frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 20 November 2025
DOI 10.3389/fphy.2025.1706465

OPEN ACCESS

EDITED BY
Salvatore Micciche',
University of Palermo, Italy

REVIEWED BY
Juan De Gregorio,
Spanish National Research Council (CSIC), Spain
Vito Domenico Pietro Servedio,
Complexity Science Hub Vienna (CSH), Austria

*CORRESPONDENCE
Xiaoming J. Zhang,
zhangxiaoming@bimsa.cn

RECEIVED 16 September 2025
REVISED 18 October 2025
ACCEPTED 21 October 2025
PUBLISHED 20 November 2025

CITATION
Zhang XJ, Hu Y and Zhang Y (2025) An affinity-based opinion dynamics model for the evolving pattern of political polarization. *Front. Phys.* 13:1706465.
doi: 10.3389/fphy.2025.1706465

An affinity-based opinion dynamics model for the evolving pattern of political polarization

Xiaoming J. Zhang[1,2]*, Yuzhong Hu[1] and Yiming Zhang[1]

[1]Group of Artificial Intelligence and Machine Learning, Beijing Institute of Mathematical Sciences and Applications, Beijing, China, [2]Institute for Applied Mathematics, Tsinghua University, Beijing, China

Political polarization has attracted many studies in recent years. We developed an opinion dynamics model with affective homophily effect and national social norm effect to describe this phenomenon. The time evolution of the polarization between the two parties and the spread of opinions within each party are affected by three factors: the repulsive effect between the two parties, the attractive and repulsive effects between the members in each party, and the national social norm effect that pulls the opinions of all members towards a common norm. The model is internally consistent and is applied to the simulation of the symmetric patterns of polarization and spread of the opinion distributions in the U.S. Congress, and the results align well with 154 years of recorded data. The time evolution of the strength of the national social norm effect is obtained and is consistent with the important historical events that occurred during the past 150 years.

KEYWORDS

strength of social norm effect, polarization, spread, opinion dynamics, ideology, affective homophily, U.S. Congress

1 Introduction

The political landscape of the United States has become increasingly polarized over the past four decades [1]. Researchers have suggested that while polarization is undermining democracy and legislative effectiveness, understanding the phenomenon could ultimately reveal strategies for bridging the divides [2–4]. However, despite past efforts in studying political polarization, most analytical models are fundamentally limited in capturing both the interactions among Congress members and the historical context. It is therefore appealing for us to raise the following questions: What drives the separation and reunion of ideologies in the U.S. political system? Is there a relationship between partisan polarization and critical historical events?

Political polarization arises as the spectrum of public opinion fractures or as differences in perspectives sharpen. Jost et al. [5] delineated that affective polarization is characterized by the strong emotional responses, whether positive or negative, that members of different social groups evoke in each other. As an important psychological mechanism that plays a pivotal role in driving polarization, group justification encompasses the collective

inclination to promote the benefit of one's own group while opposing rival groups. Cole et al [6] reviewed the works on political polarization related to climate change issues. They emphasize two types of mechanisms of political polarization: individual-level psychological processes and group-level psychological processes. The former drives polarization through ideology, personality traits, cognitive styles, and perception of risks, threats, and morality. The latter includes social identities, social norms, and affective polarization.

Researchers have been attempting to explain political polarization with agent-based models and data simulations. These models usually adopt a utility maximization approach that assumes that the agents, affected by various types of public influences, selfishly make decisions to maximize their utility. Depending on the problem of interest, these models study the temporal evolution of opinion distribution among all Congress members, either in a discrete or continuous form, by either a deterministic or stochastic process. One common assumption is that the political parties make decisions to maximize their vote counts. Downs [7] modeled the competition of two parties for voters, where the voters' opinions follow an invariant zero-mean Gaussian distribution, and each voter votes for a party to maximize an expected utility. Each party adjusts its opinion position to maximize the expected number of votes it receives. The Downsian model predicts that the two parties should reach a consensus at the median opinion position of all voters. However, this conclusion does not coincide with reality, as partisan polarization in the United States has been an often-observed phenomenon.

A notable variant of the Downsian model is the satisficing model developed by Yang et al [8]. The time evolution of a party's opinion can be described using two characteristic variables: the party's opinion as the average opinion of its members and the party's opinion spread as the standard deviation of its members. Like the Downsian model, the opinion of a party moves in a direction that maximizes the expected number of votes. At the same time, a voter decides which party to vote for by randomly selecting a satisficing party (or abstaining from voting if there is none). Through data validation, the satisficing model captures opinion polarization between the U.S. Democratic and Republican parties since 1961. The model also develops a relationship between partisan polarization and opinion spread and predicts that the opinions of the members of each party are more centralized as the two parties become more polarized, which agrees well with the observed data. However, the work does not provide an explanation of how the polarization is developed and why it exhibits a wavy ideology distribution pattern over the past 150 years of the U.S. Congress.

Recently, Lanzetti et al. [9] extended the satisficing model to predict the complete opinion distribution of a party using Wasserstein gradient flows. The model predicts that the opinion distributions of the two parties should become more polarized and homogeneous within each party with time, converging to asymmetric distributions. However, while the extended satisficing model captures the overall tendency of partisan polarization, it does not explain notable exceptions where the opinion distributions change with time in a wavy fashion over the long history of the U.S. Congress. The author indicates that these exceptions are due to impact factors such as historical context, election rounds, and political campaigns without a detailed analysis.

Jones et al. [10] introduced a new way to define voters' distribution and utilities, where the voting population is composed of two subpopulations with polarized centers. Partisan polarization would exceed subpopulation polarization either when the two subpopulations are homogeneous and polarized or when they are heterogeneous and centralized. Ferri et al. [11] introduced a three-state model that borrows ideas from thermodynamics to opinion dynamics. The system is analogized with a thermal bath of a specific temperature, representing social agitation that affects the stochastically evolving dynamics of the system. The result shows that the system converges to a disordered state with polarized opinion clusters when the temperature is high and the neutrality parameter is small or to a relatively unified state when the opposite is true. This corresponds closely to the relationship between population spread and partisan polarization in [10].

Opinion dynamics models have been used in studying affective polarization. The interactions of people with close opinions would lead to agreement and positive affection, while interactions of people with distant opinions result in distrust and negative affection [4, 12]. Iyengar et al. [13] traced affective polarization to the power of partisanship as a social identity. Finkel et al. [14] utilized a "feeling thermometer" to measure the out-party hate level and found that it is the strongest in America compared to eight other nations. Lu et al. [15] studied a dynamic system of the conversion between polarization and cooperation in political interactions. The system has used data from roll-call votes cast in the U.S. Congress and showed a growth of polarization over the recent decades. However, the model itself does not explain the cause of the wavy political polarization pattern on a longer time scale.

Leonard et al. [16] applied a nonlinear opinion dynamics model to study partisan asymmetry and polarization. In the model, the two parties adjust their opinion positions based on self-reinforcement response mechanisms, in which a party exacerbates its polarized position to gain support. The model is tested using opinion data of the U.S. Congress since 1959 by searching for optimal parameters. The article is significant because it offers a plausible explanation of the two parties' polarization asymmetry based on the changes in the public's opinion. Moreover, the model is robust and does not rely on fine-tuning specific parameters to capture the overall tendency of opinion shifts. However, the authors do not attempt to validate their model with data before 1951, which exhibit more diverse behaviors of opinion shifts. In fact, because the self-reinforcement mechanism always results in opinion polarization, the model cannot capture other types of behaviors, such as the centralization of two parties' opinions around World War II.

Baldassarria and Bearman [17] studied the paradox of the simultaneous absence and presence of attitude polarization and the paradox of the simultaneous presence and absence of social polarization. Later, Baldassarria and Page [18] reviewed this work in light of the theoretical distinction between ideological partisanship, which is generally rooted in sociodemographic and political cleavages, and affective partisanship, which is, instead, fueled mainly by emotional attachment and repulsion, rather than ideology and material interests.

This work intends to study the evolving polarization of the parties and their spread with the effects of in-party and cross-party opinion dynamics, and the effect of a time-dependent national social norm. Based on a micro-scale agent-based model, we derive an

analytical theory on how the two parties' opinions influence each other with the presence of the national social norm effect. Four parameters are used in the theory: (1) a tolerance opinion difference parameter that determines whether the mutual impact of any two interacting individuals is attractive or repulsive; (2) an influence decay parameter whose inverse determines the exponential decay rate of their mutual impact as their opinion difference increases; (3) a pre-coefficient and a rate for the exponential increase of the opinion exchange efficiency due to the fast progress of communication technology in the past; and (4) a time-dependent function characterizing the strength of the national social norm effect. These four parameters and the strength function are obtained by fitting the theory with empirical data from a frequently used dataset [19]. The obtained values of these four parameters and their social physical meanings can be reasonably well interpreted. The strength of the national social norm effect is consistent with the major historical events that occurred in the past 150 years.

This article is organized as follows. Section 2 describes the ideology dataset we used for the U.S. Congress [19]. Section 3 develops the theory for the evolution of polarization and spread with proper assumptions and approximations. Section 4 describes the numerical method used to assimilate the theoretical model and observational data. Section 5 presents the analysis results, interpretations, and the justification of the approximations used. Section 6 summarizes the contributions of this work and points out the areas for further research.

