Hegemony in Social Media and the effect of recommendations

Ana-Andreea Stoica
Columbia University
New York, USA
astoica@cs.columbia.edu

Augustin Chaintreau
Columbia University
New York, USA
augustin@cs.columbia.edu

ABSTRACT

As today's media landscape is carved by social media endorsements and built on automated recommendations, both of these are often criticized for inducing vicious dynamics, such as the filter bubble effect, echo chamber, or polarization. We introduce a new model featuring a mild version of homophily and two well-known popularity dynamics. These broadly reproduce the organic activity and the algorithmic filtering, respectively, of which the latter is now commonplace within social media or other online services. Surprisingly, we show this is all that is needed to create hegemony: a single viewpoint (or side) not only receives undue attention, but it also captures all the attention given to "top trending" items.

CCS CONCEPTS

• **Theory of computation → Network flows**; **Social networks**; *Random walks and Markov chains*; • **General and reference →** *Empirical studies.*

KEYWORDS

fairness, recommendation, homophily, random walks

ACM Reference Format:
Ana-Andreea Stoica and Augustin Chaintreau. 2019. Hegemony in Social Media and the effect of recommendations. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3308560.3317589

1 INTRODUCTION

Item recommendation is instrumental for an online marketplace to develop, for a content provider to grow its audience, and for a social media participant to navigate the deluge of information produced in real time. This is the primary form of personalization, and its effects on growth are beyond doubt: automated recommendations are the origin for 35% of sales on `amazon.com`, 50% of initial messages sent on `match.com`, and 80% of streamed hours on `netflix.com`. Their side effects have never been so scrutinized: concerns range from recommendation algorithms isolating information seekers from differing viewpoints (*filter bubble*), radicalizing citizens' attitude towards controversial issues (*polarization*), or enabling malicious actors to manipulate information spreading (*fake news*). It is increasingly common that opting out of recommendations is simply not an option: accessing news on Facebook or searching for partners on Tinder. Still, many users express a concern that the results they see may not be representative.

We describe a model to study whether diverse viewpoints are rewarded with attention from social media and recommender systems. The alternative would be a form of *hegemony* where, although multiple viewpoints exist on a controversial topic, only one of them effectively reaches a large audience. Social media are often criticized for their alleged effects on opinion formation. They claim a form of neutrality: that they add no supplementary bias to the relative proportion of those supporting different viewpoints. In particular, they vehemently deny that item recommendations or the selection of *top trending* items leverages any feature or design that could favor a particular ideology. That claim seems *a priori* credible since all media know that news readers massively oppose partisanship and ideological bias, with up to 75% claiming it is "never acceptable"[13]. But the paradox is, despite all of the above, there is a considerable distrust in the media, especially social media and algorithmic recommendations, to represent each viewpoint fairly. This lack of trust has dangerous consequences and often accompanies accusations of dishonesty. Here we explore an alternative explanation where this chasm comes from a natural hegemony emerging from root causes commonly found in most online services.

We assume no pre-existing social network connecting readers that could affect their content exposure[1]. We assume a mild form of homophily, where users tend to connect more to content they agree with. Under these (idealized) conditions, we answer the following questions: does the organic dynamics of posting and reposting in social media offer an outlet for all viewpoints to be shown? Can one viewpoint become hegemonic? And is a simple recommendation algorithm such "people who like ... also like" a deciding or a contributing factor?

- We first analyze a simple content-linking model. Readers implicitly belong to one of two groups (e.g. liberal or conservative, female or male). We prove that social media reposting *systematically under-represent* a minority viewpoint, the majority viewpoint gradually becoming the *only* visible option. (Section 4)
- We extend our model to more complex forms of homophily, and show it accurately reproduces an observed gap in the amount of likes received by items on a leading social media (i.e. pictures on Instagram). Interestingly, it shows that hegemony may be created not by majority alone but also

This material is based upon work supported by the National Science Foundation under Grant No.1254035 and 1514437.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3317589

[1]Note that exposure to ideologically diverse news on Facebook was studied empirically and shown to present a bias that was primarily attributed to that effect [3].

by minority when it exhibits exacerbated homophily. (Section 5).
- Finally, we study the effect of recommendations. We find that items recommendations *always accelerate* the hegemonic dynamics we identified, exacerbating the misrepresentation of a minority viewpoint. (Section 6).

