applied sciences

MDPI

Article

Content-Agnostic Early Detection of Fake Accounts in Growing Social Networks

Sotiria Liappi and Dimitrios Katsaros *

Department of Electrical and Computer Engineering, University of Thessaly, 38334 Volos, Greece; sliappi@uth.gr
* Correspondence: dkatsar@uth.gr

Abstract

The rapid spread of misinformation in online social networks is frequently amplified by fake or automated accounts that manipulate information diffusion and user perception. Detecting such accounts at an early stage remains a significant challenge, as traditional detection methods rely heavily on behavioral or content-based signals that are unavailable for newly created users. This work investigates whether fake accounts can be identified using only structural information derived from network growth dynamics. We propose a content-agnostic framework for early fake account detection based exclusively on topological features extracted from evolving social graphs. Networks are modeled using two preferential attachment mechanisms, namely the Barabási–Albert (BA) and Barabási–Fitness (BF) models, enabling controlled simulation of realistic network evolution. Synthetic graphs containing injected fake nodes are generated, from which interpretable node-level structural features are computed. Multiple supervised machine learning classifiers are then trained to distinguish genuine from fake accounts. Experiments under BA and BF growth show that topology-only features often yield high overall accuracy, yet precision, recall, and F1 for the fake class remain moderate because fake accounts are the minority. Gradient Boosting and Extra Trees provide the most balanced positive-class performance, whereas accuracy-leading models such as Random Forest can under-detect fakes, and KNN can degrade sharply under fitness-driven attachment at high fake prevalence. These findings suggest that early structural signals are informative but incomplete, and that content-agnostic detectors should be assessed with class-sensitive metrics rather than accuracy alone.

Keywords: fake account detection; preferential attachment; network growth; graph-based learning; social network analysis

Check for updates

Academic Editor: Guangjie Han

Received: 8 October 2026
Revised: 31 October 2026
Accepted: 5 November 2026
Published: 10 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Online social networks have become one of the primary channels for information dissemination, public discourse, and opinion formation. While these platforms enable rapid communication and large-scale connectivity, they are also increasingly exploited by malicious actors through fake or automated accounts that amplify misinformation, manipulate public opinion, and artificially influence network dynamics [1]. The early detection of such accounts has therefore emerged as a critical research challenge in network science and machine learning.

Traditional fake account and bot detection methods largely rely on content-based signals, including linguistic patterns, posting behavior, sentiment characteristics, and multimedia analysis. Although these approaches have achieved high accuracy in mature accounts with extensive activity histories, they suffer from an inherent limitation: newly created

accounts produce little or no observable content. Consequently, detection systems that depend on behavioral data often fail during the early stages of account existence, precisely when intervention would be most effective.

An alternative perspective considers the structural properties of social networks. Social systems evolve through dynamic processes in which new users establish connections according to underlying attachment mechanisms. Empirical studies have demonstrated that real-world networks are shaped by both preferential attachment and homophily, resulting in scale-free structures, pronounced community organization, and predictable growth dynamics. These structural dynamics provide valuable information even before users generate observable content, suggesting that topology alone may reveal early indicators of malicious behavior.

Fake accounts attempting to integrate into a growing network must operate within existing structural constraints. Although adversarial nodes may imitate legitimate connection strategies, their insertion timing, connectivity objectives, and adaptation to established topology can produce subtle deviations in local and global structural patterns. Such deviations motivate the hypothesis that early detection can be achieved through interpretable topological features derived solely from network structure.

In this work, we investigate content-agnostic algorithms for early fake account detection by modeling social networks as evolving graphs generated through preferential attachment mechanisms. Specifically, we study two canonical growth models, the Barabási–Albert (BA) model and the Barabási–Fitness (BF) model, to simulate realistic network evolution scenarios. Synthetic networks enable controlled experimentation, allowing the isolation of structural signals without confounding behavioral factors.

The proposed framework extracts interpretable node-level topological features and applies supervised machine learning classifiers to distinguish genuine from fake accounts during early stages of network development. By focusing exclusively on structural information, the approach evaluates whether reliable detection is possible prior to observable behavioral activity.

The contributions of this paper are summarized as follows:

- We propose a content-agnostic framework for early fake account detection based solely on network topology.
- We analyze detection performance under two preferential attachment growth mechanisms (BA and BF), enabling controlled comparison of structural dynamics.
- We evaluate multiple classical machine learning classifiers using interpretable graph-based features.
- We show that early-stage structural signals are informative but incomplete under class imbalance, and we quantify how fake–fake coordination (ρ) and hybrid randomness affect Acc/P/R/F1/PR-AUC without textual or behavioral data.

Although closely related to Breuer et al. [2], who proposed PreAttacK, a preferential-attachment-based classifier with provable guarantees for early fake-account detection on large-scale social network data, the present study adopts a complementary controlled-evaluation perspective. Their work establishes that structural signals can be informative before behavioral content becomes available. However, PreAttacK is tied to a specific multi-class attachment model and to deployment-oriented friend-request dynamics, which leaves several vulnerabilities unaddressed: the classifier may be misspecified when growth does not follow the assumed attachment process; detection can weaken if fake accounts deviate from the modeled targeting behavior; and robustness across alternative generative mechanisms, classifier families, and graph-scale parameters remains less systematically characterized. Rather than deriving a single model-specific classifier, we therefore assess whether lightweight, interpretable topological features combined with classical supervised learners

can distinguish fake from genuine accounts under two canonical growth mechanisms (BA and BF), with explicit sweeps over network size, fake prevalence, and connectivity intensity.

The remainder of the paper is organized as follows. Section 2 reviews related work on content-based and linkage-based detection methods. Section 3 presents the proposed framework. Section 4 describes the experimental setup. Section 5 reports the results, Section 6 discusses the findings and limitations, and finally Section 7 concludes the article.

2. Related Work

Fake account and bot detection in online social networks has been extensively studied across multiple research directions, primarily categorized into content-based approaches and linkage-based (graph-structural) methods. More recent studies also investigate hybrid and early-stage detection strategies.

2.1. Content-Based Bot Detection

Content-based detection methods focus on analyzing textual, behavioral, or multimedia information generated by users. Early work examined linguistic patterns, posting frequency, sentiment distribution, and semantic consistency to distinguish automated or malicious accounts from genuine users.

Ferrara et al. [3] provide a comprehensive overview of social bots, highlighting how automated accounts can be identified through abnormal content generation patterns and coordinated messaging behavior. Similarly, Varol et al. [4] introduced Botometer, a widely used framework that evaluates accounts using features derived from content, sentiment, and temporal activity signals.

Machine learning models using textual representations have also been widely explored. Yang et al. [5] demonstrated that spam accounts exhibit distinctive posting characteristics and URL-sharing behavior. Cresci et al. [6] further showed that modern bots can mimic human language effectively, reducing the reliability of purely content-based indicators.

Recent advances employ deep learning architectures for semantic analysis. Kudugunta and Ferrara [7] proposed recurrent neural networks for tweet-level bot detection, while Gilani et al. [8] evaluated feature importance across content and metadata signals. Despite strong performance, these approaches depend on sufficient activity history, making them unsuitable for early detection scenarios where little or no content exists.

2.2. Linkage-Based and Graph Structural Detection

Linkage-based methods analyze the structural properties of social networks rather than user-generated content. These approaches rely on the observation that malicious accounts often exhibit anomalous connectivity patterns.

Early graph-based detection methods exploited community structure and trust propagation. SybilGuard [9] and SybilLimit [10] leveraged random walks on social graphs to identify Sybil nodes based on limited attack edges between honest and malicious regions. Subsequent work such as SybilRank [11] improved scalability by ranking nodes according to structural trust scores.

Graph topology features have also proven effective for machine learning-based detection. Fire et al. [12] analyzed friendship networks to identify fake profiles using clustering and connectivity measures. Stringhini et al. [13] demonstrated that spam accounts form distinguishable structural interaction patterns across multiple social platforms.

More recently, graph representation learning approaches have been introduced. Techniques such as node embeddings and graph neural networks capture higher-order structural relationships for bot detection [14]. However, these methods typically require large datasets and computationally intensive training.

Finally, graph times series analysis and verification is an field of research combining graph data and time series analysis and has been used for addressing issues such as outlier and anomaly detection, diffusion processes, statistics, etc. [15,16].