2 Data description

We apply the ideology dataset from the U.S. Congress [19] for this research. The dataset comprises two-dimensional ideology scores of congressional members from 1868 to 2022 computed by the Dynamic, Weighted, Nominal Three-Step Estimation (DW-NOMINATE) algorithm [20]. We choose the scores in the first dimension (economic liberalism-conservatism) of this dataset to represent the members' opinion distributions, which has data from every 2 years with 78 time points, or Congress sessions, covering 154 years.

Figure 1a shows the opinion distribution of the Democratic (blue) and Republican (red) parties representing a wavy pattern of separation and unification process. The bipartisan polarizations can be defined as the means of the opinion distributions of the two parties. The absolute values of the polarizations are as shown in Figure 1b, which are stronger around 1880 to 1910 and 2000 to 2020 and weaker around 1930 to 1980. The opinion spreads in Figure 1c of the two parties, computed as the standard deviation of the opinion distributions about their corresponding means, is narrower when there is an intense polarization and is wider when the polarization is weak.

It is worth noting the significant asymmetric disparities in both the polarizations and spreads of the two parties. In Figure 1d, the average of the two parties' opinion means slightly deviates from the zero-opinion level. Additionally, in Figure 1e, the numbers of Congress members of the two parties fluctuate over time, with their mean increasing from 1868 to 1920 and stabilizing thereafter. This work focuses on modeling the symmetric polarization and spread patterns of the two parties, as well as their long-term temporal wavy evolutions, leaving the study of asymmetry patterns to future research.

3 Theory

Let D represent the Democratic Party and R represent the Republican Party in the U.S. Congress. An individual member i's opinion level at time t, denoted as $B_i(t)$, resides within a one-dimensional segment from −1 to +1. The opinion of each member at a given time is affected by opinion impacts from all other individuals and a national social norm effect. The general governing equation for the evolution of $B_i(t)$ is

$$\frac{dB_i(t)}{dt} = \sum_{j\in D} I_{ji} + \sum_{j\in R} I_{ji} - \alpha(t)\left[B_i(t) - B_o(t)\right] \tag{1}$$

In Equation 1, the first term on the right is the summation of impacts from all individuals in the Democratic Party (D) and the second term is from all individuals in the Republican Party (R). The opinion impact I_{ji} of an individual j on i is defined by

$$I_{ji} = A(t) D_{ji} e^{-\frac{|B_j - B_i|}{B_H}} \left(B_j - B_i\right) \tag{2}$$

The opinion impact is a product of the opinion difference between the two individuals and three additional factors. The factor $D_{ji} e^{-\frac{|B_j - B_i|}{B_H}}$ represents affinity, a measure of likeness between two individuals, which decreases as the opinion difference between them increases. D_{ji} is expressed using an opinion influence tolerance parameter, B_T:

$$D_{ji} = \begin{cases} 1 - \dfrac{|B_j - B_i|}{B_T}, & |B_j - B_i| < 2B_T \\ -1, & |B_j - B_i| \geq 2B_T \end{cases} \tag{3}$$

The tolerance parameter determines the threshold of the opinion difference between two individuals, from exerting an attractive or positive influence to a repulsive or negative influence.

The factor $e^{-\frac{|B_j - B_i|}{B_H}}$ represents the homophily effect, quantifying the extent of mutual influence between the two individuals, which is also related to the opinion difference. B_H is an influence decay parameter; its inverse determines the rate of exponential influence decay as the opinion difference increases. When the opinion difference between two individuals becomes large compared to B_H, the mutual influence power between the two becomes small. This decay parameter characterizes the range of individuals in a population that may have a significant influence on an individual, whether positive or negative. In the case of positive influence, this homophily effect has been validated by He and Zhang [21] using experimental data. In this work, we assume that each of B_T and B_H is the same for both in-party and cross-party impacts.

The proportional factor $A(t)$ relates to the interaction efficiency between individuals within the dynamic opinion system. The growth of this efficiency stems from advancements in transport technologies (such as automobiles, highways, and aircraft) and communication technologies (such as radios, TVs, phones, cell phones, and the Internet) over the past 150 years, resulting in exponential growth.

The third term in Equation 1 characterizes the national social norm effect, affecting all individuals across both parties, a topic

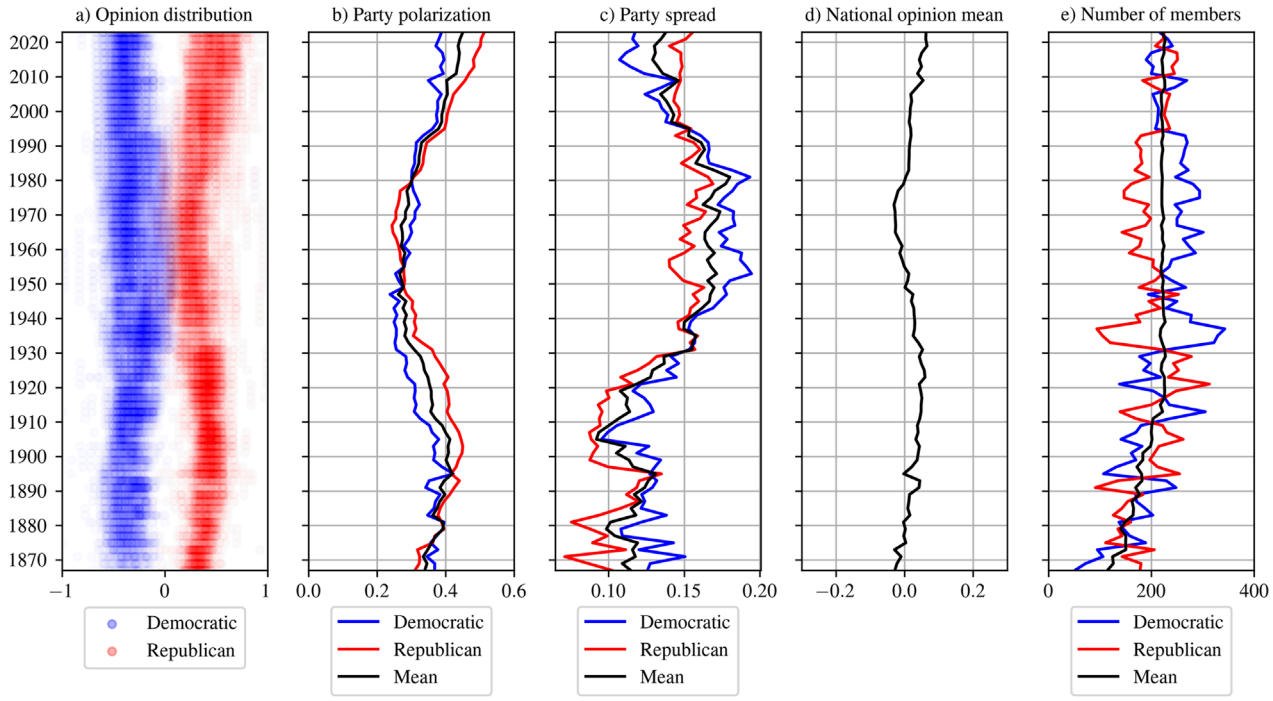


FIGURE 1
(a) Opinion distributions of the Democratic and Republican party members from 1868 to 2022, **(b)** absolute values of the polarizations of the two parties and their mean, **(c)** spreads of the two parties and their mean, **(d)** the average of the opinion means of the two parties, and **(e)** the numbers of congresspeople in the two parties and their means.

extensively studied in political science literature [6]. In Equation 1, B_o denotes the position of the norm, and $\alpha(t) > 0$ represents the strength of the norm effect. This strength is related to important world events that occurred around or before time t and is therefore time dependent. It is strong in the presence of a common national threat for both parties and weak when such threats are absent.

The model Equations 1–3 are extensions of the widely used classical DeGroot model [22] in opinion dynamics studies. The extensions are made in three aspects. (1) In addition to the opinion difference between two individuals, the impact of one person on another also depends on their mutual likeness (affinity). (2) The mutual influence of individuals' opinions also depends on the interaction efficiency, which is assumed to have increased exponentially over the past 150 years. (3) Changes in people's opinions are also regulated by the national social norm effect, which tends to strengthen when the nation faces a common threat and weaken when the nation is in a peaceful time.

The members of the U.S. Congress change from session to session, with each session lasting 2 years. It is difficult to identify a consistent counterpart across sessions for each member with a similar opinion position. This discontinuity prevents us from tracking their individual opinion trajectories over time. To overcome this limitation, we will adopt a mean-field-like approach and study the evolution of key quantities of the opinion distributions in the following analysis.

In this work, the term "partisan polarization" (or "polarization") refers to the deviation of each party's opinion center or mean opinion level from the national norm, while "opinion spread" (or "spread") indicates the standard deviation of opinion distribution within each party, also termed inclusiveness in some literature [8, 9]. Using shorthand notations B_i for $B_i(t)$, the partisan polarizations of the Democratic (D) and Republican (R) parties, respectively, are expressed as $B_D = \frac{1}{N_D}\sum_{i\in D} B_i < 0$ and $B_R = \frac{1}{N_R}\sum_{i\in R} B_i > 0$. Their time-varying absolute values and average are shown in Figure 1b above. The squares of the opinion spreads in the two parties can be represented by $\sigma_D^2 = \frac{1}{N_D}\sum_{i\in D}(B_i - B_D)^2$ and $\sigma_R^2 = \frac{1}{N_R}\sum_{i\in R}(B_i - B_R)^2$, where N_D and N_R are the numbers of congresspeople in the Democratic Party D and Republican Party R. The time-varying opinion spreads of the two parties and their average are shown in Figure 1c.