2 RELATED WORK

In spite of a large body of work studying the effect of recommendation systems, we found no model that reproduces the simple underlying dynamics of a hegemonic nature we observe. A critical novelty of our approach is that we study the effect of recommendations within the natural organic growth of the social media. Empirically, recommendations were shown to influence users in creating blockbusters [14], to expose users to a slightly narrowing set of items [10], or on the contrary to widen their interests [8]. A model of recommendation based on market share was shown to lead to higher concentration overall, while individuals enjoy higher product diversity [6]. We study dynamics that more generally emerge from the social endorsements created with time and exploited to build recommendations.

Filter bubble [12], balkanization [16] or polarization [4, 17] dynamics have been suspected to play an increasing role online for almost a decade. Polarization and echo-chambers were identified before in pre-existing structures among blogs or in the social network connecting individuals [1, 3], with important consequences on information exposure. However, recent evidence relativizes those hypotheses [5] as it finds only a modest segregation or ideological distance between news readers. Our analysis points to a different hypothesis: first, our model has no graph connecting participants that would limit or direct their information exposure; second, we do not predict that two groups of readers online end up at polar opposites, but rather explain why dissenting and minority views are suffering from a lack of representation. We note that, in contrast with previous work, these dynamics emerge in conditions that are almost always met online, which may explain why individuals holding viewpoints that are not shared by the majority found that items recommended in social networks rarely reflect their views. Note that a lack of equal access to representation in today's online services (including Twitter [11], TaskRabbit [7]) was empirically observed. This adds to a large body of work warning us on the unintended consequences of algorithms run on Big Data, with discrimination taking place in seemingly neutral settings. Since the relative representation of various viewpoints is driven by participants' homophilic behavior rather than an algorithm itself, it remains difficult to determine if that constitutes a case of active discrimination. It is however undeniably generating a disparate impact, and the fact that an algorithm reinforces this effect makes a strong case that more work is needed to prevent this harm.

The result most relevant to ours is a recent proof that rich-get-richer dynamics combined with homophily naturally exacerbate the advantage of a majority group [2]. That was recently extended to understand how social recommender system contributes to create a glass-ceiling [15], an effect that is equivalent to hegemony in a unipartite graph.

3 MODELS AND DEFINITIONS

We introduce two models - one describing the organic growth of a network and one in which recommendations are added. Both models extend previous work on network growth [2, 9, 15] by separating nodes into *individuals* and *items* in a bipartite manner.

Participants interact with items on social media by either posting their own content, or by reposting: liking, retweeting, or sharing. For simplicity we assume participants have one color (Red or Blue) that represents their opinion. Items also have colors, inherited from the person who posted it originally. Colors may represent for instance liberal and conservative views on an issue, but the model is general and can be applied to goods bought online from different brands, job ads, etc.

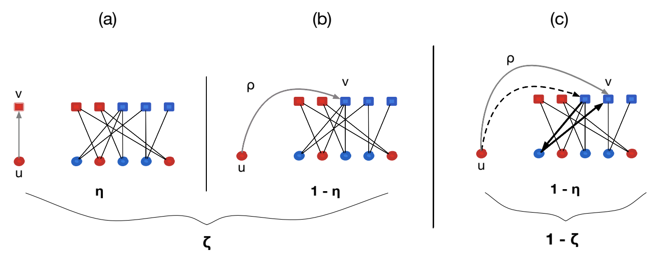

Figure 1: Illustration of the organic growth model $\mathcal{M}$ and the recommendation model $\mathcal{M} - REC$, where individuals are circles and items are squares.

3.1 Organic growth model $\mathcal{M}$

We introduce a model of growth $\mathcal{M}$, inspired by the biased preferential attachment model [2]:

- *Minority-majority partition*: for $0 \leq r \leq 1/2$, at time t, an individual enters the network and receives label R with probability r and B with probability $1 - r$.
- *Item creation*: with probability η, the new individual creates a new item that has the same color as herself, and connects to it with probability 1. This step allows the set of items to grow, capturing the dynamics or creativity in a social network - for example, an author of a certain political view writes and publishes an article that reflects her view, or composes a tweet (Figure 1 (a)).
- *Reposting*: with probability $1 - \eta$, the new individual reposts/retweets/likes one existing item. Intuitively, items that have been reposted more are more likely to be noticed and hence chosen to be reposted. We thus assume the individual chooses one item to connect to with probability proportional to that item's degree, $\mathbb{P}(v \text{ is chosen}) = \delta_t(v) / \sum_{u \in V_t} \delta_t(u)$, where $\delta_t(x)$ denotes the degree of item x at time t and V_t is the set of items in the graph at time t (Figure 1 (b)).
- *Homophily*: if the individual has a different label than the item it chooses to connect to in the previous step, the connection is accepted with probability ρ (and rejected with probability $1 - \rho$). If rejected, the process is repeated until an edge is formed. The homophily parameter $0 \leq \rho \leq 1$ captures a person's openness to repost content that does

