When Does Relational Context Help Predict Technology Entry? Evidence from U.S. Metropolitan Areas

Anonymous authors
Paper under double-blind review

Abstract

How does the inclusion of relational information improve predictions made using history? We use U.S. metropolitan technology entry during 2000–2019 to construct a panel of 281,236 observations of 100 Core-Based Statistical Areas integrating 541 CPC technology subclasses over the 10-year period. We investigate if relations with other cities forecast technology entry. We test if the benefit depends on how new the technology is relative to the city's technology level. We use citation datasets to improve our forecast and measure the benefit against a baseline using the history and location of the focal city. The benefit is strongly dependent on technology novelty. For the most novel technology, the benefit is 0.016, and for the most familiar it is -0.002, for a difference of 0.018 (95% CI [0.002, 0.036]). The same effect is observed when using collaborations between inventors and when the same inventors work for the same company. As the collaborations and company affiliations are similar—pairwise $r = 0.81$–0.97—we give greater weight to the results of the citation study and consider the effects of the collaborations and company affiliations as correlated views of a common diffusion signal rather than independent replications. Citation exposure, like geography, predicts the outcome of interest. We consider the effect of relations, the intensity of relations, and other factors. Relational networks can be randomized, but **the observed randomization achieves only Jaccard 0.645**, so strong relations and a "small world" effect constrain what can be learned. A graph neural network achieves forecasting accuracy of 0.6659. Conditional on technology novelty, we find **the network captures relational information but the gradient is not significant**; forecasts do not significantly vary from the baseline. The novelty of the technology may also condition the forecasts.

The results jointly demonstrate that relational context enhances predictions primarily for technologically recent cases of competitive entry.

1 Introduction

A common situation in machine learning is to forecast the future of an entity based on the factors that influenced its past. However, in the real world, entities are not isolated, and often the future of an entity is influenced by other entities. For example, the future of a product offered by a firm depends on the products supplied by its upstream firms. The future of a person's research interest depends on the research interests of his/her co-authors, etc.

When dealing with isolated entities, prediction models are developed based on the assumptions that the future of an entity is primarily influenced by the past of that entity. When dealing with interconnected entities, prediction models should be developed based on the assumptions that the future of an entity may be largely influenced by the past of related entities.

Here, we are interested in understanding the prediction of future states of an entity based on contextual relations, and if such prediction varies by the state being predicted. That is, are there predictors outside the entity being studied and its proximate environment that provide information about the state being forecast?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Finally, we expect that prediction using contextual relations should be more pertinent when the future state being forecast is dissimilar to the current state of the entity being studied.

A relational feature may be predictive while reproducing information already contained in the entity's own history or geographic environment. We therefore ask whether relational context provides incremental predictive information after those local and geographic signals have been accounted for, and whether that incremental information varies with the state being predicted.

Substantive motivation. When thinking about patents, "familiar" targets are CPC subclasses connected to previously patented technology by the entity in question, while "novel" targets are subclasses not connected to previously patented technology. Novel target prediction refers to the case in which an entity moves into a technology field different from the technology fields which it is traditionally active in. This situation occurs when a city, for example, traditionally active in electronics, starts patenting in the field of optoelectronics. This is the situation of interest in the existing policy research, and it is potentially the most relevant economically.

Relational Context. The term relational context is used for the general idea, relational exposure is used for the case which is of interest in the paper, and the term topology of the network is used to refer to the type of graph representing the relational connections. Relational exposure is constructed by the authors; therefore, the paper demonstrates the case for relational context. It does not, however, demonstrate that topology of the network is the case for relational context and gain in the explored field.

Rational Group Model. The authors provide a model to demonstrate the case for conditional gain. In the model, an agent (c) at time (t) decides to enter a target state (k). The agent believes that it can gain access to state (k), if state (k) is accessible. The agent uses its own experience, as well as the experience of its neighbors to form a belief about the probability of success. According to the authors, relational signals can be used to predict the outcome of actions of an agent.

Even in its present form, the model suggests that, other things being equal, more relational information should be required to forecast less familiar targets.

1.1 CONTRIBUTIONS

A tool that allows us to make time-based predictions using relational data. We define technology entry as a particular case of a city-technology prediction problem where the features available at the prediction origin are restricted to what is observed up to that time. In other words, we restrict the predictors to what is available within the city and in the technology. This restriction provides a context where we can study whether relational data lead to better predictions as compared to the history of the focal entity.

