



Università
Ca' Foscari
Venezia

Corso di Dottorato di Ricerca
in Economia

XVIII Ciclo

Tesi di Ricerca

Essays on Political Economy: Elections, Conflict & Organized Crime

SSD: SECS-P/01, SECS-P/03

Coordinatore Dottorato

prof. Pietro Dindo

Supervisors

prof. Marco Di Cataldo, prof. Agar Brugiavini

Dottorando

Adriano Amati

Matricola 956718

Acknowledgements

My sincere gratitude to my supervisors, Marco Di Cataldo and Agar Brugiavini, whose guidance was essential. I owe a great deal to Marco for his intellectual rigour; he consistently challenged me to refine my arguments, and his patient feedback taught me the value of precision and a high standard for research. I am particularly grateful for his integrity and belief in my work during challenging times. I am also deeply thankful to Agar for her steady support and encouragement, which were a constant source of motivation.

Research is never a solo endeavor, and this thesis benefited greatly from collaboration. My thanks to my co-authors, Sergio Galletta and Monica Billio. I am especially grateful to Monica and Marco for bringing me into this fascinating line of research and for their early trust in my ability to contribute. Our shared curiosity made the work a truly rewarding experience.

I owe a special thank you to Sergio for his immense practical and intellectual contribution. He generously shared his resources, expertise, and time, and our long conversations were instrumental in improving this work while sparking ideas for future collaborations. His willingness to advise on my other projects speaks volumes about his dedication as a mentor and as a person. On a personal note, I am profoundly grateful for his belief in my potential and for his crucial role in helping me take the next step in my career at ETH Zürich.

My gratitude also extends to Ca' Foscari University of Venice for providing the resources and supportive community essential for my development. I extend special thanks to Davide Raggi for his crucial feedback and encouragement, and to Irene Mammi for her support during the early stages of my research. I am also grateful to Michele Costola for the perspectives that proved invaluable for my third chapter.

I owe a special thanks to my fellow PhD colleagues, whose friendship and support—from late-night study sessions to spritz by the canals—made this journey lighter.

Special mentions go to Luca for the endless jokes, Flavio for the passionate debates about music and politics, and Andrea and Qing, for the countless conversations that challenged my thinking and helped me approach my research with greater clarity. Last but not least, a heartfelt thank you to the department’s Wednesday football crew for enduring my questionable skills and always including me—it was the perfect weekly escape.

I am also sincerely grateful to the CLE department at ETH Zürich for graciously hosting me for a year. My deepest thanks to Elliott Ash for the invitation and support throughout the year. I am also grateful to Gianmarco for the countless coffee breaks and the insightful research discussions that came with them. A special thanks is due to Claudia for her invaluable feedback on my work, to Alex, who patiently taught me everything I know about topic modeling and NLP, and to Jacob for being an invaluable guide through a complex cultural landscape essential to my research.

I would like to extend a special thank you to Giovanni Mastrobuoni, with whom I had the pleasure of working as a research assistant and now as a collaborator. His mentorship and insights have been incredibly valuable. My gratitude also goes to the many inspiring scholars I met at conferences, whose questions sharpened my thinking. I am deeply grateful to Paolo Buonanno and Ruben Durante for serving as external reviewers; their feedback has substantially improved this work, and their comments were extremely helpful.

Preface

This doctoral thesis brings together three independent essays that examine major social and political phenomena—political communication, conflict dynamics, and organized crime—and assess their impact on democratic accountability, collective violence, and the broader economic environment. Although the topics span different time periods and empirical contexts, the dissertation is unified by a shared analytical approach: understanding how political and social actors respond to institutional incentives, external shocks, and evolving network structures. Across the chapters, the thesis employs a consistent toolkit of modern quantitative methods—including quasi-experimental designs, computational text analysis, spatial data, graph theory, and machine learning—to illuminate how these forces shape behavior and generate downstream consequences for society.

The first essay, **“One Person, One Vote”: The Effect of Direct Elections on U.S. Senators**, looks into the nature of political accountability by examining how a fundamental electoral reform—the 17th Amendment—reshaped the behavior of U.S. Senators. Moving beyond traditional analyses of voting records, this study applies computational linguistics to a vast corpus of congressional speeches to distinguish between legislators’ public discourse and their formal legislative actions.

The analysis reveals that the shift to direct elections compelled Senators to publicly align their policy priorities with those of their new constituents, making their agendas more “House-like”. Yet, this profound change in rhetoric did not correspond to a significant ideological shift in their roll-call votes. The essay’s contribution lies in demonstrating that electoral accountability operates through distinct channels, powerfully shaping politicians’ policy platforms while leaving underlying legislative behavior, constrained

by party and procedure, largely intact.

The second essay, **Political Shocks, Extremism and Violence: Evidence from the West Bank**, explores the “bottom-up” dynamics of political violence, arguing that external diplomatic signals can act as powerful catalysts for local conflict. It investigates the violent aftermath of the 2017 U.S. decision to recognize Jerusalem as Israel’s capital, treating the announcement as a quasi-experimental shock. Using high-resolution spatial data, the study shows that this political act triggered an immediate and dramatic surge in violence.

This escalation was highly asymmetric, initiated by a wave of settler-perpetrated attacks that was followed, nearly four years later, by a rise in Palestinian-led violence. The project draws on this asymmetry to examine how external political shocks can generate escalation: first by enabling extremist actors on one side to act with greater impunity, and later by prompting retaliatory mobilization from the other. The paper’s central insight is that top-down diplomatic shocks can destabilize fragile environments not through direct command, but by eroding social norms and legitimizing extremist behavior, thereby sparking a durable and self-perpetuating spiral of conflict.

Building on these findings, future research will explore the feedback loop between this external shock and Israel’s domestic political landscape. A key next step is to analyze the rhetoric of Israeli politicians in the Knesset to determine whether they amplified the shock with radical language or moderated their tone in response. In parallel, a collaboration with the NGO Kerem Navot will examine the state’s material actions by analyzing new data on territorial interventions, such as land allocations to settlers. This dual analysis of rhetoric and policy aims to provide a more complete picture of the interplay between international diplomacy, local extremism, and domestic policy.

The third essay, **Mafia Network: a Machine-Learning Approach**, moves from explaining past events to predicting future illicit activity, developing an innovative early-warning system to detect mafia infiltration in the legal economy. The study’s core innovation is the construction of a dynamic corporate ownership network centered on confiscated Italian firms, which serves as the backbone for a predictive framework. By combining machine learning with features derived from both network topology and firm-level data, the model successfully forecasts confiscation up to three years in advance with exceptionally high accuracy.

The analysis yields a critical insight for law enforcement: network data and firm-level data play complementary roles. Network structure is paramount for recall, casting

a wide net to identify a broad set of potential targets, while financial distress signals are key for precision, helping to pinpoint the most acute risks. This demonstrates the value of integrating graph theory into risk-based supervision, offering a powerful and interpretable tool to combat organized crime.

A significant extension of this work, as proposed in the essay, lies in leveraging more advanced deep learning models to exploit the rich graph structure of the data. Specifically, employing Graph Neural Networks (GNNs) would allow the model to learn node embeddings that capture complex, non-linear patterns of connectivity directly from the ownership graph, rather than relying on pre-defined topological measures. Such embeddings can reveal structural fingerprints of infiltration—how criminal actors position themselves, expand their influence, and exploit vulnerabilities in corporate networks. This approach could not only improve predictive performance but also offer a deeper understanding of the dynamics of mafia penetration and its broader consequences for economic activity, shedding light on how organized crime reshapes firm behavior and distorts competitive environments.

Taken together, these essays highlight a core insight: across democratic institutions, conflict settings, and illicit markets, political and social actors adapt strategically to shifts in their environment. Electoral reforms reshape how politicians communicate with voters; international diplomatic signals alter local norms and patterns of violence; and evolving corporate structures create openings for organized crime to infiltrate legal markets. In each domain, the dissertation shows that responsiveness to changing incentives and constraints is central to understanding political and economic outcomes. By integrating diverse empirical settings with a consistent methodological toolkit, the thesis proposes a cohesive framework for studying adaptation and strategic behavior in political economy.

Table of Contents

“One Person, One Vote”: The Effect of Direct Elections on U.S. Senators

1.1	Introduction	3
1.2	Institutional Background	7
1.2.1	The 17th Amendment and its Institutional Context	7
1.2.2	Institutional Differences Between the House and Senate	10
1.3	Data and Methods	10
1.3.1	Data	10
1.3.2	Text Analysis	12
1.3.3	Empirical Strategy	17
1.4	Results	19
1.5	Additional Results and Robustness Checks	23
1.5.1	Political Selection and Senate Primaries	23
1.5.2	Topic Concentration	25
1.5.3	Roll-call Behavior	26
1.5.4	Robustness checks	27
1.6	Conclusion	28
A	Appendix	33
A.1	Figures	33
A.2	Tables	48
B	Topic Model	51
C	A Stylized Voting Model of Malapportionment and Agenda Concentration	54

Political Shocks, Extremism and Violence: Evidence from the West Bank

2.1	Introduction	61
2.2	Background	64

2.2.1	Historical Context	64
2.2.2	The 2017 Jerusalem Declaration	66
2.3	Data	69
2.4	Methodology	73
2.4.1	Spatial Data	73
2.4.2	Identification	74
2.5	Results	77
2.5.1	Main Results	77
2.5.2	Perpetrators and Targets of Violence	79
2.6	The Declaration as a Trigger for Violence	81
2.7	Palestinian Perpetrated Violence	83
2.8	Robustness Checks	87
2.9	Discussion: The Domestic Political Environment and the Post-2017 Es- calation	91
2.10	Conclusion	94
A	Appendix	102
A.1	Figures	102
A.2	Tables	132
B	Methodology for LLM-Assisted Classification	141
C	Estimating Population of Palestinian Towns	143

Mafia Network: a Machine-Learning Approach

3.1	Introduction	149
3.2	Data	155
3.3	Methodology	160
3.3.1	Data pre-processing	161
3.3.2	Classification	163
3.4	Results	165
3.4.1	Main Results	165
3.4.2	SHAP Values	169
3.5	Robustness	172
3.6	Going Forward: Graph Representation Learning for Infiltration Risk . . .	174
3.7	Embedding Analysis	181
3.8	Conclusion	186

A	Appendix	192
A.1	Figures	192
A.2	Tables	213
B	Network Visualization: Software and Settings	235
B.1	Interactive Features and Data Exploration	235
B.2	Network Layout and Physics Simulation	236
C	Methodological Details: Machine-Learning Models	237
C.1	Models	237
C.2	Hyperparameters and Randomized Cross-Validation	238
C.3	Synthetic Minority Over-sampling TEchnique (SMOTE)	239
D	Model Interpretability with SHAP	241
E	Temporal Graph Network	243

List of Figures

Chapter 1: “One Person, One Vote”: The Effect of Direct Elections on U.S. Senators

1.1	Historical Timeline: Direct Elections of U.S. Senators	8
1.2	Histogram of Speech Length & Number of Speeches	11
1.3	W-NOMINATE Scores for the House and Senate	12
1.4	Carter Glass Policy Platform	15
1.5	Cosine Similarity of Political Platforms	17
1.6	Cosine Similarity	21
1.7	β Coefficient of the Reform on Topic Relevance	22
A.1	Direct Primaries	33
A.2	Speech Counts by Chamber	34
A.3	Time Series of Alternative Policy Platform Similarity Measures	35
A.4	Policy Platform Examples	36
A.5	Event-Study: Elections and Voting Topic	37
A.6	Event-Study: Postal Services Topic	37
A.7	Event-Study: Textiles and Manufacturing Topic	38
A.8	Event-Study: Fiscal Policy and Redistribution Topic	38
A.9	Event-Study: Taxation Topic	39
A.10	Event-Study: Labor and Economic Issues Topic	39
A.11	Event-Study: Immigration and Census Topic	40
A.12	Event-Study: Community Development and Education Topic	40
A.13	Event-Study: Native American Affairs and Schooling Topic	41
A.14	Event-Study: Debt Settlements and War Finances Topic	41
A.15	Event-Study: Public Buildings and Infrastructure Topic	42

A.16	Event-Study: Entropy of Political Platform	42
A.17	Event-study: Absolute Value of W-NOMINATE	43
A.18	Cosine Similarity of Policy Platform for Incumbents.	43
A.19	Cosine Similarities excluding states with Senate primaries	44
A.20	Event-Study: Cosine Similarities with Staggered Adoption.	44
A.21	Event Studies: Alternative Measures for Platform Similarity.	45
A.22	Event Studies: LDA models, Platform Similarity	46
A.23	Cosine Similarity, Leave-One-Out Estimates	47
B.24	Training loss of ProdLDA with 36 Topics	53
C.25	Simulation results: Entropy under Electoral Regimes	56
C.26	Simulation results: Entropy under Electoral Regimes	57

Chapter 2: Political Shocks, Extremism and Violence: Evidence from the West Bank

2.1	Outposts in the Occupied West Bank	66
2.2	Number of Daily Violent Events in the West Bank.	67
2.3	Main Actors of Violence in the West Bank.	70
2.4	Cumulative Monthly Violent Events where Settlers are Involved	75
2.5	Event Study: Number of Events	78
2.6	Event Study: Violent Events with Settlers Involved	78
2.7	Event Study: Targets of Violence	80
2.8	Event Study: Perpetrators of Violence.	81
2.9	Event Studies: Declaration as a Structural Break.	82
2.10	Monthly Palestinian Attacks over Time.	84
2.11	Right & Far Right Vote Shares in the West Bank.	92
A.1	Map of the West Bank	102
A.2	Outposts, Settlements and Palestinian Communities in the West Bank	103
A.3	Demolitions in the West Bank	104
A.4	Polling Stations, Settlements and Outposts in the West Bank.	105
A.5	Knesset Vote Share by Ideology	106
A.6	Outpost Timing Example - Bentayim Farm Outpost	107
A.7	Voronoi Polygons	108
A.8	Outpost Buffer Areas	109

A.9	Outpost Buffer Areas	109
A.10	Areas with Outpost Activity	110
A.11	Histogram of Outpost Intensity	111
A.12	Event Study: Civilian Targeting	111
A.13	This is the main caption for the entire figure with two subfigures.	112
A.14	Target Event Studies: Declaration as a Structural Break	113
A.15	Perpetrator Event Studies: Declaration as a Structural Break	114
A.16	Monthly Israeli Attacks over Time	115
A.17	Robustness Check: Treatment Status and Violent Events	116
A.18	Placebo Test: Main Results	117
A.19	Placebo Test: Perpetrators and Targets	118
A.20	Alternative Thresholds: Main Results	119
A.21	Alternative Thresholds: Perpetrators and Targets	120
A.22	Synthetic Control: Main Results	121
A.23	Synthetic Control: Perpetrators and Targets	122
A.24	Non-Staggered Specification: Main Results	123
A.25	Non-Staggered Specification: Perpetrators and Targets	124
A.26	Alternative Buffers: Main Results	125
A.27	Alternative Buffers: Perpetrators and Targets	126
A.28	Leave-out Regressions: Main Results	127
A.29	Leave-out Regressions: Perpetrators and Targets	128
A.30	Geographical and Administrative Components for Spatial Clustering	129
A.31	Resulting Spatial Clusters for Standard Error Adjustment	130
A.32	Fiscal Transfers to Settlements by Israeli Government	131
B.33	Confusion Matrices for Perpetrators and Targets	142
C.34	Population of West Bank Governorates, Towns and Built-up Areas	144
C.35	Population of West Bank Governorates: Buffer Areas and Weights	145
C.36	Estimated Population of West Bank	146

Chapter 3: Mafia Network: a Machine-Learning Approach

3.1	Graph Snapshot for 2016	159
3.2	Model Performance ROC Curves by Forecast Horizon	168
3.3	XGBoost SHAP Values by Feature, $h = 0$	170

3.4	XGBoost SHAP Values by Network Feature, $h = 0$	172
3.5	Embeddings Visualization (2006-2012)	181
3.6	Histogram of Average Cosine Similarity to Nearest Confiscated Firms	183
3.7	Model Performance ROC Curves by Forecast Horizon	186
A.1	Distribution of Confiscations over time	192
A.2	Geographical Distribution of Confiscations	193
A.3	Sector Distribution of Confiscated and Non-Confiscated Firms	194
A.4	Static Graph	195
A.5	Network Snapshots over Time	196
A.6	Degree Distributions	196
A.7	Undirected Centralities	197
A.8	Directed Centralities	198
A.9	Actual and Synthetic Confiscation Dates	199
A.10	XGBoost, Confusion Matrices	200
A.11	LightGBM, Confusion Matrices	201
A.12	Model Performance, $h = 2$	202
A.13	Model Performance, $h = 3$	203
A.14	XGBoost SHAP Values by Feature, $h \in \{1, 2, 3\}$	204
A.15	XGBoost SHAP Values by Network Feature, $h \in \{1, 2, 3\}$	205
A.16	LightGBM SHAP Values by Feature, $h \in \{0, 1, 2, 3\}$	206
A.17	LightGBM SHAP Values by Network Feature, $h \in \{0, 1, 2, 3\}$	207
A.18	Earlier Cut-off Date: Actual and Synthetic Confiscation Dates	208
A.19	Network Events Over Time	208
A.20	Edge Events over Time	209
A.21	Distribution Ownership Stake Target Variables	209
A.22	Histogram of Average Cosine Similarity to Nearest Confiscated Firms	210
A.23	Matching Strategy for Synthetic Reference Year Assignment	211
A.24	Model Performance ROC Curves by Forecast Horizon	212
E.25	Training Loss Comparison: Standard TGN vs. Two-Tower TGN	248
E.26	Temporal Graph Network Architecture	249

List of Tables

Chapter 1: “One Person, One Vote”: The Effect of Direct Elections on U.S. Senators

1.1	Pre-processing of US Congressional Speeches	14
1.2	Descriptive Statistics	18
1.3	Platform Similarity Regression Results	20
1.4	Heterogeneity Analysis - TWFE Results for Platform Similarity	24
1.5	Additional Results: Concentration of Policy Platform.	25
1.6	W-Nominate Regression Results	26
A.1	Topic Labels and Words from Congressional Speeches (1879-1935)	48
A.2	Robustness Checks - Platform Similarity under alternative measurements	49
A.3	Robustness Checks - Platform Similarity Regressions: LDA with 30, 33 & 39 Topics	50
B.4	Combined NPMI and CV Scores for Different Topic Models	53

Chapter 2: Political Shocks, Extremism and Violence: Evidence from the West Bank

2.1	Effect of Outpost Establishment on Violent Events	77
2.2	Heterogeneous Effects by Target and Perpetrator.	79
A.1	Descriptive Statistics	132
A.2	Political Parties and Ideological Classifications.	133
A.3	Spatial Characteristics Summary	134
A.4	Pre-declaration Balance Test (KS)	134
A.5	Effect of Transitory Outpost Establishment on Violent Events	135

A.6	Heterogeneous Effects of Transitory Outpost Establishment by Target and Perpetrator.	135
A.7	Effect of Settler Violence and Demolitions on Palestinian Violence (Temporal Comparison)	136
A.8	Robustness Check: Effect of Outpost Establishment on Violent Events across Different Intensity Thresholds.	137
A.9	Robustness Check: Heterogeneous Effects by Target and Perpetrator across Different Intensity Thresholds.	138
A.10	Effect of Outpost Establishment on Violent Events (Non-Staggered Design) . . .	138
A.11	Heterogeneous Effects by Target and Perpetrator (Non-Staggered Design)	139
A.12	Effect of Outpost Establishment on Violent Events (IHS Transformation)	139
A.13	Heterogeneous Effects by Target and Perpetrator (IHS Transformation)	140
A.14	Effect of Outpost Establishment on Violent Events	140
A.15	Heterogeneous Effects by Target and Perpetrator.	140

Chapter 3: Mafia Network: a Machine-Learning Approach

3.1	Descriptive statistics - Directed Mafia Network	160
3.2	Descriptive Statistics by Forecast Horizon	163
3.3	Model Performance Comparison Across Feature Sets, $h = 0$	166
3.4	Model Performance Comparison Across Feature Sets, $h = 1$	167
A.1	Centrality Measures Descriptive Statistics by Network Type and Node Category	213
A.2	List of Firm-level Variables from <i>Infocamere</i> by Category.	214
A.3	Model Performance Comparison Across Feature Sets, $h = 2$	215
A.4	Model Performance Comparison Across Feature Sets, $h = 3$	216
A.5	Model Performance Comparison Across Feature Sets (Sampled Years), $h = 0$. . .	217
A.6	Model Performance Comparison Across Feature Sets (Sampled Years), $h = 1$. . .	218
A.7	Model Performance Comparison Across Feature Sets (Sampled Years), $h = 2$. . .	219
A.8	Model Performance Comparison Across Feature Sets (Sampled Years), $h = 3$. . .	220
A.9	Model Performance Comparison Across Feature Sets (No Lag Features), $h = 0$.	221
A.10	Model Performance Comparison Across Feature Sets (No Lag Features), $h = 1$.	222
A.11	Model Performance Comparison Across Feature Sets (No Lag Features), $h = 2$.	223
A.12	Model Performance Comparison Across Feature Sets (No Lag Features), $h = 3$.	224

A.13	Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 0$	225
A.14	Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 1$	226
A.15	Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 2$	227
A.16	Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 3$	228
A.17	Edge Event Breakdown and Link Duration Statistics	229
A.18	Event Type Distribution Across Data Splits	229
A.19	Performance Comparison of TGN Model Variants	230
A.20	Descriptive Statistics by Forecast Horizon	230
A.21	Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 1$	231
A.22	Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 2$	232
A.23	Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 3$	233
A.24	Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 4$	234

*And here, poor fool! With all my lore
I stand, no wiser than before:
I'm Magister-yea, Doctor-hight,
And straight or cross-wise, wrong or right,
These ten years long, with many woes,
I've led my scholars by the nose, -
And see, that nothing can be known!
That knowledge cuts me to the bone.*

— J. W. Goethe, *Faust*

Chapter 1

“One Person, One Vote”: The Effect of Direct Elections on U.S. Senators

Joint work with Sergio Galletta

Abstract: This study investigates the impact of the 17th Amendment, ratified in 1913, on the legislative behavior of U.S. Senators. By shifting Senate elections from indirect selection by state legislatures to direct popular vote, the amendment fundamentally altered senators’ electoral incentives and accountability. We analyze a comprehensive dataset of congressional speeches and roll-call votes from 1879 to 1935. Using a difference-in-differences design that leverages the already directly-elected House of Representatives as a control group, we employ natural language processing to track shifts in policy agendas. Our findings indicate that direct elections caused senators’ public policy priorities to converge significantly with those of House members. Following the reform, senators increased their focus on key issues like economic policy and taxation while substantially reducing their emphasis on infrastructure building and fiscal discipline. Furthermore, their overall policy agendas became more concentrated on a narrower set of topics. However, an analysis of their roll-call voting records reveals no significant ideological shifts. These results suggest that the transition to direct elections primarily reshaped senators’ public-facing discourse rather than their underlying legislative voting behavior.

JEL Classification: D72, D78, L51

Keywords: 17th Amendment, United States congressional elections, Representative democracy, Congressional Record, Text analysis, Machine Learning

1.1 Introduction

Accountability of politicians to the public is a cornerstone of democratic governance. Nonetheless, even within democratic settings, there are differences in institutional frameworks and electoral rules that can lead to diverse outcomes in how public officials conduct themselves. For example, whether elections are majority-based or proportional, or whether there is a runoff system instead of a single-round plurality rule, can significantly affect the types of officials elected and their behavior in office ([Alesina, 1995](#); [Persson and Tabellini, 2004](#); [Bordignon et al., 2016](#)).

Another key factor influencing politicians’ behavior and incentives is whether they are selected through direct or indirect elections. When citizens vote directly for their representatives, it establishes a clear line of accountability, potentially encouraging policies and actions that correspond with public preferences due to stronger electoral pressures. In contrast, indirect elections, where officials are appointed by other elected representatives, can weaken this accountability. This partial separation from voters often reduces the pressure to prioritize citizens’ needs and may lead to decision-making influenced by factors such as seniority or party loyalty rather than broad representativeness. While direct elections foster grassroots connections and responsiveness to public demands, the reduced stringency in indirect elections might shift priorities. In this context, the 17th Amendment to the U.S. Constitution, ratified in 1913, represents one of the most significant transitions from indirect to direct elections in American political history. This reform fundamentally altered the dynamics of political accountability by shifting the selection of U.S. Senators from state legislatures to a direct vote by the people.

In this paper, we aim to provide new evidence on the role of the 17th Amendment in shaping representatives’ behavior. To analyze this, we use standard techniques in Natural Language Processing (NLP) and quasi-experimental methods to explore how the reform influenced both legislative discourse and voting behavior. Our analysis is based on a comprehensive dataset of approximately 6.5 million speeches delivered by members of Congress from the 46th to the 73rd Congress (1879–1935). To identify the causal effects of the amendment, we employ a difference-in-differences approach that leverages variation across time (before and after the reform) and across chambers (Senate and House). This allows us to isolate the impact of transitioning to direct elections by comparing Senators’ behavior to that of House Representatives, who were already directly elected. We use Latent Dirichlet Allocation (LDA), a probabilistic topic modeling

technique, to reconstruct legislators’ policy agendas over time. LDA identifies thematic structures within speeches, enabling us to analyze shifts in discourse and measure policy agenda similarity between Senators and House Representatives (Blei et al., 2003). By examining textual data (speeches), we assess whether the amendment’s effects were affecting Senators’ public-facing policy priorities. While previous studies have highlighted associations between the amendment and shifts in senatorial behavior (Bernhard and Sala, 2006; Gailmard and Jenkins, 2009), our analysis builds on this work by employing recent causal inference and NLP methods to provide additional evidence of how the amendment influenced legislative behavior and policy priorities.

Our findings suggest that the 17th Amendment had a significant impact on Senators’ behavior, particularly in terms of their policy agendas as expressed in their speeches. Specifically, we show that the reform caused Senators to move their policy focus closer to that of House representatives, reflecting a shift in their political priorities following the introduction of popular elections. Importantly, our results are not driven by the potential selection of “different” politicians following the reform, as the findings remain consistent even when focusing exclusively on incumbent congressmen—those who served both before and after the reform. Furthermore, Senators’ policy agendas became more concentrated, with a narrower focus on fewer issues. They increasingly emphasized topics like taxation, economic policy, and democratic representation while reducing attention to areas such as infrastructure, immigration and fiscal discipline. Finally, we find no discernible effect on their roll-call voting behavior, which we proxy using their ideological alignment W-NOMINATE (Poole and Rosenthal, 1985). This finding is consistent with the idea that voting and speeches serve distinct purposes in Congress: speeches allow legislators to flexibly share priorities and connect with voters, while voting is constrained by rules, party dynamics, and negotiation, making it less responsive to direct elections (Gentzkow et al., 2019).

These patterns are consistent with the idea that direct elections reshape legislators’ incentives by altering the audience to whom they must appeal and the channels through which they can most easily signal responsiveness. Under indirect elections, Senators faced an audience composed of state legislators whose preferences were often narrow, institutionally mediated, and shaped by malapportioned districts (Schiller and Stewart, 2014). Popular elections, by contrast, expose Senators to a broad and heterogeneous statewide electorate for whom certain issues—such as taxation, labor markets, and political rights—carry greater salience. Because rhetorical communication is a low-cost

and highly flexible margin of political behavior, it provides a natural way for Senators to adapt to these new electoral pressures. The shift toward more focused agendas and greater similarity to House members therefore reflects not only a change in principals, but also a change in the informational and strategic environment in which Senators operate. More generally, our results suggest that electoral reforms can reshape the discursive dimension of representation by influencing what legislators choose to emphasize in public, even when their roll-call behavior remains largely stable. An important avenue for future work is to examine whether these rhetorical adjustments were accompanied by shifts in the substance of policymaking itself—for instance, by studying changes in the textual content or policy focus of bills introduced or discussed in Congress during this period.

Our results can be understood through several theoretical lenses. In a classical principal–agent framework ([Ferejohn, 1999](#)), the 17th Amendment replaced a narrow principal—state legislators—with a broad and heterogeneous electorate. Because speeches are far less constrained than roll-call votes, this shift in principals is most likely to manifest in rhetorical behavior: newly accountable Senators adopt public-facing agendas that appeal to the median statewide voter, thereby becoming more “House-like” in their emphasis. A complementary interpretation follows from probabilistic voting models ([Lindbeck and Weibull, 1987](#)), in which equilibrium platforms also reflect the relative weight of voter groups. Prior to the reform, malapportioned state legislatures overrepresented low-population rural districts, giving disproportionate influence to preferences that were not reflective of statewide majorities. As a result, the indirect-election equilibrium may have produced more diffuse or idiosyncratic policy emphases. By contrast, direct elections aggregate preferences at the individual level, increasing the weight of populous districts and generating stronger incentives for clear, focused appeals on broadly relevant issues. The observed decline in agenda entropy and increased similarity to House members are therefore consistent with a model in which electoral reform realigns political incentives, reshapes who matters for re-election, and alters the rhetorical—but not necessarily the behavioral—margin of legislative activity.

This paper contributes to several strands of literature. First, it adds to the broad body of work that highlights how electoral rules can influence the behavior of both politicians and voters. For instance, [Besley and Case \(1995\)](#) explores how gubernatorial term limits influence economic policy decisions. The authors find that governors facing term limits are more likely to implement policies that boost short-term economic

performance, potentially at the cost of long-term fiscal health. Similarly, [Persson et al. \(2003\)](#) examines how different electoral systems affect political corruption, finding that systems with greater individual accountability (like personal ballots in plurality elections) see less corruption, while those with higher reliance on party lists see more. In another study, [Gagliarducci et al. \(2011\)](#) uses data from the Italian House of Representatives to compare politicians’ behavior in majoritarian and proportional electoral systems. They find that politicians elected in majoritarian systems engage more in geographically targeted activities and exhibit lower absenteeism rates than their counterparts in proportional systems, indicating that electoral systems can significantly influence legislative behavior. Additionally, [Bordignon et al. \(2016\)](#) study the effects of electoral systems on political extremism. They find that runoff elections, compared to single-round elections, increase the number of political candidates and reduce policy volatility by diminishing the influence of extremist parties. This leads to more moderate policies, as the bargaining power of political extremes is reduced. Our study contributes to this literature by demonstrating how a key electoral reform, the 17th Amendment, influenced politicians’ public-facing policy agendas.

Second, we add to the growing literature examining parliamentary speeches as proxies for politicians’ behaviors and preferences. For instance, [Gentzkow et al. \(2019\)](#) investigates the evolution of partisanship in congressional speeches from 1873 to 2016, identifying a significant increase in partisanship beginning in the early 1990s. Additionally, [Gennaro and Ash \(2022\)](#) explores the use of emotional and rational language in U.S. Congressional speeches from 1858 to 2014, demonstrating that emotionality has increased over time, especially during periods of war and televised debates, and varies according to party affiliation and individual characteristics of Congress members. [Ash et al. \(2024\)](#) examines the impact of gender on the reactions to parliamentary speeches in Germany, finding that gender-conforming speeches in German state parliaments typically receive more positive reactions, suggesting that gender stereotypes significantly shape parliamentary dynamics. While our study is not the first to employ NLP methods to analyze politicians’ speeches, we propose the idea that LDA topics can be used to proxy for political agenda and to measure the similarity of policy agendas between legislative chambers.

Third, we contribute to the existing research on the effects of direct versus indirect selection of candidates. In the context of the 17th Amendment, [Bernhard and Sala \(2006\)](#) found that the reform encouraged Senators to moderate their positions to ap-

peal to median state voters, resulting in greater efforts toward re-election, particularly against opposition from antagonistic state legislatures. Similarly, [Gailmard and Jenkins \(2009\)](#) observed that Senators became more responsive to electorate preferences after the amendment, which also granted them increased legislative freedom. [Lowande and Peck \(2017\)](#) highlighted that the amendment led to more Senate investigations of the executive during periods of divided government, as Senators sought to demonstrate their commitment to protecting voters from political misconduct. Beyond the U.S. context, studies have examined comparable institutional transitions in which the shift from indirect to direct selection altered legislators’ incentives. For instance, [Micozzi \(2013\)](#) shows that directly elected Senators in Argentina became substantially more active in drafting bills addressing provincial concerns, while reforms in Italian local governments replacing indirect mayoral appointments with direct elections reshaped both accountability structures and policy priorities ([Ferlenga, 2023](#)). These cases underscore that the mechanisms we study are not unique to the United States but reflect a broader pattern in which expanding the electorate reshapes representatives’ incentives and public-facing behavior. Our analysis extends this body of work by applying rigorous causal inference and machine learning methods to examine how direct elections altered Senators’ responsiveness to voter preferences through changes in their policy agendas.

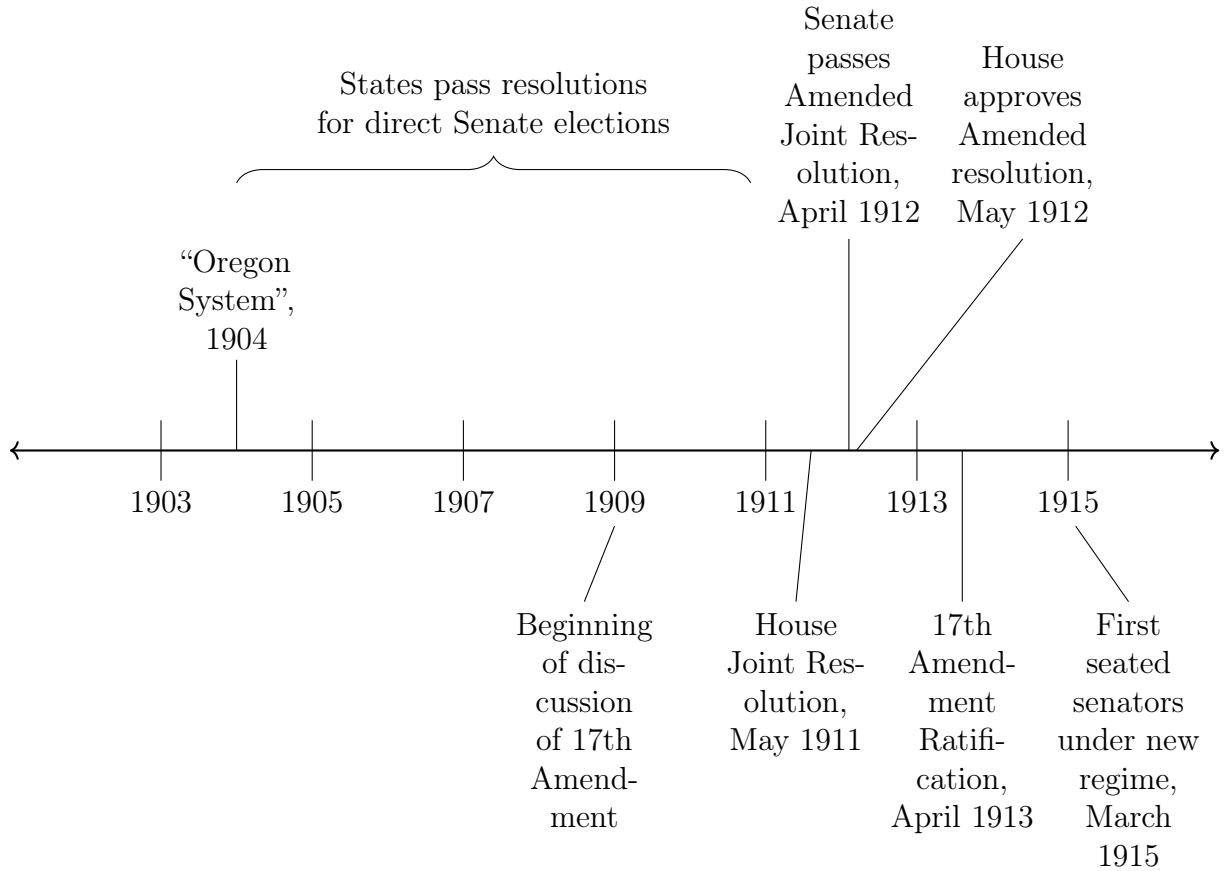
The remainder of the paper is organized as follows: Section [1.2](#) describes the institutional background. Data and methods are presented in Section [1.3](#). Section [1.4](#) presents the empirical results, while Section [1.5](#) provides additional results and robustness checks. Finally, Section [1.6](#) concludes the paper.

1.2 Institutional Background

1.2.1 The 17th Amendment and its Institutional Context

The ratification of the 17th Amendment in 1913 marked a transformative moment in American political history, introducing direct elections for U.S. Senators and fundamentally altering the balance of accountability in federal governance. Before this reform, Senators were elected indirectly by state legislatures, a system designed to insulate them from public opinion and emphasize federalism by reinforcing checks and balances ([Schiller and Stewart III, 2004](#); [Schiller et al., 2013](#); [Jenkins, 2005](#)). However, this method often led to deadlocks, delays, and vacancies in Senate appointments. Efforts to address these

Figure 1.1: Historical Timeline: Direct Elections of U.S. Senators



Notes: The figure outlines the timeline leading to the 17th Amendment, showing key events from early calls for direct Senate elections in 1903 to the seating of the first senators under the new system in 1915. Major milestones include the "Oregon System" in 1904, passage of resolutions in Congress (1911–1912), and ratification in April 1913. Congress numbers and years (in parenthesis) provide historical context.

inefficiencies began as early as 1866, when Congress passed a law to regulate Senate elections, but persistent challenges underscored the limitations of the existing system ([National Archives, 2023b](#)).

Figure 1.1 reports the path that ultimately led to the ratification of the amendment. During the Progressive Era, calls for electoral reform gained momentum as a wave of grassroots movements sought to improve government transparency and accountability. In the context of Senate elections, several states pioneered mechanisms to popularize the process. Notably, the “Oregon System”, adopted in 1904, required legislative candidates to pledge support for the plurality winner of a popular vote for Senate ([Bernhard and Sala, 2006](#)). Similar systems were implemented across other states, albeit with signifi-

cant variation in design and enforcement.¹ Despite these reforms, their effectiveness in altering senatorial behavior before the 17th Amendment remains debated (Schiller, 2006; Haynes, 1938; Gailmard and Jenkins, 2009). By 1910, 31 states passed resolutions urging Congress to amend the Constitution to allow for the direct election of senators (Bybee, 1997). In December 1911, the House passed Joint Resolution 39, which mandated the direct election of senators. This resolution included a clause barring federal intervention in case of racial discrimination among voters, the so-called “race-rider” (National Archives, 2023b). In April 1912, the Senate passed an amended joint resolution removing the “race rider” from the bill and the House approved it the following month (National Archives, 2023b). Following the usual process to change the constitution, at least 36 states—i.e. 75% of all 48 US States as of 1913, had to approve the final resolution. This was reached on April 8th, 1913, when the Amendment was officially ratified, becoming part of the Constitution and shaping all subsequent Senate elections (National Archives, 2023b).²

Notably, discussions on transitioning to direct elections began as early as 1909, during the 61st Congress, showing strong momentum for reform before formal legislative action.³ By the time of ratification, many Senators likely anticipated the reform, potentially adjusting their strategies and behavior beforehand. We account for this potential anticipation effect in our estimation strategy. It is worth noting that during the Progressive Era, there was a broad consensus, both in the House and in the Senate, on transitioning to the direct election of U.S. Senators, reflecting a nationwide push for greater democratic participation and government accountability. However, significant debate centered on the extent of federal involvement in determining voter eligibility. Many states were concerned about preserving their rights to set voting qualifications without excessive federal interference, as the initial “race-rider” clause in Joint Resolution 39 demonstrates (National Archives, 2023a). To account for this, we consider the beginning of the discussion on the 17th Amendment in our identification strategy.

¹Appendix Figure A.1 provides a bar chart illustrating the cumulative number of states adopting primary-like reforms over time.

²The time spent from 1911 to 1913, i.e., two years, matches the average time required for a constitutional amendment to be ratified.

³In that instance, the proposed reform was defeated by a margin of only four votes, largely due to the opposition of Southern Democrats who sought to maintain control over their electorate along racial lines (U.S. Congress, 1911).

1.2.2 Institutional Differences Between the House and Senate

Understanding the impact of the 17th Amendment requires examining the institutional context of the Senate and its relationship with the House of Representatives. While both chambers share legislative responsibilities, their structures and functions differ significantly, reflecting distinct roles within the federal government.

One key difference lies in electoral cycles. House representatives serve two-year terms, fostering immediate accountability to voters, while Senators serve six-year terms with staggered elections to ensure continuity and stability. Representation further distinguishes the chambers: House seats are allocated based on population, with members elected from specific districts, while the Senate provides equal representation for all states, amplifying the influence of smaller states ([Congress, 2023](#)).

The legislative roles of the two chambers also diverge. The House holds the exclusive power to initiate revenue bills, while the Senate confirms presidential nominations and approves treaties. Both chambers must agree on the same version of a bill for it to become law, but their procedural rules differ: the House prioritizes efficiency with stricter debate rules, whereas the Senate encourages extended deliberation to act as a counterbalance to rapid legislative changes ([Congress, 2023](#)).⁴

1.3 Data and Methods

1.3.1 Data

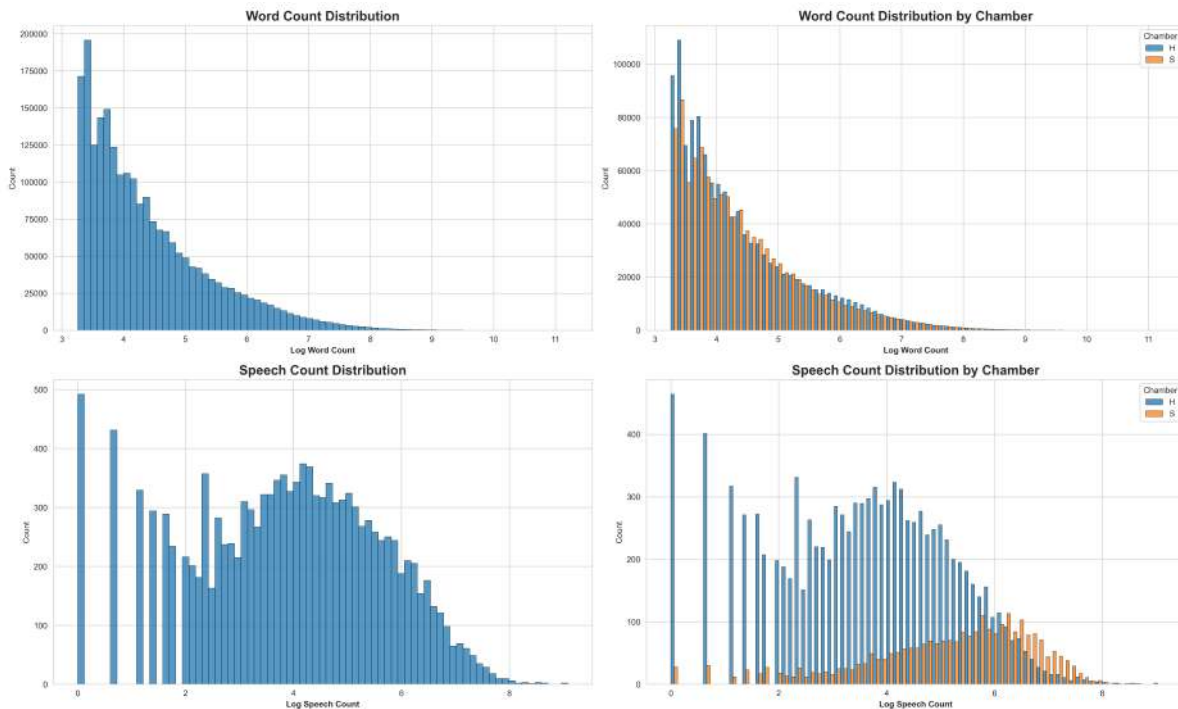
The data for this study is derived from the dataset compiled by [Gentzkow et al. \(2019\)](#), which includes transcripts from the U.S. Congressional Record covering the 19th and 20th centuries. Our focus is on the specific period from 1879 to 1935, from the 46th to 73rd Congresses. This time frame allows us to analyze the political discourse both before and after the introduction of the 17th Amendment. The dataset provides not only the text of the speeches but also detailed metadata about the speakers. To ensure the relevance of our findings, we concentrate solely on speeches delivered by voting members

⁴Regarding individual qualifications, House representatives must be at least 25 years old and have been U.S. citizens for at least 7 years. Senators must be at least 30 years old and have been U.S. citizens for at least 9 years. All members of Congress must reside in the state they represent. During the period of interest, the minimum voting age for both the Senate and the House was 21 years, until it was lowered to 18 in 1971 ([Congress, 2023](#)).

of Congress, thereby excluding contributions from non-voting participants. This refined dataset consists of 5,323,076 speeches (from a total of 6,531,892).

We count a total of 405,078,270 words and 2,276,167,184 characters in these speeches. Figure 1.2 shows the distribution of the number of speeches per speaker and the word count per speech, broken down by chamber, highlighting differences in participation and verbosity across Congress. Appendix Figure A.2 illustrates how the total number of speeches delivered by legislators has evolved over time.

Figure 1.2: Histogram of Speech Length & Number of Speeches

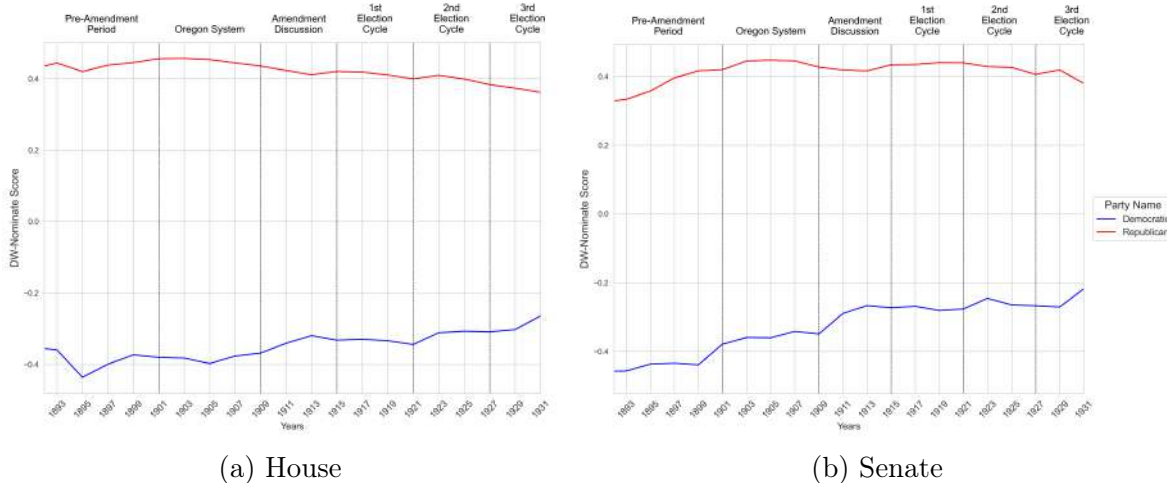


Notes: The figure presents descriptive statistics for our corpus. The top left panel displays the log of speech length, while the top right panel provides a breakdown of speech length by chamber. On the bottom row, the left panel illustrates the log of speech count per speaker, and the right panel shows the same metric distributed across chambers. We only considered speeches with more than 25 words.

We included roll-call data for both senators and representatives, obtained from [Lewis et al. \(2023\)](#). This dataset provides comprehensive information on every bill passed during the 46th to 73rd Congresses, including the votes cast by each legislator and their attendance during voting sessions. It also includes details about each legislator’s political career, such as whether they were elected or appointed, as well as notable events like contested seats, deaths, and resignations. From Wikipedia we also gathered information on whether a legislator was seeking re-election.

Furthermore, we use W-NOMINATE scores from the same dataset to measure the ideological positions of members of Congress. These scores place legislators on a liberal-conservative spectrum based on their roll-call voting behavior, with negative values indicating more liberal positions and positive values indicating more conservative ones. While the standard DW-NOMINATE scores are not ideal for direct comparisons across chambers, [Lewis et al. \(2023\)](#) also provide “Common Space” W-NOMINATE scores, which assign a constant ideology score to each legislator and allow comparisons across chambers.⁵ Figure 1.3 shows the evolution of W-NOMINATE scores by party and chamber. The data reveals a trend of moderation in roll-call behavior (indicated by a positive slope for the blue line and a negative slope for the red line) that begins well before discussions on the 17th Amendment, around the 61st Congress. Notably, while the Democratic Party exhibits clearer signs of moderation, this pattern contrasts with [Bernhard and Sala \(2006\)](#), who finds that Republican moderation was more pronounced.

