When Can Social Network Recommendations Improve Precision and Mitigate Inequality?

Elif Ece Erdem
e.e.erdem@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Maarten de Rijke
m.derijke@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Fernando P. Santos
f.p.santos@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Abstract

Link recommendation algorithms can reshape network topology and amplify structural inequalities through "rich-get-richer" dynamics: well-connected users may be disproportionately recommended, accumulating further connections at peripheral nodes' expense. Typical evaluations overlook how algorithmic choices interact with network topology behind this disparity. We introduce an auditing framework that isolates the effects of algorithmic design from those of network topology, on precision and structural equality. We define *Equality Gain*, a measure to quantify whether a recommender mitigates or amplifies structural disparities relative to a recommender-free baseline. To systematically vary topology, we introduce *Counterfactual Graph Rewiring* (CGR), which produces graphs with tunable clustering at a fixed degree sequence. We find that the precision–equality trade-off characterizing link-recommendation algorithms is contingent on network topology: sparse graphs exhibit a trade-off, whereas dense, highly clustered graphs admit a "win-win" regime in which models exploiting triadic closure around low-degree nodes attain both high precision and equality gains. These findings motivate a shift from model-centric optimization to network-aware link recommendation, aligning algorithmic mechanisms with network structure.

CCS Concepts

• **Information systems** → **Recommender systems**.

Keywords

Link recommendation, Network-aware optimization

ACM Reference Format:
Elif Ece Erdem, Maarten de Rijke, and Fernando P. Santos. 2026. When Can Social Network Recommendations Improve Precision and Mitigate Inequality?. In *Proceedings of the 35th ACM International Conference on Information and Knowledge Management (CIKM '26), November 07–11, 2026, Rome, Italy*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3799682.3840744

1 Introduction

Link recommenders are highly consequential tools deployed in online social networks: they do not merely suggest potential connections, but actively reshape network topology, redirecting information flows and altering who becomes visible to whom [14, 25]. These recommendations introduce a self-reinforcing feedback loop [50]: today's recommendations increase tomorrow's exposure, compounding advantages across successive rounds [49].

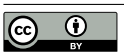
Friend-of-a-friend algorithms, the backbone of most social recommendation systems, allocate exposure proportional to individuals' degree (i.e., their number of connections). For instance, Twitter's link-recommendation algorithm increased the overall number of connections by 22% but boosted the number of connections of its most popular users by 50% [49]; similar dynamics appear on LinkedIn, Instagram, and YouTube [1, 11, 28, 55]. Organic growth, by contrast, is sublinear: the number of connections of an individual with 100,000 followers grows relatively slower than one with 1,000 [49]. Differences in degree exacerbated by algorithmic recommendations can amplify disparities by recommending already-visible nodes. Degree-based inequality carries direct social equity implications as minority groups frequently occupy less-connected network positions and degree-amplifying algorithms risk compounding their disadvantage [14]. Having fewer connections can reduce individuals' visibility, their opportunity to widely share content and to receive diverse/high-quality information.

Prior fairness-aware methods enforce parity across protected groups [24, 41]. But protected attributes are often unavailable, and group parity can obscure the structural mechanisms generating inequality. Attribute-free fairness notions exist [27, 29], yet a systematic framework for understanding precision–equality trade-offs as a function of algorithm–network interactions remains absent. The result is a fundamental ambiguity. Without understanding how structural properties (e.g., clustering coefficient, density, or degree heterogeneity) shape algorithmic fairness, we cannot determine whether observed disparities are algorithmic artifacts nor whether a model's fairness guarantees will hold across network topologies.

To address this gap, we systematically audit link recommenders varying (i) the baseline algorithm used, (ii) its hyperparameters, and (iii) the network topology itself. We define *Equality Gain* as the difference in the degree-distribution Gini coefficient [10, 44] after link-recommendations are introduced. Equality Gain measures whether a recommender amplifies heterogeneity beyond the baseline, with a neutral recommender yielding zero.

We contribute:

- *A systematic topological audit framework:* a protocol (Section 3.1) that separates algorithmic design, hyperparameter configuration, and network topology as drivers of precision–equality trade-offs, with Equality Gain as a key measure, isolating algorithmic amplification from intrinsic network heterogeneity.
- *Counterfactual Graph Rewiring (CGR):* a rewiring method (Section 3.3) that preserves the network's exact degree sequence, guarantees connectivity, and tunes the clustering coefficient



This work is licensed under a Creative Commons Attribution 4.0 International License.
CIKM '26, Rome, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2539-5/2026/11
https://doi.org/10.1145/3799682.3840744

both above and below the empirical value. These properties are not jointly satisfied by prior methods, which either approximate degree distributions [15], target assortativity [32], or perform greedy edge swaps [4, 6]. This enables controlled, degree-preserving sweeps of the clustering coefficient and density.
- *Topological drivers of the precision–equality trade-off:* we challenge the assumption of a universal trade-off, identifying a win-win regime (characterized by recommendations leading to both high precision and high alongside Equality Gains). This regime occurs in denser topologies where utility and equity align. Equality outcomes are driven primarily by topology rather than by hyperparameter choices: clustering acts as a boost for precision; density is the dominant factor behind structural equality.

Together, these results reveal that structural properties strongly constrain an algorithm's impact on degree inequality, and that gains in equality are primarily driven by mechanism–topology fit rather than by algorithms' design alone. [1]

Section 3 below presents our methodology and rewiring method; Section 4 details the experimental setup; Section 5 reports the results; Section 6 concludes.

2 Related Work

2.1 Link recommendation paradigms

Link recommendation has evolved from structural heuristics to representation learning. Local heuristics, e.g., Common Neighbors [26], exploit neighborhood overlap, while popularity-based heuristics, e.g., Preferential Attachment [26], capture degree signals. Global heuristics like SALSA [17, 23] extend this via random walks on bipartite graphs, reaching higher-order structural proximity. Despite their speed and interpretability, heuristics share a fundamental limitation: reliance on fixed structural signals restricts generalization across networks, motivating the shift to learned representations.

Embedding-based methods advanced the field by learning low-dimensional node representations to facilitate link prediction. DeepWalk [42] and node2vec [16] generate sequences of node co-occurrences and apply the skip-gram model [37]. While effectively capturing network structure, these transductive methods lack inductive capabilities, limiting their generalization to new nodes or graphs.

Graph neural networks (GNNs), on the other hand, capture complex non-linear dependencies. LightGCN [19] simplifies graph convolution via pure neighborhood aggregation, while PPRGo [7] uses approximate personalized PageRank for scalable diffusion. Recent models target specific representation learning shortcomings: SSNet [47] addresses scale distortion to prevent hub dominance, and Tail-STEAK [31] uses knowledge distillation to improve recommendations for low-degree "tail" nodes. Despite these advances, it is unclear if GNN performance gains sacrifice structural equity.

2.2 Structural bias and fairness

2.2.1 Algorithmic amplification. While early graph-theoretic studies modeled recommenders to analyze static reachability and path lengths [38], modern evaluations recognize that algorithms dynamically reshape topologies to reinforce "rich-get-richer" dynamics, or the Matthew Effect [36]. Empirical audits on social platforms confirm that algorithms amplify already popular nodes, often modulated by network homophily. [50] prove GCNs exhibit an inherent "within-group PA bias," producing scores proportional to the geometric mean of node degrees. Similarly, [3] argue standard link prediction benchmarks have implicit degree biases favoring popularity-centric models. [48] further distinguish between algorithms, finding that while PageRank mirrors existing degree distributions, HITS (and SALSA) can drastically amplify bias in homophilic networks.

Longitudinal simulations extend these one-shot analyses to iterated feedback loops, tracking how Gini, minority visibility, and homophily compound across rounds under recommender–user interaction [9, 13, 14]. These studies assume sensitive group labels and repeated acceptance. We ask a complementary, attribute-free question: whether input topology alone predicts when one recommendation round improves precision while reducing degree inequality. Closest to our protocol, Sanz-Cruzado and Castells evaluate recommenders by extended-network structural properties under uniform acceptance, targeting weak-tie diversity; we share their protocol but target degree equality and isolate topological drivers [44].

2.2.2 Fairness interventions. Existing mitigation strategies largely enforce demographic parity [24, 34, 41] or restructure networks to bridge communities [5] or optimize specific metrics like PageRank [53]. However, these require sensitive data often not publicly available. Attribute-free approaches are rising, using degree [29] or homophily-based spectral patterns [27] to proxy latent constraints. Concurrently to our work, TopoFair [33] proposes a structural bias taxonomy for fair link prediction, showing topology strongly predicts fairness, with assortativity as the dominant driver under demographic parity. No attribute-free framework alters structural properties at a fixed degree sequence to isolate topology's role.

2.3 Controlled structural auditing

Isolating topology's structural drivers requires a network generator that can tune key network properties (e.g., clustering and density) while preserving realistic degree structure (e.g., degree distribution). Standard generators do not meet this requirement. Watts-Strogatz [54] varies clustering but produces homogeneous degrees, failing to replicate scale-free topologies. Power-law models [20] generate scale-free degree distributions but couple triangle formation with network growth, preventing independent clustering control. Stochastic block models and LFR (Lancichinetti–Fortunato–Radicchi) benchmarks [22] target community structure but cannot reproduce scale-free heterogeneity without complex modifications. TopoFair [33] extends Barabási-Albert with homophily and group-imbalance controls, but as a growth-based generator it neither preserves an existing network's empirical degree sequence nor decouples clustering from density.

