How Predicted Links Influence Network Evolution: Disentangling Choice and Algorithmic Feedback in Dynamic Graphs

Mathilde Perez[1] **Raphaël Romero**[2] **Jefrey Lijffijt**[2] **Charlotte Laclau**[1]

[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
[2]Ghent University–IDLab, Belgium

Abstract

Link prediction models are increasingly used to recommend interactions in evolving networks, yet their impact on network structure is typically assessed from static snapshots. In particular, observed homophily conflates intrinsic interaction tendencies with amplification effects induced by network dynamics and algorithmic feedback. We propose a temporal framework based on multivariate Hawkes processes that decomposes these two sources and introduce an instantaneous bias measure derived from interaction intensities, capturing current reinforcement dynamics beyond cumulative metrics. We provide a theoretical characterization of the stability and convergence of the induced dynamics, and experiments show that the proposed measure reliably reflects algorithmic feedback effects across different link prediction strategies.

1 INTRODUCTION

Social networks are known to be homophilic, meaning that *similar* individuals are more likely to be connected McPherson et al. [2001]. Following Kossinets and Watts [2009], this homophily can be decomposed into two distinct components, namely *choice homophily* and *induced homophily*. Choice homophily refers to the tendency of individuals to form or maintain ties as a function of their own characteristics, such as demographic attributes, interests, values, or social status, independently of the existing network structure. This tendency does not necessarily imply a conscious or deliberate preference for similarity; it may arise from latent preferences, social dispositions, or constraints associated with individuals themselves. For this reason, choice homophily is typically viewed as intrinsic to individuals. Induced homophily, on the other hand, arises from the structure and dynamics of the system in which interactions take place. It is driven by factors such as network topology, time and order of interactions, or algorithmic mechanisms that shape visibility and opportunities for interaction. In practice, network structure and algorithmic intervention tend to reinforce homophily, even when individual preferences remain stable. Local structural effects [Asikainen et al., 2020, Granovetter, 1973, Currarini et al., 2009] and repeated exposure mechanisms [Chaney et al., 2018, Pariser, 2011, Flaxman et al., 2016] can amplify existing patterns of connections, leading to increasingly segregated interaction patterns.

Beyond its sociological interpretation, this distinction between choice and induced homophily has important implications for how learning algorithms on graphs are evaluated and designed, yet it is rarely made explicit in current graph learning pipelines. In link prediction (LP) and node classification, models are trained to exploit the connectivity patterns present in the observed graph. When interaction patterns are segregated, learning algorithms tend to reproduce and sometimes amplify these patterns. Homophily is therefore often used as a proxy for bias in graph-based learning. This perspective is also reflected in the literature, where many debiasing approaches, whether based on graph rewiring [Laclau et al., 2020, Spinelli et al., 2021], the use of explicit priors [Buyl and De Bie, 2020], or on Graph Neural Networks (GNNs) [Zheng et al., 2024, Zhang et al., 2025, for instance], explicitly aim to reduce homophily with respect to a given *sensitive* attribute.

In this paper, we study how distinguishing statistically intrinsic and induced sources of homophily matters for the design and evaluation of (fair) link prediction models. We believe that this distinction is important to reason about observed disparities and potential sources of unfairness in graph-based learning. To proceed, we move beyond a static view of the graph and consider its temporal evolution. Indeed, real-world networks are dynamic, and link prediction models directly affect the network's future structure by shaping which connections are formed. The need to adopt this dynamic perspective also stems from several recent studies that explore the long-term impact of fairness in dynamic settings,

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:5374–5392.

using simulated environments [D'Amour et al., 2020, Yin et al., 2023, Rateike et al., 2024] or the repeated empirical risk minimization framework [Liu et al., 2018, Hashimoto et al., 2018]. These works emphasize that accounting for feedback mechanisms is crucial and that fairness interventions can have delayed and counterintuitive effects over time. Next, we detail the most related contributions, which are at the intersection of fairness and dynamic graphs.

Related work. Asikainen et al. [2020] and Vega-Oliveros et al. [2019] are the studies closest to our setting, as both analyze the interaction between algorithmic mechanisms and network structure in dynamic graphs. Asikainen et al. [2020] show that weak intrinsic preferences can be amplified by endogenous processes such as triadic closure, leading to cumulative segregation. We share the distinction between intrinsic and induced homophily and a dynamic viewpoint, but depart from their framework by modeling interactions with Hawkes processes and explicitly integrating link prediction (LP) into the network evolution, enabling a fairness-oriented analysis of algorithmic mediation. Vega-Oliveros et al. [2019] study how different link recommendation strategies affect information diffusion and induce topological changes in the network, under various diffusion models. Their analysis is primarily empirical and does not explicitly model homophily or group-level disparities.

From a complementary angle, closely related work in the field of recommender systems [Akpinar et al., 2022, Fabbri et al., 2022, Cao et al., 2025] studies fairness in terms of exposure and visibility over time, highlighting the joint role of homophily, group size imbalance, and algorithmic feedback. While these approaches emphasize the importance of temporal effects, they typically abstract away the dynamics of interactions within the network itself, or rely on algorithm-dependent notions of fairness.

Finally, several works leverage Hawkes processes to model temporal interactions in dynamic graphs and to learn time-aware node representations [Liu et al., 2020, Wang et al., 2025]. Although these methods focus on representation learning rather than on homophily or fairness, they further support the relevance of Hawkes-type models for capturing the dynamics of evolving networks.

Contributions. Our work is guided by the following central question: *how do intrinsic interaction and algorithmic feedback jointly shape the long-term evolution of homophily and structural disparities in dynamic networks?*
To address this question, we make the following contributions. (i) We propose a Hawkes-process-based model of network evolution that explicitly distinguishes statistically baseline interaction propensities from reinforcement effects driven by past interactions and exposure, enabling a principled distinction between choice and induced homophily. (ii) We provide a tractable mean-field characterization of the temporal dynamics of homophily and group-level disparities, allowing us to analyze stability and long-term behaviors under different reinforcement regimes and algorithmic feedback mechanisms. (iii) We study the short- and long-term impact of fairness-aware LP strategies on network structure, highlighting how feedback loops can induce delayed effects on homophily and structural disparities. These contributions are reflected in the structure of the paper, from modeling to theory and empirical evaluation.

2 MOTIVATION AND BACKGROUND

For clarity, we first review the general setting of (fair) LP and its connection to homophily in graphs. We then briefly introduce the general framework of Hawkes processes on graphs, which forms the basis of our model.

2.1 FAIRNESS ON GRAPHS AND HOMOPHILY

Link prediction and dyadic fairness. We consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where each node $u \in \mathcal{V}$ has a sensitive attribute $S_u \in \mathbb{S}$ indicating its demographic group (e.g., $\mathbb{S} = \{0, \dots, K\}$, where K is the number of groups). The LP task consists in learning a function $h : \mathcal{V} \times \mathcal{V} \to \{0, 1\}$ that estimates the probability of an edge between node pairs.

Group fairness at the interaction level, commonly referred to as *dyadic fairness* [Li et al., 2021], aims to ensure balanced treatment of within-group and cross-group interactions. A standard metric is the demographic parity gap:

$$\Delta_{\text{DP}} = \big|\mathbb{P}(h(u,v) = 1 \mid S_u = S_v) - \mathbb{P}(h(u,v) = 1 \mid S_u \neq S_v)\big|,$$

where S_u and S_v denote the groups of nodes u and v. This metric directly quantifies disparities in predicted link rates between same-group and cross-group pairs.

Homophily as a proxy for bias. Dyadic fairness metrics are closely related to homophily with respect to the sensitive attribute S. While there is no consensus on how to measure homophily [Berry et al., 2021], various structural measures have been proposed. Assortativity [Newman, 2002] captures the tendency of nodes to connect to others in the same group, for example. Since LP models are trained to reproduce observed connectivity patterns, homophilic structures naturally lead algorithms to favor within-group predictions. For this reason, homophily is the main driver of biased LP [Marey et al., 2026], and many bias mitigation strategies explicitly target homophily reduction [Buyl and De Bie, 2020, Spinelli et al., 2021, Zheng et al., 2024, Zhang et al., 2025].

Limitations of static measures. Both homophily measures and fairness metrics like demographic parity are static snapshots that do not distinguish choice homophily (intrinsic preferences) from induced homophily (amplification

through network feedback and algorithmic recommendations [Asikainen et al., 2020, Chaney et al., 2018]). But we argue that this distinction matters: an algorithm may satisfy a certain level of demographic parity initially yet amplify bias as the network evolves, or a fairness intervention may require extended observation before showing measurable effects. This motivates us to model the temporal evolution of homophily to distinguish statistically intrinsic tendencies from feedback-driven dynamics.

2.2 HAWKES PROCESSES ON GRAPHS

Hawkes processes are a class of self-exciting temporal point processes that can be used to model interaction dynamics in evolving networks, where past events increase the likelihood of future events [Yang et al., 2017, Huang et al., 2022, Perez et al., 2025]. They are particularly well-suited to settings where interactions exhibit temporal dependence and feedback effects.

We model network dynamics as a sequence of edge activation events. Each event occurs at time $t \geq 0$ and is associated with a *mark* m from a mark space $\mathcal{M}$, encoding the type of interaction (e.g., the pair of nodes involved). The event sequence is represented as a marked point process

$$N = \sum_{\text{events}} \delta_{(t,m)}. \tag{1}$$

For any interval $[s,t)$ and subset of marks $A \subseteq \mathcal{M}$, $N([s,t) \times A)$ counts the number of events of type A occurring between times s and t. We write $N_{m'}(ds)$ for the infinitesimal increment associated with events of mark m' at time s.

The dynamics of the process are governed by its conditional intensity function $\lambda_m(t)$, which specifies the instantaneous rate at which an event with mark m occurs at time t, given the history of past events. For Hawkes processes, the intensity takes the form

$$\lambda_m(t) = \mu_m + \sum_{m' \in \mathcal{M}} \int_{[0,t)} \phi_{m,m'}(t-s)\, N_{m'}(ds),$$

where μ_m is the base rate that captures the exogenous propensity for events and ϕ is an excitation kernel that models how past events influence the future rate of events.

This formulation can capture temporal clustering, endogenous feedback, and structured interaction patterns, making Hawkes processes well-suited for studying how homophily evolves under algorithmic interventions.

3 MODELING FRAMEWORK

We present our framework for modeling the temporal dynamics of homophily while explicitly distinguishing between choice homophily and induced homophily.

