Extended data figure for "Dynamics of collective mind in online news communities"

April 10, 2025

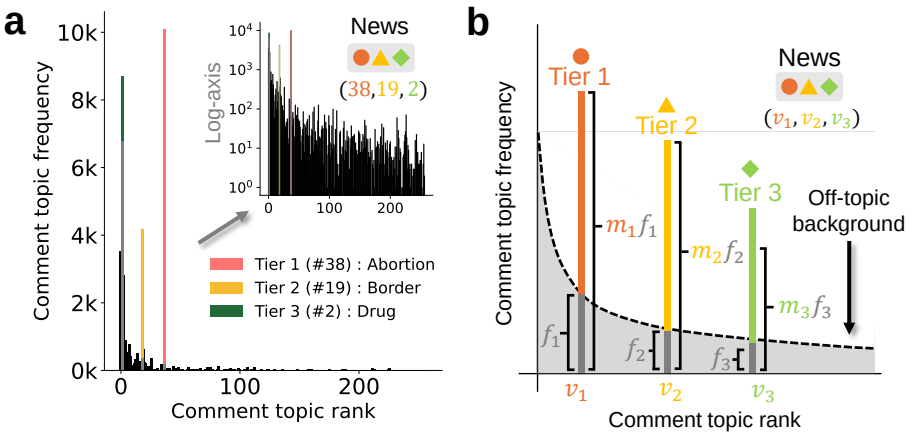

Figure E1: **Response of community semantic network to the news. a**, Illustrative example of the topic frequency distribution of the comments posted under all news characterized by certain topics (here, by the triplet $(38, 19, 2)$) and its log scale plot (inset). Gray bars represent the expected frequencies of these topics when they are not in the news. Data is taken from all-time aggregated data of The Hill. **b**, Conceptual diagram that shows the modeling of the community semantic network's response (comment frequency) to the news (v_1, v_2, v_3) in the proposed model. Each of the on-topic comments' frequencies (f_i) is multiplied by comment multiplier m_i, while other off-topic comments' frequencies follow the previous comment frequency distribution as a common background signal.

