Opinion dynamics of online social network users: a micro-level analysis

Ivan V. Kozitsin

Laboratory of Active Systems, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation; Department of Higher Mathematics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation

ABSTRACT
In this paper, we present an empirical study of the opinion dynamics of a large-scale sample of online social network users. We estimate users’ opinions as continuous scalars based on their subscriptions to information sources and analyze how friendship connections affect the dynamics of these estimations. Distinguishing between positive (toward friends’ opinions) and negative (away from friends’ opinions) opinion shifts, we find that the existence and magnitude of both types of shifts are positively related (largely through linear or inverted U-shaped form) to the distance in opinions between a user and their friends. The distance additionally moderates the balance between positive and negative movements: if the distance is within a certain moderate range, there is a relatively high chance of a positive shift.

ARTICLE HISTORY
Received 3 November 2020
Revised 13 July 2021
Accepted 14 July 2021

KEYWORDS
Opinion dynamics; online social networks; repulsive influence

1. Introduction
Individuals change their views and behavior over time and a remarkable contribution to this process is provided by social influence, a force that affects people through their peers and other sources of information. Today, online social networks (OSNs) have become a crucial channel of information diffusion. Users of these platforms can broadcast their views effectively, without the restrictions inherent in older forms of communication. They have given rise to large-scale systems of influence processes that unfold in an online environment. Investigation of such social systems is a challenging task that can, nonetheless, bring new insight into our knowledge of how individuals form their opinions (Bond et al., 2012; Ravazzi et al., 2020). By observing the opinion dynamics of users at the micro-scale, scholars can identify key micro-assumptions regarding the process of opinion formation and integrate this knowledge into existing opinion formation models (Abid et al., 2018; Castellano et al., 2009; Flache et al., 2017; Friedkin et al., 2016; Mäs, 2019; Proskurnikov & Tempo, 2017, 2018; Proskurnikov et al., 2017).
In this paper, we analyze longitudinal data on the opinion dynamics of a large-scale (1.6 M) sample of OSN users. Considering both users’ opinions (based on the information sources to which they are subscribed) and the structure of users’ friendship ties, we test competing assumptions regarding individuals’ opinion shifts against real data.

The majority of existing empirical studies suggest that the formation of continuous opinions (presented as continuous quantities) largely follow the linear positive mechanism whereby individuals tend to modify their opinions toward those of their friends with a magnitude proportional to the preexisting difference in opinions (Friedkin et al., 2021; Takács et al., 2016). However, some papers suggest a moderated positive influence: there is a maximum rate of influence if the difference is not too small and not too large; otherwise, it decreases (Moussaïd et al., 2013). Studies that reported a negative influence (opinion shifts away from the opinions of peers) either have methodological concerns (Knippenberg et al., 1990; Mazen & Leventhal, 1972) or are based on natural experiments in which researchers are unable to control for all possible confounding factors (Liu & Srivastava, 2015).

Our contributions are as follows. We distinguish between positive (toward friends’ opinions) and negative (away from friends’ opinions) opinion shifts. We find that both the probability and magnitude of positive opinion shifts are positively related (largely through a linear form or an inverted U-shaped form) to the degree of divergence in opinions between a user and their friends. Interestingly, we find that both the chance and magnitude of negative shifts tend also to increase alongside the opinion divergence between the focal user and their friends (largely through, again, a linear or an inverted U-shaped form). The balance between positive and negative shifts is moderated as well by the opinion divergence: if the opinions of the focal user and their friends are too similar or dissimilar, there is a relatively low chance of a positive shift.

These results challenge existing concerns about the presence of a negative influence; they constitute a clue for the context in which researchers should search for such an influence. Our findings indicate that scholars should pay more attention to modeling not only how individuals modify their opinions, but also when they do so. However, we also demonstrate how to explain our empirical results without referring to the notions of positive and negative influence. We concurrently investigate how the process of radicalization is connected to the ideological heterogeneity of users’ neighborhoods. We argue that individuals are less likely to radicalize if they are exposed to moderate positions; in contrast, radical views – regardless of their bias – induce a relatively high degree of radicalization.

The remainder of this paper is organized as follows. Section 1 Section 2 reviews the study’s foundational theory. In Section 2 Sections 3, we discuss the data at hand. Sections 3–5 Section 46 present our main results. In Section 6 Section 7, we discuss our results. Section 7 Section 8 offers a conclusion and suggests potential avenues for future research. Appendix includes supplemental information.
2. Theory

2.1. Social influence mechanisms

The literature on social influence models (those in which opinions are conceptualized as continuous quantities) emphasizes three main influence mechanisms (Flache et al., 2017): (1) positive (assimilative) influence; (2) positive influence coupled with the bounded confidence assumption; and (3) negative (repulsive) influence (see Figure 1). Here, we briefly discuss these three mechanisms.

In the case of positive influence, agent \(i\) aligns their opinion more closely with that of influence source \(j\). If opinions are represented on a one-dimensional axis (the only situation that we consider), agent \(i\)'s new opinion (see Figure 1, panel a) lands somewhere between influence source \(j\)'s opinion and agent \(i\)'s initial opinion (DeGroot, 1974). Another possibility that researchers may encounter while analyzing real data – despite rarely garnering attention in the literature – is the leapfrogging of a newly formed opinion past that of the influence source (Friedkin et al., 2021) (Figure 1, panel b). Of course, such “skips” could simply stem from measurement errors or external sources of influence that were not considered by researchers.

However, positive influence (without skips) in a typical case leads to an inevitable consensus in a connected social system; this result cannot explain persistent disagreement or opinion polarization in both small groups and large populations (Abelson, 1964). For this reason, the bounded confidence mechanism has been introduced, whereby agent \(i\) is influenced only if the opinion of influence source \(j\) is sufficiently similar to that of agent \(i\) (Deffuant et al., 2000;
Hegselmann & Krause, 2002); influence source \(j\)'s opinion must lie in agent \(i\)'s interval of confidence (Figure 1, panel c).

All of these models still fail to explain polarization – the growth in opinion divergence over time, whereby individuals’ opinions may even leave their initial interval (Dandekar et al., 2013). One potential solution here is the introduction of a negative influence whereby individuals with highly dissimilar opinions only intensify in divergence following an interaction (Altafini, 2013; Flache & Mäs, 2008; Macy et al.,) (Figure 1, panel d). In contrast to positive and bounded confidence influence mechanisms, which have substantial empirical support (Asch, 1961, ; Nickerson, 1998; Sherif & Hovland, 1961; Takács et al., 2016), the assumption of a negative influence suffers from a lack of definitive empirical studies in its favor (Mäs et al., 2013; Takács et al., 2016).

Other examples of models explaining polarization without the assumption of a negative influence can be found in Banisch and Olbrich (2019), Dandekar et al. (2013), and Mäs et al. (2013) These explanations are based on the idea of biased assimilation – the tendency of individuals to seek information that aligns with their views, avoid cross-cutting content, and interpret messages in a comfortable manner (Dandekar et al., 2013; Nickerson, 1998) – or on the introduction of more subtle forms of communication, such as specific arguments.

### 2.2. Mathematical formalization

Almost all the above-mentioned models can be captured by the following minimal opinion-formation rule:

\[
x_i(t+1) = x_i(t) + l(x_j(t) - x_i(t)).
\]

In equation (1), agent \(i\)'s opinion at time \(t + 1\) (assumed to be a scalar from a continuous one-dimensional space) is formed as the sum of their past opinion and the term standing for the influence from source \(j\) (representing the overall influence on \(i\)). The time scale in equation (1) is typically unspecified, and one step may stand for various timespans (e.g., day, week, month). The quantity \(l\) describes how agent \(i\) responds to the social influence, covering a wide range of micro-assumptions about influence processes. For simplicity, we do not bind opinions to a predefined interval. On the one hand, the opinion-formation rule in equation (1) can be understood as a formalization of a pairwise (one-to-one) interaction between agent \(i\) and influence source \(j\). On the other hand, the quantity \(x_j(t)\) may stand for the overall influence of peers on agent \(i\) (many-to-one interaction). One could further clarify \(x_j(t)\) by noting, for example, that \(x_j(t)\) is a convex combination of the opinions held by \(i\)'s peers.

From equation (1), it follows that \(l\) describes how the shift in agent \(i\)'s opinion is connected to the opinion divergence between \(i\) and \(j\):
If \( l \) is a positive constant (see Figure 2, which we borrowed from Takács et al., 2016), we get a positive linear influence dependency and the corresponding assumption of positive linear influence. Note that, to avoid skips, we should require \( l \leq 1 \). To account for the bounded confidence effect in its relaxed form whereby there is a possibility of opinion exchange, albeit a relatively minor one, between agents with polar positions (Kurahashi-Nakamura et al., 2016), one may consider a moderated positive influence relation. In this type of relation, the curve illustrating agent \( i \)'s opinion shift – which depends on the distance between \( i \)'s and \( j \)'s opinions – has an inverted U-shaped (negative quadratic) form (Takács et al., 2016). Assuming \( l<0 \), we see a negative influence when \( i \)'s opinion shifts away from that of \( j \). As with a positive influence, there may be negative linear influence and moderated negative influence dependencies that can be obtained from their positive counterparts through reflection on a horizontal axis (for brevity, we do not plot them). Note that the negative linear influence dependency is largely affected by the geometrical properties of the opinion space. For example, Takács et al. (2016) consider the case when opinions are restricted to lie in the interval \([0, 100]\). As a result, they have a different formalization of the negative linear influence

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Here, we plot how the magnitude of opinion change \( x_i(t+1) - x_i(t) \) may vary by the absolute difference between an individual's opinion \( x_i(t) \) and that of the influence source \( x_j(t) \). The upper half-plane represents opinion shifts toward the influence source \( l(x_i(t+1) - x_i(t)) + (x_j(t) - x_i(t)) > 0 \); the lower half-plane represents negative influence cases in which opinion shifts away from the influence source \( x_j(t) \).
assumed. In Subsection 6.2, we return to this problem. However, negative influence is rarely modeled without positive influence, so one potential solution here is to combine the moderated positive and moderated negative influence components: when the absolute opinion divergence becomes too substantial, the positive influence gives way to a negative influence that first grows in power but then disappears as individuals become so distinct in their views that there is no chance for interaction.

