Appendix

A Reproducibility

Our code and data can be downloaded from `https://anonymous.4open.science/r/conflict-awareness-00F0/README.md`.

B Broader Impacts

Our analysis reveals the types of social links that, when added to the social network, can most effectively reduce polarization and disagreement. While this result itself is for a good cause, a potential risk exists when one interprets it into the opposite direction: we now know certain types of links that, when removed from the social network, can most effectively **increase** polarization and disagreement. This could be abused by an (authoritative) adversarial to increase polarization and disagreement by diminishing social ties among certain people, or even disconnecting them.

To mitigate this risk, we suggest social platforms take more cautious steps when deciding to reduce the exposure of one person's content feed to another, such as additional algorithmic check in background, as well as more security measures to guard against the hacking of platform's administrative authority. Researchers are also encouraged to study network structures that are more robust to attacks of such kind, as well as defense measures to be taken when such attacks actually happen.

C Proofs

C.1 Proof of Theorem 1

Proof. Let L_{+e} denote the Laplacian matrix of the new social network after adding a new link $e = (i, j)$. To prove Eq.(2), we invoke the Sherman-Morrison Formula [54] for computing the inverse of rank-1 update to an invertible matrix. Notice that $G_{+e} = L + b_e b_e^T$. Therefore,

$$\begin{aligned}
\mathcal{C}(G_{+e}, s) - \mathcal{C}(G, s) &= s^T((I + G_{+e})^{-1} - (I + L)^{-1})s \\
&= s^T((I + L + b_e b_e^T)^{-1} - (I + L)^{-1})s \\
&= -s^T \frac{(I + L)^{-1} b_e b_e^T (I + L)^{-1}}{1 + b_e^T (I + L)^{-1} b_e} s \\
&= -\frac{|b_e^T (I + L)^{-1} s|_2^2}{1 + b_e^T (I + L)^{-1} b_e} \\
&= -\frac{(z_i - z_j)^2}{1 + b_e^T (I + L)^{-1} b_e}.
\end{aligned}$$

L is positive semidefinite, so $(I + L)^{-1}$ is also positive semidefinite. Therefore, $1 + b_e^T (I + L)^{-1} b_e$ is positive, and so $-\frac{(z_i - z_j)^2}{1 + b_e^T (I + L)^{-1} b_e} \le 0$.

To prove Eq.(3), we further note that

$$\begin{aligned}
\mathbb{E}_s[\mathcal{C}(G_{+e}, s) - \mathcal{C}(G, s)] &= \mathbb{E}_s[-s^T \frac{(I + L)^{-1} b_e b_e^T (I + L)^{-1}}{1 + b_e^T (I + L)^{-1} b_e} s] \\
&= \mathbb{E}_s[-\frac{b_e^T (I + L)^{-1} s s^T (I + L)^{-1} b_e}{1 + b_e^T (I + L)^{-1} b_e}] \\
&= -\frac{b_e^T (I + L)^{-1} \mathbb{E}_s[s s^T] (I + L)^{-1} b_e}{1 + b_e^T (I + L)^{-1} b_e} \\
&= -\frac{b_e^T (I + L)^{-1} (\sigma^2 I) (I + L)^{-1} b_e}{1 + b_e^T (I + L)^{-1} b_e} \\
&= -\frac{\sigma^2 |(I + L)^{-1} b_e|_2^2}{1 + b_e^T (I + L)^{-1} b_e} \le 0.
\end{aligned}$$

□

C.2 Proof of Theorem 3

Proof. Let $M = I + L$, and let matrix C be the co-factor matrix of M, then $(I+L)^{-1} = M^{-1} = |M|^{-1}C$. Therefore, $b_e^T(I+L)^{-1}b_e = |M|^{-1}(C_{ii} + C_{jj} - C_{ij} - C_{ji})$. $|M|$ is the determinant of matrix M. [55] presents a result that $|M|$ equals the total number of spanning rooted forests of G, and C_{xy} equals the total number of spanning rooted forests of G, in which node x and y belong to the same tree rooted at x. The theorem is proved by substituting this previous result back into $|M|^{-1}(C_{ii} + C_{jj} - C_{ij} - C_{ji})$. □