Figure 1.3: W-NOMINATE Scores for the House and Senate



Notes: The figure presents the evolution of W-NOMINATE scores by party over time. Panel a) refers to the House and Panel b) to the Senate

1.3.2 Text Analysis

Topic model The primary empirical challenge of this study is to reconstruct the political agendas of Congress members through their speeches. This enables us to identify

⁵Common Space scores use the voting records of legislators who served in both the House and Senate to statistically align the two chambers onto a single, unified ideological dimension, a feature that makes them comparable in a way that standard DW-NOMINATE scores are not.

their policy priorities, classify their speeches by relevant policy areas, and track the evolution of their agendas over their congressional tenure. To achieve this, we employ a topic modeling approach, specifically leveraging Product of Experts LDA (*ProdLDA*), a variation of the widely-used LDA. LDA is an unsupervised machine learning technique designed to uncover underlying topics within a collection of documents (Hoffman et al., 2010). It assumes that each document is a mixture of topics, with each word attributable to one or more of these topics. This approach provides a structured way to analyze textual data by representing documents in a latent space defined by topics. In this space, each topic is a probability distribution over words.

To better understand the model, let us consider a corpus with D documents and W unique terms. The researcher chooses a latent space of size K : each topic can be represented as a probability vector over the vocabulary, formally $\beta_k \in \Delta^{W-1}$, where each element $\beta_{k,w}$ refers to the probability of word w appearing in topic k (Hansen et al., 2018). We denote the collection of the K random vectors as $\beta_{1:K}$. Similarly, the topic proportions for the d th document are defined as θ_d and $\theta_{d,k}$ represents the “share” of topic k is in the d th document. Moreover, the topic allocations for the d th document is $\mathbf{z}_d$, where $z_{d,n}$ is the topic assignment for the n th word in document d (Blei et al., 2003)⁶. The joint distribution for hidden and observed variables is:

$$p(\beta_{1:K}, \theta_{1:D}, \mathbf{z}_{1:D}, \mathbf{w}_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\theta_d) \left(\prod_{n=1}^{N_d} p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right), \quad (1.1)$$

where N_d is the total number of words in a specific document. The marginal distribution makes the posterior calculation intractable and it is usually approximated by either sampling based methods, like MCMC algorithms such as Gibbs sampling, or through variational inference (see Appendix B for a more in-depth discussion).

Speech Pre-processing and Model Selection Topic models like LDA use dictionary methods, where a predefined vocabulary is used to analyze word occurrences across a corpus. Congressional speeches are represented in a document-term matrix, with each row as a speech and each column as a vocabulary term, where an element (d, w) indicates how often word w appears in document d . To ensure interpretability and computational

⁶Blei et al. (2003) also uses Dirichlet priors for θ and β , which model the concentration of topic proportions, e.g. prior belief on how many topics are discussed in a given document, and on topic distribution, how many words in the vocabulary should characterize a topic.

Table 1.1: Pre-processing of US Congressional Speeches

	Number of Words	Vocabulary Size	Number of Speeches
Raw Text	457,544,195	2,253,385	6,531,892
Removing Short Speeches	415,644,247	2,183,801	2,634,239
Stopwords, Stemming & OCR adjustment	108,876,306	16,108	2,557,946
Bigrams	94,518,986	1,174,329	2,557,946
Frequency Adjustment	54,681,862	37,746	2,247,690
Final revision	19,198,831	26,407	1,989,923

Notes: The table shows the pre-processing procedure on the corpus. We consider number of words, vocabulary size and number of speeches

efficiency, we preprocess the speeches to define the vocabulary and reduce dimensionality.

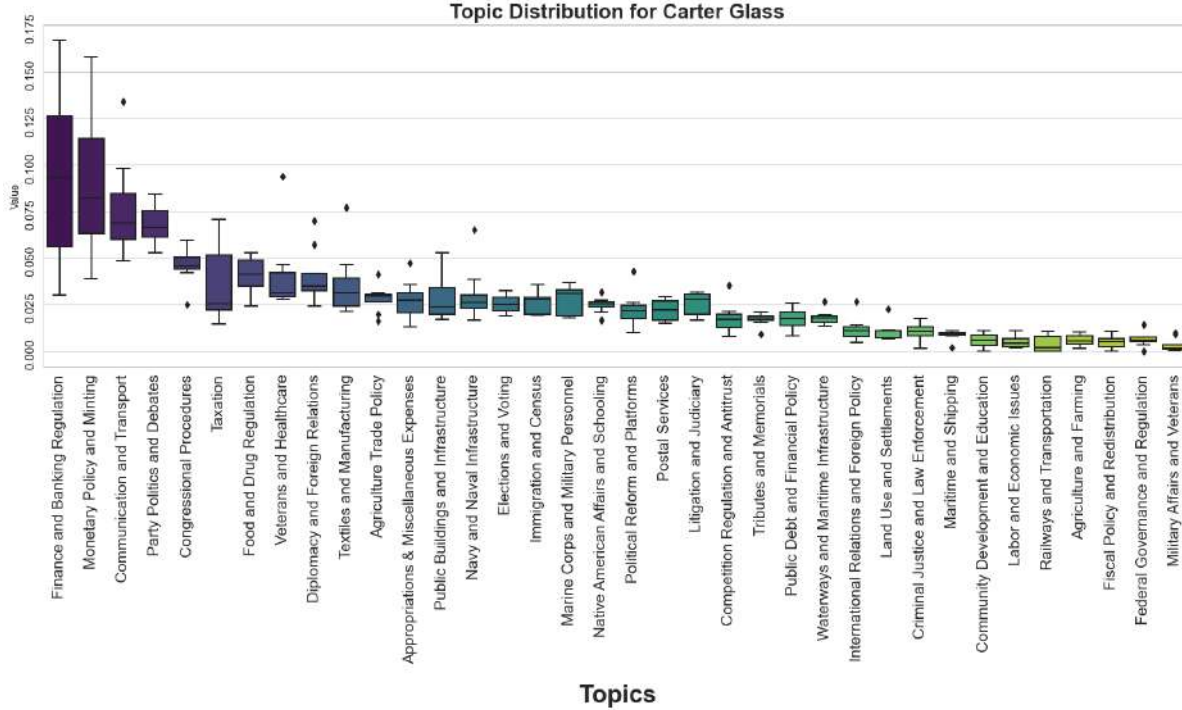
As an initial step, we remove punctuation and special characters. Following this, we tokenize and lemmatize each word. We also identify bigrams within each speech, i.e., collocations of two words that have a specific meaning such as “supreme court” or “income tax”. To achieve this, we use a part-of-speech tagger to identify sequences of two words such as *noun-noun*, *adjective-noun*, *verb-noun*, and we retain only those that occur more than 50 times throughout the entire corpus. Next, we remove stopwords and stem all words to further reduce vocabulary size.⁷ After this process, we are left with 37,033 words. We then manually revise the final vocabulary and discard words that do not carry a strong semantic connotation, e.g., words like “question”, “give”, “take”, and so forth. In the end, we are left with 26,407 unique terms. Table 1.1 summarizes the speech pre-processing procedure. It is important to note that the speeches were digitized using Optical Character Recognition (OCR), which often generates misspelled words in older documents. To address this, we compare tokens against an English dictionary to remove incorrect words, reducing noise and controlling vocabulary size.

Once we defined the vocabulary, we can proceed to train the model. Instead of random sampling, we use stratified sampling, selecting one-third of the speeches from both the House and Senate for each Congress. We then run our LDA on this sample for a different number of topics.

LDA Output and Policy Agenda From the LDA output and the topic-specific word list, we can label them in a specific set of policy themes that may characterize a congressperson’s political agenda. Appendix Table A.1 summarizes the final output

⁷Stemming is a text normalization technique where words are truncated to their root form. For instance, variants like “running,” “runner,” and “ran” are all reduced to the base word “run.”

Figure 1.4: Carter Glass Policy Platform



Notes: The figure illustrates Congressman Carter Glass’s policy platform, with each boxplot representing the distribution of a specific topic throughout his tenure as a legislator.

of our topic model.⁸ Once we have generated a topic distribution for each speech, we can link the LDA output to speaker information and aggregate the data at the speaker-congress level to create a proxy for each speaker’s political agenda.

For example, Figure 1.4 outlines the policy agenda of Congressman Carter Glass. Glass was instrumental in shaping banking and financial policies, widely known for his role in creating the Federal Reserve System with the Glass-Owen Act of 1913 ([Federal Reserve History, 2021](#)). The model performs well, as the topics extracted from his speeches clearly reflect his significant legislative focus on banking and finance.⁹

The resulting topic distribution for each legislator can be interpreted as a data-driven characterization of their policy agenda. Beyond these illustrative examples, the politician-specific topic distributions produced by the model offer an opportunity for

⁸The labels are generated using OpenAI’s GPT-4o model, asking to define a topic label based on the top 10 words for each LDA topic. We then manually revised the labels to ensure they accurately reflect the content of each topic. This was done by examining, for each topic, the top three speeches with the highest probability of belonging to that topic.

⁹Appendix Figure A.4 provide other examples from notable legislators in the late 19th and early 20th centuries, illustrating the effectiveness of our methodology in tracking politicians’ policy priorities.

further analysis. Because the LDA output provides a consistent, high-dimensional characterization of each legislator’s policy agenda, it can be used to study heterogeneity in legislative styles, specialization, constituency targeting, or agenda evolution over time. While we do not pursue these extensions here, future work could build on this structure to examine how individual-level rhetorical priorities relate to legislative productivity, bill sponsorship, or career trajectories, thereby deepening our understanding of how politicians allocate attention across issues.

Similarities in the political agenda One measure of interest is how similar the political agendas are of the House and Senate over time. To do so, we calculate the average policy platform of House representatives and compare it to the policy agenda of individual senators for each state using cosine similarity, as follows:

$$\cos(\theta_{ist}) = \frac{\mathbf{S}_{ist} \cdot \bar{\mathbf{H}}_{st}}{\|\mathbf{S}_{ist}\| \|\bar{\mathbf{H}}_{st}\|} = \frac{\sum_l^K S_{l,ist} \bar{H}_{l,st}}{\sqrt{\sum_l^K S_{l,ist}} \sqrt{\sum_l^K \bar{H}_{l,st}}},$$

which represents the cosine similarity of the political platform for senator i of state s during congress t with respect to the average political platform of House representatives in the same state and during the same period. We use the average as the House serves as a proxy for state-wide political interests rather than district-specific preferences. The K parameter represents the number of topics. Figure 1.5 displays the evolution over time of platform similarities. There’s an evident and sharp increase around the ratification of the amendment which supports the hypothesis of Senators moving their platform closer to the one of House representatives. We also consider similar measures to gauge the level of distance between Senator and House representatives, namely: Jensen-Shannon divergence, Kullback-Leibler Divergence and Bhattacharyya Distance.¹⁰

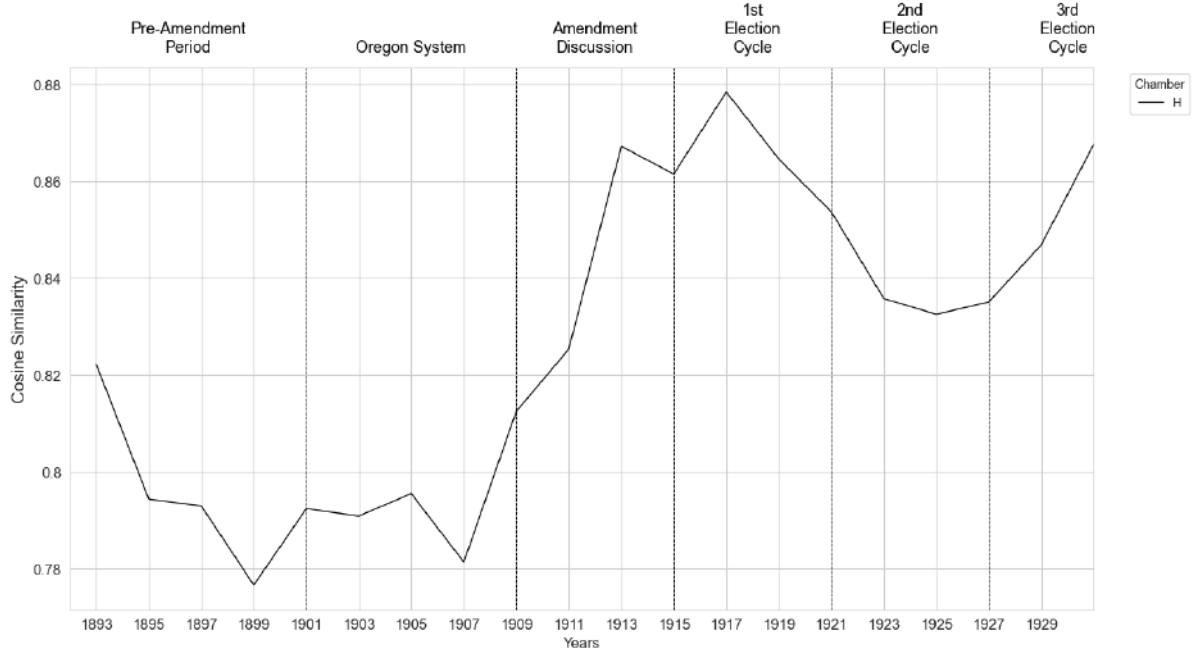
Finally, we create a measure to capture the degree of focus in legislators’ political platforms—whether these platforms are concentrated on a few key issues or spread across a broader range of topics. We do it by using a measure of topic concentration, such as entropy:

$$H_{it} = - \sum_{i=1}^{36} x_{izt} \ln(x_{izt}),$$

for legislator i during congress t . Table 1.2 summarizes the descriptive statistics for the

¹⁰As the names suggest, these variables measure dissimilarity, hence lower values would imply a larger degree of similarity between two vectors. To avoid confusion we represent these measures as $\frac{1}{1+\text{Divergence}}$.

Figure 1.5: Cosine Similarity of Political Platforms



Notes: The figure depicts average Cosine similarity over time between individual senators and the average policy platform of House representatives from their respective states.

main variables considered.

1.3.3 Empirical Strategy

With our estimation strategy, we analyze legislators' actions to examine how they changed before and after the 17th Amendment, focusing on shifts in senators' behavior and speech topics while using the House of Representatives as a control group for comparison. The reform from indirect to direct elections can be expected to influence the political priorities of Senators, moving their policy preferences more closely with those of House representatives, who have always been directly elected. The reasoning is straightforward: if the reform makes Senators more responsive to voter demands, their agendas should naturally converge with the House's. By using this approach, we avoid the need to explicitly define voter preferences or specify actions that reflect them. Instead, the behavior of House representatives offers a practical benchmark for comparison.

In our first set of estimations, the similarity measure between the policy agendas of Senators and House representatives, proxied by the cosine similarity of their speeches, acts as the first difference in a difference-in-differences (DiD) framework. Instead of a

Table 1.2: Descriptive Statistics

	Pre-Amendment		Post-Amendment		Overall	
	House	Senate	House	Senate	House	Senate
Amendment dummy					.403 (.49)	.386 (.487)
Majority Party (D = 1)	.442 (.496)	.217 (.413)	.199 (.399)	.3 (.45)	.344 (.475)	.249 (.433)
# Parties	4.594 (1.195)	2.99 (1.163)	5.203 (1.079)	2.697 (.639)	4.839 (1.19)	2.877 (1.004)
# of Speeches by Speaker	265.337 (673.202)	767.123 (947.694)	288.467 (656.471)	1083.292 (1322.65)	270.242 (86.52)	870.969 (325.917)
Word count	181.414 (344.023)	118.199 (143.18)	145.903 (204.961)	92.479 (77.872)	143.231 (246.359)	101.744 (104.894)
Number of Sessions	278.172 (77.777)	267.548 (76.549)	296.33 (89.517)	299.351 (97.858)	281.536 (81.464)	276.714 (85.928)
Congressional Tenure	5.285 (4.37)	9.494 (7.133)	7.899 (6.363)	10.665 (8.77)	6.0912 (5.288)	9.956 (7.992)
Democrats	.496 (.5)	.433 (.495)	.473 (.499)	.485 (.5)	.487 (.499)	.453 (.497)
Republicans	.524 (.499)	.538 (.498)	.517 (.499)	.501 (.5)	.495 (.5)	.524 (.499)
Other	0.023 (.147)	0.028 (.166)	0.009 (.096)	0.013 (.112)	.017 (.132)	.022 (.147)
Total Seats	359.26 (37.674)	87.248 (6.792)	437.989 (.302)	96 (0)	390.842 (48.324)	90.629 (6.817)
Majority Seats	210.589 (36.926)	49.171 (7.499)	249.781 (32.555)	53.641 (3.902)	226.379 (40.132)	50.898 (6.717)
Minority Seats	140.049 (24.442)	35.678 (4.666)	181.421 (32.219)	41.661 (3.712)	156.717 (34.448)	37.99 (5.212)
# of Speakers per Session	32.166 (11.87)	22.803 (8.321)	48.829 (9.784)	32.707 (4.713)	38.879 (13.769)	26.629 (8.621)
Observations	6,459	1,623	4,358	1,022	10,817	2,645

Notes: Descriptive statistics of the main variables used in the analysis, divided between pre- and post-amendment periods. The pre-amendment period goes from the 93rd to the 103rd Congress, while the post-amendment period goes from the 104th to the 116th Congress. The sample includes both House and Senate data. Standard deviations in parentheses.

traditional DiD strategy that compares values for both chambers before and after the reform, we compute similarity as a single value that captures the relative closeness of Senate agendas with the House, embedding the first difference directly into the response variable. The model specification is as follows:

$$\text{Similarity}_{l,it} = \alpha_i + \beta \text{Post17}_t + X'_{l,it} \gamma + \mu_l + \epsilon_{l,it} \quad (1.2)$$

where, $\text{Similarity}_{l,it}$ represents our measure of similarity, for senator l on seat i during Congress t . Post_t is a binary variable equal to 1 starting with the 61st Congress (i.e., 1909-1911), marking the onset of Congressional debates surrounding the 17th Amendment. α_i denotes the seat-level fixed effects while μ_l represents individual fixed effects for each legislator.

We complement this analysis by estimating the effects of the reform on additional dimensions of legislative behavior, employing a standard DiD approach with the following general specification:

$$Y_{l,it} = \alpha_i + \delta_t + \beta \text{Senator}_{l,i} \times \text{Post17}_t + X'_{l,it}\gamma + \mu_l + u_{l,it} \quad (1.3)$$

where $Y_{l,it}$ denotes an outcome of interest for legislator l in seat i during Congress t , which may be: (1) the share of speeches dedicated to a specific policy topic, (2) the degree of concentration in policy agendas, or (3) the absolute value of the W-NOMINATE score. In this case, $\text{Senator}_{l,i}$ is a binary variable that takes the value of 1 if the legislator l in seat i is a senator, while α_i denotes seat-level fixed effects and δ_t represents Congress fixed effects. In both equations 1.2 and 1.3, we include a vector of covariates X_{it} such as re-election bids, end-of-term status, and Congress-State trends are included when relevant. In addition to the overall effect estimate, we also run event study analyses to evaluate pre-trends and dynamic effects of the reform. We cluster standard errors by state.

As discussed in Section 1.2, we incorporate an anticipation effect of one Congress to account for Senators potentially adjusting their behavior and strategies before the 17th Amendment’s ratification. This accounts for the likelihood that the momentum for reform, which gained traction well before its formal implementation, may have influenced legislative conduct and voting patterns during the transition period.

1.4 Results

Columns (1) to (3) of Table 1.3 examine the effect of the reform on the similarity measure between Senators’ and House representatives’ policy platforms, showing a statistically significant positive effect starting from the 61st Congress. The fully specified model in column (3), which includes seat level, time, and individual fixed effects, as well as additional controls and linear time trends, shows a coefficient of 0.039 (st. error=0.010).

Table 1.3: Platform Similarity Regression Results

	Platform Similarity		
	(1)	(2)	(3)
Post	0.056*** (0.005)	0.047*** (0.009)	0.039*** (0.010)
Seat FE	✓	✓	✓
Congress FE			
Individual FE		✓	✓
Controls		✓	✓
Linear Time Trend			✓
<i>N</i>	1,861	1,784	1,784

Notes: The table shows the results for Equation 1.2. The dependent variable is Cosine Similarity. Robust standard errors in parentheses, clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

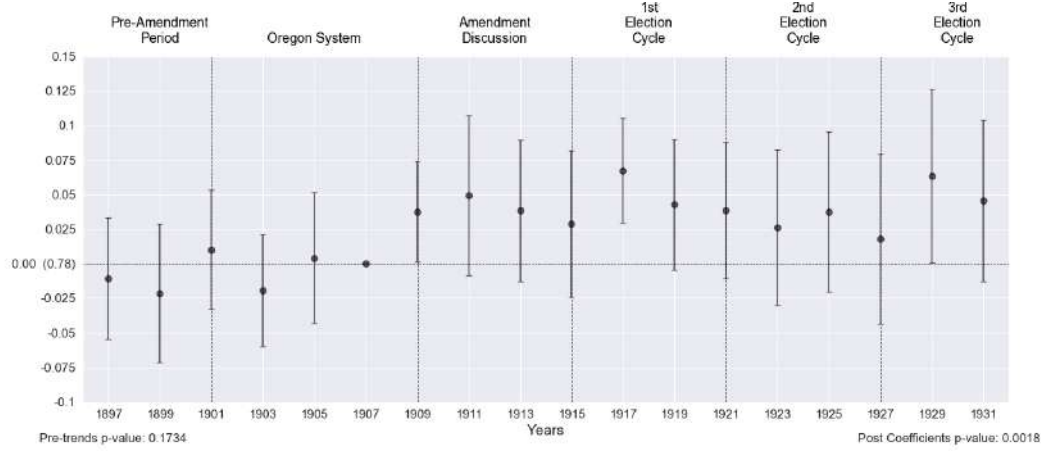
This implies an increase in similarity compared to the pre-reform period of 4.8%.

Next, we employ an event study specification to examine the dynamic effects of direct elections on platform similarity and to test for potential pre-trends. Figure 1.6 displays the results of this analysis. In the regression we use the 60th Congress as the reference year, as it was the Congress immediately preceding the extensive discussions of the reform in the 61st Congress, allowing us to account for potential anticipation effects. We find no evidence of pre-trends, as the null hypothesis of no pre-reform effects cannot be rejected (p -value = 0.17). Additionally, the post-reform coefficients are jointly significant, confirming the impact of the reform. Interestingly, the effect begins immediately with the 61st Congress and remains stable across all election cycles following the ratification of the 17th Amendment, suggesting a structural shift in Senators' policy platforms.

Next, we want to study which policy areas played the largest role in bringing Senators' political platforms closer to those of House representatives after the 61st Congress. To achieve this, we estimate Equation (1.3) separately for each policy area, using the share of speeches dedicated to specific topics as the dependent variable. The results are reported in Figure 1.7.

The results outline some topics that gained or lost momentum after the 61st Congress. Topics such as Elections and Voting, Labor and Economic Issues, Fiscal Policy, and Textiles and Manufacturing gained significant prominence, reflecting the growing importance

Figure 1.6: Cosine Similarity



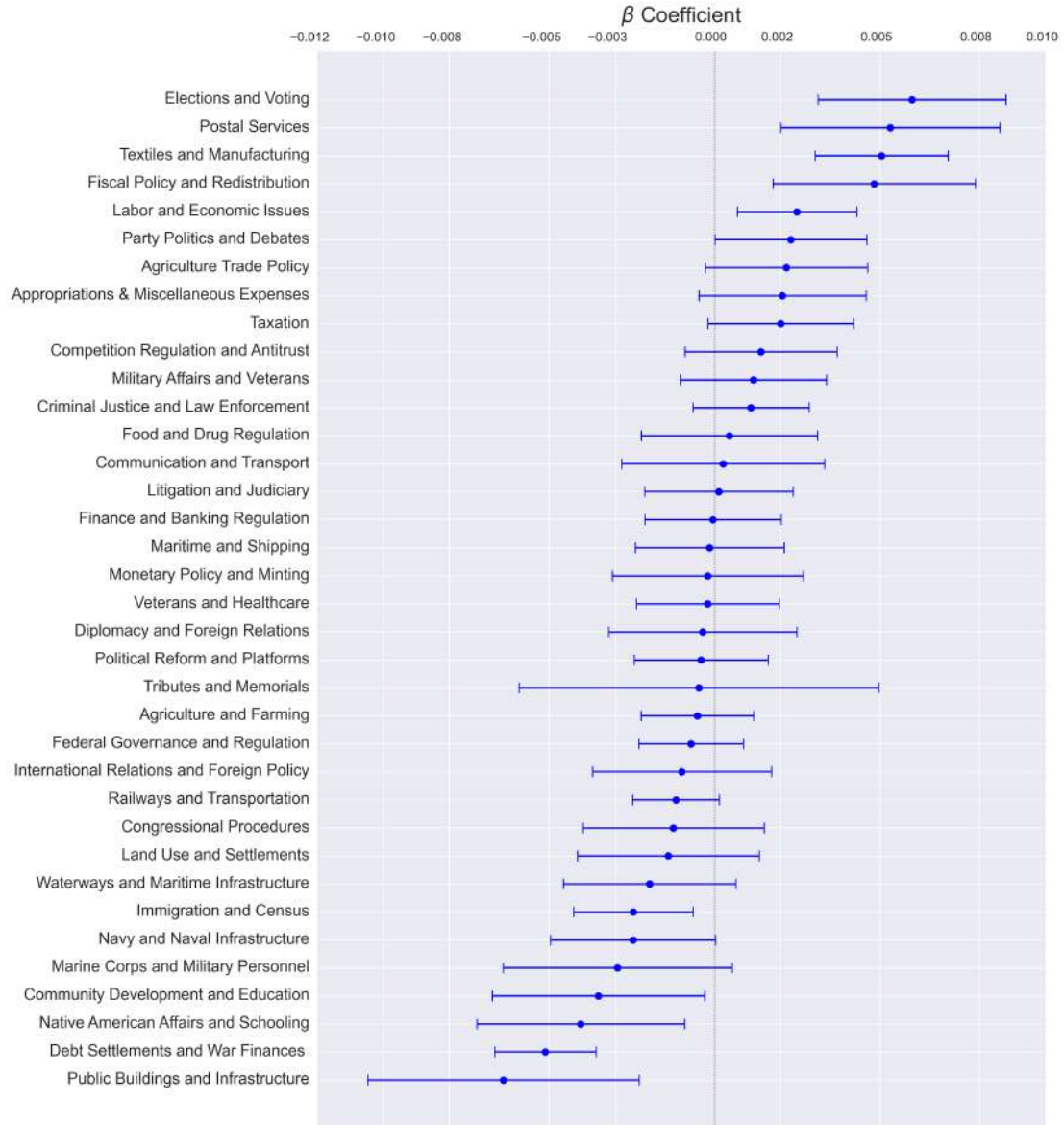
Notes: The figure shows the event-study coefficients for legislators' platform similarity along with 95% confidence intervals. The dependent variable is Cosine Similarity. Standard errors are clustered at the state level.

of issues that directly impacted the average voter. These areas address economic stability, democratic participation, and the everyday concerns of working- and middle-class citizens, signaling a recalibration of legislative agendas toward broader public interest.

At the same time, topics that lost prominence—such as Infrastructure Building, Native American Affairs, Immigration, and Debt Settlements—offer insight into the trade-offs introduced by the new electoral dynamic. These issues often require long-term planning, involve specialized constituencies, or resonate less immediately with a broad electorate. For instance, while infrastructure and community development are critical for sustained growth, they lack the immediacy and visibility of economic issues that directly influence voters' lives. Similarly, the diminished attention to Native American Affairs and Immigration reflects a shift away from niche or regional issues toward policy areas that appealed to a larger, more generalized voter base.

To assess the reform's dynamic impact on policy discussions, we conduct a binned event study, with the results for key topics visualized in Figures A.5 to A.15. This approach groups Congresses into four-session periods to smooth out noisy fluctuations, providing a clearer view of the underlying trends. We specify Congress 60 as the baseline against which treatment effects in other periods are measured as in the other event studies displayed. With the only exception of A.12, the coefficients for the pre-reform periods are statistically insignificant, confirming the absence of pre-existing trends.

Figure 1.7: β Coefficient of the Reform on Topic Relevance



Notes: The figure shows the coefficient for Equation (1.3), where the dependent variable is the relevance of a given topic in legislators' political platforms. Standard errors are clustered at the state level.

1.5 Additional Results and Robustness Checks

1.5.1 Political Selection and Senate Primaries

One possible explanation for the observed shifts in the main results is a change in the composition of the political class. The newly elected senators brought into office through the mechanisms introduced by the 17th Amendment may differ significantly in their backgrounds, priorities, and platforms compared to their predecessors. These differences in the senators themselves could directly influence the legislative agenda and the topics emphasized in Congress.

To further investigate this, we focus on understanding the extent to which the observed effects are driven by changes in the behavior of “incumbents”—congressmen who served both before and after being subjected to the new electoral framework—rather than the entry of entirely new senators. While newly elected senators, brought in during or after the process leading to the 17th Amendment, may have inherently different platforms, it is critical to assess whether the shifts in legislative priorities can also be attributed to incumbents adapting their strategies to the new political dynamics.

As shown in Table 1.4 (Panel A), there is a significant effect on platform similarity even when limiting the sample to congressmen who served across both the pre-and post-reform systems, indicating that the changes occurred in addition to the possible changes in the composition of the Senate due to the arrival of newcomers. This result strongly suggests that senators who experienced both systems have adapted their original platform to be closer to that of the House of Representatives. In Appendix Figure A.18, the event-study for this model specification illustrates this adaptation. The absence of pre-trends, the joint significance of post-event coefficients at the 1% confidence level, and the clear pattern of change confirm that even long-serving senators altered their policy platforms to accommodate the expectations of their new constituencies under the direct election framework.

A potential concern is that the observed effects might be primarily driven by states that had already implemented some form of primaries for Senate elections prior to the 61st Congress, as highlighted in Figure A.1. This raises the question of whether the early adoption of primaries played a significant role in driving the changes in platform similarity, potentially confounding the broader impact of the 17th Amendment. To address this, we run the main analysis by excluding states that had adopted primaries

Table 1.4: Heterogeneity Analysis - TWFE Results for Platform Similarity

	Platform Similarity		
	(1)	(2)	(3)
Panel A: Incumbents, serving both before and after the reform			
Post	0.047*** (0.008)	0.048*** (0.009)	0.038** (0.027)
N	721	718	718
Panel B: States without pre-reform Senate primaries			
Post	0.051*** (0.006)	0.049*** (0.010)	0.039** (0.036)
N	1,339	1,281	1,281
Seat FE	✓	✓	✓
Congress FE			
Individual FE		✓	✓
Controls		✓	✓
Linear Time-trend			✓
Panel C: All States prior to 61st Congress			
Primary $\times$ Post	0.016 (0.015)	0.012 (0.016)	0.012 (0.017)
N	1,129	1,082	1,082
Seat FE	✓	✓	✓
Congress FE	✓	✓	✓
Individual FE		✓	✓
Controls		✓	✓
Linear Time-trend			✓

Notes: The table shows the results of regression 1.2 including only senators that were in Congress before and after the 61st Congress (Panel A) and excluding states that had primaries for Senate's election prior to the 61st Congress (Panel C). The dependant variable is Cosine similarity. Panel C also includes Congress fixed-effects, using states that did not have primaries until the 61st Congress. Robust standard errors in parenthesis, clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

before the 61st Congress and re-estimate Equation (1.2). As reported in Table 1.4 (Panel B), and graphically in Appendix Figure A.19, the results remain consistent with our central findings, indicating that the presence of early primaries is not a significant driver of the observed effects.

Furthermore, we investigate whether using the implementation of primaries as the treatment itself would yield similar results. We test this using a staggered DiD design

on a sample composed exclusively of senators, considering all congresses before the 61st, where senators from states that adopted primaries form the treated group. The analysis for this specification, presented in Table 1.4 (Panel C) and Appendix Figure A.20, shows no discernible effect when primaries are used as the treatment variable. This suggests that primaries, on their own, are not sufficient to explain the observed shifts in platform similarity. Instead, the changes appear to be tied more fundamentally to the broader institutional reform brought about by the 17th Amendment.

1.5.2 Topic Concentration

We provide evidence of the reform’s effect on the concentration of Senators’ policy agendas. To examine these shifts, we use a measure of topic entropy to determine whether policy platforms became more focused on a few key issues or more evenly distributed across a broader range of topics after the reform. One possible reason for such changes is that, under the pre-amendment regime, significant malapportionment within state legislative electoral districts gave less populated districts disproportionate influence in Senate elections. With the adoption of direct elections, this over-representation ceased to play a role, potentially reshaping Senate campaigns and leading to broader or more concentrated policy platforms.

Table 1.5: Additional Results: Concentration of Policy Platform

	Entropy		
	(1)	(2)	(3)
Senator $\times$ Post	-0.027* (0.015)	-0.046** (0.022)	-0.048** (0.023)
Seat FE	✓	✓	✓
Congress FE	✓	✓	✓
Individual FE		✓	✓
Controls		✓	✓
Linear Time Trend			✓
N	9,573	8,709	8,709

Notes: The table shows the results of Equation 1.3 for concentration of the policy platform, measured by entropy. Robust standard error in parenthesis, clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.5 reports the results of estimates based on Equation (1.3), where we use as a

dependent variable a measure of the entropy of the LDA output. Across all specifications, the treatment coefficient is negative and statistically significant, suggesting that Senators narrowed their focus to fewer policy areas after the reform. These results could indicate that direct elections moved Senators to adjust their platforms to reflect the priorities of a more evenly distributed electorate, reducing the influence of over-represented districts. In Appendix Figure A.16, we report the event-study analysis. It shows that since the 61st Congress, there has been a long-lasting effect on how Senators focused on fewer policy issues in their policy platforms and reject the null hypothesis of pre-trends (p-value 0.44).

A stylized simulation presented in Appendix C illustrates how such a shift can arise mechanically from changes in the aggregation of voter preferences. When elections move from a system that weights districts equally to one that weights individuals equally, equilibrium platforms tend to become more concentrated, especially when a small number of populous districts coexist with many sparsely populated ones. While purely illustrative, the model shows how the removal of malapportionment under direct elections can naturally generate the decline in entropy observed in the data.

1.5.3 Roll-call Behavior

Table 1.6: W-Nominate Regression Results

	W-NOMINATE		
	(1)	(2)	(3)
Senator $\times$ Post	-0.003 (0.006)	-0.000 (0.004)	-0.000 (0.001)
Seat FE	✓	✓	✓
Congress FE	✓	✓	✓
Individual FE			
Controls		✓	✓
Linear Time Trend			✓
N	9,529	9,529	9,529

Notes: The table shows the results for Equation 1.3. The dependent variable is the absolute value of the DW-Nominate score. Robust standard errors in parentheses, clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.6 presents the regression results for 1.3, offering further evidence on the

reform’s impact on roll-call behavior. The dependent variable is the absolute value of legislators’ W-NOMINATE scores, used to test the moderation hypothesis proposed by [Bernhard and Sala \(2006\)](#) and to provide additional insights into the reform’s effects. These results do not reveal statistically significant effects of direct elections on roll-call behavior, suggesting that Senators’ voting patterns were less affected by the reform than their policy agendas. Appendix Figure [A.17](#) illustrates the event study for this outcome, corroborating the overall null effect. This finding contrasts with [Bernhard and Sala \(2006\)](#), who argued that the 17th Amendment moderated Senators’ ideological stances; the discrepancy may reflect our use of a different identification strategy.¹¹

These results align with evidence from [Gentzkow et al. \(2019\)](#), which indicates that voting and speeches in Congress fulfill different functions: speeches offer legislators a flexible way to convey their priorities and connect with voters, whereas voting is shaped by procedural constraints, party dynamics, and negotiations. In this context, the absence of a detectable effect on roll-call behavior is consistent with the idea that electoral incentives primarily reshape the rhetorical rather than the legislative margin of political action.

1.5.4 Robustness checks

In this section, we apply robustness checks to assess the validity of our results. First, we examine whether our findings are influenced by the choice of LDA parameters, specifically the number of topics. To test this, we select the best-performing models with comparable performance, considering three additional models with 30, 33, and 39 topics, respectively. We re-estimate Equation (1.2) using topic concentration measures as response variables, and the results are presented in Table [A.3](#). Similarly, Figure [A.22](#) illustrates the event-studies for platform similarity across these specifications. Both the table and the graphical analysis confirm the robustness of our results: the estimates remain consistent with the original specification, all post-treatment coefficients are statistically significant, and there is no evidence of pre-trends. These findings suggest that our conclusions are stable across different LDA configurations.

Second, we check whether our results hold using alternative measures of platform similarity, namely Jensen-Shannon divergence, Kullback-Leibler divergence, and Bhat-

¹¹Interestingly, Figure [1.3](#) shows some convergence of ideological scores toward zero, but this trend appears to have begun prior to the 61st Congress, suggesting that other factors may have shaped legislators’ voting behavior during this period.

tacharyya distance. As mentioned in Section 3.2, to avoid confusion, we employ the following transformation to ensure that these measures represent similarity rather than dissimilarity: $\frac{1}{1+\text{Divergence}}$. Table A.2 presents the results, showing a consistent positive and significant effect on platform similarity. Additionally, Figure A.21 illustrates the event-study results for each similarity measure, further confirming the robustness of our findings.

Finally, we examine whether our results are driven by any specific state. To test this, we run the main regression multiple times, each time excluding one state from the analysis. This leave-one-out approach allows us to assess the stability of our findings across the sample. The results reported in Appendix Figure A.23 indicate that our findings remain consistent and robust, showing no significant variation regardless of which state is excluded. This confirms that our conclusions are not disproportionately influenced by any single state.

1.6 Conclusion

In this paper, we study how the 17th Amendment, which introduced direct elections for U.S. Senators, influenced their behavior and priorities. Using a comprehensive dataset of over 6.5 million congressional speeches and roll-call votes from 1879 to 1935, we combine NLP techniques with a difference-in-differences design to isolate the causal effects of the reform. This framework allows us to assess changes in both public-facing political discourse and legislative behavior by comparing Senators to House Representatives, who were already directly elected during this period.

Our results reveal that the shift to direct elections significantly altered Senators’ political agendas, bringing them closer to the focus of House Representatives. Senators placed greater emphasis on fiscal policy, taxation, and democratic representation, while devoting less attention to infrastructure, immigration, and fiscal discipline. These changes are not driven by compositional turnover: incumbents who served both before and after the reform also adjusted their rhetorical priorities. We further document a decline in the entropy of Senators’ agendas, indicating a narrower and more concentrated set of policy concerns. By contrast, we find no corresponding shift in their roll-call behavior, consistent with the idea that speeches are more flexible and responsive than votes, which remain constrained by institutional rules and party dynamics.

Taken together, these results speak to broader debates in theories of representative

democracy. They suggest that electoral reforms may reshape political representation primarily through changes in how politicians communicate with voters, even when their formal legislative behavior remains stable. In this sense, the move to direct elections appears to have strengthened *discursive* forms of accountability: senators increasingly tailored their public rhetoric to the preferences of the mass electorate, narrowing their agendas and emphasizing issues salient to a broader constituency. This pattern implies that representation is not only enacted through policy choices or roll-call positions, but also through the way political actors structure and prioritize the issues they present to citizens. Our evidence therefore highlights a channel of democratic responsiveness that is often overlooked in institutional analyses—namely, the adaptation of rhetorical agendas in response to changes in electoral incentives. Future work could examine whether these rhetorical shifts were accompanied by corresponding changes in policymaking by analyzing the textual content of bills, thereby shedding light on the extent to which discursive responsiveness translates into substantive legislative action.

Overall, our findings show that institutional reforms can meaningfully reshape the communicative dimension of political representation, even in the absence of major ideological or behavioral shifts. This perspective opens new avenues for research on how electoral systems influence not only what legislators do, but also how they articulate and frame their policy priorities to the public.

Bibliography

- Alesina, A. (1995). *Partisan Politics, Divided Government, and the Economy*. Cambridge University Press.
- Ash, E., Krümmel, J., and Slapin, J. B. (2024). Gender and reactions to speeches in german parliamentary debates. *American Journal of Political Science*.
- Bernhard, W. and Sala, B. R. (2006). The Remaking of an American Senate: The 17th Amendment and Ideological Responsiveness. *The Journal of Politics*, 68(2):345–357.
- Besley, T. and Case, A. (1995). Does electoral accountability affect economic policy choices? evidence from gubernatorial term limits. *The Quarterly Journal of Economics*, 110(3):769–798.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Bordignon, M., Nannicini, T., and Tabellini, G. (2016). Moderating political extremism: single round versus runoff elections under plurality rule. *American Economic Review*, 106(8):2349–2370.
- Bybee, J. S. (1997). Ulysses at the mast: Democracy, federalism, and the sirens’ song of the seventeenth amendment. *Northwestern University Law Review*, 91:500.
- Congress, U. S. (2023). The legislative process. Accessed: 2023-08-03.
- Durfee, E. (1945). Apportionment of representation in the legislature: A study of state constitutions. *Michigan Law Review*, 43(6):1091–1112.
- Federal Reserve History (2021). Carter glass. Accessed: 2024-11-19.
- Ferejohn, J. (1999). Accountability and authority: toward a theory of political accountability. *Democracy, accountability, and representation*, 131:133.
- Ferlenga, F. (2023). Better to be direct? evidence from the abolition of direct elections in italian local governments. *Economics & Politics*, 35(3):822–891.
- Gagliarducci, S., Nannicini, T., and Naticchioni, P. (2011). Electoral rules and politicians’ behavior: a micro test. *American Economic Journal: Economic Policy*, 3(3):144–174.
- Gailmard, S. and Jenkins, J. A. (2009). Agency problems, the 17th amendment, and representation in the senate. *American Journal of Political Science*, 53(2):324–342.
- Gennaro, G. and Ash, E. (2022). Emotion and reason in political language. *The Economic Journal*, 132(643):1037–1059.

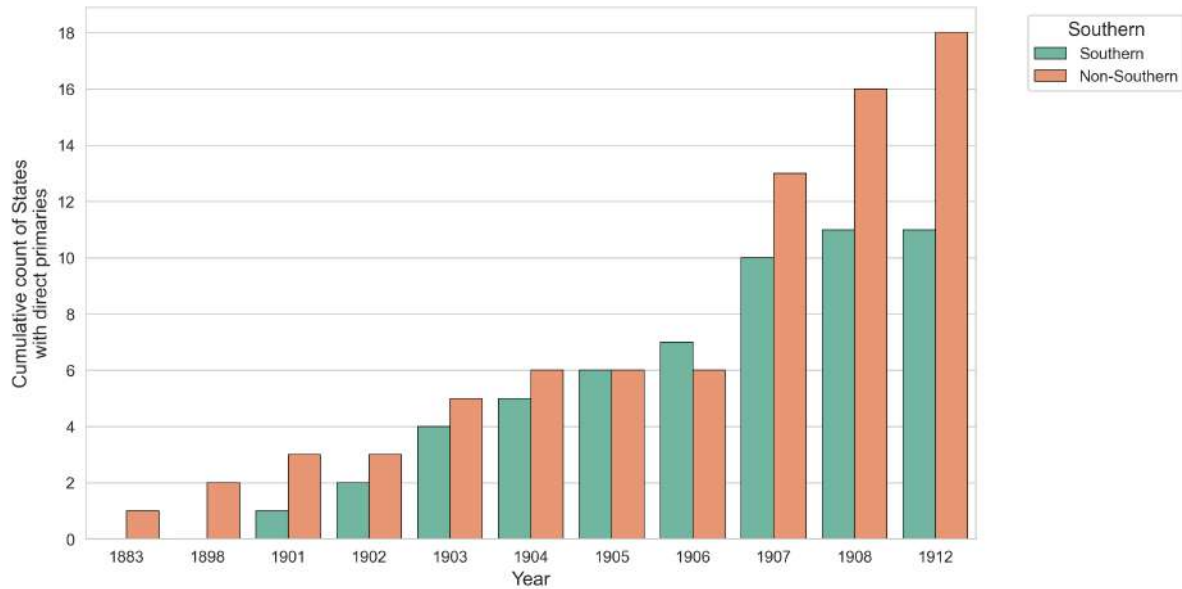
- Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340.
- Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and deliberation within the fmc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870.
- Haynes, G. H. (1938). *The Senate of the United States: Its History and Practice*. Russell & Russell, New York.
- Hoffman, M., Bach, F., and Blei, D. (2010). Online learning for latent dirichlet allocation. *advances in neural information processing systems*, 23.
- Iowa State University (2021). Urban population percentages by state. <https://www.icip.iastate.edu/tables/population/urban-pct-states>. Accessed: 2023-04-03.
- Jenkins, J. A. (2005). Partisanship and contested election cases in the Senate, 1789–2002. *Studies in American Political Development*, 19(1):53–74. Publisher: Cambridge University Press.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lapinski, J. S. (2003). The behavioral and institutional effects of direct election. In *Annual Meeting of the American Political Science Association*, Philadelphia, PA.
- Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., and Sonnet, L. (2023). Voteview: Congressional roll-call votes database. <https://legacy.voteview.com/dwnl.htm>.
- Lin, T.-M., Enelow, J. M., and Dorussen, H. (1999). Equilibrium in multicandidate probabilistic spatial voting. *Public Choice*, 98:59–82.
- Lindbeck, A. and Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. *Public choice*, pages 273–297.
- Lowande, K. and Peck, J. (2017). Congressional investigations and the electoral connection. *The Journal of Law, Economics, and Organization*, 33(1):1–27. Publisher: Oxford University Press.
- Micozzi, J. P. (2013). Does electoral accountability make a difference? direct elections, career ambition, and legislative performance in the argentine senate. *The Journal of Politics*, 75(1):137–149.
- National Archives (2023a). The 17th amendment to the constitution: Direct election of u.s. senators. Accessed: 2024-11-25.

- National Archives (2023b). 17th amendment to the u.s. constitution: Direct election of u.s. senators (1913). Accessed: 31-10-2023.
- Persson, T. and Tabellini, G. (2004). Constitutions and economic policy. *Journal of Economic Perspectives*, 18(1):75–98.
- Persson, T., Tabellini, G., and Trebbi, F. (2003). Electoral rules and corruption. *Journal of the European Economic Association*, 1(4):958–989.
- Poole, K. T. and Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):357–384.
- Schiller, W. and Stewart III, C. (2004). US Senate elections before 1914. In *annual meeting of the American Political Science Association*.
- Schiller, W. J. (2006). Building Careers and Courting Constituents: U.S. Senate Representation 1889–1924. *Studies in American Political Development*, 20(2):185–197.
- Schiller, W. J. and Stewart, C. (2014). *Electing the Senate: Indirect Democracy before the Seventeenth Amendment*, volume 146 of *Princeton Studies in American Politics: Historical, International, and Comparative Perspectives*, chapter 4. Princeton University Press.
- Schiller, W. J., Stewart, C., and Xiong, B. (2013). U.S. Senate Elections before the 17th Amendment: Political Party Cohesion and Conflict 1871–1913. *The Journal of Politics*, 75(3):835–847.
- Srivastava, A. and Sutton, C. (2017). Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488*.
- Steyvers, M. and Griffiths, T. (2007). Probabilistic topic models. In *Handbook of latent semantic analysis*, pages 439–460. Psychology Press.
- U.S. Congress (1911). Congressional Record: Proceedings and Debates of the 62nd Congress, First Session. <https://www.govinfo.gov/content/pkg/GPO-CRECB-1911-pt2-v47/pdf/GPO-CRECB-1911-pt2-v47-13-1.pdf>. [Accessed January 21, 2025].