Equation (1) represents how individuals modify their opinions under the assumption that they do modify them. In real life, it could be the case that a discrepancy in opinions does not lead to a change of opinion. As such, the question arises of what factors shape the probability of opinion changing. The bounded confidence assumption is a potential solution here. A quite general approach of leveraging this assumption was presented in the paper of Kurahashi-Nakamura et al. (2016) who introduced what they called “probability of acceptance” – the chance that an individual will accept influence from a particular peer. In a nutshell, they assumed that the probability of acceptance should decrease as \( |x_j - x_i| \) increases. In that paper, the value of \( |x_j - x_i| \) affected both the probability of acceptance and the magnitude of opinion shift (the latter through the positive linear influence assumption).

Alternatively, one can simply employ the mathematical laws used in previously introduced linear and moderated assumptions to model the probability of acceptance. For example, applying the positive linear influence assumption to the probability of acceptance, we find that the odds of being influenced will grow linearly alongside the level of discrepancy between the focal agent and their social environment.\(^\text{1}\) This hypothesis is consistent with the theory since larger differences in opinions should increase the level of an individual’s cognitive dissonance and, thus, the individual’s desire to reduce it by changing their opinion (Festinger, ). Analogously, adopting the moderated positive influence assumption, the agent is likely to change their opinion if the quantity \( |x_i - x_j| \) is not too small and not too large. Theoretical motivation for this assumption is twofold. On the one hand, small values of \( |x_i - x_j| \) can be misperceived by individual \( i \) such that \( i \) finds no difference between their own opinion and \( j \)’s (Sherif & Hovland, 1961), or, alternatively, create a strong desire to stand within their current opinion as a result of, for example, striving for uniqueness (Mäs et al., 2010). On the other hand, a large opinion distance

\(^{1}\)In the general case, the probability of acceptance and the magnitude of opinion shift could follow different assumptions and thus take different mathematical forms. For example, the probability of acceptance could depend on \( |x_i - x_j| \) linearly with a positive slope (linear assumption) while the magnitude of opinion shift may follow the moderated assumption.

\(^{2}\)If the probability of acceptance and the magnitude of opinion shift both depend on \( |x_i - x_j| \), then the effect on the expected opinion shift could be nonlinear. For example, if the probability of acceptance is defined as a linear function of \( |x_i - x_j| \) and we assume the positive linear influence assumption in (1), then the expected opinion shift will depend quadratically on \( |x_i - x_j| \).
is unlikely to lead to assimilation because of the bounded confidence effect. Similarly, one can apply linear and moderated negative influence assumptions to the probability of acceptance. That is, a larger opinion difference should intensify an individual’s desire to emphasize and increase discrepancies (linear negative assumption). However, if the opinion space is bounded, then a large distance indicates that there is little to no room to make a repulsive shift – the individual’s opinion is near one of the opinion space edges – so the chance of such shifts should be relatively small (moderated negative assumption).

A different approach is to hypothesize that expected opinion shifts should depend on $|x_i - x_j|$ in a particular form (Takács et al., 2016). In this case, the probability of acceptance and magnitude of (occurred) shift could be arbitrary, provided the expected opinion shift (which is a result of their product) depends on $|x_i - x_j|$ in a suitable fashion.

A consideration that also plays a crucial role in opinion dynamics is an individual’s willingness to express their opinions (Gaisbauer et al., 2020). However, our data gives us no opportunity to control for this effect, so we do not implement it in theoretical models and assume that all individuals homogeneously express their opinions. Note that, in the models listed above, all agents are equally involved in discussions.

### 2.3. Testing against real data

Existing empirical studies (see Table 1) that predominantly analyze how the expected opinion shifts depend on opinion discrepancy largely support the positive linear influence assumption; some of them also witness the moderated positive influence assumption (Moussaid et al., 2013) and some even support the existence of a negative influence (Knippenberg et al., 1990; Liu & Srivastava, 2015; Mazen & Leventhal, 1972). However, it is important to note that the studies that suggest a negative influence either have methodological concerns (Knippenberg et al., 1990; Mazen & Leventhal, 1972) or are based on natural experiments, which are more prone to improper conclusions than experiments in laboratory settings (Liu & Srivastava, 2015). However, it would be difficult to reproduce meaningful negative relationships, which constitute an essential source of repulsive dynamics, in a laboratory setting (Takács et al., 2016).

The current paper investigates longitudinal data on the opinion dynamics of OSN users. This data has previously been analyzed by Kozitsin (2020), who concentrated on how users – endowed with estimated opinions from the interval $[0, 1]$ – choose the direction in which they will move in the opinion space (left or right). Kozitsin finds that, after controlling for user opinion, the probability of shifting left varies with the average opinion of the user’s friends.

---

3 Moussaid et al. (2013) maintained also that the chances of opinion shifts follow the positive moderated assumption.
3. Data description

The data used in this paper initially appeared in Kozitsin et al. (). It was gathered from VKontakte (VK), the most popular online social network in Russia.

**Short overview of VK.** In terms of functionality, VK is very similar to Facebook. VK users can form follow-type (directed) connections with other users. After user \( i \) establishes a follow-type connection with user \( j \), user \( j \) receives a notification indicating that user \( i \) requests friendship. If user \( j \) accepts, the two users become friends. Otherwise, user \( i \) simply remains a follower of user \( j \). Users with more than 1,000 followers have a special status on VK (henceforth referred to as *bloggers*): information on followed bloggers is displayed on user \( i \)'s account page in a special section. Users can also follow non-user accounts, including those for *groups, public pages, and events*. The main difference between public pages (that are also visible on user \( i \)'s account page in the same section as bloggers) and other non-user account types is that, 

\[ x_{-j} \]; the target function itself has a near-sinusoidal, wave-shaped form. If source opinion \( x_{-j} \) is located more to the left than user opinion \( x_i \) \( (x_{-j} < x_i) \), the probability of shifting left is greater than when \( x_{-j} > x_i \). In this paper, we explore the remaining important questions pertaining to opinion-formation processes: when – and to what degree – do individuals change their opinions? In contrast to previous empirical studies, our paper relies on a substantially larger dataset that provides information on the trajectories of individuals’ opinions and thus gives an opportunity to analyze opinion dynamics at the microscale and characterize the existence and magnitudes of different types of opinion shifts with fine resolution. Individuals are embedded in natural settings whereby there is a higher chance of antagonistic interactions than in controlled laboratory experiments. However, this dataset also has several drawbacks that are discussed in the following section.

### Table 1. Studies testing influence assumptions against real data (listed in order of publication).

| Reference | Linear positive | Moderated positive | Approved (+) or Rejected (-) Hypothesis |
|-----------|-----------------|--------------------|----------------------------------------|
| (Mazen & Leventhal, 1972) |  |  | +* |
| (Knippenberg et al., 1990) |  |  | +** |
| (Krizan & Baron, 2007) |  |  |  |
| (Moussaid et al., 2013) |  | + |  |
| (Liu & Srivastava, 2015) |  |  | -*** |
| (Kerckhove et al., 2016) | + |  |  |
| (Takács et al., 2016) |  | + |  |
| (Clemm von Hohenberg et al., 2017) |  | + |  |
| (Friedkin & Bullo, 2017) |  | + |  |
| (Friedkin et al., 2021) |  | + |  |

*Note: * – methodological concern: did not to consider trends in opinion shifts; ** – methodological concern: did not disentangle positive in-group influence and negative out-group influence; *** – nonlinear in both cases with an increasing rate.
while users may freely follow and unfollow public pages without restrictions, group-type accounts are able to limit access to their content. Bloggers and public pages are considered to be the main information disseminators on VK. It is worthwhile to note that the majority of media accounts on VK are public pages. VK users receive information from their news feeds, where content from the accounts to which they are subscribed is listed.

The data consists of three opinion snapshots from a sample of $N = 1,660,927$ VK users. This sample was obtained through random selection among active users (at least one platform interaction per month) who are Russians of at least 18 years of age with open privacy settings. Additionally, users needed to follow at least ten and at most 200 accounts of bloggers and public pages (henceforth referred to as information sources). We explain the rationale behind this filter below. Users with no friends were excluded from the sample. The collected opinion snapshots represent users’ attitudes toward President Putin in February ($t_1$), July ($t_2$), and December 2018 ($t_3$).

The logit model introduced by Kozitsin et al. (2020) estimates users’ opinions as non-negative scalars, $x(t) \in [0, 1]$. The quantity $x_i(t)$ represents user $i$’s opinion and is, generally speaking, a projection from the set of followed information sources to the interval $[0, 1]$. The main idea of the estimation algorithm is that users’ subscriptions should reflect their political views (Barberá, 2014; Frey, 1986; Tang & Chorus, 2019). If $x_i(t) = 0$, then user $i$ is a strong oppositionist. In contrast, $x_i(t) = 1$ means that this individual is a staunch supporter of President Putin. Meanwhile, users with $x_i(t) = 0.5$ are moderate individuals in terms of the opinions’ assessing strategy. More precisely, such users are those who follow no politically relevant information sources or those who follow sources pertaining to polar ideological sides simultaneously. The argument $t$ in $x_i(t)$ reflects the moment in time at which the data used to produce the estimation was gathered. In our case, the time step $t \rightarrow t + 1$ from equation (1) corresponds to a half-year step from $t_k \rightarrow t_{k+1}$. In the analysis that follows, we do not discuss the time argument if it is clear from the context.

As pointed out by Kozitsin et al. (2020), estimations of users’ opinions are most accurate when the number of their subscriptions to information sources is between ten and 200. To ensure sustainable opinion estimation, users who do not meet this requirement were eliminated from the sample. There are two reasons for this requirement. On the one hand, users with a few subscriptions seem to not be very active on VK and thus their online accounts are unlikely to reflect their opinions. On the other hand, users who follow a huge number of information sources likely pursue purposes that are different from simple information consumption (e.g., participating in sweepstakes, advertising) or might be Internet bots.

The dataset also includes an adjacency matrix $A = [a_{ij}] \in \{0, 1\}^{N \times N}$ that represents friendship connections between users from the sample. These
connections were gathered in July 2018. The sets of information accounts and sample users have no intersections, facilitating the independence of users’ opinion estimations from interconnections. Our analysis of the friendship network reveals that it includes a giant connected component of 1,648,829 nodes (99.3% of all nodes) and many (5,535) tiny connected components (5,535) formed primarily by two or three nodes. In what follows, we focus on users from the giant connected component.

All of the data, codes, and other support information used in this study can be found at https://doi.org/10.7910/DVN/H3ZBHR.