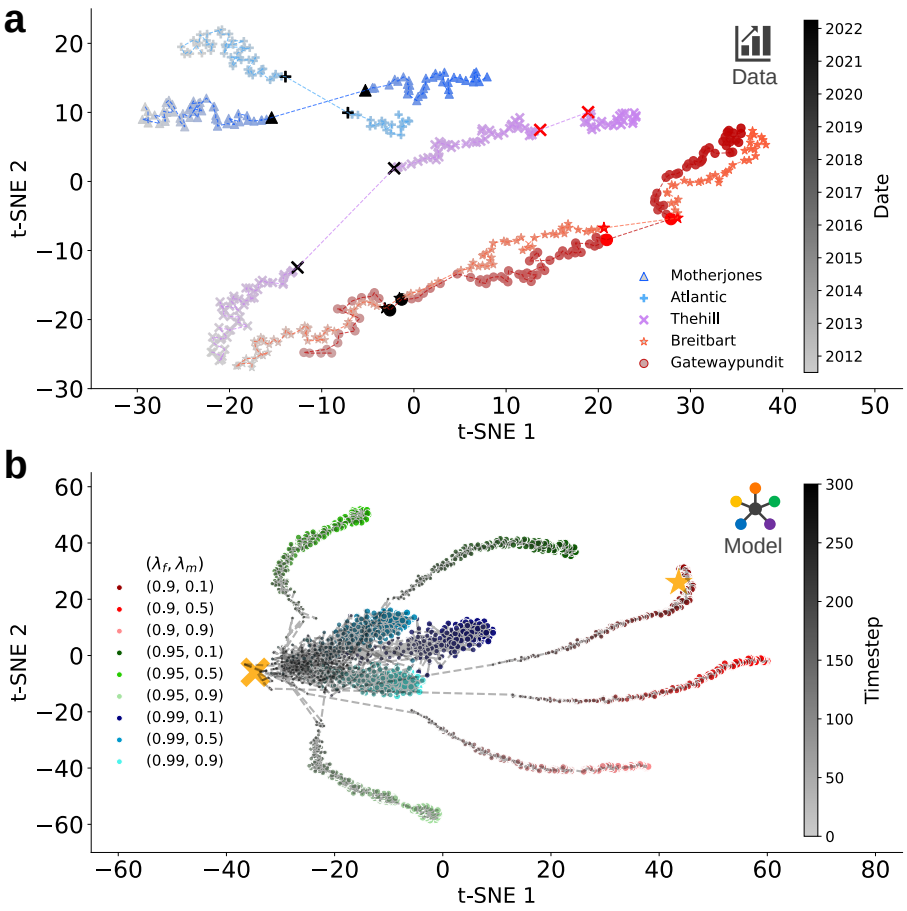

Figure E2: **Trajectories of comment topic profile a**, Empirical t-SNE plot of the trajectories of the comment topic profile for online news communities. t-SNE algorithm is employed to visualize the trajectory of the 228-dimensional comment topic profile in the 2D space. Two notable jumps in trajectories that affected all communities significantly are denoted as black (Jun/Jul 2016, Orlando mass shooting) and red (Feb/Mar 2020, COVID-19) markers. **b**, Model-based t-SNE plot of the trajectories of the comment frequency profile with diverse filter strength and memory strength, all started from the same initial frequency (orange cross) and attracted by the same general semantic network (orange star). In the model, t-SNE algorithm is applied to the 250-dimensional model comment topic profile, which corresponds to the comment topic profile of the empirical data. We can observe that the lower the filter strength and the memory strength, the more the trajectory is affected by the general semantic network and quickly converges to it.

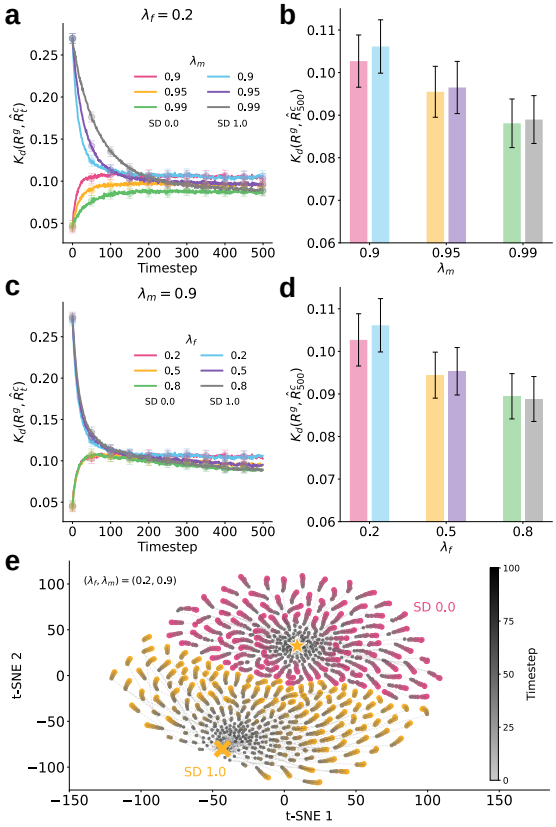

Figure E3: **Behavior of the comment topic profile from model simulations.** **a**, Kendall-tau rank distance (K_d) between relative topic frequencies of general semantic network (R^g) and comment frequencies of community semantic network at timestep t ($\hat{R}_t^c$) for various λ_m $(0.9, 0.95, 0.99)$ and different initial state (SD 0.0, 1.0) with fixed filter strength ($\lambda_f = 0.2$). The final distance at $t = 500$ is highlighted in **b**, where data with the same λ_m are grouped. **a**, Kendall-tau rank distance (K_d) between relative topic frequencies of general semantic network (R^g) and comment frequencies of community semantic network at timestep t ($\hat{R}_t^c$) for various λ_f $(0.2, 0.5, 0.8)$ and different initial state (SD 0.0, 1.0) with fixed memory strength ($\lambda_m = 0.9$). The final distance at $t = 500$ is highlighted in **d**, where data with the same λ_f are grouped. Data is gathered from $1,000$ iterations and the error bar indicates ± 1 standard deviation. For **a** and **c**, error bars are plotted every 50 timestep. **a** shares the legend with **b** and **c** shares the legend with **d**. **e** t-SNE plot of the trajectories of the comment frequency profile for 100 model simulation each, where the initial community frequencies are perturbed from general frequencies by log-normal noise with standard deviation (SD) of 0.0 (red) and 1.0 (yellow). All trajectories started from the same initial frequency (yellow cross) and attracted by the same general semantic network (yellow star), and $(\lambda_f, \lambda_m) = (0.2, 0.9)$.