We focus on the temporal evolution of polarization and the spread of the two parties instead of tracking the opinion evolution of the individual Congress members. Noticing that the total number of congress members remains fixed after 1920, while the number of each party fluctuates. These fluctuations are modest compared to the total number, apart from exceptional periods such as World War II, as shown in Figure 1e. Because the overall opinion impact on any member is the summation of the impacts from all other members, these fluctuations have a limited impact on the qualitative understanding of the mean-field opinion dynamics. Therefore, to reduce the complexity that is associated with short timescale phenomena, we assume that the numbers of congresspeople in the two parties are the same for all years, $N_D = N_R = N$.

We also assume that the opinion position of the national social norm is fixed at $B_o(t) = 0$ and that the opinion distributions of the

two parties are symmetric with respect to this national norm. The symmetry assumption is supported by the empirical observation that, over long timescales, the polarization and spread of the two parties are generally symmetric (as illustrated in Figures 1b,c). This assumption has been widely used in prior studies [7–15], given the noisy nature of social science datasets. This assumption also enables the development of a tractable model capable of capturing the dominant, long-term aggregated trends over the 154-year study period. Denoting μ and σ as the polarization and the spread of opinion distribution, we have $B_R = -B_D = \mu > 0$ and $\sigma_R^2 = \sigma_D^2 = \sigma^2$.

We further make the following four approximations, which will be justified in Section 5. Note that these approximations are not perfectly accurate for all the in-party and cross-party interactions because of the existence of members with extreme opinions. Through these approximations, we aim to achieve a model that captures the variation of the polarization and spread for the long term with simple governing equations.

Approximation 1: The in-party exponential decaying effect is negligible because $|B_j - B_i| \sim \sigma < B_H/3$, for two individuals with the same party, $i,j \in D$ or $i,j \in R$, $e^{-\frac{|B_j - B_i|}{B_H}} \approx 1$. This is a simplification that holds when most in-party opinion differences lie within a narrow range compared to B_H. We acknowledge that this assumption may not hold during several short periods with extreme party opinion spread. For the purpose of a qualitative understanding of the general long-term trend, the effect of such extreme cases is neglected.

Approximation 2: For any two individuals i and j from the same party, we assume $|B_j - B_i| < 2B_T$. Therefore, $I_{ji} = A(t)\left(1 - \frac{|B_j - B_i|}{B_T}\right)(B_j - B_i)$. This approximation allows that the members of the same party with close opinions $|B_j - B_i| < B_T$ interact attractively, those with opinion difference $B_T < |B_j - B_i| < 2B_T$ interact repulsively, and the multiplication factor D_{ji} in Equation 3 is always greater than −1.

Approximation 3: The cross-party exponential decaying effect is important because $|B_j - B_i| \approx B_R - B_D = 2\mu$ for two individuals i and j belonging to different parties, and $2\mu > 2B_T$. Thus, the opinion impact can be simplified into Equation 4 below.

$$I_{ji} = I_{RD} = A(t)\cdot(-1)\cdot e^{-\frac{|B_R - B_D|}{B_H}}(B_R - B_D) = -2A(t)\mu e^{-\frac{2\mu}{B_H}} \quad (4)$$

Approximation 4: The proportional factor $A(t)$ is assumed to be an exponential function of time t to describe the improvement of interaction efficiency due to the advancement of transportation and communication technologies. This assumption is supported by historical data showing that advances in transportation and communication—from postal mail and telegraph to telephone and the Internet—have consistently reduced the cost and increased the frequency of human interaction, exhibiting long-term exponential growth patterns in connectivity and information exchange [23, 24]. This simple approximation captures the long-term trend of steadily increasing connectivity and interaction intensity over the past 150 years.

Adopting these approximations, for an individual i in the D party, $B_i < 0$, Equation 1 becomes

$$\frac{dB_i}{dt} = A(t)\sum_{j\in D}\left(1 - \frac{|B_j - B_i|}{B_T}\right)(B_j - B_i) - 2A(t)N\mu e^{-\frac{2\mu}{B_H}} - \alpha(t)B_i \quad (5)$$

The rate equation for partisan polarization $\mu = -\frac{\sum_{i\in D}B_i}{N}$ can be obtained by averaging Equation 5 over all individuals in the D party, recognizing that the first term vanishes after summation.

$$\frac{d\mu}{dt} = \mu\left[2A(t)Ne^{-\frac{2\mu}{B_H}} - \alpha(t)\right] \quad (6)$$

The opinion centers of the two parties converge or diverge depending on the competing repulsive effect between the two parties and the attraction effect of the national norm. Specifically, the bipartisan polarization increases when the national norm strength $\alpha(t)$ is smaller than a threshold $2A(t)Ne^{-\frac{2\mu}{B_H}}$ and *vice versa*.

The equation for the spread of the parties, described by the standard deviation of the opinion distribution within each party, can be obtained as shown below, with derivations in Supplementary Appendix SA.

$$\frac{d\sigma}{dt} = \sigma\left[A(t)N\left(\frac{4\sigma}{\sqrt{\pi}B_T} - 1\right) - \alpha(t)\right] \quad (7)$$

The evolution of the spread of the party opinion is controlled by two effects. The first term is the divisive effect within the party when $\frac{4\sigma}{\sqrt{\pi}B_T} > 1$. The second term is the national social norm effect, which tends to reduce the spread. The spread increases when the national norm strength $\alpha(t)$ is smaller than $A(t)N\left(\frac{4\sigma}{\sqrt{\pi}B_T} - 1\right)$ and *vice versa*.

The mean-field-like approach, with Equations 6 and 7 as the two governing equations, gives us a mechanistic understanding of how the long-term evolutions of polarization and spread are shaped by the social–psychological processes, the effect of national social norm, which is expected to be related to important historical events, and the enhancement of interaction efficiency.

To mitigate strong polarization and spread, three strategies may be considered based on the governing Equations 6 and 7. (1) Fostering a shared civic identity and emphasizing widely endorsed societal values among all citizens, as supported by findings in political psychology and sociology [25]. This corresponds to the increase in the strength of the national social norm effect $\alpha(t)$ in Equation 2, promoting respectful discourse and exposure to ideologically diverse individuals across the two parties, which may decrease the mutual hostility [2, 4]. This corresponds to the increase in the homophily decay parameter B_H in Equation 6, for an increase of affinity. (3) Encouraging interactions within each party among those with different views as well [2, 4], which is expected to decrease the spread by fostering increased tolerance. This corresponds to an increase in the tolerance parameter B_T in Equation 7.

4 Numerical method for the assimilation of data and theory

In their discrete forms, using the forward difference scheme, Equations 6 and 7 can be written as

$$\mu(t+\Delta t) = \mu(t) + \mu(t)\left[2\overline{A}(t)e^{-\frac{2\mu(t)}{B_H}} - \alpha(t)\Delta t\right] \quad (8)$$

and

$$\sigma(t+\Delta t) = \sigma(t) + \sigma(t)\left[\overline{A}(t)\left(\frac{4\sigma(t)}{\sqrt{\pi}B_T} - 1\right) - \alpha(t)\Delta t\right] \quad (9)$$

where $\overline{A}(t) = A(t)N\Delta t$ represents the total amount of influence from all individuals impacting the system, and Δt is the incremental time step, which is 2 years in this study. The time variable t starts from 1868 and ends in 2022. We denote $t_k = 1868 + (k-1)\Delta t, 1 \le k \le K$ as the k-th time point, where $K = 78$ is the total number of time points. We assume $\overline{A}(t_k) = A_0 e^{\frac{ck}{K}}$, with two constants $A_0 > 0, c > 0$ whose values are to be determined after fitting the theory to the data.

We aim to optimally utilize information from both the empirical data and the theoretical relationships among the variables and the parameters provided by Equations 8 and 9. The model Equations 8 and 9 involve four constant parameters A_0, c, B_T, and B_H and an unknown time-dependent national norm strength function $\alpha(t)\Delta t$. By fitting the data with the model using the assimilation algorithm described below, we can obtain the estimates of the time-dependent patterns of polarization and spread, the values of the four parameters, and the intensity function of the national social norm.

For each Congress t, denote $\tilde{\mu}(t)$, $\mu^*(t)$ to be the estimated and observed partisan polarization and $\tilde{\sigma}(t)$, $\sigma^*(t)$ to be the estimated and observed spread.

The raw observed data μ^* and σ^* are very noisy, containing short-term fluctuations from session to session over 154 years. It is crucial to filter out these short-term noises with the support of the two long-term dynamics equations for the accurate capturing of the long-term trends in polarization and spread. The estimation procedure below, which simultaneously minimizes four loss terms, two for the matching of the data and two for the satisfaction of the equations, allows us to assimilate long-term information from raw observation data with the constraints of the long-term model equations. This strategy is frequently used in data-model assimilation literature and inverse modeling literature, such as the simultaneous solution of inverse problem governed by partial differential equations (PDEs) and state estimation [26], physics-informed neural networks for solving differential equations with unknown parameters using observational data [27], and variational inference using optimization for the joint estimation of system state and noise parameters [28].