2.3. Early and Semi-Supervised Detection, Network Growth and Graph Neural Networks

A growing line of research investigates detecting malicious accounts during early network evolution, before behavioral signals emerge. Preferential attachment models have been widely used to study network formation dynamics [17,18]. Leskovec et al. [19] showed that evolving networks follow predictable densification and growth patterns, suggesting structural regularities that can support early anomaly detection. Subsequently, several works investigated early detection of "events", e.g., rumors, fake news etc using mainly the network structure [20–22].

Mining of structural characteristics of networks can be efficiently performed using Graph Neural Networks. They have extensively been used for molecular structure modeling and prediction, in wireless communications, fake account detection, etc. [23–29].

Semi-supervised has been a hot research topic in machine learning in the last decade and it has found applications in computer visions, social networking, cybersecurity and many other fields [30–32]. It is very effective when we need to used unlabeled data to enhance the performance of strictly supervised learning algorithms.

The closest prior work in this direction is PreAttacK [2], which operationalizes preferential attachment for early fake-account detection at scale. As discussed in the Introduction, however, it is specialized to a particular attachment model and friend-request dynamics, leaving open how detection behaves under alternative growth assumptions and classical feature-based learners.

2.4. Class Imbalance Learning and Cost-Sensitive Learning

Class imbalance remains a great challenge in machine learning; typical algorithms assume an even class balance, but this may lead to a bias towards the majority class. This bias causes the classifier to overemphasize the majority class (negative class) while neglecting the minority class (positive class), which may be critical. A rich set of algorithms have been proposed in the literature in order to address this issue [33–36].

Another approach to address this issue is utilizing cost-sensitive learning. Robust cost-sensitive classifiers have been constructed by changing the objective functions of algorithms, including decision trees, extreme gradient boosting, random forests, etc. [37,38].

2.5. Research Gap

While content-based methods need behavioral history and many graph methods assume static graphs or heavy embeddings, we target a narrower gap: a controlled comparison of interpretable topological features with classical supervised learners under BA/BF (and hybrid) growth, with explicit sweeps of scale, prevalence, connectivity, and fake–fake coordination ρ, evaluated with class-sensitive metrics (including PR-AUC) and majority baselines. We position this as complementary to PreAttacK rather than an experimental refutation of it.

3. Proposed Framework

3.1. Problem Formulation

The focus of our research is on detecting fake accounts at an early stage in social networks that are constantly changing. The time-dependent graph G_t at growth step t is defined as (V_t, E_t), where V_t represents the set of nodes (user accounts) and E_t represents the set of edges (social connections). The network develops incrementally through a pref-

erential attachment process, where new nodes link to existing nodes based on structural characteristics such as degree or inherent fitness.

A subset of nodes $F \subset V_t$ represents the fake accounts that were added to the network as it expanded. Malicious nodes attempt to conceal themselves by adhering to the same network connectivity protocols as legitimate users. The remaining nodes, denoted by $R = V_t \setminus F$, represent authentic user accounts. In the initial stages of network development, behavioral indicators (such as posting habits or content characteristics) may not yet be present, which means structural attributes become the principal source of information for detection.

The aim is to train a classification function that maps real-valued vectors of extracted topological features and not content data, denoted by $\mathbb{R}^d$, to an output label of either 0, indicating a genuine node, or 1, indicating a fake node, where d represents the number of extracted topological features.

Each node $v \in V_t$ is described by a feature vector $\mathbf{x}_v \in \mathbb{R}^d$ that is solely based on the interpretable structural properties of the graph.

The issue is therefore formulated as a supervised binary classification problem for nodes under specific generative constraints. The objective is to establish a classifier that effectively identifies fake accounts by employing solely early-stage topological data from a network produced via a network growth model, in conjunction with authentic nodes pre-designated.

This formulation enables a systematic comparison of classifier performance across different generative mechanisms, specifically BA and BF models, and isolates the contribution of structural growth dynamics to detection feasibility.

For clarity of presentation, we include Table 1 that summarizes the notation used throughout the paper.

Table 1. Summary of notation used in the problem formulation.

Symbol	Description
$G_t = (V_t, E_t)$	Graph at time step t
V_t	Set of nodes (accounts) at time t
E_t	Set of edges (connections) at time t
t	Network growth step (time index)
v	A node in the network ($v \in V_t$)
$F \subset V_t$	Set of fake (malicious) nodes
$R = V_t \setminus F$	Set of genuine (real) nodes
$\mathbf{x}_v \in \mathbb{R}^d$	Feature vector of node v
d	Number of extracted topological features
N	Total number of nodes in the network
ϕ	Fraction of fake nodes in the network
m	Number of edges added per new node
ρ	Fraction of edges from fake nodes to other fake nodes
$f : \mathbb{R}^d \to \{0, 1\}$	Classification function
$y_v \in \{0, 1\}$	Label of node v (1: fake/positive, 0: real)
BA	Barabási–Albert preferential attachment model
BF	Barabási–Fitness model with node fitness parameter
t_0	Prediction-start timestep (online evaluation begins at $t \geq t_0$)

3.2. Proposed Solution

The proposed approach is based on the hypothesis that structural patterns emerging during network growth contain detectable—though not necessarily strong—signals for distinguishing genuine users from fake accounts at early stages. BA/BF do not encode attribute homophily; instead, coordination among fake accounts is modeled explicitly by the fake→fake edge fraction ρ, while the remaining links follow preferential attachment (camouflage). We focus exclusively on topological properties derived from the evolving social graph.

Social networks often exhibit structural regularities where nodes with similar roles or characteristics tend to occupy comparable positions in the graph. Genuine users typically form connections that follow natural growth dynamics, while fake accounts, even when attempting to imitate legitimate behavior, may introduce subtle structural inconsistencies due to their delayed insertion or strategic connectivity patterns. These deviations can be captured through interpretable topological features.

To analyze these patterns, the detection framework explicitly considers network growth mechanisms. The network is modeled using preferential attachment processes, where new nodes connect to existing nodes according to structural properties such as degree or fitness. Fake accounts are introduced during the growth process and attempt to integrate using the same attachment rules, allowing controlled analysis of their structural behavior relative to genuine nodes.

The proposed detection pipeline consists of three main stages:

- *Graph Generation:* Synthetic networks are generated using preferential attachment models to simulate realistic social network evolution.
- *Feature Extraction:* A set of interpretable node-level topological features is extracted from the generated graphs.
- *Classification:* Supervised machine learning classifiers are trained to distinguish fake accounts from genuine users based on these structural features.

By leveraging structural signals produced by network growth dynamics, the proposed solution enables early detection of fake accounts without requiring behavioral data or complex graph representation techniques.

3.3. Incorporated Network Growth Models

In our pipeline, we utilize two canonical preferential attachment models: the Barabási–Albert (BA) model and the Barabási–Fitness (BF) model.

Barabási–Albert Model (BA). The BA model [17] captures the concept of preferential attachment, wherein new nodes are more likely to connect to existing nodes with higher degrees. This process results in scale-free networks with a power-law degree distribution, reflecting structural patterns commonly observed in real-world social graphs.

Formally, when a new node is introduced at time step t, the probability $\Pi(k_i)$ that it connects to an existing node i depends on the degree k_i of that node:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{1}$$

Assuming that node degrees evolve continuously, the growth of the degree k_i over time can be modeled as:

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_j k_j} \tag{2}$$

where m is the number of edges added by each new node.

Since the total degree in the network at time t is approximately:

$$\sum_j k_j = 2mt - m \tag{3}$$

the evolution equation simplifies (for large t) to:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \tag{4}$$

This formulation explains how high-degree nodes attract more connections over time, leading to the emergence of hubs and scale-free structure.

In this study, the BA model is used to simulate organic network growth and evaluate how fake nodes attempt to integrate by preferentially linking to high-degree nodes.

Barabási–Fitness Model (BF). The BF model [18] extends the BA framework by incorporating a fitness parameter that quantifies a node's inherent attractiveness. In our experiments, each node i is assigned an independent fitness $\eta_i \sim \text{Uniform}(0, 1)$ at birth (identical sampling for genuine and fake nodes). The attachment probability is

$$\Pi(i) = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{5}$$

This allows high-fitness nodes to attract links even with moderate degree. BF is used as a controlled stress test of fitness-modulated growth rather than as a claim that fitness equals real-user popularity.