4. Ideological groups

We group individuals into five categories based on their opinions. For simplicity, we refer to individuals as “strong liberals” (SLs: \(x_i \in [0, 0.2)\)), “liberals” (Ls: \(x_i \in [0.2, 0.4)\)), “moderates” (Ms: \(x_i \in [0.4, 0.6)\)), “conservatives” (Cs: \(x_i \in [0.6, 0.8)\)), and “strong conservatives” (SCs: \(x_i \in [0.8, 1]\)). Moving forward, individuals with opinions near the edges of the opinion space are referred to as those with strong or radical opinions. We will say that the opinions of SLs are stronger or more radical than those of Ls; Ms are the individuals with the weakest or least radical opinions. If \(x_i < x_j\), then user i is said to be more liberal than user j. We refer to users with opinions of less than 0.5 as those with a liberal bias. Conversely, users having opinions greater than 0.5 are individuals with a conservative bias. While we recognize that such names of groups are not correct from a political science theory standpoint, the analogies remain pertinent. We report that there are no users with opinions strictly equal to 0.5. The populations of each group (at time \(t_2\)) are presented in Table 2. We find that, on average, more liberal users have more friends (Pearson correlation coefficient equal to \(-0.16\), \(p\)-value approximately equal to zero). To explain this result, we put forward the following hypothesis: liberal individuals are, on average, younger; thus, they should be more active on VK. We perform an additional experiment that partially confirms our hypothesis (see Appendix A for more details).

| Group          | Opinion interval | SLs    | Ls     | Ms     | Cs     | SCs    |
|----------------|------------------|--------|--------|--------|--------|--------|
| Opinion interval | \([0, 0.2)\)       | \([0.2, 0.4)\) | \([0.4, 0.6)\) | \([0.6, 0.8)\) | \([0.8, 1]\) |
| Population     | 125,714           | 312,063 | 874,327 | 266,509 | 70,216 |
| Avg. number of friends | 25.68           | 21.2    | 17.13  | 13.03  | 10.43  |
5. Homophily structure

Homophily is a well-documented phenomenon of social networks. It can be formulated as follows: if one takes a snapshot of a social network, they can observe that characteristics of connected individuals are more similar than one could expect in the case of randomly created ties (McPherson et al., 2001; M. E. J. Newman, 2018). Regarding opinions and behaviors, there are two principal explanations of such phenomena (Holme & Newman, 2006): selectivity (the tendency of individuals to form connections with those having similar traits, including opinions or behaviors) and social influence. For the analysis of opinion dynamics, it is essential to explore whether the system at hand is homophilic and to what extent as it (1) may be a result of social influence processes (and, thus, may be used to make relevant hypotheses; see discussion at the end of this section) and (2) may affect how the influence processes unfold.

Kozitsin (2020) reports that the network under consideration is homophilic with the assortativity coefficient of approximately 0.14. In this study, we perform a comprehensive analysis of homophily patterns. For brevity, we concentrate on time moment $t_2$, at which the data on friendship connections was gathered. Using this data, we calculate the ideological composition of each user’s neighborhood (represented as the fraction of their peers falling in each of five groups) and average these values across all ideological categories. We then compare our findings against the null model, in which connections between individuals appear at random. Within the null model, the list of friends of a randomly chosen user should consist of 8% SLs, 19% Ls, 53% Ms, 16% Cs, and 4% SCs. If the system is homophilic, one can expect a concentration of ties on and near the main diagonal.

Our analysis (see Table 3) reveals that users’ tendency to have homophilic ties is strongly connected to their degree of radicalism: individuals with strong positions tend to have more ties with those holding similar views than predicted by the null model. For moderate individuals, we observe little to no homophily. After controlling for the total number of friends, we find that for individuals having more friends, the connections map is significantly

| Ideological group | SLs  | LS   | MS   | CS   | SCs  |
|-------------------|------|------|------|------|------|
| SLs               | 0.15 | 0.24 | 0.46 | 0.11 | 0.03 |
| Ls                | 0.11 | 0.23 | 0.51 | 0.12 | 0.03 |
| Ms                | 0.08 | 0.20 | 0.54 | 0.15 | 0.03 |
| Cs                | 0.07 | 0.17 | 0.49 | 0.21 | 0.06 |
| SCs               | 0.07 | 0.15 | 0.45 | 0.24 | 0.10 |
| Null model        | 0.08 | 0.19 | 0.53 | 0.16 | 0.04 |
biased toward liberals (see Tables C1–C3 in Appendix C). We explain this result as follows: individuals with more friends tend to be younger and, thus, they should have younger friends (since social networks are assortative with respect to age) that, in turn, are more likely to be liberal.

In general, our results indicate that the level of homophily varies across ideological groups. As such, we must carefully disentangle individuals with different positions while analyzing patterns of social influence. The following question naturally arises: if the system at hand stems from selectivity or social influence (or a combination of both), what assumptions should we incorporate into models of selectivity or social influence (or both) to get results similar to those from Table 3? In Section 6 Section 7, we return to this problem.

6. Analysis of opinion shifts

In this section, we analyze users’ opinion shifts from the perspective of the baseline micro-assumptions detailed in Section 1 Section 2. We base our analysis on three principal aspects of opinion-formation processes:

Q1. Chance of opinion shift
Q2. Direction of opinion shift
Q3. Magnitude of opinion shift

Q2 was covered extensively by Kozitsin (2020). As a result, instead of Q2, we address the specific issue of opinion radicalization (Q2'), which is of considerable interest in the literature: under what conditions are individuals more likely to become more radical in their views?

6.1. Preliminaries

Before starting the analysis, we must define some metrics. We assume that all users exert equal influences on each other and use the average opinion of agent i’s friends to describe the ideological composition of their ego network:

\[ x_{-i} = \frac{\sum_{j=1}^{n} a_{ij} x_j}{\sum_{j=1}^{n} a_{ij}} \]

\[ \sigma_{-i} = \sqrt{\frac{\sum_{j=1}^{n} a_{ij} (x_j - x_{-i})^2}{\sum_{j=1}^{n} a_{ij}}} \]

Friends’ average opinion does not provide comprehensive insight into agent i’s ideological environment, as the same values of \( x_{-i} \) may – from the perspective of the baseline micro-assumptions – correspond to qualitatively different situations and lead to incorrect inferences (see Figure 3 for more details). For this reason, we also calculate the standard deviation of friends’ opinions:
This quantity measures the diversity of friends’ opinions.

To avoid noisy opinion shifts, we only consider shifts that exceed a pre-defined value of 0.05, which is widely used in testing statistical hypotheses:

**Definition 1.** Opinion shift \( x_i(t_k) \to x_i(t_{k+1}) \) is remarkable if its magnitude \( |x_i(t_{k+1}) - x_i(t_k)| \) is greater than 0.05.

To evaluate the probability of acceptance, we introduce the following measure:

**Definition 2.** Estimated probability of opinion change (EPOC) is defined as the proportion of individuals who make a remarkable opinion movement.

We calculate EPOC (and other metrics) for two time steps: \( t_1 \to t_2 \) and \( t_2 \to t_3 \).

Of great importance are situations where users shift their opinions toward the nearest edge of the opinion space. To emphasize these users, we introduce the following definition:

---

**Figure 3.** Illustration of how opinion diversity among peers may affect individuals’ opinion dynamics. The black stick figure icons represent agents’ past opinions and the gray stick figure icons represent their new positions formed in response to an influence (black square). The icons with non-colored heads represent friends’ opinions. **Panel a.** In the top case, the agent shifts away from their friends’ opinions (potentially as a result of negative influence). In the bottom case, the agent shifts toward the opinion held by the highlighted friend icon – the real source of (positive) influence. In both cases, observers (those who do not know about the agent’s friends and see only their average opinion) likely conclude the existence of a negative influence. **Panel b.** Assume that opinion dynamics obey the bounded confidence assumption. In the top case, there is no chance for influence, as both potential influencers are beyond the agent’s interval of confidence. In the bottom case, one of the influencers falls within the interval of confidence and exerts a positive influence. The initial settings in both cases are the same for observers – but the outcomes are different. **Panel c.** In the top case, the focal agent makes a skip over both potential influencers. In the bottom case, one of the influencers (the one located further away from the agent) is a real source of positive influence (without a skip). In both cases, observers witness a skip.
**Definition 3.** Agent $i$ becomes stronger in their opinion or radicalizes between times $t_k$ and $t_{k+1}$ if they make a remarkable opinion shift $x_i(t_k) \rightarrow x_i(t_{k+1})$ toward the nearest edge of the opinion space.

Note that Definition 3 is only correct for those users whose opinions differ from 0.5. However, as was mentioned in Section 3 Section 4, all users’ opinions differ from 0.5 (for all three waves). We disentangle opinion shifts toward and away from the average opinion of friends:

**Definition 4.** Remarkable opinion shift $x_i(t_k) \rightarrow x_i(t_{k+1})$ is said to be positive if it is directed toward the friends’ average opinion: $(x_i(t_{k+1}) - x_i(t_k)) \times (x_{-i}(t_k) - x_i(t_k)) > 0$. Remarkable opinion shift is said to be negative if it is directed away from the friends’ average opinion: $(x_i(t_{k+1}) - x_i(t_k)) \times (x_{-i}(t_k) - x_i(t_k)) < 0$.

Note that both positive and negative shifts can result in radicalization. Among positive opinion shifts, we distinguish between those with and without skips:

**Definition 5.** Positive opinion shift $x_i(t_k) \rightarrow x_i(t_{k+1})$ is said to be non-skipping if the new opinion $x_i(t_{k+1})$ lies in the closed interval between $x_i(t_k)$ and $x_{-i}(t_k)$. Positive opinion movement is said to be skipping if it is not non-skipping.

Note that negative, skipping, and non-skipping shifts collectively cover all possible remarkable opinion shifts. As such, EPOC can be naturally deconstructed as the sum of three terms standing for the probabilities of non-skipping, skipping, and negative shifts:

$$EPOC = EPOC_+ + EPOC_- = EPOC_{+}^{skip} + EPOC_{+}^{nonskip} + EPOC_-$$

In our study, we frequently explore different dependencies unfolding in the opinion space; for example, we consider the function $EPOC_+ = EPOC_+(x_i, x_{-i})$, which represents the estimated chance of a positive shift as a function of $x_i$ and $x_{-i}$. The following two definitions are thus useful in analyzing these dependencies.