Existing rewiring approaches face complementary limitations. Methods targeting assortativity [32] or greedily maximizing clustering [4] do not preserve the exact degree sequence, and the dK-series [15] targets smoothed approximations of empirical distributions rather than the empirical sequence itself. Methods that do enforce exact degree preservation [6] rely on blind edge swaps,

[1] Although our primary lens is structural, we do not treat it as a substitute for demographic fairness. A case study in Section 5.1 applies MORAL [34], a demographic-parity method, to a network dataset with explicit protected attributes, testing whether attribute-based interventions can override topological constraints and bridging structural and demographic notions of equity.

which scale poorly ($O(N^2)$) and struggle to reach high clustering targets. No existing method efficiently preserves the exact degree sequence and guarantees connectivity at high clustering targets.

3 Methodology

To examine the interplay between link recommendation algorithms, their hyperparameters, and network structure, we design a framework using both real-world and controlled (semi-)synthetic graphs.

Problem setting. We evaluate link recommenders for two outcomes: *utility* (how well a model's recommendations recover held-out test edges) and *structural equality* (the extent to which the post-recommendation graph distributes connectivity opportunities evenly across nodes). "Structural equality" refers exclusively to this attribute-free, degree-based notion (demographic and individual fairness are out of scope and briefly revisited in Section 6). We decompose variation in outcomes along three axes, and characterize how topology shapes them via common-neighbor decomposition.

3.1 Three-axis auditing protocol

Standard evaluation methods rely on measures (such as Top-k accuracy) of recommendations combining three sources of variation: *algorithmic choice*, the *hyperparameter configuration*, and the *network topology*. When varying topology and tuning models simultaneously, performance gains cannot easily be attributed to a single source. This is critical for fairness metrics, where structural changes can outweigh algorithmic effects. We propose a *three-axis audit protocol* to separate these factors.

Algorithmic axis. We hold the network and tuning protocol fixed and vary only the recommender, spanning random, heuristic, and GNN-based methods (Section 4.2). Differences in precision and Equality Gain across these methods — following the protocol detailed in Section 4.3 — are then attributable to algorithmic choice rather than to evaluation effort or topology.

Hyperparameter axis. We sweep each model's principal hyperparameters and evaluate whether changing them has a statistically significant impact in precision and equality. This addresses a gap in single-point evaluations: can these metrics be tuned, or are they determined elsewhere?

Topological axis. We evaluate models on graphs that vary in *clustering* (i.e., how closely a node's neighbors are connected to each other, indicating the degree of local interconnectedness in a network) and *density* (i.e., how many edges a graph has compared with the maximum number of edges between any pair of nodes). Since real graphs entangle these traits, performance shifts cannot be attributed to any single property. To isolate these effects, we supplement real graphs with two controlled network classes: fully synthetic graphs that vary one structural property at a time, and semi-synthetic graphs generated by our degree-preserving rewiring procedure (Section 3.3), which varies clustering at fixed degree sequences. This topological focus enables a mechanistic analysis of structural drivers in Section 3.2 and prediction testing in Section 5.3.

3.2 Network mechanism underlying precision-equality trade-offs

We argue that the divergence between precision and structural equality arises from a fundamental property of common-neighbor (CN) counts, which are the primary link-prediction signal exploited by both heuristics and GNNs. We formalize the mechanism below.

The CN decomposition. $\mathrm{CN}(i,j) = |N(i) \cap N(j)|$ counts the number of common-neighbors (or length-2 paths, also known as wedges) between i and j. Drawing on the configuration-model framework [40] and its clustered extensions [39], we approximate the expected CN count between a non-adjacent pair (i,j) via a heuristic decomposition (detailed in Appendix A.1):

$$\mathbb{E}[\mathrm{CN}(i,j)] = \underbrace{\frac{d_i\, d_j\, \kappa}{n\bar{d}}}_{\text{degree-driven}} + \underbrace{C \cdot \tau(i,j)}_{\text{clustering-driven}} + o(1), \qquad (1)$$

where $d_i = |N(i)|$ denotes the degree of node i, $\langle d \rangle = n^{-1} \sum_i d_i$ the mean degree (i.e. $\bar{d}$), and $\langle d^2 \rangle = n^{-1} \sum_i d_i^2$ the second moment. C is the global clustering coefficient, and $\tau(i,j)$ is the excess wedge density around the pair above the configuration-model baseline. The quantity $\kappa = (\langle d^2 \rangle - \langle d \rangle)/\langle d \rangle$ measures degree heterogeneity across the network. In a heavy-tailed network, $\langle d^2 \rangle$ is large and κ inflates the *overall scale* of the degree-driven term across all pairs; combined with the multiplicative $d_i\, d_j$ factor, this concentrates the strongest CN signals on hub-incident pairs; $o(1)$ absorbs finite-size scaling corrections, vanishing as $n \to \infty$, such as the probability of multi-edges and self-loops inherent to finite configuration models.

Eq. 1 conveys that $\mathbb{E}[\mathrm{CN}(i,j)]$ depends on the network's average degree and clustering, being maximized when: (a) i and j are both high-degree nodes in a heterogeneous network, so hubs accumulate many common neighbors regardless of local structure; or (b) i and j sit inside a tight local group, where the excess shared-neighborhood density $\tau(i,j)$ is high even if both nodes have low degree.

Since link-prediction (and resulting link-recommendation) relies on CN signals to identify likely edges [26], precision benefits from either term. In sparse, heavy-tailed networks, the degree-driven term dominates: κ inflates hub-incident scores faster than clustering can compensate. Precision is then achievable by recommending edges between hubs, but such recommendations flow disproportionately to well-connected nodes, amplifying inequality. *Equality therefore only improves when the clustering-driven term is strong enough to generate precise recommendations for non-hub pairs.* Which term dominates is governed by the network's degree heterogeneity, clustering, and density — not by the specific algorithm.

Eq. 1's decomposition affects differently each class of recommender, explaining the differences observed empirically:

Preferential Attachment (PA) scores links by $d_i \cdot d_j$ (Eq. 1's degree-driven term). PA's structural insensitivity to clustering concentrates recommendations on hubs regardless of local cohesion.

Common Neighbors (CN) directly counts $|N(i) \cap N(j)|$, weighting both terms based on the network structure: In sparse, low-clustering graphs the degree term dominates and CN approximates PA; in high-clustering graphs the clustering term becomes substantial.

SALSA assigns scores via a random walk on the hub-authority bipartite graph. In degree-heterogeneous networks, the stationary distribution of these walks concentrates on high-degree nodes [48], making SALSA primarily degree-driven and thus susceptible to hub-amplification in sparse topologies.

GNN behavior depends on propagation depth. At the shallow limit, Ma et al. [30] show that linear graph autoencoders with

near-orthogonal node embeddings recover a CN-like score that preserves both terms of Eq. 1. At the deep limit, repeated symmetric-normalized propagation drives node embeddings toward dominant-eigenvector alignment (oversmoothing), and [50] show that link scores collapse to $\sqrt{d_i d_j}$ – the pure degree-driven term. Shallow propagation on high-clustering graphs therefore preserves CN structure; deep propagation on sparse graphs degenerates toward the degree-only regime. These limits apply most directly to symmetric-normalised propagation (LightGCN); diffusion-based variants such as PPRGo follow a different stationary distribution and are examined empirically in Section 5.

Together, Eq. 1's decomposition and algorithm-specific behavior imply that different networks yield different precision-equality trade-offs, as Section 5.1 confirms.

3.3 Counterfactual graph rewiring

Building on the limitations identified in Section 2.3, we propose Counterfactual Graph Rewiring (CGR): given $G = (V, E)$ and target clustering coefficient $C^\star$, CGR returns G' with the G's *exact* degree sequence and clustering coefficient $\approx C^\star$ (Figure 1).

Step 1 – Topology-aware edge scoring. Since no node features are available, we use the identity matrix $\mathbf{I} \in \mathbb{R}^{N\times N}$ as input, giving each node a unique one-hot representation before propagation. A Simple Graph Convolution (SGC) encoder produces embeddings $\mathbf{Z} = \mathbf{F}^K \mathbf{I}\,\mathbf{W}$, with rows z_u, where the propagation operator:

$$\mathbf{F} \;=\; \tfrac{1}{2}\big(\hat{\mathbf{A}} + \eta\mathbf{B} + (\hat{\mathbf{A}} + \eta\mathbf{B})^\top\big) \tag{2}$$

blends two adjacency operators. Here, $\hat{\mathbf{A}}$ is the symmetrically-normalised adjacency matrix, and $\mathbf{B}$ is a row-normalised two-hop adjacency matrix: we form $\mathbf{A}^2$, zero its diagonal, and cap entries entry-wise at θ to dampen the large 2-hop multiplicities accumulated by hubs. Since $\mathbf{B}$ is asymmetric, averaging $\mathbf{F}$ with its transpose restores symmetric similarities; the scalar η controls the two-hop contribution. The projection $\mathbf{W} \in \mathbb{R}^{N\times d}$ is drawn i.i.d. from $\mathcal{N}(0, 1/N)$: only the relative ordering of similarity scores matters, and using a random rather than learned $\mathbf{W}$ avoids letting $\mathbf{W}$ absorb the very structure we are about to rewire. Candidate pairs (u, v) are scored by a degree-weighted centered cosine similarity:

$$s(u,v) = \frac{(\mathbf{z}_u - \boldsymbol{\mu})^\top (\mathbf{z}_v - \boldsymbol{\mu})}{\|\mathbf{z}_u - \boldsymbol{\mu}\|\,\|\mathbf{z}_v - \boldsymbol{\mu}\|}, \quad \boldsymbol{\mu} = \sum_i w_i \mathbf{z}_i, \quad w_i = \frac{d_i^\gamma}{\sum_j d_j^\gamma}, \tag{3}$$

where subtracting the degree-weighted mean before normalisation prevents hub embeddings from dominating the score distribution and γ controls the strength of this correction.