A behavior that holds only under certain conditions. Relational information becomes predictively more relevant as the technology target becomes less related to the city's existing technology. Among all city samples, the difference between the most related and least related technology quartiles is statistically different from zero based on the city-clustered bootstrap and permutation tests.

Consistency across various relational data. The novelty-based prediction patterns are general and hold for other types of relational data (i.e., patent citations, inventor collaborations, and corporate affiliations).

Strong correlations (0.81–0.97) between the different networks allow us to consider the result of the citation-channel network as the main case, and consider the other two networks as case studies of correlated community.

The overall relational signal. Decomposing the citation channel network, we found that removing edge weights resulted in a loss of predictive power. This suggests that the intensive (weighted) relationship between the entities is the main reason for the improvement in predictive power. As our randomization failed to achieve sufficient structural (topological) difference among the entities, we will consider the case as open.

Predictive power. A learned, graph-based representation of the data resulted in a better prediction compared to other, non-graph based, representation. Thus, a temporal graph neural network improved prediction over other, simpler, baselines. We consider the case as further evidence for

the graph neural network in improving prediction, and consider the conditional (and/or unobserved) effects as neutral.

2 RELATED WORK

Relevance of graph structure. Different works assess the contribution of graph structure by fixing the supervised learning problem and comparing graph neural networks with non-graph neural networks. Luan et al. (2022) define conditions under which graph neural networks outperform non-graph neural networks, and show that in some cases MLPs outperform graph neural networks. We study under which conditions representation and integration of relational information improve prediction, and under what conditions such information remains uninformative.

Relatedness and technology entry. Research shows regions tend to adopt new technologies that are adjacent to or related to the technologies in their portfolio. Product spaces, introduced in Hidalgo and Hausmann (2009), are a means to represent products in an economic space by relations of co-exporting goods and services. The authors showed that regions' position in this space is related to the regions' potential for economic diversification. Relatedness measures in Evolutionary Economic Geography and related fields (e.g., Boschma and Frenken, 2011) share conceptual similarities with measures of product spaces. The main contribution of this research is to treat relatedness measures as a moderator rather than a variable to be explained.

Forecasting on graphs. Recent works in the field of forecasting on graphs focused on improving architectures for modeling temporal and relational dependencies in complex systems (Xu et al., 2020; Rossi et al., 2020). Usually, the assumption is made that the temporal and relational dependencies are rich enough to allow accurate models to be developed. In this work, rather than proposing a new architecture, we assess relational structures from a complementary perspective.

Citations as knowledge transfers. Prior work used patent citations to represent knowledge transfers. Jaffe et al. (1993) documented geographic localization in patent citations, providing evidence consistent with geographically localized knowledge spillovers. In this research we use the same type of representation and define citation-related predictors to include re-citations, as well as first citations.

3 PROBLEM FORMULATION

Let c index a U.S. Core-Based Statistical Area, k index a CPC technology subclass, and t index a filing year. Let $S_{c,k,t} \in \{0, 1\}$ denote whether CBSA c produced at least one patent in CPC subclass k in year t.

For each (c, k, t) we predict:

$$Y_{c,k,t}^{(h)} = \mathbb{1}\left[S_{c,k,t} = 0 \ \wedge \ \sum_{\tau=t+1}^{t+h} S_{c,k,\tau} > 0\right] \tag{1}$$

with $h = 3$. Prediction cells are constructed only for (c, k, t) where the CBSA has no patent in subclass k in year t. This "absent-at-t" definition includes re-entries, not only first-ever entries.

Although the time period for which we are permitted to make predictions is up to and including 2019, we are free to construct the elements of the predictions up to and including 2022. Thus, the time period for which we may construct elements for predictions is three years beyond the time period for which we may make predictions.

Every feature is computable from information available at year t. Citations are restricted to $\leq t$. The similarity matrix is calculated from 2000–2009 co-occurrences. Novelty is defined using the CBSA's information as of t. The city choice rule is set to the pre-period and is fixed for the study.

3.1 RELATIONAL EXPOSURE

For each prediction cell (c, k, t) and each relational network, we define an exposure feature as the network-weighted average of presence across connected CBSAs. Let $w_{cj,t}$ denote the weight of

Table 1: Features used in the prediction models.