not match her views; $\rho = 1$ denotes an equanimous person which is not affected by that item's viewpoint in any way. We typically assume that $\rho < 1$. Intuitively, most individuals have a tendency to read broadly across the political spectrum, but are more selective in the ideology of items they repost [1].

3.2 Recommendation model $\mathcal{M} - REC$

We extend the model $\mathcal{M}$ by adding recommendations based on individuals with similar behavior. This captures the way in which, for example, an individual would see a link posted by a friend with whom he shared previous items.

We denote this new model by $\mathcal{M} - REC$ and define it below:

- *Minority-majority partition*: for $0 \leq r \leq 1/2$, at time t, an individual enters the network and receives label R with probability r and B with probability $1 - r$.
- with probability ζ, it connects to an item according to model $\mathcal{M}$ (Figure 1 (a) and (b)).
- with probability $1 - \zeta$, it chooses an item according to preferential attachment, but it does not connect to it; instead, it follows a 3-step random walk from that item by finding someone who was connected to that item and choosing another item linked to that individual. If the end item has the same color, it connects to it, and if not, it accepts it with probability ρ and rejects with probability $1 - \rho$, and repeats the process until an edge is formed (Figure 1 (c)).

Note: We assume each node has an outdegree of $m > 1$ and exclude it from the analysis it will just create a m-factor multiplication.

We adapt the analysis of the glass ceiling effect in the unipartite graph from [2], where two communities exhibit a glass-ceiling effect if their degree distributions follow a power law with different coefficients. Intuitively, this shows that the gap between them would increase as we are looking at higher-ranked individuals.

4 HEGEMONY IN ORGANIC GROWTH

We first present a theoretical result, where the item-based preferential attachment model $\mathcal{M}$ exhibits a hegemony effect, effectively suppressing the representation of the items pertaining to the minority viewpoint in top results.

A similar proof exists for a more constrained model on unipartite graph [15]. An important step is to obtain the fraction of connections to red-items as a fixed point equation. We find that, provided the fixed point equation is adapted to this new case, similar arguments can be used for a bipartite graph.

Rate at which red items receive connections: Denote by $u_t(R)$ as the sum of degrees of all red items after t steps. Since an edge is added at each step, it makes sense to introduce $u_t(R) = t \cdot \alpha_t$ and $u_t(B) = t \cdot (1 - \alpha_t)$, where α_t is equivalently defined as the sum of degree of the red items divided by the total sum of their degrees, or simply the fraction of edges created towards R items.

Lemma 4.1. *$\mathbb{E}[\alpha_{t+1}|\alpha_t] = \alpha_t + \frac{F(\alpha_t)-\alpha_t}{t+1}$, where*

$$F(x) = \eta \cdot r + (1 - \eta) \cdot \left(\frac{rx}{x + \rho \cdot (1 - x)} + \frac{x \cdot \rho \cdot (1 - r)}{x\rho + 1 - x} \right).$$

We observe similar properties of the F function than the other function described for unipartite graph in Avin et al [2]. For space considerations, we omit the complete argument of this proof.

Lemma 4.2. *For $r < 1/2$ and $0 < \rho < 1$, we have*

(1) *F is monotonically increasing.*
(2) *F has exactly one fixed point in $[0, 1]$, denoted by α^*.*
(3) *$x < \alpha^* \implies x < F(x) < \alpha^*$; $x > \alpha^* \implies \alpha^* < F(x) < x$.*
(4) *$\alpha^* < r$.*

As Lemma 4.1 illustrates, F is the expected rate of growth of red edges from time t to time $t + 1$. Thus, its fixed point α can be interpreted as the limit of the fraction of red edges when the network size grows to infinity. Such a limit exists since the fixed point is present and unique in $[0, 1]$.