We seek to reduce the mean squared errors in polarization and spread in terms of the data difference between the estimated and the observed, and the mean squared residual errors of the two equations using a minimization procedure. The four mean squared errors are

$$MSE^{\mu}_{data} = \frac{1}{K}\sum_{k=1}^{K}\left[\tilde{\mu}(t_k) - \mu^*(t_k)\right]^2 \tag{10}$$

$$MSE^{\sigma}_{data} = \frac{1}{K}\sum_{k=1}^{K}\left[\tilde{\sigma}(t_k) - \sigma^*(t_k)\right]^2 \tag{11}$$

$$MSE^{\mu}_{eq} = \frac{1}{K-1}\sum_{k=1}^{K-1}\left[\tilde{\mu}(t_{k+1}) - \tilde{\mu}(t_k) - \tilde{\mu}(t_k)\left(2\overline{A}(t_k)e^{-\frac{2\tilde{\mu}(t_k)}{B_H}} - \alpha(t_k)\Delta t\right)\right]^2 \tag{12}$$

$$MSE^{\sigma}_{eq} = \frac{1}{K-1}\sum_{k=1}^{K-1}\left[\tilde{\sigma}(t_{k+1}) - \tilde{\sigma}(t_k) - \tilde{\sigma}(t_k)\left(-\overline{A}(t_k) + \frac{4\overline{A}\tilde{\sigma}(t_k)}{\sqrt{\pi}B_T} - \alpha(t_k)\Delta t\right)\right]^2 \tag{13}$$

The unknown time-dependent national norm strength function $\alpha(t)\Delta t$ is approximated using a linear expansion using n Legendre polynomials with coefficients $(a_0, a_1, a_2, \ldots, a_n)$. Introducing a transformation $x_k = \frac{2k}{K} - 1 \in [-1, 1]$ and $\tilde{\alpha}(x_k) = \alpha(t_k)\Delta t$, we have the linear expansion $\tilde{\alpha}(x_k) = \sum_{m=0}^{n} a_m P_m(x_k)$ for $k = 1, 2, \ldots, K$, where $P_m(x)$ is the Legendre polynomial of degree m. Thus, the set of parameters to be obtained can be denoted as $\Theta = \{A_o, c, B_T, B_H, a_0, a_1, \cdots, a_n\}$. The seemingly large number of parameters is mainly for obtaining the strength function, which is not known *a priori* and cannot be described by a few parameters at this stage.

The mean polarization and the mean spread across all time points are, respectively, $\overline{\mu} = \frac{1}{K}\sum_{k=1}^{K}\mu^*(t_k) \approx 0.347$ and $\overline{\sigma} = \frac{1}{K}\sum_{k=1}^{K}\sigma^*(t_k) \approx 0.139$. The mean variations per time step for the polarization and spread are, respectively, $\overline{\mu}_d = \frac{1}{K-1}\sum_{k=1}^{K-1}|\mu^*(t_{k+1}) - \mu^*(t_k)| \approx 8.96 \times 10^{-3}$ and $\overline{\sigma}_d = \frac{1}{K-1}\sum_{k=1}^{K-1}|\sigma^*(t_{k+1}) - \sigma^*(t_k)| \approx 5.36 \times 10^{-3}$. Because the four mean-squared error terms in Equation 10 have different orders of magnitude, we use the following normalization factors in constructing a total loss function for minimization,

$$C^{\mu}_{data} = \frac{1}{K}\sum_{k=1}^{K}\left[\mu^*(t_k) - \overline{\mu}\right]^2 = 3.03 \times 10^{-3} \text{ for } MSE^{\mu}_{data},$$

$$C^{\sigma}_{data} = \frac{1}{K}\sum_{k=1}^{K}\left[\sigma^*(t_k) - \overline{\sigma}\right]^2 = 5.81 \times 10^{-4} \text{ for } MSE^{\sigma}_{data},$$

$$C^{\mu}_{eq} = \frac{1}{K-1}\sum_{k=1}^{K-1}\left[|\mu^*(t_{k+1}) - \mu^*(t_k)| - \overline{\mu}_d\right]^2 = 4.61 \times 10^{-5} \text{ for } MSE^{\mu}_{eq}, \text{ and,}$$

$$C^{\sigma}_{eq} = \frac{1}{K-1}\sum_{k=1}^{K-1}\left[|\sigma^*(t_{k+1}) - \sigma^*(t_k)| - \overline{\sigma}_d\right]^2 = 1.77 \times 10^{-6} \text{ for } MSE^{\sigma}_{eq}.$$

The total loss function can be defined as a linear combination of the four mean square errors as described in Equation 10, which is a function of the polarization $\tilde{\mu}$, spread $\tilde{\sigma}$, and the parameter set $\Theta = \{A_o, c, B_T, B_H, a_0, a_1, \cdots, a_n\}$, to be estimated by assimilating the observed data and the theory.

$$L = L(\tilde{\mu}, \tilde{\sigma}, \Theta) = \frac{\lambda^{\mu}_{data}}{C^{\mu}_{data}}MSE^{\mu}_{data} + \frac{\lambda^{\sigma}_{data}}{C^{\sigma}_{data}}MSE^{\sigma}_{data} + \frac{\lambda^{\mu}_{eq}}{C^{\mu}_{eq}}MSE^{\mu}_{eq} + \frac{\lambda^{\sigma}_{eq}}{C^{\sigma}_{eq}}MSE^{\sigma}_{eq} \tag{14}$$

where λ^{μ}_{data}, λ^{σ}_{data}, λ^{μ}_{eq}, and λ^{σ}_{eq} are weights that balance the relative importance of the four mean squared errors.

The unknowns $\tilde{\mu}(t)$ and $\tilde{\sigma}(t)$ appear in two places. The first place is in MSE^{μ}_{data} and MSE^{σ}_{data} of Equations 10, 11, and 14, where $\tilde{\mu}(t)$ and $\tilde{\sigma}(t)$ are fitted to the known raw observed data μ^*(t) and $\sigma^*(t)$. The second place is in MSE^{μ}_{eq} and MSE^{σ}_{eq} of Equations 12, 13, 14, where $\tilde{\mu}(t)$, $\tilde{\sigma}(t)$ and Θ are fitted for the satisfaction of Equations 8 and 9. These four mean squared error terms constitute the total loss in Equation 14. By minimizing this total loss, all four loss terms are minimized simultaneously. Consequently, the unknowns $\tilde{\mu}$, $\tilde{\sigma}$, and Θ are then obtained jointly when this total loss reaches its minimum.

The choice of the loss function, as defined in Equation 14, is made with the following considerations. First, political ideology data, such as the DW-DOMINATE scores, contain noise and possible biases. The computed polarization and spread data $\mu^*(t)$ and $\sigma^*(t)$ have considerable discrepancies, especially with the symmetric approximation adopted, as indicated in Figure 1. By allowing balances between data losses and equation losses, we can prevent the model from over-relying on the dataset. Second, the two

equations governing the time evolutions of polarization and spread have different levels of accuracy. For instance, we expect that the equation for $\tilde{\sigma}(t)$ is less accurate than the equation for $\tilde{\mu}(t)$. Thus, having four separate terms allows us to manually adjust weights to achieve our optimal assimilation goal when there is a need.

In minimizing the total loss function Equation 14, we employ a gradient-descent-based optimizer to obtain the optimal parameter set Θ. In this approach, we first provide an initial estimation of the polarization $\tilde{\mu}$, spread $\tilde{\sigma}$, and the parameter set Θ, and then calculate the four mean squared errors, compute the gradient of the estimated $\tilde{\mu}, \tilde{\sigma}$, and Θ, and update them using the Adam optimizer [29] for each iteration. This optimization process repeats until the total loss converges to an acceptable value.

We employ an updating rate of 0.001 to minimize the total loss with 50,000 iterations to ensure convergence. The weights for the four terms in Equation 14 are set as follows: $\lambda^{\mu}_{data} = 1, \lambda^{\sigma}_{data} = 1, \lambda^{\mu}_{eq} = 1$, and $\lambda^{\sigma}_{eq} = 1$. The four corresponding relative root mean square errors during the minimization process, and the converged relative root mean square errors of polarization and spread for degree $n = 7$ as defined below are shown in Figure 2.

$$RE^{\mu}_{data} = \sqrt{\frac{\sum_{k=1}^{K}[\tilde{\mu}(t_k) - \mu^*(t_k)]^2}{\sum_{k=1}^{K}[\mu^*(t_k)]^2}} \tag{15}$$

$$RE^{\sigma}_{data} = \sqrt{\frac{\frac{1}{K}\sum_{k=1}^{K}[\tilde{\sigma}(t_k) - \sigma^*(t_k)]^2}{\sum_{k=1}^{K}[\sigma^*(t_k)]^2}} \tag{16}$$

$$RE^{\mu}_{eq} = \sqrt{\frac{\sum_{k=1}^{K-1}\left[\tilde{\mu}(t_{k+1}) - \tilde{\mu}(t_k) - \tilde{\mu}(t_k)\left(2\overline{A}(t_k)e^{-\frac{2\tilde{\mu}(t_k)}{B_H}} - \alpha(t_k)\Delta t\right)\right]^2}{\sum_{k=1}^{K-1}[\mu^*(t_{k+1}) - \mu^*(t_k)]^2}} \tag{17}$$

$$RE^{\sigma}_{eq} = \sqrt{\frac{\sum_{k=1}^{K-1}\left[\tilde{\sigma}(t_{k+1}) - \tilde{\sigma}(t_k) - \tilde{\sigma}(t_k)\left(-\overline{A}(t_k) + \frac{4\overline{A}\tilde{\sigma}(t_k)}{\sqrt{\pi}B_T} - \alpha(t_k)\Delta t\right)\right]^2}{\sum_{k=1}^{K-1}[\sigma^*(t_{k+1}) - \sigma^*(t_k)]^2}} \tag{18}$$

Figure 2a indicates that the minimization process converges with 50,000 iterations and shows that the relative errors for equation satisfactions are smaller than for data discrepancies. Figures 2b,c show that the relative errors for equations, for almost all years, are smaller than those for data discrepancies, polarization, and spread, respectively.