3.1 SET UP AND NOTATION

We consider a temporal graph with a fixed node set $\mathcal{V}$. To analyze fairness-related structural effects, we introduce a group structure on the nodes: each node $u \in \mathcal{V}$ belongs to a latent group $g_u \in \{1, \ldots, K\}$. These groups may represent demographic attributes (e.g., gender, ethnicity), communities or other structural properties, and may be observed or latent. In the context of fairness, groups typically correspond to the sensitive attribute introduced in Section 2.1.

We also assume that the graph is undirected. Edges appear over time as instantaneous events: each edge activation is characterized by a pair of nodes $\{u, v\}$ and a timestamp. We use the notation $\{u, v\}$ to denote the unordered pair of nodes, characterized by their group memberships $\{g_u, g_v\}$. For any pair of groups $i, j \in \{1, \ldots, K\}$ with $i \leq j$, we denote by $\mathcal{E}_{ij}$ the set of edges connecting groups i and j:

$$\mathcal{E}_{ij} = \{\{u, v\} : g_u = i, g_v = j\}.$$

The collection $\{\mathcal{E}_{ij}\}_{1 \leq i \leq j \leq K}$ partitions the edge set according to group interactions. Edges with $i = j$ are *within-group edges*; edges with $i \neq j$ are *cross-group edges*. Without loss of generality, this formulation enables fine-grained analysis of structural patterns and fairness disparities across groups.

3.2 GROUP-STRUCTURED HAWKES PROCESS

To account for group structure, we extend the Hawkes process framework introduced previously. We model network evolution as a group-structured marked point process, where marks correspond to group pairs and each event represents an edge activation. This formulation assumes that the interaction dynamic is homogeneous between within-group node pairs and cross-group pairs.

Group-based marks. Since the graph is undirected, each edge activation event involves an unordered pair of nodes $\{u, v\}$, characterized by the group pair (g_u, g_v) with $g_u \leq g_v$. We denote by $\mathcal{P} = \{(i,j) : 1 \leq i \leq j \leq K\}$ the set of all possible group pairs. Group pairs $(i,j) \in \mathcal{P}$ serve as marks, so the mark space is $\mathcal{M} = \mathcal{P}$. For each group pair (i,j), we define the associated counting process

$$N_{ij}(t) = N\big([0,t) \times \{(i,j)\}\big),$$

which counts the number of interaction events between nodes from groups i and j up to time t.

Group-pair Hawkes intensities. We can model the dynamics of interactions between group pairs via a multivariate Hawkes process as follows.

Definition 3.1 (Group-Pair Hawkes Intensity Model)**.** *For each $(i,j) \in \mathcal{P}$, the conditional intensity is defined as*

$$\lambda_{ij}(t) = \mu_{ij} + \sum_{(k,l) \in \mathcal{P}} \int_{[0,t)} \phi_{(i,j),(k,l)}(t-s)\, N_{kl}(ds). \tag{2}$$

Here, $\mu_{ij} \geq 0$ denotes the baseline intensity for interactions between groups i and j, and $\phi_{(k,l),(i,j)}$ is an excitation kernel describing how past interactions between groups (k,l) influence future interactions between groups (i,j).

Within this framework, within-group interactions correspond to group pairs (i,i) and cross-group interactions to pairs (i,j) with $i \neq j$. We define the aggregated within-group and cross-group counting processes as

$$N_{\mathrm{w}}(t) = \sum_{i=1}^{K} N_{ii}(t), \qquad N_{\mathrm{c}}(t) = \sum_{1 \leq i < j \leq K} N_{ij}(t),$$

with corresponding aggregated intensities

$$\lambda_{\mathrm{w}}(t) = \sum_{i=1}^{K} \lambda_{ii}(t), \qquad \lambda_{\mathrm{c}}(t) = \sum_{1 \leq i < j \leq K} \lambda_{ij}(t).$$

This formulation serves as a proxy for heterogeneous dynamics across group pairs while preserving interpretable global measures of within- and cross-group activity.

Excitation kernel specifications. We assume exponential excitation kernels:

$$\phi_{(k,l),(i,j)}(t) = \alpha_{(i,j),(k,l)}(t)\, e^{-\beta t}\, \mathbf{1}_{t \geq 0},$$

where $\beta > 0$ controls the temporal decay of influence and $\alpha_{(i,j),(k,l)}(t) \geq 0$ denotes the time-dependent excitation strength from group pair (k,l) to group pair (i,j). Exponential kernels are a natural choice due to their tractability and widespread use in modeling temporal effects.

More generally, we allow the excitation strength to depend on time-varying features characterizing the network state:

$$\alpha_{(i,j),(k,l)}(t) = f\big(X_{(i,j),(k,l)}(t)\big)\,,$$

where $X_{(i,j),(k,l)}(t)$ is a feature representation of the edge pair at time t, and f is a non-negative modulation function. This allows structural, behavioral, or algorithmic factors to be embedded directly into the excitation dynamics. In particular, $X_{(i,j),(k,l)}(t)$ can be set to the LP score between nodes k and l at time t, computed for instance as the cosine similarity between their embeddings, thereby capturing how a LP model may amplify homophily or reinforce inter-group asymmetries over time.

Remark (Relation to prior group-Hawkes models)**.** *Group-structured Hawkes processes were first introduced by Blundell et al. [2012] for directed networks, where the intensity of each directed pair depends only on its reverse yielding a highly constrained excitation structure that can be seen as a special case of our framework. Subsequent works, including Junuthula et al. [2019], Arastuie et al. [2020] and Soliman et al. [2022] further extend this line of work with respectively simpler and richer cross-pair excitation structures, the latter moving closest to our general framework. Importantly, all these works focus on learning latent community structure, whereas we treat group memberships as known sensitive attributes in a fairness setting. Our endo/exo decomposition is empirically grounded: Fujita et al. [2018] show on Twitter data that original posts and retweets correspond to exogenous and endogenous contributions respectively.*

3.3 EMPIRICAL VS INSTANTANEOUS BIAS

As presented in Section 2.1, homophily is often represented as a function of the number of within- and cross-group edges. In our temporal setting, these quantities naturally correspond to counting processes.

Definition 3.2 (Empirical bias measure)**.** *Let $N_w(t)$ and $N_c(t)$ denote respectively the cumulative number of within-group and cross-group interaction events observed up to time t. We define the empirical bias measure as*

$$B_{\mathrm{emp}}(t) = \frac{N_w(t)}{N_w(t) + N_c(t)}, \quad B_{\mathrm{emp}}(t) \in [0,1].$$

$B_{\mathrm{emp}}(t)$ represents the proportion of observed interactions that occur within groups at a given time t. Values close to 1 indicate strong within-group concentration, while values close to 0 correspond to predominantly cross-group interactions. In addition, $B_{\mathrm{emp}}(t)$ is cumulative, meaning that as time increases, all past interactions up to time t are equally counted, making the measure increasingly insensitive to recent changes in interaction patterns. As a result, empirical bias summarizes historical imbalance but fails to capture ongoing structural changes or emerging trends in network dynamics.

This limitation motivates assessing bias at the level of the interaction dynamics rather than accumulated outcomes. In a temporal point process framework, each counting process $N_{ij}(t)$ is characterized by a conditional intensity $\lambda_{ij}(t)$.

Therefore, we introduce an instantaneous bias measure based on the aggregated within-group and cross-group interaction intensities.

Definition 3.3 (Instantaneous bias)**.** *Let $\lambda_{\mathrm{w}}(t)$ and $\lambda_{\mathrm{c}}(t)$ denote the aggregated within-group and cross-group intensities defined in Section 3.1. We define*

$$B_{\mathrm{inst}}(t) = \frac{\lambda_{\mathrm{w}}(t)}{\lambda_{\mathrm{w}}(t) + \lambda_{\mathrm{c}}(t)}, \quad B_{\mathrm{inst}}(t) \in [0,1]$$

$B_{\mathrm{inst}}(t)$ serves as a proxy for the current structural tendency of the system to favor intra-group interactions over inter-group interactions. In particular, $B_{\mathrm{inst}}(t) = 0.5$ corresponds to the case where within and cross-group rates of interaction are equal, while $B_{\mathrm{inst}}(t) \approx 1$ and $B_{\mathrm{inst}}(t) \approx 0$ correspond to homophilic and heterophilic dynamics, respectively.

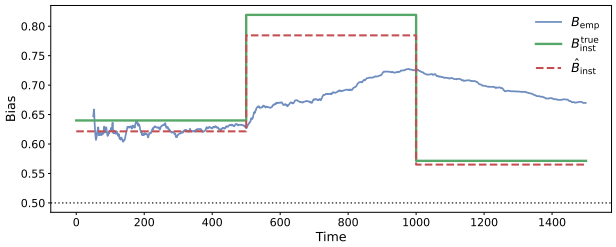


Figure 1: Empirical vs. instantaneous bias from two simulated Hawkes processes across three phases: initial moderate activity ($t < 500$), intermediate polarization ($500 \leq t < 1000$), and final alignment of group excitation ($t \geq 1000$).

Unlike $B_{\text{emp}}(t)$, the instantaneous bias reacts immediately to changes in the interaction-generating mechanism, such as variations in excitation parameters, algorithmic exposure, or fairness interventions. In that sense, it reflects what is currently being reinforced by the system, rather than what has accumulated in the past.

One of the main benefits of working with an instantaneous notion of bias is that it naturally distinguishes statistically choice homophily from induced homophily. Based on the Hawkes intensity defined in Equation 2, interaction dynamics can be decomposed into two components: a baseline intensity μ_{ij} capturing time-independent tendencies for interactions between groups i and j, and excitation terms governed by $\alpha_{(k,l),(i,j)}$, which encode endogenous reinforcement mechanisms. By defining bias at the level of instantaneous intensities, $B_{\text{inst}}(t)$ provides a direct approximation of these reinforcement dynamics, allowing us to distinguish persistent underlying trends from time-varying induced effects.

Illustrative Simulation. We illustrate the difference between empirical and instantaneous bias on a simple synthetic setting, where two groups interact according to Hawkes processes with time-varying reinforcement. The interaction dynamics go through three successive regimes: an initial phase with moderate amplification, an intermediate phase with polarized activity between groups, and a final phase where excitation levels become comparable. Details on parameter estimation are provided in Section 6, and the simulation setup is described in Appendix A.

Figure 1 compares the cumulative empirical bias B_{emp} with the instantaneous bias computed from the mean interaction intensities. As expected, the empirical bias evolves slowly and remains strongly influenced by past interactions, which smooths out regime changes. In contrast, the instantaneous bias reacts immediately to changes in the underlying interaction mechanism and clearly reflects the transitions between the three phases. Finally, the estimated instantaneous bias closely tracks the oracle bias computed from the true mean intensities, showing that our estimation procedure recovers the underlying dynamics despite stochastic fluctuations and finite observation windows.