The last property, $\alpha^* < r$, already shows that the expected degree of a red item (α^*/r) is lower than that of a random item (1). However, this does not necessarily imply a hegemonic effect. We now show that the degree distribution of the two types of items follow power laws with different coefficients, which implies hegemony as defined above.

Degree distribution: Denote by $m_{k,t}(B)$ and $m_{k,t}(R)$ the number of blue and red items of degree k at time t, and define

$$M_k(x) = \lim_{t\to\infty} \frac{\mathbb{E}(m_{k,t}(x))}{t}, \text{ for } x \in \{R, B\}. \tag{1}$$

Theorem 4.3. *In the case of organic growth according to model $\mathcal{M}$, the degree distributions of the red and blue items follow power laws, i.e. $M_k(R) \sim k^{-\beta(R)}$ and $M_k(B) \sim k^{-\beta(B)}$.*

We omit the proof, noting that in order to show this theorem we derive a recurrence relation for $m_{k,t}(R)$, $m_{k,t}(B)$ and deduce that degrees quickly converge to a power law as t gets large with coefficients $\beta(R) = 1 + \frac{1}{(1-\eta)C_R}$ and $\beta(B) = 1 + \frac{1}{(1-\eta)C_B}$, where

$$\begin{aligned} C_B &= \frac{\rho \cdot r}{\alpha + \rho(1 - \alpha)} + \frac{1 - r}{\rho \cdot \alpha + 1 - \alpha}, \\ C_R &= \frac{r}{\alpha + \rho(1 - \alpha)} + \frac{\rho \cdot (1 - r)}{\rho \cdot \alpha + 1 - \alpha}. \end{aligned} \tag{2}$$

Using the properties of the function F and these results, we find the following corollary:

Corollary 4.4. *For $r \in [0, 1/2]$ and $\rho \in [0, 1]$, we have*

$$\beta(R) > 2 > \beta(B). \tag{3}$$

Gap coefficient and normalized expected degree of a red node: Our work enables us to make a qualitative analysis of how the expected degree of a red node changes when homophily varies between 0 and 1. Firstly, notice that since α represents the expected fraction of edges towards red items, α/r is the normalized expected degree of a red note as the fraction of red items is r.

Furthermore, we are able to compute the gap between the rate at which the number of items above degree k decreases as a function of k. To do this, we are interested in computing the fraction of items with degree at least k: $\mathbb{P}(u \in R|deg(u) > k)$. We make use of the

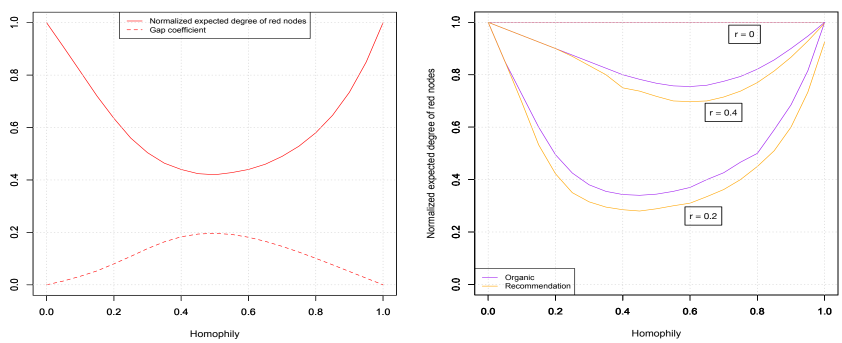

Figure 2: Effect of homophily parameter ρ as x-axis on the normalized expected degree of red nodes and the gap coefficient for $r = 0.25$ and $\eta = 0.1$ (a) and on the normalized expected degree of red nodes, for $r = 0, 0.2, 0.4, \eta = 0.1, \zeta = 0.9$ (b).

fact that the degree distribution of items of degree exactly k follows a power law with known coefficients:

$$\begin{aligned}\mathbb{P}(u \in R, deg(u) > k) &\sim C \cdot k^{-(1+\beta(R))} = C \cdot k^{-(2+1/(1-\eta)C_R)},\\ \mathbb{P}(u \in B, deg(u) > k) &\sim C \cdot k^{-(1+\beta(B))} = C' \cdot k^{-(2+1/(1-\eta)C_B)},\end{aligned} \quad (4)$$

for constants C, C'. Then, for some constants C_1, C_2,

$$\begin{aligned}\mathbb{P}(u \in R | deg(u) > k) &= \frac{\mathbb{P}(u \in R, deg(u) > k)}{\mathbb{P}(deg(u) > k)}\\ &= \frac{r \cdot C_1 \cdot k^{-(2+1/(1-\eta)C_R)}}{r \cdot C_1 \cdot k^{-(2+1/(1\eta)C_R)} + (1-r) \cdot C_2 \cdot k^{-(2+1/(1-\eta)C_B)}} \Rightarrow \\ \mathbb{P}(u \in R | deg(u) > k) &\approx k^{-\frac{1}{1-\eta}\left(\frac{1}{C_R} - \frac{1}{C_B}\right)}.\end{aligned} \quad (5)$$

