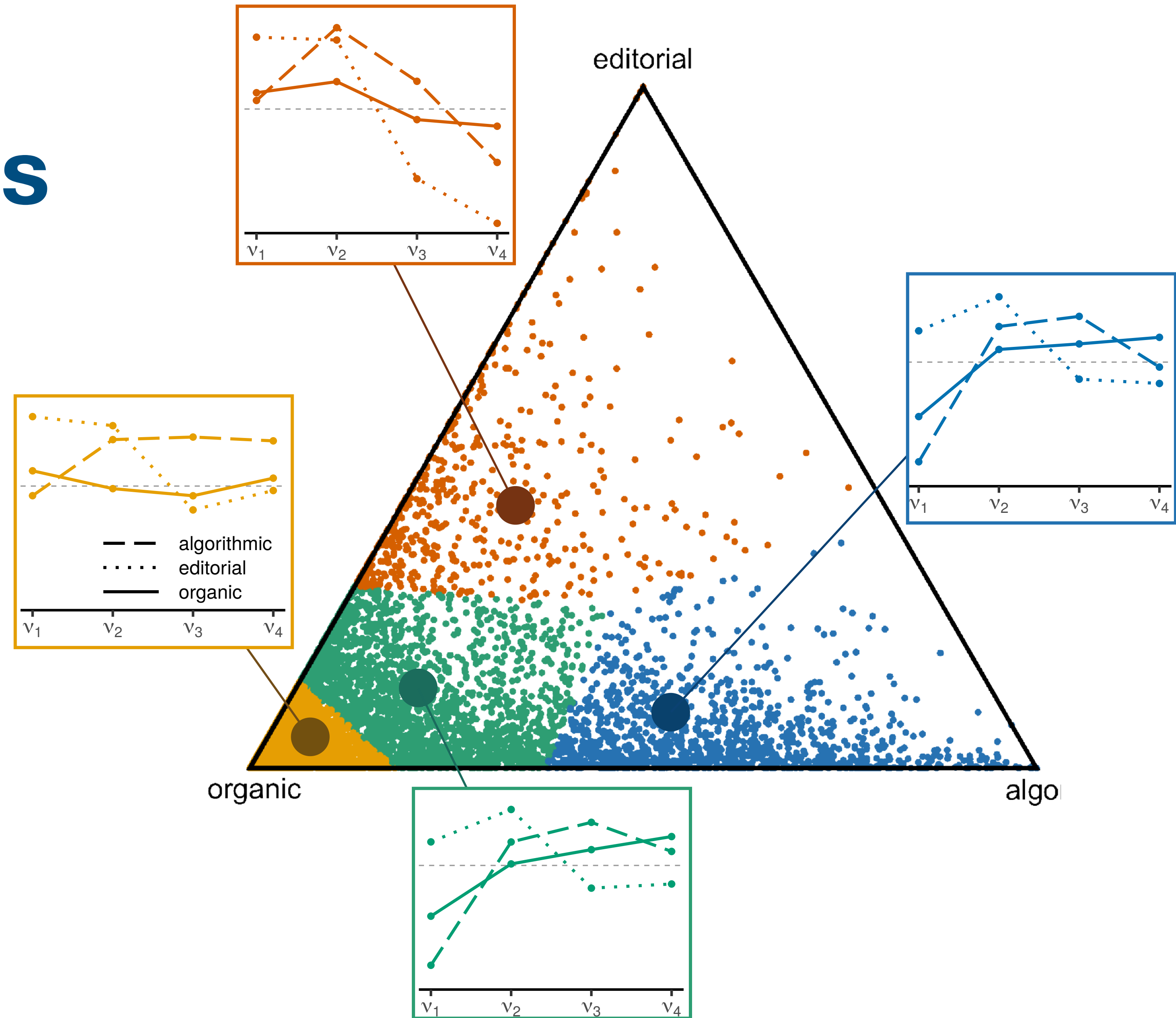
## Artist popularity and access modes



- *algorithmic*: towards niche, except for **e** users
- *editorial*: monotonous
- *organic*: it depends

