TechRank: A Graph-Based Framework for Identifying High-Growth Emerging Tech Companies

Tomas Garate Anderegg, supervised by Julian Jang-Jaccard and Andrea Cavallaro

Background and Motivation

Technological innovation evolves rapidly, moving from research labs to commercial deployment, with major implications for national security and cyber defense. Early identification of disruptive technologies is crucial for proactive strategies.

This project builds on the TechRank algorithm, modeling the ecosystem as a bipartite network of companies and technologies to quantify mutual influence. By adapting TechRank for national cyber defense forecasting, the project faces three main challenges:

1. **Rapid Evolution:** Emerging technologies evolve faster than traditional monitoring, risking reactive strategies.
2. **Dual-Use Ambiguity:** Technologies can create both opportunities and threats, requiring careful assessment.
3. **Predictive Complexity:** Forecasting disruptive technologies involves non-linear, noisy data, and integrating Temporal Graph Networks improves prediction but challenges interpretability.

To address these, TechRank is extended with temporal graph modeling and predictive analytics, transforming it into a tool for forecasting disruptive technologies and supporting early-warning cybersecurity systems.

Related Work

TechRank [4] is a recursive, network-based algorithm inspired by PageRank, designed to help investors navigate complex technological ecosystems. By ranking companies and technologies in a bipartite network, it uncovers mutual influence patterns and provides a data-driven framework to identify high-potential opportunities in fast-evolving technology landscapes.

Temporal Graph Networks for Deep Learning on Dynamic Graphs [6] Most Graph Neural Networks (GNNs) operate on static graphs, but real-world networks are often dynamic with nodes and edges appearing, disappearing or changing features over time. Temporal Graph Networks (TGN) are a framework designed specifically for deep learning on continuous-time dynamic graphs, which are represented as a sequence of timed events (e.g. a user liking a post at a specific time)

Dataset

We evaluate our approach using **Crunchbase** [1], a platform for private and public company data, which provides:

- **Comprehensive firmographic and financial data** on millions of companies, from early-stage startups to the Fortune 1000.
- **A multifaceted data collection approach** combining crowd-source, machine learning and third-party aggregation ensuring comprehensive and timely updates
- **A focus on innovative and technology oriented companies** making it a premier source for data on the tech and startup ecosystem, widely used both the venture capital industry and academic researchers. [2, 5]
- **Access via well-documented API or Daily CSV exports** allowing for integration into internal systems, workflows and model training pipelines.

Methodology and Implementation

Model Architecture for TechRank

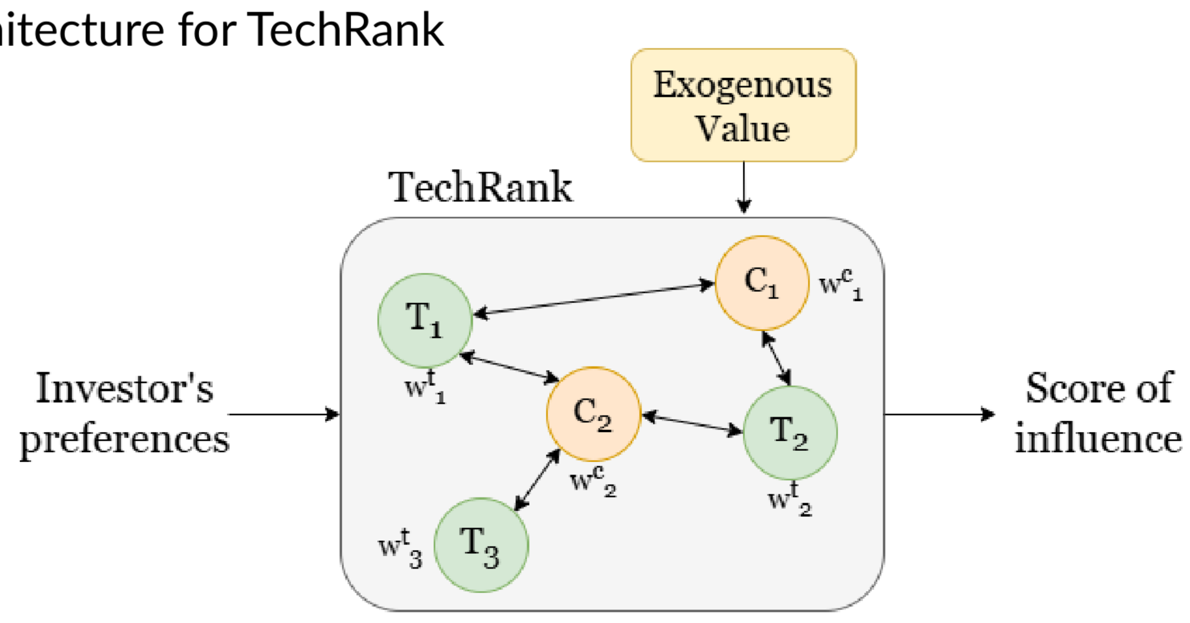

Figure 1. Framework of the TechRank algorithm.