Both models provide a controlled yet expressive framework for examining how early structural patterns—absent behavioral context—can inform classification decisions [39].

3.4. Classifiers Used

We evaluate five supervised machine learning classifiers commonly employed in graph-based node classification tasks. Each classifier is trained on structural node-level features derived from the synthetic networks. Unless stated otherwise we use scikit-learn 1.5.1 (Python 3.11.3, NetworkX 3.3, NumPy 1.26.4, and Matplotlib 3.9.0) defaults with random_state = 42, StandardScaler, feature normalization, and class_weight='balanced' for RF, Extra Trees, and Decision Tree. KNN uses $k = 5$ with Euclidean distance after scaling. Gradient Boosting uses 100 estimators. No extensive hyperparameter search is performed; the goal is a fair default comparison under the same feature set.

Random Forest. An ensemble method that constructs multiple decision trees and aggregates their outputs [40], Random Forest is known for its robustness to overfitting and effectiveness in high-dimensional feature spaces.

Extra Trees Classifier. This variant introduces additional randomness by selecting split thresholds at random [41], often yielding improved generalization and reduced training time.

Decision Tree. A single-tree classifier that partitions the feature space using rule-based decisions [42]. While interpretable, it is prone to overfitting without regularization.

K-Nearest Neighbors (KNN). A non-parametric method that assigns labels based on the majority class among a node's k nearest neighbors in feature space [43]. Its performance is sensitive to feature scaling and data sparsity.

Gradient Boosting. A sequential ensemble technique where each tree corrects the residuals of its predecessors [44]. Gradient Boosting often achieves strong performance on structured data but is sensitive to hyperparameter selection.

4. Experimental Evaluation

To evaluate the effectiveness and robustness of the proposed framework, we conduct a series of controlled experiments on networks whose growth is governed by preferential

attachment mechanisms. The evaluation focuses on how classification performance is influenced by network scale, fake account proportion, and connectivity dynamics.

4.1. Graph Generation

A systematic evaluation of the proposed methodology requires examining various data distributions, i.e., network growth rate, attachment policies, etc. So, synthetically generated data ought to be used in the evaluation so as to capture many different scenarios and truly capture the behaviour of the algorithms. On the other hand, usage of real data for the evaluation provides a solid confirmation of the performance of the algorithms in real production environments. However, in our case there exist no real data publicly available to be used in our study, so we have to confine ourselves to synthetically generated data. Our synthetically generated data follow the distribution of the corresponding real data as described in [2]. In the next paragraphs we describe how data are generated.

Networks are generated using both the Barabási–Albert (BA) and Barabási–Fitness (BF) models. Each network evolves incrementally, with one node added at each timestep. A predefined fraction ϕ of nodes is labeled as fake and injected during the growth process. Fake nodes share preferential attachment for non-coordinated edges (camouflage), but a fraction ρ of their m edges is reserved for other fake nodes (coordination). Thus "same attachment rule" refers to the preferential mechanism on the remaining $(1-\rho)m$ edges, not to $\rho = 0$. In the main Experiments 1–3 we use default $\rho = 0.3$. Experiment 4 sweeps $\rho \in \{0.1, 0.3, 0.6\}$ on hybrid networks with 70% preferential and 30% uniform-random attachment.

The synthetic network generation process is controlled by several parameters. Let N denote the total number of nodes in the network and ϕ the fraction of fake accounts. Each newly arriving node forms m edges to existing nodes according to a preferential attachment mechanism. For fake nodes, a fraction ρ of these edges is directed toward other fake nodes, modeling coordinated behavior among malicious accounts.

Algorithm 1 summarizes the graph construction process together with online prediction and evaluation. The tested values are summarized in Table 2.

Table 2. Experimental parameter settings used for synthetic network generation.

Parameter	Values Tested
Total number of nodes (N)	10,000, 25,000, 50,000
Fraction of fake nodes (ϕ)	0.05, 0.10, 0.20
Edges per new node (m)	2, 4, 10
Fake→fake fraction (ρ)	0.1, 0.3 (default), 0.6
Prediction start (t_0)	500
Hybrid PA/random (Exp. 4)	70%/30%

4.2. Performance Measures

Detection quality is assessed using four standard binary classification metrics, treating fake accounts as the positive class. Let TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively. Accuracy measures the overall fraction of correctly classified nodes:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{6}$$

Precision quantifies the reliability of positive predictions, i.e., the fraction of nodes predicted as fake that are indeed fake:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{7}$$

Recall (sensitivity) measures the fraction of actual fake accounts that are correctly identified:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{8}$$

Finally, the F1-score summarizes the trade-off between precision and recall through their harmonic mean:

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{9}$$

These metrics are computed from per-node predictions collected during network growth, from a fixed prediction-start threshold onward. We additionally report the area under the precision–recall curve (PR-AUC) using classifier scores $P(\text{fake})$, which is more informative than accuracy under class imbalance. As a trivial baseline, always predicting the majority (genuine) class yields accuracy $1 - \phi$ (e.g., 90% at $\phi = 10\%$) but zero fake-class recall.

4.3. Feature Extraction

For each node $v \in V$, we extract interpretable structural features capturing local topology and temporal position in the network. These include node degree, average neighbor degree, clustering coefficient, relative node age, and connectivity to early versus recently added nodes. We also include the fraction of out-neighbors that are already labeled fake (computed only from neighbors present before scoring v), which makes ρ-driven coordination observable without leaking v's own label.

4.4. Classification Setup

We evaluate five supervised classifiers: Decision Tree, Extra Trees, Random Forest, Gradient Boosting, and K-Nearest Neighbors (KNN). Evaluation follows Algorithm 1: from timestep $t_0 = 500$ onward, the classifier is retrained on all nodes except the newly arrived node v and then predicts v (online protocol). This avoids future-information leakage from a random 80/20 split. Section 5 reports Acc/P/R/F1 for Experiments 1–3 and Acc/P/R/F1/PR-AUC (mean $\pm$ std over three campaigns) for Experiment 4. Table 3 gives the tested range and the default used when that factor is held fixed.

Table 3. Summary of experimental variables, tested ranges, and default values.

Variable	Range of Values	Default Value
Total nodes (N)	10,000, 25,000, 50,000	25,000
Fake node fraction (ϕ)	0.05, 0.10, 0.20	0.10
Edges per node (m)	2, 4, 10	4

4.5. Experimental Design

Three experimental scenarios are considered:

- *Experiment 1: Network Growth.* Evaluation across increasing network sizes (N = 10,000, 25,000, 50,000 nodes).
- *Experiment 2: Fake Node Proportion.* Evaluation for varying fake account fractions ($\phi = 5\%, 10\%, 20\%$) at N = 25,000.
- *Experiment 3: Connectivity Patterns.* Evaluation under varying number of edges per new node ($m = 2, 4, 10$) at N = 25,000 and $\phi = 10\%$.

All experiments are conducted under both BA and BF models. For each setting we report accuracy together with precision, recall, and F1-score for the positive (fake) class.

Algorithm 1 Network Growth, Fake Account Injection, and Performance Evaluation

Require: Total nodes N, fake fraction ϕ, edges per node m, fake connection ratio ρ, prediction start t_0, classifier C
Ensure: Graph $G = (V, E)$; Accuracy, Precision, Recall, F1, PR-AUC (when scores are available)

```
Initialize graph G with a small seed network
F ← 0                                                    ▷ number of fake nodes
F_max ← φN
TP, FP, TN, FN ← 0
for t = |V| + 1 to N do
    if F < F_max then
        Determine whether new node v is fake
    else
        Assign node v as authentic
    end if
    Add node v to V
    if v is authentic then
        Connect m edges using preferential attachment
    else
        m_f ← ρm                                         ▷ edges to fake nodes
        m_r ← m − m_f                                    ▷ edges to authentic nodes
        Connect m_f edges to fake nodes
        Connect m_r edges to authentic nodes using preferential attachment
        F ← F + 1
    end if
    if t ≥ t_0 then
        Extract structural features for all nodes in G excluding v
        Train classifier C on the extracted features and labels
        Predict label ŷ_v for node v using C
        Let y_v be the true label of v                   ▷ fake = positive
        if ŷ_v = fake and y_v = fake then
            TP ← TP + 1
        else if ŷ_v = fake and y_v = authentic then
            FP ← FP + 1
        else if ŷ_v = authentic and y_v = authentic then
            TN ← TN + 1
        else
            FN ← FN + 1
        end if
    end if
end for
Accuracy ← (TP + TN)/(TP + TN + FP + FN)
Precision ← TP/(TP + FP)
Recall ← TP/(TP + FN)
F1 ← 2 · Precision · Recall/(Precision + Recall)
return G, Accuracy, Precision, Recall, F1
```

5. Results and Analysis

This section reports empirical findings along three complementary dimensions: (i) scalability with respect to network size, (ii) sensitivity to the prevalence of fake accounts, and (iii) sensitivity to the number of edges attached by each arriving node. For every experiment we present the full metric suite (accuracy, precision, recall, and F1), with BA and BF as compact side-by-side panels (*a*) and (*b*). Vertical axes are cropped or broken when needed so that differences among classifiers remain readable; absolute levels should still be read from the tick labels.