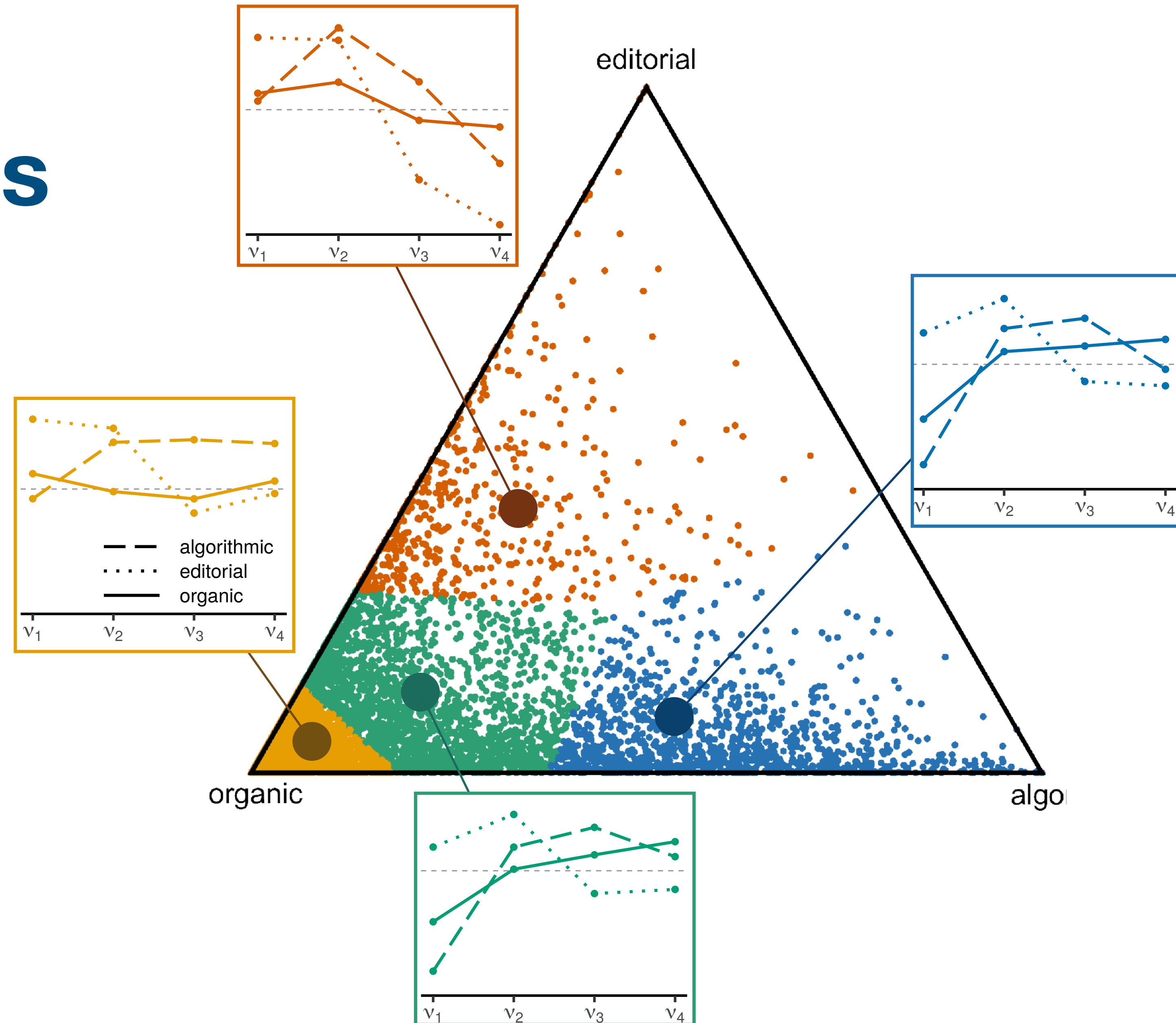
Fig. 5. Relative consumption of content from each popularity bin, average log-ratio with respect to a uniformly random baseline for each bin (0 corresponds to no deviation, the x-axis is ordered from  $v_1$  to  $v_4$  i.e., for musical content from more to less popular artists). *Top*: average over all plays. *Bottom*: breakdown by access mode.