Since K-hop SGC propagation makes embeddings of locally-cohesive nodes similar, pairs in cohesive groups score highly; using these scores in Step 2 seeds annealing with a triangle-biased topology, reducing swaps needed to reach $C^\star$.

Step 2 – Degree-constrained projection. Using the scores $s(u, v)$ as edge weights, we build an initial graph G_{init} that realises the exact degree sequence of G in two stages. First, we extract the maximum spanning tree E_T over the edges *already present in* G, weighted by $s(u, v)$, and freeze it; this guarantees connectivity. Second, residual degrees $r(u) = d_G(u) - d_{E_T}(u)$ are filled from a candidate pool of all non-frozen node pairs. Crucially, this pool includes pairs *absent from* E, allowing the generator to introduce new edges while respecting the degree sequence. We solve this degree-constrained selection greedily via a weighted Havel–Hakimi assignment [18] that, at each step, picks the node with the largest residual degree assigning its top-scoring feasible partners; if it fails, we fall back to a min-cost-flow b-matching [2] with negative scores as edge costs.

Step 3 – Simulated annealing toward $C^\star$. We refine G_{init} via degree-preserving double-edge swaps on non-frozen edges. For two edges $(a, b), (c, d)$ with four distinct endpoints, we propose either $\{(a, c), (b, d)\}$ or $\{(a, d), (b, c)\}$. Proposals are discarded if either new edge already exists in the current graph or belongs to E_T. For the swap $\{(a, b), (c, d)\} \to \{(a, c), (b, d)\}$, the change in triangle count is: $\Delta_\triangle = T(a, c) + T(b, d) - T(a, b) - T(c, d)$, with $T(\cdot)$ counts common neighbours. With clustering coefficient target $\triangle^\star = C \cdot N_{\text{wedges}}/3$ (where $N_{\text{wedges}} = \sum_v \binom{d_v}{2}$ is swap-invariant) and residual $S = |\triangle - \triangle^\star|$, swaps are accepted whenever S decreases and otherwise with probability $\exp(-\Delta S/T_{\text{temp}})$, on a geometric schedule $T_{\text{temp}}(t) = T_0(T_f/T_0)^{t/(S_{\max}-1)}$ where $S_{\max}$ is the total swap-proposal budget and $t \in \{0, 1, \ldots, S_{\max} - 1\}$ indexes the proposal.

Complexity. The pipeline is dominated by Step 3 (simulated annealing) at $O(S_{\max}\langle d\rangle)$; Steps 1–2 run in $O(Kmd_{\text{feat}})$ and $O(n^2 + mK\log K)$ respectively, where K is the SGC propagation depth.

Limitations. The achievable clustering ceiling is bounded by the input degree sequence; empirically, the Reed98-based generator saturates near $C \approx 0.65$; the sparser Wiki-based near $C \approx 0.4$ (Appendix, Figure 6). Since Section 5.3 plots results against the *achieved* clustering, imperfect target tracking does not bias conclusions. Two fidelity properties further qualify this design. CGR is generative rather than perturbative: meeting the degree and clustering constraints requires re-selecting a substantial fraction of edges, so collateral properties (e.g., modularity, assortativity, average path length) can shift alongside clustering, particularly in dense sweeps. The sweep therefore spans a coherent local-cohesion regime rather than clustering in isolation, mirroring how these properties co-vary in real networks [8]. We attribute dense-network effects to this composite regime rather than to clustering alone. Separately, we omit a per-step ablation: each step enforces a distinct constraint: connectivity (Step 1–2), exact degree preservation (Step 2), and target clustering coefficient (Step 3). Removing any step breaks a guarantee of the pipeline rather than isolating its contribution. Hyperparameter information is reported in Appendix A.3.

4 Experimental Setup

4.1 Datasets

Real-world datasets. We conduct experiments across four real-world networks: Twitter-9K [46], Pokec [51], (Wiki)pedia [43], and (Facebook) Reed98 [43]; see Appendix A.2 for details.

Synthetic datasets. To isolate the effects of topological properties, we first generate fully synthetic graphs using the Watts–Strogatz (WS) [54] and power-law (PL) [20] models. This allows us to systematically vary one structural property (e.g., clustering coefficient) while holding others (e.g., degree sequence and density) constant. For WS models, we generate graphs with $N = 1000$ nodes, varying either the rewiring probability $p \in [0.05, 1.0]$ (with fixed mean degree $k = 4$) or the mean degree $k \in \{2, 4, 6, 8, 10, 12\}$ (with fixed $p = 0.1$). For PL models, we varied either the triangle-formation

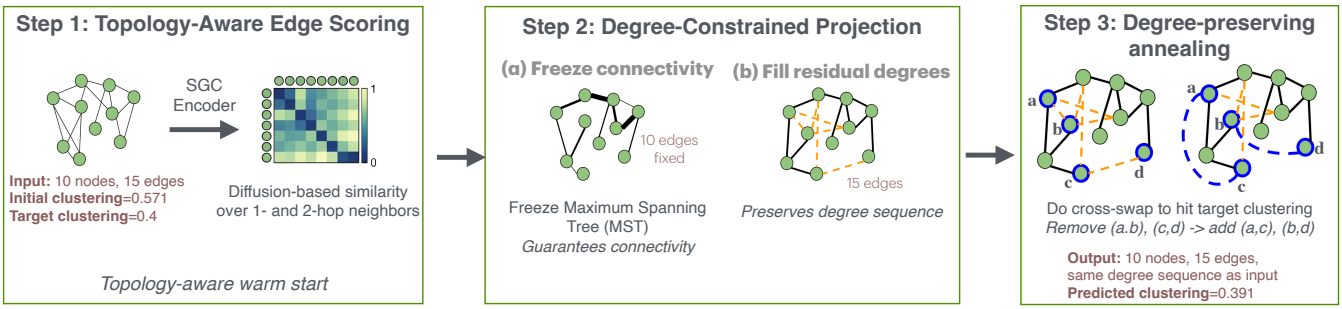


Figure 1: Counterfactual Graph Rewiring (CGR). Given input graph G with degree sequence d and target clustering coefficient $C^\star$, CGR (1) scores candidate pairs via diffusion-based similarity; (2) constructs a degree-feasible initial graph by freezing a maximum spanning tree (guaranteeing connectivity) and assigning residual degrees via weighted Havel–Hakimi; and (3) drives global clustering coefficient toward $C^\star$ via degree-preserving double-edge swaps under simulated annealing. The output preserves G's exact degree sequence and remains connected throughout. CGR can drive $C^\star$ above or below C_0.

probability $p \in [0, 1]$ (with fixed $m = 4$) or the number of initial connections $m \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ (with fixed $p = 0.1$). Increasing p generates networks with a higher clustering, while keeping the density constant. We report the average results over 10 independent networks for each parameter setting.

Semi-synthetic datasets. To study structural effects on realistic topologies, we created semi-synthetic networks based on the degree-distribution of real-world sparse (Wiki) and dense (Reed98) graphs. We used our generator to hit target clustering values. We generate 10 instances for each configuration (Table 3).

4.2 Baseline models

Heuristics. We include the three heuristic families of Section 2.1: Common Neighbors, Preferential Attachment, and SALSA plus a random recommender as a sanity check.

GNN backbones. We evaluate the four GNN paradigms of Section 2.1: LightGCN, PPRGo, SSNet, and Tail-STEAK.

Demographic fairness method. MORAL [34], a demographic-parity method (Section 2.2.2), serves as our attribute-based reference for the case study in Section 5.1

4.3 Evaluation

4.3.1 Recommendation generation and ranking. We treat all networks as undirected, enabling symmetric GNN message passing and an undirected degree-based notion of equality; no node features are used, so all signal comes from topology. Each model is trained on the training graph $G_{\text{train}} = (V, E_{\text{train}})$ and produces, for every node $u \in V$, a ranked list of all candidate non-neighbors ordered by predicted link score $f(u, \cdot)$. We retain the top $K = 10$ per node (reflecting typical social platform suggestion counts [21, 35, 52]) and aggregate into the recommended edge set E_{pred}, keeping a single edge when both u and v recommend each other. This yields the post-recommendation graph $G_{\text{pred}} = (V, E_{\text{train}} \cup E_{\text{pred}})$, which we evaluate against the control graph $G_{\text{control}} = (V, E_{\text{train}} \cup E_{\text{test}})$. During training, we follow the negative-sampling scheme of [47]: for each held-out positive edge $(u, v^+) \in E_{\text{test}}$ we sample 99 negatives uniformly from u's 2-hop neighborhood in G_{train}, matching realistic candidate pools and providing harder negatives than uniform sampling. Implementation details are in Appendix A.3.

Assumption (uniform acceptance). Equality Gain treats every top-K recommendation as an accepted edge (the extended-network protocol of [44]). They acknowledge this is "not realistic as a literal assumption" but use it to assess how recommendation may reshape the network. We adopt this established protocol and, going beyond prior work, explicitly characterize the two biases it induces. *(i) Acceptance bias.* In deployed platforms, acceptance rates correlate with the target node's degree [49]; since hubs receive disproportionately real acceptances, actual post-deployment concentration is likely *worse* than G_{pred} reports. *(ii) Edge-budget bias.* Models add $\approx K|V|$ edges, typically exceeding $|E_{\text{test}}|$, and the resulting density increase compresses the degree distribution, lowering $\text{Gini}_{\text{pred}}$ relative to $\text{Gini}_{\text{control}}$. Both effects bias Equality Gain upward, so negative values are a lower-bound estimate of amplification and positive values reflect genuinely distributive behaviour beyond mechanical compression. Since both biases enter as approximately constant factors across methods on the same graph, rankings hold.

