6^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2020 75 Amherst St. Cambridge, MA USA

The Effects of Disaggregated Factors of YouTube Recommendations in Diversity

Pedro Ramaciotti Morales[†], Bilel Benbouzid[‡], Emma Gauthier[‡], Alexis Perrier[‡]

[†] médialab, Sciences Po [‡] LISIS UMR 9003 CNRS, UPEM, Université Gustave Eiffel

Keywords: YouTube, algorithmic recommendations, diversity of recommendations, agent-based models, consequences of recommender systems.

We are in the midsts of an active debate about the effects that algorithms have on the digital public space. As they filter, select, and influence what comes to our attention, the effects of algorithms are political, social, cultural, as well of economic. This debate has also come to encompass important phenomena such as radicalization, and the so-called *echo chambers* and *filter bubbles*. Most interactions between users and with content take place in a few platforms that account for a wealth of scientific works. YouTube, an important content provider (the second most popular site as of 2019), has been comparatively less investigated.

In this study we set out to identify how different relations between YouTube channels may affect the diversity of content watched by users. We propose agent-based simulations to explore the consequences of privileging some of these relations (generated by explicit human interactions or computed algorithmically) in recommendations. We consider a set of YouTube channels from a single country –in our case France– to better account for its representativeness within an informational ecosystem. This set includes the channels of political parties and personalities, news outlets, *youtubers*, independent media, companies, and social movements that have the most subscribers, reaching nearly 1400 channels. We then proceed to establish relations between channels by consulting YouTube's API and by web-crawling during Jan-May 2019. Four relations are considered: two of them (*human*) depending on explicit user actions, and two of them (*algorithmic*) computed algorithmically by YouTube. These relations are represented in four networks of channels: **a) Featured network** (*human*), in which channel A is linked to channel B if the administrator of channel A decided to include channel B in the *featured* section of channel A, **b) CoCommented network** (*human*), in which channel A is linked to channel B if there is a significant number of users commenting videos of both channels, **c) Recommended network** (*algorithmic*), in which channel A is linked to a channel B if a video from B was recommended in the lateral bar in a video from channel A, **d) Related network** (*algorithmic*), in which channel A is linked to a channel B if it appears in the *related* section in channel B (this amounts to channel recommendation and was discontinued in May 2019). In order to have a classification over which to measure diversity of video consumption in simulations, these channels were then classified manually into one of 12 categories: political comment, complotism, militant groups, investigative journalism, entertainment, grassroots movements, traditional media, local media, theme-specialized media, conferences/interviews, public relations, and political vulgarization.

To study the effects of following hypothetical recommendations computed with each one of these four networks, we propose an agent-based simulation [2]. On each network we consider random

walks of different lengths (to simulate different user engagement levels), and for each length we perform 10.000 simulations, storing the categories of channels visited during the walks. Then, for each network, and each length of walk we compute the diversity of categories visited, using Perplexity [3] (a measure of uncertainty of visited channel category, closely related to Shannon entropy) to account for both the number of categories and their apportionment. Crucially, we distinguish collective diversity (computed with the categories visited by all 10.000 agents) from the individual diversities (computed for each agent) to separate the system-wide effects from those on individual watching experience.

One possible organization of the factors that influence actual recommendations in YouTube can be organized in two groups: users' characteristics (personalized, or user-centric) and videos' characteristics (video-centric, or *cold-start*) [1]. Our agent-based simulations provide insights about the latter. Simulations using **Featured** and **CoCommented networks** show what the effects would be of weighting more of the traces of human actions underlying these networks in the computation of the video-centric part of recommendations. In comparison, simulations obtained using the **Related network** use only algorithmic similarity computed by YouTube, and **Recommended network** serves as baseline representing the effects that actual recommendations have on diversity.

In this work, we analyze the effects that these disaggregated factors have to show that including more traces of human actions in cold-start recommendations could –depending on the mixing of user- and video-centric factors– improve the diversity of contents watched by a collective of people, but not necessarily the diversity experienced by individuals. We also suggest that both collective and individual diversities resulting from actual recommendations are higher that those that would result from recommending based purely in YouTube's similarity. In our study we also address the algorithmic visibility of the different categories of channels that result from giving more weight to factors here identified. Finally, we discuss how network clusterizations compare between these four networks when contrasted with our expert categorization to comment on the relation between the so-called *filter bubbles* and contents offered by channels.

References

[1] P. Covington, J. Adams, and E. Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, 2016.