C.3 Proof of Theorem 4

Proof.

$$\begin{aligned}\sigma^2 |(I+L)^{-1}b_e|_2^2 &= \sigma^2 \sum_{k\in V} (|M|^{-1}C_{ik} - |M|^{-1}C_{jk})^2 \\ &= \sigma^2 |M|^{-2} \sum_{k\in V} (C_{ik} - C_{jk})^2\end{aligned}$$

Since M is symmetric, we have $C_{ik} + N_{ik} = C_{ki} + \mathcal{N}_{ki} = C_{kk}$, $C_{jk} + \mathcal{N}_{jk} = C_{kj} + \mathcal{N}_{kj} = C_{kk}$, where C_{kk} according to [55] is equal to the total number of spanning rooted forests where node k is at the root of the tree to which k belongs. Joining the two equations, we have $C_{ik} - C_{jk} = \mathcal{N}_{ik} - \mathcal{N}_{jk}$. Therefore,

$$\sigma^2 |(I+L)^{-1}b_e|_2^2 = \sigma^2 \mathcal{N}^{-2} \sum_{k\in V} (\mathcal{N}_{ik} - \mathcal{N}_{jk})^2$$

□

C.4 Proof of Corollary 1

Proof. The correctness quickly follows from substituting Equations (5, 6) into Equation (2). □

C.5 Proof of Proposition 1

Proof. To show that the objective is convex, we resort to the result in [56], Example 9: X^{-1} is a matrix convex on the set of all nonnegative invertible Hermitian matrices. Obviously $I + L + L_f$ is nonnegative, invertible and symmetric, so it is a matrix convex. Therefore, the objective is convex. Any convex combination of Laplacians is still a Lapalacian. The trace of any convex combination of of matrices cannot exceed the trace of any members. Therefore, the feasible region is also convex. □

C.6 Expected Conflict Awareness

Definition 2. *Given a social network G and a budget $\beta > 0$, the* ***conflict awareness over Expectation (CAE)*** *of a link addition function $f(e; G, \beta)$ is likewise defined as:*

$$\textbf{CAE}(f) \equiv \frac{\Delta_f \mathbb{E}_s[\mathcal{C}]}{\Delta_{f^*} \mathbb{E}_s[\mathcal{C}]} \tag{14}$$

where

$$\Delta_f \mathbb{E}_s[\mathcal{C}] \;\equiv\; \sigma^2[\mathrm{Tr}((I + L + L_f)^{-1}) - \mathrm{Tr}((I+L)^{-1})] \tag{15}$$

$$\Delta_{f^*} \mathbb{E}_s[\mathcal{C}] \;\equiv\; \min_{L_f} \quad \Delta_f \mathbb{E}_s[\mathcal{C}] \tag{16}$$

$$\text{subject to} \quad L_f \in \mathcal{L} \textit{ (Laplacian constraint)} \tag{17}$$

$$\mathrm{Tr}(L_f) \le 2\beta \textit{ (budget constraint)} \tag{18}$$

Proposition 2. *The definition of $\Delta_f \mathbb{E}_s[\mathcal{C}]$ above is consistent with that of $\Delta_f \mathcal{C}$ in Definition 1 in the sense that they satisfy $\Delta_f \mathbb{E}_s[\mathcal{C}] \equiv \int_s \rho(s)\; \Delta_f \mathcal{C}\; d_s$.*

Proof. Let A be any square matrix of the same shape as L. Then $\int_s \rho(s) s^T A s\, d_s = \int_s \rho(s) s^T (As)\, d_s = \int_s \rho(s) \mathrm{Tr}((As)s^T)\, d_s = \int_s \rho(s)\mathrm{Tr}(A(ss^T))\, d_s = \mathrm{Tr}(A \int_s \rho(s)(ss^T)\, d_s) = \mathrm{Tr}(A(\sigma^2 I)) = \sigma^2 \mathrm{Tr}(A)$. By substituting $A = (I + L + L_f)^{-1}$ and $A = (I+L)^{-1}$ into Eq. (9) respectively, the proposition is proved.