4.3.2 Data split. We employ random 80/10/10 splits for Pokec, Wiki, and Reed98. For Twitter, we use a temporal split based on interaction timestamps to capture directional evolution. To account for stochasticity, we averaged over multiple seeds (10 for real-world, 2 for synthetic variants). We verified that Figure 2's precision-equality trade-off patterns remain stable across random splits, confirming they reflect structural properties rather than train-test artifacts.

4.3.3 Metrics. Utility is measured by Precision@10 (P@10). For each user u, we score $P@10(u)$ as the fraction of u's top-10 recommended items that are held-out positives. We report the mean over users. We choose P@10 because it directly reflects practical utility by focusing on the quality of the most visible results. We prefer it over Recall for its top-k emphasis and over NDCG because our link recommendation task has binary relevance (a link is correct or not), for which P@10 is a clear and interpretable metric.

Equality Gain. To quantify link-recommenders' effect on degree inequality, we propose Equality Gain, which measures this effect relative to a control baseline. We quantify inequality using the Gini coefficient of the degree distribution [10], the standard measure of distributional inequality, which has likewise been adopted to measure structural dispersion in link recommendation [44]. Three properties make it the appropriate instrument here: (i) it is scale-invariant, so values remain comparable across networks of different sizes and densities – essential since recommendation mechanically grows the graph (Section 4.3.1); (ii) it increases precisely when

connectivity is transferred from low-degree to high-degree nodes, matching the amplification phenomenon Equality Gain is designed to detect; and (iii) it is attribute-free, consistent with our structural scope (Section 3): Gini $= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|d_i - d_j|}{2n\sum_{i=1}^{n} d_i}$, where d_i node i's degree and n is the node count. $Gini = 0$ means perfect equality, while $Gini = 1$ maximum inequality.

Equality Gain is the percentage improvement of the model's predicted graph ($\text{Gini}_{\text{pred}}$) over the control baseline ($\text{Gini}_{\text{control}}$):

$$\text{Equality Gain} = \left(\frac{\text{Gini}_{\text{control}} - \text{Gini}_{\text{pred}}}{\text{Gini}_{\text{control}}}\right) \times 100, \tag{4}$$

where $\text{Gini}_{\text{control}}$ uses the training graph plus observed test edges, and $\text{Gini}_{\text{pred}}$ uses the training graph plus the model's recommendations, representing the system's predicted future. Equality Gain is therefore a structural diagnostic of how a recommender redistributes connectivity relative to a recommender-free counterfactual (not an absolute or user-level fairness measure); Section 4.3.1 characterizes its two inherent biases.

5 Results

5.1 Algorithmic axis

Figure 2 shows the utility vs. Equality Gain landscape for four real-world datasets, revealing two utility-equality trade-off regimes.

Regime 1: The traditional trade-off. On sparse networks (Twitter, Pokec), a clear Pareto front emerges: higher-precision methods yield negligible or no equality improvement. The random recommender reduces degree disparities most effectively, (∼25% equality gain on Twitter, ∼18% on Pokec), while high-precision models fail to match it. LightGCN achieves near-zero Equality Gain; Tail-STEAK occupies a middle ground (∼+3% on Twitter) but still falls short of the random baseline, suggesting that, on these sparse networks, even tail-focused designs do not overcome the structural barrier.

On Pokec, heuristic-based methods worsen inequality; PA and CN yield Equality Gains of (∼−10% and −5%), amplifying "rich-get-richer" dynamics. Inversely, GNN-based methods (LightGCN, PPRGo) achieve near-neutral equality (∼0% Equality Gain) with higher precision, yet remain well below the random baseline's (∼+18%). Tail-STEAK mirrors its Twitter behavior on Pokec, achieving only ∼+7% Equality Gain at lower precision than LightGCN/PPRGo.

Twitter appears anomalous, with nearly all models showing positive Equality Gains. This is a sampling artifact, not algorithmic equity: the dataset is constructed via snowball sampling of interaction edges (retweets, mentions) [45], whose recursive inclusion probability quadratically inflates the sampled degree distribution. This pushes $\text{Gini}_{\text{control}}$ to 0.6798 — 13.9% above Pokec's 0.5967, though Twitter is ∼235× denser. Because Equality Gain is a relative metric, this inflated control baseline makes hub-concentrating PA appear fairer; applied to the more complete Pokec graph, PA yields strongly negative gains. Twitter's positive Equality Gains do not represent genuine algorithmic equity improvements.

MORAL case study (group fairness vs. structural position). To test whether demographic-parity objectives overcome topological constraints, we evaluate MORAL [34] on Pokec-n using *region* as the sensitive attribute. Region 0 is the demographic majority (71.1% of users) but region 1 holds the structural advantage, with ∼12.7% higher average degree (16.89 vs. 14.98) and a mild over-representation in the top 1% of high-degree nodes (1.04× vs. 0.98× population share). This inversion decouples demographic majority status from structural advantage. MORAL delivers the larger *relative* equality gain to the structurally disadvantaged majority (+78.6% vs. +67.0%), yet the structurally advantaged minority retains a higher per-user utility (Precision@10: 0.218 vs. 0.185, a ∼17% relative gap; Figure 3), while aggregate equality slightly regresses. Demographic parity thus redistributes equity without equalizing per-user utility, which tracks connectivity rather than labels: link recommendation fairness must target structural position.

Regime 2: Alignment of utility and equity. The denser Wiki (7.3×10^{-3}) and Reed98 (4.0×10^{-2}) datasets exhibit a qualitatively different dynamic. We call an outcome win-win when a model matches the best precision on that network while matching or exceeding the random recommender's Equality Gain. In these networks, the "rich-get-richer" effect weakens: hubs already share many connections with candidates, so strong CN signals are no longer exclusive to hub pairs. GNNs exploit multi-hop *structural bridges* to simultaneously boost precision and distribute recommendations more equitably.

On Wiki, the effect is most pronounced: LightGCN and PPRGo achieve Equality Gains of ∼+12 to +15%, matching or exceeding the Random Recommender while delivering accurate recommendations. The denser Reed98 network reveals an architectural divergence. PPRGo and PPRGo+SSNet reach a Pareto-optimal position: high precision alongside ∼+10% Equality Gain, suggesting PageRank-based diffusion effectively escapes dense clusters to find tail nodes. Conversely, LightGCN falls into the negative equality regime (∼−8%) with lower precision: its standard message passing appears susceptible to over-smoothing or capture by the core, reinforcing recommendations toward already-central nodes. Notably, SSNet does not rescue LightGCN here, indicating the message-passing mechanism itself—rather than embedding magnitude—is the primary bottleneck for equality in dense topologies. Tail-STEAK is an instructive third case: despite explicit tail-targeting via knowledge distillation, it lands in the negative equality regime alongside LightGCN, with lower precision than both GNN families. Consistent with Eq. 1, the win-win regime arises from clustering-driven CN signals benefiting non-hub pairs, a structural signal that Tail-STEAK's node-local reweighting cannot exploit and that PPRGo captures via global diffusion across the core. Equality outcomes in this regime thus track mechanism–topology fit more than algorithmic intent.

These findings show that algorithms vary substantially in both utility and degree-equality. Section 5.2 tests whether these differences stem from tuning, and Section 5.3 from topology.

5.2 Hyperparameter axis

We test whether precision-equality trade-offs reflect fundamental constraints or suboptimal tuning, systematically varying key hyperparameters: embedding dimension (GNNs), degree thresholds (Tail-STEAK), k_circle_of_trust (SALSA), and α/topk (PPRGo). We assessed significance via Friedman tests followed by pairwise Wilcoxon signed-rank tests (Table 1).

Sparse networks show trade-offs when both metrics respond to tuning. On Twitter and Pokec, no sweep yields simultaneous gains. On Pokec, every sweep falls into one of two patterns:

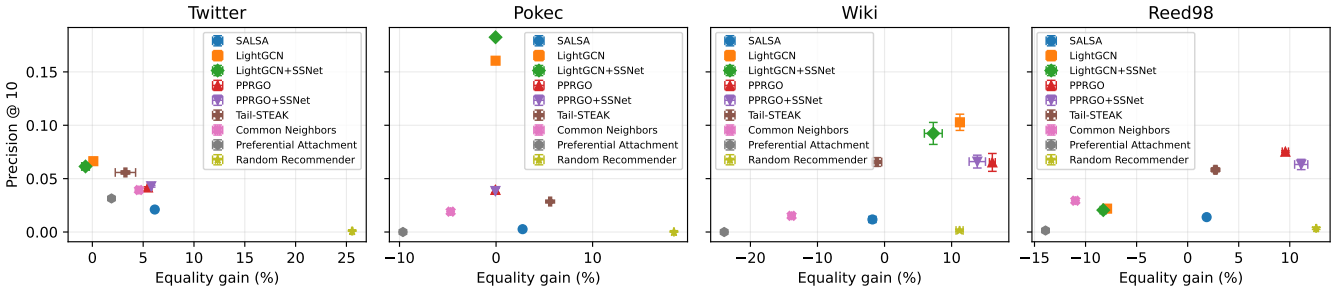


Figure 2: Precision–equality results. Each point shows mean ± std (over 10 seeds). Sparse networks (Twitter, Pokec) exhibit a trade-off where high precision limits equality gain relative to the Random baseline. Denser networks (Wiki, Reed98) reveal a "win-win" regime where models like PPRGo achieve high precision while matching the structural equity of random recommendations.