4 THEORETICAL ANALYSIS

Now, our goal is to characterize how homophily and group-level biases evolve under different interaction and reinforcement regimes, and to identify the conditions under which these dynamics stabilize, persist, or diverge.

4.1 MEAN-FIELD APPROXIMATION.

We adopt a mean-field perspective [Bacry et al., 2015, Delattre et al., 2016] in order to obtain a deterministic description of the average interaction dynamics at the group pair level. The central idea is to approximate stochastic interaction increments by their expected behavior, which allows us to average out microscopic fluctuations while preserving the macroscopic structure of reinforcement between groups. This approximation gives rise to a tractable system that allows us to analyze the asymptotic behavior of interaction intensities and derive interpretable stability conditions.

For each group pair $(i,j) \in \mathcal{P}$, with $\mathcal{H}_{t^-}$ the history up to time t, the counting process increment satisfies

$$\mathbb{E}[dN_{ij}(t) \mid \mathcal{H}_{t^-}] = \lambda_{ij}(t)\, dt.$$

We therefore approximate

$$dN_{kl}(s) \;\approx\; \bar{\lambda}_{kl}(s)\, ds, \qquad \bar{\lambda}_{kl}(t) := \mathbb{E}[\lambda_{kl}(t)]\,.$$

Substituting this approximation into the Hawkes dynamics yields a closed system of deterministic integral equations.

Proposition 1 (Group-level mean-field dynamics)**.** *Under a mean-field approximation of the multivariate Hawkes process, the expected intensity for each group pair $(i,j) \in \mathcal{P}$ satisfies*

$$\bar{\lambda}_{ij}(t) = \mu_{ij} + \sum_{(k,l)\in\mathcal{P}} \int_0^t \phi_{(i,j),(k,l)}(t-s)\, \bar{\lambda}_{kl}(s)\, ds.$$

The derivation follows from taking conditional expectations in Equation 2 and replacing stochastic increments by their mean-field approximation.

Recall that we consider a network with K latent clusters and denote by $\lambda_{ij}(t)$ the intensity of edges between clusters i and j ($1 \leq i \leq j \leq K$). Assuming exponential excitation kernels,

$$\phi_{(i,j),(k,l)}(t) = \alpha_{(i,j),(k,l)}\, e^{-\beta t}, \quad \beta > 0,$$

we can define the number of group pairs as $G := \frac{K(K+1)}{2}$.

We introduce the following vector and matrix representations $\bar{\boldsymbol{\lambda}}(t) = [\bar{\lambda}_{11}(t), \ldots, \bar{\lambda}_{KK}(t)]^T \in \mathbb{R}^G$, $\boldsymbol{\mu} =$

$[\mu_{11}, \ldots, \mu_{KK}]^T \in \mathbb{R}^G$ and $\mathbf{A} = (\alpha_{(i,j),(k,l)})_{(i,j),(k,l)} \in \mathbb{R}_+^{G\times G}$. With these definitions, the mean-field dynamics of the system can be written as:

$$\bar{\boldsymbol{\lambda}}(t) = \boldsymbol{\mu} + \int_0^t \mathbf{A}\, e^{-\beta(t-s)}\, \bar{\boldsymbol{\lambda}}(s)\, ds. \tag{3}$$

This mean-field system captures the average evolution of interaction intensities across group pairs. It is essential to note that it preserves group-specific differences in baseline tendencies and excitation forces, allowing us to analyze how structural reinforcement mechanisms shape the temporal evolution of homophily and biases.

4.2 MEAN AND CONVERGENCE.

In this section, we analyze the temporal behavior of the mean intensities defined in Equation 3. Our goal is to characterize the stationary mean of and the convergence rate toward this equilibrium under different spectral regimes of the excitation matrix. This analysis is essential to assess the accuracy of the instantaneous bias 3.3 and to understand how feedback mechanisms influence long-term homophily.

Stationary regime. For a multivariate Hawkes process with exponential kernel and fixed excitation matrix $\mathbf{A}$, stationarity holds if $\rho(\frac{\mathbf{A}}{\beta}) < 1$ (Bacry et al. [2015]).

In that case, the stationary mean intensity is given by

$$\bar{\boldsymbol{\lambda}}^* = (\mathbf{I} - \frac{\mathbf{A}}{\beta})^{-1}\boldsymbol{\mu}.$$

Additionally, from Definition 3.3 we have:

$$B^*_{inst} = \frac{\bar{\lambda}^*_{\mathrm{w}}}{\bar{\lambda}^*_{\mathrm{w}} + \bar{\lambda}^*_{\mathrm{c}}} \tag{4}$$

To assess the reliability of B^*_{inst} as a measure of homophily, it is essential to quantify how quickly $\bar{\boldsymbol{\lambda}}(t)$ converges to its stationary value $\bar{\boldsymbol{\lambda}}^*$.

Proposition 2 (Exponential convergence rate)**.** *Assume that $\rho(\mathbf{A}/\beta) < 1$. Then there exist constants $C > 0$ and $\kappa > 0$ such that*

$$\|\bar{\boldsymbol{\lambda}}(t) - \bar{\boldsymbol{\lambda}}^*\| \le Ce^{-\kappa t}\|\bar{\boldsymbol{\lambda}}(0) - \bar{\boldsymbol{\lambda}}^*\|.$$

Moreover, the rate can be chosen such that

$$\kappa < \beta\big(1 - \rho(\mathbf{A}/\beta)\big).$$

The proof is given in Appendix B. Intuitively, dissipation dominates excitation, ensuring that the mean intensity rapidly stabilizes. When $\rho(\frac{\mathbf{A}}{\beta}) \ge 1$, the system enters critical or supercritical regimes where no stationary means exist; these cases are addressed later in the paper.

Locally stationary regime. In realistic settings, exposure policies evolve over time, inducing a time-dependent excitation matrix $\mathbf{A}(t)$. A natural modeling assumption is that $\mathbf{A}(t)$ is piecewise constant on intervals $[\tau_k, \tau_{k+1})$:

$$\mathbf{A}(t) = \mathbf{A}^{(k)}, \qquad t \in [\tau_k, \tau_{k+1}).$$

On each interval, the system behaves as a multivariate Hawkes process with fixed parameters, allowing us to reason locally [Roueff et al., 2016], which is particularly relevant for LP analysis. If convergence to equilibrium is rapid relative to the rate at which policies change, the system operates near its local equilibrium $\bar{\boldsymbol{\lambda}}^{*(k)}$, and bias measures calculated from this equilibrium are meaningful. However, if convergence is slow, persistent transient imbalances may arise, leading to systematic exposure disparities that are not captured by steady-state analyses. Thus, the convergence rate κ_k determines whether measures calculated from the local equilibrium accurately reflect the system's behavior, or whether transient dynamics dominate.

Proposition 3 (Exponential convergence rate - locally stationary case)**.** *Consider an interval $[\tau_k, \tau_{k+1})$ on which the excitation matrix is constant: $\mathbf{A}(t) = \mathbf{A}^{(k)}$.*

Assume that $\rho(\mathbf{A}^{(k)}/\beta) < 1$. Then there exist constants $C_k > 0$ and $\kappa_k > 0$ such that, for all $t \in [\tau_k, \tau_{k+1})$,

$$\|\bar{\boldsymbol{\lambda}}(t) - \bar{\boldsymbol{\lambda}}^{*(k)}\| \le C_k e^{-\kappa_k(t-\tau_k)}\|\bar{\boldsymbol{\lambda}}(\tau_k) - \bar{\boldsymbol{\lambda}}^{*(k)}\|,$$

where $\bar{\boldsymbol{\lambda}}^{*(k)} = (\mathbf{I} - \mathbf{A}^{(k)}/\beta)^{-1}\boldsymbol{\mu}$ is the local equilibrium. Moreover, the rate can be chosen such that

$$\kappa_k < \beta\big(1 - \rho(\mathbf{A}^{(k)}/\beta)\big).$$

Consequently, the instantaneous bias $B_{inst}(t)$ converges to a well-defined limit within $O(\kappa_k^{-1})$ on each interval.

Proof. The first part follows from applying Proposition 2 on each interval $[\tau_k, \tau_{k+1})$. The convergence of $B_{\text{inst}}(t)$ follows from the fact that it is a continuous function of the intensities $\bar{\lambda}(t)$. □

Implications for algorithmic interventions. Proposition 3 provides theoretical insight to interpret the temporal effects of algorithmic interventions on bias.

Interventions modify exposure dynamics over time, yielding a sequence of excitation matrices $\mathbf{A}^{(k)}$ across policy windows. The spectral radius $\rho(\mathbf{A}^{(k)}/\beta)$, determines whether the dynamics admit a stable regime and how fast transients decay. When $\rho(\mathbf{A}^{(k)}/\beta) < 1$, bias indicators can be meaningfully interpreted once the system has reached its local equilibrium. As the system approaches the critical regime, convergence becomes slow and short-term observations are dominated by transient effects; beyond this threshold, no stable regime is expected, and bias measures may fluctuate indefinitely or diverge, making reliable evaluation difficult.

Additionally, this result implies that stability does not imply bias reduction. For instance, standard LP models tend to reinforce within-group interactions, while fairness-oriented interventions promote cross-group exposure; although they act in opposite directions in terms of bias at equilibrium, both increase the overall level of excitation in the system, pushing the dynamics closer to critical regimes. As a result, interventions designed to reduce bias may require longer observation windows before their long-term effects can be reliably assessed. This also points to a limitation of purely static fairness evaluations in LP: when predictions are fed back into the interaction process, fairness properties become inherently dynamic and may not be captured by metrics computed on a fixed graph.

5 EXAMPLE

We consider a professional network with two demographic groups: group 1 (e.g., men (M)) and group 2 (e.g., women (W)). Interactions in this context correspond to repeated professional exchanges such as endorsements, messages, or recommendations. The mark space reduces to three group pairs: $(1,1)$ (M–M), $(2,2)$ (F–F), and $(1,2)$ (M–F).

Model parameters. Baseline intensities reflect a moderate choice homophily: men interact more frequently among themselves ($\mu_{11} = 0.8$), women less so ($\mu_{22} = 0.5$), and cross-gender interactions are rare in the absence of algorithmic intervention ($\mu_{12} = 0.2$). The decay parameter is set to $\beta = 1$.