□

Proposition 3. *In Definition 2, $\Delta_{f^*}\mathbb{E}_s[\mathcal{C}]$ is also the objective of a convex optimization problem.*

Proof. From the proof for Proposition 1, it suffices to only show that the $\Delta_f\mathbb{E}_s[\mathcal{C}]$ in Equation 15 is convex in L_f given other variables fixed. Notice that we mentioned $\Delta_f\mathbb{E}_s[\mathcal{C}] \equiv \int_s \rho(s)\ \Delta_f\mathcal{C}\ d_s$, in which $\rho(s) \geq 0$, $\Delta_f\mathcal{C}$ can be viewed as a function of L_f and s, and is convex in L_f given s to be further fixed. Therefore, the integral $\Delta_f\mathbb{E}_s[\mathcal{C}]$ is also convex in L_f. □

C.7 Proof of Theorem 2

Proof. Let $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be eigenvalues of L in ascending order; the eigen decomposition of $L = U\Lambda U^T$ where $\Lambda = \text{diag}([\lambda_1, ...\lambda_n])$ and U is the corresponding orthornormal matrix satisfying $UU^T = I$. Notice that $(I+L)^{-1} = U(I+\Lambda)^{-1}U^T$.

$$\frac{\mathcal{C}(G_0, s)}{\mathcal{C}(G, s)} = \frac{s^T s}{s^T(I+L)^{-1}s} = \frac{s^T UU^T s}{s^T U(I+\Lambda)^{-1}U^T s}$$

let $s' = U^T s$, and further notice that since we have assumed s to be zero-centered (see Section2), $s'_1 = 1^T s = 0$. We can further rewrite:

$$\frac{\mathcal{C}(G_0, s)}{\mathcal{C}(G, s)} = \frac{s'^T s'}{s'^T(I+\Lambda)^{-1}s'} = \frac{\sum_{i=1}^n {s'_i}^2}{\sum_{i=1}^n (1+\lambda_i)^{-1}{s'_i}^2} = \frac{\sum_{i=2}^n {s'_i}^2}{\sum_{i=2}^n (1+\lambda_i)^{-1}{s'_i}^2}$$

It is not hard to see that

$$1+\lambda_n \geq \frac{\mathcal{C}(G_0, s)}{\mathcal{C}(G, s)} \geq 1+\lambda_2$$

For the upper bound, [57] shows that $\lambda_n \leq \max_{(i,j)\in E}(d_i + d_j)$; for the lower bound, we know from Lemma A.1 of [38] that $\lambda_2 \geq \frac{1}{2}d_{\min}h_G^2$, where d_{min} is the minimum node degree in G; h_G is the Cheeger constant of G. Substituting these back into expression above, we have $1 + \max_{(i,j)\in E}(d_i + d_j) \geq \frac{\mathcal{C}(G_0,s)}{\mathcal{C}(G,s)} \geq 1 + \frac{1}{2}d_{\min}h_G^2 \geq 1$. □

C.8 Proof of Theorem 5

Proof. Notice that $\mathcal{C} = s^T(I+L)^{-1}s$, and $\mathcal{U} = \mathcal{P} + \mathcal{I} = s^T(I+L)^{-2}s + s^T(I-(I+L)^{-1})^2 s = s^T(I-(I+L)^{-1})s$. Therefore, $\mathcal{C} + \mathcal{U} = s^T s$ which is a constant. □

D Experiments

D.1 Verifying the Direction of Conflict Change (Theorem 1)

We computationally verify that opinion conflict always gets reduced when a new link is added to the network. We use six datasets, including three synthetic networks and three real-world social networks. The synthetic networks are, a Erdős–Rényi Graph ($n = 100, p = 0.5$), a path graph ($n = 100$), a 10 by 10 2D-grid graph. The real-world networks are, the Karate club social network, Reddit, and Twitter (as introduced in Sec.D.3). For each network, we compute the amount of conflict change caused by adding a link between every pair of disconnected node in the graph, with each link replaced one at a time. Figure 4 shows the distributions of the amounts of conflict conflict for all the six datasets. We can see that they are all on the negative side of the axis. This result validates the negative sign in Theorem 1 and demonstrates its broad applicability.