Table 1: Statistically significant hyperparameter effects. We employ a two-stage protocol: a Friedman test ($p < 0.05$) followed by pairwise Wilcoxon signed-rank tests to isolate specific configurations deviating from the baseline. Significance levels: $^{**}p < 0.01$, $^{*}p < 0.05$, ns = not significant. ρ: Spearman correlation. † Non-monotonic. PPRGo: no significant effects. Sparse networks (Twitter, Pokec) exhibit trade-offs, while dense networks (Wiki, Reed98) allow parameter alignment.

Dataset	Hyperparameter	p@10 Effect	Equality Effect	Trade-off
Twitter	LightGCN: emb_dim	✓** 512 ($\rho = +0.8$)	✓** 128 ($\rho = +0.4^{\dagger}$)	✓ Exists
	Tail-STEAK: deg_t_low	✓** 5 ($\rho = -0.7$)	✓** 20 ($\rho = +0.6$)	✓ Exists
	Tail-STEAK: emb_dim	ns	✓* 32 ($\rho = -0.4$)	-
	SALSA: k_circle_of_trust	✓* 500 ($\rho = +0.85$)	✓* 5 ($\rho = -0.95$)	✓ Exists
Pokec	LightGCN: emb_dim	✓*256 ($\rho = +0.7$)	ns	-
	Tail-STEAK: deg_t_low	✓**7 ($\rho = -0.6$)	ns	-
	Tail-STEAK: emb_dim	ns	✓* 32 ($\rho = -0.4$)	-
	SALSA: k_circle_of_trust	✓*500 ($\rho = +0.6$)	✓* 5 ($\rho = -0.3$)	✓ Exists
Reed98	LightGCN: emb_dim	ns	✓* 64 ($\rho = -0.6$)	-
	Tail-STEAK: deg_t_low	✓** 5 ($\rho = -1.0$)	✓** 20 ($\rho = 0.0^{\dagger}$)	✓ Exists
	Tail-STEAK: emb_dim	ns	ns	-
	SALSA: k_circle_of_trust	✓* 5 ($\rho = -0.9$)	✓* 5 ($\rho = -1.0$)	× **Aligned**
Wiki	LightGCN: emb_dim	✓** 512 ($\rho = +1.0$)	✓** 512 ($\rho = +1.0$)	× **Aligned**
	Tail-STEAK: deg_t_low	✓** 5 ($\rho = -1.0$)	✓** 5 ($\rho = -1.0$)	× **Aligned**
	Tail-STEAK: emb_dim	ns	ns	-
	SALSA: k_circle_of_trust	✓* 5 ($\rho = -0.9$)	✓* 500 ($\rho = +0.95$)	✓ Exists

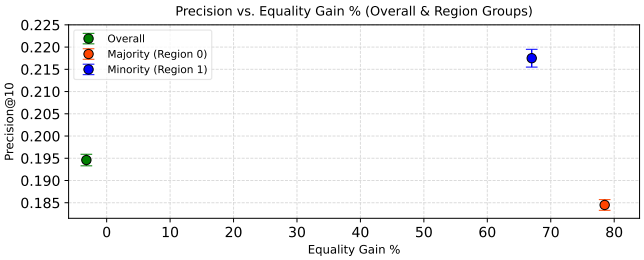


Figure 3: Precision@10 vs. equality gain under MORAL on Pokec-n (region attribute). Points are means over 10 runs; error bars denote one SD. The structurally disadvantaged majority (Region 0) gains more nomination equity (+78.6%) than the structurally advantaged minority (Region 1, +67.0%), yet retains lower per-user utility (Precision@10: 0.185 vs. 0.2180, a ∼17% gap). Aggregate equality slightly regresses (≈−3%, Overall): group-label parity redistributes nomination equity without closing topology-driven utility gaps.

either explicit trade-off (SALSA: precision peaks at k = 500, fairness at k = 5) or decoupling (LightGCN emb_dim moves precision, $\rho = +0.7$, but not fairness; Tail-STEAK emb_dim moves fairness but not precision). Twitter follows the same logic — Tail-STEAK requires deg_t_low = 5 for precision but deg_t_low = 20 for equality. The structural conclusion is that in sparse topologies, capacity either fits the biased structure (precision without redistribution) or trades one metric for the other; alignment is unavailable.

Dense networks enable alignment where tuning benefits both metrics. On Wiki, specific settings (e.g., deg_t_low = 5 for Tail-STEAK) maximize both precision and equality ($\rho = -1.0$ both, $p < 0.001$), supporting our Section 5.1 finding: in dense settings, accurate links are often bridging links.

Equality is largely insensitive to tuning. Of the 16 sweeps in Table 1, four show no significant fairness effect; PPRGo additionally shows no significant effect on either metric across all four datasets, consistent with its stationary distribution being topology-dominated rather than diffusion-parameter-driven in heterogeneous graphs. Even where tuning does move fairness, the effect

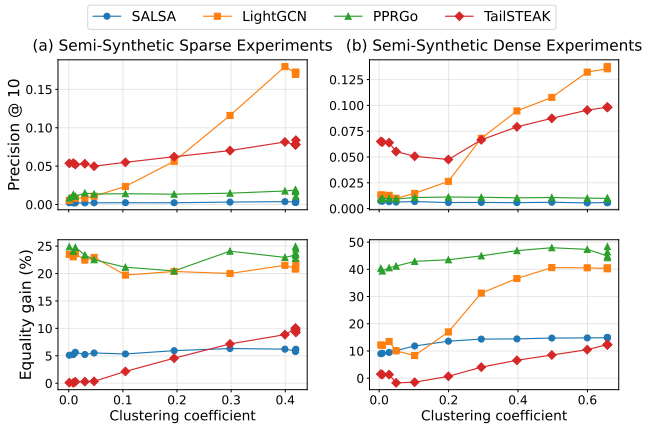


Figure 4: Impact of clustering on semi-synthetic networks. (a) Sparse (Wiki-based): clustering boosts precision for LightGCN and Tail-STEAK; equality remains roughly stable. (b) Dense (Reed98-based): High clustering improves precision and fairness for LightGCN, and fairness for PPRGo, showing the "win-win" alignment inherent to Reed98's topology.

can run counter to precision: Reed98 LightGCN degrades fairness with larger embedding dimensions without precision benefit: added capacity amplifies hub concentration without utility gain.

5.3 Topological axis

Section 5.2 held topology fixed and varied tuning; here, we hold tuning fixed and vary topology. We focus on clustering and density on both methodological and theoretical grounds. Methodologically, these are the two structural properties that can be varied independently while exactly preserving the empirical degree sequence: density is fixed by the degree sum, and clustering can be tuned via degree-preserving rewiring (Section 3.3). Other candidates (assortativity, modularity, average path length) co-vary with clustering and density, so independently manipulating them without confounding shifts in the degree distribution is infeasible [8]. Theoretically, clustering underlies triadic closure, the dominant link-prediction signal exploited by both heuristics and GNNs [25, 26], while density governs the global distribution of exposure opportunities across nodes [12]. Section 3.2 formalizes this connection via Eq. 1, predicting a trade-off regime in sparse, low-clustering graphs and a win-win regime in dense, high-clustering ones. We test these predictions empirically using synthetic graphs (Watts–Strogatz, power-law) and our semi-synthetic graphs (Wiki- and Reed98-derived) to vary clustering and density while controlling confounders. We present results for LightGCN, PPRGo, SALSA, and Tail-STEAK; we omit SSNet, as its outcomes mirrored its base models, and non-learning baselines, which lack the mechanisms central to our investigation.

Local structure: Clustering as a precision multiplier. Higher clustering amplifies the CN-driven signal that both heuristics and GNNs exploit (Section 3.2): triangles densify the 2-hop paths around each candidate pair, mutually reinforcing valid-link signals. Figure 5a and Figure 4 (top row) confirm a monotonic precision gain for LightGCN across all evaluated topologies. The other models' responses are strictly modulated by degree heterogeneity: PPRGo shows monotonic precision gains across fully synthetic topologies (Figure 5a) but remains flat on semi-synthetic graphs (Figure 4). Conversely, Tail-STEAK gains precision alongside LightGCN on semi-synthetic graphs (Figure 4), but remains completely flat on homogeneous WS networks (Figure 5a, left). In dense Reed98-based networks (Figure 4b), this signal also benefits equality: LightGCN's Equality Gain surges (10% → 40%) as clustering coefficient grows alongside high precision gains, since high-clustering CN signals reach non-hub pairs. As Section 3.3 notes, these high-clustering sweeps also shift collateral properties, so we attribute the effect to the clustering-anchored regime rather than to clustering alone.

Two model-specific contrasts clarify this mechanism's activation. Tail-STEAK's response is highly suppressed in homogeneous WS — its tail-targeting has minimal signal when degrees are uniform — but responds to heavy-tailed PL and semi-synthetic graphs. PPRGo is the mirror: it tracks clustering on synthetic graphs (Figure 5a) but its precision flattens completely on semi-synthetic graphs (Figure 4). While PPRGo succeeds on empirical networks by leveraging global community macro-structures (Section 5.1), our degree-preserving random annealing destroys these long-range bridges. Since PageRank diffusion relies on this global macro-structure rather than local triadic closure, forcing local triangles (clustering) fails to rescue its predictive utility. Equality outcomes thus reflect mechanism–topology fit, not algorithmic intent.