Class	Features
Local history	`local_patents_5y`, `local_recency`
CBSA aggregate	`cbsa_total_patents_5y`, `cbsa_peak_annual_tech_count_5y`
Geographic	`geo_exposure`
Relational (citation)	`network_exposure`
Relational (co-inventor)	`co_inventor_exposure`
Relational (shared-assignee)	`shared_assignee_exposure`
Technology	`novelty`

the edge from CBSA c to CBSA j in the network aggregated through year t, and let $S_{j,k,t}$ denote presence of technology k in CBSA j at time t. Relational exposure is:

$$E_{c,k,t} = \frac{\sum_j w_{cj,t} \cdot S_{j,k,t}}{\sum_j w_{cj,t} + \epsilon} \tag{2}$$

where ϵ is a small smoothing constant. $E_{c,k,t}$ is a weighted average of neighbors' presence in the target technology, with weights proportional to the strength of the relational tie.

The citation exposure uses citation-network edge weights; the co-inventor exposure uses fractional co-inventor edge weights; the shared-assignee exposure uses fractional shared-assignee edge weights.

Unit-weight variant. For the decomposition analysis in Section 6.6, we construct a unit-weighted alternative in which every positive edge receives weight one:

$$E^{\text{unit}}_{c,k,t} = \frac{\sum_j \mathbb{1}(w_{cj,t} > 0) \cdot S_{j,k,t}}{\sum_j \mathbb{1}(w_{cj,t} > 0) + \epsilon} \tag{3}$$

The unit-weight variant preserves the observed topology (which CBSAs are connected to c) but discards edge multiplicity. Comparing the predictive performance of E and E^{unit} therefore isolates the contribution of weighted relational intensity from the contribution of topological connectivity.

3.2 Novelty

Novelty is defined in two steps.

First, we compute a technology similarity matrix from pre-period (2000–2009) co-occurrence across all CBSAs. For each pair of CPC subclasses (k, l), let $S(k, l)$ denote the cosine similarity of their pre-period CBSA presence vectors. For each target CPC k, we then compute the percentile rank $\rho(k, l) \in [0, 1]$ of $S(k, l)$ within the distribution $\{S(k, l') : l' \in \mathcal{K}\}$, where $\mathcal{K}$ is the set of 541 CPC subclasses. Lower values of ρ indicate that l is more similar to k; higher values indicate that l is more dissimilar.

Second, for each prediction cell (c, k, t), let $\mathcal{P}_{c,t}$ denote the set of CPC subclasses in which CBSA c has produced at least one patent in any year up to and including t. Novelty is defined as the mean percentile rank of the target k across the portfolio:

$$\text{Novelty}_{c,k,t} = \frac{1}{|\mathcal{P}_{c,t}|} \sum_{l \in \mathcal{P}_{c,t}} \rho(k, l) \tag{4}$$

Higher values indicate that the target is, on average, less similar to technologies previously present in the CBSA's portfolio. In the empirical sample, novelty ranges from 0.20 to 0.64.

Design of the measure. The similarity matrix is deliberately fixed at 2000–2009, while the portfolio against which each target is evaluated expands through the prediction year t. A fixed pre-period matrix is the conservative choice: it ensures that the technology-relatedness structure used to define novelty cannot encode any information from the prediction window.

We considered several formulations of technological novelty, including $1 - \max_l S(k,l)$, $\min_l \rho(k,l)$, and the mean percentile rank used here. The mean-percentile formulation preserves variation across prediction cells and avoids the compression and degeneracy observed in the alternative formulations. We use this measure as the primary operationalization of novelty and treat the stratified analysis as conditional on this measurement choice.

Novelty is not used as a predictor in the paper's main models; it is used to stratify prediction cells into quartiles for evaluating the heterogeneity of the relational-feature contribution.

Evaluation. Average precision (AP) is the primary metric. Because the entry rate varies substantially across novelty quartiles (47.8% in Q1, 13.6% in Q4), AP's baseline varies across strata. This is a feature rather than a defect for our purposes: AP rewards improvements over the stratum-specific baseline, and each quartile is evaluated against its own base rate. We also report ROC-AUC and Recall@K in the replication package.

Prediction cells are split by year into training (2010–2016), validation (2017–2018), and test (2019). Novelty quartile boundaries are computed on the training set.

4 DATA

Patents. PatentsView 2024 release (Zenodo DOI 10.5281/zenodo.15783125). Granted patents from 1976 onward with disambiguated inventors, inventor locations, CPC classifications, assignees, and U.S. patent citations. We restrict to utility patents with U.S. inventor locations and valid CPC subclass assignments. Filing year is the temporal index.