We adopt the sparse neighbourhood structure introduced above: only group pairs sharing a common group index have non-zero excitation. Self-excitation is strongest for M–M interactions, reflecting the strong tendency of men to engage with other men once a connection is established. Cross-gender self-excitation is weaker, capturing the limited tendency of cross-gender interactions to self-reinforce without external intervention. Under the fairness-aware scenario, within-group self-excitation is reduced and cross-gender self-excitation is boosted.

The non-zero entries of the excitation matrix $\mathbf{A}$ are reported below under the standard and fairness-aware scenarios. The resulting matrix $\mathbf{A}$, whose entry $[\mathbf{A}]_{(i,j),(k,l)} = \alpha_{(i,j),(k,l)}$, ordered as $(1,1), (2,2), (1,2)$, takes the form:

$$\mathbf{A}_{\text{std}} = \begin{bmatrix} 0.60 & 0 & 0.10 \\ 0 & 0.40 & 0.10 \\ 0.05 & 0.05 & 0.20 \end{bmatrix}, \mathbf{A}_{\text{fair}} = \begin{bmatrix} 0.40 & 0 & 0.10 \\ 0 & 0.30 & 0.10 \\ 0.05 & 0.05 & 0.45 \end{bmatrix}$$

where rows correspond to target group pairs and columns to source group pairs. Blue entries involve M–M dynamics, red entries F–F dynamics, and green entries cross-gender dynamics. The corresponding kernel matrix is then $\phi(t) = \mathbf{A}\, e^{-\beta t}$.

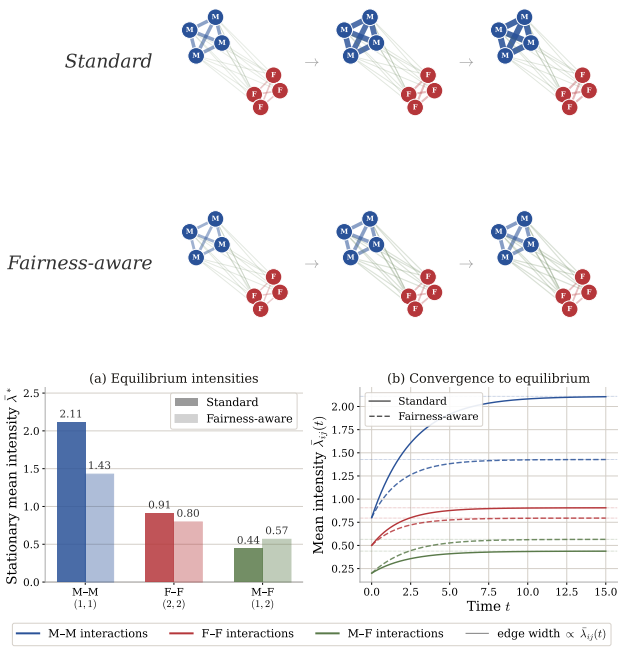


Figure 2: Network evolution and equilibrium intensities under both scenarios. *Top:* network snapshots at three time points; edge width $\propto \bar{\lambda}_{ij}(t)$. *Bottom:* (a) stationary intensities $\bar{\lambda}^*$; (b) mean-field convergence trajectories $\bar{\lambda}_{ij}(t)$.

Algorithmic scenarios. We compare a standard scenario that amplifies within-group connections with a fairness-aware scenario that boosts cross-group exposure by increasing $\alpha_{(1,2),(1,2)}$. Both satisfy $\rho(\mathbf{A}/\beta) < 1$, ensuring convergence to a well-defined equilibrium. Figure 2 illustrates the resulting dynamics. Under the standard scenario, within-group interactions progressively dominate at equilibrium ($\bar{\lambda}^*_{11} = 2.11$, $\bar{\lambda}^*_{22} = 0.91$), reflecting the amplification of baseline homophily through self-excitation. The fairness-aware intervention instead reinforces cross-group interactions, relatively reducing $\bar{\lambda}^*_{11}$ and increasing $\bar{\lambda}^*_{12}$.

6 EXPERIMENTS

The research questions motivating our experiments are:

- **RQ1:** Is the B^*_{inst} measure appropriate for capturing the dynamic evolution of homophily in a network?
- **RQ2:** How do different approaches to fairness influence the temporal dynamics of homophily?
- **RQ3:** What is the impact of the retraining frequency of prediction models on the evolution of homophily?

We conduct two types of experiments. First, we use simulated networks to validate our B^*_{inst} measure and evaluate various fair LP models, with and without model retraining. This allows us to answer RQ1–RQ3.

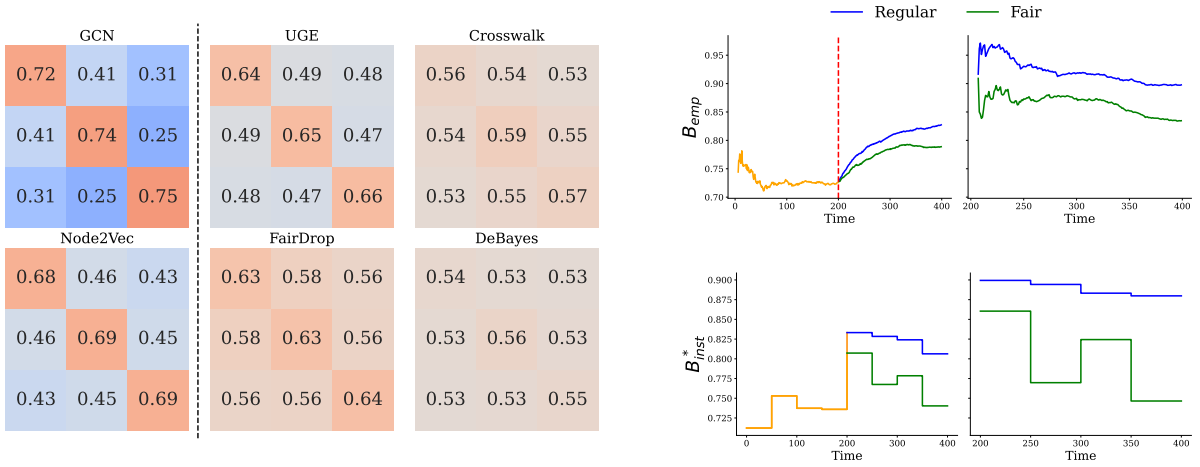


Figure 3: Results on synthetic data. Left: Mean self-excitation coefficients $\alpha_{(i,j)}$ estimated for each model, where entry (i, j) represents the estimated self-excitation of the interaction process between groups i and j. Right-top: $B_{emp}(t)$ evolution: full timeline v.s. post-intervention phase. Right-bottom: B^*_{inst} evolution: full timeline v.s. only LP-intervention phase.

Parameter estimation. First, we assume that cross-pair excitations are negligible, i.e. $\alpha_{(i,j),(k,l)} = 0$ for $(i, j) \neq (k, l)$.This diagonal assumption is motivated by identifiability constraints given the finite observation windows available per pair, and by interpretability, as diagonal coefficients directly quantify the self-reinforcing tendency of each group pair independently. Figure 6 (Appendix C.1) shows that B^*_{inst} computed from the full multivariate model remains consistent with the univariate estimates across all models, confirming that the diagonal approximation does not distort the homophily measure. Under this simplification, each $\alpha_{(i,j)}$ is estimated via maximum likelihood, maximizing the log-likelihood of the observed event sequence, and the spectral radius reduces to $\rho(A/\beta) = \max_{(i,j)} \alpha_{(i,j)}/\beta$, recovering a closed-form stationary intensity $\lambda^*_{ij} = \mu_{ij}/(1 - \alpha_{ij}/\beta)$. The β is fixed at 1 (see Appendix C.1) and μ is estimated via its closed-form MLE $\hat{\mu} = N/T$, where N is the number of observed events over $[0, T]$, following standard practice in the Hawkes process literature Daley and Vere-Jones [2003].

6.1 SIMULATION PROTOCOL

For the simulations, we draw inspiration from stochastic block models (SBMs) to model homophily of choice. We first simulate a pre-network of 1000 nodes partitioned into sensitive groups, where each node is associated with an embedding drawn from a latent cluster structure representing latent characteristics, independent of the sensitive group attribute. At each timestep, active nodes are sampled at random, and three candidate nodes are proposed based on cosine similarity in the embedding space. Link formation probabilities depend on the candidates' ranking, the SBM group structure, and node popularity. This construction ensures that the initial phase reflects homophily of choice only, without any algorithmic intervention.

After this phase, a LP model is trained on the pre-network to generate more informative embeddings. Depending on the retraining policy, the model can be updated periodically or remain fixed, allowing us to study how algorithmic recommendations, and their potential adaptation over time, interact with the initial network structure. This framework allows simulation of the evolution of a social network under the combined influence of its initial structure and recommendations. Details are provided in Appendix C.1.

6.2 B^*_{inst} FOR DYNAMIC HOMOPHILY (RQ1)

First, we highlight the interest of B^*_{inst} compared to the traditional model B_{emp}. After generating the pre-network, we run a fair LP model and a regular one (retrained several times) starting from $t = 200$. Figure 3 shows that B_{emp} fails to capture the rapid changing homophilic dynamics induced by the models and smooths out the changes over time. In contrast, B^*_{inst} is largely insensitive to whether it is measured throughout the timeline or only during the LP activity period, confirming its reliability as a measure of homophilic dynamics. For all calculations, α is estimated in the retraining windows, and μ is estimated from the pre-network phase; when limited to the second phase, we set $\mu_w = \mu_c$ to isolate homophilic dynamics. We further observe that the fair LP model is less polarizing than the unfair model, which motivates the analysis in the next section.

Regarding retraining, the significant increase in B^*_{inst} observed for GCN confirms that, without fairness constraints, periodic model updates amplify homophily in the network. Notably, UGE loses its bias mitigation capacity after retraining, whereas DeBayes remains stable (Appendix C.1).

The gap between AUC_{own} and AUC_{ref} reveals a self-reinforcing feedback: each LP model predicts links that

Table 1: Synthetic SBM ($N = 1{,}000$, 3 groups, 50 runs). $\pm$: 95 % CI. AUC_{ref}: on neutral graph; AUC_{own}: on model's own evolved graph (gap $\Rightarrow$ self-reinforcing feedback). B^*_{inst}: Hawkes homophily bias; ΔDP: demographic parity gap. *Retrain*: embeddings recomputed at $t = 750$.