D.2 Verifying Conflict Contraction (Theorem 2)

We start with an empty graph with N nodes. In each iteration, one edge is randomly added between two disconnected nodes; we then compute the the lower bound, the upper bound, and the conflict contraction rate as given in Theorem 2. The iterations stop when no pair of nodes are left disconnected (*i.e.* the graph is complete). We choose $N = 20$ in this experiment as computing the Cheeger constant term is NP-hard.

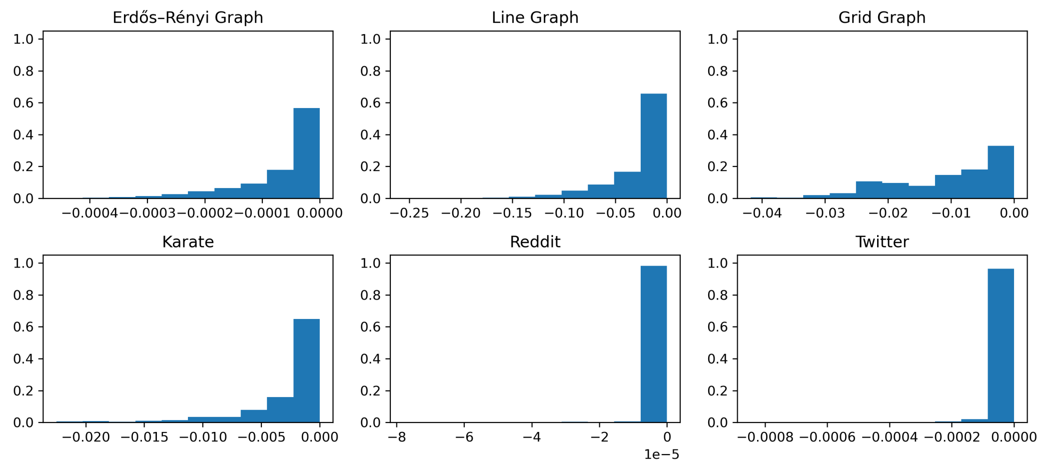

Figure 4: Computational validation for Theorem 1.

Figure 5 plots the lower bounds, the lower bounds, the upper bounds, and the conflict contraction rates, with respect to the increasing numbers of edges in the graph. We can see that the conflict contraction rates are indeed lying in between the two bounds. The gap exists because we cannot exhaust all the possible graphs on 20 nodes. Nevertheless, this experiment provides a good piece of evidence that Theorem 2 is correct.

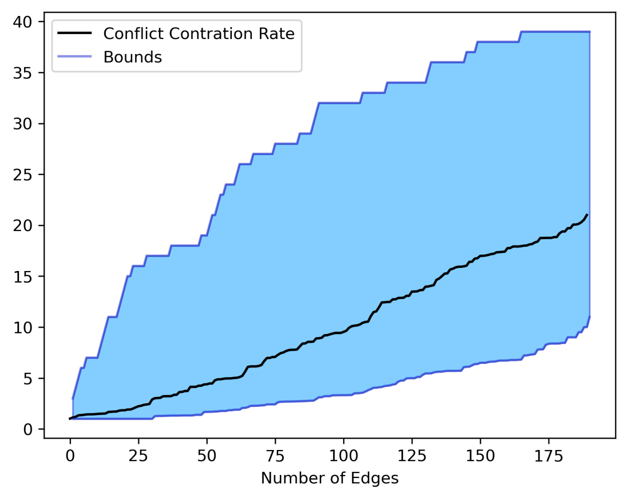

Figure 5: Computational validation for Theorem 2.