We call $\frac{1}{C_R} - \frac{1}{C_B}$ the *gap coefficient*. Figure 2 thus shows how the normalized expected degree of a red node (straight red line) and the gap coefficient reach a minimum and a maximum, respectively, for an intermediate value of ρ. Thus, there is a critical value of homophily for which the red items lose the most of their edges.

An example of the effect this can have in the real world is to consider Youtube videos and analyze their revenue as a function of their number of likes. Given Figure 2, one could expect that homophily can reduce the revenue for videos pertaining to a minority view by up to 60% of what they should receive in an equal setting.

5 VALIDATION AND EXTENSIONS

5.1 Empirical study of Hegemony

We validate our results using data collected from Instagram to study the effect of homophily among gender on the popularity of items. To the best of our knowledge, no prior studies of this case exists. The data will be made publicly available. For the purpose of this study, we use a subset of this dataset, consisting of 92,935 users with labeled gender, who posted in total 44,725,839 photos. We consider each of these photo as an "item", and we attribute it the gender of its author.

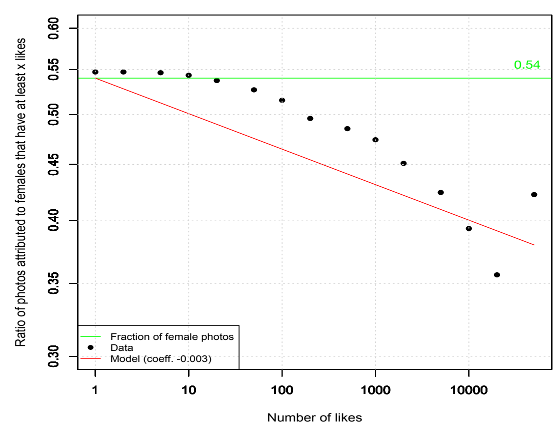

Figure 3: Fraction of photos posted by females among those that have at least x likes for the Instagram dataset (black dots) and model $\mathcal{M}$ (red line), where the power law coefficient is estimated to be a= -0.0325, plotted in log-log scale. The green line is the true fraction of female photos, i.e. 54%.

Figure 3 presents the observed fraction of female photos among those with at least k likes. As k grows, we focus on those that are most likely to be noticed and feature in a “top trending” set of items. Unfortunately, that seems to have a detrimental effect on the observed fraction of female pictures, which drops from 54% (the fraction of female accounts) to 40%. We note that the trend, mild at first, seems to accelerate towards the top.

At first, this seems the exact opposite of our model, which predicts that female should be more represented since they are a majority. In order to reconcile those results with this empirical analysis, we extend the present theoretical framework by introducing *differentiated homophily*, case in which the two subpopulations exhibit different homophily parameters.

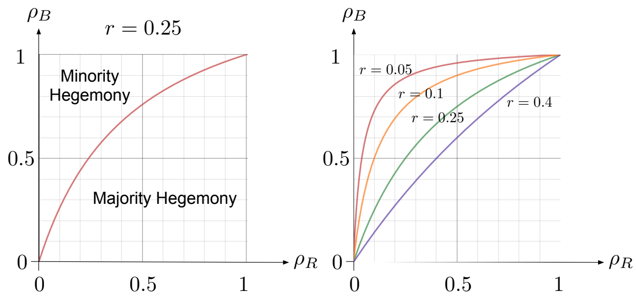

Figure 4: The two regions of hegemony appearing for different values of differentiated homophily ρ_B and ρ_R, plotted here for $r = 0.25$ (left), and various values of r (right).

5.2 Extension to differentiated homophily

As our main result, we find that hegemony is the norm, rather than the exception, for a more general model of homophily.