Global structure: Density as mechanism-dependent baseline. We examine density through three complementary lenses, each removing a different confound. (i) Matched-pair semi-synthetic comparison (Figure 4, realistic heavy-tailed degrees). Pairing low-clustering points across Figure 4a and 4b, sp_1/d_1 ($C = 0.001/0.007$), sp_3/d_2 ($C = 0.012/0.016$), sp_6/d_3 ($C = 0.042/0.047$), sp_7/d_4 ($C = 0.070/0.068$) isolates density at near-fixed clustering. Here, the effect of density is highly model-dependent. For random-walk based models, the dense series sits higher on Equality Gain at every matched C (PPRGo jumps from ~20% sparse to ~40% dense; SALSA lifts from ~5% sparse to ~10% dense). This indicates that, for diffusion mechanisms, density expands reach and raises the equality baseline even at matched clustering. Conversely, for LightGCN, the sparse series actually provides better equality at these low matched points (~20% sparse vs. ~10% dense). LightGCN's Equality Gain only surges to ~40% later, as clustering grows *beyond* the matched pairs. So density universally expands the pool of reachable non-hub pairs, but GNNs and heuristics need clustering to convert that reach into equality gains. (ii) Continuous WS sweep (Figure 5b left, homogeneous degrees). Removing degree heterogeneity to test density in isolation, increasing density lifts Equality Gain across models by creating alternative paths that bypass hubs confirming the matched-pair finding under continuous control. (iii) Continuous PL sweep (Figure 5b right, scale-free degrees, clustering not co-scaled). This isolates the failure mode: density without proportional clustering adds non-triadic edges, lowering the GNN signal-to-noise ratio and pushing LightGCN/PPRGo back toward degree centrality (consistent with Section 5.1's findings for LightGCN on the empirical Reed98 graph). Density alone is therefore not universally beneficial: while it inherently raises the equality floor for diffusion mechanisms by expanding random-walk reach, it lowers the GNN signal-to-noise ratio. For GNNs, density only raises the equality ceiling if clustering is proportionally high enough to fill the new edge slots with non-hub-driven CN signals.

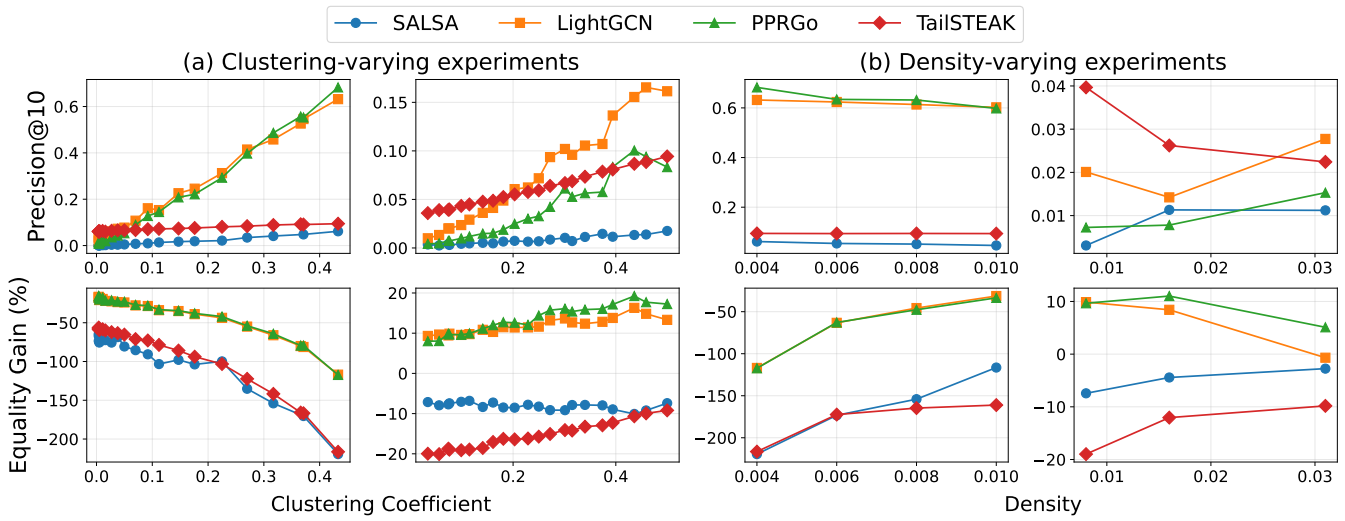


Figure 5: Impact of structure on fully synthetic networks. (a) Clustering: in homogeneous WS (left), it boosts precision but degrades equality (trade-off); in scale-free PL (right), it lifts both for GNNs (win-win). (b) Density: in WS (left), it universally improves equality at some precision cost; in PL (right), the effect splits by mechanism: density improves precision but harms equality for GNNs (LightGCN, PPRGo), while it improves equality but reduces precision for Tail-STEAK, and improves both precision and equality for SALSA, confirming density alone is insufficient without proportional clustering.

Heterogeneity constraints. Finally, degree heterogeneity imposes a clear performance ceiling: all models show significantly lower precision and equality stability on power-law graphs than on homogeneous Watts–Strogatz ones, indicating that extreme degree inequality limits equitable recommendation.

6 Conclusion

We examined utility–equality trade-offs in link recommenders across three axes: algorithm, hyperparameters, and topology. We proposed Equality Gain as a fairness measure and Counterfactual Graph Rewiring (CGR) to vary topology at a fixed degree sequence.

Main findings. (i) Algorithmic axis: the precision–equality relationship is topology-dependent: While sparse networks exhibit a tension where accuracy costs equity, denser networks reveal a "win-win" regime. (ii) Hyperparameter axis: precision is tunable, but equality is largely insensitive to tuning; in sparse graphs, no hyperparameter sweep aligned both metrics; the trade-off is not an artifact of suboptimal tuning. (iii) Topological drivers: clustering improves precision, density sets the equality baseline for diffusion mechanisms, and degree heterogeneity imposes a clear performance ceiling, equality is driven primarily by topology. (iv) Equality interventions: the MORAL case study shows that dense high-degree graphs allow demographic-parity gains to flow to the disadvantaged group, but not to close per-user utility gaps; structural position, not labels, bounds such interventions.

Implications. The actionable rule is mechanism–topology alignment. Sparse networks demand explicit structural regularization or blended random exploration. Dense, highly clustered networks admit triadic-closure GNNs (e.g., LightGCN) for genuine win-win outcomes and diffusion models (e.g., PPRGo) for robust equitable baselines. Equitable link recommendation depends on aligning algorithmic mechanisms with network structure, not model-centric optimization alone.

Limitations and future work. First, our setup is a controlled structural abstraction: we assume uniform acceptance and analyze one-shot predictions on static, undirected graphs without node features. This masks asymmetric exposure dynamics and longitudinal feedback loops, which can compound inequality [9, 13, 14]. Second, our negative sampling protocol, though standard for GNN benchmarks [47], may favor local-closure methods over global heuristics. Third, CGR's clustering intervention entangles with modularity and assortativity in dense sweeps (Section 3.3), so dense-network attributions target the clustering-anchored regime; a full fidelity characterization is deferred. Fourth, Equality Gain is a relative directional metric: complementary indices (Theil, top-share) and exposure- or user-level fairness notions could test our rankings' robustness. Finally, degree-based equity is necessary but not sufficient for social fairness; demographic-parity interventions inherit a ceiling set by the connectivity distribution. Future work can validate our results with explicit user acceptance models, layered group/individual-fairness objectives, and more complex recommendation systems eventually based on generative models.

Acknowledgments

This work is supported by the ERC (RE-LINK, 101116987, https://doi.org/10.3030/101116987; UNITE, 101201510), and by NWO (024.004.022, NWA.1389.20.183, and KICH3.LTP.20.006). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of their respective employers, funders and/or granting authorities.

A Technical Appendix

A.1 Derivation of Eq. 1

In standard configuration models, network topology is strictly locally tree-like, and global clustering vanishes ($C \to 0$) in the thermodynamic limit $n \to \infty$ [40]. Consequently, expected common neighbors (CN) between non-adjacent nodes i and j rely entirely on

Table 2: Structural properties of the empirical networks.

Property	Twitter	Pokec	Wiki	Reed98
# Nodes	9,528	1.6M	889	962
# Edges	204.2K	22.3M	2.9K	18.8K
Density	$4.0e^{-3}$	$1.7e^{-5}$	$7.3e^{-3}$	$4.07e^{-2}$
Avg. degree	42.85	27.32	6.56	39.11
Clustering	0.248	0.109	0.153	0.318
Avg. Path	3.048	5.2	4.096	2.461
$Gini_{control}$	0.6798	0.5967	0.5319	0.4611

the random matching of degree "stubs," yielding the degree-driven expectation $\frac{d_i d_j \kappa}{n\bar{d}}$, where $\kappa = (\langle d^2 \rangle - \langle d \rangle)/\langle d \rangle$.

To theoretically ground the $E[CN(i,j)]$ decomposition for clustered networks (where $C > 0$), we use [39]'s generalized configuration model. This model proves that non-vanishing transitivity requires separating the generative process by assigning nodes a joint degree sequence (s_i, t_i) of single edges (s_i) and triangles (t_i).