# Filter niches rather than bubbles



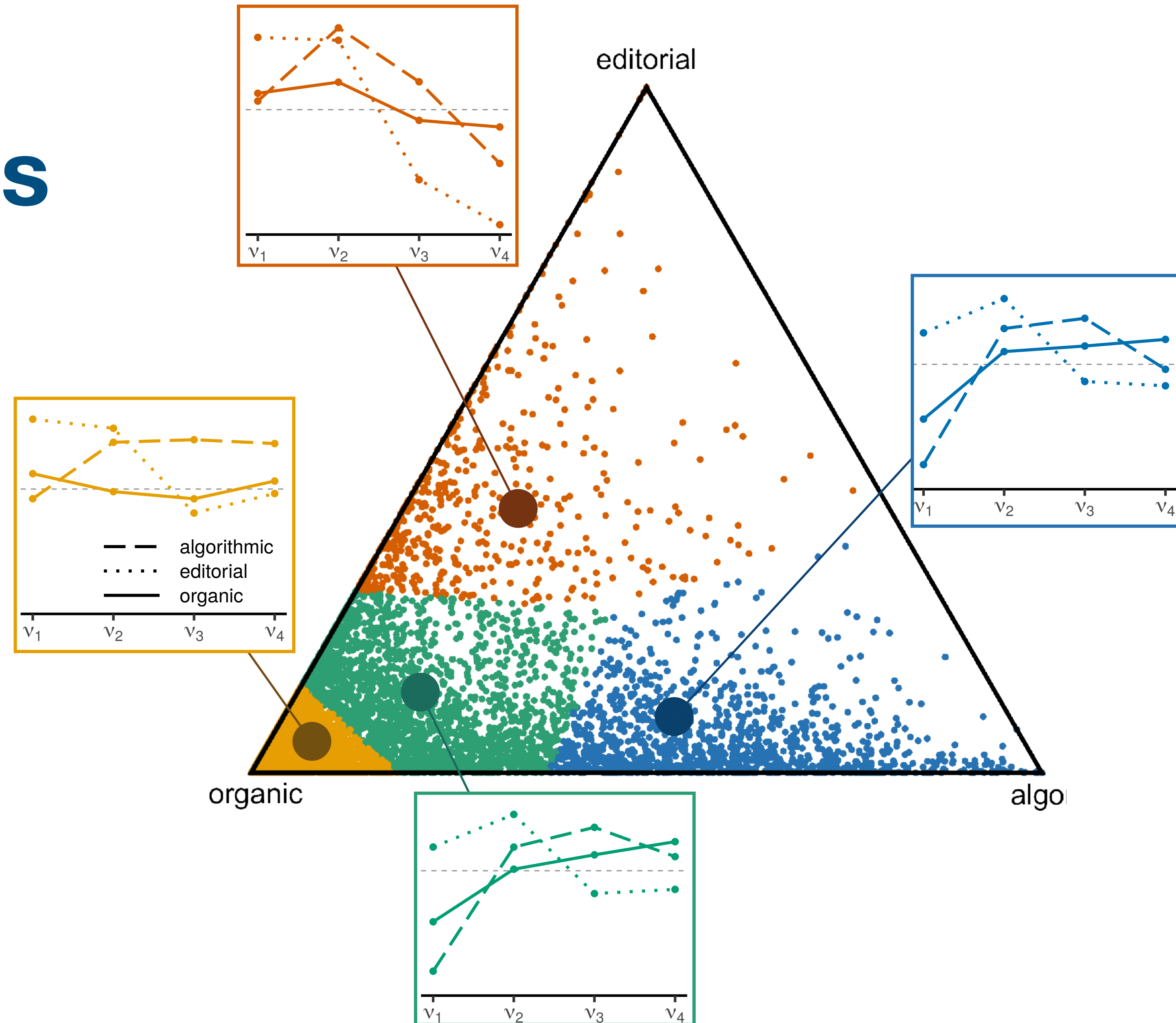
# Filter niches rather than bubbles

- **Radio playlists** generally **less dispersed than user types** (more exploitation), while they are wide variations in terms of **artist popularity**