We experiment with four choices of degree n, ranging from 5 to 8, for the approximation of $\tilde{\alpha}(x_k)$ with the expansion using Legendre polynomials. During the minimization process, we obtain the polarization $\tilde{\mu}(t)$, the party spread $\tilde{\sigma}(t)$, as well as the parameter set $\Theta = \{A_o, c, B_T, B_H, a_0, a_1, \cdots, a_n\}$, as shown in Table 1. It can be seen that the estimated values of all parameters do not vary significantly for $n = 7$ and $n = 8$, which indicates that the expansion with $n = 7$ is sufficiently accurate for the approximation of $\tilde{\alpha}(x_k)$. It should be noted that the seemingly larger number of coefficients, n = 7, is only for the accurate quantification of the strength of the national norm effect $\tilde{\alpha}(x_k)$, which is an unknown function of time. The political meaning of this obtained function will be provided in the next section in light of the major historical events over the past 150 years.

Table 2 lists four relative errors defined in Equations 15–18 for evaluating model accuracy and robustness. RE^{μ}_{data} and RE^{σ}_{data} (Equations 15,16,) quantify the deviations of the simulated polarization $\tilde{\mu}(t)$ and intra-party spread $\tilde{\sigma}(t)$ from the empirical DW-NOMINATE data $\mu^*(t)$ and $\sigma^*(t)$. RE^{μ}_{eq} and RE^{σ}_{eq} (Equations 17,18,) measure the consistency of the results with respect to the governing mean-field equations, thus reflecting the degree to which the results satisfy the dynamic equations for polarization and spread.

The relative errors RE^{μ}_{data} and RE^{σ}_{data} for polarization $\tilde{\mu}(t)$ and spread $\tilde{\sigma}(t)$, respectively, are below 0.12, and remain consistently small for all tested polynomial degrees $n = 5 - 8$. The equation-consistency errors RE^{μ}_{eq} and RE^{σ}_{eq} are also consistently small for all n. These results demonstrate that the model not only achieves high quantitative accuracy in reproducing the observed long-term polarization patterns but also preserves internal dynamical consistency across different polynomial expansions of the national norm function $\alpha(t)$. The stability of the error values further confirms the model's robustness with respect to the choice of degree n and validates that the model's performance is insensitive to n. As indicated in Figure 3 below, the profiles of polarization and spread do not change with n. Having n = 7 is mainly for the more accurate quantification of $\alpha(t)$ and $A(t)$.

5 Results and interpretations

5.1 Polarization and spread

Figures 3a,b compared the estimated polarization $\tilde{\mu}(t)$ and spread $\tilde{\sigma}$(t) with the observed data. When the polarization increases, the spread decreases, and *vice versa*, which is consistent with the empirical data and research work of Cole et al [6]. Our model yields a more accurate estimation of polarization than spread. The estimated polarization captures the wavy evolution pattern over the past 154 years, and the estimated spread captures the overall wavy trend as well, with the exceptions of notable deviations from 1900 to 1930 and from 1970 to 2000. With the symmetric approximations adopted in Section 3, we are content with these results at this stage because the objective of this work is only on the overall long timescale wavy patterns of the polarization and spread for the entire recorded historical dataset. The modeling of the asymmetries will be investigated in the future.

Figure 3c shows the estimated strength function of the national social norm effect $\tilde{\alpha}$(t), and Figure 3d shows the estimated interaction parameter factor $\overline{A}$(t). All four plots in Figure 3 show that the estimated parameters and functions have insignificant differences for the degree of approximation of $\tilde{\alpha}$ (t) using Legendre polynomials from $n = 7$ to $n = 8$. For simplicity, we will use $n = 7$ for the discussions and interpretations below.

The comparison between the evolutions of the observed data and the estimated opinion distributions is shown in Figure 4. Typically, stronger polarization corresponds to a narrower spread, and *vice versa*. The slight differences in distributions are primarily due to the symmetry approximation we imposed and the noisy nature of the observed data.

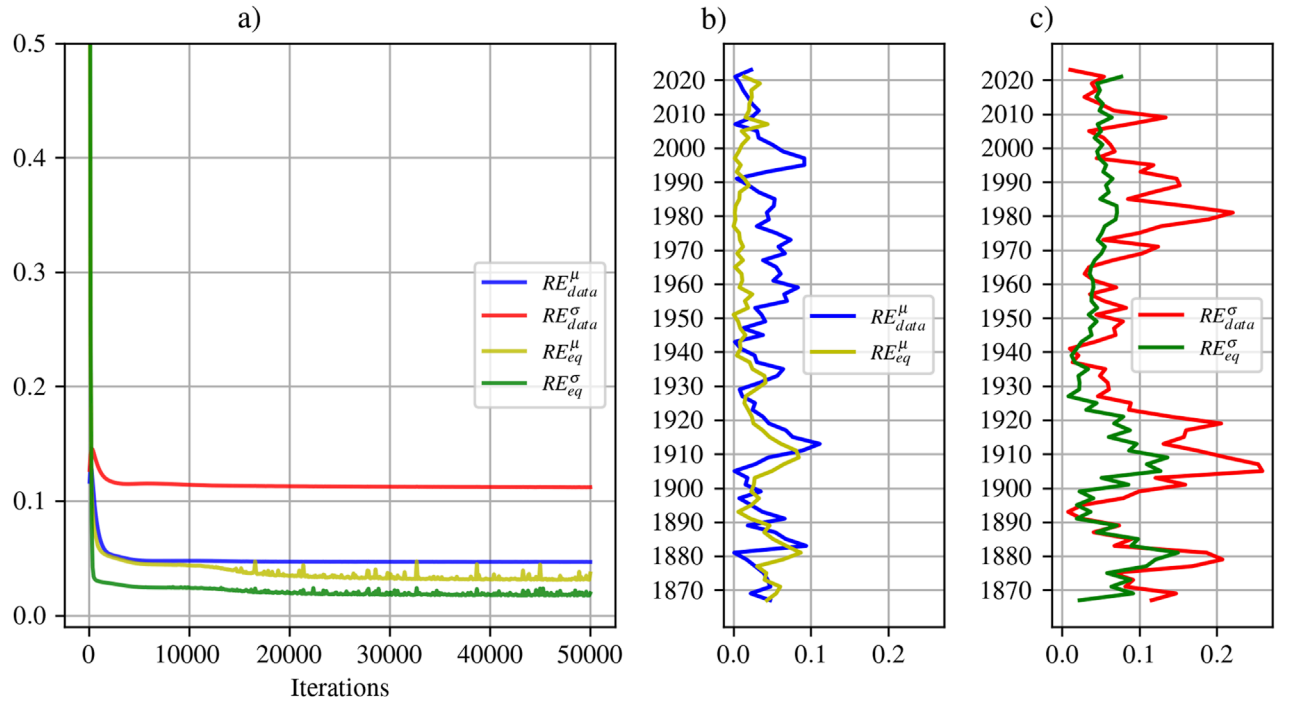


FIGURE 2
(a) The minimization processes of the four relative root mean squared errors. **(b,c)** The temporal distributions of the four relative root mean square errors are shown when the number of iterations reaches 50,000. All plots are with degree n = 7.

5.2 Strength of the national social norm effect

The national social norm strength $\alpha(t)$ can be viewed as the strength of nationwide consensus. Strong consensus occurs when the nation faces a common threat, resulting in ideology clusters attracted to the national norm with a stronger convergence force. Weak consensus occurs during peaceful times when the attractive influence from the national norm is weak, when severe cross-party conflicts emerge, and distant ideology clusters are more likely to diverge. In Figure 5, we graph the obtained $\alpha(t)$ next to a timeline of notable historical events in the past 154 years. We categorize the events into those that lead to agreements with green color and those that lead to conflicts with red color. A more complete list of events with descriptions is in the Supplementary Materials. Two turning points are presented: the first point around 1880 and the second around 1972. The wavy evolution of the strength of the national social norm is consistent in a historical context by splitting the 154 years into the five periods shown in Figure 5, and the interpretations are provided below.