While an exact analytical expectation requires restricting chance overlaps strictly to single edges ($s_i s_j$), Eq. 1 uses [39]'s theoretical separation as a first-order heuristic approximation for practical link prediction: (i) The Degree-Driven Term: We estimate the baseline of random chance wedges using the full observed degree sequence $\left(\frac{d_i d_j \kappa}{n\bar{d}}\right)$ [40]. (ii) The Clustering-Driven Term: While the baseline configuration model assumes uniformly random connections globally, real networks exhibit high local clustering. To capture non-vanishing clustering, it is necessary to parameterize the formation of triangles [39] via $\tau(i,j)$, which represents the localized excess wedge density around nodes i and j. Modulated by the global clustering coefficient C, the multiplicative term $\tau(i,j)$ corrects for the non-vanishing expected common neighbors arising strictly from local cohesion. Combining these components yields:

$$E[CN(i,j)] = (d_i d_j \kappa)/n\bar{d} + C \cdot \tau(i,j) + o(1). \quad (5)$$

Using the total degree d rather than s implicitly double-counts edges involved in explicit triangles. This heuristic decomposition serves as a practical upper bound dual structural origins of common neighbors without requiring the intractable inference of unobservable (s_i, t_i) sequences for real-world graphs. The $o(1)$ term absorbs finite-size scaling corrections that vanish as $n \to \infty$ [39].

A.2 Datasets

Structural properties of the four empirical networks and those of the semi-synthetic variants are summarized in Table 2 and Table 3.

Real-world networks. We use four datasets: Twitter-9K [46], Pokec [51] (directed friendship graph), Wiki [43] (directed Wikipedia voting network), and Reed98 [43] (undirected friendship network).

Semi-synthetic networks. Using our rewiring procedure, we generate variants of Wiki (sparse) and Reed98 (dense) while preserving the source graph's exact degree sequence. Each target clustering value is realized via 10 independent seeds. Figure 6 shows target-vs-achieved clustering tracking.

A.3 Implementation details

We use official implementations for LightGCN, PPRGo, SSNet, and Tail-STEAK; SALSA is implemented from scratch. MORAL uses official default parameters. Dataset-specific hyperparameters: SALSA

Table 3: Per-instance clustering and average path length for generated graphs; means (std) over 10 independent seeds. Static metrics ($|V|$, $|E|$, density, Gini) are invariant by construction and reported in Table 2.

Sparse (Wiki)			*Dense (Reed98)*		
ID	Clustering	Avg. path	ID	Clustering	Avg. path
sp_1	0.001 (0.001)	3.55 (0.01)	d_1	0.007 (0.001)	2.26 (0.00)
sp_2	0.010 (0.002)	3.56 (0.01)	d_2	0.016 (0.001)	2.26 (0.00)
sp_3	0.012 (0.003)	3.56 (0.01)	d_3	0.047 (0.001)	2.28 (0.00)
sp_4	0.024 (0.001)	3.59 (0.01)	d_4	0.068 (0.002)	2.23 (0.00)
sp_5	0.030 (0.004)	3.63 (0.02)	d_5	0.114 (0.002)	2.29 (0.00)
sp_6	0.042 (0.003)	3.84 (0.02)	d_6	0.122 (0.002)	2.41 (0.00)
sp_7	0.070 (0.004)	4.08 (0.03)	d_7	0.187 (0.004)	2.47 (0.00)
sp_8	0.110 (0.005)	4.44 (0.04)	d_8	0.285 (0.005)	2.55 (0.00)
sp_9	0.151 (0.005)	5.10 (0.09)	d_9	0.384 (0.006)	2.66 (0.01)
sp_10	0.162 (0.004)	5.27 (0.09)	d_10	0.508 (0.005)	2.92 (0.02)
			d_11	0.571 (0.004)	3.79 (0.16)

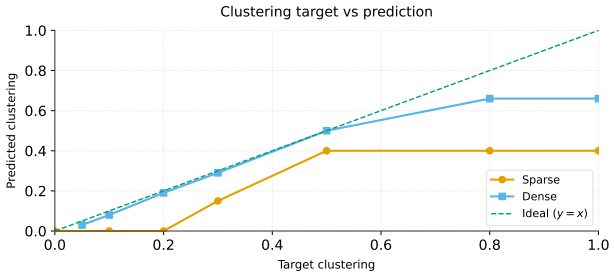


Figure 6: Semi-synthetic networks from our rewiring procedure. Target vs. achieved clustering on Wiki-based (sparse) and Reed98-based (dense) networks. Both saturate at an upper plateau; sparse undershoots low non-zero targets due to limited annealing budgets.

$k \in \{5, 20, 100\}$, PPRGo $\alpha \in \{0.5, 0.7, 0.5, 0.5\}$, Tail-STEAK deg_t $\in \{15, 7, 5, 10\}$ for Twitter/Pokec/Wiki/Reed98, respectively.

The graph rewiring pipeline uses NetworkX/NumPy/SciPy with: SGC depth $K = 10$, embedding dimension $d = 64$, two-hop weight $\eta = 0.6$, $\mathbf{A}^2$ row-capping threshold $\theta = 0.05$, degree-weighted centering $\gamma = 1.0$, and a geometric annealing schedule from $T_0 = 1.5$ to $T_f = 0.005$ over 100,000 swap proposals.

Experiments ran on a Linux server (Xeon E5-2630 v3 CPU, 11 GB NVIDIA GeForce GTX 1080 Ti GPU, and Python 3.9). Most runtimes were under 5 hours; Tail-STEAK on Pokec took ~6 days. Semi-synthetic generation took 25 minutes for all $21 \times 10 = 210$ graphs. Our code is available at: https://github.com/ElifEceErdem/salsa-baselines-and-graph-rewiring under an MIT license.

GenAI Usage Disclosure

We used Claude Opus 4.6 for grammar and style editing, to improve the clarity of some technically dense text written by us, to address LaTeX formatting issues, and for assistance with structuring figures and tables. No AI tools were used for data analysis, experimentation, or the formulation of conclusions. We acknowledge the contributions of AI-Assisting Tools in enhancing the writing process while maintaining full academic integrity.

References

[1] Ayan Acharya, Siyuan Gao, Borja Ocejo, Kinjal Basu, Ankan Saha, Keerthi Selvaraj, Rahul Mazumdar, Parag Agrawal, and Aman Gupta. 2023. Promoting Inactive Members in Edge-Building Marketplace. In *Companion Proceedings of the ACM Web Conference 2023* (Austin, TX, USA) *(WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 945–949.

[2] Ravindra K Ahuja, Thomas L Magnanti, and James B Orlin. 1994. *Network Flows: Theory, Algorithms and Applications*. Prentice hall.

[3] Rachith Aiyappa, Xin Wang, Munjung Kim, Ozgur Can Seckin, Yong-Yeol Ahn, and Sadamori Kojaku. 2025. Implicit Degree Bias in the Link Prediction Task. In *Proceedings of the 42nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 267)*, Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (Eds.). PMLR, 874–908. https://proceedings.mlr.press/v267/aiyappa25a.html

[4] Jeff Alstott, Christine Klymko, Pamela B Pyzza, and Mary Radcliffe. 2018. Local Rewiring Algorithms to Increase Clustering and Grow a Small World. *Journal of Complex Networks* 7, 4 (2018), 564–584. https://doi.org/10.1093/comnet/cny032

[5] Adrian Arnaiz-Rodriguez, Georgina Curto Rex, and Nuria Oliver. 2025. Structural Group Unfairness: Measurement and Mitigation by Means of the Effective Resistance. *Proceedings of the International AAAI Conference on Web and Social Media* 19, 1 (Jun. 2025), 83–106. https://ojs.aaai.org/index.php/ICWSM/article/view/35805

[6] Shweta Bansal, Shashank Khandelwal, and Lauren Ancel Meyers. 2009. Exploring Biological Network Structure with Clustered Random Networks. *BMC Bioinformatics* 10 (Dec. 2009), 405.

[7] Aleksandar Bojchevski, Johannes Gasteiger, Bryan Perozzi, Amol Kapoor, Martin Blais, Benedek Rózemberczki, Michal Lukasik, and Stephan Günnemann. 2020. Scaling Graph Neural Networks with Approximate PageRank. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Virtual Event, CA, USA) *(KDD '20)*. 2464–2473.

[8] Gergana Bounova and Olivier de Weck. 2012. Overview of Metrics and their Correlation Patterns for Multiple-metric Topology Analysis on Heterogeneous Graph Ensembles. *Phys. Rev. E* 85 (Jan 2012), 016117. Issue 1. https://link.aps.org/doi/10.1103/PhysRevE.85.016117

[9] Federico Cinus, Marco Minici, Corrado Monti, and Francesco Bonchi. 2022. The Effect of People Recommenders on Echo Chambers and Polarization. *Proceedings of the International AAAI Conference on Web and Social Media* 16, 1 (May 2022), 90–101. doi:10.1609/icwsm.v16i1.19275

[10] Robert Dorfman. 1979. A Formula for the Gini Coefficient. *The Review of Economics and Statistics* 61, 1 (1979), 146–149. http://www.jstor.org/stable/1924845

[11] Brooke Erin Duffy and Colten Meisner. 2023. Platform Governance at the Margins: Social Media Creators' Experiences with Algorithmic (In)visibility. *Media, Culture & Society* 45, 2 (2023), 285–304.