Construction of the bipartite graph

In graph theory, an **adjacency matrix** represents all connections in a graph. For TechRank, this matrix models the ecosystem of companies and technologies.

$$w_c^{(0)} = \sum_{t=1}^{n_t} M_{c,t}^{CT}, \quad w_t^{(0)} = \sum_{c=1}^{n_c} M_{c,t}^{CT} \tag{1}$$

$$w_c^{(n+1)} = \sum_{t=1}^{n_t} G_{c,t}(\beta)\, w_t^{(n)}, \quad w_t^{(n+1)} = \sum_{c=1}^{n_c} G_{t,c}(\alpha)\, w_c^{(n)} \tag{2}$$

where $M_{c,t}^{CT} = 1$ if company c uses technology t (0 otherwise). $w_c^{(0)}$ and $w_t^{(0)}$ are the initial scores. The exogenous variable incorporates external context, improving TechRank's predictive accuracy.

Recursive Scoring (Reflexion Method)

The algorithm iteratively updates company and technology scores based on neighbors (Eq. (1) and Eq.(2)). Transition probabilities $G_{c,t}(\beta)$ and $G_{t,c}(\alpha)$ control the influence between nodes.

Model Architecture for Temporal Graph Networks (TGN)

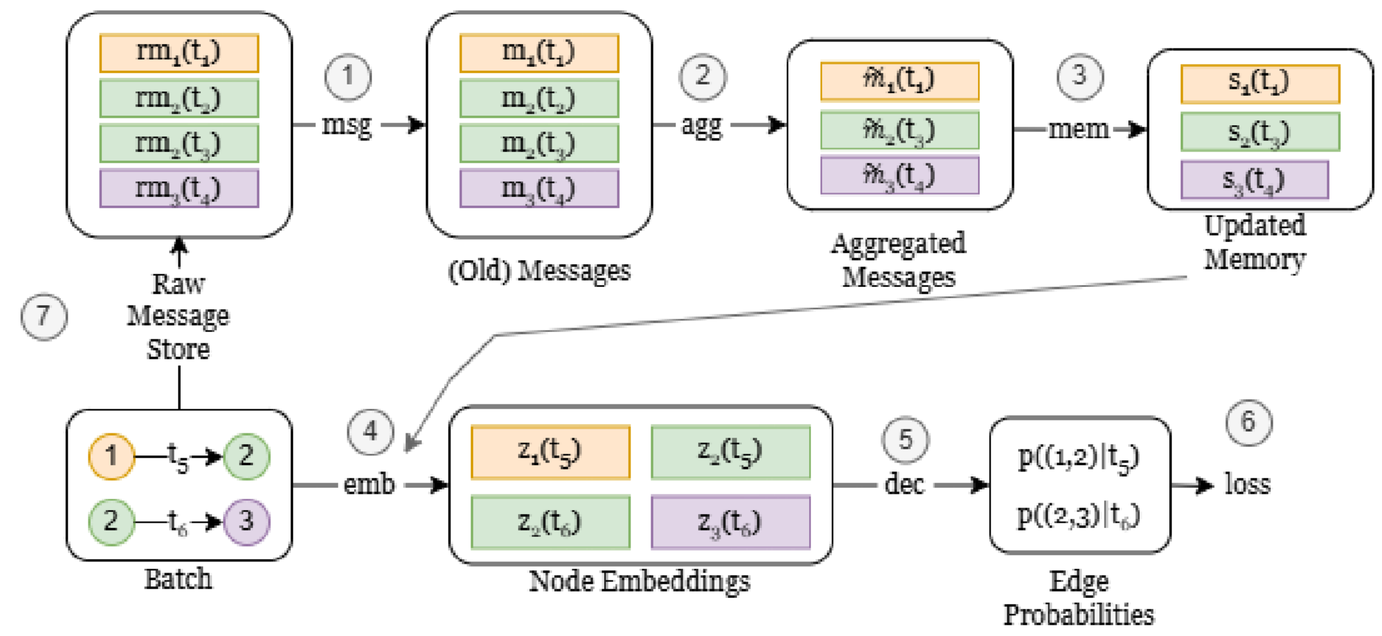

Figure 2. Framework of the TGN algorithm.

TGN models and predicts how relationships evolve over time by combining a *memory module*, capturing each node's historical state, with *graph-based operators* that model node interactions.

In this project, TGN transforms TechRank into a **predictive framework**, forecasting technologies likely to become disruptive.

Why TGN?

- **Temporal Forecasting:** Handles continuous event streams to anticipate shifts in the cyber-technology landscape.
- **Long-Term Dependency:** Memory mechanisms capture temporal context, improving predictions of emerging technologies.