**Definition 6.** Let us consider ideological group $G$, where $G \in \{SL, L, M, C, SC\}$. Equality $x_i = G$ indicates that agent $i$ belongs to ideological group $G$. Inequality $x_i > G$ means that agent $i$’s opinion is located further to the right than $G$ in the opinion space (for example, inequality $x_i > M$ means that $x_i = C$ or $x_i = SC$).
Definition 7. Let us consider movement \( G_1 \rightarrow G_2 \), where \( G_1, G_2 \in \{SL, L, M, C, SC\} \) and \( G_1 < G_2 \) (the case \( G_2 < G_1 \) is elaborated analogously). We determine the area of negative shifts by \( x_{-i} < G_1 \) and the area of positive shifts by \( x_{-i} > G_1 \). Inequality \( G_1 < x_{-i} < G_2 \) defines the area of skipping movement; if \( G_2 \leq x_{-i} \), we get the area of non-skipping movements. If \( G_1 = G_2 \), the shift \( G_1 \rightarrow G_2 \) is called static.

It is worthwhile to mention that static movements may constitute remarkable shifts. However, note that Definition 7 prohibits highlighting skipping movements if \( G_1 \) and \( G_2 \) are neighboring ideological groups.

6.2. Core assumptions and expectations

In Table 4, for each quantity of interest, we present which mathematical forms it may (but is not necessarily limited to) take. We investigate magnitudes of opinion shifts in two ways: (1) we analyze how the magnitude of occurred (remarkable) movements (or, simply, conditional magnitude) depends on \( |x_i - x_{-i}| \) and (2) we study how expected magnitude (which is obtained as the product of conditional magnitude and EPOC) depends on the same quantity. We consider three main forms which are motivated theoretically (see Subsection 2.2): linear (with a positive slope), negative quadratic, and decreasing. The linear form stems from the linear influence assumption and the negative quadratic form appears from the moderated assumption. Both linear and negative quadratic forms are applicable for positive and negative movements and may describe the behavior of EPOC and magnitudes of opinion changes. Instead, the decreasing form (that comes from the bounded confidence assumption) appears only for positive movements and is applicable solely to EPOC. Note that we concentrate on how the metrics should look if we consider positive and negative shifts separately. The basis behind this separation is that positive and negative shifts are different from the “geometrical intuition.” A simple example: increasing the value of \( |x_i - x_{-i}| \), we make more room for a positive shift and, concurrently, reduce room for a negative shift. It is worthy to note that, in the general case, different metrics may take different mathematical forms. For example, EPOC of positive movements may depend on \( |x_i - x_{-i}| \) linearly (with positive slope), and conditional magnitude of positive movements could depend on \( |x_i - x_{-i}| \) quadratically (with a negative coefficient). In this case, the expected magnitude of positive movements will depend on \( |x_i - x_{-i}| \) cubically, a situation that is more demonstrative than realistically possible since it conflicts with empirics that largely witness that expected magnitude should follow the positive linear assumption (see Subsection 2.3).

However, it is also important to understand how positive and negative shifts coexist in opinion dynamics. The literature suggests that positive influence is
active when the opinion difference \(|x_i - x_{-i}|\) is not too large; high \(|x_i - x_{-i}|\) values likely result in negative influence. In other words, when opinion divergence is large enough, positive influence is replaced by negative influence. This means that high values of \(|x_i - x_{-i}|\) should be associated with negative movements, whereas if \(|x_i - x_{-i}|\) is not too large, then one should expect positive movements.

### 6.3. Probability of opinion change

In this Subsection, we try to address the question “Under what conditions do individuals change their opinions more readily?” We begin our analysis by calculating the map of cross-group movements (see Table 5). We find that the most popular user “strategy” is to stay within the current ideological group. Most cross-group shifts occur between neighboring ideological groups. Our analysis reveals that the populations of SLs and Ls increase, whereas SCs are nearly balanced and the populations of Ms and Cs clearly decrease.

The next part of our analysis is structured around the investigation of EPOC as a function of both users’ and friends’ opinions: \(EPOC = EPOC(x_i, x_{-i})\). For time step \(t_1 \to t_2\), we observe 147,344 remarkable shifts (60% positive). For time step \(t_2 \to t_3\), we observe 84,210 remarkable movements (57.9% positive). In Figure 4, we depict how EPOC varies with friends’ average opinion \(x_{-i}\) across different values of

**Table 4.** Metrics and theoretically motivated expectations.

| Quantity of interest | Linear with a positive slope (quantity of interest grows linearly with the value of \(|x_i - x_{-i}|\)) | Negative quadratic (quantity of interest first increases then decreases as \(|x_i - x_{-i}|\) goes up) | Decreasing (quantity of interest declines with the value of \(|x_i - x_{-i}|\)) |
|----------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| \(EPOC_i\)           | \(+\) (radicalization most likely when \(0.5 < x_i < x_{-i}\) or \(x_{-i} < x_i < 0.5\)) | \(+\) (radicalization most likely when \(0.5 < x_i < x_{-i}\) or \(x_{-i} < x_i < 0.5\) and \(|x_i - x_{-i}|\) has a “moderate” value) | \(+\) (radicalization most likely when \(0.5 < x_i < x_{-i}\) or \(x_{-i} < x_i < 0.5\) and \(|x_i - x_{-i}|\) is small) |
| \(EPOC_{-i}\)        | \(+\) (radicalization most likely when \(x_{-i} < x_i\) and \(0.5 < x_i\) or when \(x_{-i} > x_i\) and \(0.5 > x_{-i}\)) | \(+\) (radicalization most likely when \(x_{-i} < x_i\) and \(0.5 < x_i\) or when \(x_{-i} > x_i\), \(0.5 > x_{-i}\) and \(|x_i - x_{-i}|\) has a “moderate” value) | \(+\) |

Note: * – the word “moderate” is somewhat general here, meaning only that \(|x_i - x_{-i}|\) is not too small and not too large (depending on the context).
Table 5. Map of movements.

|       | -> SLs | -> Ls  | -> Ms   | -> Cs   | -> SCs  | Group growth | Group decrease | Growth rate (group growth – group decrease)/group size |
|-------|--------|--------|---------|---------|---------|--------------|---------------|-------------------------------------------------------|
| $r_1 \rightarrow r_2$ |        |        |         |         |         |              |               |                                                       |
| SLs   | 113,081| 5,662  | 991     | 117     | 25      | 12,633       | 6,795         | 0.049                                                 |
| Ls    | 11,141 | 269,961| 21,032  | 448     | 49      | 42,102       | 32,670        | 0.031                                                 |
| Ms    | 1,307  | 35,724 | 830,633 | 16,706  | 233     | 43,694       | 53,970        | −0.012                                                |
| Cs    | 147    | 636    | 21,245  | 4,863   | 22,393  | 26,891       | −0.011        |                                                       |
| SCs   | 38     | 80     | 5122    | 65046   | 5,170   | 5,666        | −0.007        |                                                       |
| $r_2 \rightarrow r_3$ |        |        |         |         |         |              |               |                                                       |
| SLs   | 121,036| 3,915  | 645     | 98      | 20      | 8,865        | 4,678         | 0.033                                                 |
| Ls    | 8,071  | 291,304| 12,483  | 186     | 19      | 30,433       | 20,759        | 0.031                                                 |
| Ms    | 690    | 26,107 | 835,230 | 12,194  | 106     | 27,909       | 39,097        | −0.013                                                |
| Cs    | 83     | 351    | 14,454  | 3,926   | 5,170   | 15,983       | 18,814        | −0.011                                                |
| SCs   | 21     | 60     | 327     | 3,505   | 4,071   | 3,913        | 0.002         |                                                       |

Note: each cell in first five columns represents the population of users who make a particular movement.
xi, separated by movement type (positive or negative). To avoid situations in which both the focal user’s opinion and the opinions of their friends are altered (since we do not know which change came first), we calculate EPOC by only considering individuals whose friends’ average opinion changes by less than 0.05. In doing so, we attempt to disentangle real cases of influence from potential confounders such as the presence of common stimuli (see Example 1 in Discussion) or general trends in opinion dynamics. We report that 97.7% of all users meet this condition for time step \( t_1 \rightarrow t_2 \) (only 3.2% of all remarkable movements are excluded) and 98.8% of all users meet it for time step \( t_2 \rightarrow t_3 \) (1.7% of all remarkable movements are excluded). In what follows, we calculate metrics only for such users. In Figure C1 (Appendix C), we demonstrate how \( EPOC(t_k \rightarrow t_{k+1}) \) is moderated by the value of \(| x_{-i}(t_{k+1}) - x_{-i}(t_k) |\) (the general trend is that higher values of \(| x_{-i}(t_{k+1}) - x_{-i}(t_k) |\) increase the chance of opinion shift).

**Positive movements.** \( EPOC_+ \) is positively connected with the distance \(| x_i - x_{-i} |\); if we bring \( x_{-i} \) away from \( x_i \) (left or right), \( EPOC_+ \) increases, excepting a number of zones near \( x_{-i} = 0 \) and \( x_{-i} = 1 \). The minima are located at \( x_{-i} \approx x_i \). We highlight clear linear segments (for example: \( t_1 \rightarrow t_2, x_i = M, x_{-i} \in [0.6, 1] \); \( t_1 \rightarrow t_2, x_i = SC, x_{-i} \in [0, 0.4] \); \( t_2 \rightarrow t_3, x_i = C, x_{-i} \in [0, 0.4] \)) and sharp decreases at the edges of the opinion space (for example: \( t_1 \rightarrow t_2, x_i = SL, x_{-i} \in [0.7, 1] \); \( t_2 \rightarrow t_3, x_i = SC, x_{-i} \in [0, 0.2] \)). For non-moderate individuals we notice intervals where \( EPOC_+ \) does not change considerably \( (t_1 \rightarrow t_2, x_i = SL, x_{-i} \in [0.2, 0.4]) \).

**Negative movements.** If we bring \( x_{-i} \) away from \( x_i \) (left or right), \( EPOC_- \) tends to increase, sometimes falling near the interval boundaries. Minima are not located at \( x_{-i} \approx x_i \); instead, they are usually slightly shifted from \( x_i \) (rightward for SLs and Ls and leftward for Ms, Cs and SCs). Some parts of curves have clear linear (for example: \( t_1 \rightarrow t_2, x_i = L, x_{-i} \in [0.5, 1] \); \( t_1 \rightarrow t_2, x_i = C, x_{-i} \in [0, 0.5] \); \( t_2 \rightarrow t_3, x_i = M, x_{-i} \in [0, 0.35] \) or inverted U-shaped \( (t_1 \rightarrow t_2, x_i = SL, x_{-i} \in [0.7, 1]) \); \( t_2 \rightarrow t_3, x_i = SC, x_{-i} \in [0, 0.4] \)) forms.