THEOREM 5.1. *If model $\mathcal{M}$ exhibits differentiated homophily ρ_R and ρ_B, in which a red node rejects a blue node with probability ρ_R and a blue node rejects a red node with probability ρ_B, then the network exhibits:*

- *A hegemonic majority iff* $(1-r) \cdot \left(\frac{1}{\rho_B} - 1\right) > r \cdot \left(\frac{1}{\rho_R} - 1\right)$.
- *A hegemonic minority iff* $(1-r) \cdot \left(\frac{1}{\rho_B} - 1\right) < r \cdot \left(\frac{1}{\rho_R} - 1\right)$.
- *No hegemony iff* $(1-r) \cdot \left(\frac{1}{\rho_B} - 1\right) = r \cdot \left(\frac{1}{\rho_R} - 1\right)$.

The proof of this theorem leverages new properties of the function F to show that the above comparison determines whether $F(r) < r$ holds. This theorem confirms the previous empirical observations that, in certain conditions, a minority with a stronger homophily may overcome its disadvantage and obtain an hegemonic advantage. We present in Figure 4 the various regions of (ρ_R, ρ_B) under which the two types of hegemony occurs. Note that majority hegemony occurs across a larger range of parameters.

We empirically observed that female users on Instagram are less homophilic: they are more likely to like a picture of a male than vice-versa. The homophily parameter is hence smaller for males than for females ($\rho_R < \rho_B$) which would be consistent with Theorem 5.1.

6 HEGEMONY UNDER RECOMMENDATIONS

In the extended model of recommendation, computing the evolution equation for the degree of the items becomes more complicated since we include the case of the random walk of length 3. Thus, we make a couple of assumptions for ease of work:

- edges in the random walk are assumed to have formed through model $\mathcal{M}$.
- edges in the random walk are independent of each other.
- edges that already exist are naturally formed at different steps in time; however, when computing an evolution equation, we set their time of formation to the current time, t.

We first start by investigating again the equilibrium state when the number of red edges converges as the fixed point of a function that describes their growth.

Rate at which red items receive connections: Denote by X_0, X_1, X_2, X_3 the vertices of the random walk created through recommendation, where X_0 is the newly added node to the network (thus, X_0 and X_2 are individuals and X_1 and X_3 are items).

Denoting again by R_{t+1} the number of red balls at time t and by $\alpha_{2,t}$ the fraction of red balls at time t, we observe the following cases in computing $\mathbb{P}(R_{t+1} = 1)$:

- a red item obtain a connection through model $M2$, with probability

$$\zeta \cdot \left(\eta \cdot r + (1-\eta) \cdot \left(\frac{r \cdot \alpha_{2,t}}{\alpha_{2,t} + \rho(1-\alpha_{2,t})} + \frac{(1-r) \cdot \rho\alpha_{2,t}}{\rho\alpha_{2,t} + 1 - \alpha_{2,t}}\right)\right). \quad (6)$$

- a red item obtain a connection through recommendation, with probability

$$(1-\zeta) \cdot \Bigg(\mathbb{P}(X_0 \in R)\frac{\mathcal{P}(X_3 \in R|X_0 \in R)}{\mathcal{P}(X_3 \in R|X_0 \in R) + \rho \cdot \mathcal{P}(X_3 \in B|X_0 \in R)} + \mathbb{P}(X_0 \in B)\frac{\rho \cdot \mathcal{P}(X_3 \in R|X_0 \in B)}{\mathcal{P}(X_3 \in B|X_0 \in B) + \rho \cdot \mathcal{P}(X_3 \in R|X_0 \in B)}\Bigg) \quad (7)$$

For ease of computation, consider the probabilities from above as functions of $\alpha_{2,t}$ (so for example, $\mathcal{P}(X_3 \in R|X_0 \in R) = \mathcal{P}(X_3 \in R|X_0 \in R)(\alpha_{2,t})$). Thus, we define a new function F_2, similar as for the organic growth model, which takes the following form (for detailed computation of the probabilities, refer to appendix A):

$$F_2(x) = \zeta \cdot F(x) + (1-\zeta) \cdot \Bigg(r \cdot \frac{\mathbb{P}(X_3 \in R|X_0 \in R)(x)}{\mathbb{P}(X_3 \in R|X_0 \in R)(x) + \rho \cdot \mathbb{P}(X_3 \in B|X_0 \in R)(x)} + (1-r) \cdot \frac{\rho \cdot \mathbb{P}(X_3 \in R|X_0 \in B)(x)}{\mathbb{P}(X_3 \in B|X_0 \in B)(x) + \rho \cdot \mathbb{P}(X_3 \in R|X_0 \in B)(x)}\Bigg), \quad (8)$$

where $F(x)$ is the function from model $\mathcal{M}$.