Geographic aggregation. Inventor locations are aggregated to U.S. Core-Based Statistical Areas via the Urban Institute CBSA crosswalk (OMB July-2023 delineation). Municipalities are assigned to CBSAs through their county FIPS codes. The sample contains 100 CBSAs selected by total patent weight during 2000–2009. This selection rule ensures that the sample is fixed before the prediction window, but it also means the results speak to established innovation hubs. Extrapolation to smaller or emerging metro areas is not tested and is left as open.

Technology unit. CPC subclasses with at least 100 pre-period fractional patents: 541 subclasses.

Three relational networks. The patent-citation network aggregates city-to-city citations to the CBSA level. The co-inventor network connects CBSAs whose inventors appear together on the same patent, with fractional weight $1/\binom{n}{2}$ per patent where n is the number of distinct CBSAs. The shared-assignee network connects CBSAs when the same organizational assignee has patents from both—companies and government entities, excluding individual assignees—with the same fractional weighting.

5 METHOD

Models. Gradient-boosted trees (HistGradientBoostingClassifier, 300 iterations, learning rate 0.05, 31 leaves, early stopping on a 10% internal split) over feature sets H, H+GEO, H+NET, H+GEO+NET, and combinations with CBSA aggregate features and the three relational channels. Logistic regression with and without class weighting is reported as a robustness check.

Residualization. Each network feature is residualized on `geo_exposure` using a train-only linear regression, producing `*_exposure_resid`. The residual is orthogonal to geography in the training set by construction.

Stratified evaluation. We evaluate models on the full test sample and within each quartile of novelty and geographic exposure. ΔAP = AP(with relational features) − AP(without). Quartile boundaries are computed on training data only.

Formal test of the novelty gradient. City-clustered bootstrap of the difference ΔAP(Q4) − ΔAP(Q1), resampling CBSAs rather than prediction cells. Permutation test shuffling novelty labels within CBSA (500 permutations). With 100 clusters, block resampling is at the smaller end of the range where bootstrap coverage is reliable.

Multiple correlated channels. The similarity of correlations leads us to treat the results from the citation channel as the primary specification and the other two channels as supplemental, but similar, evidence. The effective number of supplemental channels is less than three. The presence of the effect in other channels supports the results are not specific to the legal citation channel, but does not provide three independent replications of the same outcome.

Decomposition of the relational signal. The unit-weighted variant E^{unit} (Section 3.1) is applied to all three relational channels. A directed degree-preserving edge-swap randomization tests whether the observed topology contributes beyond degree.

Graph neural network. A heterogeneous GNN with two rounds of message passing over the sparsified relational graph (top-15 outgoing edges per CBSA) and learned 32-dimensional embeddings for CBSAs and CPC subclasses. For each prediction year, the relational graph is constructed using observations available through that year, with the five-year rolling window ending at t.

6 RESULTS

6.1 MODEL COMPARISON

Table 2: Model comparison on the test set (2019).

Model	Features	GB test AP
H	2	0.6051
H+CBSA	4	0.6337
H+GEO	3	0.6428
H+NET	3	0.6412
H+GEO+NET	4	0.6486
H+GEO+NETresid	4	0.6503
H+GEO+CBSA	5	0.6463
H+GEO+CBSA+NET	6	0.6609
H+GEO+ALL3	6	0.6529
H+GEO+ALL3resid	6	0.6524
GNN (full)	—	0.6659
GNN (local head only)	—	0.6286

The strongest hand-crafted configuration is `H+GEO+CBSA+NET` at 0.6609. Adding the co-inventor and shared-assignee networks does not improve performance; `H+GEO+ALL3` achieves 0.6529, consistent with the three networks being largely redundant. The GNN achieves test AP 0.6659, improving on the strongest hand-crafted configuration by $+0.0050$.

Practical significance. The pooled citation-network increment of $+0.006$ AP translates, at a fixed false-positive budget of 20% of predicted positives, to approximately 3 additional true entries retained per 1,000 test cells. At the Q4 novelty stratum, the citation increment of $+0.016$ AP corresponds to approximately 8 additional true entries retained per 1,000 cells.

6.2 THE NOVELTY GRADIENT

Table 3: Incremental predictive gain from relational exposure over H+GEO, by novelty quartile, for each channel.