**Mutual positioning of curves.** If \( x_{-i} < x_i \), the probability of a positive shift is higher than that of a negative shift:

\[
EPOC_+(x_i, x_{-i}) > EPOC_-(x_i, x_{-i})
\]

If \( x_{-i} > x_i \), the situation is rather unclear: for SLs and Ls, \( EPOC_+ \) usually dominates \( EPOC_-(x_i, x_{-i}) \); for other groups, we observe a minimal advantage in \( EPOC_- \).

We also notice the following inequalities (see Figure C2 in Appendix C):
Figure 4. EPOC as a function of friends’ average opinion $x_{-i}$ across ideological groups, separated by movement type. Left panels represent time step $t_1 \rightarrow t_2$ and right panels represent time step $t_2 \rightarrow t_3$. Colored columns represent the value of $x_i$. Black dots plot $EPOC_+$ and blue stars plot $EPOC_-$. 
EPOC+ (x_i SL, x_i = L) < EPOC+ (x_i SL, x_i = SL);
EPOC+ (x_i SL, x_i = M) < EPOC+ (x_i M, x_i = SL);
EPOC+ (x_i M, x_i = M) < EPOC+ (x_i M, x_i = L);
EPOC+ (x_i C, x_i = M) > EPOC+ (x_i M, x_i = C);
EPOC+ (x_i M, x_i = SC) > EPOC+ (x_i SC, x_i = M);
EPOC+ (x_i C, x_i = SC) > EPOC+ (x_i SC, x_i = C).

(3)

Note that inequality EPOC+ (x_i L, x_i = M) < EPOC+ (x_i M, x_i = L) in equation (3) holds only for time step t_2 \rightarrow t_3 (others are true for both time steps).

**Influence of opinion diversity.** We have not found any clear trends in how the diversity of friends’ opinions affects the probability of opinion shift (see online supplementary materials).

### 6.4. Radicalization of opinions

To investigate how the process of opinion radicalization relates to users’ opinions and their neighborhood’s ideological composition, we concentrate on two cross-group transitions that definitively meet the definition of radicalization (Definition 3): L -> SL and C -> SC. They are plotted in Figure 5. We do not consider individuals with initially strong positions, as they have little to no room in the opinion space for radicalization. We furthermore do not attend to the radicalization of moderate individuals, as radicalization in these cases may be confused with political mobilization. Instead, we focus on identifying how environments can prompt individuals with clear political identities to become stronger in their views.

The curves have approximately the same character. For each, the probability of radicalization has two areas of relatively high values: (1) if x_i \in \{ SL, L \} (for Ls) or x_i \in \{ C, SC \} (for Cs) and (2) if x_i \in \{ C, SC \} (for Ls) or x_i = SL (for Cs). Minima are reached at x_i \approx 0.5. The radicalization of Ls is featured with the more pronounced minimum. This means that individuals radicalize more frequently (1) if they are exposed to similar views or more radical opinions with a similar bias or (2) when they observe clearly opposite opinions. Users are less prone to radicalization if they are exposed to moderate views. We also observe that conservatives are less prone to radicalization than liberals.

**Influence of opinion diversity.** After controlling for opinion diversity (see Figure C3 in Appendix C), we find that liberals embedded in networks with higher diversity (but a liberal or moderate average opinion) radicalize more frequently.
6.5. *Magnitude of opinion change*

In this subsection, we analyze how the magnitude of opinion change is connected to users’ and friends’ opinions. Figure 6 illustrates how the conditional opinion shift magnitude depends on $x_{-i}$ after controlling for $x_i$ and movement type.

**Positive shifts.** Figure 6 reveals that the magnitude of positive opinion shifts tends to be positively correlated with the value of $|x_i - x_{-i}|$: friends who are more distant in the opinion space from the focal user tend to induce larger opinion shifts toward their positions. Observing dependencies, we emphasize segments with an inverted U-shaped form.

**Negative shifts.** We report that, for all users except Ms, the global maxima are reached at the nearest edges of the opinion space. If we bring $x_{-i}$ away from $x_i$ toward the furthest edge, the magnitude tends to change through an inverted U-shaped form.

**Influence of opinion diversity.** We obtain no clear picture of how opinion diversity is connected to opinion shift magnitude (see online supplementary materials).

We also plot the expected magnitude of opinion shift (see Figure 7), which is obtained as the product of conditional magnitude and EPOC. We observe that expected magnitude inherits many features from Figure 6 (including linear and negative quadratic segments).\(^4\) In general, the tendency remains the same: larger

\(^4\)This observation does not contradict previously obtained results on EPOC and conditional magnitude – a puzzle may be because, after multiplication of two linear/quadratic functions, we should get a quadratic/cubic/tetrad function. The reason is that all curves that represent EPOC and conditional magnitude do not have a strictly linear/quadratic form.
Figure 6. Average conditional opinion shift magnitude as a function of $x_{i-j}$ across different values of $x_i$, separated by movement type. Left panels represent time step $t_1 \rightarrow t_2$ and right panels represent time step $t_2 \rightarrow t_3$. Colored areas represent the value of $x_i$. Black dots plot magnitudes of positive movements and blue stars plot magnitudes of negative movements.
Figure 7. Average expected opinion shift magnitude (conditional magnitude * EPOC) as a function of $x_{i,j}$ across different values of $x_i$, separated by movement type. Left panels represent time step $t_1 \rightarrow t_2$ and right panels represent time step $t_2 \rightarrow t_3$. Colored areas represent the value of $x_i$. Black dots plot magnitudes of positive movements and blue stars plot magnitudes of negative movements.
values of \(|x_i - x_{-i}|\) are associated with more distant opinion shifts (for both positive and negative shifts).

6.6. Positive or negative?

The final part of our analysis concentrates on the following question: “Given we know that an individual will make a (remarkable) movement, will it be positive or negative?” This question has the following motivation. Assume that positive and negative shifts largely stem from the interplay between positive influence and negative influence. How can we predict which component will operate for a given user (if they will make a remarkable movement)? From a theoretical perspective, as we previously mentioned in Subsection 6.2, one should expect that if the distance between users’ and friends’ opinions is not too large, a positive influence is most likely. When \(|x_i - x_{-i}|\) becomes large enough, positive influence gives way to negative influence, meaning we can expect a negative shift to be more likely.

We plot the ratio of the number of positive movements to the number of negative movements as a function of \(x_{-i}\) across different values of \(x_i\) (see Figure 8). It is clear that, if one increases the value of \(|x_i - x_{-i}|\), the positive/negative ratio persistently features an inverted U-shaped form, indicating that positive shifts have a (relatively) higher chance to occur if the distance \(|x_i - x_{-i}|\) is not too large and not too small.

7. Discussion

An immediate conclusion from our findings is that users’ most popular “strategy” is not to change their positions. The general opinion trend is to drift in the left direction in the opinion space. As a result, the population of individuals located on the left side of the opinion space increases whereas number of individuals located in the middle and on the right side decreases. Exceptions are SCs whose population seems to fluctuate. It either declines \((t_1 \rightarrow t_2)\) or increases \((t_2 \rightarrow t_3)\) with a relatively small speed. In a nutshell, we observe “asymmetrical” opinion polarization. The drastic difference in user activity (measured as the number of remarkable movements) between time steps \(t_1 \rightarrow t_2\) and \(t_2 \rightarrow t_3\) may be explained by the political context: the presidential election was conducted between times \(t_1\) and \(t_2\).

How the probability and the average magnitude of a positive shift depend on users’ and friends’ opinions (precisely, on the distance between them) largely aligns with the linear and moderated (positive) assumptions. Our findings on the magnitudes of positive shifts are in line with existing empirical studies. The size of the dataset has given us an opportunity to make a quite precise characterization of how the chance of a positive shift depends on opinion discrepancy \(|x_{-i} - x_i|\), a result that is rarely obtained in empirical
Figure 8. Positive/negative ratio as a function of $x_i$, separated by $x_i$. Left panels represent time step $t_1 \rightarrow t_2$ and right panels represent time step $t_2 \rightarrow t_3$. Colored areas represent the value of $x_i$. 
studies and which challenges the hypothesis that larger values of $|x_{-i} - x_i|$ should decrease the chance of assimilative influence. Our results on how the probability of a positive shift varies by users’ and friends’ opinion strength (inequalities in equation (3)) also match the theory and can be understood as follows: more radical individuals are less susceptible to peer influence than those with weaker positions. This explains (partially, at least – we should not forget about negative shifts and, of course, selectivity) both the structure of homophily (Table 3) and the map of cross-group opinion shifts (Table 5): individuals with strong opinions frequently attract their peers, inducing cascades of positive shifts. As a result, these individuals tend to accumulate a nearby concentration of individuals with similar views. Note that the dynamics of connections between users (that are beyond the scope of this article) can also explain such patterns of homophily. For example, radical individuals are highly engaged in political discourse. As such, two radical users that follow similar information sources (and thus having similar estimated opinions) are highly likely to create a friendship connection through mutual conversations.

The inequalities in equation (3) are quite non-trivial from the perspective of geometrical intuition. Let us consider, for example, the following inequality:

$$EPOC_{+}(x_i = SL, x_{-i} = L) < EPOC_{+}(x_i = L, x_{-i} = SL).$$

This means that the probability that a strong liberal user will move right (given their friends are, on average, liberals) is less than the probability that a liberal user will move left (given their friends are, on average, strong liberals). What is non-trivial here is the fact that, by default, SLs have far more space to move right than Ls do to move left. As a result, one might expect $EPOC_{+}(x_i = SL, x_{-i} = L)$ to be greater.

It is worthwhile to note that the role of assimilative influence in positive shifts is not entirely clear, as other effects may also contribute to positive shifts.

**Example 1.** The presence of a common stimulus (Aral & Nicolaides, 2017) is a potential confounder that may be confused with positive influence. Befriended users tend to follow similar information sources and, in turn, be influenced in a similar way; this effect is not accounted for in this study, as we assume that information sources serve only as indicators of individuals’ opinions – not the sources. As a result, we may observe a positive shift stemming from the non-simultaneous reaction of befriended individuals to the same stimuli (simultaneous reaction is controlled for by only considering individuals whose friends’ average opinion shifts by less than 0.05).