A Appendix

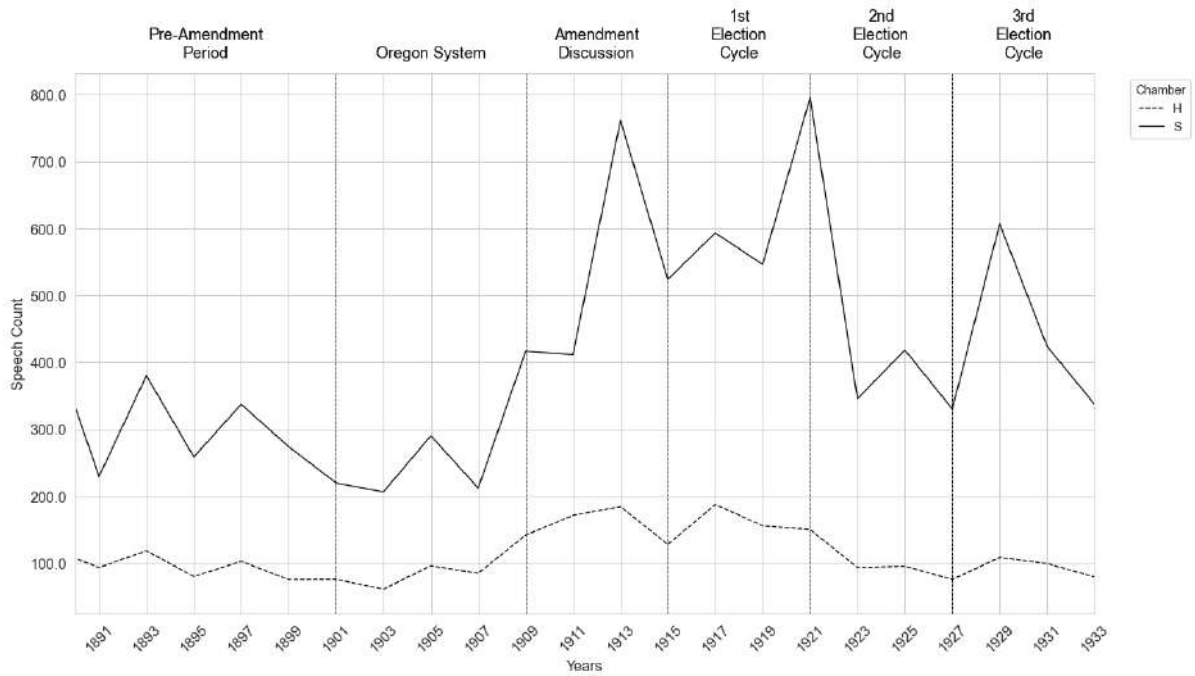
A.1 Figures

Figure A.1: Direct Primaries



Notes: The figure shows the adoption of Senate primaries by state legislatures (see 1.2). The vertical axis indicates the cumulative number of states that adopted primaries. The analysis distinguishes states based on whether they implemented Jim Crow laws to disenfranchise voters. Southern states include Alabama, Arkansas, Florida, Georgia, Louisiana, Missouri, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia. Data on Senate primaries is sourced from [Lapinski \(2003\)](#).

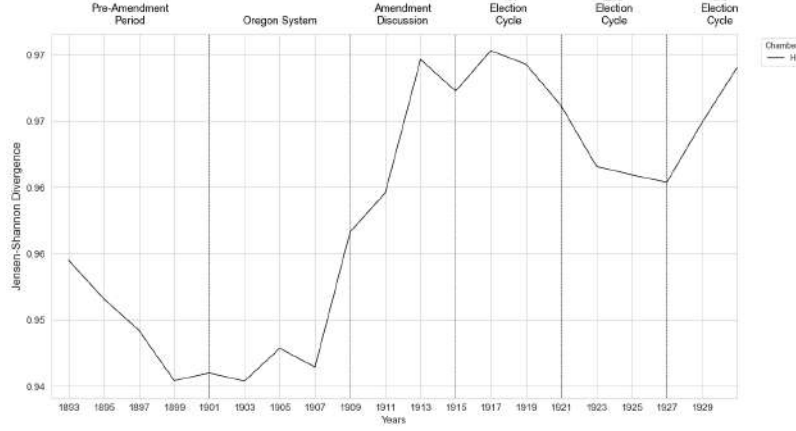
Figure A.2: Speech Counts by Chamber



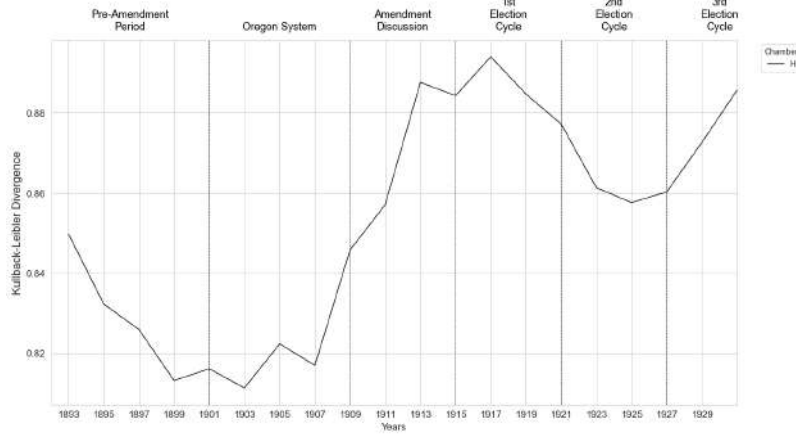
Notes: The figure depicts the average number of speeches by speaker over time. The solid line refers to the Senate and dashed line refers to the House

Figure A.3: Time Series of Alternative Policy Platform Similarity Measures

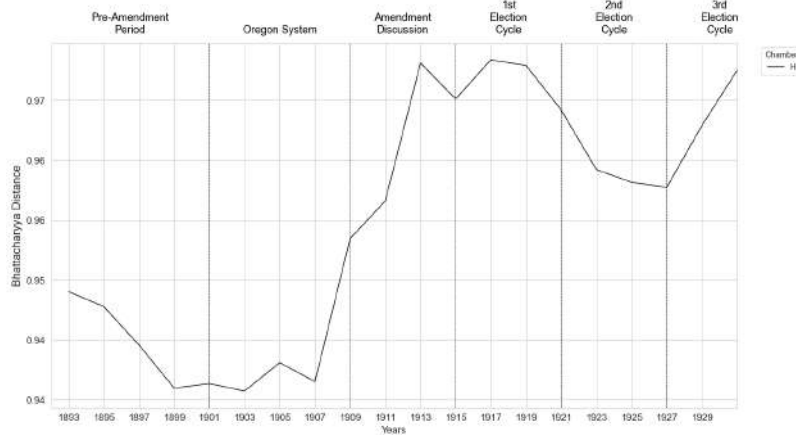
(a) Time Series of Jensen-Shannon Divergence (Inverted)



(b) Time Series of Kullback-Leibler Divergence (Inverted)



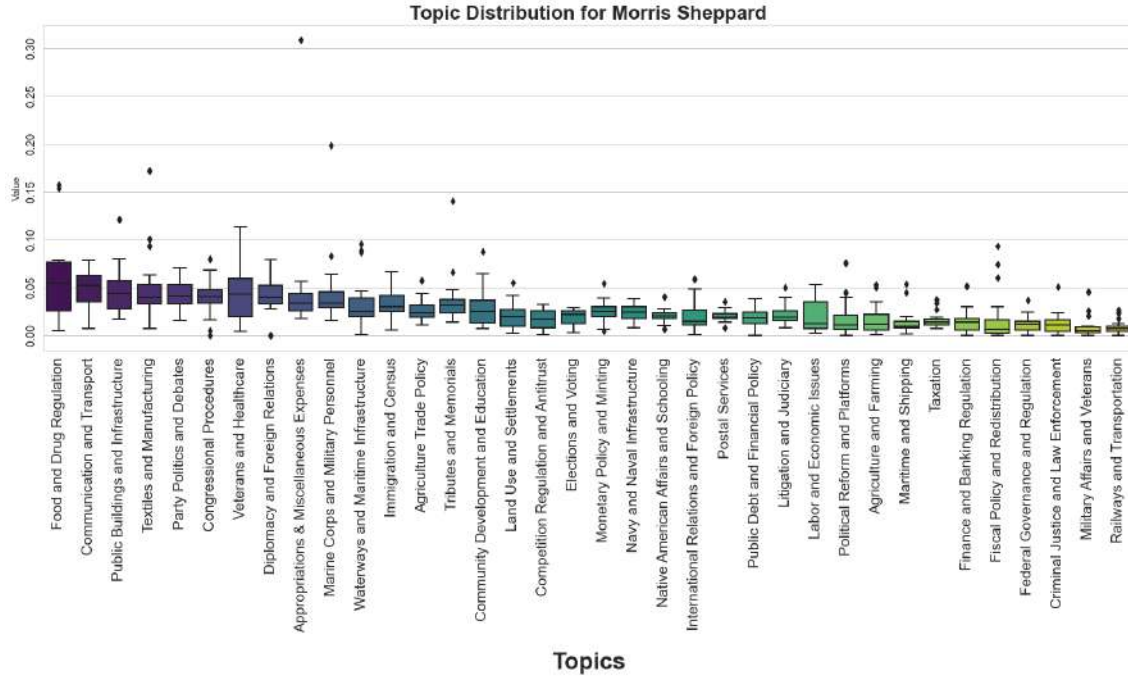
(c) Time Series of Bhattacharyya Distance (Inverted)



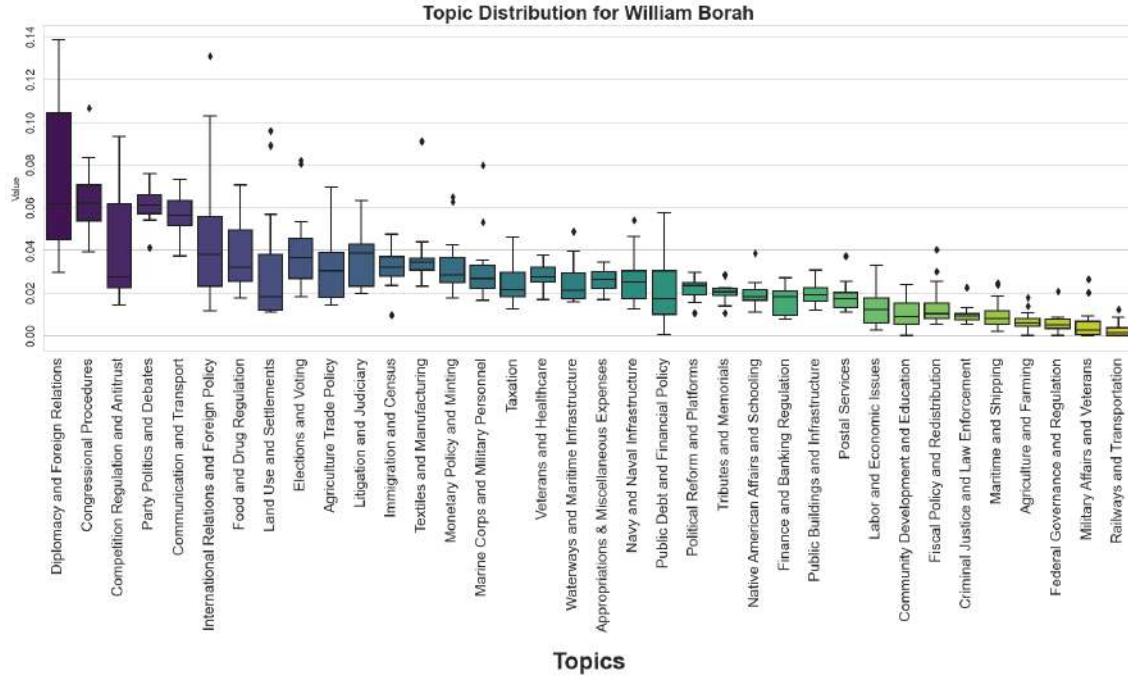
Notes: The figure shows the time series of our alternative measures of similarities. The measures have been transformed to depict similarity rather than divergence. Panel a) refers to $\frac{1}{1+\text{Jensen-Shannon Divergence}}$, Panel b) to $\frac{1}{1+\text{Kullback-Leibler Divergence}}$ and Panel c) to $\frac{1}{1+\text{Bhattacharyya Distance}}$.

Figure A.4: Policy Platform Examples

(a) Morris Sheppard

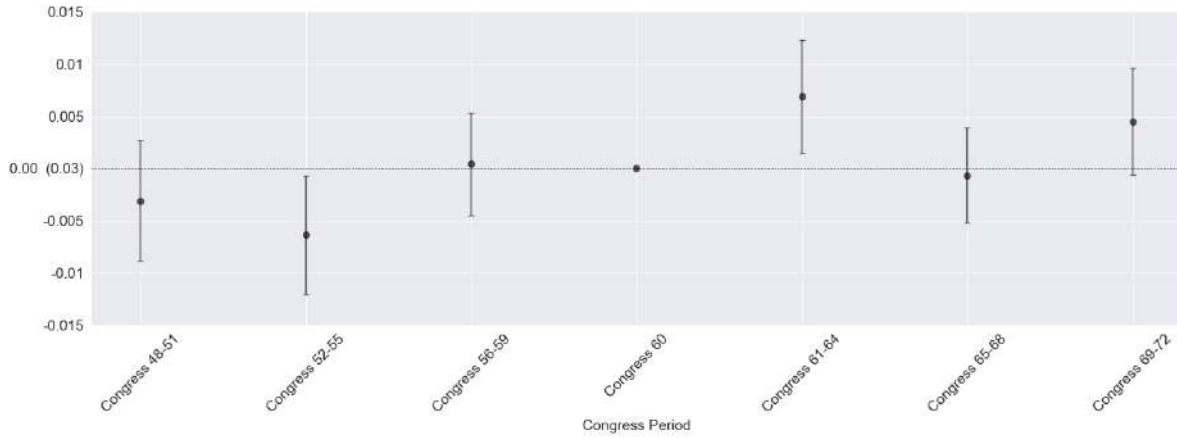


(b) William Borah



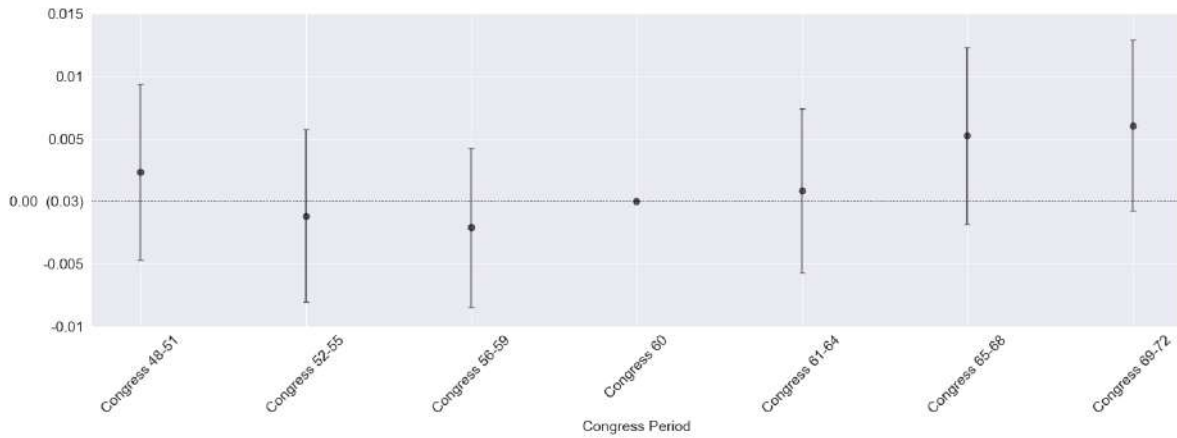
Notes: In Panel (a), we present the policy platform of William Borah, a Republican Senator known for his staunch isolationist stance. Borah opposed the League of Nations and championed anti-imperialist policies, reflecting his commitment to limiting international entanglements. In Panel (b), the policy platform of Morris Sheppard, a Democratic Congressman from Texas. Famously referred to as the "Father of National Prohibition", he sponsored the 18th Amendment, which prohibited alcohol in the United States, and supported Prohibition laws throughout the 1920s and 1930s.

Figure A.5: Event-Study: Elections and Voting Topic



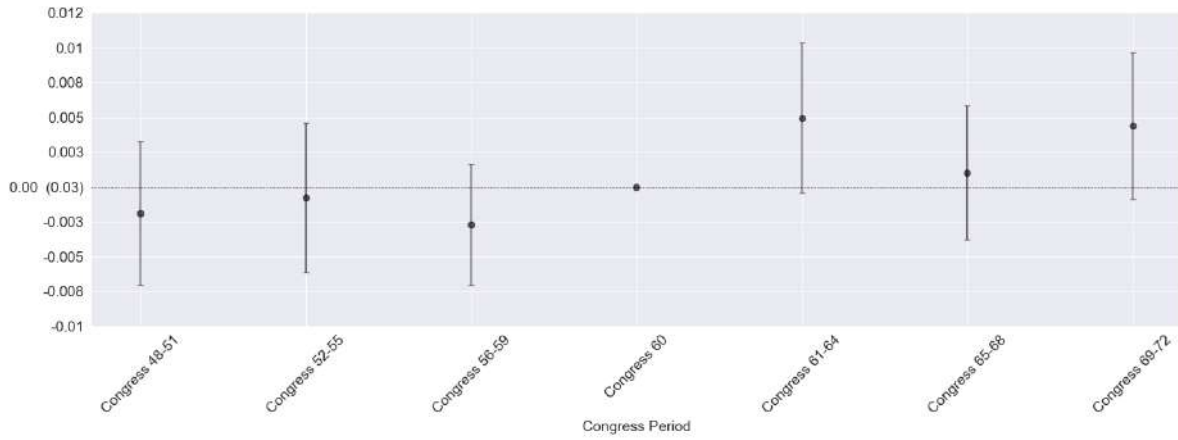
Notes: The figure shows the event-study of Equation 1.3 with Elections and Voting topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.6: Event-Study: Postal Services Topic



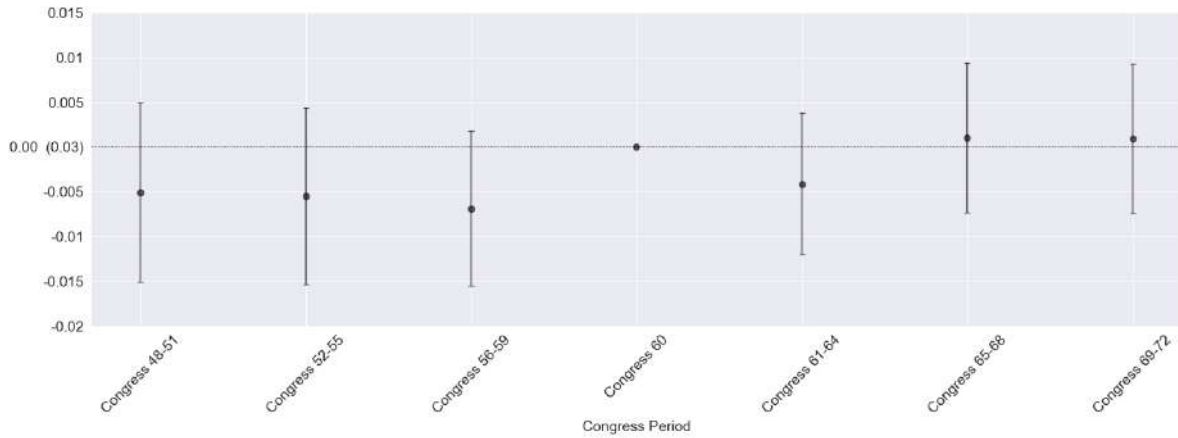
Notes: The figure shows the event-study of Equation 1.3 with Postal Services topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.7: Event-Study: Textiles and Manufacturing Topic



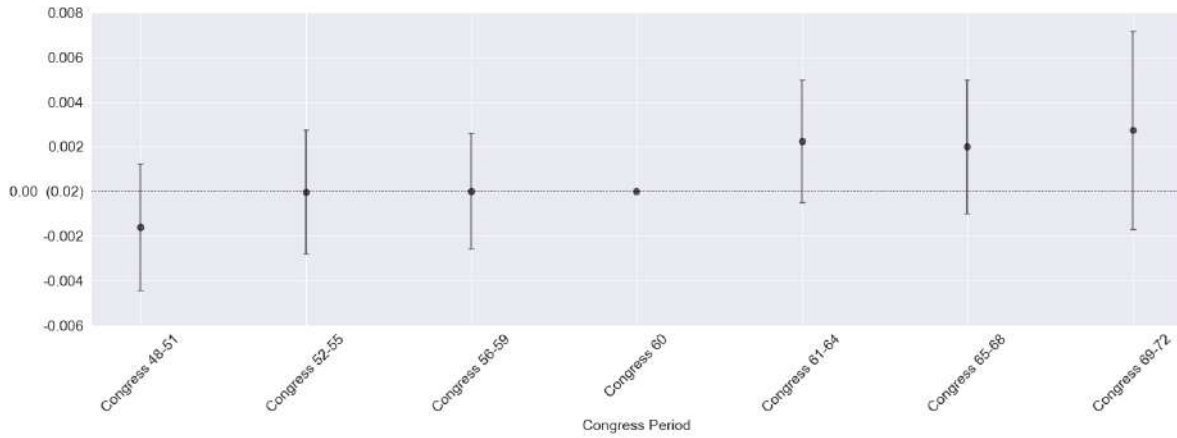
Notes: The figure shows the event-study of Equation 1.3 with Textiles and Manufacturing topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.8: Event-Study: Fiscal Policy and Redistribution Topic



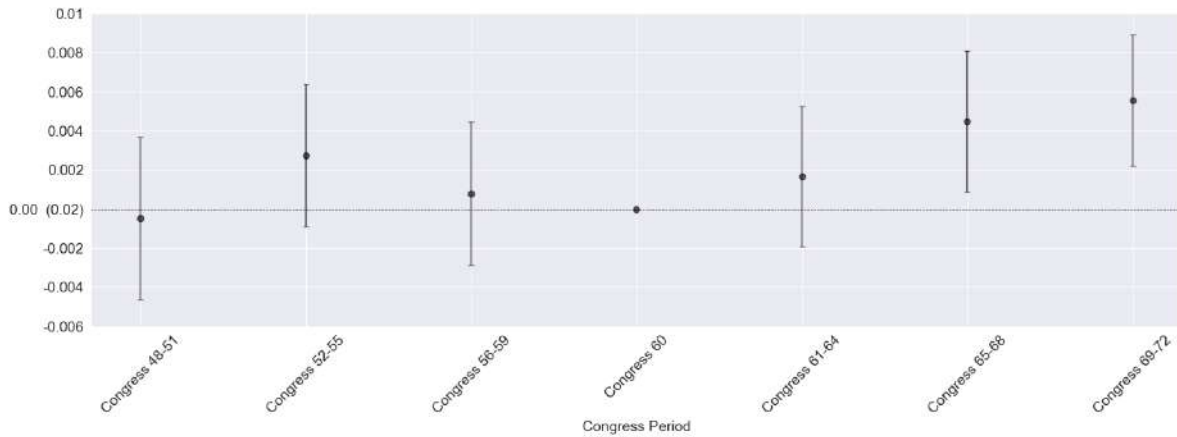
Notes: The figure shows the event-study of Equation 1.3 with Fiscal Policy and Redistribution topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.9: Event-Study: Taxation Topic



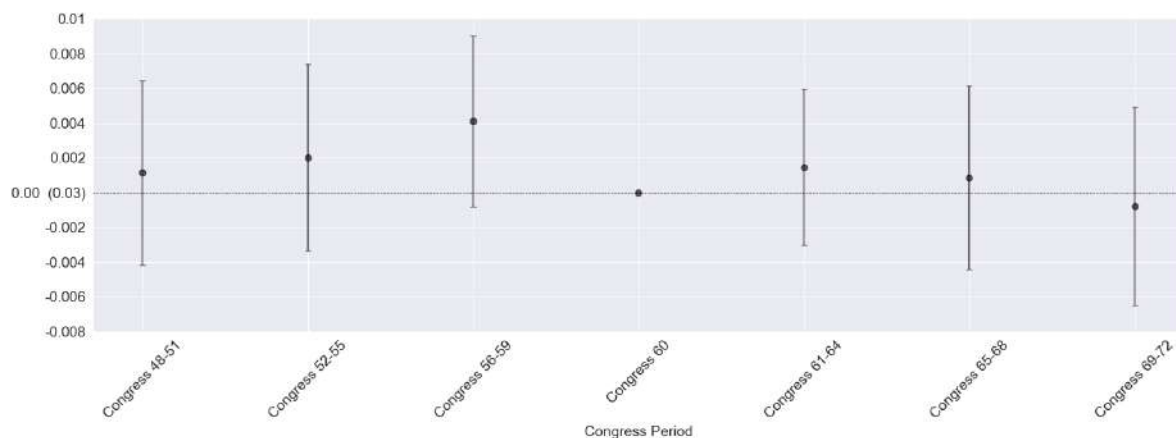
Notes: The figure shows the event-study of Equation 1.3 with Taxation topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.10: Event-Study: Labor and Economic Issues Topic



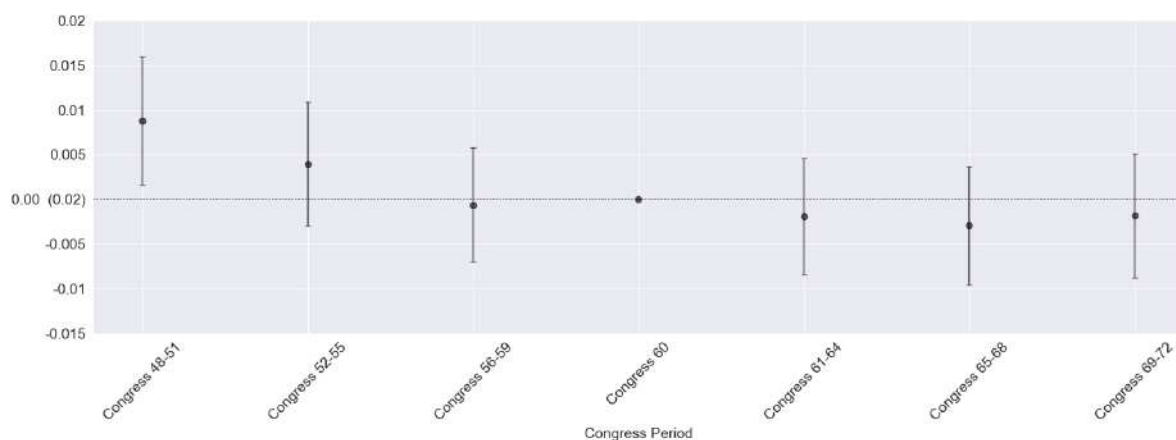
Notes: The figure shows the event-study of Equation 1.3 with Labor and Economic Issues topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.11: Event-Study: Immigration and Census Topic



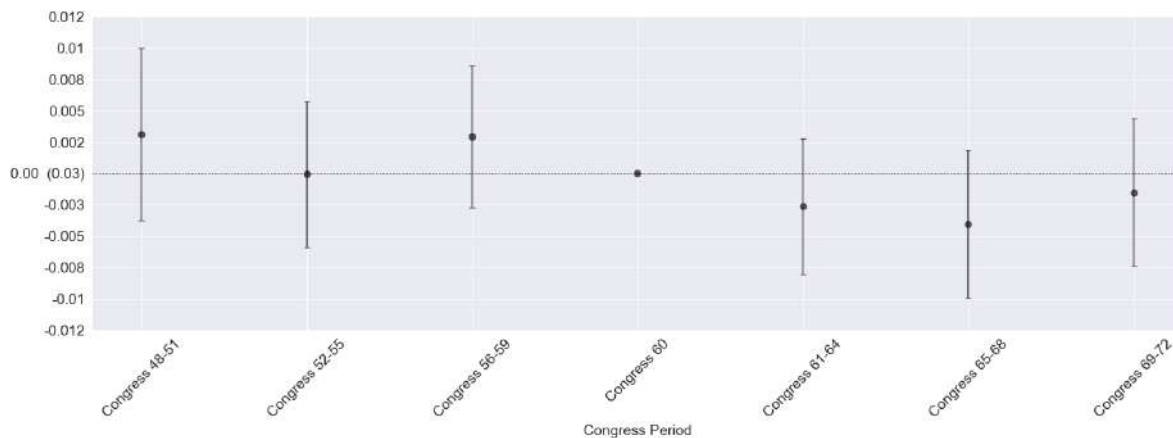
Notes: The figure shows the event-study of Equation 1.3 with Immigration and Census topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.12: Event-Study: Community Development and Education Topic



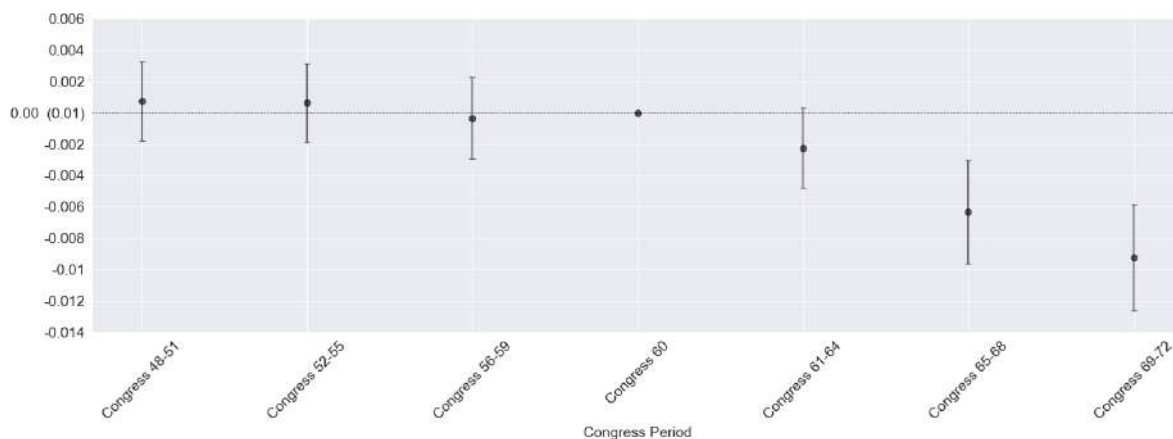
Notes: The figure shows the event-study of Equation 1.3 with Community Development and Education topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.13: Event-Study: Native American Affairs and Schooling Topic



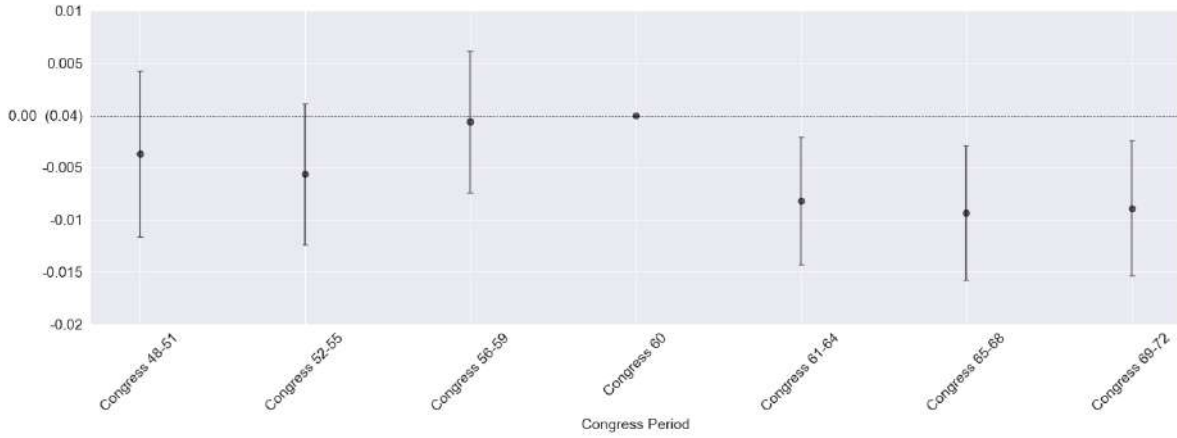
Notes: The figure shows the event-study of Equation 1.3 with Native American Affairs and Schooling topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.14: Event-Study: Debt Settlements and War Finances Topic



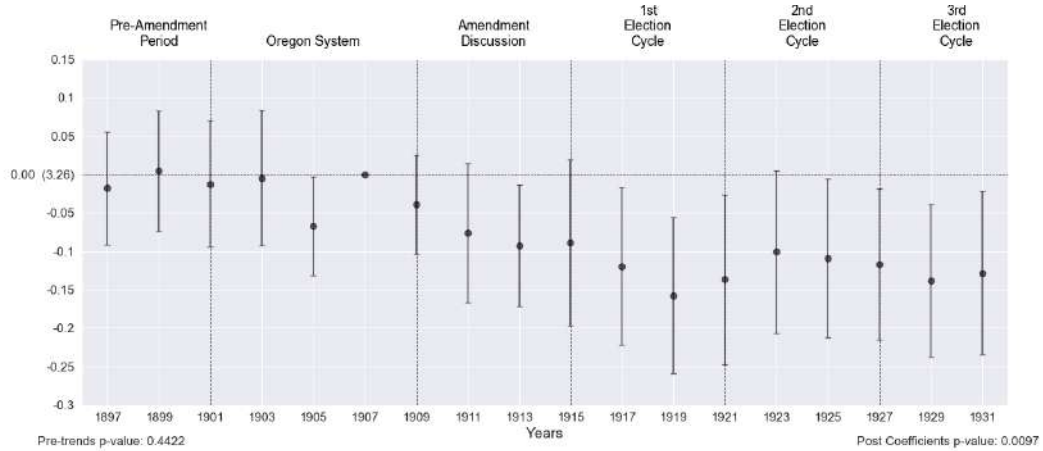
Notes: The figure shows the event-study of Equation 1.3 with Debt Settlements and War Finances topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.15: Event-Study: Public Buildings and Infrastructure Topic



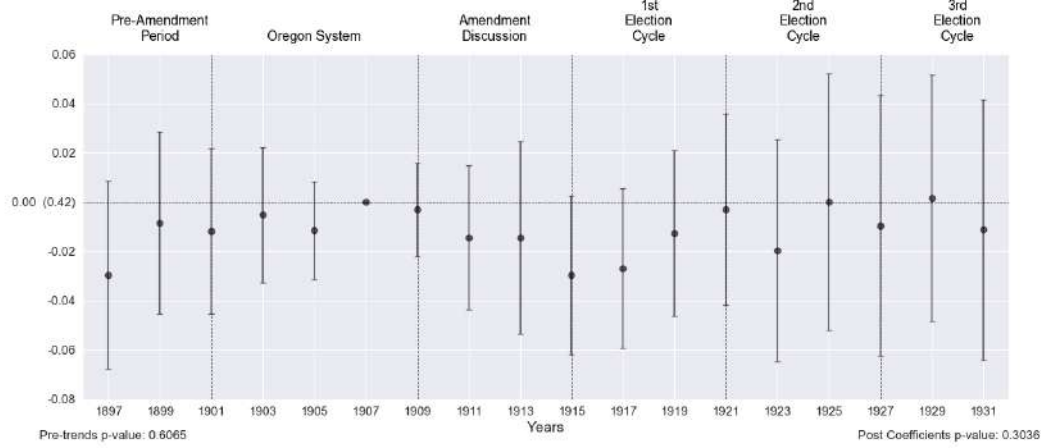
Notes: The figure shows the event-study of Equation 1.3 with Public Buildings and Infrastructure topic relevance as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.16: Event-Study: Entropy of Political Platform



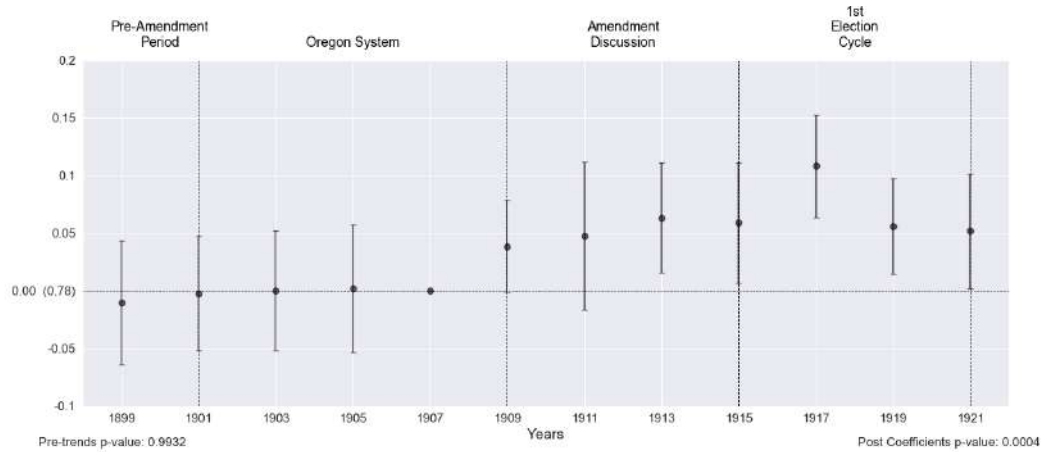
Notes: The figure shows the event-study of Equation 1.3 with concentration of the policy platform, i.e. entropy, as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.17: Event-study: Absolute Value of W-NOMINATE



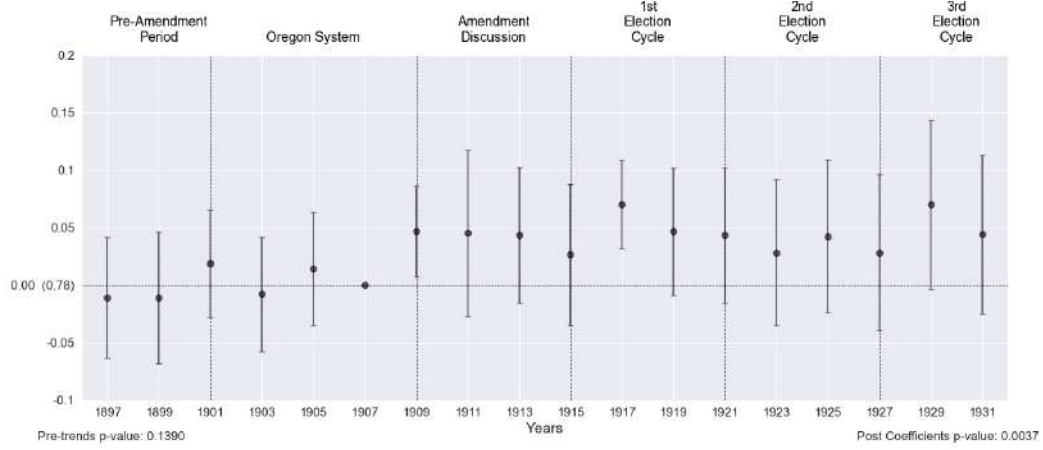
Notes: The figure shows the event-study of Equation 1.3 with W-NOMINATE as the dependent variable. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.18: Cosine Similarity of Policy Platform for Incumbents



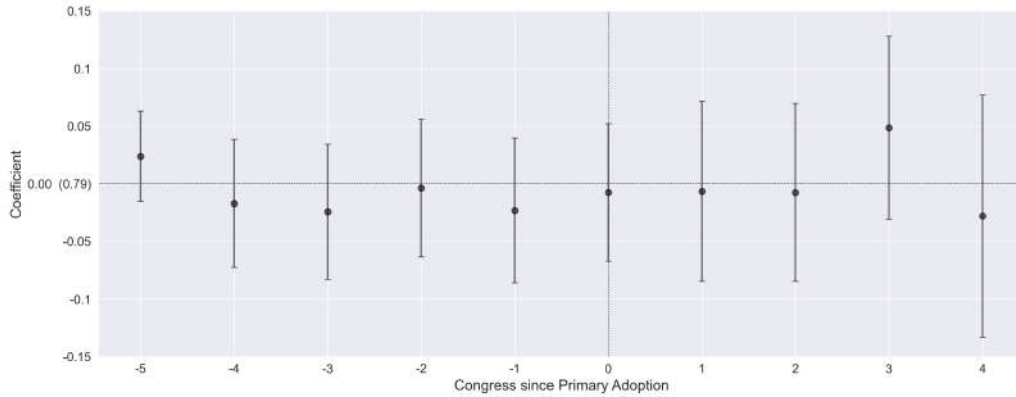
Notes: The figure shows the event-study of Equation 1.2. The dependent variable is cosine similarity. The estimation only considers incumbents, i.e. Senators that were present before and after the the discussion of the 17th Amendment. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.19: Cosine Similarities excluding states with Senate primaries



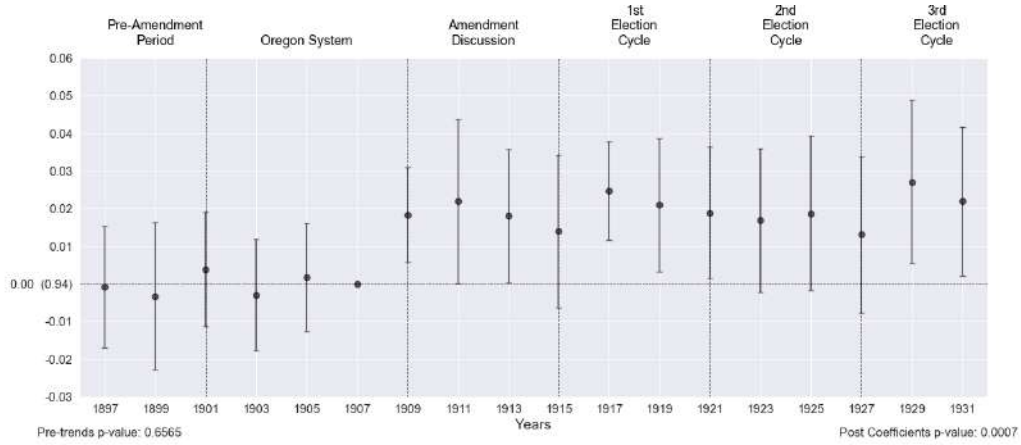
Notes: The figure shows the event-study of Equation 1.2. The dependent variable is cosine similarity. The estimation only considers states that did not adopt Senate primaries prior to the 61st Congress. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.20: Event-Study: Cosine Similarities with Staggered Adoption.

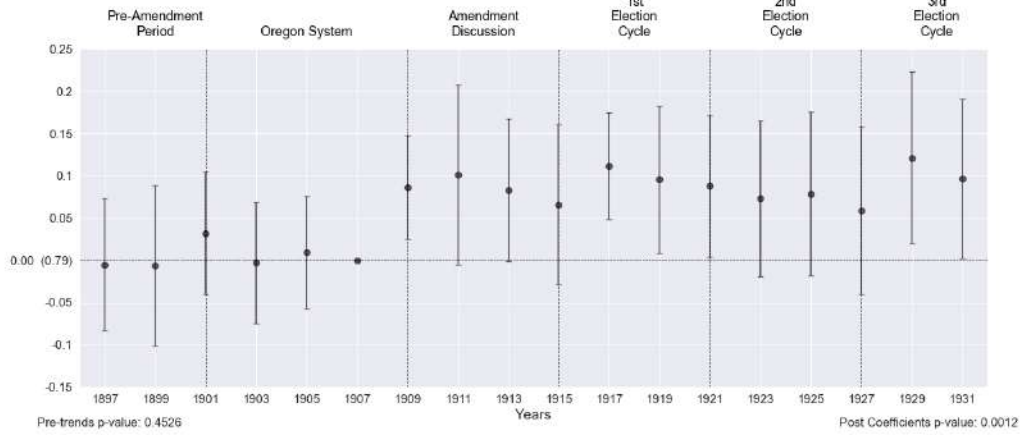


Notes: The figure shows the event-study of Equation 1.3. The dependent variable is cosine similarity. Here we consider the adoption of a Senate primary for a given state as the treatment time, using states that did not adopt primaries prior to the 61st Congress as control. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

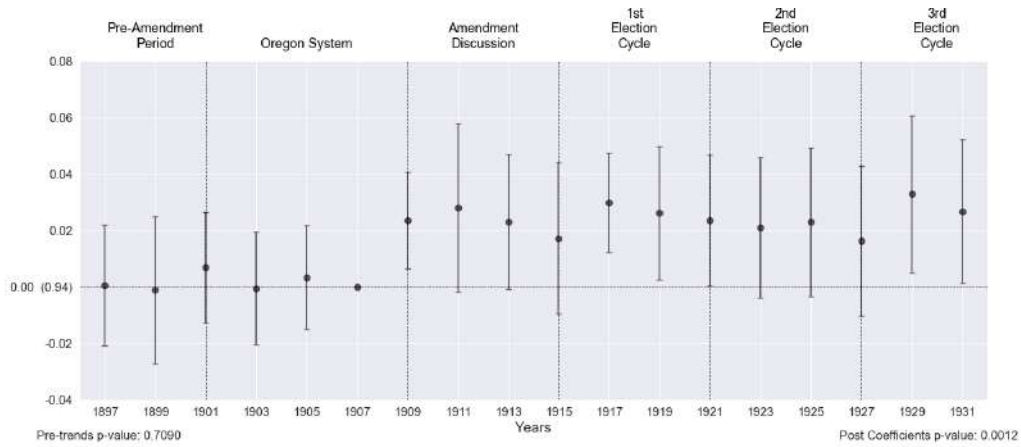
Figure A.21: Event Studies: Alternative Measures for Platform Similarity



(a) Jensen-Shannon Divergence (Inverted)



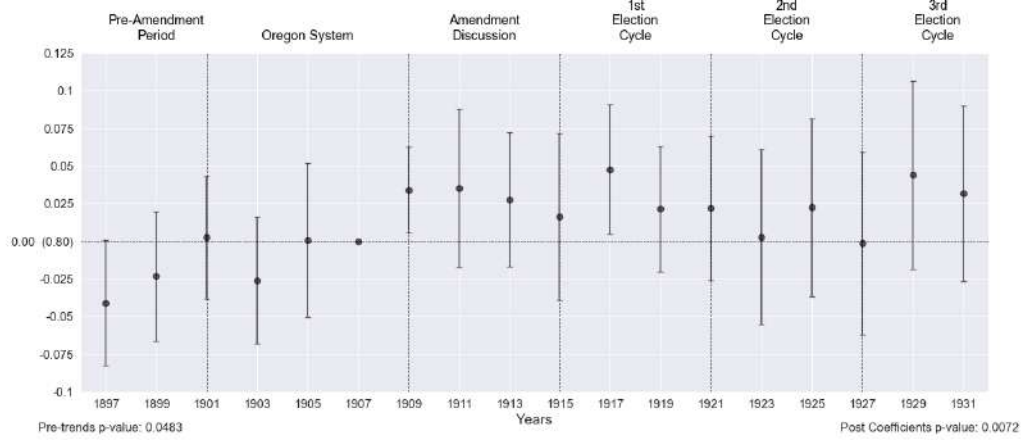
(b) Kullback-Leibler Divergence (Inverted)



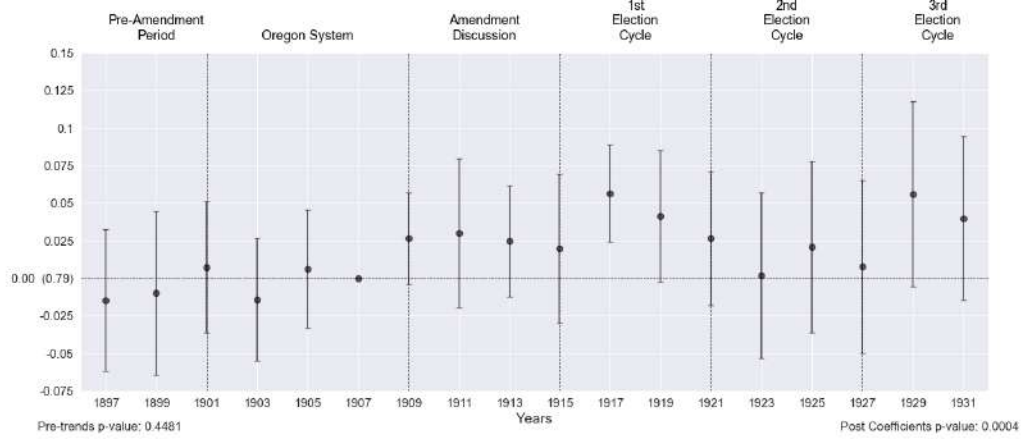
(c) Bhattacharyya Distance (Inverted)

Notes: The figure shows the event-study of Equation 1.2, considering alternative measures of similarity. The dependent variable is inverted Jensen-Shannon divergence in panel a), inverted Kullback-Leibler Divergence in panel b) and Inverted Bhattacharyya Distance in panel c). Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

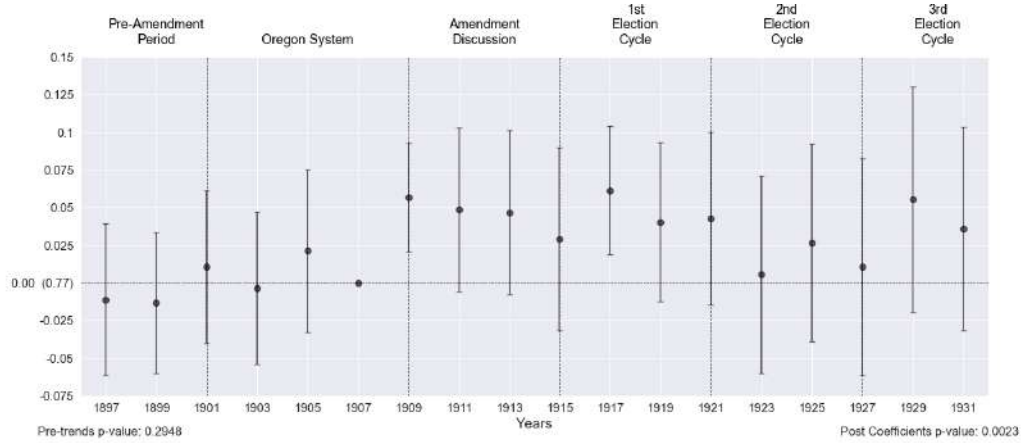
Figure A.22: Event Studies: LDA models, Platform Similarity



(a) 30 Topics



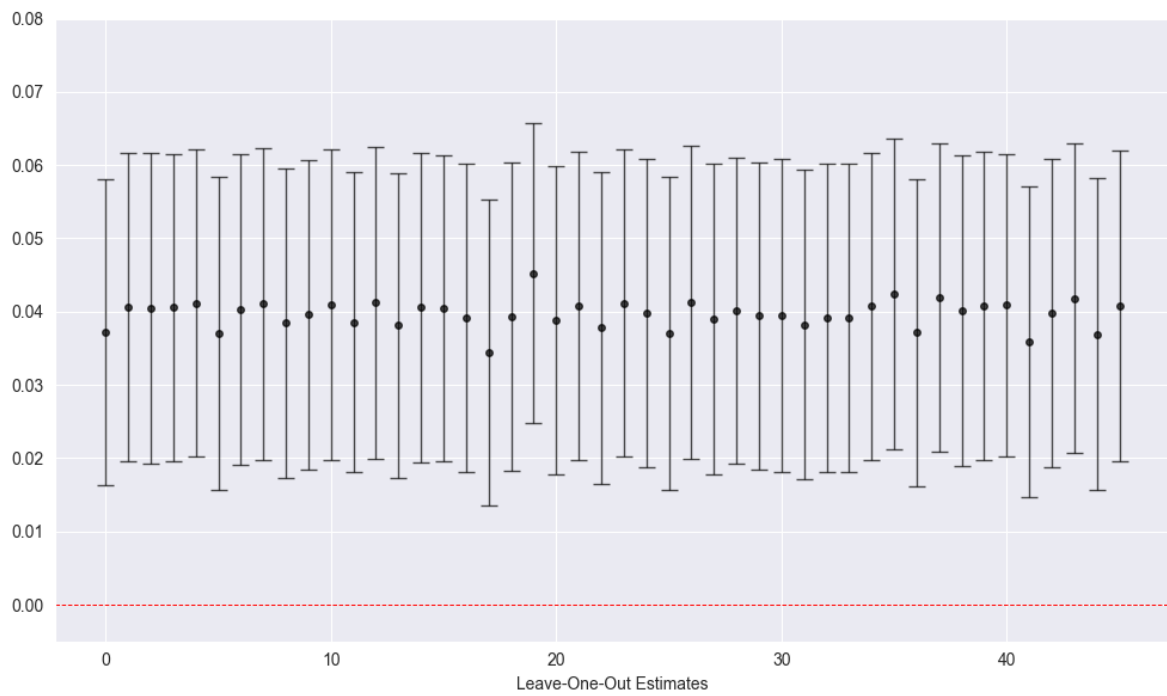
(b) 33 Topics



(c) 39 Topics

Notes: The figure shows the event-study of Equation 1.2, considering alternative topic models with different choice of the number of topics parameter. The dependent variable is cosine similarity. Panel a) displays the result for 30 topics, panel b) for 33 topics and Panel c) for 39 topics. Coefficients are displayed along with 95% CI. Robust standard error, clustered at the state level.