[2] A. Dash, A. Mukherjee, and S. Ghosh. A network-centric framework for auditing recommendation systems. In *2019-IEEE Conference on Computer Communications*. IEEE, 2019.

[3] P. Ramaciotti Morales, L. Tabourier, S. Ung, and C. Prieur. Role of the website structure in the diversity of browsing behaviors. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 2019.

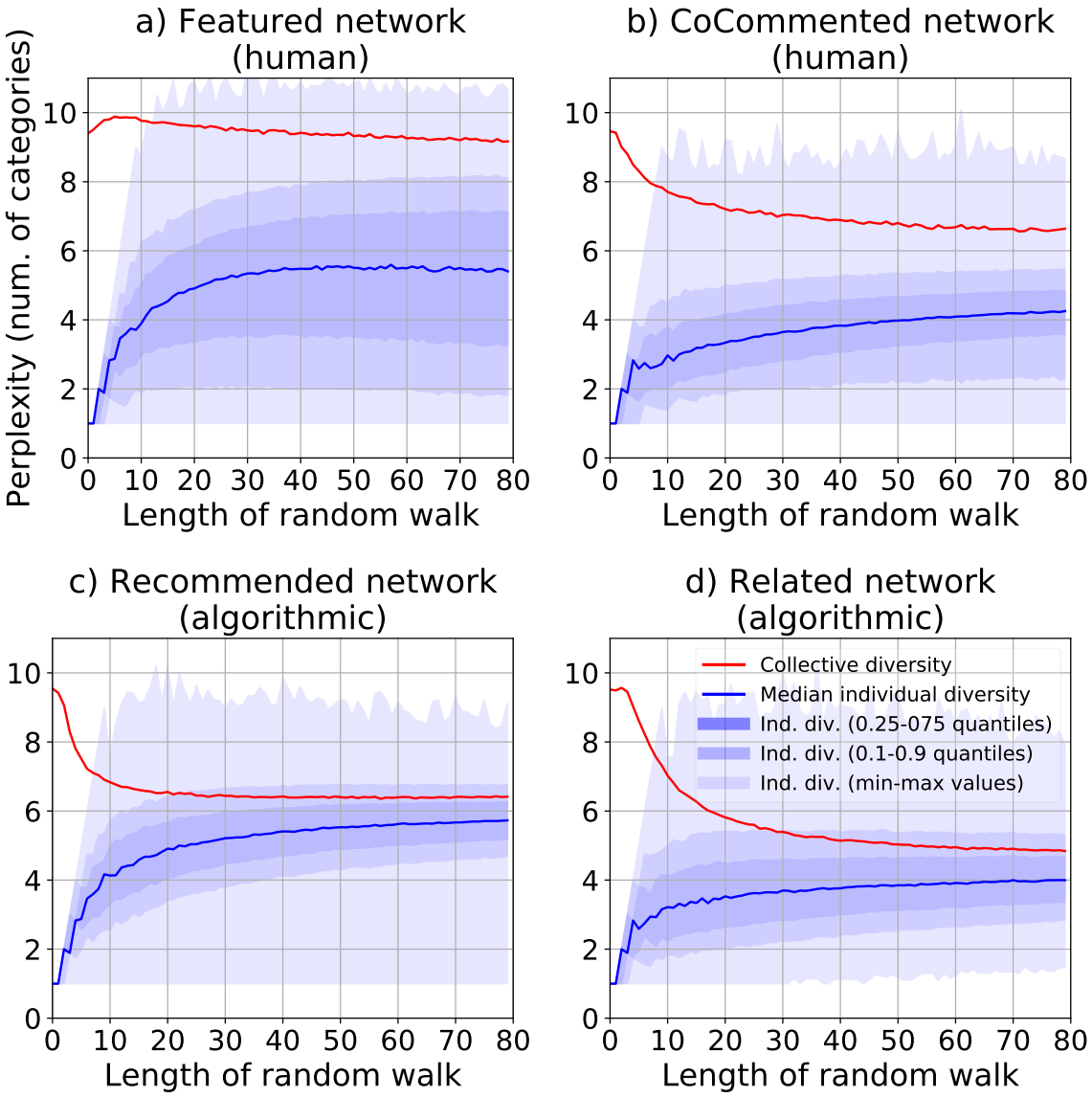

Figure 1: Collective and individual category diversity (measured with Perplexity) of simulated agents (10.000 per length of browsing) randomly browsing YouTube channels following one of four relations between channels encoded in four networks. Some relations lead to faster loss of collective diversity as users adopt more recommendations. Weighting more human interaction factors in recommendations can improve collective diversity but not necessarily individual diversity.