Period 1: Post-Civil War (1868–1914) The Post-Civil War years for the United States were peaceful, as cities were reconstructed, and industries began to thrive. While the rise of industrialization produced a class of wealthy industrialists and a prosperous middle class, the working class continued to agonize from unemployment, minimal wages, and pressure from immigrants. Negative sentiment rose as more minor societal conflicts were magnified, resulting in a weak influence from the national norm in the late 19th and early 20th century, as depicted in Figure 5. The first turning point of $\alpha(t)$ is around 1880 before the Progressive Movement, when a group of activists tried to advocate democracy, expand civil rights, and regulate the higher social classes. These political activities slowly strengthened the national norm effect, which agrees with the increasing trend of $\alpha(t)$ from 1890 to 1914 in our result.

Period 2: World Wars (1914–1945) The time from 1914 to 1945 witnessed World War I, the Great Depression, and World War II, when Americans united to face a series of crises. During World Wars I and II, the national norm strength increased as Americans formed a consensus on defending their country against foreign military forces. On the other hand, studies have shown that during the Great Depression, most Americans were unified and optimistic, mainly due to the enactment of the New Deal. These events together explained the steeply increasing trend of $\alpha(t)$ between 1914 and 1945.

Period 3: Early Cold War (1945–1972) In the first 30 years after World War II, a new political consensus was formed concerning the Cold War and anti-communism, causing the national norm strength to continue rising. This consensus peaked around 1972 (the second turning point), indicated by a series of events that included the construction of the Berlin Wall, the Cuban Missile Crisis, the space race, and the Vietnam War, which increased the national norm strength.

Period 4: Late Cold War (1972–1992) The once rigid anti-communism consensus began to fragment as the protracted conflict between the United States and the Soviet Union edged toward its conclusion. This period was punctuated by a series of transformative events. In 1972, a notable stride was made when President Nixon signed the SALT I agreement, establishing a framework for limiting

TABLE 1 The estimated values of all parameters for different degrees n.

Parameters	$n = 5$	$n = 6$	$n = 7$	$n = 8$
A_0	0.551	0.535	0.591	0.603
c	1.760	1.914	1.843	1.817
B_H	0.452	0.443	0.437	0.439
B_T	0.218	0.219	0.221	0.221
a_0	0.669	0.696	0.720	0.730
a_1	0.497	0.555	0.552	0.554
a_2	−0.231	−0.233	−0.253	−0.253
a_3	−0.275	−0.294	−0.316	−0.314
a_4	0.127	0.131	0.138	0.142
a_5	0.156	0.143	0.143	0.144
a_6		0.037	0.006	0.012
a_7			−0.037	−0.023
a_8				0.016

TABLE 2 Relative errors for different degree n.

Error	n = 5	n = 6	n = 7	n = 8
RE^{μ}_{data}	0.046	0.047	0.047	0.047
RE^{σ}_{data}	0.113	0.112	0.112	0.112
RE^{μ}_{eq}	0.035	0.036	0.035	0.031
RE^{σ}_{eq}	0.021	0.020	0.019	0.017

strategic armaments. The subsequent year, 1973, witnessed the United States withdraw from the Vietnam War, signaling a retreat from one of the most contentious battlegrounds of the Cold War. The thaw in relations continued as the United States established diplomatic ties with mainland China in 1979. The latter years of this period saw a continued diminishment in the confrontations between major superpowers. In 1989, the Berlin Wall fell. In 1991, the United States and the Soviet Union signed the START I treaty, agreeing to reduce strategic nuclear arms further. Later that year, the Soviet Union broke up, signaling the denouement of the Cold War. These events show that the conflicts between the United States and communist countries gradually became smaller. With fewer threats outside the country, inner conflicts emerged, and public views began to divide. The slowly decreasing trend of $\alpha(t)$ in this period corresponded to these developments.

Period 5: Post-Cold War (1992–2022) The years after the Cold War were relatively peaceful. A few global and national crises still emerged, but the scales of the impacts were much smaller. As a result, we see a decreasing trend in the graph of $\alpha(t)$ from 1991 to 2014. Intriguingly, the resulting $\alpha(t)$ indicates an upward trend in the most recent decade, which corresponds to very strong asymmetric opinion distributions of the two parties. Future modeling work is needed to address these asymmetricities.

In summary, a common trend is that national threats usually correspond to a growth in the strength of the national norm, a decrease in polarization, and an increase in spread. In contrast, peaceful times usually correspond to a decay in the strength, a rise in polarization, and a decrease in spread.

5.3 Justifications of the approximations used in Section 3

Justification of Approximation 1: The estimated inverse exponential decay parameter B_H is around 0.44, which significantly surpasses the range of σ from 0.10 to 0.18. Consequently, for most pairs of individuals within the same party, their opinion difference $|B_j - B_i| \sim \sigma < B_H/3$. This approximation is appropriate for capturing the general trend, and the effect of a few extreme cases with a short timescale can be neglected.

Justification of Approximation 2: The estimated tolerance parameter B_T is around 0.22, and thus, $|B_j - B_i| \sim \sigma < 2B_T$. This suggests that any two individuals in the same party could potentially attract or repel each other based on their pairwise opinion difference. Hence, Approximation 2 is justified.

Justification of Approximation 3: For individuals i and j belonging to different parties, $|B_j - B_i| \approx B_R - B_D = 2\mu$, ranging from 0.54 to 0.90, and is larger than $2\ B_T = 0.44$. While Figure 1a indicates that a small number of pairs from different parties can have similar opinions that allow attractive interactions, most of the pairs from different parties interact repulsively. Because our model focuses on a qualitative understanding of the general trend of the polarization and spread, it is appropriate to exclude the small portion of individuals whose ideologies lie very close to the national social norm.

Justification of Approximation 4: The proportional factor, $A(t)$, determines the interaction efficiency in terms of communications among Congress members, with parameters A_0 and c estimated at 0.6 and 1.8, respectively. We conceptualize interactions among Congress members as an integration of cross-district interactions among all citizens. A Congress member's opinion level represents a district's average opinion, and an influence exerted by another Congress member represents influence from one district to another. Due to rapid advancements in communication technology over the past 150 years, interaction efficiency has exponentially increased, doubling every $\frac{Kln2}{c}\Delta t = \frac{78\times ln\,2}{1.8} \times 2 = 60$ years. Although quantitative justification is lacking, this trend aligns with previous qualitative observations [4, 23]. The proportional parameter of one individual influencing another in 2022 is $\frac{A_0 e^c}{N\Delta t} = \frac{0.6\times e^{1.5}}{225\times 2} = 0.006$ per year per person for ideology defined in the range from −1 to +1.

6 Discussions and areas for further studies

This work develops a mathematical model aimed at describing the evolution patterns of the opinion distributions within the

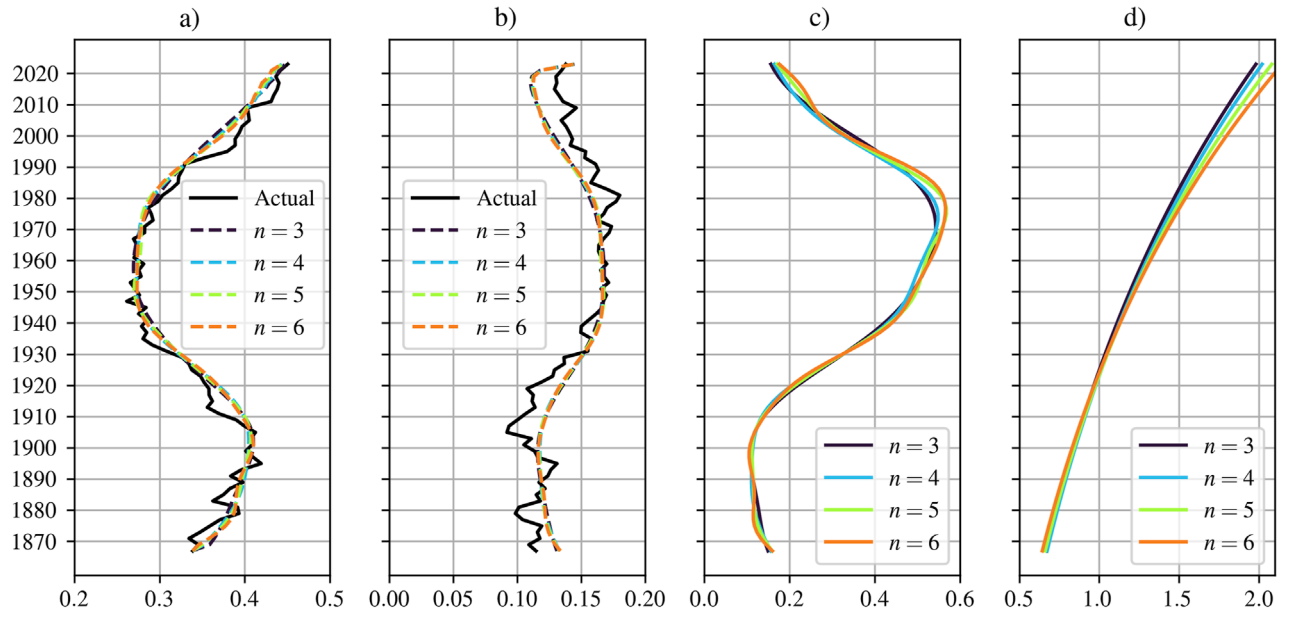


FIGURE 3
(a) The estimated polarization $\tilde{\mu}(t)$ compared with the corresponding actual data. **(b)** The estimated spread $\tilde{\sigma}(t)$ compared with the actual data. **(c)** The estimated strength function of the national social norm effect $\alpha(t)$. **(d)** The estimated influence parameter factor $\bar{A}(t)$. In all plots, n is the number of polynomials used in approximating the strength function of social norm $\alpha(t)$.