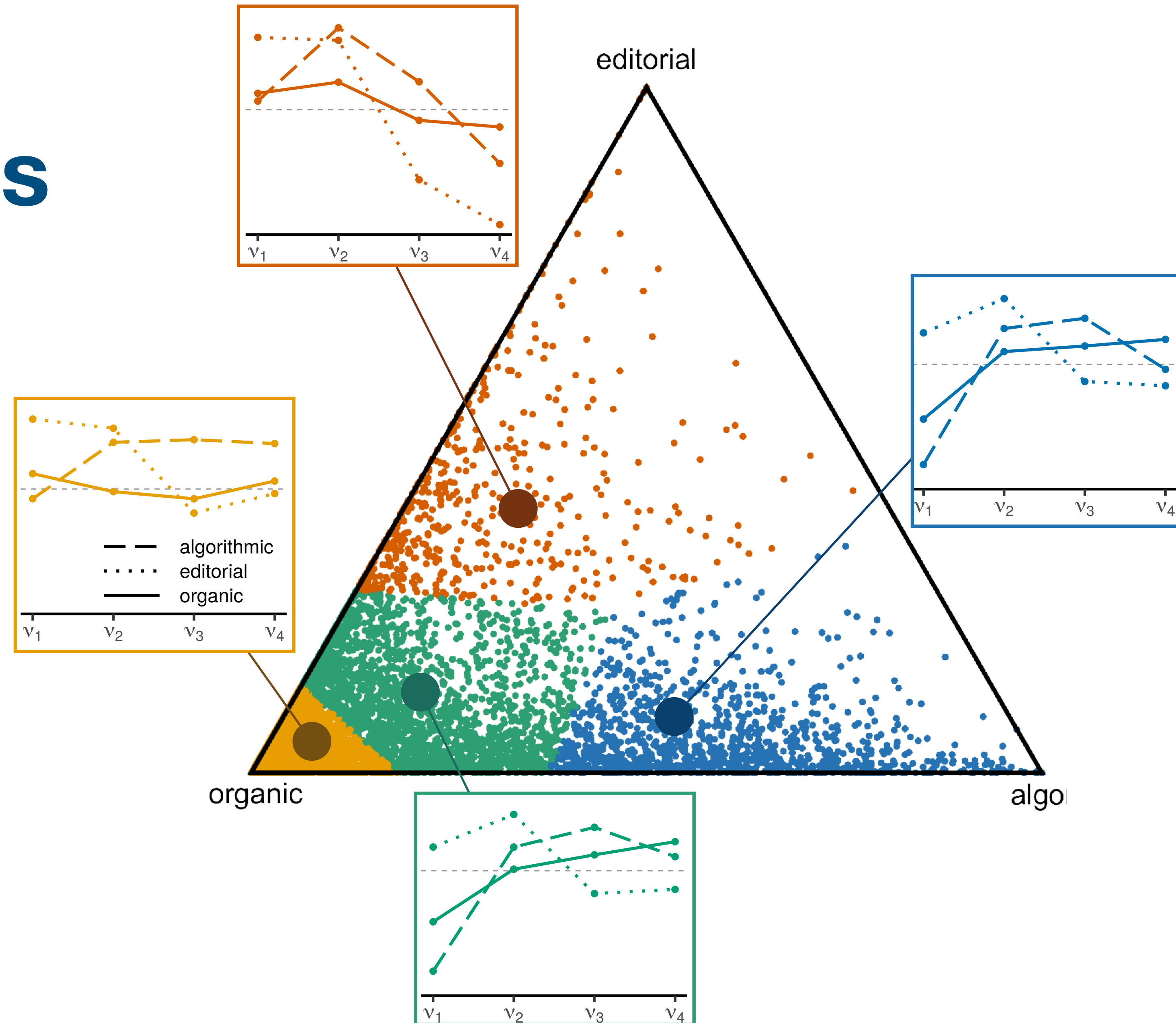
# Filter niches rather than bubbles

- **Radio playlists** generally **less dispersed than user types** (more exploitation), while they are wide variations in terms of **artist popularity**
- **Algorithmic access modes** generally **avoid popular content**, but not for editorial users