5.1. Network Growth and Scalability

In our first experiment, we study the impact of network size on detection performance. Keeping the remaining parameters fixed ($\phi = 10\%$), we evaluate all classifiers at three growth milestones ($N = 10{,}000$, 25,000, and 50,000) under both the BA and BF models.

Figure 1 shows that accuracy remains in a narrow high band (roughly 86–89%) across all three milestones for both BA and BF. Tree-based and ensemble classifiers do not exhibit a systematic accuracy collapse as N grows, and KNN stays in the same range. Random Forest is typically the most accurate method, but the gap to the other classifiers is small.

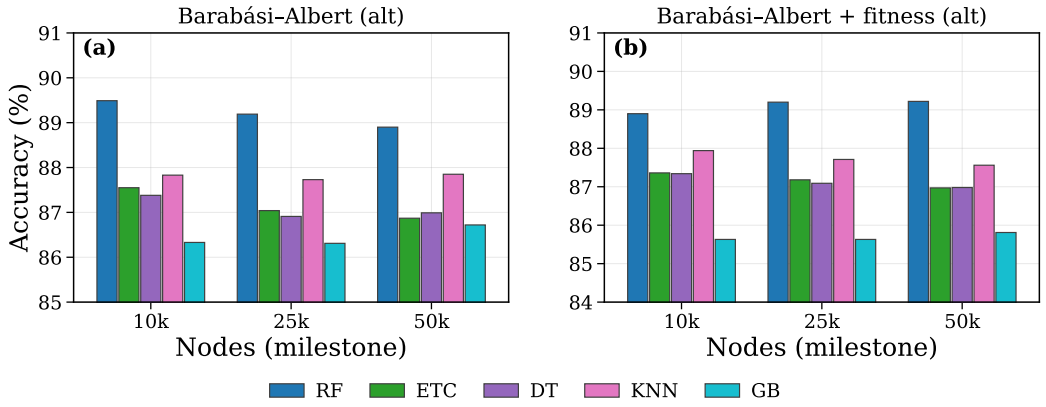


Figure 1. Experiment 1: average accuracy versus milestone network size (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

Precision in Figure 2 is likewise stable across network size, remaining in the mid-20% range under both generative models. This indicates that, among nodes labeled as fake, only about one in four predictions is correct—a direct consequence of class imbalance and of conservative positive predictions by several classifiers.

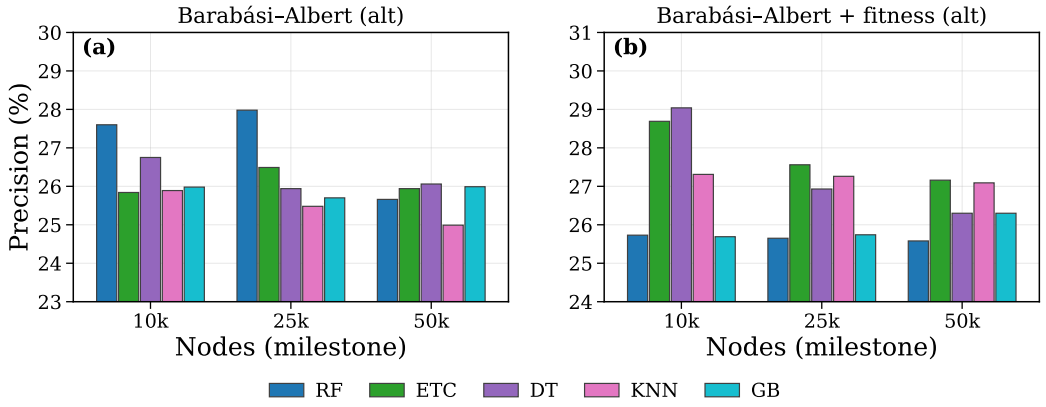


Figure 2. Experiment 1: precision (positive class = fake) versus milestone network size (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

The class-sensitive metrics in Figures 3 and 4 reveal a clearer ranking than accuracy alone. Gradient Boosting attains the highest recall and F1 (approximately 19–24% F1), Decision Tree and Extra Trees follow closely, whereas Random Forest—despite leading on accuracy—posts the lowest recall and F1 (often near 5–6% recall and about 8–10% F1). High accuracy can therefore overstate detection quality when fake accounts remain the minority class.

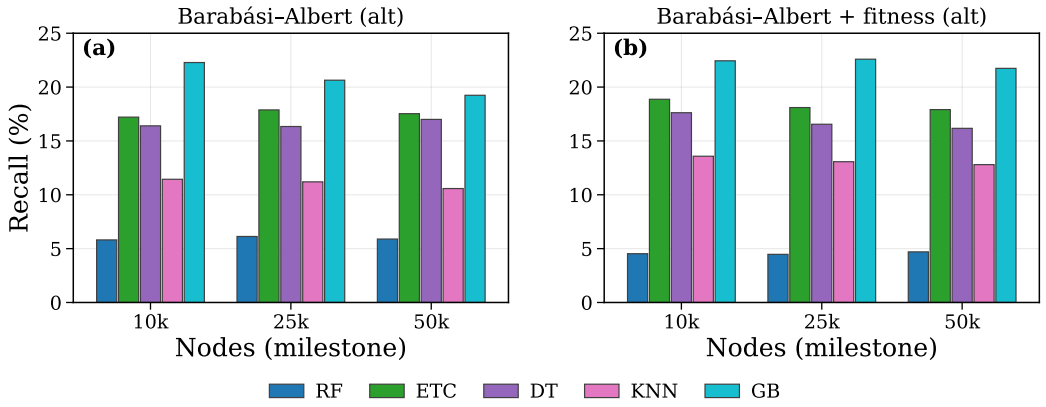


Figure 3. Experiment 1: recall (positive class = fake) versus milestone network size (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

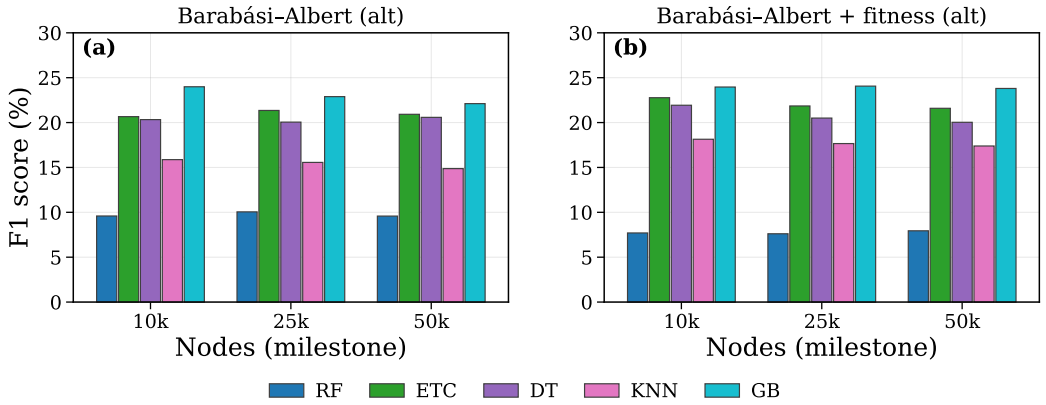


Figure 4. Experiment 1: F1-score (positive class = fake) versus milestone network size (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

5.2. Impact of Fake Account Prevalence

In our second experiment, we fix the network size at $N = 25{,}000$ and vary the fake-node fraction over $\phi \in \{5\%, 10\%, 20\%\}$, again under both BA and BF growth.