Democratic and Republican parties in the U.S. Congress. Grounded in opinion dynamic theory, the model captures the interplay of three time-dependent effects: the national social norm effect, cross-party interactions, and in-party interactions. These effects, whose relative importance varies over time, govern the evolution of polarization and spread within each party. Utilizing an algorithm for theory and data assimilation, we assimilate the model using the U.S. Congressional DW-DOMINATE dataset, which spans more than 150 years of recorded data. The time-varying polarization and spread patterns outlined by the model compare well with observations and are consistent with previous studies [4, 6, 11].

Notably, while many prior models focus solely on the modeling of either polarization or spread evolution, our model simultaneously captures these two interrelated quantities. For example, the "satisficing" model proposed by Yang et al. [8] models the party polarization with known party spread. Our model models both polarization and spread simultaneously and achieves better estimation on polarization. Moreover, whereas some previous models, such as the nonlinear feedback dynamic model [16], compared their results with datasets of limited temporal scope of less than 70 years, our model demonstrates the ability to explain trends across the entire 154 years.

In the model, cross-party interactions augment polarization, while in-party interactions foster spread. The national social norm effect always works to reduce both polarization and spread. By fitting the theory to observational data, we obtain a time-dependent strength function for the national social norm. This function is greater when the nation faces a severe threat and smaller when the nation experiences a peaceful time. Remarkably, periods of heightened threat correspond to lower polarization and greater spread, whereas periods of peace correspond to higher polarization and reduced spread. This finding aligns well with the important events that occurred in history, at least qualitatively.

It should be noted that future polarization $\tilde{\mu}$ and spread $\tilde{\sigma}$ cannot be determined based on the current dataset. According to Equations 8 and 9, the future values of these two variables depend on the future influence factor $A(t)$ and the future strength of the national social norm effect $\alpha(t)$, both of which are currently unknown. If the nation were to encounter a significant threat that leads to an increase in $\alpha(t)$ while $A(t)$ remains relatively constant, both polarization and spread are expected to decrease, and *vice versa*. Nonetheless, caution is warranted when applying Equations 8 and 9, as they are derived based on long-term dynamics and may overlook the impact of short-term effects.

The model maintains internal consistency, with the approximations made in deriving the theory being well-justified. The meanings and values of the four social physical parameters, namely, the tolerance parameter, the homophily influence decay parameter, and the two parameters governing the exponentially growing proportional impact factor, are interpretable. From a micro-scale perspective, the model suggests that the repulsion between pairs of individuals when their opinion difference exceeds the tolerance parameter is the root cause of opinion polarization and spread in the political system.

Based on our model, we suggest three potential strategies to mitigate excessive polarization and spread, strengthening shared civic identity to enhance the national norm effect $\alpha(t)$ [25]; promoting cross-party engagement to increase the homophily decay parameter B_H and thus lower mutual hostility [2, 4]; and encouraging in-party dialogue across differing views to increase the tolerance parameter B_T, thereby reducing opinion spread [2, 4].

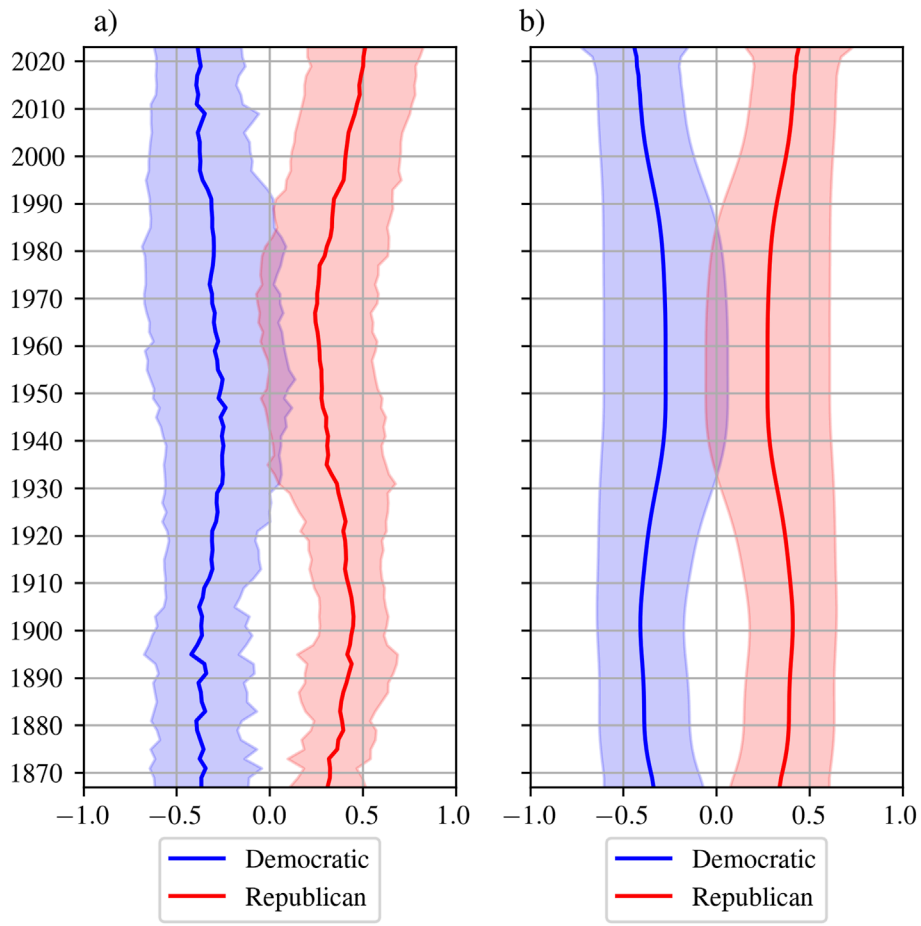
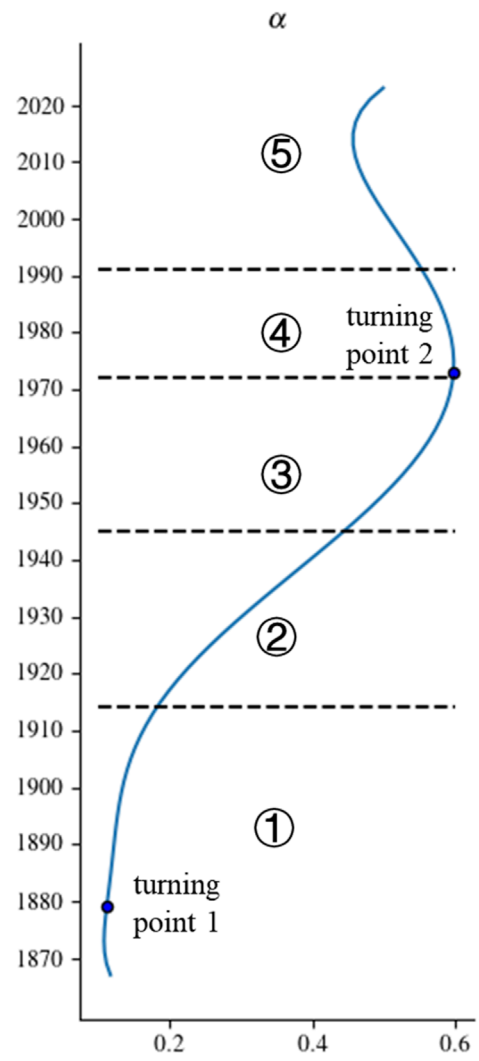


FIGURE 4
(a) The actual opinion evolution over time. The solid lines represent the party means, and the shaded area represents two standard deviations away from the means. **(b)** The estimated opinion evolution when $n = 7$.

Our work still contains weaknesses. Significant discrepancies between the model results and the actual data exist, most notably in the asymmetric opinion distribution of the two parties. For example, Figures 1b, 4a show a rightward shift of the Republican Party compared with the Democratic Party in recent decades. Modeling such asymmetric patterns may need to remove the symmetric approximation and introduce additional parameters, such as asymmetric tolerance and homophily decay parameters, for both in-party and cross-party dynamics. Studies [16, 30, 31] indicate that Republican supporters tend to exhibit lower in-party ideological tolerance, while Democratic supporters encompass a more ideologically diverse base. These well-documented insights merit further investigation in subsequent model refinement. However, any future extension, including the addition of asymmetric tolerance and decay parameters, would only be considered if those parameters are supported by robust empirical evidence and lead to significant improvements in the model's explanatory power.

Considering the nature of this study, we can only limit our conclusions in a qualitative manner before more work is done. To refine the model and develop operational strategies, additional studies are needed, especially with empirical experiments. We need to use measurable quantities for the quantification of the tolerance parameter describing the threshold of acceptance to rejection or the reverse, the homophily decay parameter describing the affinity change when people interact with each other, and the strength function of the national social norm effect. For example, the strength of the national social norm effect may be related to critical factors such as economic development, military capability, ideological frameworks, and public health systems.