Figure A.23: Cosine Similarity, Leave-One-Out Estimates



Notes: The figure presents the coefficients from Equation 1.2, iteratively re-estimating the regression by excluding one state at a time. The dependent variable is cosine similarity. Coefficients are shown alongside 95% confidence intervals. Robust standard errors clustered at the state level..

A.2 Tables

Table A.1: Topic Labels and Words from Congressional Speeches (1879-1935)

Label	Word List
1 Litigation and Judiciary	litig, plaintiff, attorney, circuit, suit, equity...
2 Military Affairs and Veterans	war_depart, troop, regiment, enlist, command, camp...
3 Agriculture and Farming	plant, animal, expert, farm, quantity, crop...
4 Competition Regulation and Antitrust	interest_commerce, antitrust, common_carrier, penalty, ...
5 Elections and Voting	ballot, voter, legislature, elector, cast, ticket...
6 Public Buildings and Infrastructure	public_building, avenue, architect, erect, rent, block...
7 Waterways and Maritime Infrastructure	harbor, channel, project, engineer, navigate, dam...
8 Land Use and Settlements	acre, settler, tract, homestead, area, public_domain...
9 Marine Corps and Military Personnel	marine_corp, enlist_men, retire_list, lieutenant, corp, ...
10 Labor and Economic Issues	wage, factory, competition, american_labor, ...
11 Postal Services	postmaster, mail, postal_service, postmaster_general, ...
12 Finance and Banking Regulation	loan, stock, asset, director, borrow, deposit...
13 Criminal Justice and Law Enforcement	newspaper, testify, dare, kill, accuse, crime...
14 Debt Settlements and War Finances	budget, program, world_war, unemploy, owe, bonus...
15 Veterans and Healthcare	hospital, doctor, treatment, patient, public_health, physician...
16 Food and Drug Regulation	bottle, alcohol, gallon, liquor, oleomargarine, drug...
17 Community Development and Education	community, educate, contribute, school, born, devote...
18 Federal Governance and Regulation	custom, revise_statute, expire, assign, collector, transact...
19 Native American Affairs and Schooling	indian_school, superintend, tribe, pupil, teacher indian_affair
20 Taxation	deduct, taxable, refund, personal_property, exempt, income...
21 Congressional Procedures	unfinish_business, manual, parliamentary_inquiry...
22 Tributes and Memorials	decease, dead, tribute, widow, die, son...
23 Maritime and Shipping	vessel, marine, crew, sail, foreign_trade, american_ship...
24 Fiscal Policy and Redistribution	prosper, democrat_party, evil, debt, tax, wealth...
25 Textiles and Manufacturing	wool, yarn, iron, sheet, thread, rag...
26 Communication and Transport	telephone, telephone_company, telegraph_company, omaha...
27 International Relations and Foreign Policy	philippin, great_britain, alli, liberty, sovereignty, ratify...
28 Navy and Naval Infrastructure	navy_yard, yard, naval, navy_depart, naval_affair,...
29 Railways and Transportation	car, rail, railroad_company, distance, railway, director...
30 Appropriations & Miscellaneous Expenses	contingent_fund, travel_expense, necessary_expense, seed, nitrate, ...
31 Immigration and Census	quota, alien, immigr, chinese, census, italian...
32 Monetary Policy and Minting	silver, circulate, redeem, coin, mint, denomination...
33 Agriculture and Trade Policy	wheat, bale, corn, flour, grower, agriculture_product...
34 Diplomacy and Foreign Relations	consular_service, ambassador, embassy, canal_zone, ...
35 Political Reform and Platforms	reform, membership, platform, democrat_platform...
36 Party Politics and Debates	distinguish_friend, teller, lecture, lobby, untrue, ...

Notes: The table shows topic numbers and associated labels. Most important keywords for each topic in the third column

Table A.2: Robustness Checks - Platform Similarity under alternative measurements

	Platform Similarity		
	(1)	(2)	(3)
Panel A: Jensen-Shannon Divergence (Inverted)			
Post	0.017*** (0.001)	0.015*** (0.002)	0.013*** (0.0028)
Panel B: Kullback-Leibler Divergence (Inverted)			
Post	0.049*** (0.009)	0.045*** (0.014)	0.039*** (0.017)
Panel C: Bhattacharyya Distance (Inverted)			
Post	0.038*** (0.008)	0.019*** (0.001)	0.015*** (0.0044)
N	1,861	1,784	1,784
Seat FE	✓	✓	✓
Congress FE			
Individual FE		✓	✓
Controls		✓	✓
Linear Time-trend			✓

Notes: The table shows the results of Equation 1.2 with our alternative measures of similarity: Jensen Shannon Divergence (Panel A), Kullback-Leibler Divergence (Panel B) and Bhattacharyya Distance (Panel C). All variables have been transformed as follows for clarity: $\frac{1}{1+\text{Divergence}}$. Robust standard error in parenthesis, clustered at the state level *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.3: Robustness Checks - Platform Similarity Regressions: LDA with 30, 33 & 39 Topics

	Platform Similarity		
	(1)	(2)	(3)
Panel A: ProdLDA 30 Topics			
Post	0.048*** (0.0048)	0.042*** (0.015)	0.035*** (0.011)
Panel B: ProdLDA 33 Topics			
Post	0.051*** (0.00499)	0.036*** (0.009)	0.031*** (0.0106)
Panel C: ProdLDA 39 Topics			
Post	0.053*** (0.0051)	0.048*** (0.094)	0.045*** (0.0112)
N	1,861	1,784	1,784
Seat FE	✓	✓	✓
Congress FE			
Individual FE		✓	✓
Controls		✓	✓
Linear Time-trend			✓

Notes: The table shows the results of Equation 1.2 LDA models with different number of topics. The dependent variable is Cosine-Similarity. Robust standard error in parenthesis, clustered at the state level *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

B Topic Model

This section will go through the details of the topic modeling approach used in this paper¹². As mentioned in Section 1.3.2, the main objective of LDA consists in recovering a distribution of topics over documents (θ_k) and a distribution of words over topics (β_k). Let us consider the matrix $\beta = (\beta_1, \dots, \beta_K)$, representing a learnable parameter of word probabilities for each topic, as in (Blei et al., 2003). Then the joint distribution of a topic mixture θ , a set of topics $\mathbf{z}$ and words w will be given by:

$$p(\theta, \mathbf{z}, \mathbf{w} | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^{N_d} p(z_n | \theta) p(w_n | z_n, \beta), \quad (\text{B.1})$$

where α is a hyperparameter regulating the Dirichlet prior on the topic mixture θ , i.e. the prior on how many topics can be contained in a single document. By integrating over the topic mixture θ and summing over the topic assignments $\mathbf{z}$, we obtain the marginal distribution of a document:

$$p(\mathbf{w} | \alpha, \beta) = \int_{\theta} \left(\prod_{n=1}^{N_d} \sum_{z_n=1}^k p(w_n | z_n, \beta) p(z_n | \theta) \right) p(\theta | \alpha) d\theta. \quad (\text{B.2})$$

Finally, the joint probability of the entire corpus $\mathcal{D}$ is given by:

$$p(\mathcal{D} | \alpha, \beta) = \prod_{d=1}^D \int_{\theta_d} \left(\prod_{n=1}^{N_d} \sum_{z_{d,n}=1}^k p(w_{d,n} | z_{d,n}, \beta) p(z_{d,n} | \theta_d) \right) p(\theta_d | \alpha) d\theta_d. \quad (\text{B.3})$$

Equation B.3 provides a comprehensive overview of the statistical dependencies of the LDA model, e.g. the topic assignment $z_{d,n}$ depends on the topic proportions θ_d and similarly the word assignment $w_{d,n}$ depends on $z_{d,n}$ and β . In other words, the model describes a generative process where documents are formed by first establishing a mixture of topics (θ_d), and then for each word, a specific topic is drawn from that mixture ($z_{d,n}$), which in turn determines the final choice of the word itself ($w_{d,n}$).

The main problem with Equation B.1 is that the posterior distribution $p(\theta, \mathbf{z}, \mathbf{w} | \alpha, \beta)$ is intractable. To overcome this issue several solutions have been proposed, such as traditional MCMC methods –Gibbs Sampling (Blei et al., 2003) or Variational Inference (VI) (Hoffman et al., 2010). VI reframes the problem of calculating the posterior distribution,

¹²The *ProdLDA* implementation follows the source code at <https://pyro.ai/examples/prodllda.html>.

$p(z|w)$, as an optimization problem. The goal is to find the best possible approximation to the true posterior from within a simpler, predefined family of distributions, $q_\phi(z)$, known as the variational family (see Hoffman et al. (2010); Steyvers and Griffiths (2007) for a more in-depth discussion).

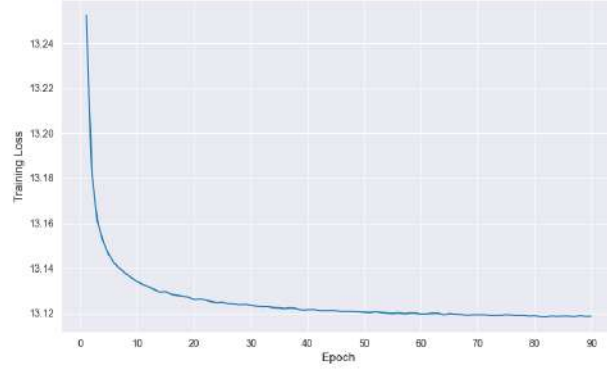
In our study, we employ *ProdLDA*, a topic modeling approach that proposes a novel inference framework by leveraging the Auto-Encoding Variational Bayes (AEVB) framework and a neural network architecture (Srivastava and Sutton, 2017). *ProdLDA* improves the robustness and coherence of topic representations by combining multiple simpler distributions, or "experts," to model the complex patterns within the data. This approach provides a better fit for large and complex datasets like congressional speeches, enhancing interpretability and performance.

The network architecture proposed by Srivastava and Sutton (2017) consists of an Encoder and a Decoder. The Encoder is composed of two fully connected hidden layers with 200 neurons each. It receives the document-term matrix as input and outputs the parameters—the mean and variance—of a logistic-normal distribution in the latent space, which models the document’s topic proportions. The Decoder then uses this latent representation to reconstruct the original document: at this point the loss is computed. The model was implemented in Python using PyTorch for the encoder-decoder neural network architecture and Pyro for the probabilistic programming components. The model has been trained for 90 epochs, using the ADAM optimizer (Kingma and Ba, 2014) with 0.001 learning rate, a batch size of 64 and a dropout rate of 10%. Using T4 GPU it takes a total of roughly 45 minutes to train a single model.

Figure B.24 displays the computed loss over the 90 epochs. We also show the differences in performance between traditional LDA and *ProdLDA* in Table B.4. The results indicate that *ProdLDA* outperforms traditional LDA in both PMI and CV coherence metrics¹³, suggesting that the topics generated by *ProdLDA* are more coherent and interpretable. This improvement is likely due to the neural network architecture of *ProdLDA*, which allows for a more flexible and powerful modeling of the underlying data structure compared to the probabilistic framework of traditional LDA. Through a prediction module in the model class, we used the trained model to infer the topic dis-

¹³NPMI Coherence measures the degree of association between words in a topic, with higher values indicating that the words frequently appear together in the corpus. CV Coherence combines several metrics, including NPMI, to evaluate how well the words in a topic form a meaningful concept based on their distribution and semantic similarity in the corpus. These are two standard measures to evaluate the performance of topic models

Figure B.24: Training loss of ProdLDA with 36 Topics



Notes: The figure illustrates the training loss for *ProdLDA* with 36 topics, evaluated over 90 training epochs

tributions for each document in our corpus. For the final choice of number of topics, we observed that models accounting for 30 to 39 topics were generating more interpretable results.

Table B.4: Combined NPMI and CV Scores for Different Topic Models

Topics	ProdLDA		LDA	
	NPMI	CV	NPMI	CV
20	0.094	0.652	-0.114	0.365
25	0.120	0.649	-0.096	0.393
27	0.098	0.631	-0.101	0.421
30	0.125	0.667	-0.142	0.388
33	0.126	0.657	-0.141	0.366
35	0.113	0.643	-0.137	0.385
36	0.116	0.639	-0.155	0.388
39	0.125	0.667	-0.163	0.387
40	0.112	0.635	-0.166	0.395
45	0.115	0.629	-0.195	0.381
50	0.122	0.646	-0.208	0.392

Notes: The table shows scores for *ProdLDA* (columns 1 and 2) and *LDA* (columns 3 and 4), for different numbers of topics.

C A Stylized Voting Model of Malapportionment and Agenda Concentration

This Appendix presents a very simple probabilistic voting framework, adapted from [Lin et al. \(1999\)](#), to illustrate how differences in preference aggregation under direct and indirect elections can mechanically produce differences in the concentration of political platforms. The goal is purely interpretative: the model is not intended as a structural account of senatorial behavior but as a conceptual supplement to the entropy results in the main text.

In the standard multi-candidate probabilistic spatial model, voter i has ideal point $\mathbf{v}_i \in \mathbb{R}^K$, and candidate j chooses a platform $\mathbf{c}_j \in \mathbb{R}^K$. Voter utility is

$$U_{ij} = \eta_j - \|\mathbf{v}_i - \mathbf{c}_j\|,$$

where η_j captures random non-policy shocks. Under standard assumptions, [Lin et al. \(1999\)](#) show that candidates converge to the *mean voter position*:

$$\mathbf{c}_j^* = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i,$$

which provides a benchmark for understanding how equilibrium platforms change when the aggregation rule over voters differs across electoral systems.

To capture features of the pre-17th Amendment environment, we introduce two modifications:

1. Voters are grouped into D districts with populations $(n_1, \dots, n_D)$, where $\sum_d n_d = N$. Consistent with the historical setting, urban districts are few but populous, while rural districts are numerous but small.
2. Elections may be **direct** or **indirect**.

Under *direct elections*, candidates maximize expected votes by aggregating preferences at the individual level, yielding the standard convergent equilibrium. In the context of *indirect elections*, the second stage of the process would, in reality, involve strategic behavior by state legislators, party coordination, and potentially more sophisticated bargaining than the sincere probabilistic voting assumed in the baseline model. Treating

both direct and indirect elections as sharing the same form of equilibrium therefore abstracts from these complexities. We adopt this simplification deliberately: the aim of the exercise is not to model the strategic intricacies of multi-stage elections, but to isolate how different *aggregation rules*—individual-level versus district-level—affect the location and concentration of equilibrium platforms. For this purpose, assuming comparable behavioral primitives across stages offers a transparent way to highlight the mechanical effect of malapportionment on the resulting equilibrium.

Under *indirect elections*, voters elect state legislators, and each district receives equal weight in choosing the senator. Let $\mathbf{v}_d$ denote the mean preference vector in district d . The indirect-election equilibrium platform is

$$\mathbf{c}_{\text{indirect}}^* = \frac{1}{D} \sum_{d=1}^D \mathbf{v}_d,$$

while the direct-election equilibrium is

$$\mathbf{c}_{\text{direct}}^* = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i.$$

Because district populations vary, these two aggregators need not coincide: when rural districts are overweighted, the indirect-election platform will be mechanically pulled away from areas containing most of the population. To illustrate this mechanism, we simulate voter preferences as draws from district-specific Dirichlet distributions, allowing urban and rural districts to differ systematically. For each configuration of:

- the share of voters in urban districts, θ , and
- the number of urban districts,

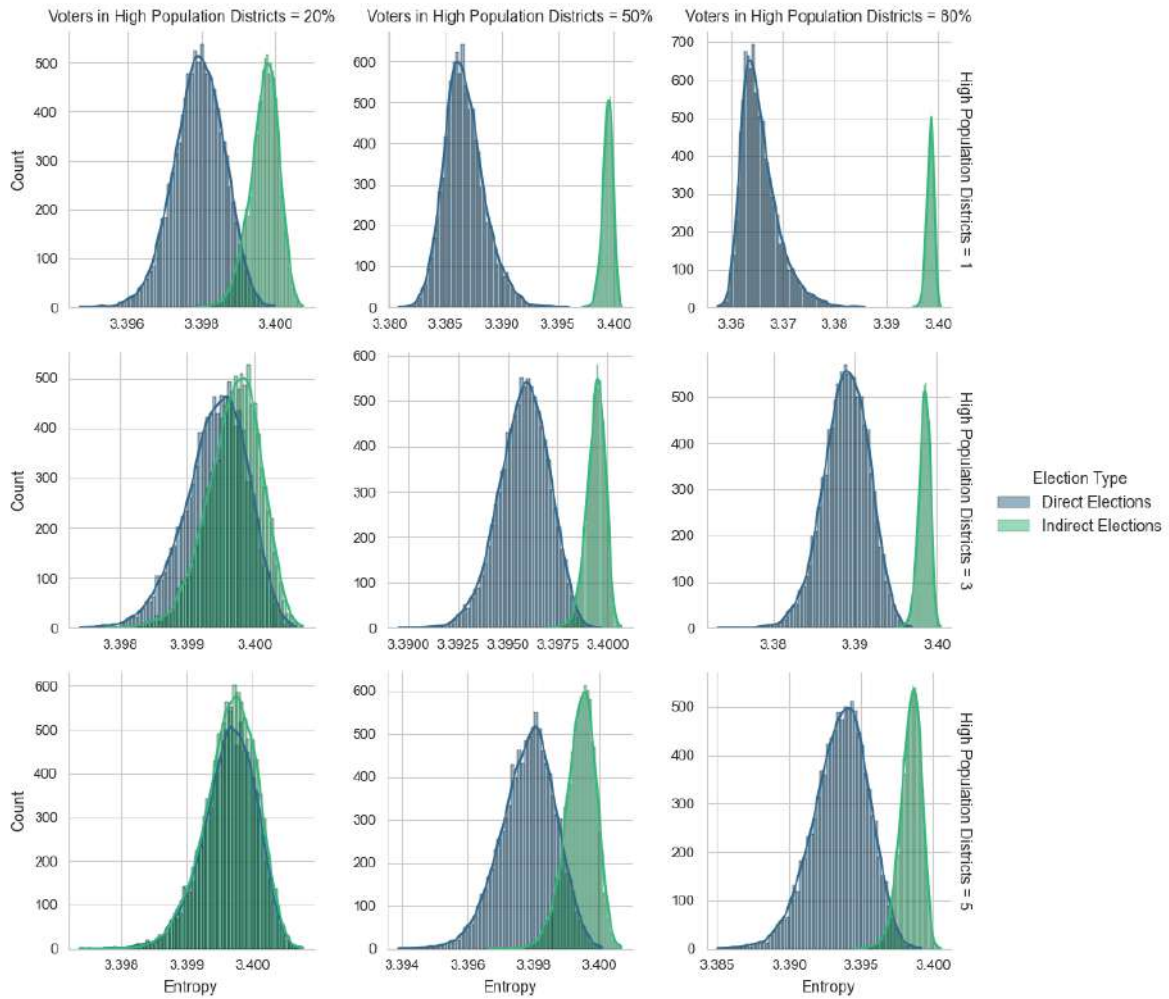
we compute equilibrium platforms under both electoral systems and measure their *entropy*, the same measure of agenda concentration used in the main paper. Figure C.25 displays the distribution of platform entropies across simulations. Two findings stand out:

1. Platforms under direct elections are systematically **more concentrated** (lower entropy) than under indirect elections.

2. Differences are largest when urbanization is high but the number of urban districts is small—conditions that maximize malapportionment.

These results mirror the early 20th century United States, where rapidly growing urban populations were represented through rural-dominated legislative districts (Durfee, 1945).

Figure C.25: Simulation results: Entropy under Electoral Regimes

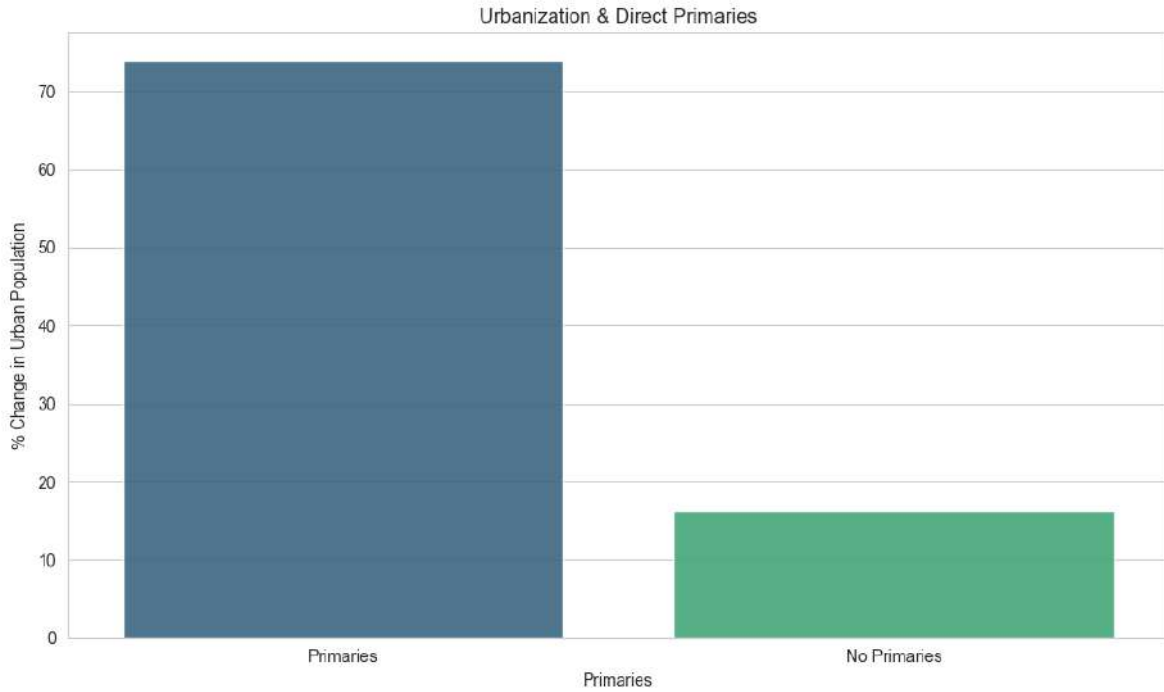


Notes: The figure illustrates the distributions of equilibrium platforms across simulations under direct and indirect electoral systems. The simulation considers 10,000 voters, a platform dimensionality of 30 and 50 total districts.

Historical documentation highlights substantial population disparities across districts, with rural counties often enjoying disproportionate representation and re-apportionment occurring infrequently (Durfee, 1945). Moreover, states experiencing greater urbaniza-

tion were more likely to adopt primary elections prior to the 17th Amendment, consistent with increasing pressure for more proportionate representation, as shown in Figure C.26, although our results show that this early primary adoption did not impact senatorial agenda concentration.

Figure C.26: Simulation results: Entropy under Electoral Regimes



Notes: Change in urban population (1890-1920) for states adopting direct primaries prior to the 17th Amendment versus those that did not. States with primaries experienced substantially greater urbanization. Data on urban population by state over time has been sourced from [Iowa State University \(2021\)](#).

This stylized framework illustrates how unequal representation across districts can yield *more diffuse equilibrium agendas* under indirect elections. By shifting to direct statewide voting, the 17th Amendment reweighted voter preferences toward populous districts, generating incentives for *more focused, voter-oriented rhetorical agendas*. This mechanism complements the empirical decline in entropy documented in Section 1.4 of the main text.

Chapter 2

Political Shocks, Extremism and Violence: Evidence from the West Bank

Abstract: This study investigates the impact of political shocks on extremist violence. It analyzes the aftermath of the December 2017 U.S. recognition of Jerusalem as Israel’s capital, treating the decision as a quasi-experimental shock. Using a staggered difference-in-differences methodology, the paper examines changes in violence comparing areas proximate to illegal Israeli outposts with those that are not. The findings reveal that the U.S. declaration triggered a sharp and significant increase in violence in areas near these outposts, where violent events rose by roughly 150%. The analysis shows that incidents targeting civilians were particularly affected. The escalation of violence was asymmetrical; there was an immediate surge in violence perpetrated by Israeli settlers in the first years following the announcement. Conversely, a significant rise in violence from Palestinians emerged with a delay of nearly four years. These results suggest the diplomatic shock acted as a catalyst for radical behavior, initiating a self-perpetuating cycle of conflict in the West Bank.

JEL Classification: C21, C31, D72, D74, F51, Z13

Keywords: Political violence, territorial conflicts, Israeli-Palestinian conflict, extremism, settlements, outposts, diplomatic shocks, event-study

2.1 Introduction

Political violence rarely emerges from a vacuum. Scholars have devoted significant attention to understanding the conditions under which such conflicts escalate, trying to disentangle their drivers and dynamics. Traditional explanations have focused on state failure, ethnic grievances, as well as economic or political competition frameworks ([Besley and Persson, 2011](#); [Acemoglu et al., 2013](#)). Yet, another important category of political violence operates through a different logic entirely: the emergence of radical behavior through the erosion of social norms.

A growing body of research suggests that extremist behavior often escalates not through gradual radicalization, but through sudden shifts in the perceived acceptability of violent conduct ([Bursztyn et al., 2020](#); [Giani and Méon, 2021](#); [Müller and Schwarz, 2023](#); [Albornoz et al., 2022](#)). When political elites signal tolerance of –or support for –radical positions, they can effectively lift the social sanctions that previously constrained such actions. This erosion of behavioral norms emboldens individuals who already hold extreme views. Previously constrained by social costs, they suddenly find themselves operating in an environment that seems to legitimize their actions and reduce the likelihood of punishment.

This paper tests this mechanism in the context of the Israeli-Palestinian conflict, focusing on a specific manifestation of political extremism: illegal settler outposts in the West Bank. Unlike authorized settlements, outposts are established without formal Israeli government approval by ideologically-driven activists and are deemed illegal under Israeli domestic law ([Israeli Supreme Court, 2014, 2011](#); [Kurtzer, 2016](#)).¹ Thus, these outposts represent radical right-wing ideological pockets within the broader settler population. This context motivates the study’s central question: how can political signals escalate extremist violence? To answer this, I exploit the unexpected announcement by the Trump administration on December 6, 2017, to recognize Jerusalem as Israel’s capital –a move that marked a significant departure from decades of U.S. diplomatic policy.

¹Israeli settlements in the West Bank are deemed illegal under international law by the United Nations ([United Nations Security Council, 1979, 2016](#)) and a 2024 International Court of Justice advisory opinion ([International Court of Justice, 2024](#)), both agreeing they are in violation of the Fourth Geneva convention ([Fourth Geneva Convention, 1949](#)). However, they are recognized as legal by the Israeli judicial system, which advances arguments including claims that the West Bank is not technically occupied territory under international law, assertions of historical Jewish rights to the land, military necessity and security justifications, and legal frameworks established through military orders ([Levy Commission, 2012](#); [Israeli Military Government in the West Bank, 1967](#)).

The declaration, which signaled high-level U.S. support for maximalist territorial claims, could be perceived by radical individuals as a validation of their own political agenda.

This paper tests the interplay between the 2017 U.S. declaration on Jerusalem and extremist violence using a staggered difference-in-differences (DiD) design. The central challenge is to separate the effect of the political announcement from the ongoing violence already associated with extremist outposts. To achieve this, I compare changes in violence in areas proximate to illegal outposts with those in areas where outposts had not yet been established, before and after the U.S. declaration. The analysis leverages an entirely novel combination of high-resolution satellite imagery to precisely date outpost establishment ([Copernicus Sentinel, 2025](#); [Maxar Technologies, 2023](#)), daily geo-referenced data on violent events from the Armed Conflict Location and Event Data Project (ACLED) ([ACLED, 2024b](#)) and detailed geographic information from Israeli NGOs like Peace Now ([Peace Now, 2024](#)) and B’Tselem ([B’Tselem, 2024](#)). The integration of these sources provides a uniquely granular view, making it possible to precisely trace how the conflict intensified in specific, high-friction areas surrounding illegal outposts.

The findings demonstrate that the U.S. declaration acted as a critical catalyst, triggering a sharp and generalized escalation in political violence. This effect was particularly pronounced in areas near outposts, where the announcement appears to have been interpreted by radical settlers as a powerful political signal or “green light” for action. The data reveal that this was not a temporary spike but the beginning of a sustained period of heightened conflict: total violent events increased by approximately 150%, while incidents involving Israeli settlers surged by roughly 190% relative to pre-treatment levels. Crucially, however, the analysis reveals a stark asymmetry in this aggression. The initial wave of violence was immediate, one-sided, and perpetrated by Israeli settlers, particularly against civilians, whereas Palestinian-perpetrated violence remained comparatively low for nearly four years—a divergence that suggests the declaration did not immediately spark a bilateral exchange, but rather unleashed a specific, unilateral mobilization that would only later evolve into a broader cycle.

Further analysis confirms that the diplomatic announcement, rather than the mere physical presence of outposts, was the primary trigger. Outposts established prior to the declaration generated no significant violence until the political signal effectively “activated” them. This activation set in motion a distinct temporal pattern, where Palestinian-perpetrated violence remained sporadic for several years before undergoing

a fundamental shift following the sharp spike in Israeli-perpetrated violence and military operations in mid-2021. Prior to this inflection point, Palestinian-perpetrated violence was largely unresponsive to lagged settler attacks; however, in the subsequent period, attacks became strongly predicted by prior settler aggression. These findings trace the evolution from a single diplomatic act to a durable retaliatory cycle, providing a sobering account of how international politics can destabilize fragile local realities and trigger an intensifying conflict that persists long after the initial shock.

These results speak directly to a growing body of research that examines how shifts in the social acceptability of radical views can unleash extremist behavior. The mechanism by which elite signaling empowers latent extremism is clearly not unique to the Israeli-Palestinian context, it is becoming increasingly common throughout the world. Recent studies have brought forth compelling evidence showing that Donald Trump’s electoral success increased Americans’ willingness to publicly express xenophobic views and reduced the negative judgment associated with such expressions ([Bursztyn et al., 2020](#)). Similarly, other research finds that Trump’s election was followed by a significant increase in the reporting of racially biased attitudes in Europe ([Giani and Méon, 2021](#)), and that his tweets acted as a direct predictor of subsequent anti-Muslim hate crimes in the U.S. ([Müller and Schwarz, 2023](#)). This process does not necessarily persuade individuals to adopt new extremist beliefs, but rather encourages those with pre-existing radical preferences to act on them by lowering the expected social costs associated with such actions ([Müller and Schwarz, 2023](#)). While this strand of literature has focused on contexts with relatively low baseline levels of violence, this study also contributes by exploring such dynamics in a setting of endemic conflict. The Trump administration’s Jerusalem declaration provides an ideal case study for this purpose, as the announcement functioned as an unambiguous signal from a global superpower on a deeply polarizing issue.

This paper also adds to the literature on political violence: it offers a complementary perspective to the traditional “top-down” literature on political violence, which emphasizes strategic decisions by domestic state elites. Foundational theories, such as [Besley and Persson \(2011\)](#), model conflict as a rational contest for state power, where elites engage in violence based on potential prizes and institutional constraints. This logic is particularly evident where states actively direct non-state violence, for instance by using state radio to incite genocide participation ([Yanagizawa-Drott, 2014](#)), or where they allow paramilitaries to persist in exchange for electoral support ([Acemoglu et al., 2013](#)). In

contrast, this paper provides a “bottom-up” perspective, arguing that exogenous political shocks can empower non-state actors at the local level. Rather than treating violence as a state-driven tool, my analysis focuses on ideologically motivated individuals whose behavior is enabled by shifts in perceived political opportunity and social legitimacy. By analyzing how this diplomatic signal impacted radical settler violence, I show how conflict can emerge from the periphery, not sparked by domestic strategic calculations, but by the external validation of a radical agenda.²

This approach also offers a new lens for the Israeli-Palestinian conflict, since existing empirical work has largely focused on the strategic calculations of recognized actors and the structural conditions of the occupation. For instance, studies have shown that state actors in the conflict behave with nuanced strategic considerations; [Durante and Zhuravskaya \(2018\)](#) find that the Israeli military is more likely to launch attacks that risk civilian casualties on days when the US news cycle is occupied by other major events, suggesting an effort to manage international political costs. Other research has examined the internal political dynamics of violence, with [Jaeger et al. \(2012\)](#) showing that Israeli violence temporarily radicalizes Palestinian political preferences. Furthermore, in contrast to existing work on the conflict’s impact on Palestinians—such as how Israeli-imposed mobility restrictions fragment the economy and harm the Palestinian labor market ([Cali and Miaari, 2018](#); [Adnan, 2015](#))—this study focuses on the behavioral triggers of violence within Israeli society. Hence, this approach provides a novel, bottom-up perspective on the micro-dynamics of conflict escalation.

The remainder of this paper proceeds as follows. Section [2.2](#) provides historical context, and Section [2.3](#) describes the data sources. I then outline the empirical strategy in Section [2.4](#), present the main findings in Section [2.5](#), and validate them with a series of robustness checks in Section [2.8](#).

2.2 Background

2.2.1 Historical Context

The Israeli-Palestinian conflict is one of modern history’s most protracted disputes, with deep historical, religious, and nationalistic roots. A pivotal moment in this conflict

²While the Israeli government under Prime Minister Netanyahu publicly welcomed the U.S. decision, the move itself represented a unilateral departure from decades of U.S. foreign policy.

occurred during the 1967 Six-Day War, when Israel occupied the West Bank, the Gaza Strip, the Golan Heights and Sinai. Following this occupation, Israeli citizens began establishing civilian settlements –communities built on occupied land–in the West Bank and, for a time, the Gaza Strip.³ While considered illegal under international law,⁴ Israeli law treats settlement construction as legal and formally authorized by the state. Settlement activity began under left-wing governments in the years following the war and intensified significantly after 1977 with the Likud party’s victory (Pappe, 2017).⁵

As settlement construction continued through the following decades, the 1993 Oslo Accords aimed to establish a path toward a two-state solution by creating the Palestinian Authority and dividing the West Bank into three administrative zones, with Area C remaining under full Israeli control, which includes roughly 60% of the whole West Bank, Area B under joint Israeli-Palestinian control, and Area A under full Palestinian control. However, the Oslo process ultimately failed, resulting in the fragmentation of Palestinian territory and the entrenchment of the Israeli occupation.

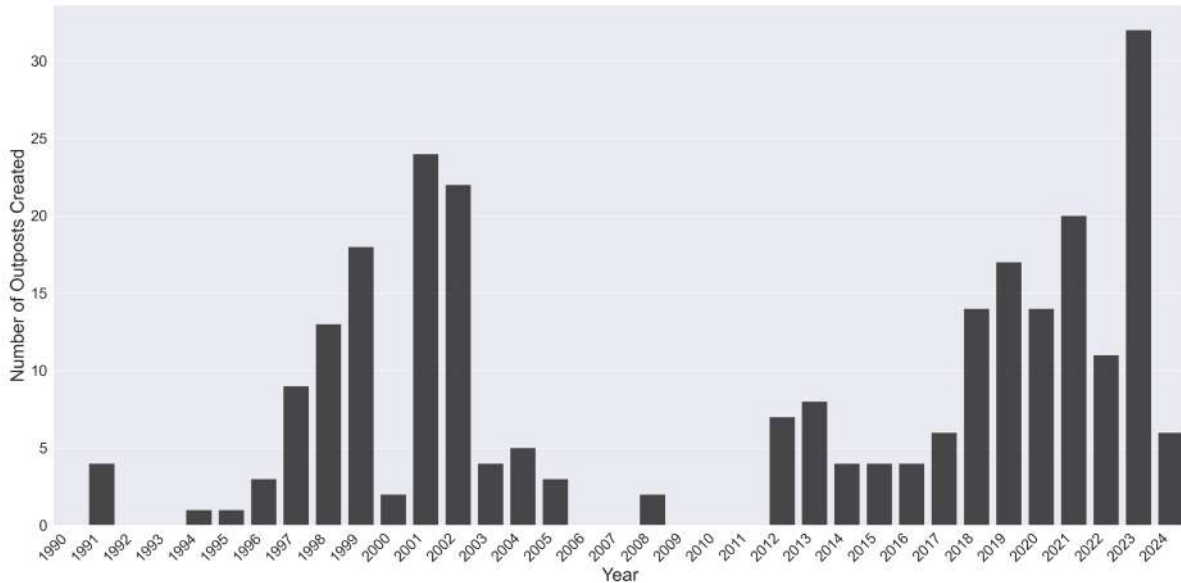
From the 1990s onward, increased international pressure halted the creation of new settlements, although existing ones still continue to expand (Pappe, 2017). This led to the emergence of unauthorized outposts—small clusters of structures built without official government approval by ideologically motivated settlers. These outposts are illegal under Israeli domestic law (Israeli Supreme Court, 2014, 2011), yet have proliferated across the West Bank over the last decades (Peace Now, 2017). Outposts typically begin as small-scale ventures with temporary structures on strategically chosen hilltops, often developing into agricultural or shepherding farms that enable control over vast areas of Palestinian land (B’Tselem, 2021). Figure A.2 in the Appendix provides some examples of outposts using high resolution satellite imagery. Despite their illegal status, these outposts frequently expand with infrastructure support and are established on both declared “state land” and privately owned Palestinian land (Peace Now, 2017; B’Tselem, 2021). Figure 2.1 illustrates the emergence of outposts over time, while Figure A.1 provides a comprehensive map of the West Bank showing outposts, settlements, and Palestinian communities, along with the division of the territory into Areas A, B, and C as defined

³Settlements in the Gaza Strip were dismantled in 2005 as part of Israel’s unilateral disengagement plan.

⁴United Nations Security Council (1979, 2016); Fourth Geneva Convention (1949); International Court of Justice (2024)

⁵Likud (Hebrew for ‘Consolidation’) is the major center-right to right-wing political party in Israel. Its 1977 electoral victory marked a historic political upheaval, ending nearly three decades of political dominance by left-wing parties.

Figure 2.1: Outposts in the Occupied West Bank



Notes: Number of new outposts created over time. Source [B'Tselem \(2024\)](#) and [Peace Now \(2024\)](#).

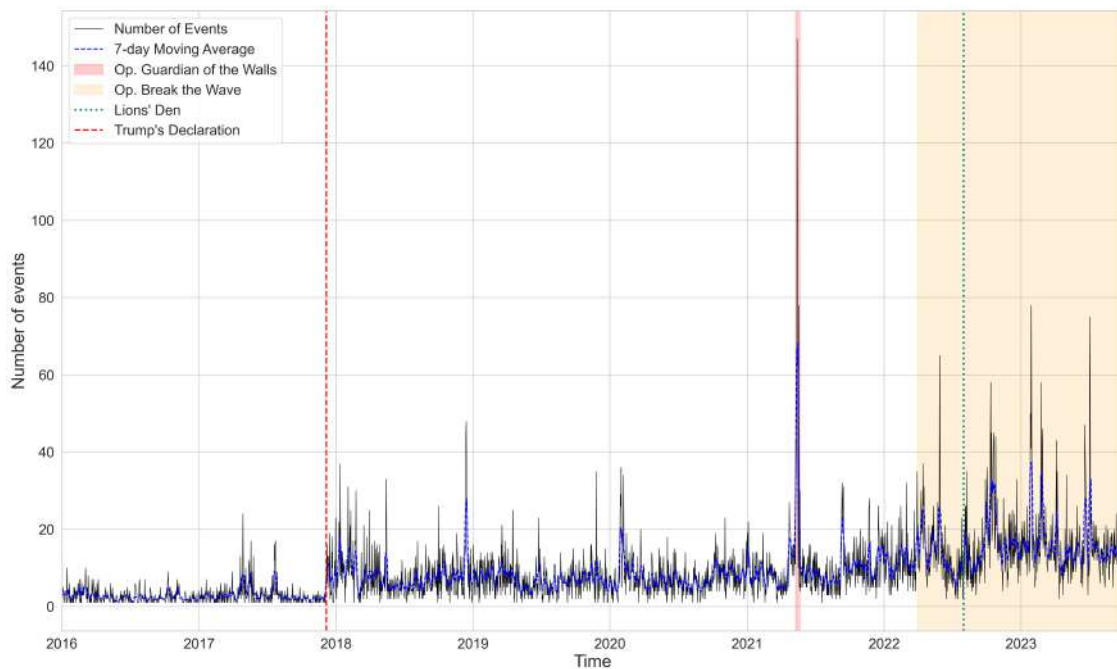
by the Oslo Accords.

2.2.2 The 2017 Jerusalem Declaration

Against this backdrop of expanding outpost activity, a major external shock altered the conflict dynamics. On December 6, 2017, the United States under President Trump officially recognized Jerusalem as Israel’s capital, marking a significant departure from decades of U.S. diplomatic policy. This diplomatic move sent immediate shockwaves through the region, igniting a surge of violence in the West Bank over the following months ([OCHA, 2018](#)), while also reshaping the trajectory of the Israeli-Palestinian conflict in the following years. This decision was also met with widespread international criticism, with the international community formally rebuking the U.S. policy shift. In a rare emergency session, the United Nations General Assembly overwhelmingly condemned the unilateral decision, with a vast majority of countries adopting a resolution that affirmed Jerusalem’s status as a final issue to be negotiated by both sides and asking other nations not to locate diplomatic missions there ([United Nations General Assembly, 2017](#)).⁶ Nonetheless, the declaration seems to have sparked a cycle of violence that lasted

⁶As of 2025, only six countries have relocated their embassies in Jerusalem: United States and Guatemala since May 2018, Kosovo since March 2021, Honduras since June 2021, Papua New Guinea since September 2023 and Paraguay since December 2024.

Figure 2.2: Number of Daily Violent Events in the West Bank



Notes: Time series of daily violent events in the West Bank, retrieved from ACLED. The dashed vertical line indicates the date of the U.S. recognition of Jerusalem as Israel’s capital. The dashed blue line is a 7-day moving average. The series stops on September 2023.

for the years to follow. Figure 2.2 shows the number of daily violent events in the West Bank, highlighting the surge in violence following the U.S. recognition of Jerusalem.

The symbolic weight of this declaration cannot be overstated, as Jerusalem lies at the crucible of the conflict’s most intractable religious, national, and political claims. For Israel, it is the biblical heart of the Jewish people and the site of their ancient Temples. On the other hand, for Palestinians, it is the home of the Al-Aqsa Mosque, the third holiest site in Islam, and the intended capital of their future state ([European Parliament, 2012](#)).⁷ The U.S. recognition, therefore, did more than just relocate an embassy; it preemptively validated the most maximalist Israeli claims over the city. For radical settlers, whose ideology is predicated on indivisible Jewish sovereignty over the entire land, the American declaration was perceived as a step toward eschatological fulfillment—a divine endorsement of their mission to reclaim the entirety of *Eretz Yisrael*.⁸

⁷East Jerusalem is the portion of the city captured and occupied by Israel during the 1967 Six-Day War. Under international law, it is considered occupied territory ([United Nations Security Council, 1980, 2016](#)).

⁸Hebrew for “Land of Israel”, referring to the biblical promised land ([Havelock, 2011](#)). In modern political use, particularly by nationalists, it often denotes a territory that includes the West Bank ([Likud, 2012](#)).

Similarly, the decision provoked widespread anger among Palestinians, who viewed it as a direct negation of their national aspirations and a fatal blow to the peace process.

It is important to note that the unilateral decision was also unexpected. Although fulfilling a campaign promise, President Trump's recognition of Jerusalem was a diplomatic shock due to its abrupt timing and manner. Since the passage of the Jerusalem Embassy Act in 1995,⁹ a presidential waiver had been consistently used by his predecessors—from Clinton to Obama—to delay the embassy's move. Trump himself upheld this two-decade precedent by signing the waiver in June 2017 ([Cortellessa, 2017](#)), making his sudden reversal just six months later an unexpected break from established U.S. policy. The first reports anticipating this decision emerged on December 1, 2017, six days before the official announcement: citing U.S. officials, Reuters reported that President Trump was weighing a plan to formally recognize Jerusalem as Israel's capital ([Holland and Spetalnick, 2017](#)). Given the monthly resolution of the empirical analysis, this short anticipation period should not be considered as a threat to identification.

Beyond the immediate consequences of the U.S. recognition, the years that followed were marked by additional shocks that further intensified conflict dynamics on the ground. In May 2021, the escalation surrounding Jerusalem and Gaza culminated in Operation Guardian of the Walls (10–21 May 2021), which—although concentrated in Gaza—triggered widespread unrest and heightened confrontations across the West Bank ([Israel Defense Forces, 2021](#)).¹⁰ The operation also featured intense exchanges with Gaza-based militias, as Hamas and Palestinian Islamic Jihad (PIJ) launched large barrages of rockets toward Israeli cities, further amplifying regional instability ([Israel Defense Forces, 2021](#)). These rocket attacks resulted in several civilian deaths and numerous injuries inside Israel ([Israel Defense Forces, 2021](#)). Israel responded with extensive airstrikes across the Gaza Strip, causing significant civilian casualties and widespread destruction of residential buildings, infrastructure, and essential services, while simultaneously intensifying its operations in the West Bank, including arrests and raids ([United Nations Office for the Coordination of Humanitarian Affairs, 2021b](#)). This operation marks the highest peak of violence after Trump's declaration, reaching over 140 violent events in May 14

⁹A bipartisan law passed by Congress in 1995 mandating the relocation of the U.S. embassy in Israel from Tel Aviv to Jerusalem ([U.S. Congress, 1995](#)).