Quartile	n	Citation	Co-inventor	Shared-assignee
Q1 (familiar)	5,906	-0.002	-0.001	-0.006
Q2	6,882	$+0.013$	$+0.003$	$+0.005$
Q3	7,105	$+0.005$	$+0.002$	$+0.004$
Q4 (novel)	6,800	$+0.016$	$+0.019$	$+0.019$

All three channels show the same shape: near-zero or negative in Q1, positive in Q4. The entry rate decreases monotonically with novelty across the pooled test set, from 47.8% in Q1 to 13.6% in Q4.

6.3 JOINT (NOVELTY × GEOGRAPHY) STRUCTURE

Table 4: Incremental predictive gain from citation-network exposure over H+GEO, by joint novelty × geographic-exposure cell.

	geo Q1	geo Q2	geo Q3	geo Q4
nov Q1	-0.039	-0.010	$+0.006$	$+0.005$
nov Q2	-0.006	$+0.018$	$+0.014$	$+0.019$
nov Q3	$+0.000$	$+0.009$	$+0.003$	$+0.003$
nov Q4	$+0.012$	$+0.023$	$+0.023$	$+0.020$

The gain is concentrated in the Q4 novelty row across all four geography columns.

6.4 CROSS-CHANNEL ROBUSTNESS

Table 5: Pairwise correlations between the three relational networks.

Pair	Pearson r
shared-assignee vs co-inventor	0.97
shared-assignee vs citation	0.87
citation vs co-inventor	0.81

The similarity of correlations leads us to treat the results from the citation channel as the primary specification and the other two channels as supplemental, but similar, evidence. The effective number of supplemental channels is less than three. The presence of the effect in other channels supports the results are not specific to the legal citation channel, but does not provide three independent replications of the same outcome.

6.5 FORMAL TEST OF THE NOVELTY GRADIENT

Table 6: Formal test of the novelty gradient. $\Delta = \Delta\text{AP(Q4)} - \Delta\text{AP(Q1)}$.

Network	Δ	95% CI	permutation p
Citation (primary)	$+0.0180$	$[+0.0015, +0.0357]$	< 0.001
Co-inventor (supplemental)	$+0.0196$	$[+0.0073, +0.0326]$	< 0.001
Shared-assignee (supplemental)	$+0.0248$	$[+0.0078, +0.0446]$	< 0.001

Three estimates, based on resampling and using the methods described in this chapter, show that the Q4–Q1 difference in predicted value is statistically significant. We do not present these as three independent validation cases; see Section 6.4.

6.6 DECOMPOSITION OF THE RELATIONAL SIGNAL

In all three channels, the majority of the predictive signal was derived from relational structure, not geographic location. The unit-weight ablation holds true across all three of the correlated relational representations.

Degree-preserving randomization. After a directed edge-swapping randomization performed 100 times each, the observed configuration yields Jaccard overlap of 0.645 and a coefficient ratio of

Table 7: Unit-weight ablation across the three relational channels.

Channel	Weighted	Unit-weighted	Δ (weighted $-$ unit)
Citation	0.6486	0.6422	$+0.0064$
Co-inventor	0.6466	0.6431	$+0.0035$
Shared-assignee	0.6456	0.6429	$+0.0027$

0.949. For the purposes of this analysis, we consider the random edge swaps to be a weak null hypothesis. Therefore, the topology vs. intensity trade-off is considered undecided. A more stringent null hypothesis, e.g. configuration model randomization that more extensively rewires edges, or a randomization that breaks city clustering, is beyond the scope of this research.

6.7 Graph neural network

We investigate the efficacy of incorporating graph representation learning to improve prediction. The best configuration of the GNN outperformed the best hand-crafted configurations. The GNN achieved an absolute improvement of 0.005 in average precision (AP) in the test set, achieving an AP of 0.6659. Compared to predictions made by a local attention head, the improvement in the novelty quartile (Q4: 0.6692, Q1: 0.6350) was greater than the improvement in the familiar quartile (Q4: 0.6648, Q1: 0.6350). The city clustered bootstrap did not find a significant difference from zero ($\Delta = 0.007$, 95% CI: $[-0.015, 0.028]$). Because the gate does not exhibit monotonic behavior, differences in prediction are not due to a learned monotonic novelty representation. Thus, the GNN predominantly captures relational representations.

This shows that using a learned relational representation improves prediction, and is not direct evidence for the conditional mechanism.