Other examples of potential confounders that could contribute to the observed positive shifts are selectivity, shifts in the political bias of information sources, and the assumption of static ties (see Kozitsin, 2020).
Formal models of opinion formation usually constrain positive movements to be non-skipping – a newly formed opinion must lie in between the former opinion and the opinion of the influence source (non-skipping movement) and must not skip over the influence source opinion (skipping movement). Despite failing to disentangle skipping and non-skipping movements (see Appendix A), we advance the following potential explanation for skipping movements based on the profound nature of the considered process.

**Example 2 (explanation of skipping movements).** Let us consider agent i making shift $M \rightarrow SC$ in the opinion space. This shift is essentially driven by some change in agent i’s subscriptions, presumably as a result of influence from a friend, source j. Let us assume that source j follows public pages $P_1$, $P_2$, and $P_3$: the first two ($P_1$ and $P_2$) with a conservative bias and the third ($P_3$) with a strong conservative bias. Naturally, source j’s estimated opinion should be rather conservative. If agent i starts to follow only $P_3$ (perhaps influenced by source j’s repost from $P_3$), their opinion may be estimated as strongly conservative. As a result, we observe a skip in the opinion space performed by agent i over source j’s opinion.

However, alongside positive shifts, we also observe negative shifts that constitute a significant fraction of all remarkable movements. Our analysis reveals that negative shifts have several interesting properties. Figure 4 serves as clear evidence that SLs and SCs are associated with remarkable rates of negative shifts among their peers. On the one hand, this observation is in line with the theory – the furthest opinions are the most repulsive. However, the theory suggests that negative influences should have a minimal likelihood if friends’ opinions are close to those of the focal user. In contrast, we observe minima at $x_{-i}$ that differ from – but are close to – $x_i$. For all individuals except Ms, this may be attributed to geometrical constraints. For example, Cs apriori have more space to move left than to move right. As such, the lowest rate of negative movements for such individuals should be observed at $x_{-i} < C$. For Cs, the minima are observed at $x_{-i} = M$, the nearest leftward position. For Ms, we observe a clear divergence between theory and empirics that cannot be explained by geometrical constraints (Ms are located strictly in the middle, so; both directions should have equal priority). Nonetheless, magnitudes of negative opinion shifts tend to follow the theory (the moderated assumption). Global maxima in Figure 6 at $x_{-i}$, observed at the nearest edges (which violates the theory), may be explained by geometrical intuition: the magnitude of a negative shift likely has the highest value when an individual directs it toward the farthest edge.

How the chance and magnitude of a negative shift are related to opinion divergence is somewhat non-trivial (in a way that, importantly, largely matches the theoretical expectations); it may be considered as an argument in favor of repulsive influence. If repulsive influence is really the cause of
negative shifts, Ms’ behavior clearly indicates a relatively high rate of repulsive influence induced by individuals with a conservative bias. This observation may be attributed to political affairs (in the spirit of ideas presented by Böttcher & Gersbach, 2020). For instance, it may constitute a display of citizens’ negative attitudes toward Russian pension reform (and, correspondingly, those who support the government that launched this reform) in 2018\textsuperscript{5}, when this data was gathered.

More arguments in favor of the negative influence hypothesis are that (1) the negative influence is theorized to be more likely when observing attitude – rather than belief – dynamics (exactly our case) and (2) “there are psychological theories that predict that individuals develop increasingly positive views on a political candidate when they learn that members of the opposite social category dislike that person” (Mäs, 2019). One critical point against the negative influence hypothesis is that “negative influence may be very unlikely when individuals do not communicate face-to-face as students do in a classroom, but in a computer-mediated setting like a comment board on the Internet” (Mäs, 2019).

We must note, however, that observed patterns of negative shifts may be partially explained in a different way, without referring to the notion of repulsive influence.

**Example 3 (explanation of patterns of negative shifts without negative influence).** Here, we provide an example of how to explain negative shifts among individuals in friendships with users holding strong positions. Let us assume that radical positions are associated with some degree of discomfort such that some users holding them are inclined to tone down their opinions. This discomfort could be driven by various elements of political affairs, such as political talk shows (Petrov & Proncheva, 2020). As a result, individuals with strong positions are pushed to reconsider their views. Additionally, let us suppose that estimations of users’ opinions are not fully correct (an assumption which is quite realistic). Befriended individuals tend to have similar real views. As a result, we should expect the following regularity: the more friends that agent i has with a particular position $G \in \{ SL, L, M, C, SC \}$, the higher the chance that agent i has the same real position (regardless of agent i’s estimated position). If $G = SL$ or $G = SC$, there is a higher chance of individuals toning down their real opinion; in turn, this may lead to subscriptions to information sources that shift agent i’s estimated opinion away from their friends’ (estimated) opinions.

The next question is one of the balance between positive and negative shifts. Our results indicate that, for a given individual with opinion $x_i$, left-located opinion $x_{-i} < x_i$ is more likely to be associated with a positive shift than with a negative shift. Interestingly, if a (remarkable) movement has occurred for an individual with opinion $x_i$, that movement is highly likely to be positive if the

\textsuperscript{5}https://www.bbc.com/news/world-europe-45342721
value of opinion divergence $|x_i - x_{-i}|$ is not too small and not too large. The prevalence of negative shifts for high $|x_i - x_{-i}|$ values is consistent with the theory; their prevalence if $x_i$ and $x_{-i}$ are close may be attributed, for example, to users striving for uniqueness (Mäs et al., 2010). In fact, this result gives another perspective on those of Kozitsin (2020), who found that, after controlling for user opinion, the probability of shifting left varies with the average opinion of the user’s friends $x_{-i}$, and the target function has a near-sinusoidal wave-shaped form whereby if source opinion $x_{-i}$ is less than user opinion $x_i$, the probability is greater than that for $x_{-i} > x_i$.

So far, we have not focused on the very important issue of how the absence of challenging opinions in the user’s neighborhood (or, simply, opinion isolation) affects opinion radicalization. Scholars assert that individuals with no access to discrepant opinions are more prone to radicalization. Such problem arises especially frequently in the context of so-called echo-chambers – cohesive groups of individuals holding similar positions⁶ (Del Vicario et al., 2017). However, the opinion-formation mechanisms considered in this paper cannot capture the situation in which individuals surrounded by those with the same opinions radicalize; in such a case, under any of our core assumptions, individuals’ opinions will be stable. To obtain polarization stemming from communication between individuals with similar opinions, one could, for example, clarify how these individuals communicate by specifying arguments in the model (Mäs et al., 2013). If the theory that opinion isolation reinforces radicalization is correct, we must expect individuals to radicalize more frequently when exposed to similar (or more radical yet having the same bias) opinions.

We have found mixed support for the theory that opinion isolation reinforces radicalization. More specifically, we have demonstrated that individuals have a high likelihood of radicalizing if they (1) are surrounded by individuals with similar or stronger views yet having the same bias or (2) are exposed to opposing radical views. After observing any kind of radical opinion – liberal or conservative – individuals tend to radicalize with more enthusiasm. The proliferation of OSNs may lead to opinion isolation, as OSNs are driven by algorithms that are supposed to accommodate communication between like-minded individuals (Bakshy et al., 2015; Geschke et al., 2019; Kozyreva et al., 2020; Mäs & Bischofberger, ; Perra & Rocha, 2019; Rossi et al., 2019). If opinion isolation does truly drive radicalization, then personalization systems likely contribute to opinion radicalization (that in turn may lead to opinion polarization). Our results only partially confirm these ideas. We argue that the most efficient way to mitigate radicalization is to avoid exposing individuals to radical opinions, instead exposing them only to moderate positions.

⁶Note that an individual whose friends’ opinions are not different from their own is not necessarily in an echo-chamber. To ensure this, one must additionally claim that their friends also tend to be isolated, and friends of their friends are as well, and so on.
8. Conclusion

This paper presents an empirical study of the opinion dynamics of a large-scale sample of online social network users. Users’ opinions were estimated via their subscriptions to information sources, and we have analyzed how friendship connections affect the dynamics of these estimations.

Most existing empirical studies suggest that the formation of continuous opinions tends to follow assimilative influence in linear or sometimes negative quadratic (with respect to the preexisting difference in opinions between an individual and their acquaintance) forms (Friedkin et al., 2021; Moussaïd et al., 2013; Takács et al., 2016). Studies that reported a negative influence either have methodological concerns (Knippenberg et al., 1990; Mazen & Leventhal, 1972) or are based on natural experiments in which researchers are unable to control for all possible confounding factors (Liu & Srivastava, 2015).

In this paper, we distinguish between positive (toward friends’ opinions) and negative (away from friends’ opinions) opinion shifts. We find that the existence and magnitude of positive opinion shifts are positively related (largely through a linear form or an inverted U-shaped form) to the degree of divergence in opinions between a user and their friends. Interestingly, we find that both the chance and magnitude of negative shifts increase alongside the opinion divergence between the focal user and their friends (largely through, again, a linear form or an inverted U-shaped form). In other words, larger values of opinion discrepancy between the target user and their friends lead to larger opinion shifts (both positive and negative) that occur more frequently. The balance between positive and negative shifts is also moderated by opinion divergence: if the opinions of the focal user and their friends are too similar or dissimilar, there is a relatively low chance of a positive shift.

These results challenge existing concerns about the presence of a negative influence; they constitute a clue for the context in which researchers should search for such an influence. Our findings indicate that scholars should pay more attention to modeling not only how individuals modify their opinions, but also when they do so. We demonstrate, as well, how to explain our empirical results without referring to the notions of positive and negative influence. Concurrently, we investigate how the process of radicalization is connected to the ideological heterogeneity of users’ neighborhoods. We argue that individuals are less likely to radicalize if they are exposed to moderate positions; in contrast, radical views – regardless of their bias – induce a relatively high degree of radicalization.

One crucial aspect that was not identified in this research is how users create and delete ties with each other. The evolution of the social graph has a potentially significant impact on opinion dynamics as its structure determines largely how information flows between users. A greater understanding as to how individuals
form connections at the micro-level can improve our knowledge of social graph formation processes (M. E. J. Newman, 2001; Snijders, 2017). Additionally, it would be interesting to control for users’ demographic characteristics, such as age, gender, and location, as these characteristics may affect how influence processes unfold on OSNs (Kovanen et al., 2013). For example, individuals with similar characteristics likely have more communication contacts. Additionally, age and gender may control for sensitivity to peer influence (Peshkovskaya et al., 2019).