Computing each of these probabilities in closed form, we are able to show that F_2 exhibits the exact same properties as the function F (Lemma 4.2), having a fixed point α_2^*. This shows that the fraction of red edges also converges towards a constant that depends on r, ρ, η, ζ, as it is a stable point at the intersection of $F_2(x)$ and $f(x) = x$. Most importantly, empirical analysis reveals that $\alpha_2 < \alpha$ in this case as well, showing that the red items have less power under recommendation than under organic growth.

We continue by showing that the degree distribution also follows a power law and by showing that the coefficients are even further apart than in organic growth.

Degree distribution: Denote by $m^2_{k,t}(B)$ and $m^2_{k,t}(R)$ the number of blue and red items of degree k at time t in the model $\mathcal{M} - REC$, and define

$$M^2_k(x) = \lim_{t \to \infty} \frac{\mathbb{E}(m^2_{k,t}(x))}{t}, \text{ for } x \in \{R, B\}. \quad (9)$$

THEOREM 6.1. *In the case of recommendation growth according to model $\mathcal{M} - REC$, the degree distribution of the red and blue items also follows a power law, i.e.* $M^2_k(R) \sim k^{-\beta_2(R)}$ *and* $M^2_k(B) \sim k^{-\beta_2(B)}$ *for coefficients* $\beta_2(R)$ *and* $\beta_2(B)$.

We prove this by deriving again a recurrence relation for $m^2_{k,t}(R)$ and $m^2_{k,t}(B)$ (omitted here) and computing the coefficients as

$$\beta_2(R) = 1 + \frac{1}{(1-\eta)C_{2,R}} \text{ and} \beta_2(B) = 1 + \frac{1}{(1-\eta)C_{2,B}},$$

where $C_{2,R}$ and $C_{2,B}$ can again be computed explicitly.

Theorem 6.2. *For $r \in [0, 1/2]$ and $\rho \in [0, 1]$, we have*

$$\beta_2(R) > \beta(R) > 3 > \beta(B) > \beta_2(B), \quad (10)$$

implying again that the tail and strong glass ceiling effects are even more pronounced under algorithmic recommendations (model $\mathcal{M}$ – REC) than under organic growth (model $\mathcal{M}$).

Proof. In proving this, we derive an invariant equation for the rate of growth of red edges and use that $\alpha^*_2 < \alpha^*$. □

Lemma 6.3. *Given α and α_2 the limits of the fraction of red balls as defined above, the following holds:*

$$(1-\eta) \cdot \alpha \cdot C_R + \eta \cdot r = \alpha, \text{ and } (1-\eta) \cdot \alpha_2 \cdot C_{2,R} + \eta \cdot r = \alpha_2. \quad (11)$$

Proof. As C_R encapsulates the rate at which red edges appear, the system reaches an equilibrium state. Due to this, the rate at which red edges appear must equal the current fraction of red edges, as it does not evolve anymore. Furthermore, as η represented the probability for an item to be created, $\eta \cdot r$ is the rate of growth of edges towards red items in the case where they are created, and $(1-\eta) \cdot \alpha \cdot C_R$ is the rate of growth of edges towards red items in the preferential attachment case. The sum of these is then equal to α as this is the number of red edges at equilibrium. For the second part of equation 11, note that while ζ plays a role, our assumption is that the edges form sparsely through recommendation, so we may approximate the rate of growth of red edges at equilibrium. In a sense, since this result holds even when "ignoring" such edges formed through recommendation in this invariant equation, we conclude that their influence is even higher in reality. □

Subtracting the equations in 11, we obtain

$$\begin{aligned} (1-\eta) \cdot (\alpha \cdot C_R - \alpha \cdot C_{2,R}) &= \alpha - \alpha_2 \Leftrightarrow \\ (1-\eta) \cdot (C_R \cdot (\alpha - \alpha_2) + \alpha_2 \cdot (C_R - C_{2,R})) &= \alpha - \alpha_2 \Leftrightarrow \\ (1-\eta) \cdot \alpha_2 \cdot (C_R - C_{2,R}) &= (\alpha - \alpha_2) \cdot (1 - (1-\eta) \cdot C_R). \end{aligned} \quad (12)$$

As shown before, $C_R < 1$ and $\alpha_2 < \alpha$, which yields $C_R > C_{2,R}$, since $\eta < 1$. Since $\beta(R) = 1 + \frac{1}{(1-\eta)C_R}$ and $\beta_2(R) = 1 + \frac{1}{(1-\eta)C_{2,R}}$, Theorem 6.2 is proved. The case for C_B and $C_{2,B}$ can be proved similarly. Thus, we show that the tail and moment hegemony are even more exacerbated in item recommendations than in the organic network.