It is important to emphasize that the present model is primarily interpretive rather than predictive. The model is designed to elucidate the underlying mechanisms linking polarization and spread with social–psychological processes rather than to forecast future trends of ideological development. Because the communication factor $A(t)$ and the national social norm strength $\alpha(t)$ are strongly influenced by unpredictable socio-political-technological events, it is difficult to predict polarization over the long term. Future extension of the model to have predictive capability will rely on the establishment of the link between $\alpha(t)$ to the dynamic interplay of measurable indicators such as the changes in culture, economy, politics, technology, and global interaction.



Periods	Historical Events
5	Osama Bin Laden Killed (2011)
	9/11 Attack (2001)
4	Breakup of the Soviet Union (1991)
	START I Treaty (1991)
	Berlin Wall Fell (1989)
	The U.S.'s Diplomatic Ties with China (1979)
	End of Vietnam War (1975)
	SALT I (1972)
	U.S. President First Visit to China (1972)
3	U.S. Military Intervened in the Vietnam War (1965)
	The Cuban Missile Crisis (1962)
	Berlin Wall Built (1961)
	Space Race—Soviet Union Launched 1st Satellite (1957)
	Start of Vietnam War (1955)
2	World War II (1939-1945)
	The Great Depression (1929-1939)
	World War I (1914-1918)
1	The Progressive Era (1890-1920)
	Rise of Industrial America (1876-1890)
	End of Civil War (1865)

Peace/Towards more polarizations
Threat/Towards less polarizations

FIGURE 5
The figure on the left displays the estimated strength function of the national norm effect $\alpha(t)$ divided into five periods. The table on the right lists the important historical events that occurred in each period.

We invite collaboration with sociologists and political scientists to improve the model by including more features and offering better interpretations of the relationship between the evolution of opinion distributions and the important social and political events, both past and current. For example, the mechanisms for the important findings of Jahani et al [32] that exposure to common enemies can increase political polarization need to be studied.

With proper extensions, our theory holds promise to study the opinion interactions of three or more groups of people, including communities, parties, countries, etc. By incorporating realistic communication network structures, the theory may be applied to explore dynamics in online opinion spaces. Much work needs to be done, including investigating the influence of opinion elites, the impact of social media, the effects of group norms, and the interplay of multiple interactive opinion topics. Collaboration across disciplines is essential for advancing our understanding of these complex dynamics.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.5281/zenodo.15274244.

Author contributions

XZ: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing. YH: Data curation, Software, Validation, Writing – original draft, Writing – review and editing. YZ: Formal Analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing.

Funding

The authors declare that financial support was received for the research and/or publication of this article. This work is partially supported by the Beijing Institute of Mathematical Sciences and Applications.

Acknowledgements

We appreciate the useful assistance from and discussions with our colleagues Prof. Miao He and Prof. Wuyue Yang.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2025.1706465/full#supplementary-material

References

1. McCarty N, Poole KT, Rosenthal H. *Polarized America: the dance of ideology and unequal riches. Chapter 2. The polarization of the politicians*. London, UK: MIT Press (2016).

2. Axelrod R, Daymude JJ, Forrest S. Preventing extreme polarization of political attitudes. *Proc Natl Acad Sci U S A* (2021) 118(50):e2102139118. doi:10.1073/pnas.2102139118

3. Hill SJ, Tausanovitch C. A disconnect in representation? Comparison of trends in congressional and public polarization. *J Polit* (2015) 77(4):1058–75. doi:10.1086/682398

4. Stefano B, Lise G, Daniel GG, Duncan JW. Reducing opinion polarization: effects of exposure to similar people with differing political views. *Proc Natl Acad Sci* (2021) 118(52):e2112552118. doi:10.1073/pnas.2112552118

5. Jost JT, Baldassarri D, Druckman JN. Cognitive–motivational mechanisms of political polarization in social-communicative contexts. *Nat Rev Psychol* (2022) 1(10):560–76. doi:10.1038/s44159-022-00093-5

6. Cole JC, Gillis AJ, van der Linden S, Cohen MA, Vandenbergh MP. Social psychological perspectives on political polarization: insights and implications for climate change. *Perspect Psychol Sci* (2023) 20(1):115–141. doi:10.1177/17456916231186409

7. Downs A. An economic theory of political action in a democracy. *J Polit Econ* (1957) 65:135–50. doi:10.1086/257897

8. Yang VC, Abrams DM, Kernell G, Motter AE. Why are US parties so polarized? A "satisficing" dynamical model. *SIAM Rev* (2020) 62(3):646–57. doi:10.1137/19m1254246

9. Lanzetti N, Hajar J, Dörfler F. *Modeling of political systems using wasserstein gradient flows*. In: IEEE 61st conference on decision and control (CDC); 2022 December 06–09; Mexico, Cancun: IEEE (2022). p. 364–9.

10. Jones MI, Sirianni AD, Fu F. Polarization, abstention, and the median voter theorem. *Humanit Soc Sci Commun* (2022) 9(43):43. doi:10.1057/s41599-022-01056-0

11. Ferri I, Díaz-Guilera A, Palassini M. *Equilibrium and dynamics of a three-state opinion model* (2022). doi:10.48550/arXiv.2210.03054

12. Liu CC, Srivastava SB. Pulling closer and moving apart: interaction, identity, and influence in the U.S. senate, 1973 to 2009. *Behav Modification* (2015) 39(1):192–217. doi:10.1177/0003122414564182

13. Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ. The origins and consequences of affective polarization in the United States. *Annu Rev Polit Sci* (2019) 22:129–146. doi:10.1146/annurev-polisci-051117-07303

14. Finkel EJ, Bail CA, Cikara M, Ditto PH, Iyengar S, Klar S, et al. Political sectarianism in America. *Science* (2020) 370:533–6. doi:10.1126/science.abe1715

15. Lu X, Gao J, Szymanski BK. The evolution of polarization in the legislative branch of government. *J. R. Soc.* (2019) 16.

16. Leonard NE, Lipsitz K, Bizyaeva A, Franci A, Yphtach LY. The nonlinear feedback dynamics of asymmetric political polarization. *Proc Natl Acad Sci* (2021) 118(50):e2102149118. doi:10.1073/pnas.2102149118

17. Baldassarri D, Bearman P. Dynamics of political polarization. *Am Sociological Rev* (2007) 72(5):784–811. doi:10.1177/000312240707200507

18. Baldassarri D, Page SE. The emergence and perils of polarization. *Proc Natl Acad Sci U S A* (2021) 118:e2116863118. doi:10.1073/pnas.2116863118

19. Lewis JB, Poole K, Rosenthal H, Boche A, Rudkin A, Sonnet L. *Voteview: congressional roll-call votes database*. Los Angeles: voteview (2021). Available online at: https://voteview.com/ (Accessed November 8, 2025).

20. Boche A, Lewis JB, Rudkin A, Sonnet L. The new voteview.com: preserving and continuing keith Poole's infrastructure for scholars, students and observers of congress. *Public Choice* (2018) 176:17–32. doi:10.1007/s11127-018-0546-0

21. He M, Zhang XJ. Affinity, value homophily, and opinion dynamics: the co-evolution between affinity and opinion. *PLoS One* (2023) 18:e0294757. doi:10.1371/journal.pone.0294757

22. DeGroot MH. Reaching a consensus. *J Am Stat Assoc* (1974) 69(345):118–21. doi:10.1080/01621459.1974.10480137

23. Odlyzko A. The history of communications and its implications for the internet. *SSRN Electron J* (2000). doi:10.2139/ssrn.235284

24. Montesinos HM, Connors J, Gwartney J. The transportation-communication revolution: 50 years of dramatic change in economic development. *Cato J* (2020) 40(1):153–198. doi:10.36009/CJ.40.1.9

25. Van Der Linden S, Leiserowitz A, Maibach E. Communicating the scientific consensus on human-caused climate change is an effective and depolarizing public engagement strategy: experimental evidence from a large national replication study. *SSRN Electron J* (2016). doi:10.2139/ssrn.2733956

26. Flavien A, Ben Mansour D, Mazen S. Joint state-parameter estimation and inverse problems governed by reaction–advection–diffusion type PDEs with application to biological keller–segel equations and pattern formation. *J Comput Appl Mathematics* (2025) 461:116454. doi:10.1016/j.cam.2024.116454

27. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* (2019) 378:686–707. doi:10.1016/j.jcp.2018.10.045

28. Lan H, Zhao S, Hu J, Wang Z, Fu J. Joint state estimation and noise identification based on variational optimization. *IEEE Trans Automatic Control* (2025) 70:4500–15. doi:10.1109/TAC.2024.3524270

29. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: *Proceedings of the 3rd international conference on learning representations*; 2015 May 7–9; San Diego, CA, USA (2015).

30. Rawlings CM, Childress C. The polarization of popular culture: tracing the size, shape, and depth of the "Oil Spill". *Social Forces* (2023) 102:1582–607. doi:10.1093/sf/soad150

31. Lelkes Y, Sniderman PM. The ideological asymmetry of the American party system. *Br J Polit Sci* (2016) 46:825–44. doi:10.1017/s0007123414000404

32. Jahani E, Gallagher N, Merhout F, Cavalli N, Guilbeault D, Leng Y, et al. An online experiment during the 2020 US-Iran crisis shows that exposure to common enemies can increase political polarization. *Sci Rep* (2022) 12:19304. doi:10.1038/s41598-022-23673-0