The findings presented in this research shed some additional light on the problem of linking theoretical models of opinion formation with real data. However, our findings should not be overestimated due to a large number of dataset limitations. From this perspective, it would be interesting to evaluate the extent to which these limitations may affect the outcome that is visible to an observer (who analyzes the data). One could introduce a model in which agents form connections with each other and subscribe to information sources. These sources would have their own political biases and agents would decide whether to follow a particular information account by comparing this bias with their own opinion and considering how many of their friends follow this account (in the fashion of complex contagion models (Centola, 2010, 2019; Christakis & Fowler, 2013; Monsted et al., 2017; Ugander et al., 2012)). Of course, it would be beneficial to allow agents to follow information sources whose biases differ from agents’ views. In this model, an observer would estimate agents’ opinions using their subscriptions to information sources with some error (also a parameter that could be varied) and then analyze these estimations, as we have done in this paper. By considering the different rules as to how agents communicate with each other (including the impact of personalization algorithms) and how they form connections (with other agents and information sources), we can obtain different outcomes and compare them against the real data.

**Acknowledgments**

The authors are grateful to anonymous reviewers for their invaluable comments.

**Funding**

This work was supported by the Russian Foundation for Basic Research, project no. 18-29-22042.

**ORCID**

Ivan V. Kozitsin http://orcid.org/0000-0002-2638-0131
References

A. Petrov, O. Proncheva. Identifying the Topics of Russian Political Talk Shows // Proceedings of the Conference on Modeling and Analysis of Complex Systems and Processes 2020, October 22–24 (MACSPro 2020). pp. 79–86. CEUR-WS.org, online. http://ceur-ws.org/Vol-2795/short1.pdf

Abelson, R. P. (1964). Mathematical models of the distribution of attitudes under controversy. Contributions to Mathematical Psychology, 142–160.

Abid, O., Jamoussi, S., & Ayed, Y. B. (2018). Deterministic models for opinion formation through communication: A survey. Online Social Networks and Media, 6, 1–17. https://doi.org/10.1016/j.osnem.2018.02.002

Altafini, C. (2013). Consensus problems on networks with antagonistic interactions. IEEE Transactions on Automatic Control, 58(4), 935–946. https://doi.org/10.1109/TAC.2012.2224251

Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. Nature Communications, 8(1), 14753. https://doi.org/10.1038/ncomms14753

Asch, S. E. (1961). Effects of group pressure upon the modification and distortion of judgments. In Documents of gestalt psychology (pp. 222–236). University of California Press. https://doi.org/10.1525/9780520301351-017

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Banisch, S., & Olbrich, E. (2019). An Argument Communication Model of Polarization and Ideological Alignment. ArXiv:1809.06134 [Physics]. http://arxiv.org/abs/1809.06134

Barberá, P. (2014). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. Political Analysis, 23(1), 76–91. https://doi.org/10.1093/pan/mpu011

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295. https://doi.org/10.1038/nature11421

Böttcher, L., & Gersbach, H. (2020). The great divide: Drivers of polarization in the US public. EPJ Data Science, 9(1), 1–13. https://doi.org/10.1140/epjds/s13688-020-00249-4

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. Reviews of Modern Physics, 81(2), 591. https://doi.org/10.1103/RevModPhys.81.591

Centola, D. The spread of behavior in an online social network experiment. (2010). Science, 329 (5996), 1194–1197. Scopus. https://doi.org/10.1126/science.1185231

Centola, D. (2019). Influential networks. Nature Human Behaviour, 3(7), 664–665. https://doi.org/10.1038/s41562-019-0607-5

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. Statistics in Medicine, 32(4), 556–577. https://ssrn.com/abstract=2974413 or

Clemm von Hohenberg, Bernhard and Más, Michael and Pradeski, Bary, Micro Influence and Macro Dynamics of Opinion Formation (October 29, 2017). Available at SSRN: https://ssrn.com/abstract=2974413 or https://doi.org/10.2139/ssrn.2974413

Dandekar, P., Goel, A., & Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. Proceedings of the National Academy of Sciences, 110(15), 5791–5796. https://doi.org/10.1073/pnas.1217220110

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. Advances in Complex Systems, 3, 87–98. https://doi.org/10.1142/S0219525900000078

DeGroot, M. H. (1974). Reaching a consensus. Journal of the American Statistical Association, 69(345), 118–121. https://ssrn.com/abstract=2974413 or
Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific Reports*, 7(1), 40391. https://doi.org/10.1038/srep40391

Flache, A., & Mäs, M. (2008). How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams. *Computational and Mathematical Organization Theory, 14*(1), 23–51. https://doi.org/10.1007/s10588-008-9019-1

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Defuant, G., Huet, S., & Lorenz, J. (2017). Models of Social Influence: Towards the Next Frontiers. *Journal of Artificial Societies & Social Simulation, 20*(4), 4. https://doi.org/10.18564/jasss.3521

Frey, D. (1986). Recent research on selective exposure to information. *Advances in Experimental Social Psychology*, 19, 41–80. https://doi.org/10.1016/S0065-2601(08)60212-9

Friedkin, N. E., & Bullo, F. (2017). How truth wins in opinion dynamics along issue sequences. *Proceedings of the National Academy of Sciences, 114*(43), 11380–11385. https://doi.org/10.1073/pnas.1710603114

Friedkin, N. E., Proskurnikov, A. V., & Bullo, F. (2021). Group dynamics on multidimensional object threat appraisals. *Social Networks, 65*, 157–167. https://doi.org/10.1016/j.socnet.2020.12.009

Friedkin, N. E., Proskurnikov, A. V., Tempo, R., & Parsegov, S. E. (2016). Network science on belief system dynamics under logic constraints. *Science, 354*(6310), 321–326. https://doi.org/10.1126/science.aag2624

Gaisbauer, F., Olbrich, E., & Banisch, S. (2020). Dynamics of opinion expression. *Physical Review E, 102*(4), 042303. https://doi.org/10.1103/PhysRevE.102.042303

Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modeling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology, 58*(1), 129–149. https://doi.org/10.1111/bjso.12286

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation, 5*(3). https://www.jasss.org/5/3/2.html

Holme, P., & Newman, M. E. J. (2006). Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E, 74*(5), 056108. https://doi.org/10.1103/PhysRevE.74.056108

Kerckhove, C. V., Martin, S., Gend, P., Rentfrow, P. J., Hendrickx, J. M., Blondel, V. D., & Sánchez, A. (2016). Modelling Influence and Opinion Evolution in Online Collective Behaviour. *PLOS ONE, 11*(6), e0157685. https://doi.org/10.1371/journal.pone.0157685

Knippenberg, D. V., Vries, N. D., & Knippenberg, A. V. (1990). Group status, group size and attitude polarization. *European Journal of Social Psychology, 20*(3), 253–257. https://doi.org/10.1002/ejsp.2420200306

Kovanen, L., Kaski, K., Kertész, J., & Saramäki, J. (2013). Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences, 110*(45), 18070–18075. https://doi.org/10.1073/pnas.1307941110

Kozitsin, I. V. (2020). Formal models of opinion formation and their application to real data: Evidence from online social networks. *The Journal of Mathematical Sociology, 1–28*. https://doi.org/10.1080/0022250X.2020.1835894

Kozitsin, I. V., Chkhartishvili, A. G., Marchenko, A. M., Norkin, D. O., Osipov, S. D., Uteshev, I. A., Goiko, V. L., Palkin, R. V., & Myagkov, M. G. (2020). Modeling Political Preferences of Russian Users Exemplified by the Social Network Vkontakte. *Mathematical Models and Computer Simulations, 12*(2), 185–194. https://doi.org/10.1134/S2070048220020088
Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. (2020). Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the US. PsyArXiv. https://doi.org/10.31234/osf.io/3q4mg

Krizan, Z., & Baron, R. S. (2007). Group polarization and choice-dilemmas: How important is self-categorization? European Journal of Social Psychology, 37(1), 191–201. https://doi.org/10.1002/ejsp.345

Kurahashi-Nakamura, T., Mäs, M., & Lorenz, J. (2016). Robust clustering in generalized bounded confidence models. Journal of Artificial Societies and Social Simulation, 19(4), 4. https://doi.org/10.18564/jasss.3220

Liu, C. C., & Srivastava, S. B. (2015). Pulling Closer and Moving Apart: Interaction, Identity, and Influence in the U.S. Senate, 1973 to 2009. American Sociological Review, 80(1), 192–217. https://doi.org/10.1177/0003122414564182

Mäs, M. (2019). Challenges to Simulation Validation in the Social Sciences. A Critical Rationalist Perspective. In C. Beisbart & N. J. Saam (Eds.), Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives (pp. 857–879). Springer International Publishing. https://doi.org/10.1007/978-3-319-70766-2_35

Mäs, M., Flache, A., Helbing, D., & Bergstrom, C. T. (2010). Individualization as driving force of clustering phenomena in humans. PLoS Computational Biology, 6(10), 10. https://doi.org/10.1371/journal.pcbi.1000959

Mäs, M., Flache, A., & Perc, M. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. PloS One, 8(11), e74516. https://doi.org/10.1371/journal.pone.0074516

Mazen, R., & Leventhal, H. (1972). The influence of communicator-recipient similarity upon the beliefs and behavior of pregnant women. Journal of Experimental Social Psychology, 8(4), 289–302. https://doi.org/10.1016/0022-1031(72)90019-4

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mønsted, B., Sapieżyński, P., Ferrara, E., Lehmann, S., & Lambiotte, R. (2017). Evidence of complex contagion of information in social media: An experiment using Twitter bots. PLOS ONE, 12(9), e0184148. https://doi.org/10.1371/journal.pone.0184148

Moussaid, M., Kämmer, J. E., Analytis, P. P., Neth, H., & Szolnoki, A. (2013). Social Influence and the Collective Dynamics of Opinion Formation. PLOS ONE, 8(11), e78433. https://doi.org/10.1371/journal.pone.0078433

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. Physical Review E, 64(2), 025102. https://doi.org/10.1103/PhysRevE.64.025102

Newman, M. E. J. (2018). Networks. Oxford university press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), 175. https://doi.org/10.1037/1089-2680.2.2.175

Perra, N., & Rocha, L. E. (2019). Modelling opinion dynamics in the age of algorithmic personalisation. Scientific Reports, 9(1), 1–11. https://doi.org/10.1038/s41598-019-43830-2

Peshkovskaya, A., Babkina, T., & Myagkov, M. (2019). Gender effects and cooperation in collective action: A laboratory experiment. Rationality and Society, 31(3), 337–353. https://doi.org/10.1177/104346319858788

Proskurnikov, A. V., & Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. Annual Reviews in Control, 43, 65–79. https://doi.org/10.1016/j.arconstr.2017.03.002

Proskurnikov, A. V., & Tempo, R. (2018). A tutorial on modeling and analysis of dynamic social networks. Part II. In Annual Reviews in Control.
Appendix

Appendix A. Testing Interrelation Between Ideological Position, Age, and Number of Friends

We focus on two public pages on VK that are official media accounts: “Медуза” (Meduza) and “Новости RT на русском” (RT). RT is a state-controlled Russian media outlet that is often accused of being the “Kremlin’s propaganda tool.” In contrast, Meduza is an online newspaper that is not controlled by the state; in fact, according to state-controlled outlets, Meduza generally opposes the current Russian government. Recently, Meduza has been given the status of “foreign agent” in the Russian Federation. If the estimations of users’ opinions are correct, we should expect that an individual following only RT is estimated as one with conservative polarity $x_{RT} > 0.5$ while an individual following only Meduza is estimated as one with liberal polarity $x_{M} < 0.5$. To check this, we create two artificial users that meet the above requirements and estimate their opinions. We find that the user following Meduza has $x_{M} = 0.38$ while the one following RT has $x_{RT} = 0.56$; these results confirm our expectations.