	Without retrain				*With retrain*	
Model	AUC_{ref}	AUC_{own}	B^*_{inst}	ΔDP	B^*_{inst}	ΔDP
GCN	$0.93_{\pm 0.01}$	$0.93_{\pm 0.01}$	$0.81_{\pm 0.01}$	$0.82_{\pm 0.08}$	$0.83_{\pm 0.01}$	$0.91_{\pm 0.09}$
CrossWalk	$0.92_{\pm 0.00}$	$0.94_{\pm 0.00}$	$0.65_{\pm 0.00}$	$0.35_{\pm 0.04}$	$0.65_{\pm 0.00}$	$0.39_{\pm 0.05}$
DeBayes	$0.90_{\pm 0.00}$	$0.90_{\pm 0.00}$	$0.64_{\pm 0.00}$	$0.34_{\pm 0.04}$	$0.64_{\pm 0.00}$	$0.34_{\pm 0.05}$
UGE-R	$0.90_{\pm 0.00}$	$0.92_{\pm 0.00}$	$0.72_{\pm 0.01}$	$0.55_{\pm 0.06}$	$0.77_{\pm 0.01}$	$0.65_{\pm 0.06}$
Node2Vec	$0.93_{\pm 0.00}$	$0.94_{\pm 0.00}$	$0.75_{\pm 0.00}$	$0.57_{\pm 0.06}$	$0.77_{\pm 0.00}$	$0.59_{\pm 0.06}$
FairDrop	$0.91_{\pm 0.00}$	$0.93_{\pm 0.00}$	$0.67_{\pm 0.00}$	$0.56_{\pm 0.06}$	$0.68_{\pm 0.00}$	$0.60_{\pm 0.06}$

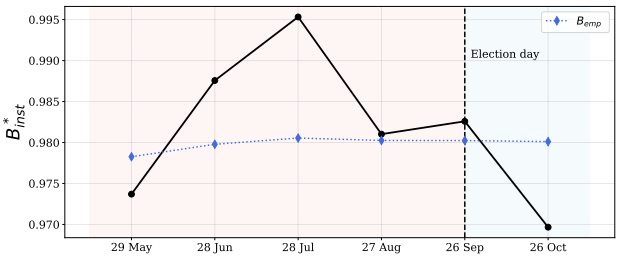


Figure 4: B^*_{inst} around the German federal election of 2021.

reshape the graph toward a structure it was trained to recognize, yielding higher link prediction performance on its own evolved graph than on the graph without any algorithmic intervention.

Finally, fairness-aware models tend to yield smaller spectral radii $\rho(\mathbf{A}/\beta)$, implying faster convergence to equilibrium (Proposition 2). This provides an additional theoretical argument in favour of fairness interventions: beyond reducing bias at equilibrium, they also accelerate the system's stabilization, making B^*_{inst} a more reliable indicator over shorter observation windows.

6.3 REAL-WORLD NETWORKS

We consider a subset of the Twitter/X 2021 dataset proposed by Pournaki et al. [2025], focusing on interactions related to political content. This period is of particular interest as it coincides with an election year, during which polarization dynamics are known to be amplified. To define the sensitive attribute, we infer users' political orientation from hashtag usage, distinguishing coarse-grained ideological groups. The labeling protocol is described in Appendix C.2.

We observe a marked peak of B^*_{inst} (Figure 4) in July 2021, well before the election, consistent with a transient amplification of within-group self-excitation. This peak coincides with the Rhineland floods (July 14–15, 2021), a major climate disaster that triggered intense partisan debate on Twitter, likely driving users to interact predominantly within their political group. The high baseline level of B^*_{inst} throughout the window indicates that the network is already strongly polarized outside the peak, and that our measure mainly captures variations in reinforcement on top of an existing polarized structure. We do not claim a causal link, but note that the election period itself is associated with a sustained elevation of B^*_{inst}, consistent with heightened within-group mobilization during the campaign.

7 CONCLUSION AND PERSPECTIVES

We analyzed how link prediction models, when coupled with network dynamics, shape the evolution of homophily and group-level disparities over time. Modeling interactions with Hawkes processes allows us to separate baseline interaction tendencies from reinforcement effects driven by past interactions and algorithmic exposure, and to define an instantaneous bias measure that captures ongoing structural tendencies beyond cumulative homophily indicators.

Our theoretical results clarify when such bias measures are meaningful in terms of stability and convergence of the induced dynamics, and experiments highlight contrasted homophily trajectories across link prediction strategies and retraining policies. Overall, this suggests that fairness assessments based on static snapshots can be misleading when predictions actively shape future interactions. Relaxing simplifying assumptions on the excitation structure would allow for finer-grained modeling of cross-group reinforcement effects. A natural perspective to achieve this would be to leverage Hawkes graphical models [Embrechts and Kirchner, 2018, Yu et al., 2020] to formalize structured sparsity patterns on $\mathbf{A}$ beyond our diagonal simplication. More broadly, our results suggest that designing and evaluating graph learning methods requires explicit attention to their dynamic, feedback-driven impact on network structure.

Acknowledgements

This work is supported by the ANR through the grant ANR-23-CE23-0026 (Project ReFAIR) and through the PEPR IA FOUNDRY (ANR-23-PEIA-0003), by the BOF of Ghent University (BOF20/IBF/117), by the European Union (ERC, VIGILIA, 101142229), the Special Research Fund (BOF) of Ghent University (BOF20/IBF/117), the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme, and the FWO (project no. G073924N). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. For the purpose of Open Access the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

Nil-Jana Akpinar, Cyrus DiCiccio, Preetam Nandy, and Kinjal Basu. Long-term dynamics of fairness intervention in connection recommender systems. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22, page 22–35, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392471. doi: 10.1145/3514094.3534173. URL https://doi.org/10.1145/3514094.3534173.

Makan Arastuie, Subhadeep Paul, and Kevin S. Xu. CHIP: A hawkes process model for continuous-time networks with scalable and consistent estimation. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems*, 2020.

Aili Asikainen, Gerardo Iñiguez, Javier Ureña-Carrión, Kimmo Kaski, and Mikko Kivelä. Cumulative effects of triadic closure and homophily in social networks. *Science Advances*, 6(19):eaax7310, 2020.

Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(1):1550005, 2015.

George Berry, Antonio Sirianni, Ingmar Weber, Jisun An, and Michael Macy. Estimating homophily in social networks using dyadic predictions. *Sociological Science*, 8: 285–307, 2021. doi: 10.15195/v8.a14.

Charles Blundell, Katherine A. Heller, and Jeffrey M. Beck. Modelling reciprocating relationships with hawkes processes. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, pages 2609–2617, 2012.

Maarten Buyl and Tijl De Bie. Debayes: a bayesian method for debiasing network embeddings. In *Proceedings of the 37th International Conference on Machine Learning, PMLR 119*, pages 1220–1229, 2020.

Meng Cao, Hussain Hussain, Sandipan Sikdar, Denis Helic, Markus Strohmaier, and Roman Kern. Recommendation fairness in social networks over time. In *2025 12th International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 116–123. IEEE, 2025.

Allison JB Chaney, Brandon M Stewart, and Barbara E Engelhardt. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of RecSys*, pages 224–232, 2018.

Sergio Currarini, Matthew O Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D. Sculley, and Yoni Halpern. Fairness is not static: Deeper understanding of long-term fairness via simulation studies. In *Proceedings of ICML*, volume 119, pages 1988–1997. PMLR, 2020.

Sylvain Delattre, Nicolas Fournier, and Marc Hoffmann. Hawkes processes on large networks. *The Annals of Applied Probability*, 26(1):216–261, 2016. ISSN 10505164. URL http://www.jstor.org/stable/43859597.

P. Embrechts and M. Kirchner. Hawkes graphs. *Theory of Probability & Its Applications*, 62(1):132–156, 2018. doi: 10.1137/S0040585X97T988538.

Francesco Fabbri, Francesco Bonchi, Carlos Castillo, and Aristides Gionis. Exposure inequality in people recommender systems: The long-term effects. In *Proceedings of SIGIR*, pages 1694–1704, 2022.

Seth Flaxman, Sharad Goel, and Justin M Rao. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320, 2016.

Kazuki Fujita, Alexey Medvedev, Shinsuke Koyama, Renaud Lambiotte, and Shigeru Shinomoto. Identifying exogenous and endogenous activity in social media. *Physical Review E*, 98(5):052304, 2018.

Mark Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, and Percy Liang. Fairness without demographics in repeated loss minimization. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1929–1938. PMLR, 2018.

Zhipeng Huang, Hadeel Soliman, Subhadeep Paul, and Kevin S Xu. A mutually exciting latent space hawkes process model for continuous-time networks. In *Uncertainty in artificial intelligence*, pages 863–873. PMLR, 2022.

Ruthwik R. Junuthula, Maysam Haghdan, Kevin S. Xu, and Vijay K. Devabhaktuni. The block point process model for continuous-time event-based dynamic networks. In Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia, editors, *The World Wide Web Conference, WWW*, pages 829–839. ACM, 2019. doi: 10.1145/3308558.3313633.

Bo Kang, Jefrey Lijffijt, and Tijl De Bie. Conditional network embeddings. In *International Conference on Learning Representations*, 2019.

Ahmad Khajehnejad, Moein Khajehnejad, Mahmoudreza Babaei, Krishna P. Gummadi, Adrian Weller, and Baharan Mirzasoleiman. Crosswalk: Fairness-enhanced node representation learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11): 11963–11970, Jun. 2022. doi: 10.1609/aaai.v36i11.21454. URL https://ojs.aaai.org/index.php/AAAI/article/view/21454.

Gueorgi Kossinets and Duncan J. Watts. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–450, 2009. ISSN 00029602, 15375390. URL http://www.jstor.org/stable/10.1086/599247.

Charlotte Laclau, Ievgen Redko, Manvi Choudhary, and Christine Largeron. All of the fairness for edge prediction with optimal transport. In *AIStats*, pages 1774–1782, 2020.

Peizhao Li, Yifei Wang, Han Zhao, Pengyu Hong, and Hongfu Liu. On dyadic fairness: Exploring and mitigating bias in graph connections. In *International conference on learning representations*, 2021.

Lydia T Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. In *International Conference on Machine Learning*, pages 3150–3158. PMLR, 2018.

Meng Liu, Ziwei Quan, and Yong Liu. Network representation learning algorithm based on neighborhood influence sequence. In Sinno Jialin Pan and Masashi Sugiyama, editors, *Proceedings of The 12th Asian Conference on Machine Learning*, volume 129 of *Proceedings of Machine Learning Research*, pages 609–624. PMLR, 18–20 Nov 2020. URL https://proceedings.mlr.press/v129/liu20a.html.