Figure 5 shows that accuracy is highest at $\phi = 5\%$ (typically mid-90%), decreases at $\phi = 10\%$, and is lowest at $\phi = 20\%$ (near 77–81% for most classifiers). KNN is the clear exception under BF at $\phi = 20\%$, where its accuracy collapses to about 21%—well below its BA counterpart and the other BF classifiers at the same prevalence.

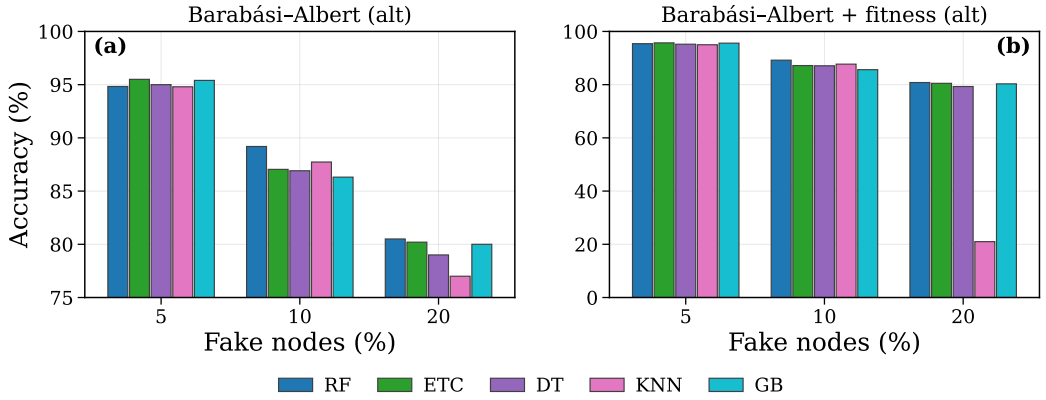


Figure 5. Experiment 2: average accuracy versus fake-node proportion at $N = 25{,}000$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

Figures 6–8 complement the accuracy trend. Precision remains comparatively flat in the mid-20% range across ϕ, whereas recall and F1 of the stronger detectors (especially Gradient Boosting and Extra Trees) improve modestly as fake accounts become less rare, consistent with a less skewed positive class. Random Forest again trades high accuracy for weak fake-class recall. Overall, higher fake prevalence makes accuracy harder to sustain, and fitness-driven attachment can interact with instance-based classification in a destabilizing way.

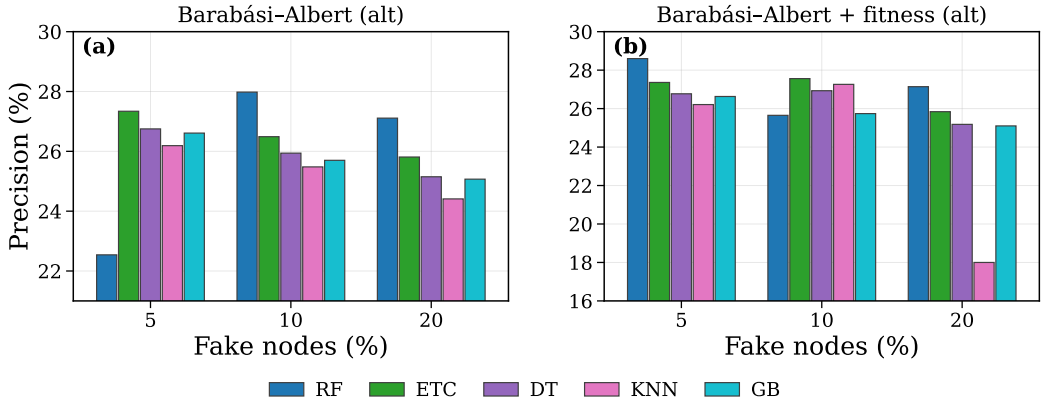


Figure 6. Experiment 2: precision (positive class = fake) versus fake-node proportion at $N = 25{,}000$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

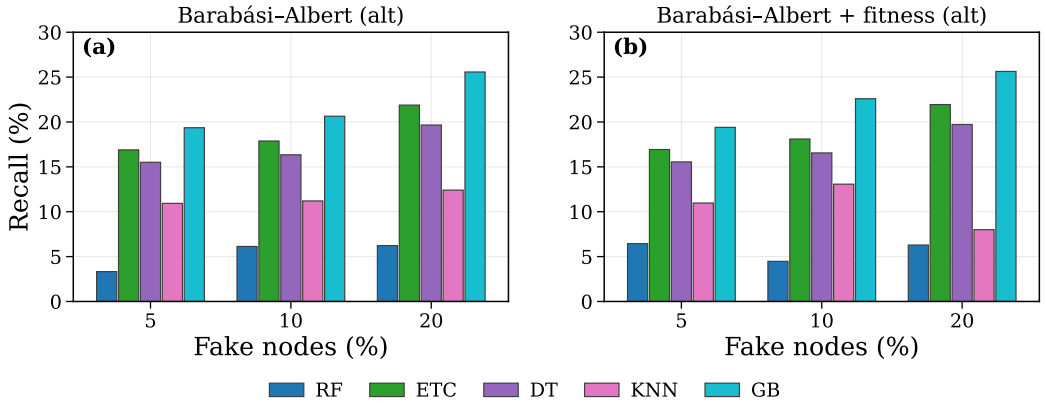


Figure 7. Experiment 2: recall (positive class = fake) versus fake-node proportion at $N = 25{,}000$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

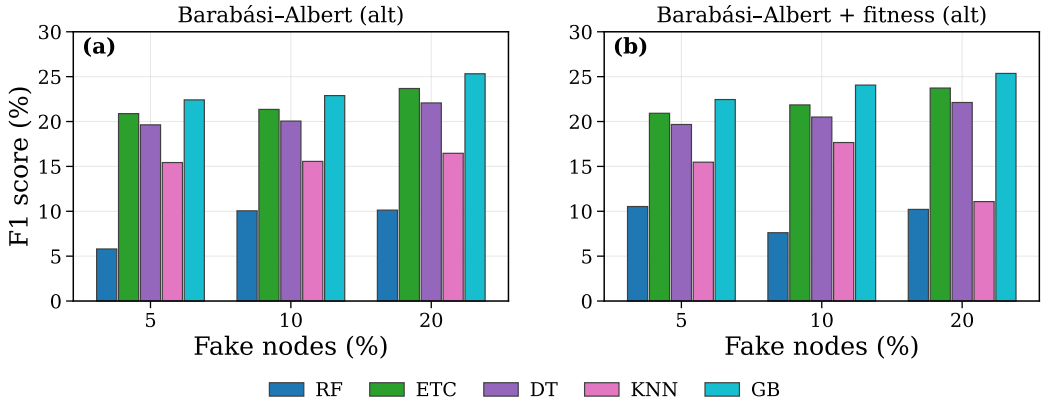


Figure 8. Experiment 2: F1-score (positive class = fake) versus fake-node proportion at $N = 25{,}000$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

5.3. Effect of Connectivity Intensity

In our third experiment, we keep $N = 25{,}000$ and $\phi = 10\%$ fixed, and vary the number of edges created by each arriving node over $m \in \{2, 4, 10\}$.

Figure 9 indicates that varying m has only a modest effect on accuracy for most classifier–model pairs: values stay near 86–91%, close to the Experiment 1 setting at $N = 25{,}000$ and $\phi = 10\%$. Under BA, all five classifiers are essentially flat across $m \in \{2, 4, 10\}$.