Project Framework

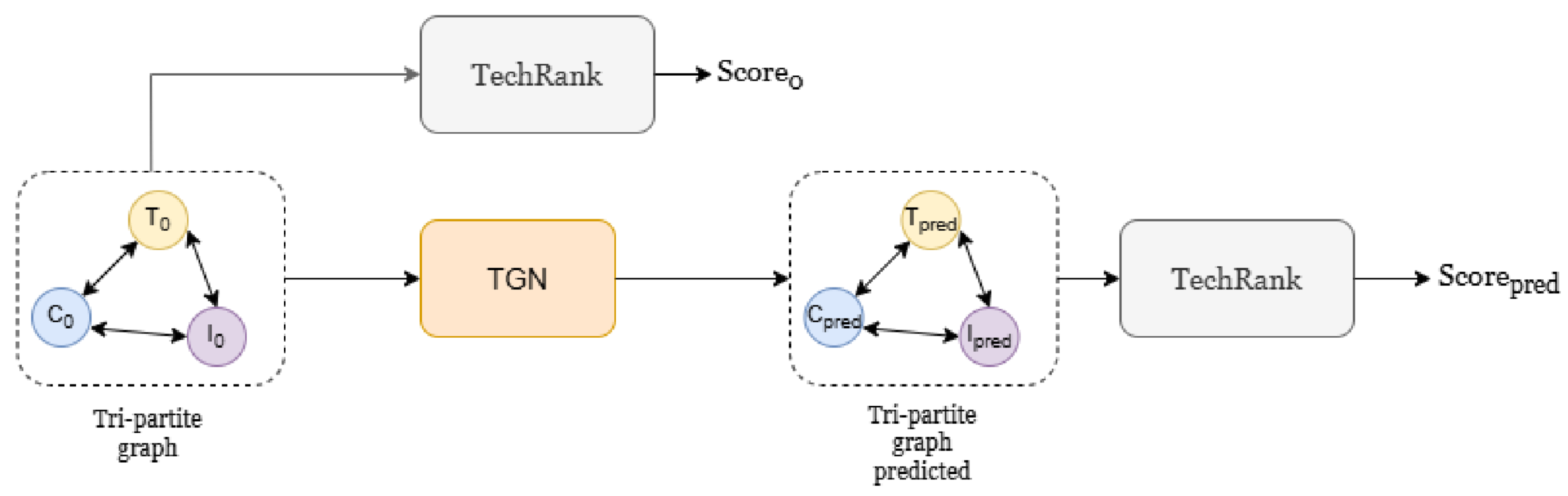

Figure 3. Framework of the project's algorithm.

This project aims to build a tripartite graph of Companies (C), Technologies (T), and Investments (I). A TechRank, adapted to 3 entities, score will first be computed, and the graph will then serve as input to the Temporal Graph Network (TGN), which predicts a new graph with potential new nodes or edges. Recalculating TechRank on the updated graph will allow us to assess technological disruptiveness, ideally using a scale to quantify its degree.

Limitations

Several limitations should be noted:

- The Temporal Graph Network (TGN) produces probabilistic outputs, so predicted links are not guaranteed.
- Recomputing TechRank on an uncertain graph may cause instability if uncertainties are not properly handled.
- For forecasting labels like technological disruption, using TGN's dynamic embeddings, possibly combined with initial TechRank scores, is often sufficient.
- The role of TechRank in the pipeline can be re-evaluated to determine the configuration that yields the best results.

Conclusion

This project presents a methodology to model interactions between Companies, Technologies, and Investments using a tripartite graph. By combining TechRank with the predictive capabilities of the Temporal Graph Network (TGN), we aim to identify and quantify potentially disruptive technologies. Despite probabilistic uncertainties, the approach offers a flexible framework for analyzing technological evolution. Future work will refine the integration of TechRank and TGN, explore alternative scoring strategies, and validate the model against real-world forecasts.

References

- Crunchbase. https://www.crunchbase.com. Accessed: 18-Oct-2025.
- Jean-Michel Dalle, Matthijs den Besten, and Carlo Menon. Using crunchbase for economic and managerial research. https://www.crunchbase.com/, October 31 2017. Preprint.
- Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19)*, pages 1–10, Anchorage, AK, USA, 2019. ACM.
- Anita Mezzetti, Loïc Maréchal, Dimitri Percia David, William Lacube, Sébastien Gillard, Michael Tsesmelis, Thomas Maillart, and Alain Mermoud. Techrank: Modelling portfolios of cyber-related emerging technologies. *arXiv preprint arXiv:2210.07824*, 2022.
- Mark Potanin, Andrey Chertok, Konstantin Zorin, and Cyril Shtabtsovsky. Startup success prediction and vc portfolio simulation using crunchbase data. https://www.crunchbase.com/, 2023. Preprint, September 28, 2023.