¹⁰The May 2021 escalation was triggered by rising tensions in East Jerusalem, particularly over the scheduled forced evictions of Palestinian families in Sheikh Jarrah, i.e. Palestinian neighbourhood in East Jerusalem, and clashes at the Al-Aqsa Mosque compound during Ramadan ([United Nations Office for the Coordination of Humanitarian Affairs, 2021a](#)).

2021, with over 2,500 casualties ([ACLED, 2024b](#)). Moreover, beginning on 31 March 2022, the Israeli military launched Operation Break the Wave, an ongoing counterinsurgency campaign characterized by near-daily raids in Jenin, Nablus, and Tulkarm, leading to persistently elevated levels of violence ([Israel Defense Forces, 2023](#)). During this period, a new generation of decentralized Palestinian armed groups also emerged, such as the Lions’ Den in Nablus (August 2022), alongside the intensification of activity of long-standing paramilitary groups such as the Al-Aqsa Martyrs’ Brigade, as well as locally organized factions aligned with Hamas and PIJ ([ACLED, 2023, 2024b](#)). All of these developments—reflected in Figure 2.2 through shaded intervals and vertical lines show how the post-2017 environment was shaped not only by the diplomatic shock of the U.S. declaration but also by a series of overlapping military and organizational shifts that contributed to the sustained increase in violence across the West Bank.

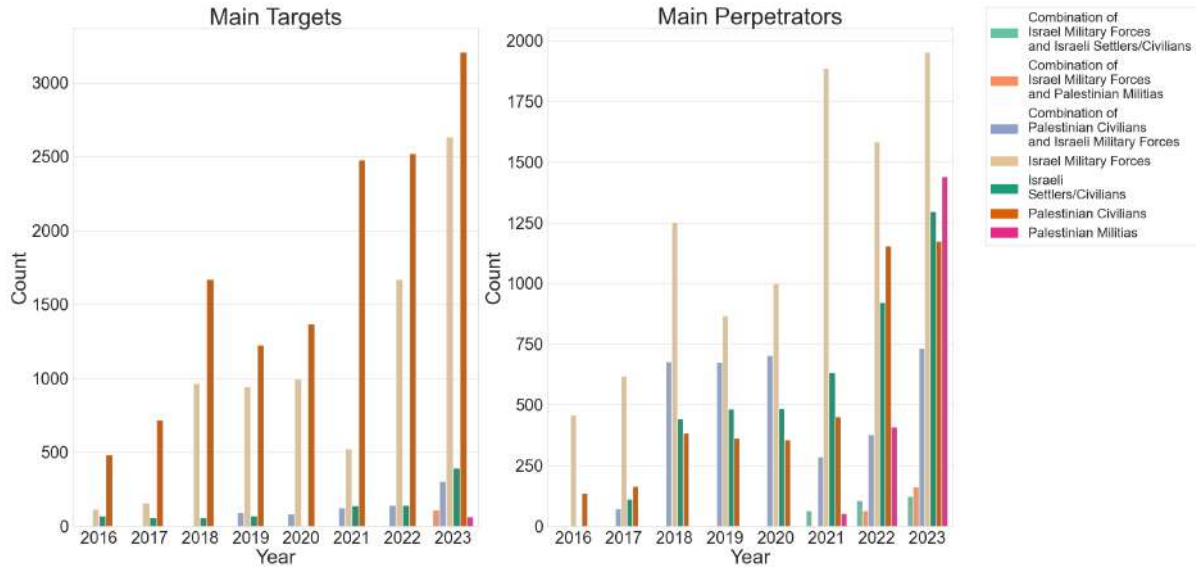
2.3 Data

This section will go through all the data sources used for this study, which include daily geo-referenced violent events from the [ACLED \(2024b\)](#) project, satellite imagery for precise outpost timing ([European Union, Copernicus Land Monitoring Service, 2025](#); [Maxar Technologies, 2023](#)), and comprehensive spatial data from Israeli NGOs ([Peace Now, 2025](#); [B’Tselem, 2024](#)).

Political Violence Data [ACLED \(2024b\)](#) maintains daily, geo-referenced records of political violence events, demonstrations, and other politically relevant developments across the West Bank ([ACLED, 2024a](#)). The database systematically compiles information from diverse local, national, and international news sources to create a comprehensive repository of political violence incidents, capturing detailed information on location, timing, and event characteristics, along with a short textual description of each event ([ACLED, 2024a](#)).¹¹ The data starts from the 1st of January 2016 and is updated daily, although the study presented here only considers events until the end of December 2022. For the remainder of the paper, a “violent event” is defined broadly to include not only

¹¹ACLED employs a rigorous data collection methodology, including a tailor-made sourcing process for individual regions and countries. More details can be found on the [organization’s website](#). In recent years, several studies have relied on this data source for research purposes ([Kafando and Sakurai, 2024](#); [Rohner et al., 2013](#); [Bertoni et al., 2019](#); [Berman and Couttenier, 2015](#)), making ACLED a trustworthy source for gathering data on political violence.

Figure 2.3: Main Actors of Violence in the West Bank



Notes: Bar chart showing the distribution of violent events by perpetrator and target categories, as classified by ChatGPT 4o-mini. Source: [ACLEDD \(2024b\)](#).

acts against individuals but also those directed at property, such as looting, property destruction, and arson.

To obtain additional information on the nature of the violent events, I feed the textual description to ChatGPT 4o-mini. The model is prompted to classify the main perpetrators of the event, the main targets and the number of casualties involved. Section B in the Appendix describes the procedure. Figure 2.3 shows the output of the classification. The graph displays a general increase in violent events in the West Bank since 2018, as well as the sharp rise in violence perpetrated by Israeli settlers over time, with a notable jump in 2018. It also appears that a new actor emerges in 2022 –i.e. Palestinian Militias, with roughly 2000 events perpetrated between 2022 and 2023, compared to the previous years in which they do not even appear.

Geographic Data In order to get detailed spatial data on the West Bank, I rely on the datasets from two key Israeli civil society organizations. The first is Peace Now, through its Settlement Watch portal that monitors and campaigns against settlement construction. The second is B’Tselem, the Israeli Information Center for Human Rights in the Occupied Territories, which documents a wide range of human rights violations

and maintains an updated database on settlements and outposts in the area.¹² I merged these two data sources to obtain an exhaustive list of settlements and outposts in the area, alongside information on Palestinian towns and villages. The data include location, population and year of establishment of Israeli settlements, location and year of birth of outposts and location of Palestinian towns. After merging, I manually revised the data to remove duplicate information. The map in Figure A.1 provides an overview of merged data. I also retrieve population data of the eleven governorates of Palestine from the [Palestinian Central Bureau of Statistics \(2025\)](#). To disaggregate the official governorate-level data, I developed a method to estimate the population for individual towns and villages, as detailed in Appendix C.

From [B’Tselem \(2024\)](#) I also retrieve data on demolitions of Palestinian dwellings from 2016 to 2023. The data shows the amount of housing units demolished by the Israeli authorities in Palestinian, the amount of number left homeless and the number of minors left homeless. The data also reports the location and date of each demolition event, which I geocode to precise locations in the West Bank. According to [B’Tselem \(2024\)](#) demolitions may occur for mainly three reason: (1) lack of building permits, (2) punitive demolitions as a response to attacks against Israelis, and (3) military needs. Figure A.3 illustrates the quarterly evolution of demolition events by type. Unlawful-construction demolitions dominate the series throughout the period, whereas punitive and military demolitions remain comparatively rare but display occasional spikes associated with periods of heightened tension.¹³ Panel A.3b shows the corresponding quarterly number of housing units destroyed and the number of people—adults and minors—left homeless. The two trends closely track each other, with pronounced surges during episodes of intensified demolition activity (notably in 2016, 2021, and 2023).

Satellite Imagery To obtain precise information on the timing of outpost establishment, I rely on satellite imagery from the European Space Agency’s Copernicus program, specifically the Sentinel-2 Level 2A ([European Union, Copernicus Land Monitoring Service, 2025](#)), as well as satellite data from [Maxar Technologies \(2023\)](#). These two sources

¹²Data from B’Tselem was also used in [Durante and Zhuravskaya \(2018\)](#) to gather information on IDF and Palestinian militias attacks.

¹³Demolitions for lack of building permits occur almost exclusively in Area C, where the Israeli Civil Administration retains full planning and construction authority. Because Palestinians face extremely restrictive planning policies and very low permit approval rates, they are effectively prevented from building legally. As a result, much of the Palestinian residential stock in Area C exists in a state of official unauthorization, which in turn provides the legal basis for demolition orders ([B’Tselem, 2017](#)).

offer complementary strengths. While the Copernicus Sentinel-2 imagery provides a moderate spatial resolution of 10 meters per pixel, its key advantage lies in its high temporal resolution, with frequent flyovers of the study area.¹⁴ In contrast, Maxar data offers a very high spatial resolution of 0.5 meters per pixel, which is crucial for identifying small structures, but its publicly available imagery has limited and infrequent temporal coverage. My approach leverages both datasets in a two-step process. Initially, the high-resolution Maxar imagery is used to precisely identify and locate the specific structures characteristic of new outposts. Following this identification, I code the date of establishment by manually reviewing the imagery and identifying the first appearance of the outpost. This process allows for a high degree of accuracy in determining the timing of outpost establishment. Figure A.6 in the Appendix provides an example of how these two data sources were used.

Election Data To account for local political preferences, I also gather data on Israeli national elections from the [Central Elections Committee for the Knesset \(2025\)](#) database. The data provides detailed results at the polling station level for all national elections. To isolate polling stations located in the West Bank settlements, I geocode the precise location of the polling station with Google Maps API and manually revise inconsistencies. The result is displayed in Figure A.4. For each party, I classify its ideological orientation along a five-category scale—*far left*, *left*, *center/moderate*, *right*, and *far right*—drawing on available historical information and cross-validating coding decisions with Israeli political experts. This classification captures the dominant political blocs in Israeli politics and allows me to track how support for different ideological positions evolves over time across settlements.¹⁵ Table A.2 in the Appendix reports the full classification of parties used in the analysis. Figure A.5 presents vote shares by ideology for elections between 2006 (17th Knesset) and 2022 (25th Knesset). The figure documents a clear rightward shift in national political preferences, with far-right parties increasing their vote share from roughly 30% in 2015 to over 40% in 2022. By contrast, political preferences in the West Bank remained relatively stable over the period, with

¹⁴2 to 5 days depending on the cloud cover threshold parameter and the weather conditions.

¹⁵The ideological classification used in this paper is based *solely* on parties' positions regarding the Israeli–Palestinian conflict and, in particular, the West Bank and settlement expansion. For example, Naftali Bennett's *Yamina* party is coded as "far right" in this framework due to its consistently hawkish positions on settlement growth and annexation, even though on several domestic economic issues it is often viewed as center-right.

support for right and far-right parties consistently amounting to around 50% of the total vote share.

2.4 Methodology

This section outlines the empirical strategy employed to assess the causal impact of post-Trump outpost activity on political violence in the West Bank. I will first describe how the spatial data is structured and then present the identification strategy.

2.4.1 Spatial Data

The first challenge when working with spatial data is to define the spatial units of analysis. To address this, I build a spatial grid starting from the point geometries of the Palestinian towns and villages, as well as Israeli settlements, which remain fixed over time. I then create Voronoi polygons around each point, constituting my spatial units of observation.¹⁶ This approach ensures that each polygon contains a single point, which is either a Palestinian town or an Israeli settlement, and enables me to control for unobserved heterogeneity across localities in the West Bank. Figure A.7 in the Appendix illustrates the Voronoi polygons created around the points of interest.

Having established these spatial units, the next step is to measure the influence of illegal outposts within each polygon. This is based on the theoretical premise that any impact these locations have on violence is likely to be spatial in nature. If outposts are indeed associated with increased friction, their influence would plausibly extend beyond their physical structures and into the surrounding landscape, creating a “friction zone”. To systematically investigate this, a plausible area of influence is measured by first converting outpost locations into a continuous surface by rasterizing the data. Then, for each outpost, a 2 km buffer zone is generated, and the intensity value within it is calculated using a spatial kernel with linear decay, according to the formula:

$$K(d) = \begin{cases} 1 - \frac{d}{d_{max}} & \text{if } d < d_{max} \\ 0 & \text{if } d \geq d_{max} \end{cases}$$

¹⁶For each initial point, a Voronoi polygon is created by enclosing the entire area of surrounding space that is closer to that point than to any other. The boundaries of these polygons are the lines located exactly halfway between each pair of neighboring points.

where d is the distance from the outpost, and d_{max} is the maximum distance of the kernel, i.e. 2 km. This process creates a per-pixel measure of influence that is highest at the outpost's location and declines to zero at the buffer's edge. In cases where the buffer zones of multiple outposts overlap, the final intensity value for any given pixel is the sum of the kernel values from each contributing outpost. As a final step, I overlay the raster data with the Voronoi polygons and calculate the average intensity value for each polygon: the result is a spatially explicit measure of the influence of outpost presence on the surrounding areas. Figure A.8 in the Appendix shows the rasterized outpost influence data and Figure A.9 illustrates the average intensity values for each Voronoi polygon. As a final step, I discretize the intensity values to a binary measure as follows:

$$\text{Outpost}_{i,t} = \begin{cases} 1 & \text{if Intensity}_{i,t} \geq 0.1 \\ 0 & \text{otherwise,} \end{cases} \quad (2.1)$$

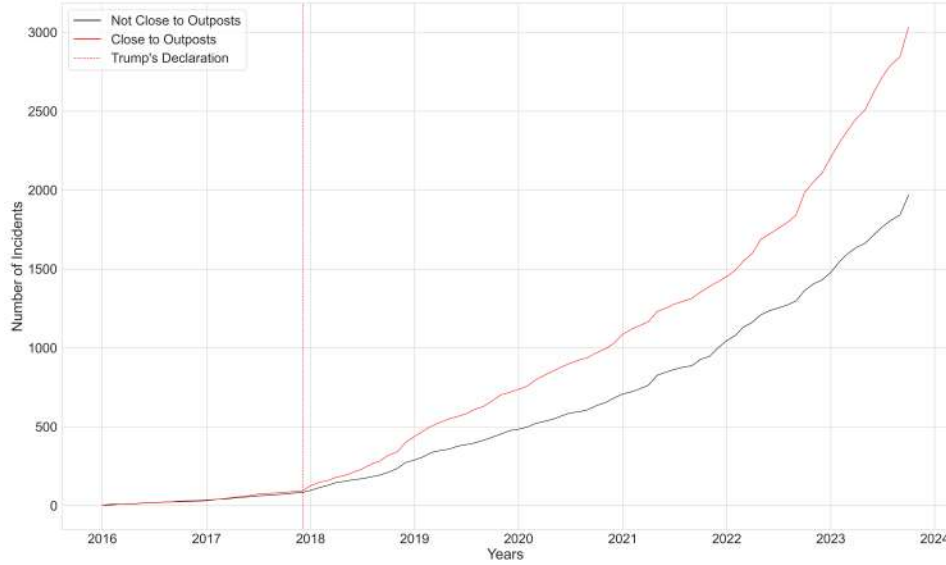
where $\text{Outpost}_{i,t}$ is the binary intensity measure for polygon i at time t . In section 2.5, I will show that the results are robust to different thresholds. Figure A.11 in the Appendix shows the distribution of outpost intensity together with the established threshold. The resulting data structure is a balanced panel, comprising monthly observations for each Voronoi polygon throughout the analyzed timeframe, i.e. from January 2016 to December 2022.

The final dataset reports for each Voronoi polygon a binary measure of outpost influence and violent events occurring in the polygon: this will constitute the backbone of my analysis. From the resulting data, I can characterize violent events happening in high or low outpost intensity areas. Figure 2.4 shows the cumulative number of violent events where settlers are involved over time, both outside and inside the 2 km buffer zone around outposts. From the figure, there is a clear break in violent events following the U.S. recognition of Jerusalem, with a marked increase in violence in the areas close to outposts.

2.4.2 Identification

The identification strategy employed in this study is based on a staggered DiD design, which allows me to exploit the temporal variation in outpost presence relative to the timing of the U.S. recognition of Jerusalem. The main challenge in this context is that outposts are not randomly assigned, and their establishment is likely correlated with

Figure 2.4: Cumulative Monthly Violent Events where Settlers are Involved



Notes: Cumulative number of violent events where settlers are involved over time. The dashed vertical line indicates the date of the U.S. recognition of Jerusalem. The red line indicates events happening within 2 km of an outpost, while the blue line indicates events happening outside the 2 km buffer zone.

unobserved factors that also influence political violence. To address this issue, I restrict the sample only to areas that either already have an outpost present or will eventually see an outpost established. A polygon is then assigned a binary treatment status for a given month if this intensity score crosses the threshold defined in equation 2.1. The control group at any point in time consists of all polygons that have not yet crossed this threshold, which includes both “not-yet-treated” and “never-treated” areas. This approach allows me to control for the unobserved characteristics correlated with the “treatment” status. Robustness checks will be provided to further validate this claim.¹⁷

The empirical analysis covers the period from January 2016 to December 2022. To ensure a never-treated cohort for comparison, data on outpost formation was extended through March 2024. Furthermore, the sample deliberately excludes East Jerusalem, as its distinct political geography and dense urban landscape make the reliable identification of illegal outposts exceptionally difficult. This selection process results in a strongly balanced panel of 493 Voronoi polygons, each observed for 92 months. Detailed descriptive statistics and the spatial distribution of this sample are provided in the Ap-

¹⁷In Section 2.8 I also consider a non-staggered specification where I drop “not-yet-treated” Voronoi polygons just before switching to treatment.

pendix (Tables A.1 and A.3). To validate the comparison, a pre-treatment balance check between the treated and control groups reveals no statistically significant differences in prior violence, as confirmed by a Kolmogorov-Smirnov (KS) test (Table A.4).¹⁸ The one notable distinction is that treated areas are, on average, more populated.

In order to estimate the causal effect of post-Trump outpost activity on political violence, I estimate the following regression model:

$$y_{i,m} = \alpha_i + \theta_m + \beta(\text{Outpost}_{i,m} \times \text{Declaration}_m) + \gamma \text{Pop}_{i,m} + \lambda_i \times k + \epsilon_{im} \quad (2.2)$$

where $y_{i,m}$ is the outcome variable of interest, α_i are the Voronoi polygon fixed effects, θ_m are the month-year fixed effects, $\text{Outpost}_{i,m}$ is a binary indicator that takes the value of 1 if polygon i has an outpost intensity larger than the threshold defined in 2.1 in month-year m , and Declaration_m is a binary indicator that takes the value of 1 if month-year m is after the U.S. recognition of Jerusalem. In an additional specification, I also control for a Voronoi polygon and Knesset linear trend ($\lambda_i \times k$), to account for other domestic political shocks that may have affected the violence in the area, and population estimates ($\text{Pop}_{i,m}$). Here, β is the coefficient of interest, which captures the differential effect of outpost activity on political violence after the U.S. recognition of Jerusalem. The outcome variables considered are the total number of violent events, the number of violent events involving Israeli settlers, and the number of violent events targeting civilians. Leveraging the classification of incident explained in Section 2.3, I can also disentangle the dynamics of violence by distinguishing perpetrators and targets.

Another important aspect of the identification strategy is the staggered design implied by the model. Although many outposts were established well before December 2017, many others emerged after (cf. Figure 2.1). Hence, there will be a first cohort of treated units that will be exposed to the treatment immediately after the declaration—the large majority, while other units will be treated later.¹⁹ To account for this staggered treatment, I will use the event study approach proposed by Borusyak et al. (2024). The counterfactual will consider both not-yet-treated and never-treated units. Event studies

¹⁸A KS test was chosen because the variables follow a power-law distribution. Unlike a standard t-test, which assumes data is normally distributed, the non-parametric KS test makes no such assumption about the underlying distribution, making it appropriate for the case at hand.

¹⁹326 of the Voronoi polygons are treated as the first cohort, 100 are treated later in time, and the remaining 67 are never-treated, as reported in Figure A.10.

will also be considered to visualize the dynamic effect of outpost emergence and to check for pre-trends.

2.5 Results

This section presents the main empirical findings. I will first consider the impact of outpost emergence post-Trump on political violence in the West Bank, and I will later present results looking at perpetrators and targets of violence.

2.5.1 Main Results

The empirical results in Table 2.1 show that outpost-proximate areas experienced a sharp and statistically significant increase in political violence after Trump’s declaration, with estimates being robust to the inclusion of Voronoi polygon-Knesset linear trends and population controls. The findings reveal a staggering 150% (column (2)) increase relative to the pre-treatment mean in violent events. The event study plot in Figure 2.5 confirms this finding, illustrating a sharp, immediate, and sustained jump in violence right after the political shock, with no evidence of pre-existing trends.

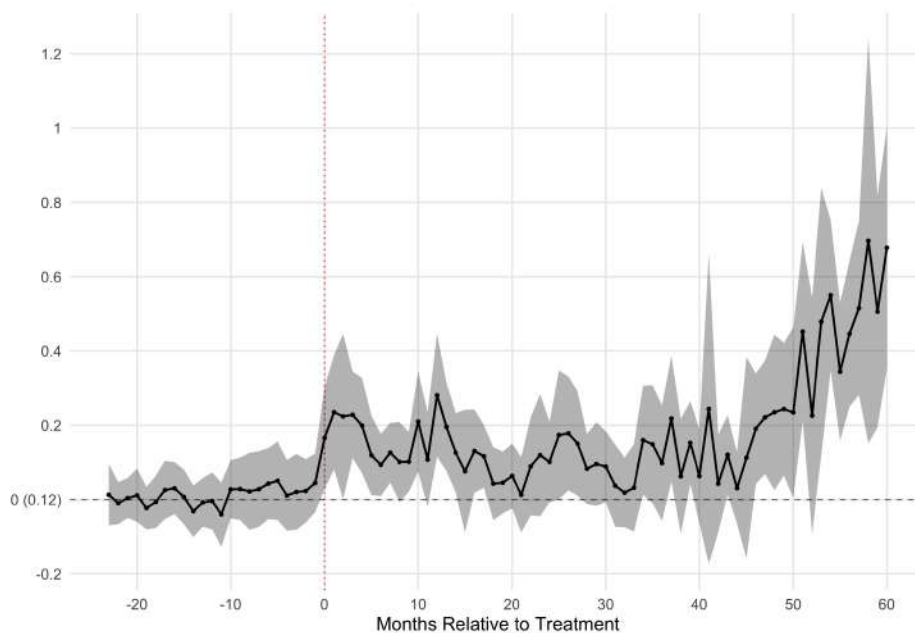
Table 2.1: Effect of Outpost Establishment on Violent Events

	Violent Events					
	All Events		Setter Involved		Civilian Targeting	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Outpost	0.153*** (0.046)	0.172*** (0.042)	0.060*** (0.016)	0.065*** (0.015)	0.043*** (0.010)	0.042*** (0.010)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓
Pre-treatment Mean	0.115	0.115	0.033	0.033	0.031	0.031
<i>N</i>	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

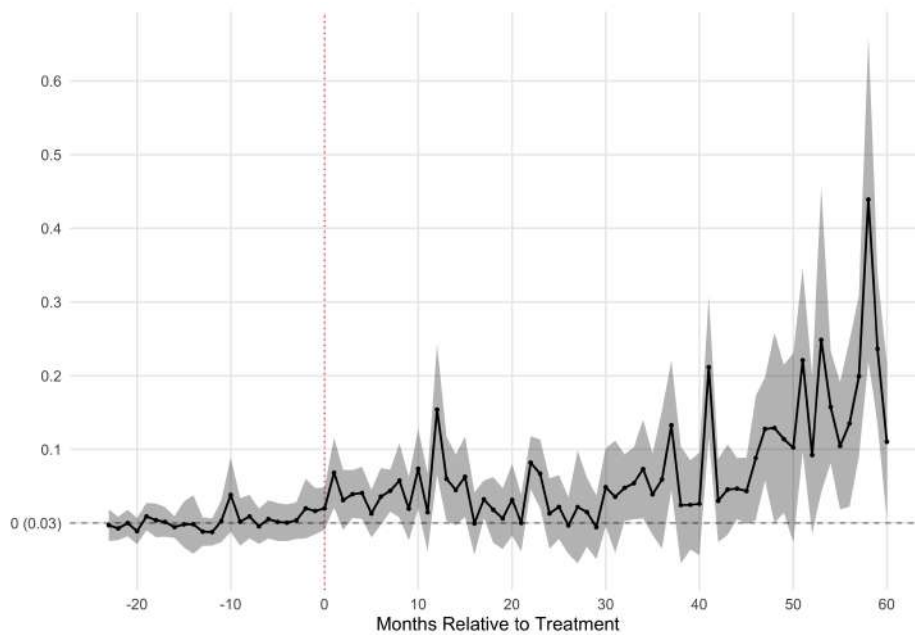
The effect is also significant for violence involving settlers and the targeting of civilians. Incidents targeting civilians increased by approximately 140% from their pre-declaration levels. The most dramatic rise was observed in events involving Israeli settlers, which surged by roughly 190% relative to the pre-treatment average. The event

Figure 2.5: Event Study: Number of Events



Notes: This figure presents the event study for the total number of violent events as outcome variable from Table 2.1. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

Figure 2.6: Event Study: Violent Events with Settlers Involved



Notes: This figure presents the event study for events in which settlers are involved from Table 2.1. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

study plots in Figures 2.6 and A.12 confirm this pattern, showing a clear and immediate spike in violence after the announcement, with no evidence of pre-existing trends. A common feature of these plots is an initial sharp increase that stabilizes for about three and a half years, followed by an exponential rise, suggesting the shock may have initiated a longer-term escalation of violence.

2.5.2 Perpetrators and Targets of Violence

To better understand these dynamics, the analysis disaggregates the violence by perpetrator and target. The results, presented in Table 2.2, reveal a clear and asymmetric pattern in the escalation of violence following Trump’s declaration. The analysis focuses specifically on violence perpetrated by or targeting the conflict’s primary non-state actors: Israeli settlers, Palestinian civilians, and Palestinian paramilitary formations. This approach allows for an examination of grassroots conflict dynamics, distinct from official state-led operations. It is worth noting that I only consider instances where there is a clear attribution of responsibility, excluding events with ambiguous or unknown perpetrators or targets, e.g. ignoring cases where the classification outcome is “Combination of actors”.²⁰ This ensures that the analysis accurately captures the specific dynamics between these key groups.

Table 2.2: Heterogeneous Effects by Target and Perpetrator

	Main Target				Main Perpetrator			
	Settlers		Palestinian Civilians		Settlers		Palestinians	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post $\times$ Outpost	0.004 (0.003)	0.003 (0.002)	0.115*** (0.025)	0.121*** (0.024)	0.058*** (0.013)	0.062*** (0.013)	0.014 (0.017)	0.024** (0.012)
Unit FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓		✓
Pre-treatment Mean	0.009	0.009	0.082	0.082	0.035	0.035	0.012	0.012
<i>N</i>	41,412	41,412	41,412	41,412	41,412	41,412	41,412	41,412

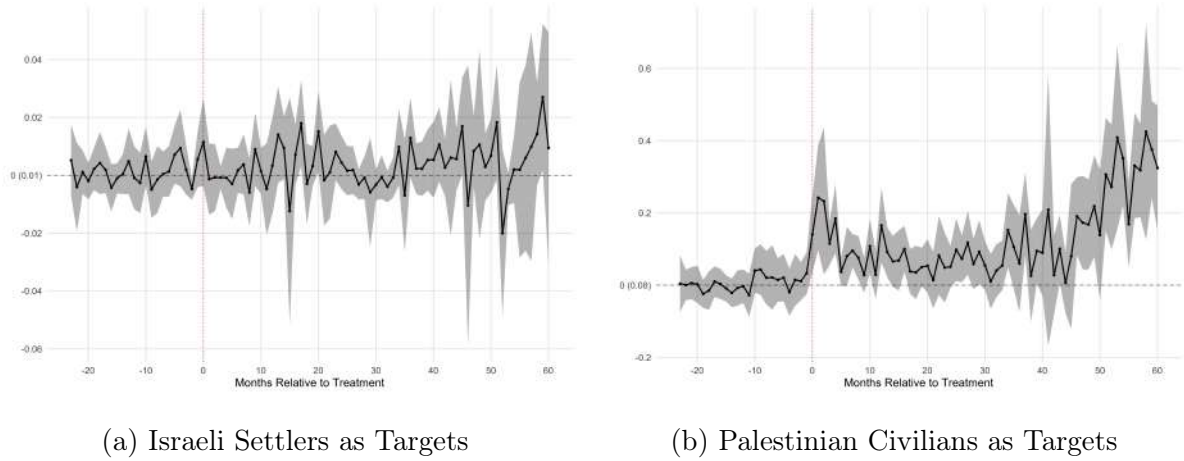
Notes: Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

The most immediate and substantial effect was a surge in violence perpetrated by Israeli settlers. Settler-perpetrated violence in outpost-proximate areas rose by over 175% compared to the pre-declaration average. The corresponding event study plot

²⁰See Figure 2.3

in Figure 2.8a clearly illustrates this dynamic, showing an immediate, sharp increase in violence by settlers right after the shock, which was sustained and ultimately grew over the final years. This violence was primarily directed at Palestinian civilians, who experienced a 140% rise in attacks in these areas from the baseline. The event study in Figure 2.7b confirms that attacks on Palestinian civilians followed the same immediate and sustained upward trend.

Figure 2.7: Event Study: Targets of Violence

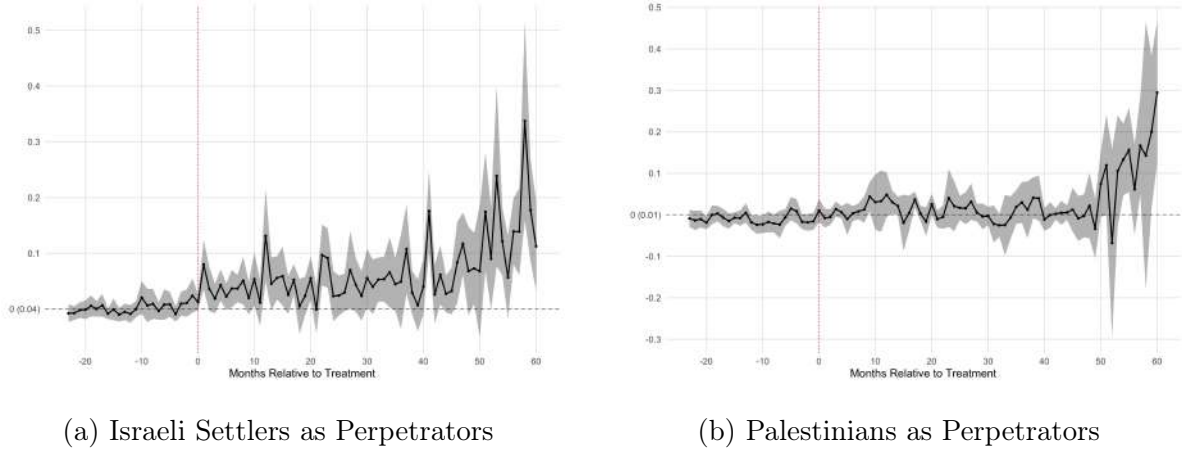


Notes: This figure presents the event studies for different targets of violence from Table 2.1. Panel (a) shows events where Israeli settlers are targeted, while Panel (b) shows events where Palestinian civilians are targeted. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

In contrast, the increase in violence perpetrated by Palestinians is delayed. While the coefficient is statistically significant, the event study in Figure 2.8b shows that there is no discernible effect for nearly 50 months after the treatment time. Only in the later period of the analysis does a significant upward trend emerge. Although the coefficient is insignificant in Table 2.2, it is possible to observe a similar pattern for events targeting settlers in Figure 2.7a. This suggests that the initial shock did not immediately trigger a parallel rise in Palestinian-led violence but rather that such actions emerged later.

This escalating pattern suggests that the initial shock did more than just cause a one-time increase in violence; it appears to have triggered a self-perpetuating cycle of conflict. The sharp, initial increase, largely driven by settlers, may have created an environment of heightened tension and grievance. This could explain the subsequent, steadier rise in violence observed after nearly four years. It is unlikely, however, that the eventual uptick in Palestinian-perpetrated violence was directly caused by the declaration itself; rather, it seems to have emerged from the violent climate the declaration helped generate. In

Figure 2.8: Event Study: Perpetrators of Violence



Notes: This figure presents the event studies for different perpetrators of violence from Table 2.1. Panel (a) shows events where Israeli settlers are perpetrators, while Panel (b) shows events where Palestinians are perpetrators. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

the Section 2.7, I provide descriptive evidence to better understand these dynamics and to shed light on the evolution of Palestinian-originated violence over time.

2.6 The Declaration as a Trigger for Violence

To isolate the declaration’s impact from the simple presence of new outposts, this section analyzes the period before the diplomatic shock. I examine whether the creation of outposts prior to December 2017 was, in itself, sufficient to trigger the levels of violence observed afterward. This approach directly tests the alternative explanation that the surge in violence was merely a continuation of a pre-existing trend, rather than a distinct reaction to the U.S. announcement.

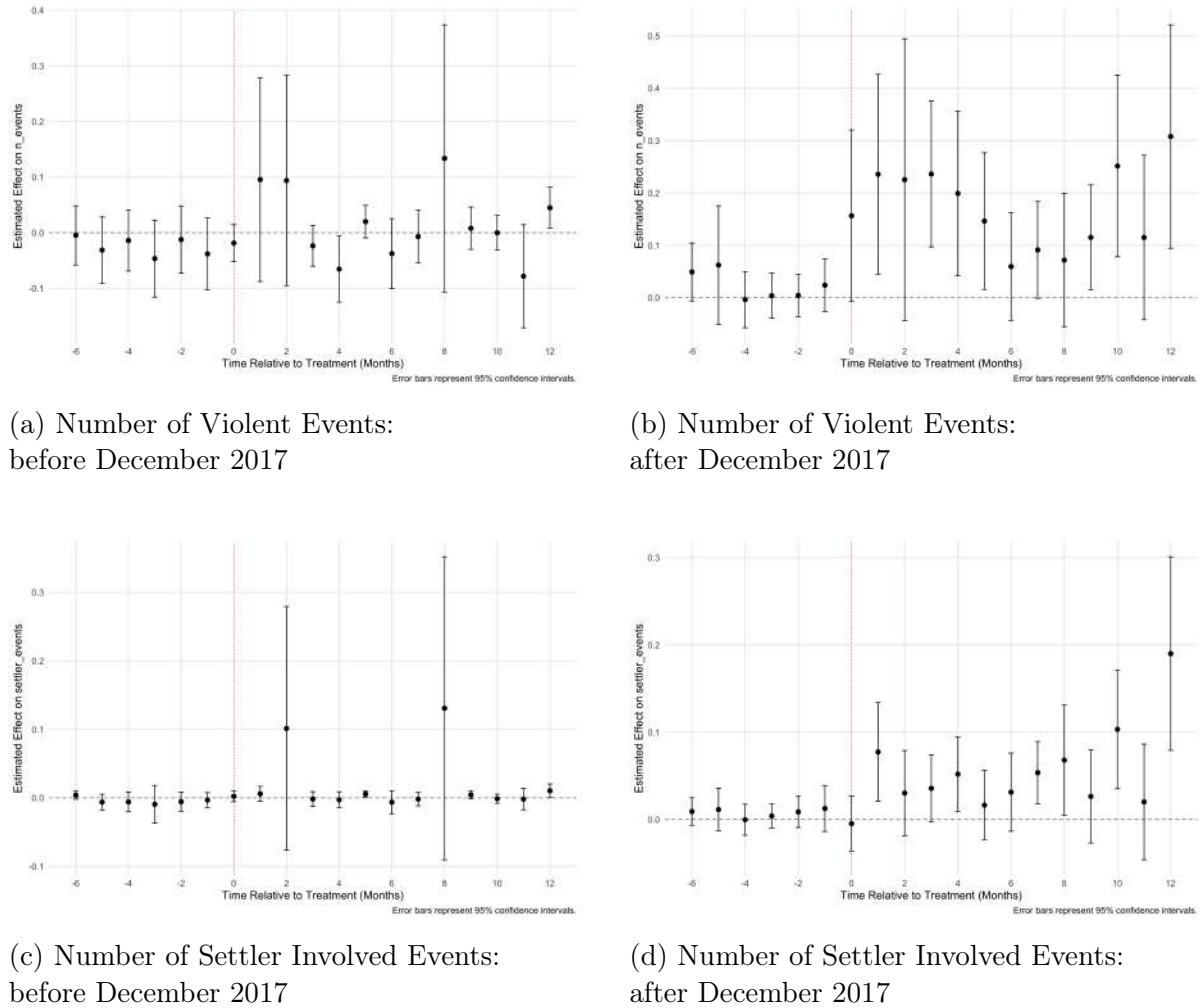
To do this, I only consider a subset of outposts: “the pre-shock cohort”. These are all the Voronoi polygons that saw an outpost intensity growing substantially²¹ between January 2016 and November 2017 (included).²² For this cohort, the treatment status is the establishment of the outpost in a period of relative diplomatic status quo. The control group comprises areas where outposts were established only after the declaration. This design allows us to observe what happens when a new outpost appears without

²¹According to equation 2.1

²²Due to ACLED’s data limitations, the first data point is the 1st of January 2016.

the powerful political endorsement provided by the Trump administration. This design accounts for 10 Voronoi polygons being treated and 177 used as control units.

Figure 2.9: Event Studies: Declaration as a Structural Break



Notes: This figure presents the event studies for two outcomes: the total number of violent events (top row) and the number of violent events involving settlers (bottom row). Panels (a) and (c) show the event studies considering only outposts established before December 2017, while Panels (b) and (d) show the event studies displayed from the main analysis considering for a shorter time frame, i.e. six lags and twelve leads. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The estimation equation includes the linear time trend.

To facilitate a comparison, I re-estimate the event study from Section 2.5, focusing on a shorter horizon of six months before and twelve months after the treatment. In this specification, the control group of never-treated units is substantially larger, as it now includes all polygons where outposts were established only after December 2018. Figure 2.9 displays the comparative event study for two outcomes: total violent events and

those involving settlers. The results for both are unambiguous. In the months after an outpost was established, the estimated coefficients hover near zero and are statistically insignificant. This null finding indicates that while these outposts were undoubtedly sources of local tension, their mere presence was not enough to independently trigger a systematic escalation of violence.

Figures [A.15](#) and [A.14](#) display the results for the other outcomes.²³ The pattern is identical: with the exception of some isolated incidents (high magnitude but statistically insignificant coefficient), the effect of outpost emergence before Trump’s declaration is negligible. This suggests that any significant reactions were confined to a small, non-representative minority of cases, thus not reflecting a systematic trend. The evidence suggests that only following the U.S. recognition of Jerusalem did outpost creation become a hotspot for widespread violence.

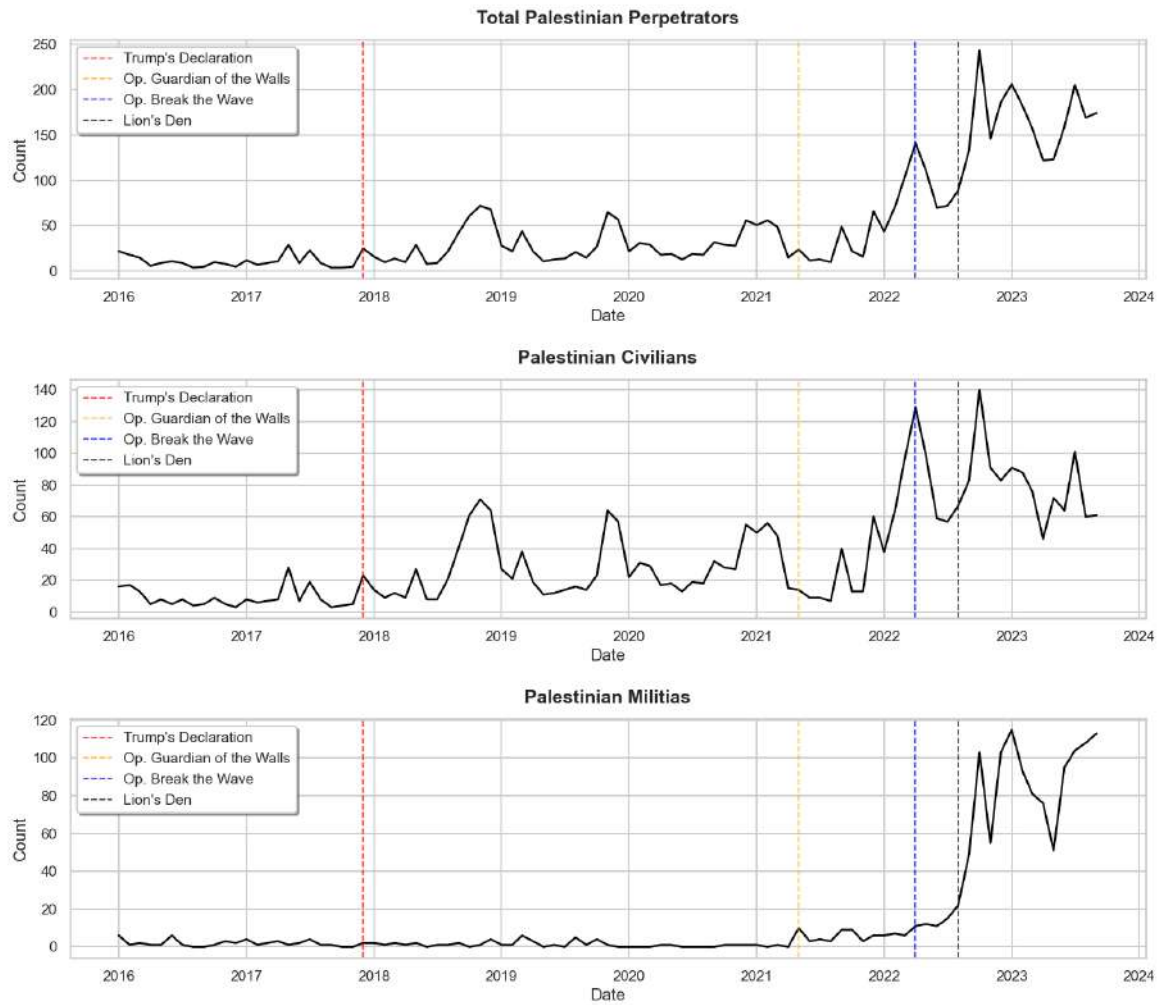
2.7 Palestinian Perpetrated Violence

The event–study estimates highlight a clear temporal pattern: Palestinian violence does not rise immediately after Trump’s declaration, nor does it rise significantly after an outpost is built before December 2017. However, the results show a substantial and significant increase only several years later, a pattern that still calls for further descriptive investigation. To bring descriptive evidence to this pattern, Figure [2.10](#) plots the monthly evolution of Palestinian attacks in the West Bank, along with some important dates in the years following Trump’s declaration: the IDF operations Guardian of the Walls in May 2021 and Break the Wave starting March 2022, as well as the emergence of the Palestinian militia Lions’ Den, officially in August 2022. Similarly, Figure [A.16](#) shows Israeli perpetrated violence by actor type, i.e. Settlers and IDF.

Figure [2.10](#) shows that Palestinian violence experiences a mild increase following Trump’s declaration, though the rise is far less pronounced than the escalation observed in Israeli-perpetrated violence (see [A.16](#)). The months after the declaration display several small peaks, consistent with a tit-for-tat dynamic in response to Israeli settler attacks and, to a lesser extent, IDF-perpetrated incidents, both of which exhibit a substantial and sustained upward trend. Palestinian militias remain largely inactive during this period. In May 2021, however, both settler and IDF violence spike sharply. This

²³See also Tables [A.5](#) and [A.6](#) in the Appendix.

Figure 2.10: Monthly Palestinian Attacks over Time



Notes: This figure presents the monthly number of Palestinian perpetrated attacks broke down by Palestinian actor type, i.e. civilians and militias. Vertical dashed lines indicate important dates: the Trump's declaration (December 2017), IDF operations in the West Bank and Gaza strip, the emergence of the paramilitary group Lions' Den.

episode appears to mark the beginning of two processes: (i) a gradual upward trend in violence carried out by Palestinian civilians, and (ii) the reactivation of Palestinian militias, which show a noticeable but still modest increase in attacks. By the onset of the Break the Wave operation, civilian-perpetrated violence has already risen to levels significantly higher than those observed before 2021, and militia activity reaches an inflection point. When the Lion’s Den militia formally emerges in late 2022, its involvement grows rapidly, with militia violence increasing exponentially and remaining at very high levels until the end of the sample period in September 2023.

The monthly series of violent events suggests that the post–Trump-declaration period is not homogeneous. In particular, there is a first phase between early 2018 and the May 2021 escalation, during which Palestinian violence remains comparatively subdued and mostly takes the form of small peaks, and a second phase from mid-2021 onwards, in which both Palestinian civilian and militia violence increase markedly and display a persistent upward trend. In what follows, I therefore distinguish between these two regimes and explicitly model the dynamics of Palestinian-perpetrated violence before and after May 2021.

To focus on places where Palestinian and Israeli actors effectively live “toe to toe”, I restrict the sample to Voronoi polygons that either lie within Area C or border it. These are the localities where Israeli settlements, Palestinian villages and IDF presence are most spatially intertwined, and where one would most plausibly expect retaliatory dynamics to operate. For each Voronoi cell i and quarter t , I estimate two-way fixed-effects models of the form

$$y_{it}^{(k)} = \alpha_i + \gamma_t + \beta_1 \text{Settler}_{i,t-1} + \beta_2 \text{Settler}_{i,t-2} + \delta_1 \text{IDF}_{i,t-1} + \delta_2 \text{IDF}_{i,t-2} + \theta_1 \text{Demo}_{i,t-1} + \theta_2 \text{Demo}_{i,t-2} + \varepsilon_{it}. \quad (2.3)$$

where $y_{it}^{(k)}$ is alternatively (i) total Palestinian-perpetrated events, (ii) events perpetrated by Palestinian militias, or (iii) events perpetrated by Palestinian civilians. $\text{Settler}_{i,t}$ and $\text{IDF}_{i,t}$ denote the number of violent events perpetrated by settlers and by the IDF in cell i at time t , while $\text{Demo}_{i,t}$ is the number of housing units demolished. All covariates enter with one- and two-quarter lags to capture delayed reactions to past Israeli actions. The specification includes Voronoi fixed effects α_i and time fixed effects γ_t , and standard errors are clustered at the Voronoi level. This specification will help to bring descriptive

evidence on the retaliatory dynamics of Palestinian violence after Trump’s declaration

To allow for spatial diffusion of retaliation, I then augment this baseline model with spatial lags of the lagged settler and IDF variables.²⁴ This extended specification allows Palestinian violence in a given locality to respond not only to past Israeli actions in the same cell but also to recent Israeli violence in neighbouring cells, which is a natural way of capturing local diffusion and indirect retaliation. I estimate these models separately for the two post-declaration periods. Panel A of Table A.7 covers the interval from January 2018 to May 2021, while Panel B covers June 2021 to September 2023.

The estimates in Table A.7 reveal a clear structural break in the dynamics of Palestinian-perpetrated violence before and after May 2021. In the earlier period (Panel A), Palestinian violence appears largely unresponsive to lagged settler attacks, IDF actions, or demolitions. Coefficients are small, imprecisely estimated, and often negative, suggesting either the absence of systematic retaliation or the dominance of other channels of response during this phase. This aligns with the descriptive evidence showing only mild and sporadic peaks in Palestinian activity between 2018 and early 2021.

In the period following May 2021 (Panel B), the pattern changes markedly. Lagged settler violence becomes a strong and statistically significant predictor of subsequent Palestinian attacks, particularly for civilian-perpetrated violence: both the one-quarter and two-quarter lags of settler events are positive and sizeable in magnitude. These effects are an order of magnitude larger than in the pre-2021 period, consistent with the emergence of a more pronounced tit-for-tat or grievance-accumulation mechanism. Lagged IDF violence remains generally negative, especially for militia activity, suggesting that direct military operations may temporarily suppress organised armed groups even as settler attacks fuel broader civilian mobilisation. Spatial lags, by contrast, do not exhibit consistent effects, indicating that these dynamics are primarily local rather than diffused across neighbouring communities.