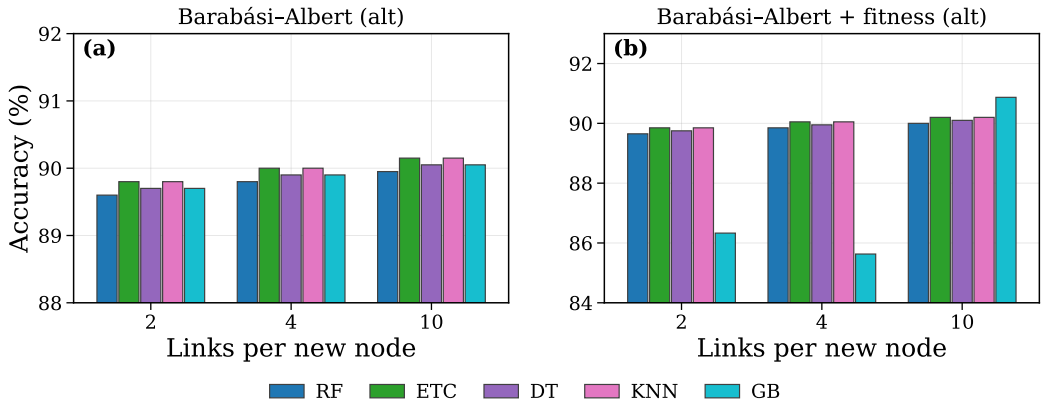


Figure 9. Experiment 3: average accuracy versus links per new node at $N = 25{,}000$ and $\phi = 10\%$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

Figures 10–12 show the same near-flat behavior under BA for precision, recall, and F1. Under BF, Gradient Boosting is a notable exception at $m = 10$, where precision, recall, and F1 rise sharply (F1 about 54%), while the other BF classifiers remain near their lower baseline. We treat this BF/$m = 10$/GB spike as a regime-sensitive outlier rather than a universal gain; the multi-run ρ-sweep below provides more stable evidence across coordination levels. Connectivity intensity is therefore largely secondary to fake prevalence, except in this denser BF setting where GB benefits substantially.

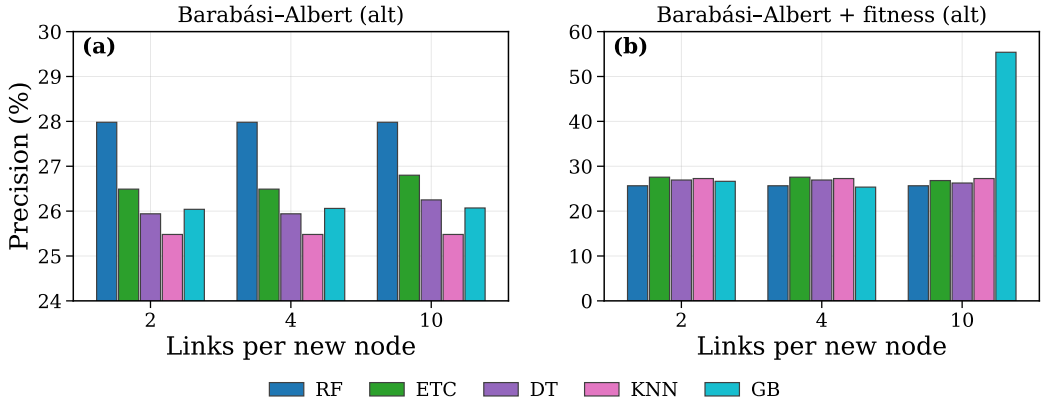


Figure 10. Experiment 3: precision (positive class = fake) versus links per new node at $N = 25{,}000$ and $\phi = 10\%$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

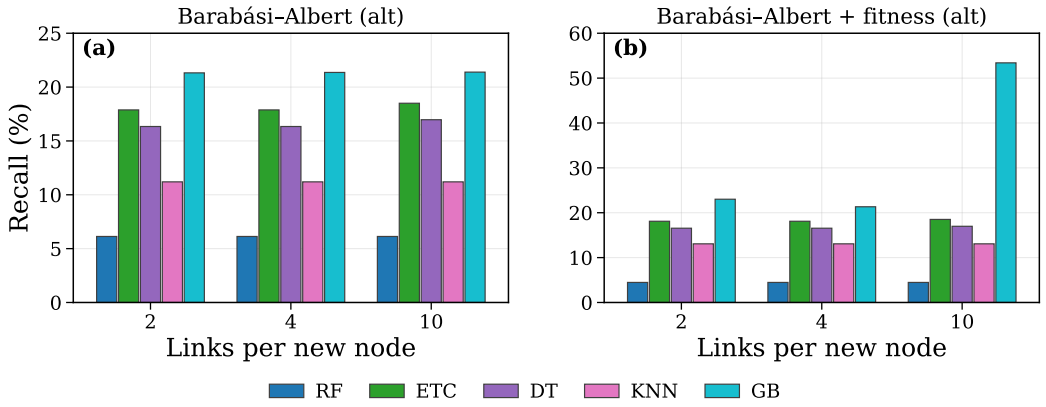


Figure 11. Experiment 3: recall (positive class = fake) versus links per new node at $N = 25{,}000$ and $\phi = 10\%$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

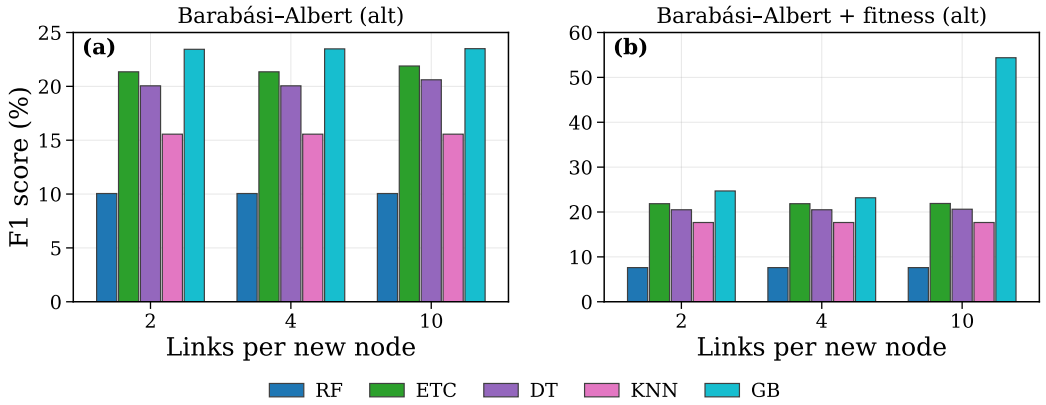


Figure 12. Experiment 3: F1-score (positive class = fake) versus links per new node at $N = 25{,}000$ and $\phi = 10\%$ (**a**) Barabási–Albert; (**b**) Barabási–Fitness.

5.4. Effect of Fake–Fake Coordination (ρ)

To address coordinated collusion among fake accounts, Experiment 4 fixes $N = 10{,}000$, $\phi = 10\%$, $m = 4$, and $t_0 = 500$, and sweeps $\rho \in \{0.1, 0.3, 0.6\}$. Growth is hybrid: each attachment decision uses preferential attachment with probability 0.7 and uniform random selection with probability 0.3, under both BA and BF preferential kernels. Every configuration is repeated in three independent experimental campaigns; we report mean ± std for Accuracy, Precision, Recall, F1, and PR-AUC.

Figures 13 and 14 show a clear monotone trend: at low coordination ($\rho = 0.1$), fake-class F1 and PR-AUC remain low (often below 0.15 PR-AUC), whereas at $\rho = 0.6$ several classifiers exceed 0.55–0.65 PR-AUC and F1 near 50–66%. Gradient Boosting and Extra Trees again provide balanced positive-class performance; Random Forest can retain high accuracy while lagging in recall. Wall-clock times for a single online run at $N = 10{,}000$ typically range from about 0.4–2.5 hours depending on the classifier (tree ensembles being slowest), confirming the computational cost of Algorithm 1.

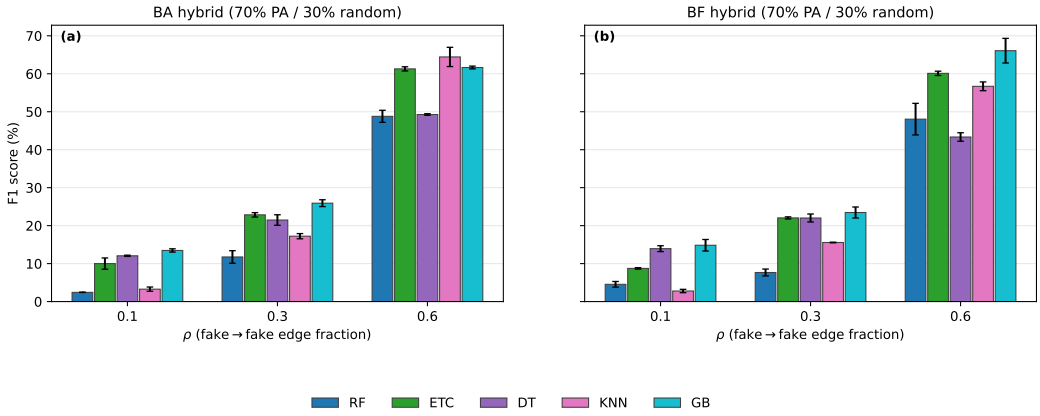


Figure 13. Experiment 4: F1 (positive=fake) versus ρ under hybrid growth (**a**) BA; (**b**) BF. Bars show mean ± std over three campaigns.