# Filter niches rather than bubbles

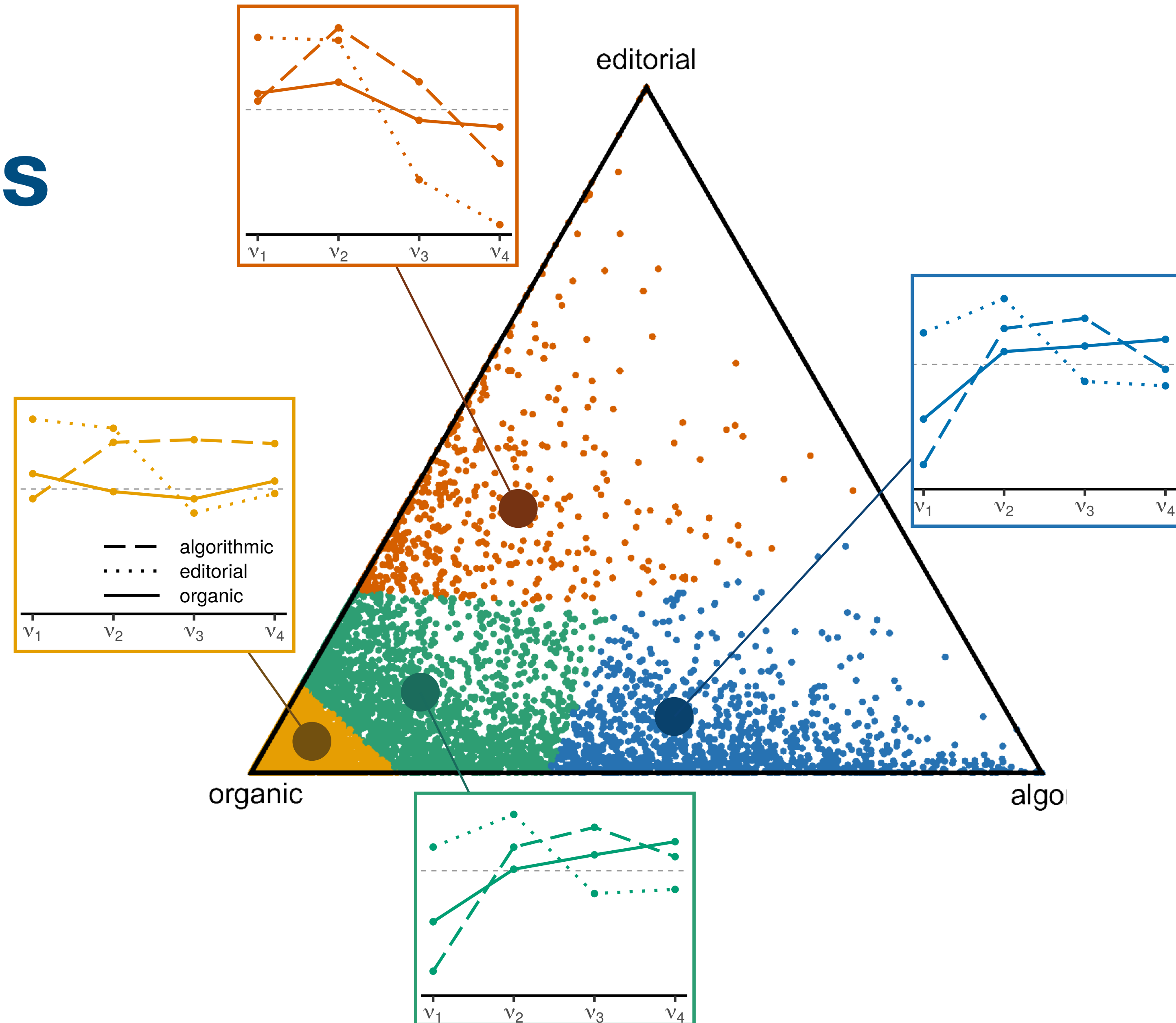
- **Radio playlists** generally **less dispersed than user types** (more exploitation), while they are wide variations in terms of **artist popularity**
- **Algorithmic access modes** generally **avoid popular content**, but not for editorial users
- **Organic users** focus most on **less popular content** — yet, these users precisely exhibit a relatively balanced diet of platform affordances,
  - *hypothesis*: these are the most “expert” users who best exploit platform affordances





# Filter niches rather than bubbles

- **Radio playlists** generally **less dispersed than user types** (more exploitation), while they are wide variations in terms of **artist popularity**
- **Algorithmic access modes** generally **avoid popular content**, but not for editorial users
- **Organic users** focus most on **less popular content** — yet, these users precisely exhibit a relatively balanced diet of platform affordances,
  - *hypothesis*: these are the most “expert” users who best exploit platform affordances
- **Editorial access**, by contrast, even more so for **editorial users**, fulfills a role traditionally ascribed to radios in terms of mainstream exploration, yet with more exploration / higher dispersion



# **No blanket answer to the impact of recommendation**

but more clear-cut answers  
if we distinguish beforehand  
persona, user types, affordance use

# **No blanket answer to the impact of recommendation**

but more clear-cut answers  
if we distinguish beforehand  
persona, user types, affordance use

—  
thanks!