These findings help clarify how the impact of Trump’s declaration unfolded over time. In the immediate post-announcement years, Palestinian retaliation mechanisms appear either largely dormant or institutionally constrained, with no systematic connec-

²⁴That is estimating the following equation:

$$y_{it}^{(k)} = \alpha_i + \gamma_t + \beta' \mathbf{X}_{i,t-1:t-2} + \phi' W \mathbf{X}_{i,t-1:t-2} + \varepsilon_{it}, \quad (2.4)$$

where $\mathbf{X}_{i,t-1:t-2}$ stacks the own-cell lags of settler and IDF violence and demolitions, and $W \mathbf{X}_{i,t-1:t-2}$ are their corresponding spatial lags, constructed using a queen-contiguity weights matrix over the Voronoi tessellation.

tion between past Israeli actions and subsequent Palestinian violence. By contrast, after May 2021 the dynamics shift sharply: lagged settler attacks become a strong and highly localised predictor of Palestinian-perpetrated violence, while IDF operations exhibit suppressive effects on militia activity. The magnitude, sign, and temporal structure of the coefficients all point to the emergence of a more active grievance–response mechanism in the later period, consistent with both the descriptive evidence and the timing of new militia mobilisation. As emphasised above, these patterns are associative rather than causal, but they provide a suggestive map of how Palestinian behaviour evolved in the years following the declaration.

Viewed through the lens of [Besley and Persson \(2011\)](#), the evolution of violence in the West Bank can be interpreted as a sequence of movements across different conflict equilibria. Before the Trump declaration, conditions broadly resembled their peaceful equilibrium: although low-level violence and repression existed, institutional constraints were relatively stronger, political polarization had not reached levels that made systematic violence attractive, and neither side had clear incentives to escalate. The declaration appears to have disrupted this balance by eroding social norms, increasing political polarization within Israel—particularly by empowering radical settler actors—and signalling a greater degree of tolerance for settler violence (see Figure [A.16](#), before May 2021). After the events of May 2021, polarization between settlers and Palestinians intensified further, and the IDF increasingly engaged directly in the conflict rather than constraining settler actions, weakening institutional constraints even more. The rise of new Palestinian armed groups in this period is consistent with a movement toward an equilibrium in which both sides find violence strategically attractive, aligning with [Besley and Persson \(2011\)](#) conditions for the emergence of mutually reinforcing political violence.

2.8 Robustness Checks

This section presents a series of robustness checks to validate the main findings of the study. The goal is to ensure that the results are not driven by specific assumptions or methodological choices.

Endogeneity in Outpost Emergence There is a major concern regarding the potential endogeneity of outpost emergence through violent events against Israeli settlers. If outposts are established in response to violence targeting settlers, this could bias the

results. To address this issue, I consider the following alternative specification:

$$\begin{aligned} \text{Outpost}_{i,m} = & \alpha_i + \theta_m + \sum_{k=1}^{12} \mu_k \text{Casualties}_{i,m-k} \\ & + \sum_{k=1}^{12} \delta_k \text{Settlers}_{i,m-k} + \epsilon_{im} \end{aligned} \quad (2.5)$$

where $\text{Outpost}_{i,m}$ is a binary variable as defined in equation 2.1. $\text{Casualties}_{i,m}$ is the number of Israeli casualties, and $\text{Settlers}_{i,m}$ is the number of Israeli settlers targeted in violent events. I consider 12 lags to check for any delayed effects. The results displayed in Figure A.17 show that no coefficient is statistically significant, suggesting that the emergence of outposts is not driven by previous violent events against Israeli settlers.

Placebo Test To further validate the causal interpretation of the results, I conduct a placebo test by artificially shifting the treatment date six months earlier. If the observed effects were driven by unobserved trends or other confounding factors rather than the Jerusalem declaration itself, there should be a significant effect even with this falsified timing. Figures A.18 and A.19 report the results of the placebo tests. From the output, there is no discernible treatment effect of this falsified treatment timing, corroborating the main findings.

Alternative Thresholds for Outpost Intensity Another aspect that may be driving the results is the choice of intensity thresholds for defining outposts. To explore this, I re-estimate the main models using different thresholds (e.g., 0.01, 0.05, and 0.12) and assess the robustness of the findings. Tables A.8 and A.9 present the results for these alternative specifications. The findings in Table 2.1 and in Table 2.2 are robust to changes in the outpost intensity threshold. The only main difference is on column (7) of Panel A in Table A.9, where the coefficient is statistically significant, as compared to the same column on Table 2.2. After controlling for population and introducing a linear trend, i.e. column (8), the coefficient becomes very similar to the original estimate. Figures A.20 and A.21 display the event studies. Although in the aggregate the ATE is rather similar, I can find some differences on the estimated effects, most notably for the last 10 to 15 months in the sample. Aside from these small differences, the results seem globally

robust when considering alternative thresholds.

Synthetic Control To check for potential violations of the parallel trends assumption, I also consider using a synthetic control approach with staggered adoption, as proposed by Ben-Michael et al. (2022). To mitigate the risk of weight extrapolation, the donor pool for the synthetic control is restricted to units that either have existing outposts or will have them established in the future. Results are displayed in Figures A.22 and A.23. Not only do the pre-trends display a good match between control and treatment areas, but also the recovered ATT is generally consistent with the main findings. The main differences with respect to the baseline results concern Figure A.23b and A.23c where there is no significant effect. This is mainly due to the fact that the increase around post-treatment month 50 is not significant in this case. Conversely, the initial jump in violent events and attacks against Palestinian civilians is still significant and comparable in magnitude.

Non-Staggered Specification To address concerns about staggered treatment adoption, I employ an alternative non-staggered DiD model. In this approach, not-yet-treated units serve as the control group, and are subsequently dropped from the analysis once they switch to treatment status. This method provides a cleaner comparison by avoiding issues related to treatment effect heterogeneity over time. The results displayed in Tables A.10 and A.11 and in Figures A.24 and A.25 from this specification are largely consistent with the original findings, confirming the immediate impact of the declaration. Minor differences emerge in the long-term estimates, but the core result of a significant increase in violence remains robust.

Alternative Buffer Areas The choice of a 2 km buffer area around outposts is somewhat arbitrary. Figures A.26 and A.27 show the results for all outcomes with different buffer areas, namely 1, 2 –e.g. the baseline results, 3 and 4 kilometers. The results do not drastically change, although as expected, there is a decline in the magnitude of the estimated coefficient as the buffer size increases. The only two exceptions concern A.27b and A.27a, where the estimated impact on violence perpetrated by settlers and Palestinians does not uniformly decline with distance. This non-linear pattern might suggest a displacement of conflict to a wider periphery around the outposts, or it could capture different types of violent interactions occurring at varying distances from the

outpost itself.

Leave-Out Regression Analysis To rigorously test the robustness of the findings, I conduct a leave-out validation analysis. The purpose of this test is to ensure that the main results are not driven by a small or specific subset of treated observations. The procedure involves iteratively re-estimating the main specification, i.e. Equation 2.2. In each iteration, I randomly select and exclude a different group of five treated units (Voronoi polygons) without replacement.²⁵ Each regression specification is estimated on the full dataset minus only the five units excluded in that specific iteration. Figures A.28 and A.29 display the results of this exercise by plotting the distribution of the estimated β coefficients. The results remain consistent with the baseline results in terms of sign and magnitude. This robustness indicates that the exclusion of any random group of treated units does not meaningfully alter the conclusions.

Inverse Hyperbolic Sine Transformation Given the count nature of the outcome variables, I also re-estimate the main specification using an inverse hyperbolic sine (IHS) transformation of the dependent variable. This transformation is particularly useful for handling zero values while still allowing for interpretation similar to a log transformation. The IHS transformation is defined as:

$$IHS(y) = \log(y + \sqrt{y^2 + 1})$$

This transformation helps to stabilize variance and make the distribution of the outcome variable more symmetric. The results of the IHS transformation are presented in Tables A.12 and A.13. Overall, the findings remain consistent with the main results, further validating the robustness of the analysis.

Alternative Clustering of Standard Errors To ensure the inferences are robust to potential spatial autocorrelation, I also employ an alternative clustering strategy for the standard errors. This approach aggregates the base Voronoi polygons into larger clusters based on both geographic contiguity and political-administrative jurisdiction. The rationale is that unobserved shocks affecting political violence may not be confined

²⁵The number of iterations is set to $\min(100, \lfloor N_{treated}/5 \rfloor)$, where $N_{treated}$ is the total number of treated units.

to a single polygon but could spill over to neighboring areas that share a similar political or security environment.

Operationally, I define a cluster as a set of polygons that satisfy two conditions: 1) they must be spatially contiguous, and 2) they must belong to the same administrative area. I define spatial contiguity using the Queen criterion, where polygons are considered neighbors if they share at least one edge or vertex. The administrative areas are defined by the unique combination of their Oslo Accord designation (i.e., Area A, B, C, or the special case of Hebron) and the Palestinian Governorate they fall within.

The algorithm identifies these clusters using a Breadth-First Search (BFS). It begins at an arbitrary polygon and explores its neighbors. Any neighboring polygon belonging to the same administrative area is added to the current cluster. The search then expands outward from these newly added polygons until no more contiguous polygons with the same classification can be found. This process is repeated until every polygon in the dataset is assigned to a unique, spatially contiguous administrative cluster. The procedure yields 118 unique clusters. Clustering standard errors at this level provides a more conservative test by accounting for correlated shocks within these geographically and politically coherent regions. Figures [A.30](#) and [A.31](#) give a graphical illustration of the clustering procedure and the final clusters used for standard error adjustment. Results are displayed in Tables [A.14](#) and [A.15](#). The results are robust to this alternative clustering strategy.

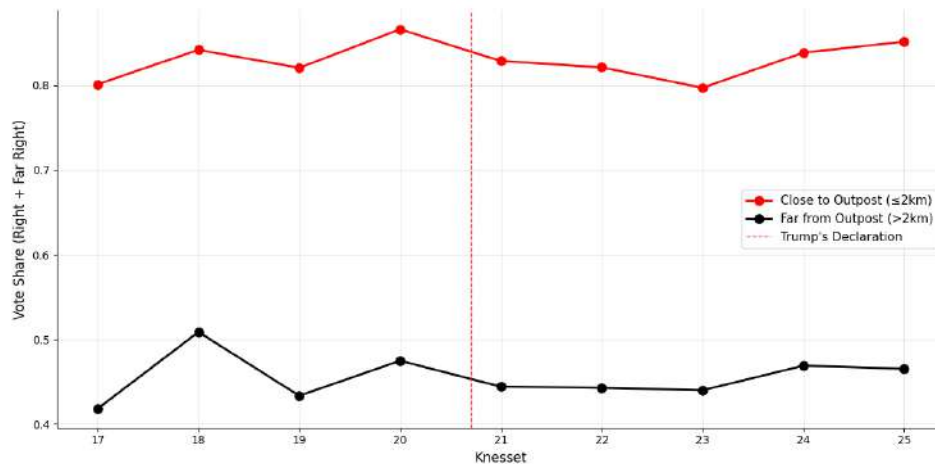
2.9 Discussion: The Domestic Political Environment and the Post-2017 Escalation

To address the concern that the observed escalation might reflect domestic political dynamics rather than the diplomatic shock itself, this section examines long-run political trends and key Israeli policy developments surrounding the period of analysis. The goal is not to treat these as competing explanations but to assess whether identifiable shifts in state behavior or political preferences could account for the timing of the post-2017 increase in violence. I therefore draw on descriptive evidence on electoral patterns, fiscal transfers and administrative practices to contextualize the empirical findings.

An important alternative explanation for the post-2017 escalation is that it reflects concurrent political developments within Israel rather than the effect of the U.S. decla-

ration. To evaluate this possibility, I examine several domestic channels. First, I look at polling-station-level election results from 2006–2022 to assess whether there were any sizable shifts in political preferences over the period. Using Figure A.4, I identify polling stations located near outposts and those farther away, and I track the evolution of vote shares for right-wing and far-right parties across electoral cycles. If domestic ideological realignments were driving the escalation, we would expect to see marked increases in support for these parties, especially in outpost-proximate areas, in the years immediately preceding or following the declaration. Instead, the trends are remarkably stable: while support for right-wing blocs is consistently high in outpost areas, there is no evidence of a discrete ideological shift that coincides with the timing of the observed increase in settler-perpetrated violence. That said, parties themselves may have shifted their rhetorical or ideological positioning over time. To address this, future work will employ text analysis on a comprehensive corpus of Knesset speeches spanning 1993–2024²⁶ to detect any relevant shifts in the ideological slant concerning Palestinians and the occupied territories.

Figure 2.11: Right & Far Right Vote Shares in the West Bank



Notes: Share of votes in the West Bank for right-wing and far-right parties, differentiated by proximity to outposts. The red dashed line indicates the U.S. recognition of Jerusalem.

A second way to assess Israel’s domestic political environment is to examine observable state actions in the occupied territories, particularly fiscal transfers to settlements and demolition activity in Area C. Transfers to settlements display a marked rise in the years surrounding Donald Trump’s election in late 2016, consistent with the increasingly

²⁶Retrieved from [Goldin et al. \(2024\)](#)

pro-settlement orientation of the Netanyahu governments during this period, although the data I employ end in 2018 and therefore cannot speak to subsequent developments. However, from 2016 to 2017, government transfers to settlements increased by roughly 50% and remained stable in 2018 (see Figure [A.32](#)).

Demolitions, by contrast, show no discrete break around 2017/2018 but instead exhibit a gradual upward trend beginning in 2020–2021, a period that coincides with two major political developments: the unveiling of the Trump Administration’s “Peace to Prosperity” plan in January 2020—which explicitly entertained Israel’s annexation of large parts of the West Bank—and the subsequent Israeli government statements preparing for possible *de jure* annexation later that same year (ultimately suspended in the context of the Abraham Accords) ([Immenkamp, 2020](#)). These shifts suggest a progressively hardening state posture toward the occupied territories, yet demolitions evolve slowly and do not match the immediate and localized discontinuity in violence identified after the December 2017 declaration. Instead, it appears to form part of a broader permissive environment within which the diplomatic shock operated, rather than constituting independent triggers for the sharp post-declaration escalation.

A further development in this broader political backdrop is the formation, at the end of 2022, of the most right-wing governing coalition in Israel’s history, including parties and figures whose public statements on the conflict became markedly more confrontational ([BBC News, 2022](#); [The Times of Israel, 2025](#)). Although this political realignment occurs toward the end of the sample period and cannot account for the initial post-2017 discontinuity in violence, it underscores the longer-run trajectory of domestic politics: rather than moving toward de-escalation, the political environment increasingly featured actors whose rhetoric signalled a more assertive approach to the West Bank. This trend is consistent with the descriptive evidence of a gradually hardening policy climate.

Taken together, these patterns provide circumstantial but informative evidence about the broader political environment in which the escalation unfolded. While they do not allow us to isolate a definitive causal role for domestic political dynamics in triggering the post-2017 increase in violence, they likewise offer no indication of any discrete policy shift aimed at curbing settler–Palestinian tensions in the relevant period. Instead, the stability of electoral preferences, the gradual nature of administrative and fiscal trends, and the slow-moving character of demolition activity suggest a permissive and increasingly assertive policy backdrop rather than an active effort to de-escalate. These observations reinforce the interpretation that the diplomatic shock interacted with an

existing political environment but was not merely a proxy for contemporaneous domestic shifts.

2.10 Conclusion

This study examines the dynamics of political violence in the West Bank through the lens of a major diplomatic shock. I leverage the sudden 2017 U.S. recognition of Jerusalem as Israel’s capital as a quasi-experimental event to isolate its impact on local tensions. By employing a staggered difference-in-differences methodology, the analysis reveals that this high-level political signal had profound and immediate consequences on the ground. The findings demonstrate a sharp, significant, and sustained increase in violence in areas proximate to illegal Israeli outposts, which embody the most radical elements of the settler movement. Specifically, violent events in these areas surged by roughly 150%, with attacks on civilians being particularly affected.

The ensuing escalation was critically asymmetric: an immediate wave of violence was perpetrated by Israeli settlers, while a significant rise in Palestinian-led hostilities emerged only after a delay of nearly four years. This causal interpretation is reinforced by the finding that the establishment of outposts *before* the 2017 announcement did not, on its own, lead to a similar surge in violence, confirming that the diplomatic shock effectively “activated” these localities as hotspots of conflict. Crucially, the analysis traces a distinct temporal pattern in the Palestinian response: violence remained sporadic and largely unresponsive to settler attacks for several years before undergoing a fundamental shift following the sharp spike in Israeli-perpetrated violence and military operations in mid-2021. In this subsequent period, attacks became strongly predicted by prior settler aggression, confirming the delayed emergence of a sustained retaliatory dynamic. Taken together, this evidence suggests the diplomatic shock did not merely intensify existing tensions but acted as a catalyst that unleashed previously constrained extremist behavior, initiating a self-perpetuating cycle of aggression.

The results contribute to our understanding of how political violence emerges and escalates, particularly in protracted territorial disputes. By focusing on the erosion of social norms as a key mechanism, this paper offers a “bottom-up” perspective that complements traditional state-centric theories of conflict. It shows how an external political signal can empower non-state actors, validating their radical agenda and lowering the perceived social costs of violence. The initial, one-sided surge in settler violence, fol-

lowed by a delayed but durable retaliatory cycle, illustrates how such shocks can lock communities into an intensifying dynamic of conflict. This highlights the tangible and dangerous consequences of international diplomacy in fragile contexts, where symbolic political acts can destabilize local realities and fuel extremist violence long after the initial event has passed.

A final caveat concerns potential confounders and data limitations that qualify the interpretation of the results. First, violence data of this kind is inherently complex: incidents cluster in both space and time, and even with extensive robustness checks the possibility of unobserved spatial spillovers or temporal correlation cannot be fully eliminated. Second, the empirical strategy estimates a reduced form of a broader causal structure in which the U.S. declaration may have affected violence directly, but also indirectly through its impact on outpost proliferation, while outposts themselves may independently shape local conflict dynamics: disentangling these channels remains an important task for future work. Third, although the identification isolates a sharp diplomatic shock, the conflict environment is shaped by simultaneous domestic and international developments, shifting alliances, and long-standing historical grievances that are difficult to fully observe or code. These limitations do not overturn the main findings, but they underscore the need for caution and highlight the value of future research aimed at resolving these intertwined mechanisms.

Looking ahead, a crucial next step is to explore the potential feedback loop between this externally-triggered violence and domestic political discourse. To this end, I plan to conduct a textual analysis of Knesset speeches, using data provided by [Goldin et al. \(2024\)](#). This analysis will examine whether the U.S. declaration and the subsequent spike in settler violence prompted a corresponding shift in the rhetoric of Israeli politicians. By investigating how domestic political elites responded to the shock—i.e., whether by adopting more radical language that legitimized extremist actions, or by moderating their rhetoric to curb the violence—this future research aims to uncover the nature of the interplay between the diplomatic signal from abroad and the political climate at home. Furthermore, in collaboration with the Israeli NGO Kerem Navot, I will analyze newly available data on territorial interventions by the Israeli government, including the declaration of state land and the allocation of land plots to Israeli settlers. This will allow for an examination of the material relationship between the state, the broader settler population, and the radical actors at the forefront of outpost creation. Such a finding would provide a more complete picture of the dynamics of this cycle of violence, offering

critical insights into the interplay between international politics, local extremism, and domestic policy.

Bibliography

- Acemoglu, D., Santos, R. J., and Robinson, J. A. (2013). The monopoly of violence: Evidence from colombia. *Journal of the European Economic Association*, 11(S1):5–44.
- ACLED (2023). Resurgence of armed groups in the west bank and their connections to gaza. <https://acleddata.com/report/resurgence-armed-groups-west-bank-and-their-connections-gaza>. Accessed: 2025-02-07.
- ACLED (2024a). Acled codebook. Technical report, Armed Conflict Location & Event Data Project (ACLED). Accessed on July 11, 2025.
- ACLED (2024b). Acled dataset. Accessed on 2024-04-17.
- Adnan, W. (2015). Who gets to cross the border? the impact of mobility restrictions on labor flows in the west bank. *Labour Economics*, 34:86–99.
- Albornoz, F., Bradley, J., and Sonderegger, S. (2022). Updating the social norm: the case of hate crime after the brexit referendum. Working Paper 203, Red Nacional de Investigadores en Economía (RedNIE).
- BBC News (2022). Israel’s most right-wing government agreed under Benjamin Netanyahu. Accessed: 2025-12-09.
- Ben-Michael, E., Feller, A., and Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):351–381.
- Berman, N. and Couttenier, M. (2015). External shocks, internal shots: the geography of civil conflicts. *Review of Economics and Statistics*, 97(4):758–776.
- Bertoni, E., Di Maio, M., Molini, V., and Nistico, R. (2019). Education is forbidden: The effect of the boko haram conflict on education in north-east nigeria. *Journal of Development Economics*, 141:102249.
- Besley, T. and Persson, T. (2011). The logic of political violence. *The Quarterly Journal of Economics*, 126(3):1411–1445.
- Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285.
- B’Tselem (2017). Summary of 2016 demolitions. https://www.btselem.org/planning_and_building/20170213_2016_demolitions. Accessed: 2025-02-07.
- B’Tselem (2021). This is ours and this too: Israel’s settlement policy in the west bank.
- B’Tselem (2024). B’tselem statistics. Map data. Accessed on 2024-04-17.

- B'Tselem (2025). Interactive map of the west bank. Online Interactive Map. Accessed: 2025-04-17.
- Bursztyn, L., Egorov, G., and Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11):3522–3548.
- Calì, M. and Miaari, S. H. (2018). The labor market impact of mobility restrictions: Evidence from the west bank. *Labour Economics*, 51:136–151.
- Central Elections Committee for the Knesset (2025). Knesset election results dataset. <https://data.gov.il/he/datasets/central-election-committee/votes-knesset>. Accessed: 2025-02-07.
- Copernicus Sentinel (2025). Sentinel-2 data. Accessed on 2025-04-17. Data range: 2015-2025.
- Cortellessa, E. (2017). Trump signs waiver to keep us embassy in tel aviv. *The Times of Israel*.
- Durante, R. and Zhuravskaya, E. (2018). Attack when the world is not watching? us news and the israeli-palestinian conflict. *Journal of Political Economy*, 126(3):1085–1133.
- European Parliament (2012). Jerusalem: The Heart of the Israeli-Palestinian Conflict. Technical report, Directorate-General for External Policies of the Union. Briefing Note.
- European Union, Copernicus Land Monitoring Service (2025). Copernicus sentinel-2 level-2a data. Accessed via openEO Platform API (e.g., openeo.cloud or specific backend). European Union’s Copernicus Land Monitoring Service information.
- Fourth Geneva Convention (1949). Adopted in Geneva. Official text available from the International Committee of the Red Cross (ICRC) or United Nations Treaty Series (UNTS Vol. 75, No. 973). Commonly known as the Fourth Geneva Convention.
- Giani, M. and Méon, P.-G. (2021). Global racist contagion following donald trump’s election. *British Journal of Political Science*, 51(4):1332–1343. Documents international spillover effects of Trump’s election on racist behavior and hate crimes globally.
- Goldin, G., Howell, N., Ordan, N., Rabinovich, E., and Wintner, S. (2024). The knesset corpus: An annotated corpus of hebrew parliamentary proceedings.
- Havelock, R. (2011). *River Jordan: The Mythology of a Dividing Line*. University of Chicago Press, Chicago.
- Holland, S. and Spetalnick, M. (2017). Trump weighs recognizing jerusalem as israel’s capital: officials. *Reuters*.

- Immenkamp, B. (2020). US 'Peace Plan' for the Middle East. European Parliamentary Research Service, European Parliament, At a Glance PE 646.166.
- International Court of Justice (2024). Legal consequences arising from the policies and practices of israel in the occupied palestinian territory, including east jerusalem. Case 186.
- Israel Defense Forces (2021). Operation guardian of the walls. <https://www.idf.il/en/mini-sites/wars-and-operations/operation-guardian-of-the-walls/operation-guardian-of-the-walls-1/>. Accessed: 2025-02-07.
- Israel Defense Forces (2023). One year since operation break the wave. <https://www.idf.il/en/articles/2023/one-year-since-operation-break-the-wave/>. Accessed: 2025-02-07.
- Israeli Military Government in the West Bank (1967). Military order no. 158 - order concerning abandoned property (private property) (judea and samaria) (no. 58), 5727-1967. Military Order.
- Israeli Supreme Court (2011). Migron dismantling order. Supreme Court Ruling.
- Israeli Supreme Court (2014). Amona evacuation order. Supreme Court Ruling.
- Jaeger, D. A., Klor, E. F., Miaari, S. H., and Paserman, M. D. (2012). The struggle for palestinian hearts and minds: Violence and public opinion in the second intifada. *Journal of Public Economics*, 96(3-4):354–368.
- Kafando, W. A. and Sakurai, T. (2024). Armed conflicts and household food insecurity: Effects and mechanisms. *Agricultural Economics*, 55(2):313–328.
- Kurtzer, D. C. (2016). Sleight of hand: Israel, settlements, and unauthorized outposts. Policy Focus 24, Middle East Institute.
- Levy Commission (2012). Report on the legal status of building in judea and samaria. Government commission report, Government of Israel.
- Likud (2012). Likud – platform. Knesset official website. Archived from the original on 4 February 2012.
- Maxar Technologies (2023). Maxar satellite imagery of the West Bank. <https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9>. Imagery acquired October 21, 2023. Accessed July 11, 2025.
- Müller, K. and Schwarz, C. (2023). From hashtag to hate crime: Twitter and antiminority sentiment. *American Economic Journal: Applied Economics*, 15(3):270–312.
- OCHA (2018). Increase in settler violence during the first four months of 2018. Report,

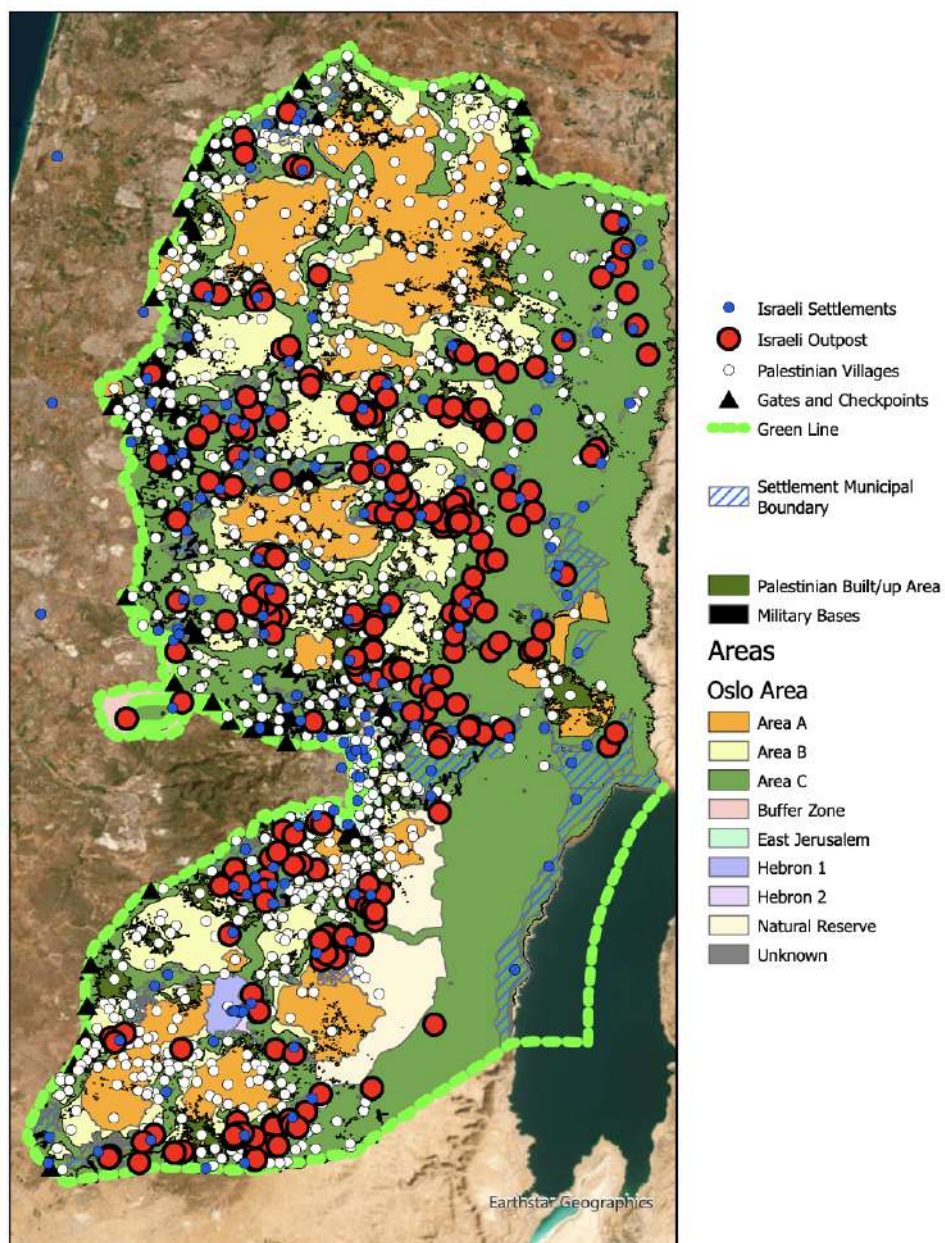
- OCHA oPt. Part of The Monthly Humanitarian Bulletin, May 2018. Accessed: 2025-04-17.
- Palestinian Central Bureau of Statistics (2025). Estimated population in the palestine mid-year by governorate, 1997-2026. Accessed: 2024-07-16.
- Pappe, I. (2017). *The Biggest Prison on Earth: A History of the Occupied Territories*. Oneworld Publications, London.
- Peace Now (2017). Unraveling the mechanism behind illegal outposts - full report. Full report, Peace Now.
- Peace Now (2024). Settlement watch data and reports. Settlement Watch. Accessed on 2024-04-17.
- Peace Now (2025). Settlement watch data and reports. Online Data Portal / Project Website. Provides ongoing data and reports on Israeli settlements and outposts. Accessed: 2025-04-17.
- Rohner, D., Thoenig, M., and Zilibotti, F. (2013). Seeds of distrust: Conflict in uganda. *Journal of Economic Growth*, 18(3):217–252.
- The Times of Israel (2025). Far-right MKs sport noose-shaped pins at hearing on controversial death penalty bill. Accessed: 2025-12-09.
- United Nations General Assembly (2017). General Assembly Overwhelmingly Adopts Resolution on Status of Jerusalem. Press Release GA/11995.
- United Nations Office for the Coordination of Humanitarian Affairs (2021a). Escalation of hostilities and unrest in the opt: Flash appeal. <https://www.unocha.org/publications/report/occupied-palestinian-territory/escalation-hostilities-and-unrest-opt-flash-appeal-27-may-2021>. Accessed: 2025-02-07.
- United Nations Office for the Coordination of Humanitarian Affairs (2021b). Protection of civilians report: 24–31 may 2021. <https://www.ochaopt.org/poc/24-31-may-2021>. Accessed: 2025-02-07.
- United Nations Security Council (1979). Resolution 446 (1979). UN Document S/RES/446.
- United Nations Security Council (1980). Resolution 478 (1980) / [adopted by the security council at its 2245th meeting, on 20 august 1980]. Available at: <https://digitallibrary.un.org/record/25618?v=pdf>.
- United Nations Security Council (2016). Resolution 2334 (2016). UN Document S/RES/2334.

- U.S. Congress (1995). Jerusalem embassy act of 1995. <https://www.govinfo.gov/content/pkg/PLAW-104publ45/pdf/PLAW-104publ45.pdf>. Public Law 104-45, 104th Congress. Accessed: 2024-10-26.
- Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics*, 129(4):1947–1994.

A Appendix

A.1 Figures

Figure A.1: Map of the West Bank



Notes: Map of the West Bank. Point geometries include Palestinian towns and villages (white dots), Israeli settlements (blue dots), Israeli outposts (red dots). Polygon geometries include Oslo Areas partitions, settlement municipality areas, Palestinian built-up areas and military bases. The map has been compiled by combining sources from ([B'Tselem, 2025](#)) and ([Peace Now, 2024](#))

Figure A.2: Outposts, Settlements and Palestinian Communities in the West Bank



(a) Ahiya Outpost, 1997



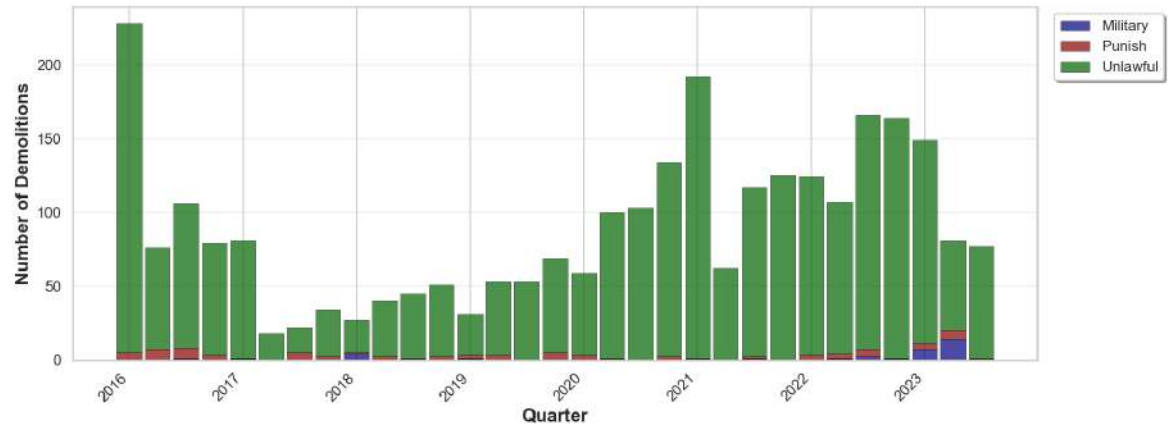
(b) Nahal Tirza Farm Outpost, 2020



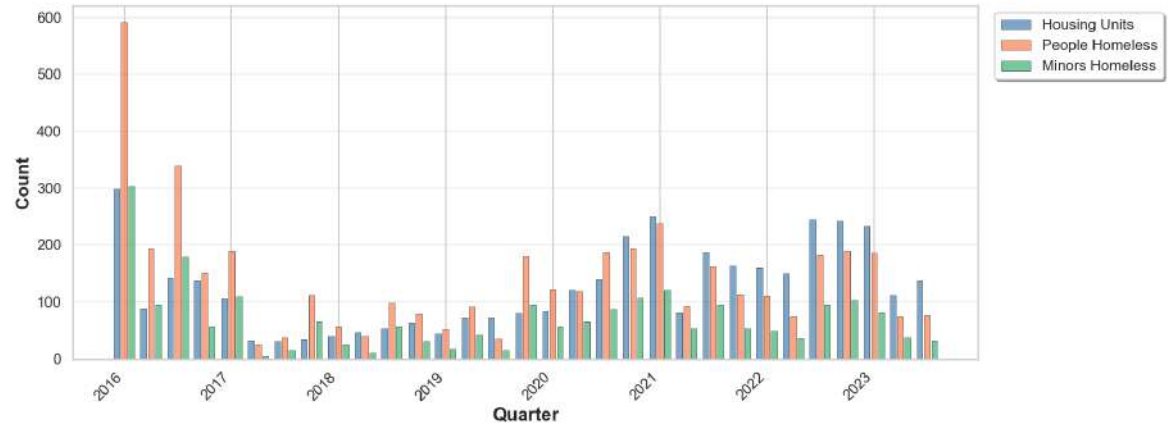
(c) Outposts (red), Settlement (blue) and Palestinian Towns (white)

Notes: Satellite images of different outposts in the West Bank. Resolution: 0.5m. Source: [Maxar Technologies \(2023\)](#).

Figure A.3: Demolitions in the West Bank



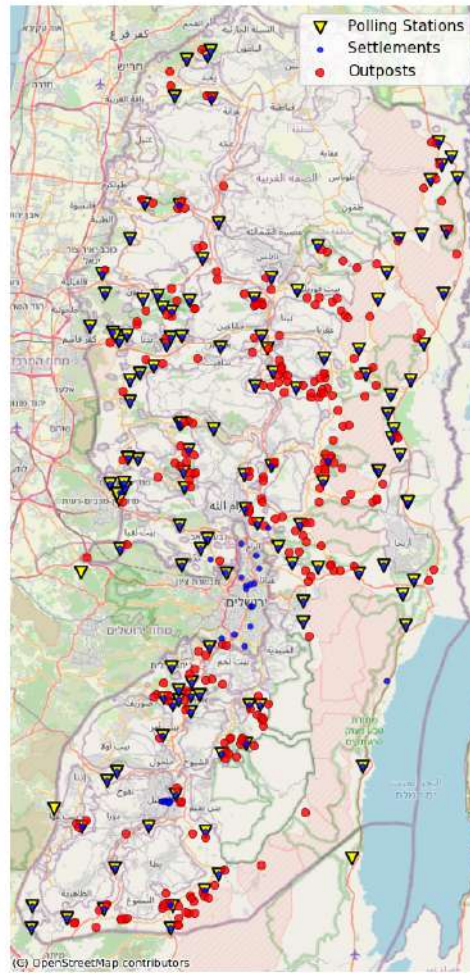
(a) Demolitions by Type



(b) Demolitions by Variable Breakdown

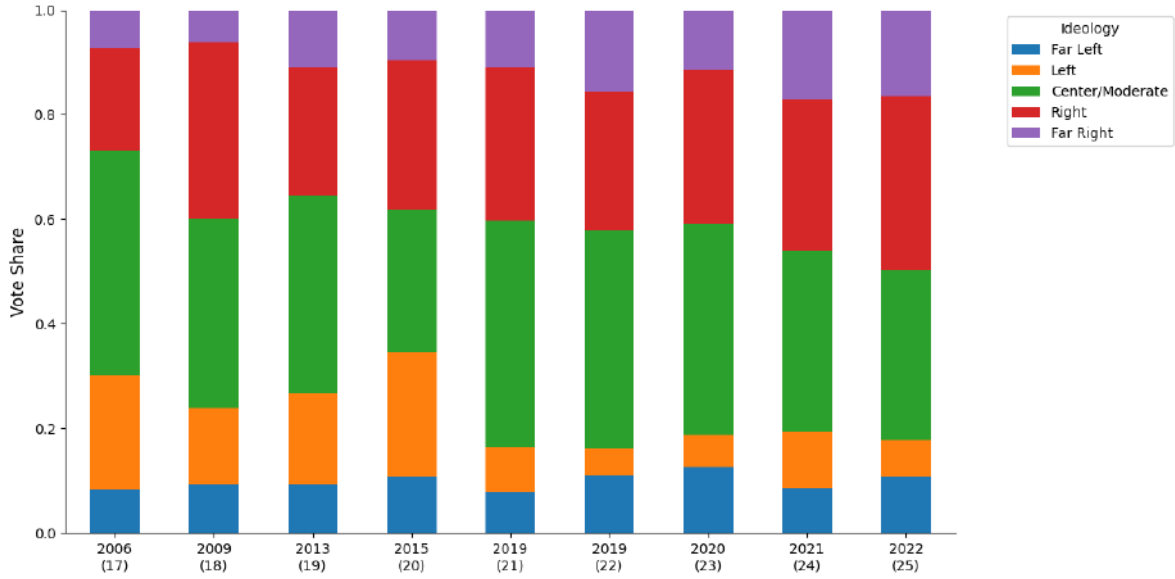
Notes: Quarterly demolition activity in the West Bank, 2016–2023. Panel A.3a reports the number of demolition events by demolition type (unlawful-construction, punitive, military). A.3b reports the quarterly counts of housing units demolished and resulting homelessness (people and minors).

Figure A.4: Polling Stations, Settlements and Outposts in the West Bank

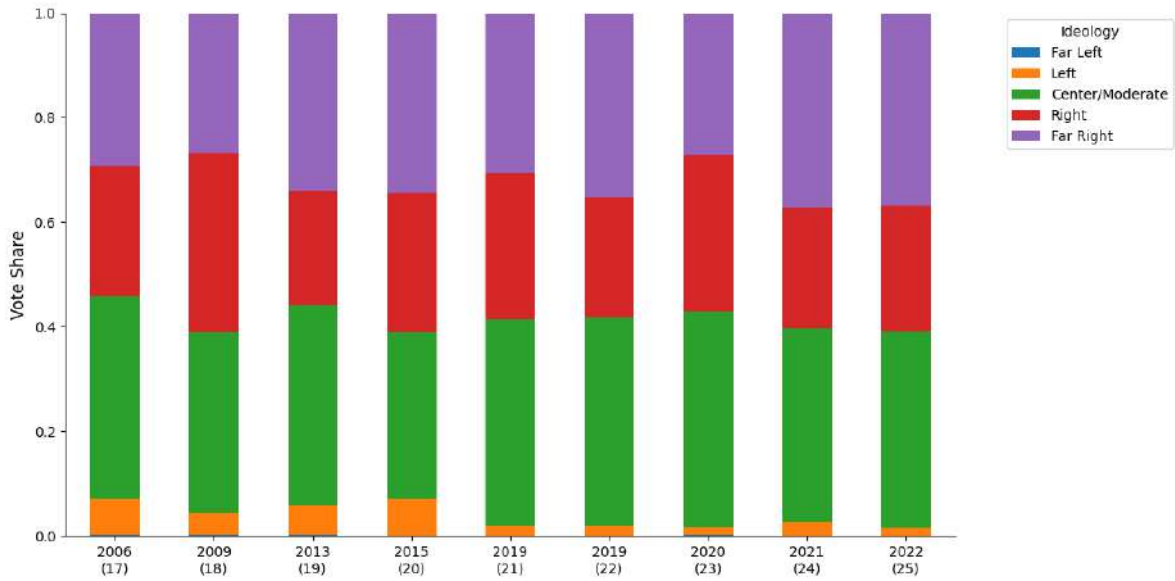


Notes: Map of the West Bank with geocoded polling stations (yellow triangles), Israeli settlements (blue dots) and Israeli outposts (red dots). The map has been compiled by combining sources from (B'Tselem, 2025), (Peace Now, 2024) and (Central Elections Committee for the Knesset, 2025).

Figure A.5: Knesset Vote Share by Ideology



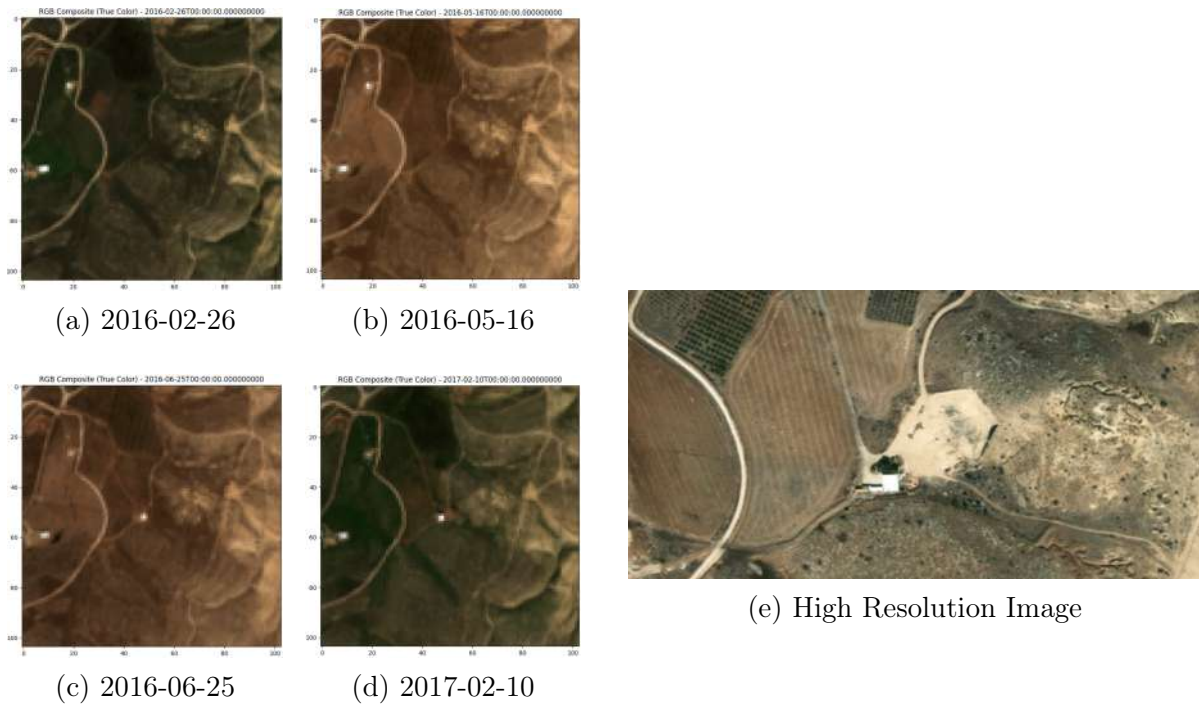
(a) Knesset Vote Share by Ideology in Israel



(b) Knesset Vote Share by Ideology in the West Bank

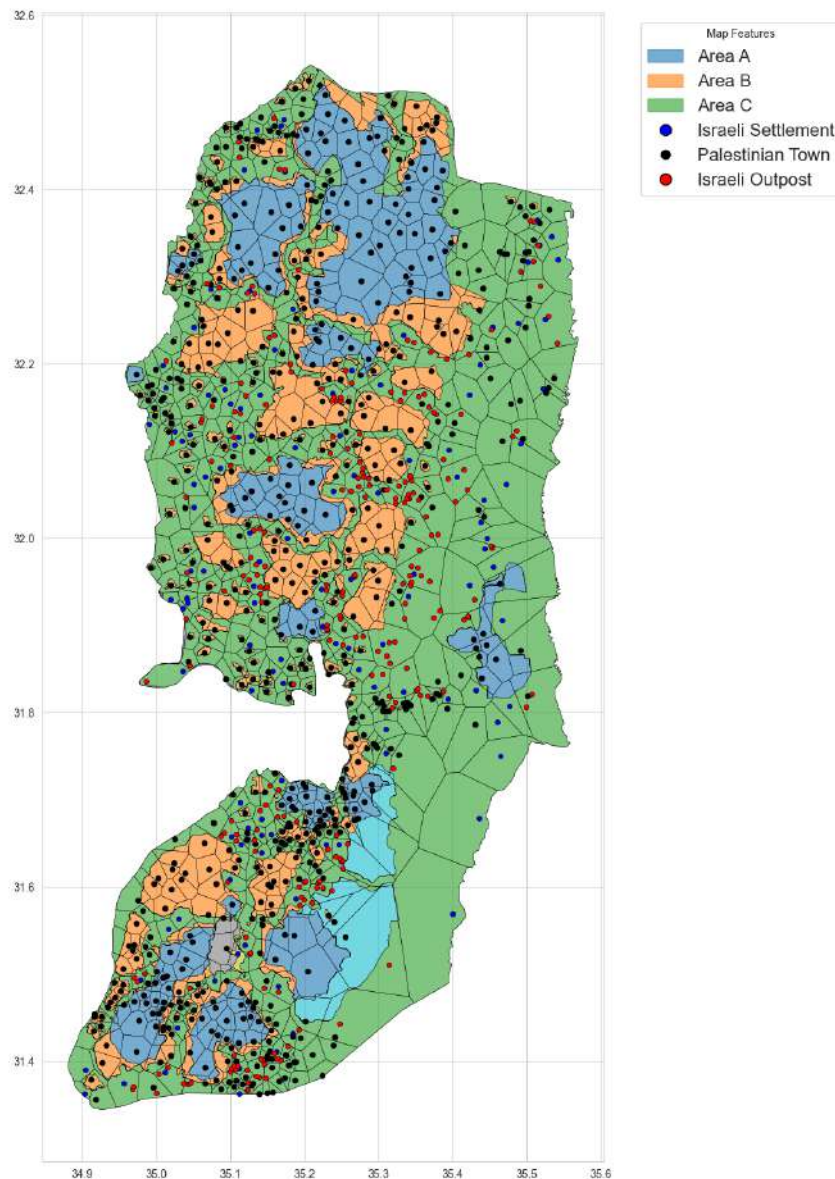
Notes: Vote share by ideology in Israeli national elections from 2005 to 2022. Panel A.5a shows the national-level results, while panel A.5b shows the results for polling stations located in the West Bank. The ideological classification follows Table A.2.