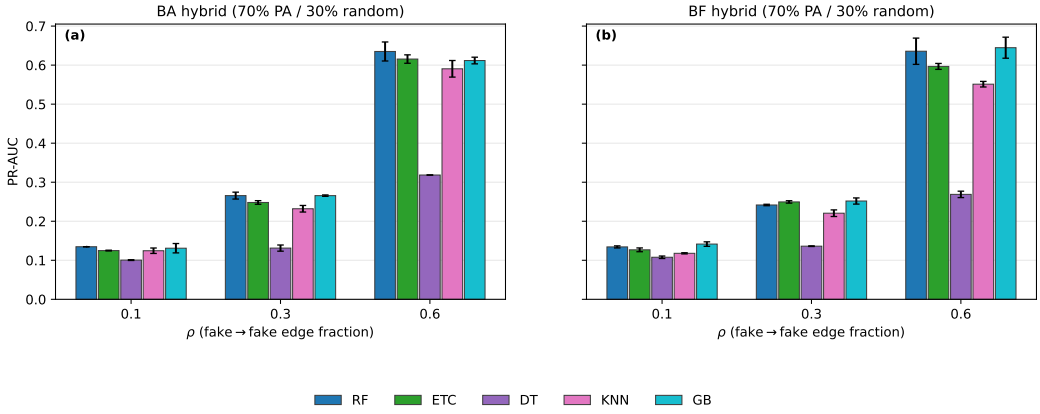


Figure 14. Experiment 4: PR-AUC (positive=fake) versus ρ under hybrid growth (**a**) BA; (**b**) BF.

Feature Importance

On a hybrid BA snapshot with $\rho = 0.3$, Random Forest importances rank fake out-neighbor fraction first ($\approx$0.51), followed by normalized age ($\approx$0.21) and average neighbor degree ($\approx$0.16). Raw degree features contribute little. This supports interpretability: coordination-sensitive neighborhood composition dominates, enabling future feature pruning for lighter deployment.

5.5. Overall Observations

Taken together, the experiments indicate that tree-based and boosting models provide the most balanced trade-off on the positive class: Gradient Boosting and Extra Trees typically lead in recall and F1, while Random Forest can post high accuracy with comparatively weak fake-class recall. KNN remains competitive on accuracy in many BA settings but can fail abruptly under BF at high fake prevalence.

BA and BF outcomes are often aligned, which supports the claim that early structural signals transfer across preferential-attachment mechanisms; the BF outliers (KNN at $\phi = 20\%$, GB at $m = 10$) nevertheless show that generative assumptions and classifier choice should be evaluated jointly.

Finally, because fake accounts are the minority class, accuracy alone is an incomplete success criterion. The precision, recall, and F1 panels should be read together with the accuracy plots, and apparent separations among methods should be interpreted in light of the controlled synthetic generation process. In particular, a majority-class baseline already achieves accuracy $1 - \phi$; methods should be judged primarily by fake-class recall/F1/PR-

AUC. Experiment 4 further shows that structural detectability rises with ρ: coordination creates stronger topological footprints, whereas low-ρ camouflage remains difficult.

6. Discussion

The experimental results highlight the potential of topology-based features for early fake account detection. Ensemble classifiers, particularly Gradient Boosting and Extra Trees, offered the strongest fake-class recall and F1 across both generative models, while high accuracy alone (e.g., for Random Forest) did not imply strong detection of the minority class.

KNN and Structural Sensitivity. KNN remained competitive on accuracy under many BA settings and across network-size milestones, but degraded sharply under BF at high fake prevalence ($\phi = 20\%$). This points to sensitivity to the interaction between instance-based classification and fitness-modulated attachment rather than to network size alone.

Fitness-Based Models and Realism. BA and BF often produced similar rankings, but BF exposed classifier-specific failure and success modes (KNN collapse at high ϕ; GB gains at denser m). Fitness-driven attachment therefore remains a useful stress test, mirroring real-world dynamics where influence depends on latent attributes beyond degree alone.

Metric Trade-offs. Across experiments, accuracy often exceeds 85%, yet precision stays near the mid-20% level and F1 rarely exceeds the mid-20% range except in the BF denser-connectivity case for Gradient Boosting. This gap is expected under class imbalance: many genuine nodes are classified correctly, which inflates accuracy, while recovery of fake accounts (recall) remains the harder objective. Reporting precision, recall, and F1 alongside accuracy is therefore essential for interpreting detection quality.

Feature Simplicity and Effectiveness. Despite relying on lightweight and interpretable features, several classifiers still achieve useful structural discrimination without textual or behavioral signals, demonstrating that early-stage detection does not necessarily require complex embeddings or deep architectures.

Relation to Prior Preferential-Attachment Detection. Our findings complement rather than supersede the approach of Breuer et al. [2]. PreAttacK targets deployment-scale early detection with theoretical guarantees derived from friend-request behavior; the present controlled synthetic study instead isolates how generative assumptions (BA vs. BF), hybrid randomness, and coordination ρ affect interpretable feature-based classifiers. We do not replicate PreAttacK here (different assumptions and unavailable aligned public setting in this study); therefore we do not claim experimental coverage of its failure modes, only complementary controlled evidence.

Limitations. All results remain synthetic: real OSNs involve mixed attacks, communities, noisy edges, and distribution drift. We did not validate on real traces. Adaptive adversaries beyond a fixed ρ are not modeled. Online retraining is computationally expensive. While we added PR-AUC, majority baselines, runtimes, and feature importances, ROC-AUC/confusion matrices and a full PreAttacK baseline remain future work. Topology-only signals are informative but incomplete for minority-class detection at low coordination.

7. Conclusions

We studied content-agnostic early fake-account detection under BA/BF (and hybrid) growth using interpretable topological features and five classical classifiers. High accuracy is common but can approach the majority-class baseline; fake-class Precision/Recall/F1 are moderate unless coordination ρ is high. Gradient Boosting and Extra Trees offer the best positive-class trade-offs, PR-AUC rises with ρ, and fake-neighborhood features dominate importance. Structural early detection is therefore promising as a complementary signal, not a standalone solution, pending real-network validation.

https://doi.org/10.3390/app162211115

There are several avenues for future work, such as deploying small or large language models to detect fake accounts [45,46], and also to use federated learning [47] to exploit knowledge dispersed across different actors/owners of social sites.

Author Contributions: Conceptualization, S.L. and D.K.; methodology, S.L. and D.K.; software, S.L.; validation, S.L.; formal analysis, S.L.; investigation, S.L.; resources, S.L.; data curation, S.L.; writing original draft preparation, S.L. and D.K.; writing review and editing, S.L. and D.K.; visualization, S.L.; supervision, D.K.; project administration, D.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: No such approval is required.

Informed Consent Statement: The study does not involve humans.