[12] Lisette Espín-Noboa, Claudia Wagner, Markus Strohmaier, and Fariba Karimi. 2022. Inequality and Inequity in Network-based Ranking and Recommendation Algorithms. *Scientific Reports* 12, 1 (07 Feb 2022), 2012. https://doi.org/10.1038/s41598-022-05434-1

[13] Francesco Fabbri, Maria Luisa Croci, Francesco Bonchi, and Carlos Castillo. 2022. Exposure Inequality in People Recommender Systems: The Long-Term Effects. *Proceedings of the International AAAI Conference on Web and Social Media* 16, 1 (May 2022), 194–204. doi:10.1609/icwsm.v16i1.19284

[14] Antonio Ferrara, Lisette Espin-Noboa, Fariba Karimi, and Claudia Wagner. 2022. Link Recommendations: Their Impact on Network Structure and Minorities. In *Proceedings of the 14th ACM Web Science Conference 2022* (Barcelona, Spain). 228–238.

[15] Minas Gjoka, Maciej Kurant, and Athina Markopoulou. 2013. 2.5K-Graphs: From Sampling to Generation. In *2013 Proceedings IEEE INFOCOM*. 1968–1976. doi:10.1109/INFCOM.2013.6566997

[16] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. 855–864.

[17] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. 2013. WTF: The Who to Follow Service at Twitter. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. 505–514.

[18] S. L. Hakimi. 1962. On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph. I. *J. Soc. Indust. Appl. Math.* 10, 3 (1962), 496–506. doi:10.1137/0110037

[19] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. 639–648.

[20] Petter Holme and Beom Jun Kim. 2002. Growing Scale-free Networks with Tunable Clustering. *Phys. Rev. E* 65 (Jan 2002), 026107. Issue 2. https://link.aps.org/doi/10.1103/PhysRevE.65.026107

[21] Emanuel Lacic, Dominik Kowald, Denis Parra, Michael Kahr, and Christoph Trattner. 2014. Towards a Scalable Social Recommender Engine for Online Marketplaces. In *Proceedings of the 23rd International Conference on World Wide Web*. 817–822. doi:10.1145/2567948.2579245

[22] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. 2008. Benchmark Graphs for Testing Community Detection Algorithms. *Phys. Rev. E* 78 (Oct 2008), 046110. Issue 4. https://link.aps.org/doi/10.1103/PhysRevE.78.046110

[23] Ronny Lempel and Shlomo Moran. 2001. SALSA: The Stochastic Approach for Link-Structure Analysis. *ACM Trans. Inf. Syst.* 19, 2 (April 2001), 131–160. doi:10.1145/382979.383041

[24] Yanying Li, Xiuling Wang, Yue Ning, and Hui Wang. 2022. FairLP: Towards Fair Link Prediction on Social Network Graphs. *Proceedings of the International AAAI Conference on Web and Social Media* 16, 1 (May 2022), 628–639. https://ojs.aaai.org/index.php/ICWSM/article/view/19321

[25] Zhepeng (Lionel) Li, Xiao Fang, and Olivia R. Liu Sheng. 2017. A Survey of Link Recommendation for Social Networks: Methods, Theoretical Foundations, and Future Research Directions. *ACM Trans. Manage. Inf. Syst.* 9, 1, Article 1 (Oct. 2017), 26 pages.

[26] David Liben-Nowell and Jon Kleinberg. 2003. The Link Prediction Problem for Social Networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management* (New Orleans, LA, USA) *(CIKM '03)*. 556–559. https://doi.org/10.1145/956863.956972

[27] David Liu, Virginie Do, Nicolas Usunier, and Maximilian Nickel. 2023. Group fairness without demographics using social networks. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1432–1449. https://doi.org/10.1145/3593013.3594091

[28] Ying Chieh Liu and Min Qi Huang. 2021. Examining the Matthew Effect on YouTube Recommendation System. In *2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*. 146–148.

[29] Zemin Liu, Trung-Kien Nguyen, and Yuan Fang. 2023. On Generalized Degree Fairness in Graph Neural Networks. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23)*. AAAI Press, Article 505, 9 pages. doi:10.1609/aaai.v37i4.25574

[30] Weishuo Ma, Yanbo Wang, Xiyuan Wang, and Muhan Zhang. 2025. Reconsidering the Performance of GAE in Link Prediction. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (Seoul, Republic of Korea) *(CIKM '25)*. Association for Computing Machinery, New York, NY, USA, 2052–2062. doi:10.1145/3746252.3761108

[31] Yijun Ma, Chaozhuo Li, and Xiao Zhou. 2024. Tail-STEAK: Improve Friend Recommendation for Tail Users via Self-training Enhanced Knowledge Distillation. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*. Article 989, 9 pages.

[32] Shane Mannion, Padraig MacCarron, Akrati Saxena, and Frank W. Takes. 2024. Fast Degree-preserving Rewiring of Complex Networks. *arXiv preprint arXiv:2401.12047* (2024). https://arxiv.org/abs/2401.12047

[33] Lilian Marey, Mathilde Perez, Tiphaine Viard, and Charlotte Laclau. 2026. TopoFair: Linking Topological Bias to Fairness in Link Prediction Benchmarks. arXiv:2602.11802 [cs.LG] https://arxiv.org/abs/2602.11802

[34] João Mattos, Debolina Halder Lina, and Arlei Silva. 2025. Breaking the Dyadic Barrier: Rethinking Fairness in Link Prediction Beyond Demographic Parity. *arXiv preprint arXiv:2511.06568* (2025).

[35] Masoud Mazloom, Bregt Hendriks, and Marcel Worring. 2017. Multimodal Context-Aware Recommender for Post Popularity Prediction in Social Media. In *Proceedings on Thematic Workshops of ACM Multimedia 2017*. 236–244. doi:10.1145/3126686.3126731

[36] Robert K. Merton. 1968. The Matthew Effect in Science. *Science* 159, 3810 (1968), 56–63. https://www.science.org/doi/abs/10.1126/science.159.3810.56

[37] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs.CL] https://arxiv.org/abs/1301.3781

[38] Batul J. Mirza, Benjamin J. Keller, and Naren Ramakrishnan. 2003. Studying Recommendation Algorithms by Graph Analysis. *Journal of Intelligent Information Systems* 20, 2 (01 Mar 2003), 131–160. doi:10.1023/A:1021819901281

[39] Mark Newman. 2009. Random Graphs with Clustering. *Phys. Rev. Lett.* 103 (Jul 2009), 058701. Issue 5. doi:10.1103/PhysRevLett.103.058701

[40] Mark Newman. 2018. *Networks*. Oxford University Press. doi:10.1093/oso/9780198805090.001.0001

[41] Manjish Pal, Sandipan Sikdar, and Niloy Ganguly. 2025. Fair Link Prediction With Overlapping Groups. *IEEE Transactions on Computational Social Systems* 12, 3 (2025), 998–1012.

[42] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, New York, USA) *(KDD '14)*. 701–710.

[43] Ryan A. Rossi and Nesreen K. Ahmed. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. https://networkrepository.com

[44] Javier Sanz-Cruzado and Pablo Castells. 2018. Enhancing structural diversity in social networks by recommending weak ties. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 233–241. doi:10.1145/3240323.3240371

[45] Javier Sanz-Cruzado and Pablo Castells. 2020. Beyond Accuracy in Link Prediction. In *Bias and Social Aspects in Search and Recommendation*, Ludovico Boratto, Stefano Faralli, Mirko Marras, and Giovanni Stilo (Eds.). 79–94.

[46] Javier Sanz-Cruzado and Pablo Castells. 2022. RELISON: A Framework for Link Recommendation in Social Networks. In *SIGIR 2022: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11-15, 2021*. 2992––3002.

[47] Xiran Song, Jianxun Lian, Hong Huang, Mingqi Wu, Hai Jin, and Xing Xie. 2022. Friend Recommendations with Self-Rescaling Graph Neural Networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington DC, USA) *(KDD '22)*. 3909–3919.

[48] Ana-Andreea Stoica, Nelly Litvak, and Augustin Chaintreau. 2024. Fairness Rising from the Ranks: HITS and PageRank on Homophilic Networks *(WWW '24)*. 2594–2602. https://doi.org/10.1145/3589334.3645609

[49] Jessica Su, Aneesh Sharma, and Sharad Goel. 2016. The Effect of Recommendations on Network Structure. In *Proceedings of the 25th International Conference on World Wide Web* (Montréal, Québec, Canada) *(WWW '16)*. 1157–1167.

[50] Arjun Subramonian, Levent Sagun, and Yizhou Sun. 2024. Networked Inequality: Preferential Attachment Bias in Graph Neural Network Link Prediction. In *Proceedings of the 41st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 235)*. 46891–46925.

[51] Lubos Takac and Michal Zabovsky. 2012. Data Analysis in Public Social Networks. In *International Scientific Conference & International Workshop Present Day Trends of Innovations*.

[52] Antonela Tommasel and Filippo Menczer. 2022. Do Recommender Systems Make Social Media More Susceptible to Misinformation Spreaders?. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 550–555. doi:10.1145/3523227.3551473

[53] Sotiris Tsioutsiouliklis, Evaggelia Pitoura, Konstantinos Semertzidis, and Panayiotis Tsaparas. 2022. Link Recommendations for PageRank Fairness. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 3541–3551. doi:10.1145/3485447.3512249

[54] Duncan J. Watts and Steven H. Strogatz. 1998. Collective Dynamics of 'Small-world' Networks. *Nature* 393, 6684 (01 Jun 1998), 440–442. https://doi.org/10.1038/30918

[55] Qiannan Yin, Yan Wang, Divya Venugopalan, Cyrus Dicicco, Heloise Logan, Preetam Nandy, Kinjal Basu, and Albert Cui. 2021. *Optimizing People You May Know (PYMK) for Equity in Network Creation*. LinkedIn Engineering Blog. https://www.linkedin.com/blog/engineering/member-customer-experience/optimizing-pymk-for-equity-in-network-creation