Figure A.6: Outpost Timing Example - Bentayim Farm Outpost



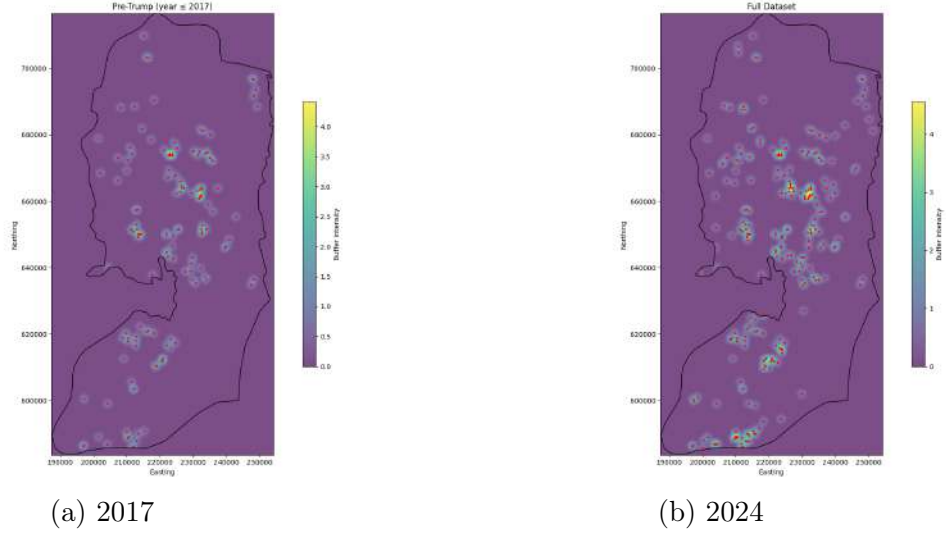
Notes: Panels A.6a to A.6d show the evolution of the Bentayim Farm outpost from February 2016 to February 2017, with the first image showing no structures and the last one showing the presence of a farm. The high resolution image in panel A.6e shows the same outpost in 2023. The images are sourced from Copernicus Sentinel-2 Level 2A data and Maxar Technologies.

Figure A.7: Voronoi Polygons



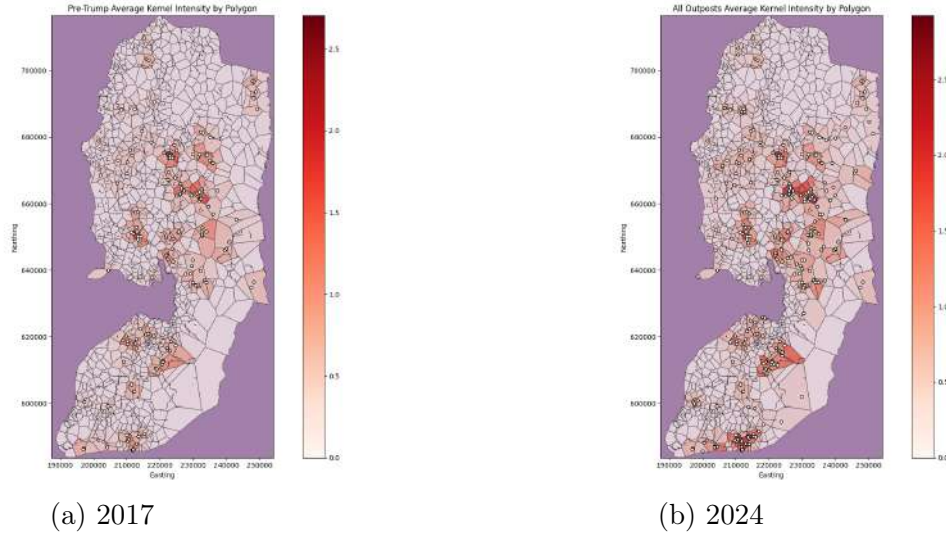
Notes: Voronoi polygons of West Bank settlements and Palestinian towns. Black dots represent Palestinian towns and villages, blue dots represent Israeli settlements and red dots represent Israeli outposts. The Voronoi tessellation is computed using these point geometries. Polygon colors indicate the Oslo Areas classification, where the light blue area is part of Area C, and grey area is Hebron. Dark blue area is area A, orange area is Area B and green area is Area C. East Jerusalem is excluded.

Figure A.8: Outpost Buffer Areas



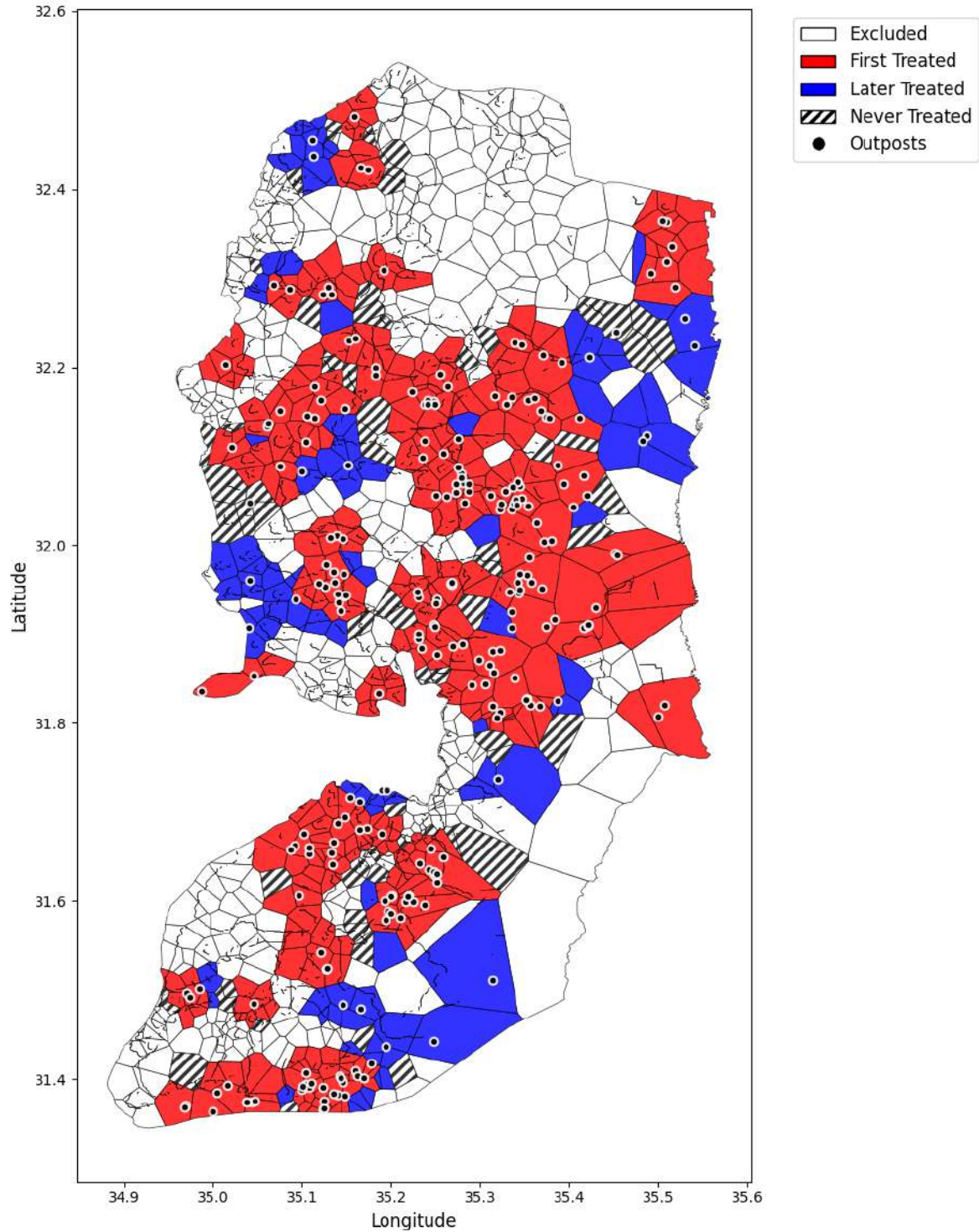
Notes: Raster data of the outposts buffer zones. The intensity value of each pixel is calculated using a spatial kernel with linear decay, according to the formula: $K(x) = 1 - \frac{d}{d_{max}}$ if $d < d_{max}$, where d is the distance from the outpost and d_{max} is the maximum distance of the kernel, i.e. 2 km. Panel (a) shows the buffer areas before Trump's declaration, while panel (b) shows the buffer areas as of 2024.

Figure A.9: Outpost Buffer Areas



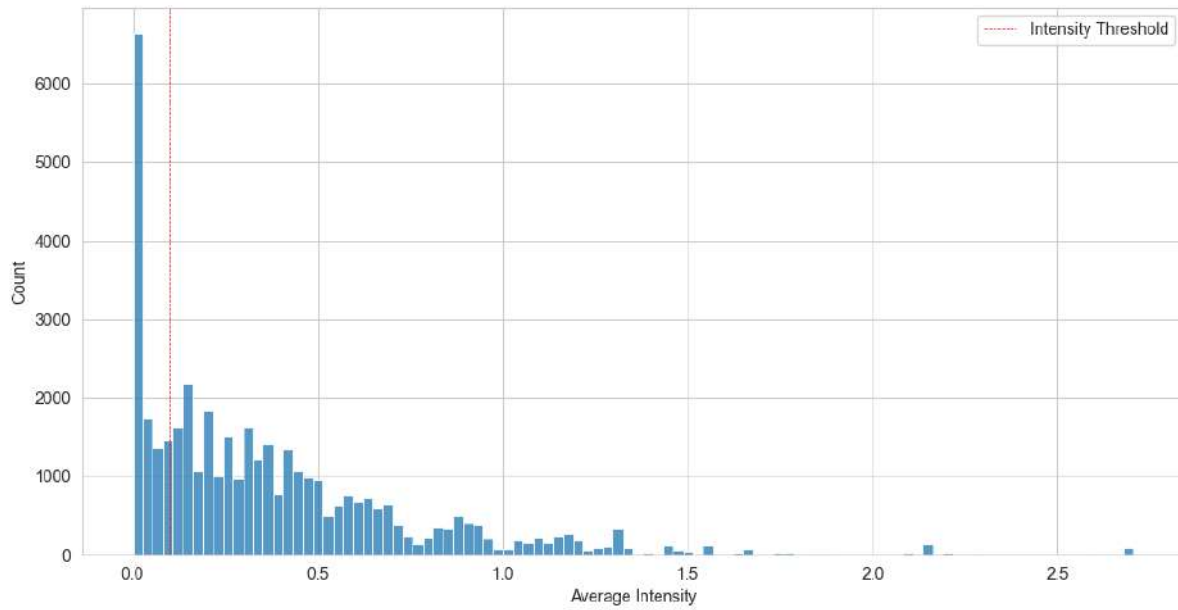
Notes: Average outpost intensity of Voronoi polygons in the West Bank. Panel (a) shows the intensity distribution in 2017, while panel (b) shows the distribution in 2024.

Figure A.10: Areas with Outpost Activity



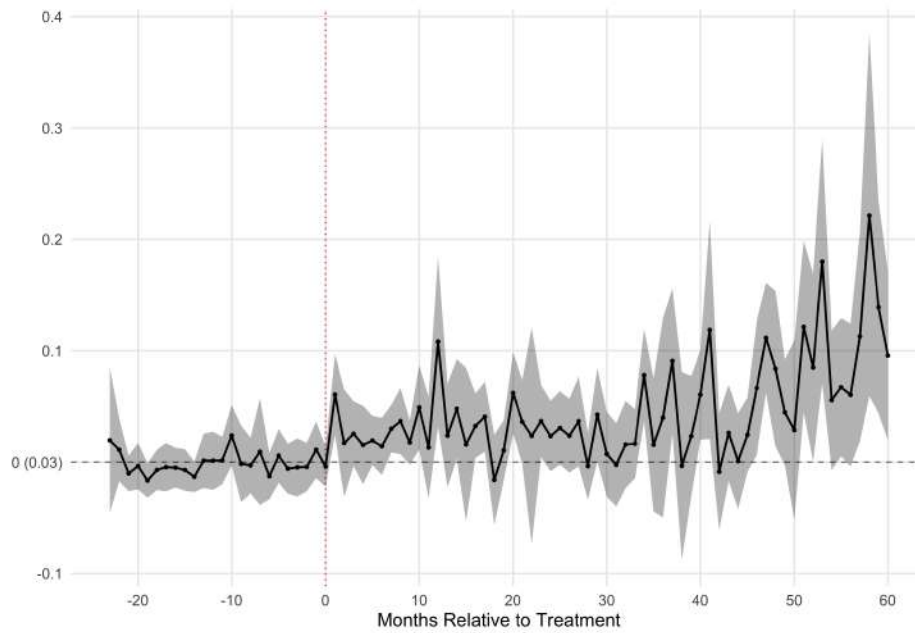
Notes: Areas with outpost presence in the West Bank. Red colored areas indicate polygons that developed an outpost before December 2017, blue colored areas indicate polygons that developed an outpost after December 2017. White polygons are those that never developed an outpost, while polygons with the hatched pattern are those that developed an outpost after December 2022.

Figure A.11: Histogram of Outpost Intensity



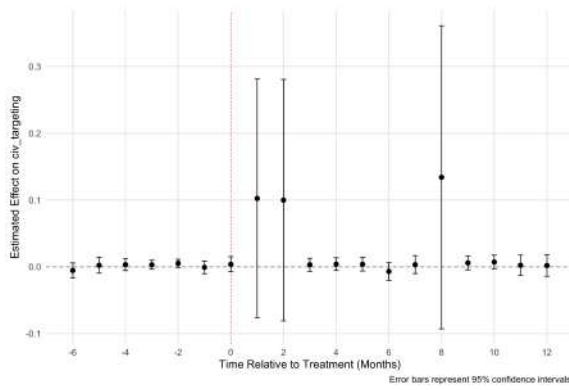
Notes: Histogram of the average outpost intensity in the Voronoi polygons. The red dashed line indicates the threshold of 0.1 used to define the treatment status.

Figure A.12: Event Study: Civilian Targeting

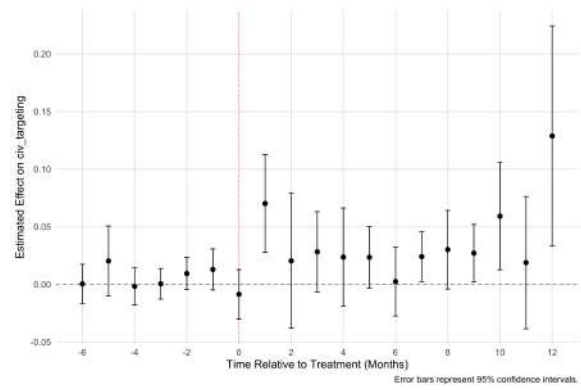


Notes: This figure presents the event study for results in which the target of violence is a civilian from Table 2.1. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

Figure A.13: This is the main caption for the entire figure with two subfigures.



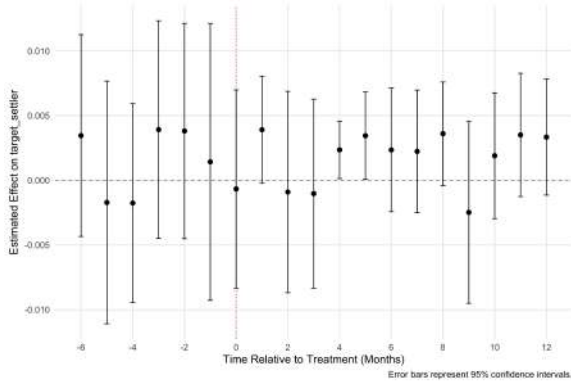
(a) Civilian Targeting:
before December 2017



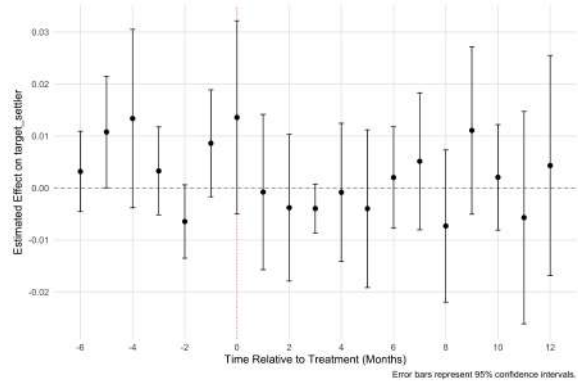
(b) Civilian Targeting:
after December 2017

Notes: This figure presents the event studies for events in which civilians were targeted. Panel (a) shows the event studies considering only outposts established before December 2017, while Panel (b) shows the event studies displayed from the main analysis considering for a shorter time frame, i.e. six lags and twelve leads. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The estimation equation includes the linear time trend.

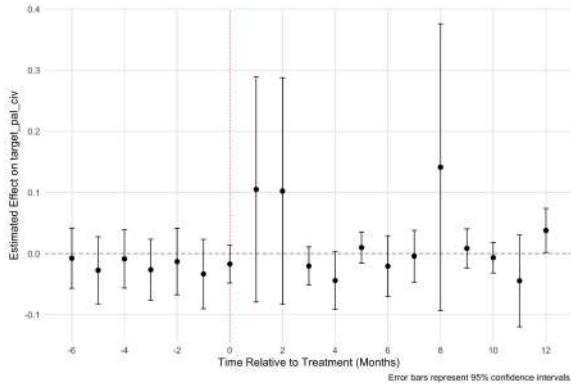
Figure A.14: Target Event Studies: Declaration as a Structural Break



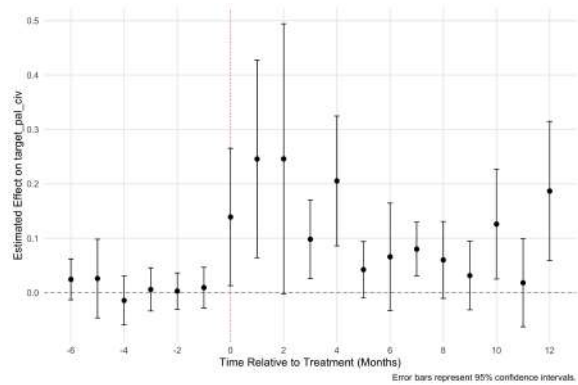
(a) Settlers as Targets:
before December 2017



(b) Settlers as Targets:
after December 2017



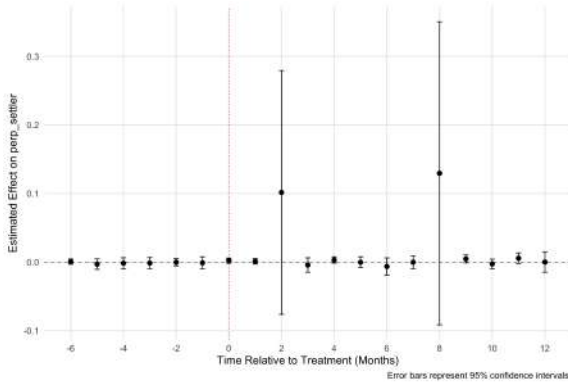
(c) Palestinian Civilians as Targets:
before December 2017



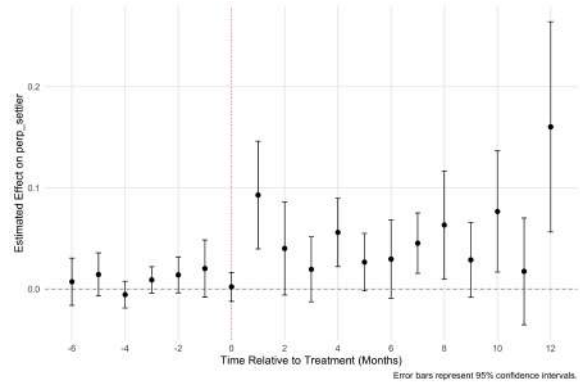
(d) Palestinian Civilians as Targets:
after December 2017

Notes: This figure presents the event studies for two outcomes: the total number of violent events where settlers were identified as targets (top row) and the number of violent events where palestinians were identified as targets (bottom row). Panels (a) and (c) show the event studies considering only outposts established before December 2017, while Panels (b) and (d) show the event studies displayed from the main analysis considering for a shorter time frame, i.e. six lags and twelve leads. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The estimation equation includes the linear time trend.

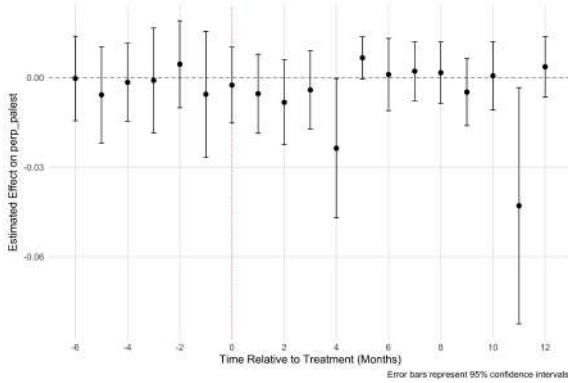
Figure A.15: Perpetrator Event Studies: Declaration as a Structural Break



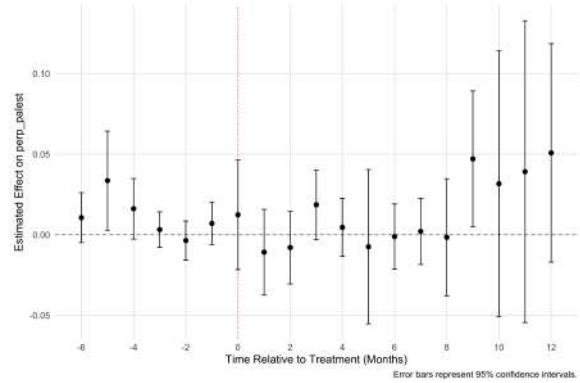
(a) Settlers as Perpetrators:
before December 2017



(b) Settlers as Perpetrators:
after December 2017



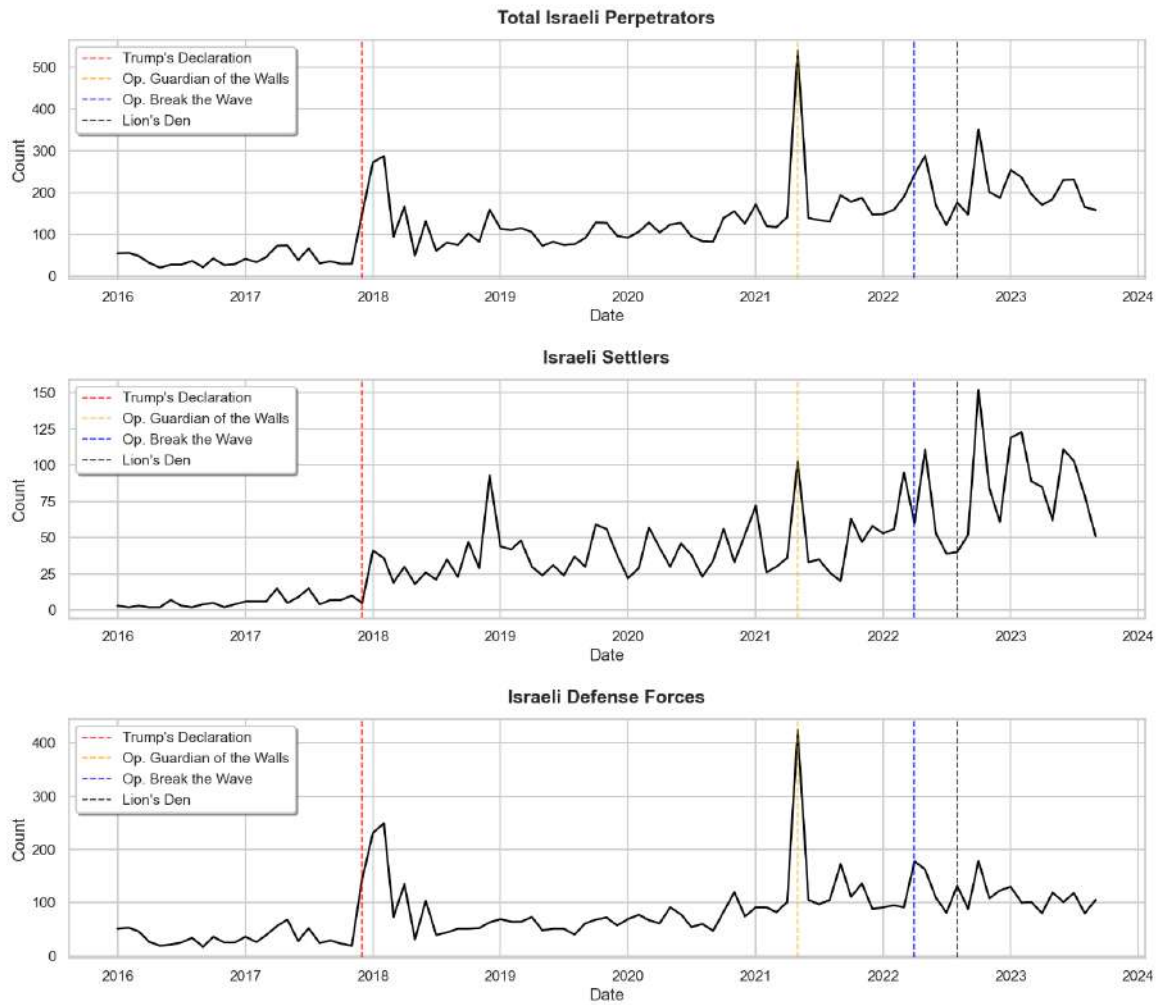
(c) Palestinian Civilians as Perpetrators:
before December 2017



(d) Palestinian Civilians as Perpetrators:
after December 2017

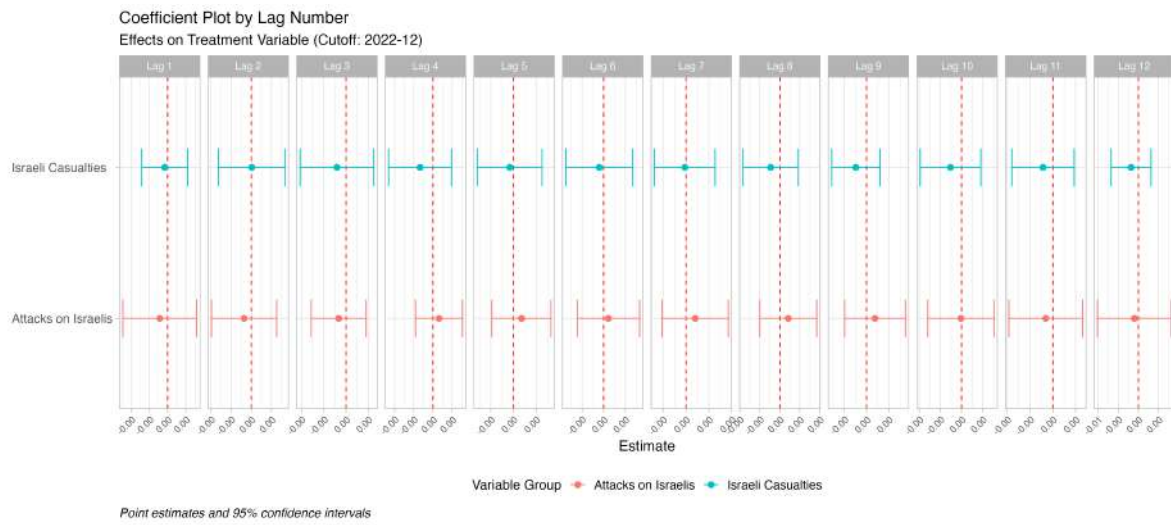
Notes: This figure presents the event studies for two outcomes: the total number of violent events where settlers were identified as perpetrators (top row) and the number of violent events where palestinians were identified as perpetrators (bottom row). Panels (a) and (c) show the event studies considering only outposts established before December 2017, while Panels (b) and (d) show the event studies displayed from the main analysis considering for a shorter time frame, i.e. six lags and twelve leads. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The estimation equation includes the linear time trend.

Figure A.16: Monthly Israeli Attacks over Time



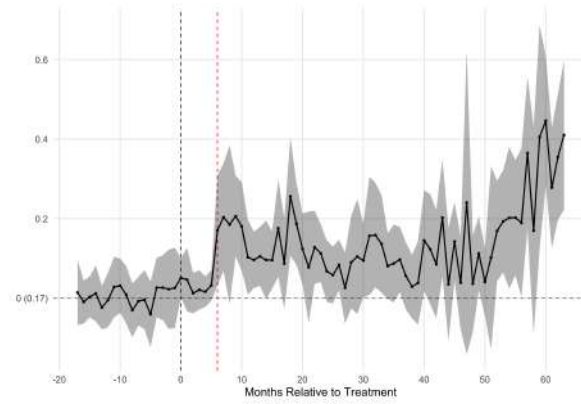
Notes: This figure presents the monthly number of Israeli perpetrated attacks broke down by Palestinian actor type, i.e. civilians and militias. Vertical dashed lines indicate important dates: the Trump's declaration (December 2017), IDF operations in the West Bank and Gaza strip, the emergence of the paramilitary group Lions' Den.

Figure A.17: Robustness Check: Treatment Status and Violent Events

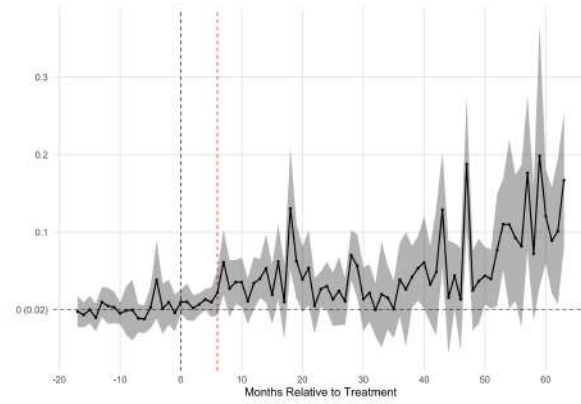


Notes: OLS Regression results of equation 2.5. The dependent variable is the binary treatment status of the Voronoi polygons. The coefficients are estimated using a linear probability model with 12 lags of violent events, casualties and settlers targeted in violent events.

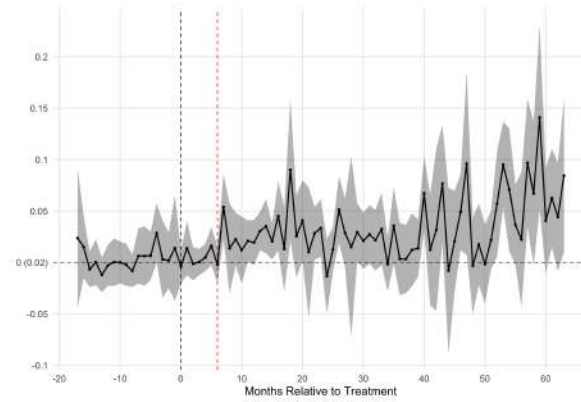
Figure A.18: Placebo Test: Main Results



(a) Number of Events



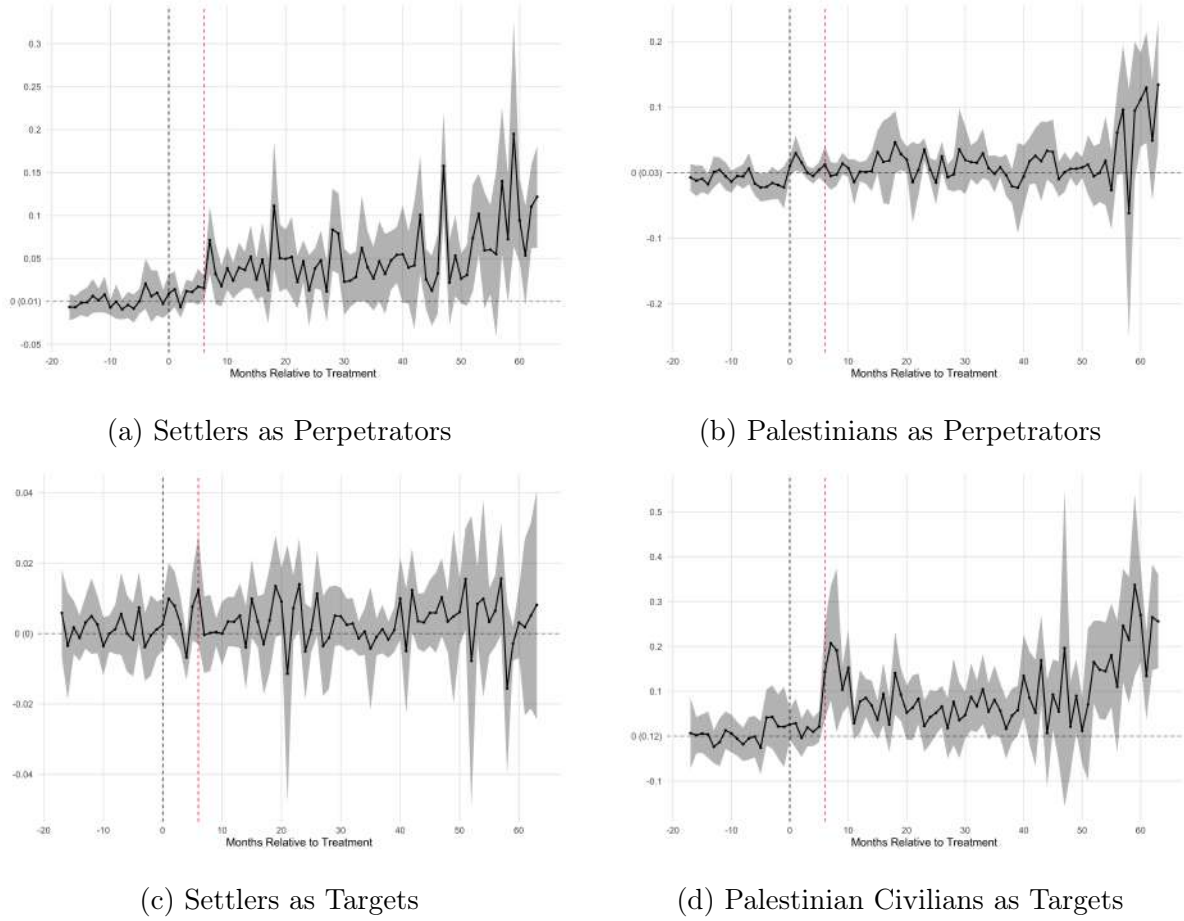
(b) Settler Events



(c) Civilian Targeting

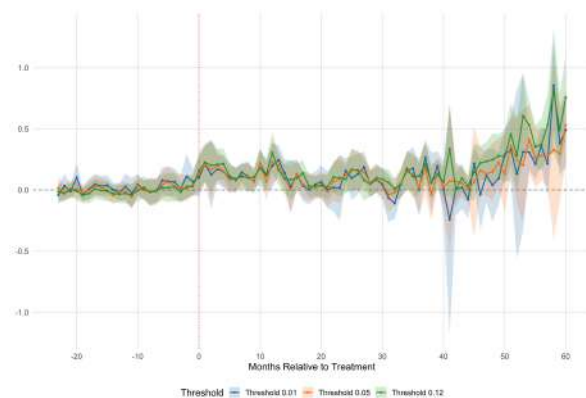
Notes: This figure presents the placebo test results where the treatment date is artificially shifted six months earlier. Panel (a) shows the total number of violent events, Panel (b) shows events involving settlers, and Panel (c) shows events targeting civilians. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The black dashed line indicates the placebo treatment date, the red one the actual treatment date.

Figure A.19: Placebo Test: Perpetrators and Targets

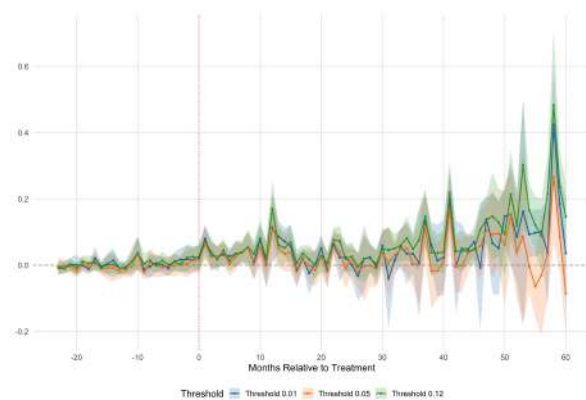


Notes: This figure presents the placebo test results for different perpetrators and targets where the treatment date is artificially shifted six months earlier. Panel (a) shows events where settlers are perpetrators, Panel (b) shows events where Palestinians are perpetrators, Panel (c) shows events where settlers are targets, and Panel (d) shows events where Palestinian civilians are targets. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The black dashed line indicates the placebo treatment date, the red one the actual treatment date.

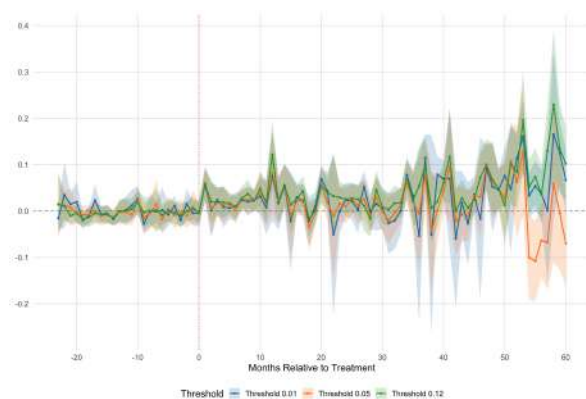
Figure A.20: Alternative Thresholds: Main Results



(a) Number of Events



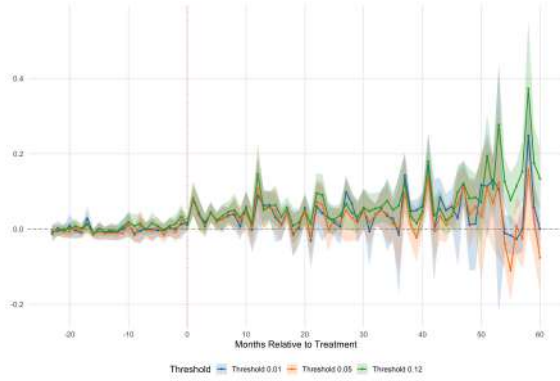
(b) Settler Events



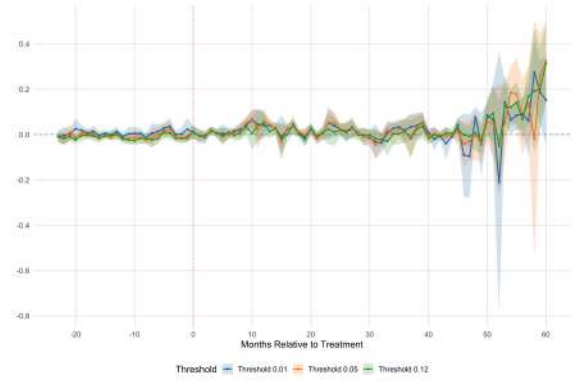
(c) Civilian Targeting

Notes: This figure presents the results the alternative thresholds are used to binarize the outpost intensity. Panel (a) shows the total number of violent events, Panel (b) shows events involving settlers, and Panel (c) shows events targeting civilians. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The blue line displays the results for a threshold of 0.01, the orange line for a threshold of 0.05 and the green line for a threshold of 0.12.

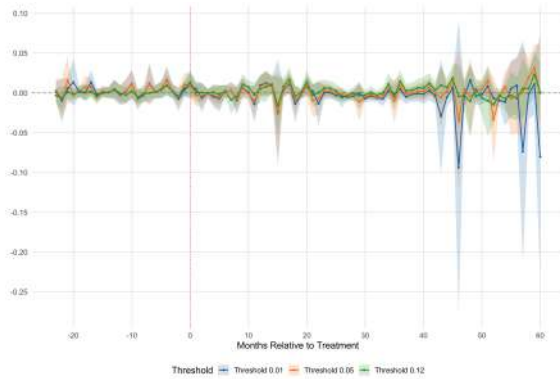
Figure A.21: Alternative Thresholds: Perpetrators and Targets



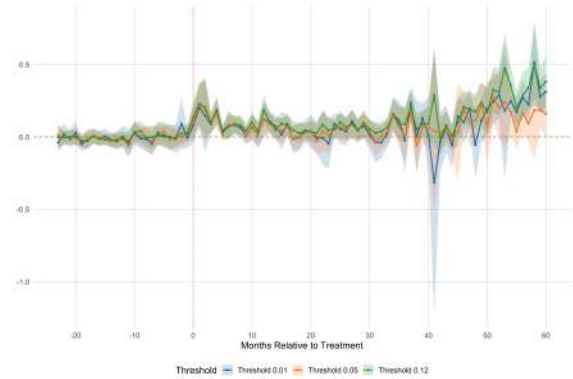
(a) Settlers as Perpetrators



(b) Palestinians as Perpetrators



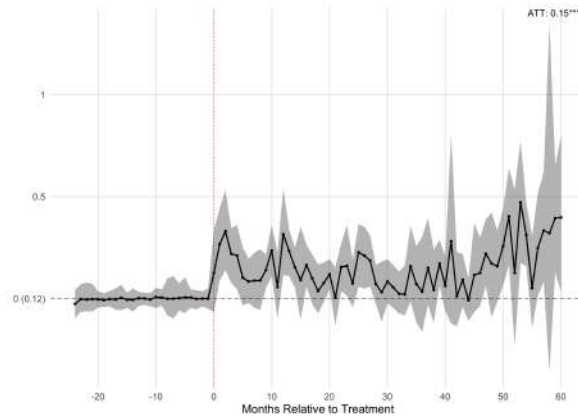
(c) Settlers as Targets



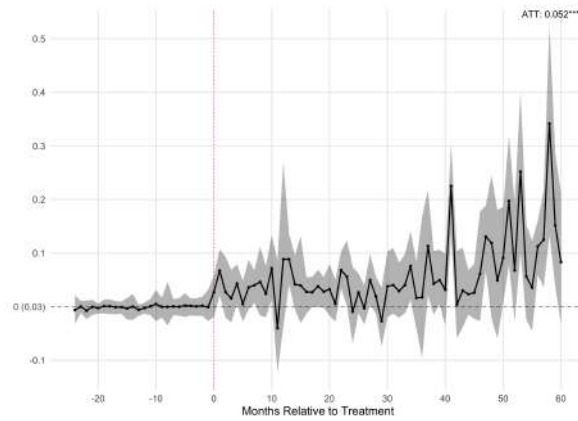
(d) Palestinian Civilians as Targets

Notes: This figure presents the placebo test results for different perpetrators and targets where the treatment date is artificially shifted six months earlier. Panel (a) shows events where settlers are perpetrators, Panel (b) shows events where Palestinians are perpetrators, Panel (c) shows events where settlers are targets, and Panel (d) shows events where Palestinian civilians are targets. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The blue line displays the results for a threshold of 0.01, the orange line for a threshold of 0.05 and the green line for a threshold of 0.12.

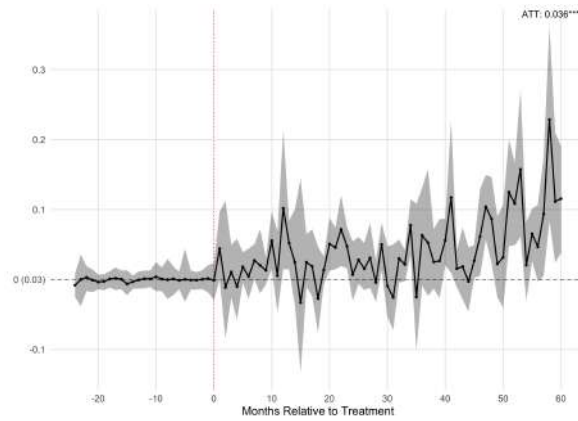
Figure A.22: Synthetic Control: Main Results



(a) Number of Events



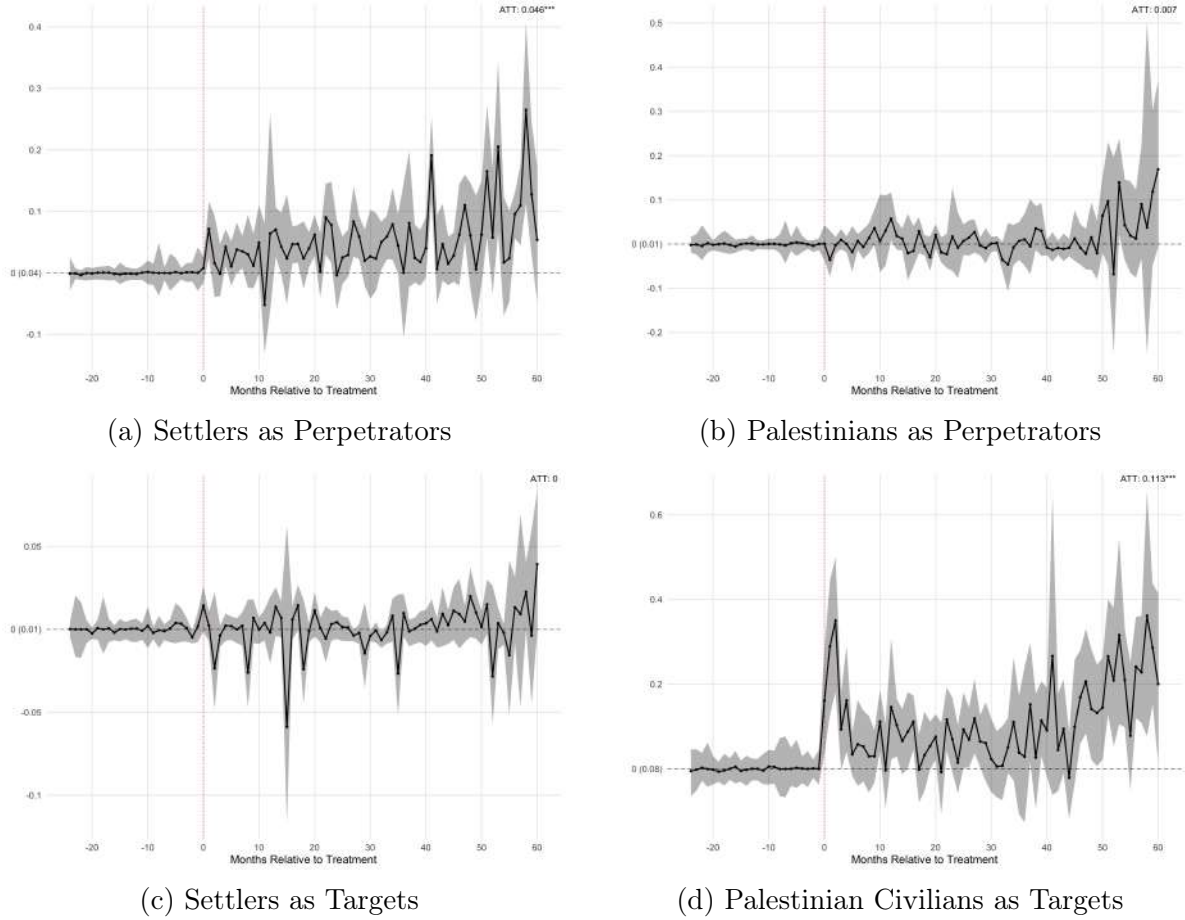
(b) Settler Events



(c) Civilian Targeting

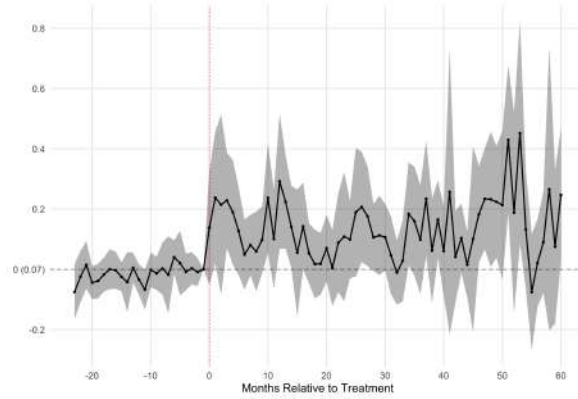
Notes: This figure presents the synthetic control results. Panel (a) shows the total number of violent events, Panel (b) shows events involving settlers, and Panel (c) shows events targeting civilians. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The ATT is displayed on top right of each panel: ** $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.23: Synthetic Control: Perpetrators and Targets

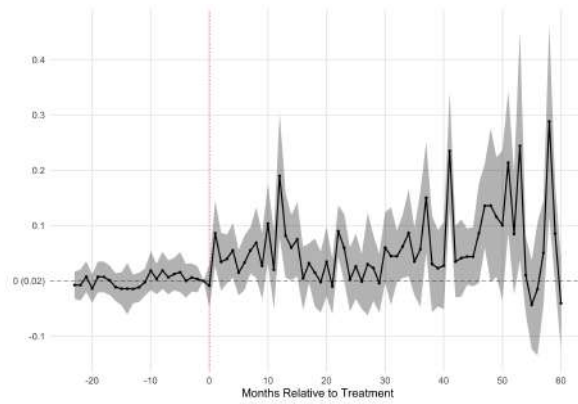


Notes: This figure presents the synthetic control results. Panel (a) shows events where settlers are perpetrators, Panel (b) shows events where Palestinians are perpetrators, Panel (c) shows events where settlers are targets, and Panel (d) shows events where Palestinian civilians are targets. 95% confidence intervals clustered at the Voronoi polygon level are displayed. The ATT is displayed on top right of each panel: ** $p < 0.01$, * $p < 0.05$, *** $p < 0.01$.