Lilian Marey, Mathilde Perez, Tiphaine Viard, and Charlotte Laclau. Structural bias beyond homophily: A study of fairness in link prediction, 2026. URL https://arxiv.org/abs/2602.11802.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001. doi: 10.1146/annurev.soc.27.1.415.

M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20), October 2002. ISSN 1079-7114. doi: 10.1103/physrevlett.89.208701. URL http://dx.doi.org/10.1103/PhysRevLett.89.208701.

Eli Pariser. *The filter bubble: What the Internet is hiding from you*. penguin UK, 2011.

Mathilde Perez, Raphaël Romero, Bo Kang, Tijl De Bie, Jefrey Lijffijt, and Charlotte Laclau. Simhawnet: a modified hawkes process for temporal network simulation. *Data Mining and Knowledge Discovery*, 39 (5):48, Jul 2025. ISSN 1573-756X. doi: 10.1007/s10618-025-01119-1. URL https://doi.org/10.1007/s10618-025-01119-1.

Armin Pournaki, Felix Gaisbauer, and Eckehard Olbrich. How influencers and multipliers drive polarization and issue alignment on twitter/x. *Proceedings of the International AAAI Conference on Web and Social Media*, 19 (1):1599–1615, Jun. 2025. doi: 10.1609/icwsm.v19i1.35890. URL https://ojs.aaai.org/index.php/ICWSM/article/view/35890.

Miriam Rateike, Isabel Valera, and Patrick Forré. Designing long-term group fair policies in dynamical systems. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 20–50, 2024.

François Roueff, Rainer {Von Sachs}, and Laure Sansonnet. Locally stationary hawkes processes. *Stochastic Processes and their Applications*, 126(6):1710–1743, June 2016. ISSN 0304-4149. doi: 10.1016/j.spa.2015.12.003.

Hadeel Soliman, Lingfei Zhao, Zhipeng Huang, Subhadeep Paul, and Kevin S. Xu. The multivariate community hawkes model for dependent relational events in continuous-time networks. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML*, volume 162 of *Proceedings of Machine Learning Research*, pages 20329–20346. PMLR, 2022.

Indro Spinelli, Simone Scardapane, Amir Hussain, and Aurelio Uncini. Fairdrop: Biased edge dropout for enhancing fairness in graph representation learning. *IEEE Transactions on Artificial Intelligence*, 2021.

Didier A. Vega-Oliveros, Liang Zhao, and Lilian Berton. Evaluating link prediction by diffusion processes in dynamic networks. *Scientific Reports*, 9(1): 10833, Jul 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-47271-9. URL https://doi.org/10.1038/s41598-019-47271-9.

Nan Wang, Lu Lin, Jundong Li, and Hongning Wang. Unbiased graph embedding with biased graph observations. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 1423–1433. ACM, April 2022. doi: 10.1145/3485447.3512189. URL http://dx.doi.org/10.1145/3485447.3512189.

Zhiqiang Wang, Baijing Hu, Kaixuan Yao, and Jiye Liang. Hawkes point process-enhanced dynamic graph neural network. In *Proceedings of the Eighteenth*

ACM International Conference on Web Search and Data Mining, WSDM '25, page 401–409, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713293. doi: 10.1145/3701551.3703520. URL https://doi.org/10.1145/3701551.3703520.

Jiasen Yang, Vinayak A Rao, and Jennifer Neville. Decoupling homophily and reciprocity with latent space network models. In *UAI*, 2017.

Tongxin Yin, Reilly Raab, Mingyan Liu, and Yang Liu. Long-term fairness with unknown dynamics. In *Proceedings of NeurIPS*, Red Hook, NY, USA, 2023. Curran Associates Inc.

Xiufan Yu, Karthikeyan Shanmugam, Debarun Bhattacharjya, Tian Gao, Dharmashankar Subramanian, and Lingzhou Xue. Hawkesian graphical event models. In Manfred Jaeger and Thomas Dyhre Nielsen, editors, *Proceedings of the 10th International Conference on Probabilistic Graphical Models*, volume 138 of *Proceedings of Machine Learning Research*, pages 569–580. PMLR, 23–25 Sep 2020. URL https://proceedings.mlr.press/v138/yu20a.html.

Guixian Zhang, Guan Yuan, Debo Cheng, Lin Liu, Jiuyong Li, and Shichao Zhang. Towards fair graph representation learning by overcoming social homophily. *ACM Trans. Intell. Syst. Technol.*, December 2025. ISSN 2157-6904. doi: 10.1145/3785503. URL https://doi.org/10.1145/3785503. Just Accepted.

Yilun Zheng, Sitao Luan, and Lihui Chen. What is missing in homophily? disentangling graph homophily for graph neural networks. In *Proceedings of NeurIPS*, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

How Predicted Links Influence Network Evolution: Disentangling Choice and Algorithmic Feedback in Dynamic Graphs (Supplementary Material)

Mathilde Perez[1] **Raphaël Romero**[2] **Jefrey Lijffijt**[2] **Charlotte Laclau**[1]

[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
[2]Ghent University–IDLab, Belgium

A DETAILS ON THE ILLUSTRATIVE SIMULATION

Simulation setup. We simulate two independent point processes corresponding to within-group interaction streams for groups w and c, using Hawkes processes with exponential kernel. Events are generated with Ogata's thinning algorithm over a time horizon $T = 1500$ and decay rate $\beta = 1$. Baseline intensities are set to $\mu_w = 0.8$ and $\mu_c = 0.6$.

To emulate regime changes in interaction dynamics, excitation parameters are chosen to be piecewise constant with two change points at $t = 500$ and $t = 1000$. Specifically, for group w, the self-excitation coefficient increases in the intermediate phase before decreasing in the final phase, while for group c it initially decreases and then increases, leading to three successive regimes with contrasted levels of endogenous reinforcement. This construction is meant to reflect qualitative shifts in interaction dynamics rather than to model a specific real-world system.

Parameter Estimation. The Table 2 shows that the estimation successfully recovers both the baseline intensities μ and the piecewise self-excitation coefficients α, with estimates closely matching the true values for each regime.

Table 2: Self-excitation coefficients α and baseline intensities μ for each regime and their estimates.

Group	True α values			Estimated values by window		
	Regime 1	Regime 2	Regime 3	Window 1	Window 2	Window 3
w	0.40	0.75	0.50	0.435	0.733	0.543
c	0.20	0.15	0.50	0.240	0.198	0.490

Parameter	Estimated value	True value
μ_w	0.717	0.800
μ_c	0.600	0.600

B TECHNICAL PROOF

Proof. We aim to show that if $\rho(\frac{\mathbf{A}}{\beta}) < 1$, the error $e(t) = \bar{\boldsymbol{\lambda}}(t) - \bar{\boldsymbol{\lambda}}^*$ decays exponentially and satisfies

$$\|e(t)\| \leq Ce^{-\kappa t}\|e(0)\|,$$

with a rate $\kappa < \beta(1 - \rho(\frac{\mathbf{A}}{\beta}))$.

The proof proceeds in two steps: we first reformulate the mean-field dynamics as a linear ODE, then analyze the error dynamics using spectral properties of the excitation matrix.

Step 1: Differential formulation. Starting from Equation 3, we differentiate with respect to time using the Leibniz integral rule. Since μ is constant and the integrand has the form $f(t,s) = Ae^{-\beta(t-s)}\bar{\lambda}(s)$, we obtain

$$\dot{\bar{\lambda}}(t) = Ae^0\bar{\lambda}(t) + \int_0^t (-\beta)Ae^{-\beta(t-s)}\bar{\lambda}(s)\,ds = A\bar{\lambda}(t) - \beta\left[\bar{\lambda}(t) - \mu\right].$$

Rearranging Equation 3, we recognize that $\int_0^t Ae^{-\beta(t-s)}\bar{\lambda}(s)\,ds = \bar{\lambda}(t) - \mu$. Substituting this into the above expression yields the linear ODE

$$\dot{\bar{\lambda}}(t) = A\bar{\lambda}(t) - \beta[\bar{\lambda}(t) - \mu] = (A - \beta I)\bar{\lambda}(t) + \beta\mu. \tag{5}$$

At equilibrium, $\dot{\bar{\lambda}}^* = 0$ implies $(A - \beta I)\dot{\bar{\lambda}}^* + \beta\mu$, which give the well-known stationary mean:

$$\bar{\lambda}^* = (I - A/\beta)^{-1}\mu. \tag{6}$$

Note that this invertibility of $(I - A/\beta)$ is guaranteed by the assumption $\rho(A/\beta) < 1$, since $||(A/\beta)^n|| \to 0$ ensures the Neumann series $\sum_{n=0}^{\infty}(A/\beta)^n$ converges.

Step 2: Exponential decay of the error. Starting from

$$\dot{\bar{\lambda}}(t) = (A - \beta I)\,\bar{\lambda}(t) + \beta\,\mu, \tag{7}$$

we can define the error $e(t) = \bar{\lambda}(t) - \bar{\lambda}^*$. Differentiating and using Equations 5 and 6, we obtain

$$\begin{aligned}
\dot{e}(t) &= \dot{\bar{\lambda}} - \dot{\bar{\lambda}}^* = \dot{\bar{\lambda}} \\
&= (A - \beta I)\bar{\lambda}(t) + \beta\mu \\
&= (A - \beta I)(e(t) + \bar{\lambda}^*) + \beta\mu \\
&= (A - \beta I)e(t) + \underbrace{(A - \beta I)\bar{\lambda}^* + \beta\mu}_{=0}.
\end{aligned}$$

Thus, the error statisfies the homogeneous linear system:

$$\dot{e}(t) = (A - \beta I)\,e(t), \qquad e(0) = \bar{\lambda}(0) - \bar{\lambda}^*. \tag{8}$$

The solution is $e(t) = \exp[(A - \beta I)t]e(0)$. To bound the matrix exponential, we analyse the spectrum of $A - \beta I$. If v is an eigenvector of A/β with eigenvalues ν, then:

$$(A - \beta I)v = Av - \beta v = \beta\nu v - \beta v = \beta(\nu - 1)v.$$

Hence the eigenvalues of $A - \beta I$ are $\{\beta(\nu_i - 1)\}$ where $\{\nu_i\}$ are eigenvalues of A/β. Since $\rho(A/\beta) < 1$, we have $|\nu_i| \le \rho(A/\beta) < 1$ for all i, which implies:

$$\Re[\beta(\nu_i - 1)] = \beta(\Re(\nu_i) - 1) < 0,$$

making $A - \beta I$ a Hurwitz matrix. By standard results on linear ODEs, there exist constants $C, \kappa > 0$ such that

$$||\exp[(A - \beta I)t]|| \le Ce^{-\kappa t}, \quad \forall t \ge 0,$$

with $\kappa < -\max_i\Re[\beta(\nu_i - 1)] = \beta(1 - \max_i\Re(\nu_i))$. When A/β is nonnegative (as is natural in our setting), the Perron-Frobenius theorem guarantees that $\rho(A/\beta) = \max_i\Re(\nu_i)$, yielding:

$$\kappa < \beta(1 - \rho(A/\beta)).$$

Consequently,

$$\|\bar{\boldsymbol{\lambda}}(t) - \bar{\boldsymbol{\lambda}}^*\| \le Ce^{-\beta(1-\rho(A/\beta))t}\|\bar{\boldsymbol{\lambda}}(0) - \bar{\boldsymbol{\lambda}}^*\|, \tag{9}$$

which can be written as $||e(t)|| \le Ce^{-\kappa t}||e(0)||$, completing the proof.