Normalized expected degree of red nodes: We analyze comparatively the effect of homophily on the expected degree of red nodes for the organic growth model $\mathcal{M}$ and the recommendation model $\mathcal{M}$ – *REC*, for different values of r. As compared to Figure 2 (a), we observe in Figure 2 (b) that recommendations exacerbate the gap between the two subgroups. Computing the normalized degree of red nodes as α/r (purple line) and α_2/r (orange line), recommendations decrease the expected degree of red items by 5% for both $r = 0.2$ and $r = 0.4$.

7 CONCLUSION

Our theoretical analysis unveils the subtle relation between network structure, homophily, and hegemony. By exploring a model of growth based on item recommendations, we show that in the case of a bi-populated network, the minority's viewpoint loses representation. It is interesting to notice that the minority viewpoint is not excluded from the network, as the ratio of items pertaining to the minority remains constant throughout, but it is only at the top of the hierarchy where this effect occurs. This effect has great ramifications in the way we interact with social media and the information that is disseminated through popular items. Future work should focus on extended empirical analysis and on the design of algorithms that alleviate such effects by taking into consideration network structure.

REFERENCES

[1] J An, D Quercia, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. 2014. Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science* (2014).
[2] C Avin, B Keller, Z Lotker, C Mathieu, and David Peleg. 2015. Homophily and the glass ceiling effect in social networks. *ITCS* (2015).
[3] E. Bakshy, S Messing, and L.A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
[4] Pranav Dandekar, Ashish Goel, and David T Lee. 2013. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* 110, 15 (April 2013), 5791–5796.
[5] S Flaxman, S Goel, and J M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quaterly* 80, 1 (2016), 298–320.
[6] D Fleder and K Hosanagar. 2009. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science* 55, 5 (May 2009), 697–712.
[7] Aniko Hannak, C Wagner, David Garcia, Markus Strohmaier, and Christo Wilson. 2016. Bias in Online Freelance Marketplaces: Evidence from TaskRabbit. *DAT '16: Proceedings of The Workshop on Data and Algorithmic Transparency* (Nov. 2016), 1–8.
[8] K Hosanagar, D Fleder, D Lee Management Science, and 2013. 2008. Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science* 60, 4 (2008), 805–823.
[9] Ravi Kumar, Prabhakar Raghavan, S Rajagopalan, D Sivakumar, Andrew Tomkins, and E Upfal. 2000. Stochastic models for the Web graph. (2000), 57–65.
[10] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *WWW '14: Proceedings of the 23rd international conference on World wide web*. International World Wide Web Conferences Steering Committee.
[11] S Nilizadeh, A Groggel, P Lista, and S Das. 2016. Twitter's Glass Ceiling: The Effect of Perceived Gender on Online Visibility. *Proceedings of the International Conference Weblogs and Social Media (ICWSM)* (2016).
[12] E Pariser. 2011. *The filter bubble: What the Internet is hiding from you.* Penguin.
[13] K Purcell, L Rainie, A Mitchell, T Rosenstiel, and K Olmstead. 2010. Understanding the participatory news consumer. *Pew Research Center's for Excellence in Journalism* (2010).
[14] M J Salganik, P S Dodds, DJ Watts science, and 2006. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311 (2006).
[15] Ana Stoica, Christopher J Riederer, and Augustin Chaintreau. 2018. Algorithmic Glass Ceiling in Social Networks. *WWW '18: Proceeding of the 25th International Conference on World Wide Web* (2018), 1–9.
[16] Marshall Van Alstyne and Erik Brynjolfsson. 2005. Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. *Management Science* 51, 6 (June 2005), 851–868.
[17] Ingmar Weber Venkata Rama Kiran Garimella. 2017. A Long-Term Analysis of Polarization on Twitter. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media ICWSM* (April 2017), 1–4.