Data Availability Statement: Data available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Katsaros, D.; Stavropoulos, G.; Papakostas, D. Which machine learning paradigm for fake news detection? In Proceedings of the IEEE/WIC/ACM Web Intelligence Conference (WI), Thessaloniki, Greece, 14–17 October 2019; pp. 383–387.
2. Breuer, A.; Khosravani, N.; Tingley, M.; Cottel, B. Preemptive Detection of Fake Accounts on Social Networks via Multi-Class Preferential Attachment Classifiers. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), Long Beach, CA, USA, 6–10 August 2023; pp. 105–116.
3. Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. The rise of social bots. *Commun. ACM* **2016**, *59*, 96–104. [CrossRef] [PubMed]
4. Varol, O.; Ferrara, E.; Davis, C.; Menczer, F.; Flammini, A. Online human-bot interactions: Detection, estimation, and characterization. In Proceedings of the 11th International Conference on Web and Social Media (ICWSM), Montreal, QC, Canada, 15–18 May 2017; pp. 280–289.
5. Yang, C.; Harkreader, R.; Gu, G. Empirical evaluation and new design for fighting evolving Twitter spammers. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1280–1293. [CrossRef]
6. Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; Tesconi, M. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion), Perth, Australia, 3–7 April 2017; pp. 963–972.
7. Kudugunta, S.; Ferrara, E. Deep neural networks for bot detection. *Inf. Sci.* **2018**, *467*, 312–322. [CrossRef]
8. Gilani, Z.; Kochmar, E.; Crowcroft, J. Classification of Twitter accounts into automated agents and human users. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Sydney, Australia, 31 July–3 August 2017; pp. 489–496.
9. Yu, H.; Kaminsky, M.; Gibbons, P.; Flaxman, A. SybilGuard: Defending against Sybil attacks via social networks. In Proceedings of the 2006 ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Pisa, Italy, 11–15 September 2006; pp. 267–278.
10. Yu, H.; Gibbons, P.; Kaminsky, M.; Xiao, F. SybilLimit: A near-optimal social network defense against Sybil attacks. In Proceedings of the 2008 IEEE Symposium on Security and Privacy (S&P), Oakland, CA, USA, 18–21 May 2008; pp. 3–17.
11. Cao, Q.; Sirivianos, M.; Yang, X.; Pregueiro, T. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI), San Jose, CA, USA, 25–27 April 2012; pp. 197–210.
12. Fire, M.; Kagan, D.; Elyashar, A.; Elovici, Y. Friend or foe? Fake profile identification in online social networks. *Soc. Netw. Anal. Min.* **2014**, *4*, 194. [CrossRef]
13. Stringhini, G.; Kruegel, C.; Vigna, G. Detecting spammers on social networks. In Proceedings of the 26th Annual Computer Security Applications Conference (ACSAC), Austin, TX, USA, 6–10 December 2010; pp. 1–9.
14. Hamilton, W.; Ying, Z.; Leskovec, J. Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.* **2017**, *40*, 52–74.
15. Hall, T.; Rasheed, K. A Survey of Machine Learning Methods for Time Series Prediction. *Appl. Sci.* **2025**, *15*, 5957. [CrossRef]

16. Hayashi, T.; Nakamura, A. Propagation graph estimation from individuals' time series of observed states. *Sci. Rep.* **2022**, *12*, 6078. [CrossRef] [PubMed]
17. Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef] [PubMed]
18. Bianconi, G.; Barabási, A.-L. Competition and multiscaling in evolving networks. *Europhys. Lett.* **2001**, *54*, 436–442. [CrossRef]
19. Leskovec, J.; Kleinberg, J.; Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data* **2007**, *1*, 2.
20. Qi, R.; Chen, S.; Li, W.; Zhang, C.; Lei, J.; Lv, H.; Sun, Y. Early Detection of Fake News via Structured Social Interaction Simulation and Hierarchical Cross-Modal Fusion. *Appl. Sci.* **2026**, *16*, 6001. [CrossRef]
21. Song, C.; Shu, K.; Wu, B. Temporally evolving graph neural network for fake news detection. *Inf. Process. Manag.* **2021**, *58*, 102712. [CrossRef]
22. Zhou, Q.; Lin, X.; Xu, L.; Sun, Y. Bi-directional temporal graph attention networks for rumor detection in online social networks. *Computing* **2025**, *107*, 45. [CrossRef]
23. Huang, Z.; Wang, Z.; Han, Z. Graph Neural Network Empowered Wireless Communications: Fundamentals, State-of-the-Art, Challenges, and Opportunities. *IEEE Wirel. Commun.* **2025**, *32*, 162–168. [CrossRef]
24. Kose, H.T.; Nunez-Yanez, J.; Piechocki, R.; Pope, J. A Survey of Computationally Efficient Graph Neural Networks for Reconfigurable Systems. *Information* **2024**, *15*, 377. [CrossRef]
25. Lin, C.-H.; Jian, J.-Y. Identifying Fake Accounts on Social Media Using Graph Neural Networks. In Proceedings of the 2024 IEEE 6th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), Tainan, Taiwan, 14–16 June 2024; pp. 387–389.
26. Lu, Y.; Li, Y.; Zhang, R.; Chen, W.; Ai, B.; Niyato, D. Graph Neural Networks for Wireless Networks: Graph Representation, Architecture and Evaluation. *IEEE Wirel. Commun.* **2025**, *32*, 150–156. [CrossRef]
27. Murgod, T.R.; Reddy, P.S.; Gaddam, S.; Sundaram, S.M.; Anitha, C. A Survey on Graph Neural Networks and its Applications in Various Domains. *SN Comput. Sci.* **2025**, *6*, 26. [CrossRef]
28. Sharma, K.; Lee, Y.-C.; Nambi, S.; Salian, A.; Shah, S.; Kim, S.-W.; Kumar, S. A Survey of Graph Neural Networks for Social Recommender Systems. *ACM Comput. Surv.* **2024**, *56*, 265. [CrossRef]
29. Zhang, X.; Li, S.; Li, H.; Gong, P.; Shi, J.; Shen, J.; Tian, C.; Zhang, D.; Zhu, Q. A Survey on Graph Neural Networks. *IEEE Access* **2026**, *14*, 27491–27512. [CrossRef]
30. Peláez-Vegas, A.; Aguilera-Martos, I.; Mesejo, P.; Luengo, J. A Survey on Deep Learning-Based Semi-Supervised Semantic Segmentation. *ACM Comput. Surv.* **2026**, *58*, 308. [CrossRef]
31. van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [CrossRef]
32. Yang, X.; Song, Z.; King, I.; Xu, Z. A Survey on Deep Semi-Supervised Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 8934–8954. [CrossRef]
33. Gao, X.; Xie, D.; Zhang, Y. A comprehensive survey on imbalanced data learning. *Front. Comput. Sci.* **2026**, *20*, 2011622. [CrossRef]
34. Ma, Y.; Tian, Y.; Moniz, N.; Chawla, N. Class-Imbalanced Learning on Graphs: A Survey. *ACM Comput. Surv.* **2025**, *57*, 1–16. [CrossRef]
35. Ormeño-Arriagada, P.; Puraivan, E.; Kloss, S.; Cofré-Morales, C.; Rodriguez, M. Interpretable Fake News Detection Using Linguistic Indicators Under Imbalanced and Low-Resource Conditions. *Appl. Sci.* **2026**, *16*, 5080. [CrossRef]
36. Zhao, L.; Han, F.; Ling, Q.; Han, H.; Liu, W. A Survey on Class Imbalance Learning Algorithms in Complex Scenarios. *IEEE Access* **2025**, *13*, 180799–180833. [CrossRef]
37. Lomax, S.; Vadera, S. A survey of cost-sensitive decision tree induction algorithms. *ACM Comput. Surv.* **2013**, *45*, 1–35. [CrossRef]
38. Safi, S.A.-D.; Castillo, P.A.; Faris, H. Cost-Sensitive Metaheuristic Optimization-Based Neural Network with Ensemble Learning for Financial Distress Prediction. *Appl. Sci.* **2022**, *12*, 6918. [CrossRef]
39. Leskovec, J.; Suri, S. SnapGen: Benchmarking Synthetic Graph Generators for Graph ML. In Proceedings of the NeurIPS 2022 Workshop on Synthetic Data for Machine Learning, New Orleans, LA, USA, 2 December 2022.
40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
41. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
42. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
43. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
44. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
45. Gulati, A.; Kumar, E.; Agarwal, V.; Sharma, A. Weak Links in LinkedIn: Enhancing Fake Profile Detection in the Age of LLMs. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Niagara Falls, ON, Canada, 25–28 August 2025; pp. 372–382.

46. Kalamakis, I.; Liappi, S.; Papakostas, D.; Katsaros, D. Language Models for Disinformation Detection: To be Large or to be Small? In Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 19–22 October 2025.
47. Fragkou, E.; Chini, E.; Papadopoulou, M.; Papakostas, D.; Katsaros, D.; Dustdar, S. Distributed federated deep learning in clustered Internet of Things wireless networks with data similarity-based client participation. *IEEE Internet Comput. Mag.* **2024**, *28*, 53–61. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.