□

Figure 5 illustrates this result empirically: across three successive intervals with different excitation matrices, the normalized distance to the local equilibrium consistently stays below the theoretical exponential bound, confirming that the convergence rate κ_k is well captured by $\beta(1 - \rho(\mathbf{A}^{(k)}/\beta))$.

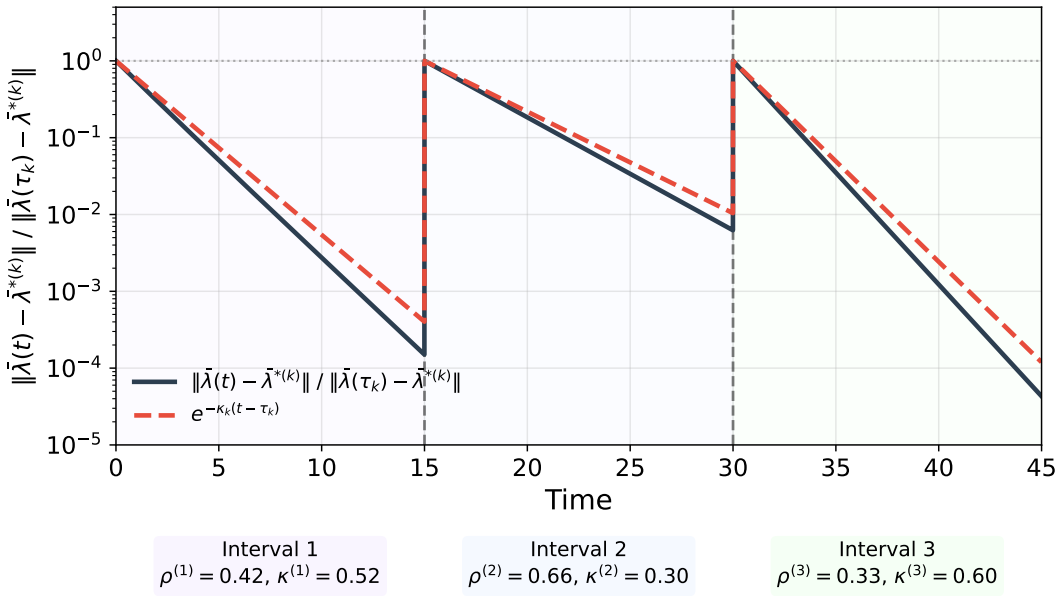


Figure 5: Empirical verification of Proposition 2. The normalized distance $\|\bar{\boldsymbol{\lambda}}(t) - \bar{\boldsymbol{\lambda}}^{*(k)}\| / \|\bar{\boldsymbol{\lambda}}(\tau_k) - \bar{\boldsymbol{\lambda}}^{*(k)}\|$ (solid line) remains below the exponential bound $e^{-\kappa_k(t-\tau_k)}$ (dashed line) on each interval $[\tau_k, \tau_{k+1})$, with $\kappa_k = 0.9\,\beta(1-\rho(\mathbf{A}^{(k)}/\beta))$. The excitation matrix $\mathbf{A}^{(k)}$ changes twice, inducing a new local equilibrium $\bar{\boldsymbol{\lambda}}^{*(k)}$ and a fresh convergence phase at each transition.

C ADDITIONAL DETAILS ON EXPERIMENTS

C.1 SIMULATION ON SYNTHETIC DATA

Simulation parameter and procedure. We present in Table 3 the hyper-parameters present in our simulation and their role. In addition, Algorithm 1 details the step to generate the temporal network used as a starting point for our experiments.

Table 3: Hyperparameters used in the simulation.

Concept	Parameter	Description
Network size	$N = 1000$	Total number of nodes
Sensitive groups	`groups`	Node group assignments
Group link probability	`prob_matrix`	Matrix of link formation probabilities between group pairs
Node activity	`activity_rate`	Probability a node is active at each time step
Popularity influence	`popularity`	Initial popularity score of each node in link formation
Retrain policy	L	Frequency of link prediction updates (periodic or single)
Recommendations	`top_probs`	Probabilities for the top k recommended candidates

Algorithm 1 Pre-network Simulation

Require: G, $prob\_matrix$, top_probs, $popularity$
$\mathbf{e}_i \leftarrow$ GETSOCIALEMBEDDINGS($nodes$) ▷ Latent cluster embeddings, independent of sensitive groups
for $t = 0$ **to** T **do**
  $active\_nodes \leftarrow$ nodes activated according to $activity\_rate$
  for all $i \in active\_nodes$ **do**
    $candidates \leftarrow$ nodes not connected to i
    $\text{sim}_{ij} \leftarrow \frac{\mathbf{e}_i \cdot \mathbf{e}_j}{\|\mathbf{e}_i\| \|\mathbf{e}_j\|}$ for all $j \in candidates$
    probs $\leftarrow \text{softmax}(\text{sim} - \min(\text{sim}))$ ▷ Weighted sampling
    $top\_idx \leftarrow$ sample 3 indices without replacement with $p =$ probs
    for $rank, idx \in top\_idx$ **do**
      $j \leftarrow candidates[idx]$
      $p_{\text{sbm}} \leftarrow prob\_matrix[g_i][g_j]$
      $acceptance \leftarrow popularity[j]$
      $p \leftarrow (p_{\text{sbm}} + top\_probs[rank]) \cdot acceptance$
      if rand() $< p$ **then**
        Add edge (i, j) to G with timestamp t
      end if
    end for
  end for
end for
return $G, edges, \mathbf{e}$

Multivariate vs Univariate. Figure 6 compares B^*_{inst} estimated under the diagonal (univariate) and full multivariate models across all LP methods and runs. The left panel shows that points cluster tightly around the $y = x$ line, confirming that both estimators are highly consistent. The right panel reports the signed differences $\Delta B^* = \text{mv} - \text{uv}$: all values are negative and small in magnitude (ranging from -0.0007 for DeBayes to -0.0216 for Node2Vec), indicating that the multivariate model yields slightly lower homophily estimates, as expected when cross-pair excitations absorb part of the within-pair signal. Crucially, the relative ordering of models is fully preserved: fairness-aware methods (CrossWalk, DeBayes, FairDrop) consistently exhibit lower B^*_{inst} than standard baselines (GCN, Node2Vec), under both estimation schemes.

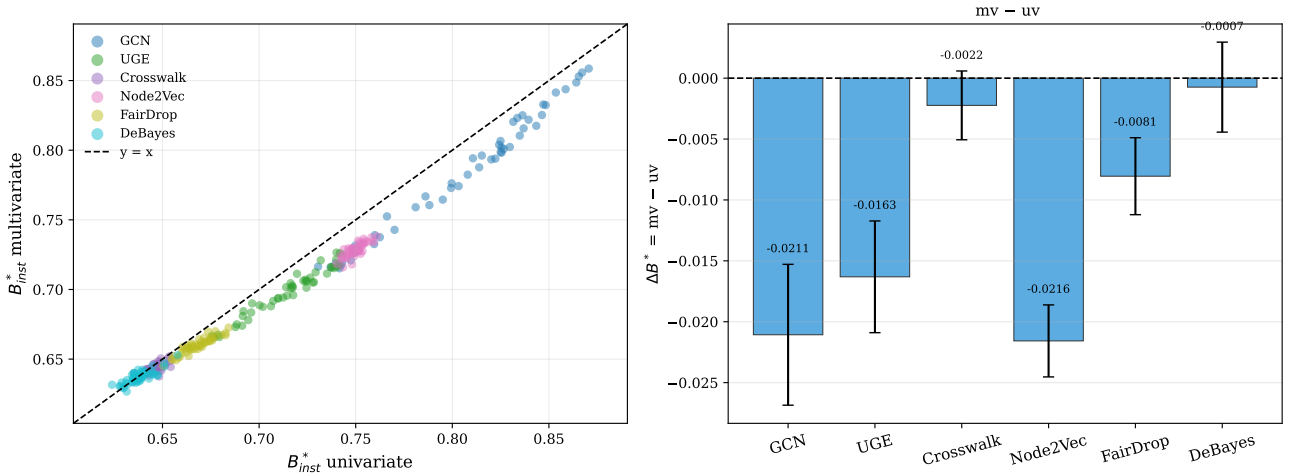


Figure 6: B^*_{inst} under diagonal vs. full multivariate Hawkes ($N = 1{,}000$, 50 runs). *Left:* multivariate vs. univariate estimates; dashed line $y = x$. *Right:* signed differences $\Delta B^* = \text{mv} - \text{uv}$ ($\pm$ 95 % CI). All differences are small ($|\Delta B^*| \leq 0.022$) and the model ranking is preserved.

β sensibility analysis. Fixing β is standard practice (Zhou et al., 2013; Bacry et al., 2015) and is motivated by three reasons. First, jointly estimating α and β yields $\hat{\beta} = 0.5$ with identical $B^*_{inst} = 0.629$ across all models, making LP

comparison impossible. Second, a common β is a prerequisite for fair model comparison. Third, $\beta = 1$ is principled given our discrete-time simulation: the kernel decays by e^{-1} per timestep, aligning estimation with the generative process. Figure 7 confirms that model rankings are fully preserved for all $\beta \geq 0.8$.