D.3 Dataset and Preprocessing

Twitter. The dataset is extracted from a number of tweets relevant to the Delhi Assembly elections 2013. In the preprocessing, only the largest strongest-connected component (SCC) gets retained, which contains 548 users and 3638 undirected edges; each edge represents a pair of follower and followee. The initial opinions (s) were mapped by a sentiment analysis tool designed for Twitter [58], based on each user's first-hour tweets in the record window.

Reddit. The dataset is extracted from the subreddit of "Politics" between 07/2013 and 12/2013. Similar to Twitter, only the largest SCC is retained, containing 556 users and 8969 edges. An edge exists between two users if both of them posted in the same subreddit other than "Politics" during the aforementioned time period. The initial opinions were mapped using the standard linguistic analytics tool LIWC [59].

D.4 Linear Scaling of the Output

To make sure that the weights of all recommended links sum up to β, we linearly scale each link recommendation algorithm's output by a normalizing constant: Notice that each link recommendation algorithms is essentially a scoring function on the links. For a model h, its output weight $w_h(e)$ of each recommended link e follows the normalized form $w_h(e) = \beta \frac{s_h(e)}{\sum_e s_h(e)}$, where $s_h(e)$ is the original score that model h assigns to link e.

D.5 Precision@10

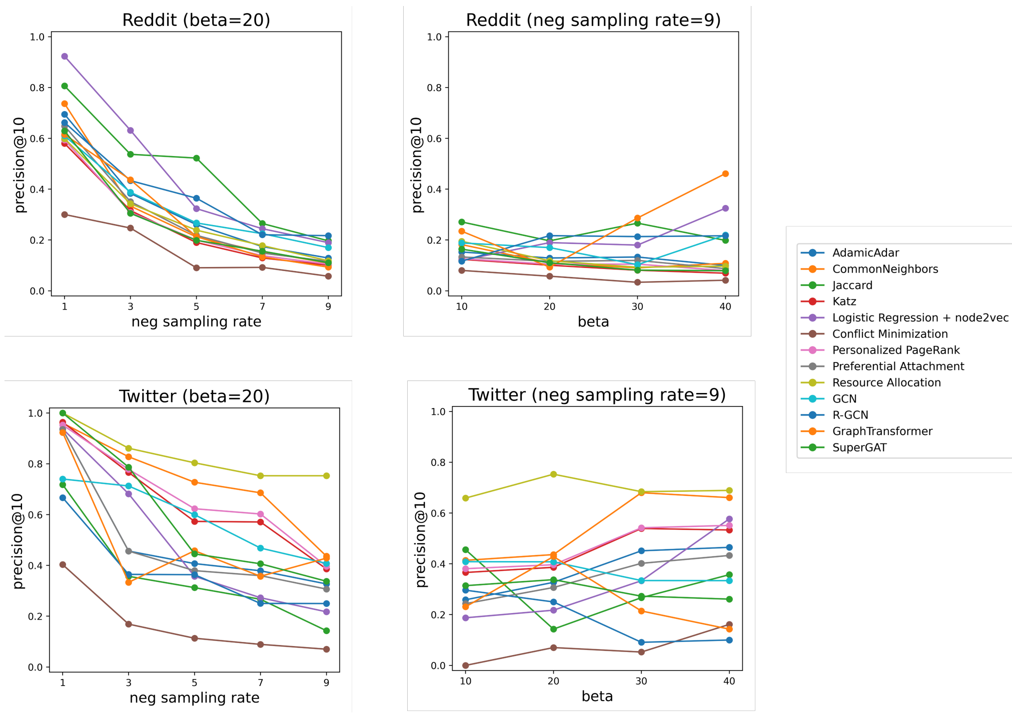

Figure 7: Precision@10 of 13 link recommendation algorithms on samples of Reddit (upper) and Twitter (lower) social network. These plots supplement the recall measurement in Fig. 2 as another proxy for "relevance".