7 Discussion

In the case study considered here, relational data provides added predictive value for targeted entries in the considered time windows. For already familiar (or old) targets in the local window, the predictive value of relational data is low. For new targets, the value is high. The observed pattern is in line with the signal combination hypothesis.

The same pattern holds for the other two relational networks, and hence provides some level of confidence in the result.

Intensity of relational data, and not particular relational structures, explains the bulk of the pooled signal for the case study. For the case study sample, the main effect of relational data is due to a CBSA's relations with other CBSAs that are downstream in relational flow to city-level units that contain the technology. It is not known if other structures beyond relational intensity explain the signal, especially if the sample includes more distantly related case studies. This is an open research question.

7.1 Limitations

The effect of citation network data is small, approximately equal to the effect of CBSA level aggregated data and geography.

We study the predictive value of relational context and find that in the case of the U.S. metropolitan technology citation network, exposure to similar technologies in other cities can enhance the likelihood of future technology entry. This is especially the case for novel technologies. Relational networks are correlated between 0.81 and 0.97. Robustness is more comparable between related views than between independent tests. Degree-preserving edge swaps do not isolate topology-specific components in this case, and thus, we cannot distinguish topology-specific components from relational-intensity components for the case of the Q4 novelty effect.

The GNN's inability to learn the conditional structures results in a novelty gradient that is not different from zero, city-clustered bootstrap confidence. The GNN provides a stronger prediction confidence, but not a stronger evidence confidence, for the case of the conditional novelty effect.

The main study case (2019) is documented in the main body. Other study years (2017 and 2018) are evaluated under a rolling-origin scheme. The qualitative case holds for all years studied, but individual years exhibit larger confidence intervals. The pooled estimate represents a stronger effect.

The 100 CBSA sample represents large and established innovation centers. Other, smaller and/or developing metro areas are not represented.

We aggregate at the CBSA level. Cross-municipality collaborations within a CBSA do not generate edges.

8 Conclusion

The difference between the most innovative and least innovative quartiles is statistically significant and the pattern holds for citations, collaborations and co-assignees.

The citation-network shows that information related to the relative strength of ties improves the prediction of outcomes and information related to absolute strength of ties does not. This is true for all three relationships, thus it can be concluded that a collection of weighted relationships accounts for the overall improvement in prediction. In the case of network randomization, it is not possible to fully decouple topology from other aspects of the structure, and thus it is not possible to fully determine if citing a particular entity adds predictive value beyond a weighted relationship. Therefore, we also consider the case to be open.

Relational information often improves the prediction of the introduction of new technology. In this case, predicting the introduction of technologies most related to existing technologies does not benefit from the inclusion of relational information. Relational information improves prediction when the technology most closely related to existing technologies is not the most similar to the technology being predicted.

The broader result is therefore not that relational information is uniformly useful for temporal prediction. Rather, its incremental predictive value depends on the state being predicted. The result suggests that relational context should be evaluated not only by whether it improves aggregate predictive performance, but also by whether its information value varies systematically across the states an entity is attempting to reach.

AI use statement

On this work, we used generative AI tools for prose drafting. We have not used generative AI tools for data collection or the design of the study.

Ethics statement

This work uses public patent records and public economic data. No human subjects were involved. The paper studies technology diffusion patterns that are matters of public record. We do not make causal claims, and we do not recommend policy. No identifying information is used.

Reproducibility statement

Data sources are the PatentsView 2024 release (DOI 10.5281/zenodo.15783125) and the Urban Institute CBSA crosswalk (OMB July-2023 delineation). Prediction cells are constructed with strict temporal leakage controls: features at prediction origin t use only information available through t, and the technology similarity matrix used to define novelty is computed from a fixed 2000–2009 window.

REFERENCES

Boschma, R. and Frenken, K. Technological relatedness and regional branching. In Bathelt, H., Feldman, M. P., and Kogler, D. F. (eds.), *Beyond Territory: Dynamic Geographies of Knowledge Creation, Diffusion, and Innovation*, pp. 64–81. Routledge, 2011.

Hidalgo, C. A. and Hausmann, R. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575, 2009.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598, 1993.

Luan, S., Hua, C., Lu, Q., Zhu, J., Chang, X., and Precup, D. When do we need graph neural networks for node classification? *arXiv preprint arXiv:2210.16979*, 2022.

Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., and Bronstein, M. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2020.

Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., and Achan, K. Inductive representation learning on temporal graphs. In *International Conference on Learning Representations (ICLR)*, 2020.