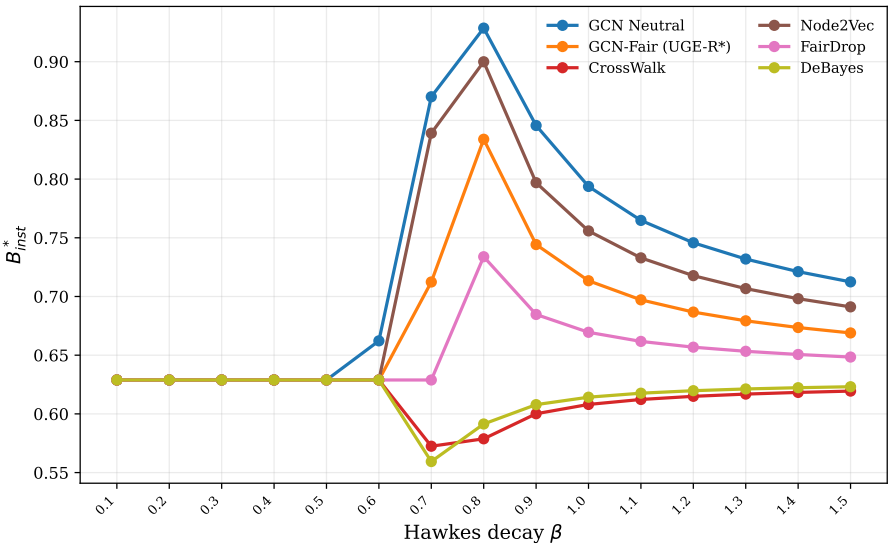


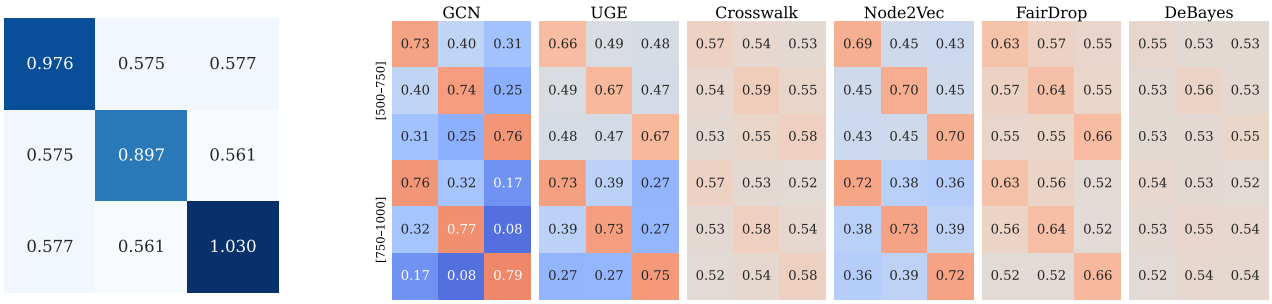
Figure 7: Sensitivity of B^*_{inst} to the Hawkes decay parameter β. Model rankings are fully preserved for $\beta \geq 0.8$.

Link prediction models. We provide technical details on the link prediction models used in the synthetic experiments.

As non-fairness-aware baselines, we consider node2Vec and a Graph Convolutional Network (GCN). The GCN serves as a backbone encoder to learn node representations from the observed graph, which are then used for link prediction. We additionally consider several fairness-aware methods that operate at different stages of the pipeline (random-walk biasing, graph pre-processing, and debiasing at the representation level).

- **CrossWalk** [Khajehnejad et al., 2022] extends random-walk-based node embedding methods (such as node2Vec) by biasing the sampling of random walks so as to encourage transitions across group boundaries. This is achieved by assigning higher weights to edges that lie close to group peripheries or that connect nodes from different sensitive groups, thereby reducing the tendency of walks to remain within the same demographic group.
- **UGE** [Wang et al., 2022] builds on a GCN backbone and enforces fairness at the representation level by encouraging the learned node embeddings to be invariant with respect to the sensitive attribute. This is achieved through an additional regularization or adversarial objective that discourages the encoder from encoding group-specific information.
- **FairDrop** [Spinelli et al., 2021] is a pre-processing strategy that modifies the adjacency matrix during training in order to compensate for homophily with respect to the sensitive attribute. The method relies on a biased edge dropout mechanism: at each training step, edges are removed according to a randomized response scheme that assigns higher removal probabilities to edges connecting nodes with the same sensitive attribute value, thereby attenuating within-group reinforcement in the observed graph.
- **DeBayes** [Buyl and De Bie, 2020] is a fairness-aware adaptation of Conditional Network Embedding (CNE) [Kang et al., 2019], a Bayesian framework that incorporates prior knowledge about the network structure (e.g., density, degree distribution, or block structure) through an explicit prior distribution. In DeBayes, sensitive group information is encoded in the prior, such that the learned embeddings are encouraged to be independent of the sensitive attribute while still capturing the underlying structural properties of the graph.

For all methods producing node embeddings, link prediction is performed using a cosine similarity-based decoder. Specifically, for each active node i, candidate neighbors are scored by their cosine similarity with i in the embedding space. These scores are then passed through a softmax with temperature $\tau = 0.1$ to obtain a probability distribution over candidates, from which the top-3 candidates are sampled without replacement. This choice ensures that differences in observed behavior across methods primarily stem from the learned representations rather than from the expressiveness of the link prediction head. For GCN-based models, the decoder is applied to the final-layer node embeddings.



(a) (b)

Figure 8: Estimated Hawkes parameters: (a) μ estimated; (b) Estimated excitation matrices α over two windows [500–750] and [750–1000].

Since our objective is not to compare predictive performance but to analyze the qualitative impact of these methods on structural bias and homophily dynamics, we use the default hyperparameters from the original implementations for all methods. For GCN-based models (GCN and UGE), we use the same number of layers, hidden dimensions, and optimizer settings, so that observed differences can be attributed to the fairness mechanisms rather than architectural choices.

Additional results. Figure 8a shows the baseline intensity matrix $\hat{\mu}$, estimated on the pre-network phase. The diagonal values confirm the presence of structural homophily before any algorithmic intervention.

Figure 8b shows the matrices α estimated over two successive windows [300–450] and [450–600], as part of retraining. Compared to the single-window estimates in Figure 3, the values are significantly higher, which is normal: shorter windows capture a more concentrated signal immediately after each model update, while a single long window dilutes the instantaneous influence of the model over time. GCN and UGE consistently show diagonal coefficients that are highly concentrated on both windows (0.58–0.69), confirming the persistent algorithmic reinforcement of interactions within the group. Crosswalk and DeBayes produce nearly uniform matrices (0.38–0.46), reflecting low and stable group-level excitation. The models remain largely consistent between the two windows for all models, suggesting that homophilic dynamics stabilize quickly after each intervention. These observations raise a broader question concerning the temporal granularity at which fairness should be evaluated. A coarse measurement window may mask short-term spikes in homophilic excitement that occur immediately after a model update, while a window that is too fine may capture transient noise rather than structural trends.

C.2 SOCIAL NETWORK (TWITTER/X) APPLICATION

Data extraction. We use a retweet network extracted from the dataset released by Pournaki et al. [2025], covering the year 2021 and the German context. Nodes correspond to user accounts and an edge represents a retweet interaction between two accounts. The original dataset includes topic modeling over tweet content; we restrict our analysis to interactions associated with the german political topic. The resulting graph contains $70{,}384$ nodes and $581{,}443$ edges.

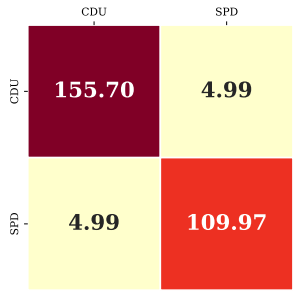
Labelisation procedure. To identify the political orientation of users, we adopted a three-step approach. First, keywords characteristic of each political camp were selected to define *anchor* accounts (*seeds*): the most active users on these terms serve as known political reference points. Second, a retweet graph is constructed and submitted to the Louvain community detection algorithm, which blindly identifies groups of heavily interconnected users. Finally, the detected communities are labeled by counting the seeds from each political side they contain: the community predominantly populated by CDU seeds is tagged *right-wing*, while the one populated by SPD seeds is tagged *left-wing*.

We emphasize that this procedure provides only a reductive proxy for political orientation. Our goal here is not to draw substantive conclusions about political behavior or electoral dynamics, but rather to rely on real-world interaction data to assess whether the proposed model can be learned on non-synthetic networks exhibiting meaningful temporal and structural patterns. The list of seed keywords was established in collaboration with the AI assistant Claude in Table 4.

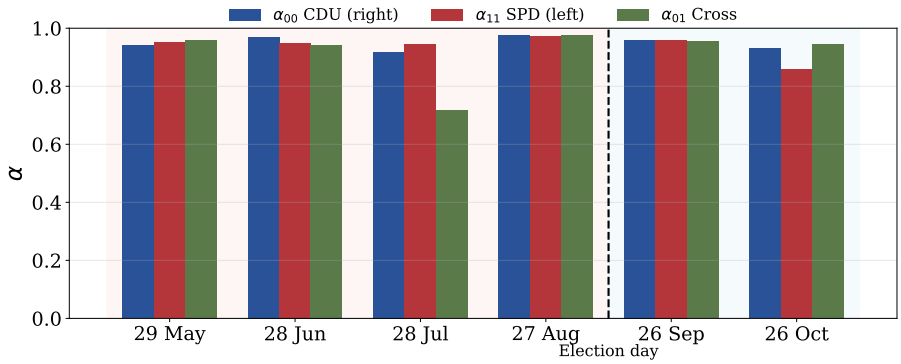
Additional Results. Figure 9b shows that the estimated coefficients α are consistently high for all windows, reflecting strong self-excitation within both political camps. Figure 9a shows the reference intensity matrix μ : the cross-group values

Table 4: Keywords used to initialise the political seeds.

Camp	Keywords
CDU (right-wing)	`laschet`, `maassen`, `cduwahlprogramm`, `union`
SPD (left-wing)	`scholz`, `brueckenlockdown`, `pimmelgate`, `spd`



(a)



(b)

Figure 9: Estimated parameters for the German network: (a) μ estimated; (b) Excitation coefficients computed on the tweet extraction from May to Oct. 2021.

are about 30 times higher than the within-group term, confirming that the network exhibits structural polarization. These two observations are consistent with the echo chamber effect characteristic of retweet-based platforms, where information circulates mainly within ideologically homogeneous communities. A drop in α_{01} is observed in the July window, coinciding with the Rhineland floods of mid-July. This transient reduction in cross-camp excitation may reflect diverging reactions to the disaster within each political camp, though caution is warranted in drawing causal conclusions from a single window. After the election (26 Sep, 26 Oct), α values remain elevated and stable across all pairs.