Next, we download a new, independent snapshot of Meduza and RT followers. We consider only active users (at least one platform interaction per month) who are Russians of at least 18 years of age with open privacy settings. We collect information on 117,736 users following only

---

Proskurnikov, A. V., Tempo, R., Cao, M., & Friedkin, N. E. (2017). Opinion evolution in time-varying social influence networks with prejudiced agents. *IFAC-PapersOnLine, 50*(1), 11896–11901. https://doi.org/10.1016/j.ifacol.2017.08.1424

Ravazzi, C., Dabbene, F., Lagoa, C., & Proskurnikov, A. V. (2020). Learning hidden influences in large-scale dynamical social networks: A data-driven sparsity-based approach. *ArXiv:2007.06713 [Cs, Eess]. http://arxiv.org/abs/2007.06713*

Rossi, W. S., Polderman, J. W., & Frasca, P. (2019). The closed loop between opinion formation and personalised recommendations. *ArXiv:1809.04644 [Physics]. http://arxiv.org/abs/1809.04644*

Sherif, M., & Hovland, C. I. (1961). Social judgment: Assimilation and contrast effects in communication and attitude change. Yale Univer. Press.

Snijders, T. A. B. (2017). Stochastic Actor-Oriented Models for Network Dynamics. *Annual Review of Statistics and Its Application, 4*(1), 343–363. https://doi.org/10.1146/annurev-statistics-060116-054035

Takács, K., Flache, A., Más, M., & Amblard, F. (2016). Discrepancy and Disliking Do Not Induce Negative Opinion Shifts. *PLOS ONE, 11*(6), e0157948. https://doi.org/10.1371/journal.pone.0157948

Tang, T., & Chorus, C. G. (2019). Learning Opinions by Observing Actions: Simulation of Opinion Dynamics Using an Action-Opinion Inference Model. *Journal of Artificial Societies and Social Simulation, 22*(3), 2. https://doi.org/10.18564/jasss.4020

Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012). Structural diversity in social contagion. *Proceedings of the National Academy of Sciences, 109*(16), 5962–5966. http://ceur-ws.org/Vol-2795/short1.pdf
RT, 84,283 users following only Meduza, and 3,656 users who follow both accounts. We observe that Meduza followers are, on average, younger and have more friends than RT followers (see Figure A1). However, we find a non-trivial dependency between user age and number of friends within follower groups. For those who follow only RT, there is a weak negative correlation (Pearson correlation coefficient equal to −0.078, p-value near zero). For those who follow only Meduza, there is a weak positive correlation (0.096, p-value near zero). For those who follow both Meduza and RT, there is an insignificant (p-value equal to 0.09) and negligible negative correlation (−0.02). This suggests that our initial hypothesis holds only for conservative individuals; liberal users with more friends are generally older than those with fewer friends. This observation may be explained as follows. From Figure A1, we can see that the audience of Meduza is dominated by persons under 40. Recalling that VK was established in 2006, we can hypothesize that it should have the most popularity among those who were young in the mid-2000s. However, young people today may divert their attention to newer platforms (e.g., TikTok, Clubhouse); thus, they may not be relatively active on VK (but still more active than people over 40, who constitute a remarkable portion of RT followers).

Appendix B. Skipping and Non-skipping Movements

To disentangle the nature of skipping (just 9.9% of all positive movements for time step $t_1 \rightarrow t_2$ and 7% for time step $t_2 \rightarrow t_3$) and non-skipping movements, we first investigate how the value of $\sigma_{-i}$ differs between those who made skipping movements and those who made non-skipping movements. The intuition behind this is that skips may be attributed to high values of opinion diversity among friends (see Figure 3, panel c). However, we find a negligible difference (in both cases, the average of $\sigma_{-i}$ is approximately 0.141). We then plot the probabilities of making each of 25 possible (remarkable) movements in the opinion space (accounting for static movements but not imposing the requirement to be remarkable on them) as functions of $x_{-i}$, aiming to investigate how the character of a curve behaves while transitioning from the area of skipping movements to the area of nonskipping movements. If skipping movements are somewhat inadvertent, we should expect a sound change in behavior (presumably, growth with
an increasing rate). Unfortunately, the dependencies (presented in Figure B1 and in the online supplementary materials) do not enable us to assert that skipping movements have a strictly different nature from non-skipping movements.

Figure B1. These subplots represent how the probability of a particular opinion shift depends on the average opinion of friends (here, we consider time step $t_1 \rightarrow t_2$; the same subplots for time step $t_2 \rightarrow t_3$ are included in the online supplementary materials). The dashed lines illustrate the boundaries between areas of negative (dark green), positive skipping (light green) and positive nonskipping (lime green) movements. Some subplots demonstrate that the curves do change their behavior after transitioning from the area of skipping movements to the area of nonskipping movements (see, for example, SL -> C or SC -> M whereby the curves start to grow with a higher rate after entering the nonskipping zone). However, similar behavior (growth with an increasing rate) may be observed for movements that imply no skips (C -> M, SC -> C).

Appendix C. Additional Figures and Tables
In this plot, we demonstrate how $EPOC(t_k \rightarrow t_{k+1})$ is moderated by the value of $|x_{i}(t_{k+1}) - x_{i}(t_k)|$. Left panels represent time step $t_1 \rightarrow t_2$ and right panels represent time step $t_2 \rightarrow t_3$. Blue stars plot $EPOC_-(t_k \rightarrow t_{k+1})$ if $|x_{i}(t_{k+1}) - x_{i}(t_k)|$ is under 0.05; purple dots plot $EPOC_-(t_k \rightarrow t_{k+1})$ if $|x_{i}(t_{k+1}) - x_{i}(t_k)| \geq 0.05$. Black stars plot $EPOC_+(t_k \rightarrow t_{k+1})$ if $|x_{i}(t_{k+1}) - x_{i}(t_k)|$ is under 0.05; green dots plot $EPOC_+(t_k \rightarrow t_{k+1})$ if $|x_{i}(t_{k+1}) - x_{i}(t_k)| \geq 0.05$. The tendency is that higher values of $|x_{i}(t_{k+1}) - x_{i}(t_k)|$ are correlated with a higher chance of an opinion shift.
**Table C1** Average ideological composition of neighborhoods across different ideological groups among users with fewer than five friends (at time moment $t_2$).

| Ideological group | SL  | L   | M   | C   | SC  |
|-------------------|-----|-----|-----|-----|-----|
| SL                | 0.14| 0.22| 0.44| 0.15| 0.05|
| L                 | 0.09| 0.21| 0.51| 0.15| 0.04|
| M                 | 0.07| 0.18| 0.53| 0.18| 0.05|
| C                 | 0.06| 0.15| 0.46| 0.24| 0.09|
| SC                | 0.06| 0.13| 0.40| 0.27| 0.14|
| Null model        | 0.08| 0.19| 0.53| 0.16| 0.04|

**Table C2** Average ideological composition of neighborhoods across different ideological groups among users with fewer than 25 friends and more than or equal to five friends (at time moment $t_2$).

| Ideological group | SL  | L   | M   | C   | SC  |
|-------------------|-----|-----|-----|-----|-----|
| SL                | 0.15| 0.24| 0.47| 0.12| 0.03|
| L                 | 0.10| 0.22| 0.52| 0.13| 0.03|
| M                 | 0.08| 0.20| 0.54| 0.15| 0.03|
| C                 | 0.07| 0.17| 0.50| 0.20| 0.05|
| SC                | 0.07| 0.16| 0.47| 0.23| 0.08|
| Null model        | 0.08| 0.19| 0.53| 0.16| 0.04|

**Table C3** Average ideological composition of neighborhoods across different ideological groups among users with more than or equal to 25 friends (at time moment $t_2$).

| Ideological group | SL  | L   | M   | C   | SC  |
|-------------------|-----|-----|-----|-----|-----|
| SL                | 0.17| 0.27| 0.47| 0.08| 0.02|
| L                 | 0.13| 0.26| 0.50| 0.09| 0.02|
| M                 | 0.11| 0.23| 0.54| 0.11| 0.02|
| C                 | 0.09| 0.20| 0.53| 0.15| 0.03|
| SC                | 0.08| 0.20| 0.51| 0.17| 0.04|
| Null model        | 0.08| 0.19| 0.53| 0.16| 0.04|
Figure C2. EPOC_+ as a function of x_i, separated by x_i. Upper panels represent time step t_1 → t_2; bottom panels represent time step t_2 → t_3. We emphasize the following regularity: for two groups, G_1 and G_2, where without loss of generality G_1 < G_2, it tends to hold that EPOC_+(x_i = G_1, x_{−i} = G_2) < EPOC_+(x_i = G_2, x_{−i} = G_1) if G_2 ≤ M and EPOC_+(x_i = G_1, x_{−i} = G_2) > EPOC_+(x_i = G_2, x_{−i} = G_1) if G_1 ≥ M. Exceptions here are: (1) Ls and Ms for time step t_1 → t_2 and (2) Ms and Cs for both time steps.
Figure C3. Probability of opinion shifts $L \rightarrow SL$ and $C \rightarrow SC$ as functions of $x_{i,j}$, separated by $\sigma_{i,j}$. Upper panels represent time step $t_1 \rightarrow t_2$ and bottom panels represent time step $t_2 \rightarrow t_3$. Liberals embedded in networks with higher diversity (but a liberal or moderate average opinion) radicalize more frequently.