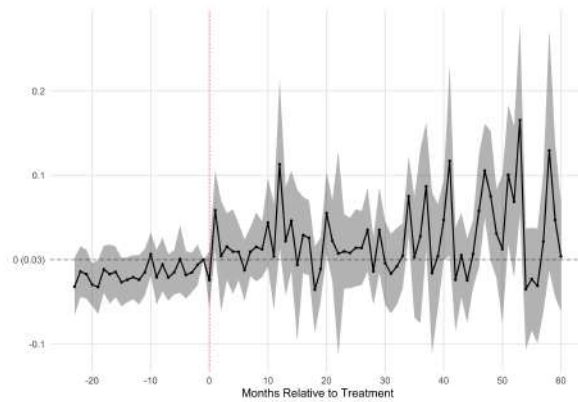
Figure A.24: Non-Staggered Specification: Main Results



(a) Number of Events



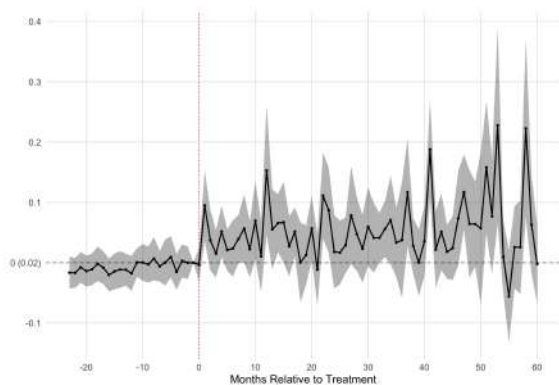
(b) Settler Events



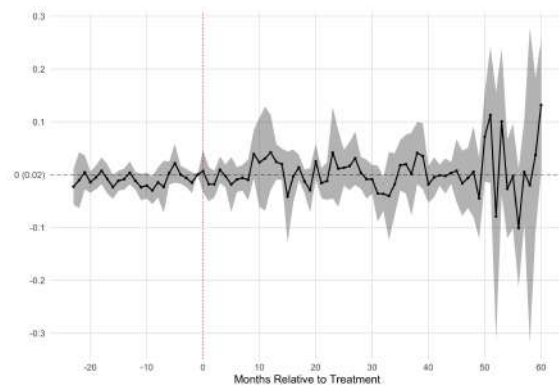
(c) Civilian Targeting

Notes: This figure presents the results considering not yet treated units as controls and drop them when they switch the treatment status. Panel (a) shows the total number of violent events, Panel (b) shows events involving settlers, and Panel (c) shows events targeting civilians. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

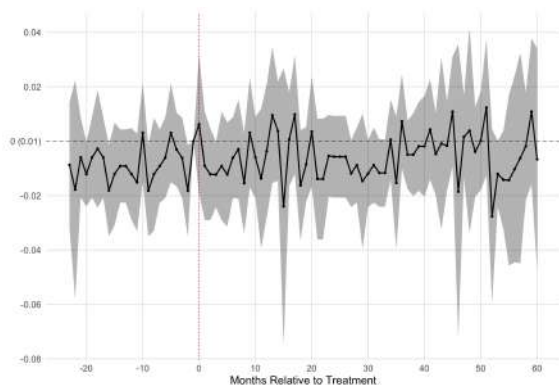
Figure A.25: Non-Staggered Specification: Perpetrators and Targets



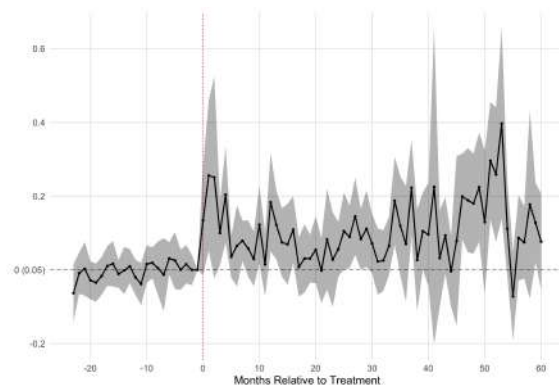
(a) Settlers as Perpetrators



(b) Palestinians as Perpetrators



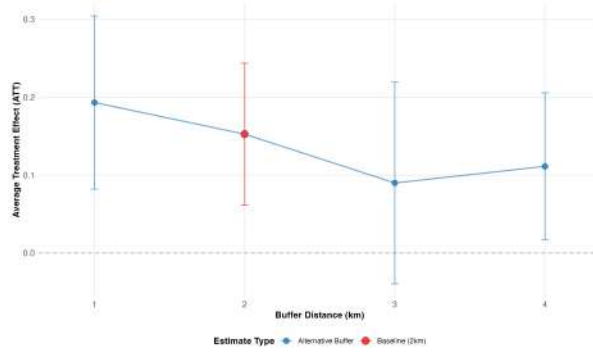
(c) Settlers as Targets



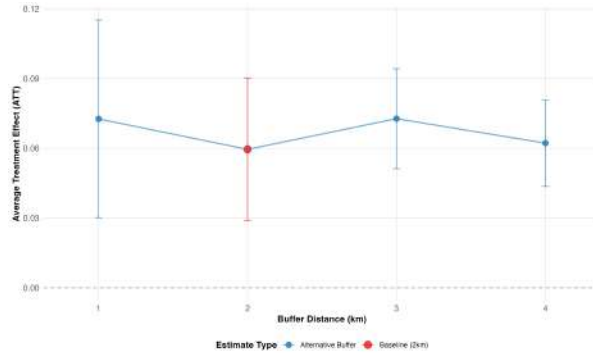
(d) Palestinian Civilians as Targets

Notes: This figure presents the results considering not yet treated units as controls and drop them when they switch the treatment status. Panel (a) shows events where settlers are perpetrators, Panel (b) shows events where Palestinians are perpetrators, Panel (c) shows events where settlers are targets, and Panel (d) shows events where Palestinian civilians are targets. 95% confidence intervals clustered at the Voronoi polygon level are displayed.

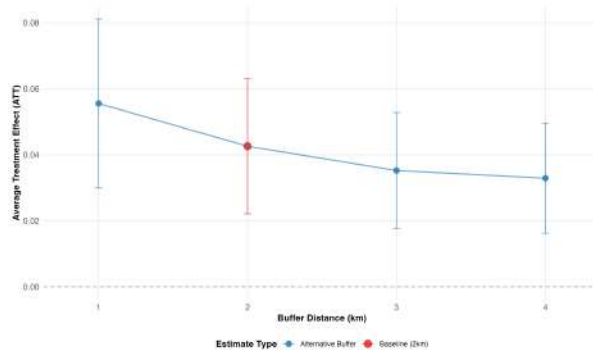
Figure A.26: Alternative Buffers: Main Results



(a) Number of Events



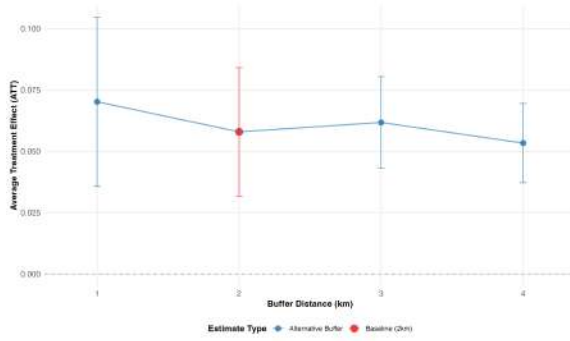
(b) Settler Events



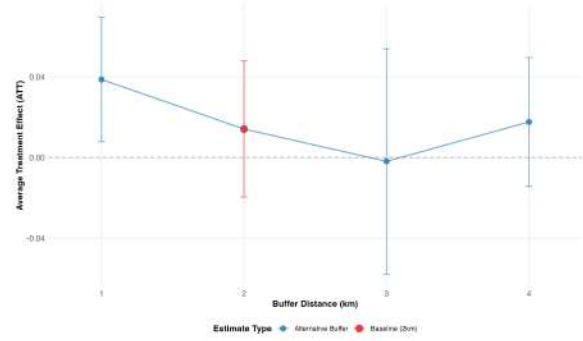
(c) Civilian Targeting

Notes: This figure presents the results of the robustness check where I consider different buffer areas of 1, 3 and 4 km for the main outcomes of Table 2.1. Panel (a) shows the total number of violent events, Panel (b) shows events involving settlers, and Panel (c) shows events targeting civilians. Confidence bands indicate 95% confidence intervals, clustered at the Voronoi polygon level.

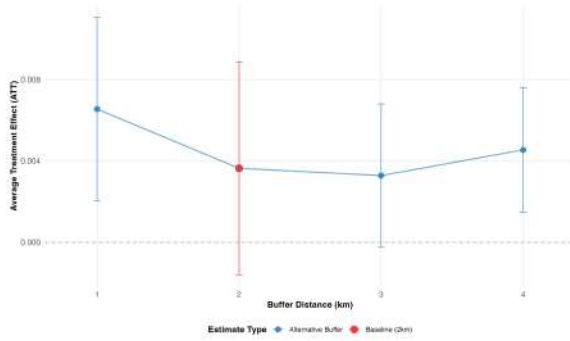
Figure A.27: Alternative Buffers: Perpetrators and Targets



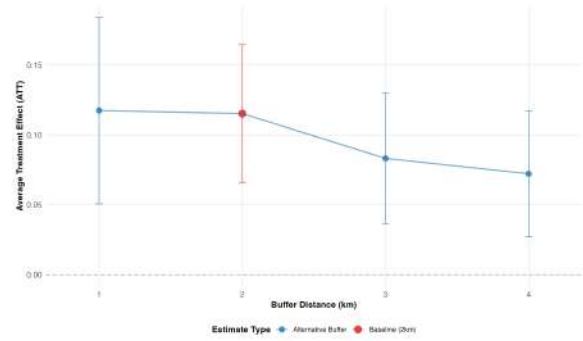
(a) Settlers as Perpetrators



(b) Palestinians as Perpetrators



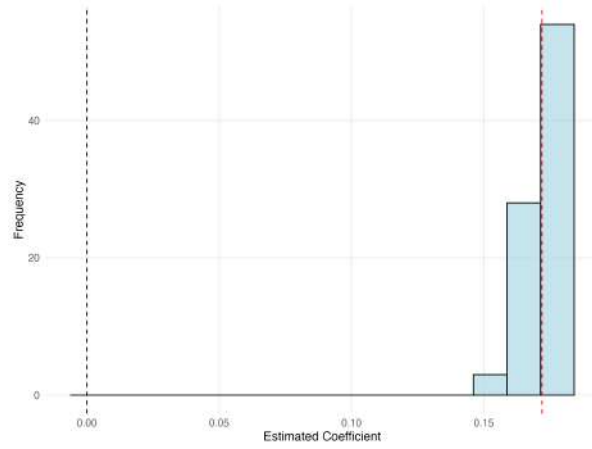
(c) Settlers as Targets



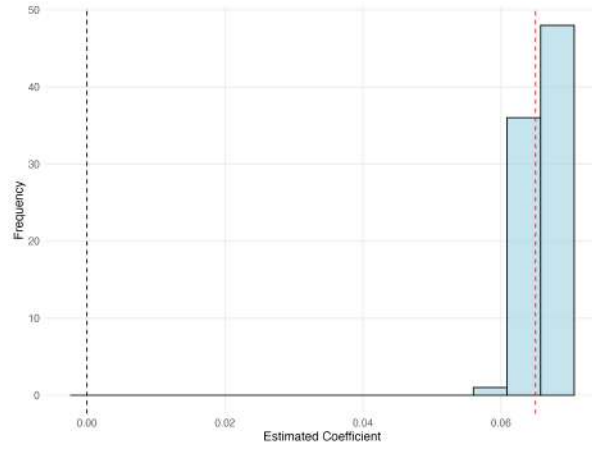
(d) Palestinian Civilians as Targets

Notes: This figure presents the results of the robustness check where I consider different buffer areas of 1, 3 and 4 km for the main outcomes of Table 2.2. Confidence bands indicate 95% confidence intervals, clustered at the Voronoi polygon level. Panel (a) shows events where settlers are perpetrators, Panel (b) shows events where Palestinians are perpetrators, Panel (c) shows events where settlers are targets, and Panel (d) shows events where Palestinian civilians are targets.

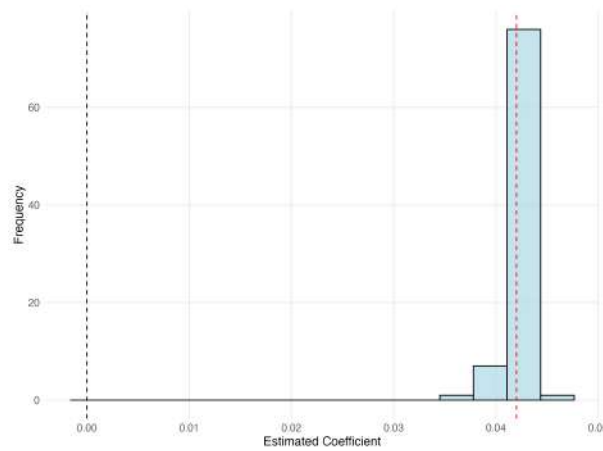
Figure A.28: Leave-out Regressions: Main Results



(a) Number of Events



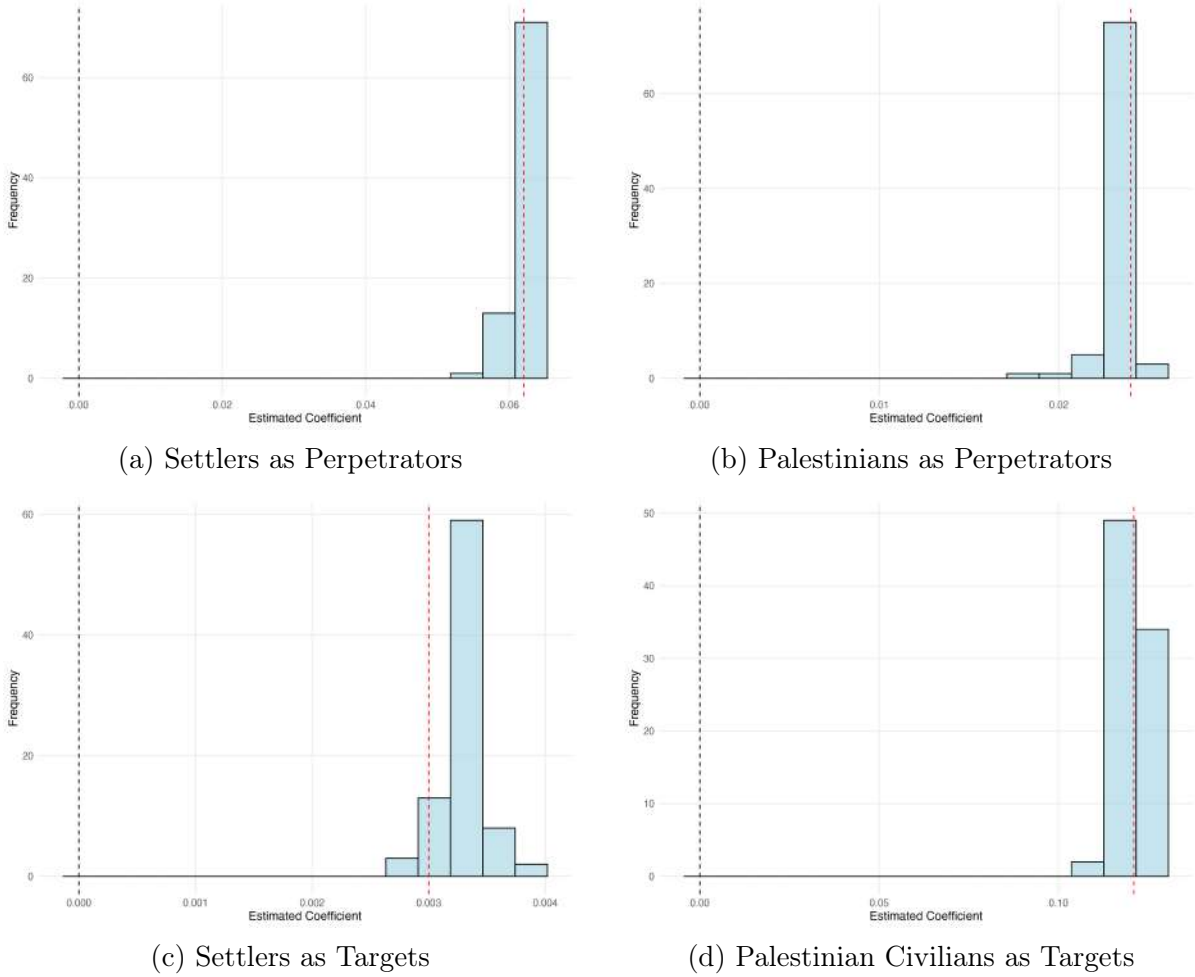
(b) Settler Events



(c) Civilian Targeting

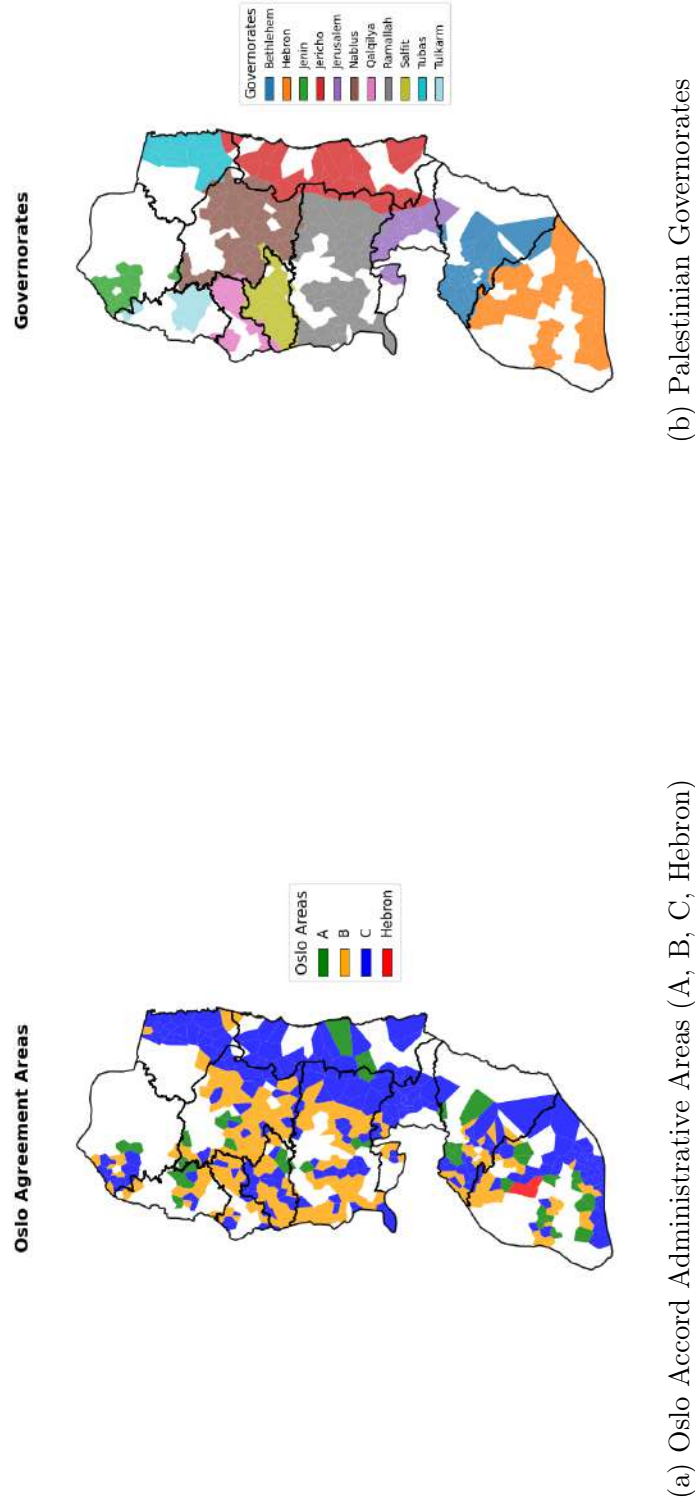
Notes: This figure presents the results of the leave-out regression robustness. Panel (a) shows the total number of violent events, Panel (b) shows events involving settlers, and Panel (c) shows events targeting civilians. The results display the distribution of the β coefficients for 85 iterations of sampling without replacement. The red dashed line indicates the baseline estimate from Table 2.1.

Figure A.29: Leave-out Regressions: Perpetrators and Targets



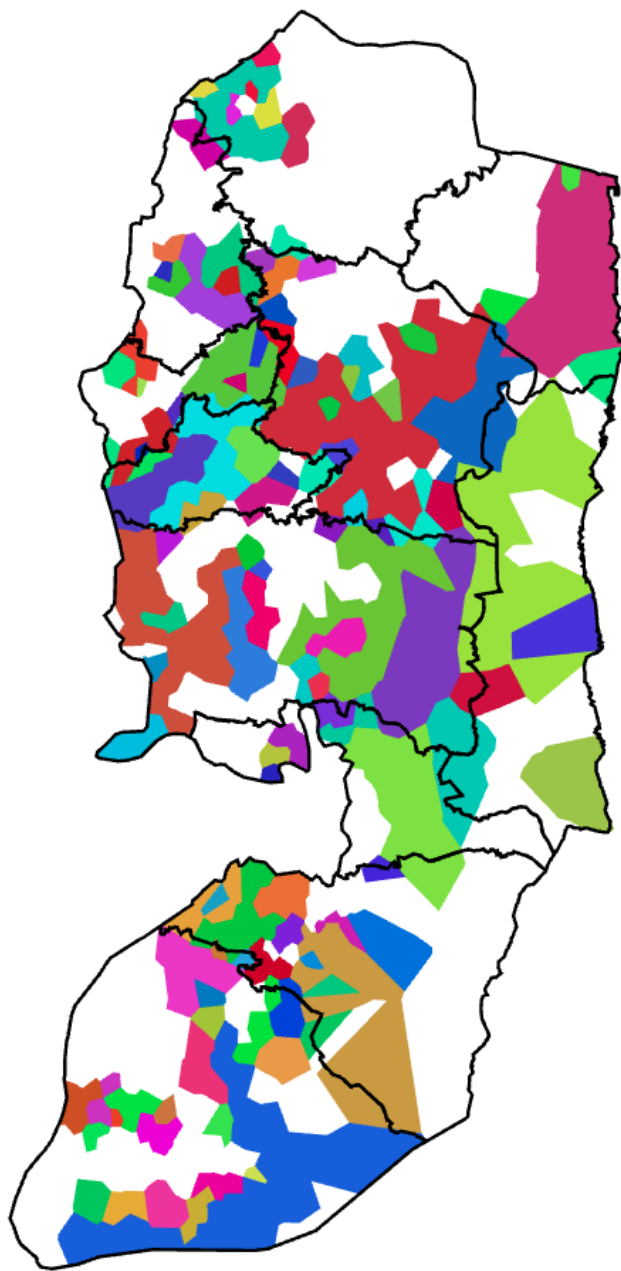
Notes: This figure presents the results of the leave-out regression robustness. Panel (a) shows events where settlers are perpetrators, Panel (b) shows events where Palestinians are perpetrators, Panel (c) shows events where settlers are targets, and Panel (d) shows events where Palestinian civilians are targets. The results display the distribution of the β coefficients for 85 iterations of sampling without replacement. The red dashed line indicates the baseline estimate from Table 2.2.

Figure A.30: Geographical and Administrative Components for Spatial Clustering



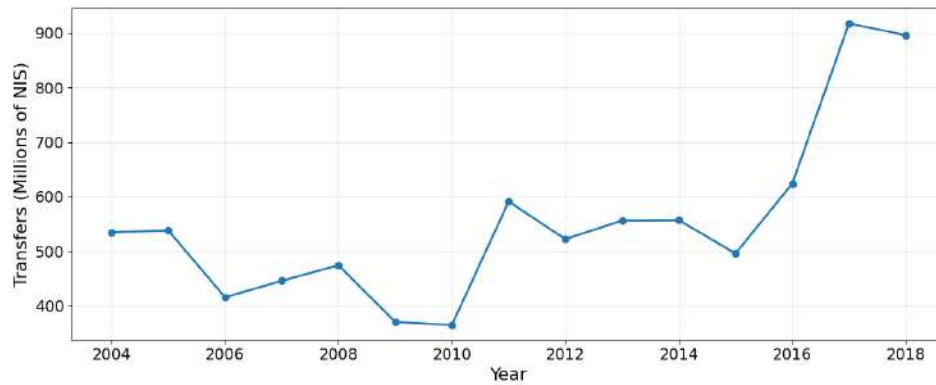
Notes: This figure displays the two distinct geographical layers used to define the spatial clusters. Panel (a) shows the administrative divisions from the Oslo Accords. Panel (b) shows the governorate boundaries as per the [Palestinian Central Bureau of Statistics \(2025\)](#). A polygon must be contiguous with another and share the same classification in *both* of these dimensions to be included in the same cluster. Only the Voronoi Polygons included in the regressions are displayed.

Figure A.31: Resulting Spatial Clusters for Standard Error Adjustment



Notes: This map illustrates the final spatial clusters generated by the algorithm. Each distinct, colored area represents a unique cluster, comprising all contiguous Voronoi polygons that fall within the same Oslo Area and the same Palestinian Governorate. Only the Voronoi Polygons included in the regressions are displayed. Colors are chosen to maximize contrast between adjacent clusters, although some similar colors may appear in non-adjacent clusters because of the final number of clusters (118).

Figure A.32: Fiscal Transfers to Settlements by Israeli Government



Notes: This figure presents the fiscal transfers to settlements by the Israeli government from 2013 to 2018. Since the last data point is from the 2nd quarter of 2018, I consider for each year just the first two quarters. The data is obtained from the Israeli Settlement Division reports on budget allocations to settlements in the occupied territories and disclosed through the Movement for Freedom of Information. Data source: <https://www.meida.org.il/9427>

A.2 Tables

Table A.1: Descriptive Statistics

	Pre-declaration		Post-declaration	
	No Outpost	Outpost	No Outpost	Outpost
Violent Events	0.065 (0.399)	0.088 (0.444)	0.201 (0.808)	0.381 (1.375)
Total Casualties	0.085 (1.653)	0.169 (3.367)	0.379 (6.121)	0.787 (16.826)
Target: Israeli	0.012 (0.126)	0.021 (0.162)	0.072 (0.421)	0.115 (0.651)
Perpetrator: Israeli	0.048 (0.324)	0.065 (0.378)	0.118 (0.560)	0.256 (0.952)
Target: Palestinian	0.049 (0.319)	0.062 (0.361)	0.114 (0.552)	0.244 (0.923)
Perpetrator: Palestinian	0.011 (0.122)	0.018 (0.153)	0.040 (0.304)	0.065 (0.475)
Casualties: Israeli	0.018 (0.231)	0.030 (0.656)	0.044 (0.506)	0.076 (0.643)
Casualties: Palestinian Civilians	0.076 (1.626)	0.143 (3.266)	0.357 (6.002)	0.755 (16.777)
Casualties: Israeli Civilians	0.012 (0.186)	0.017 (0.265)	0.027 (0.305)	0.055 (0.472)
Casualties: IDF	0.006 (0.107)	0.013 (0.459)	0.017 (0.394)	0.021 (0.377)
Casualties: Palestinian Militias	0.000 (0.016)	0.004 (0.272)	0.002 (0.069)	0.005 (0.353)
Population	3101.078 (4464.366)	3481.596 (4972.616)	3620.043 (5071.282)	3651.577 (5090.746)
N	3,976	7,363	6,755	23,318

Notes: Standard deviations in parentheses. 'Outpost' and 'No Outpost' refer to the binary classification in the methodology section.

Table A.2: Political Parties and Ideological Classifications

Party	Ideology	Party	Ideology	Party	Ideology
Da'am (Workers' Party)	Far Left	Tsav HaSh'a'a	Center/Moderate	Lehem	Right
National Arab Party	Far Left	Other	Center/Moderate	Hetz	Right
Da'am	Far Left	New Independents	Center/Moderate	HaLev	Right
Balad	Far Left	Nativ	Center/Moderate	Atid Ehad	Right
Ra'am - Balad	Far Left	Koach Lehashpia	Center/Moderate	Jewish Heart	Right
Hadash	Far Left	Every Vote Counts	Center/Moderate	United Religious Front	Right
Ra'am	Far Left	Ani Ve'Ata - Israel's People Party	Center/Moderate	Likud	Right
The Joint List	Far Left	Ale Yarok	Center/Moderate	Eretz Hadasha	Right
The Arab List	Far Left	30-40 List	Center/Moderate	National Religious Party	Right
Ra'am - Ta'al	Far Left	The Impossible - Possible	Center/Moderate	Haikva Leshimui	Right
Hadash - Ta'al	Far Left	Rafa - Rak Brit	Center/Moderate	National Unity	Right
Brit Olam	Left	Olam Chadash	Center/Moderate	Am Ehad	Right
Mishpat Tzedek	Left	New Economy	Center/Moderate	Israel Ahavat	Right
Strength of the Poor	Left	Chetz	Center/Moderate	The Israelis	Right
Labor-Gesher-Meretz	Left	Atzmenu	Center/Moderate	Biblical Bloc Party	Right
Kol HaNashim	Left	Anachnu	Center/Moderate	Yisrael Beiteinu	Right
Justice for All (Tzedek Lakol)	Left	Vision	Center/Moderate	Ha'achdut Ha'Amami	Right
Labor-Gesher	Left	Uvitzchuan	Center/Moderate	New Hope (Gideon Sa'ar)	Right
Mitkademet	Left	Seder Chadash	Center/Moderate	The Secular Right	Right
Justice	Left	Koach LeHashpi'a	Center/Moderate	Ahavat Israel	Right
Green Economy - One State	Left	HaBrit HaMeshutefet	Center/Moderate	Economic Freedom	Right
Justice to All	Left	Red and White	Center/Moderate	Yisrael Chazaka	Right
Kvod HaAdam	Left	New Order	Center/Moderate	Nivcheret Ha'am Hazmanit	Right
Zkhuyotenu Bekolenu	Left	Kavod Veshivyon	Center/Moderate	Hayisraelim	Right
Betach - Social Security	Left	Christian Liberal Movement	Center/Moderate	Tzabar	Right
Green Movement	Left	Sikhrut Bekavod	Center/Moderate	Moreshet Avot	Right
Holocaust Survivors and Green Leaf Alumni	Left	Perach	Center/Moderate	HaYisraelim	Right
Green Leaf	Left	Manhigut Chevrutit	Center/Moderate	Tzaon	Right
Labor Party	Left	Am Shalem	Center/Moderate	Zehut	Right
Meretz	Left	Yisrael HaMithadeshet	Center/Moderate	Lazuz	Right
Tzedek Hevrat	Left	Or	Center/Moderate	N Noach - Mamlachtit List	Right
Progressive National Alliance	Left	Men's Rights in the Family (Ra-ash)	Center/Moderate	Magen	Right
Social Justice	Left	Koach LeHashpi'a	Center/Moderate	Brave for You	Right
Greens	Left	Tafnit	Center/Moderate	Herut	Far Right
Green Party	Left	Lev	Center/Moderate	Tzomet	Far Right
Labor	Left	Kadima	Center/Moderate	Lahava	Far Right
Brit Olam Legendat Yisrael	Left	Gil	Center/Moderate	Yachad	Far Right
Brit HaShitufit	Left	Leeder	Center/Moderate	Yisrael Ba-Aliya	Far Right
Zionist Union	Left	Center Party	Center/Moderate	Eretz Yisrael Shelanu	Far Right
Burning Young	Left	Men's Right in the Family	Center/Moderate	Ma'au (Yachad) - Lel'dan Hadash	Far Right
Environmental and Animal Voice	Left	Citizen and State	Center/Moderate	The Jewish Home	Far Right
Green Leaf - Liberal List	Left	Shinui	Center/Moderate	Otzma LeYisrael	Far Right
Adrayut LaMevasdim	Center/Moderate	Shas	Center/Moderate	New Zionism	Far Right
Ani Ve'Ata - Israel's People Party	Center/Moderate	Haim Bekavod	Center/Moderate	Yamina - New Right	Far Right
Gesher	Center/Moderate	Hatma	Center/Moderate	Otzma Yehudit	Far Right
Simply Love	Center/Moderate	Alrayut	Center/Moderate	Otzma Kalkalit	Far Right
Education	Center/Moderate	Kulanu Haverim	Center/Moderate	Biblical Bloc Party	Far Right
Blue and White	Center/Moderate	Leader (Progressive Liberal Democratic)	Center/Moderate	Jewish Home	Far Right
Shavim	Center/Moderate	Yesh Atid	Center/Moderate	National Union	Far Right
Mehatchala	Center/Moderate	The Pirates	Center/Moderate	Yamina	Far Right
Ichud Bnei HaBrit	Center/Moderate	Koach Lehashpia	Center/Moderate	The New Right	Far Right
Yashar - True Democracy	Center/Moderate	Kulanu	Center/Moderate	Yisrael Beiteinu	Far Right
United Torah Judaism	Center/Moderate	The Economics Party	Center/Moderate	Religious Zionism	Far Right
Shma	Center/Moderate	Orr (Light)	Center/Moderate	Otzma Liberalit-Kalkalit	Far Right
Kana	Center/Moderate	Democratatura Party	Center/Moderate		
Yesh Kivun	Center/Moderate	National Unity Party	Center/Moderate		
			Right		

Table A.3: Spatial Characteristics Summary

Characteristic	Value
Total Area (km ²)	3,273.67
Area A (%)	8.2
Area B (%)	18.6
Area C (%)	72.4
Hebron (%)	0.7
Palestinian Towns (%)	78.5
Palestinian Population	1,432,217
Israeli Population	269,272

Notes: Area percentages calculated relative to total area. Palestinian Towns shows the percentage of Voronoi polygons that are Palestinian towns.

Table A.4: Pre-declaration Balance Test (KS)

Variable	Control	Treatment	Difference	P-value
Violent Events	0.065 (0.399)	0.088 (0.444)	0.023	0.470
Total Casualties	0.085 (1.653)	0.169 (3.367)	0.084	0.816
Target: Israeli	0.012 (0.126)	0.021 (0.162)	0.009	0.991
Perpetrator: Israeli	0.048 (0.324)	0.065 (0.378)	0.017	0.906
Target: Palestinian	0.049 (0.319)	0.062 (0.361)	0.013	0.985
Perpetrator: Palestinian	0.011 (0.122)	0.018 (0.153)	0.006	1.000
Casualties: Israeli	0.018 (0.231)	0.030 (0.656)	0.013	1.000
Casualties: Palestinian Civilians	0.076 (1.626)	0.143 (3.266)	0.067	0.945
Casualties: Israeli Civilians	0.012 (0.186)	0.017 (0.265)	0.005	1.000
Casualties: IDF	0.006 (0.107)	0.013 (0.459)	0.007	1.000
Casualties: Palestinian Militias	0.000 (0.016)	0.004 (0.272)	0.004	1.000
Population	3101.078 (4464.366)	3481.596 (4972.616)	380.518	0.000***

Notes: Standard deviations in parentheses. Kolmogorov-Smirnov test used for distribution comparison. *** p<0.001, ** p<0.01, * p<0.05

Table A.5: Effect of Transitory Outpost Establishment on Violent Events

	Violent Events					
	All Events		Settler Involved		Civilian Targeting	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Outpost	0.025 (0.022)	0.016 (0.024)	0.017 (0.016)	0.019 (0.015)	0.032* (0.017)	0.029* (0.018)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓
Pre-treatment Mean	0.000	0.000	0.000	0.000	0.000	0.000
<i>N</i>	4,071	4,071	4,071	4,071	4,071	4,071

Notes: Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. Results based on transitory treatment definition (outposts existing for limited periods). *p<0.1, **p<0.05, ***p<0.01.

Table A.6: Heterogeneous Effects of Transitory Outpost Establishment by Target and Perpetrator

	Main Target				Main Perpetrator			
	Settlers		Palestinian Civilians		Settlers		Palestinians	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post $\times$ Outpost	0.003 (0.003)	0.001 (0.002)	0.025 (0.022)	0.022 (0.023)	0.016 (0.016)	0.018 (0.015)	0.002 (0.004)	-0.004 (0.004)
Unit FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓		✓
Pre-treatment Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>N</i>	4,071	4,071	4,071	4,071	4,071	4,071	4,071	4,071

Notes: Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. Results based on transitory treatment definition (outposts existing for limited periods). *p<0.1, **p<0.05, ***p<0.01.

Table A.7: Effect of Settler Violence and Demolitions on Palestinian Violence (Temporal Comparison)

	Palestinian Violent Events					
	All Palestinian		Palestinian Militias		Palestinian Civilians	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: January 2018 – May 2021</i>						
Settler Perpetrators _{<i>t</i>-1}	0.006 (0.017)	0.005 (0.017)	-0.002 (0.003)	-0.002 (0.003)	0.007 (0.017)	0.007 (0.017)
Settler Perpetrators _{<i>t</i>-2}	-0.028 (0.018)	-0.028 (0.019)	0.003 (0.005)	0.003 (0.005)	-0.031* (0.017)	-0.031* (0.017)
IDF Perpetrators _{<i>t</i>-1}	-0.037** (0.016)	-0.037** (0.016)	-0.001 (0.003)	-0.001 (0.003)	-0.036** (0.015)	-0.036** (0.015)
IDF Perpetrators _{<i>t</i>-2}	0.029 (0.023)	0.029 (0.023)	0.002 (0.002)	0.002 (0.002)	0.028 (0.023)	0.027 (0.023)
Demolitions _{<i>t</i>-1} (housing units)	0.006 (0.005)	0.006 (0.005)	-0.000 (0.000)	-0.000 (0.000)	0.006 (0.005)	0.006 (0.005)
Demolitions _{<i>t</i>-2} (housing units)	-0.008 (0.006)	-0.008 (0.006)	-0.001 (0.000)	-0.000 (0.000)	-0.007 (0.006)	-0.008 (0.006)
W(Settler Perpetrators) _{<i>t</i>-1}		-0.038 (0.032)		0.003 (0.005)		-0.041 (0.032)
W(Settler Perpetrators) _{<i>t</i>-2}		-0.026 (0.025)		0.009 (0.007)		-0.035 (0.027)
W(IDF Perpetrators) _{<i>t</i>-1}		-0.006 (0.015)		0.001 (0.003)		-0.008 (0.015)
W(IDF Perpetrators) _{<i>t</i>-2}		0.007 (0.015)		0.003 (0.004)		0.004 (0.014)
<i>N</i>	9,516	9,516	9,516	9,516	9,516	9,516
<i>Panel B: June 2021 – September 2023</i>						
Settler Perpetrators _{<i>t</i>-1}	0.197** (0.085)	0.197** (0.085)	0.076 (0.050)	0.076 (0.050)	0.120*** (0.043)	0.120*** (0.044)
Settler Perpetrators _{<i>t</i>-2}	0.060** (0.030)	0.060** (0.031)	0.052 (0.032)	0.051 (0.032)	0.007 (0.028)	0.009 (0.028)
IDF Perpetrators _{<i>t</i>-1}	-0.140* (0.074)	-0.139* (0.073)	-0.087* (0.048)	-0.086* (0.048)	-0.052 (0.041)	-0.052 (0.041)
IDF Perpetrators _{<i>t</i>-2}	0.008 (0.058)	0.008 (0.058)	-0.026 (0.048)	-0.025 (0.048)	0.033 (0.040)	0.033 (0.040)
Demolitions _{<i>t</i>-1} (housing units)	0.018 (0.022)	0.019 (0.022)	0.008 (0.012)	0.008 (0.013)	0.010 (0.021)	0.010 (0.021)
Demolitions _{<i>t</i>-2} (housing units)	0.008 (0.008)	0.008 (0.008)	0.006 (0.006)	0.006 (0.006)	0.002 (0.005)	0.002 (0.005)
W(Settler Perpetrators) _{<i>t</i>-1}		0.000 (0.048)		0.021 (0.032)		-0.021 (0.049)
W(Settler Perpetrators) _{<i>t</i>-2}		0.073 (0.106)		0.005 (0.032)		0.068 (0.103)
W(IDF Perpetrators) _{<i>t</i>-1}		-0.011 (0.043)		-0.032 (0.030)		0.022 (0.031)
W(IDF Perpetrators) _{<i>t</i>-2}		-0.003 (0.059)		-0.064* (0.037)		0.059 (0.067)
<i>N</i>	7,320	7,320	7,320	7,320	7,320	7,320
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Spatial Lags		✓		✓		✓

Notes: Two-way fixed effects panel regressions. Outcomes are counts of Palestinian violent events. Covariates are lagged settler violence events and housing unit demolitions at $t - 1$ and $t - 2$. Even-numbered columns include spatial lags (W) of all covariates using queen contiguity weights. Panel A covers the period January 2018 to May 2021. Panel B covers June 2021 to September 2023. Standard errors clustered at the unit level in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A.8: Robustness Check: Effect of Outpost Establishment on Violent Events across Different Intensity Thresholds

	Violent Events					
	All Events		Settler Involved		Civilian Targeting	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Outpost Intensity ≤ 0.01						
Post $\times$ Outpost	0.130*** (0.045)	0.119*** (0.044)	0.049*** (0.017)	0.051*** (0.016)	0.030** (0.014)	0.027** (0.013)
Pre-treatment Mean	0.119	0.119	0.031	0.031	0.025	0.025
Panel B: Outpost Intensity ≤ 0.05						
Post $\times$ Outpost	0.118** (0.049)	0.123*** (0.044)	0.040** (0.018)	0.033** (0.016)	0.031** (0.012)	0.018 (0.011)
Pre-treatment Mean	0.109	0.109	0.035	0.035	0.031	0.031
Panel C: Outpost Intensity ≤ 0.12						
Post $\times$ Outpost	0.159*** (0.046)	0.171*** (0.044)	0.066*** (0.015)	0.071*** (0.015)	0.044*** (0.011)	0.042*** (0.011)
Pre-treatment Mean	0.120	0.120	0.037	0.037	0.029	0.029
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓
N	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Kneset linear trend is included. Panel A includes outposts with intensity ≤ 0.01 , Panel B with intensity ≤ 0.05 , and Panel C with intensity ≤ 0.15 . *p<0.1, **p<0.05, ***p<0.01.

Table A.9: Robustness Check: Heterogeneous Effects by Target and Perpetrator across Different Intensity Thresholds

	Main Target				Main Perpetrator			
	Settlers		Palestinian Civilians		Settlers		Palestinians	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Outpost Intensity ≤ 0.01								
Post $\times$ Outpost	-0.002 (0.003)	-0.005 (0.004)	0.082*** (0.027)	0.080*** (0.028)	0.052*** (0.014)	0.044*** (0.013)	0.028** (0.012)	0.021* (0.011)
Pre-treatment Mean	0.008	0.008	0.073	0.073	0.031	0.031	0.021	0.021
Panel B: Outpost Intensity ≤ 0.05								
Post $\times$ Outpost	0.000 (0.003)	-0.001 (0.003)	0.089*** (0.029)	0.077*** (0.027)	0.043*** (0.015)	0.034** (0.014)	0.012 (0.017)	0.024** (0.011)
Pre-treatment Mean	0.011	0.011	0.076	0.076	0.037	0.037	0.013	0.013
Panel C: Outpost Intensity ≤ 0.12								
Post $\times$ Outpost	0.004* (0.002)	0.002 (0.002)	0.120*** (0.025)	0.124*** (0.025)	0.063*** (0.013)	0.067*** (0.013)	0.017 (0.016)	0.022* (0.012)
Pre-treatment Mean	0.010	0.010	0.083	0.083	0.039	0.039	0.012	0.012
Unit FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓		✓
N	41,412	41,412	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. Panel A includes outposts with intensity ≤ 0.01 , Panel B with intensity ≤ 0.05 , and Panel C with intensity ≤ 0.12 . *p<0.1, **p<0.05, ***p<0.01.

Table A.10: Effect of Outpost Establishment on Violent Events (Non-Staggered Design)

	Violent Events					
	All Events		Settler Involved		Civilian Targeting	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Outpost	0.157*** (0.046)	0.151*** (0.044)	0.060*** (0.015)	0.056*** (0.014)	0.041*** (0.010)	0.039*** (0.010)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓
Pre-treatment Mean	0.074	0.074	0.021	0.021	0.028	0.028
N	36,888	36,888	36,888	36,888	36,888	36,888

Notes: Standard errors clustered at the local entity level in parentheses. Non-staggered design using only first treated cohort (month 24) vs never treated and later cohorts before treatment. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

Table A.11: Heterogeneous Effects by Target and Perpetrator (Non-Staggered Design)

	Main Target				Main Perpetrator			
	Settlers		Palestinian Civilians		Settlers		Palestinians	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post $\times$ Outpost	0.003 (0.003)	0.003 (0.003)	0.118*** (0.025)	0.110*** (0.025)	0.061*** (0.013)	0.058*** (0.013)	0.011 (0.013)	0.012 (0.010)
Unit FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓		✓
Pre-treatment Mean	0.012	0.012	0.052	0.052	0.025	0.025	0.015	0.015
<i>N</i>	36,888	36,888	36,888	36,888	36,888	36,888	36,888	36,888

Notes: Standard errors clustered at the local entity level in parentheses. Non-staggered design using only first treated cohort (month 24) vs never treated and later cohorts before treatment. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

Table A.12: Effect of Outpost Establishment on Violent Events (IHS Transformation)

	Violent Events					
	All Events		Settler Involved		Civilian Targeting	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Outpost	0.062*** (0.021)	0.075*** (0.019)	0.037*** (0.010)	0.042*** (0.009)	0.030*** (0.007)	0.031*** (0.007)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓
Pre-treatment Mean	0.115	0.115	0.033	0.033	0.031	0.031
<i>N</i>	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Dependent variables are transformed using inverse hyperbolic sine (IHS). ATT estimates are recovered to original scale. Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

Table A.13: Heterogeneous Effects by Target and Perpetrator (IHS Transformation)

	Main Target				Main Perpetrator			
	Settlers		Palestinian Civilians		Settlers		Palestinians	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post $\times$ Outpost	0.003 (0.002)	0.003 (0.002)	0.062*** (0.014)	0.066*** (0.014)	0.038*** (0.009)	0.042*** (0.008)	0.006 (0.010)	0.013* (0.007)
Unit FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓		✓
Pre-treatment Mean	0.009	0.009	0.082	0.082	0.035	0.035	0.012	0.012
<i>N</i>	41,412	41,412	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Dependent variables are transformed using inverse hyperbolic sine (IHS). ATT estimates are recovered to original scale. Standard errors clustered at the local entity level in parentheses. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

Table A.14: Effect of Outpost Establishment on Violent Events

	Violent Events					
	All Events		Setter Involved		Civilian Targeting	
	(1)	(2)	(3)	(4)	(5)	(6)
Post $\times$ Outpost	0.153*** (0.055)	0.172*** (0.050)	0.060*** (0.022)	0.065*** (0.022)	0.043*** (0.014)	0.042*** (0.013)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓
Pre-treatment Mean	0.115	0.115	0.033	0.033	0.031	0.031
<i>N</i>	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Standard errors clustered at the neighbor-area-governorate level. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

Table A.15: Heterogeneous Effects by Target and Perpetrator

	Main Target				Main Perpetrator			
	Settlers		Palestinian Civilians		Settlers		Palestinians	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post $\times$ Outpost	0.004 (0.002)	0.003 (0.002)	0.115*** (0.037)	0.121*** (0.036)	0.058*** (0.020)	0.062*** (0.020)	0.014 (0.017)	0.024** (0.012)
Unit FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Population & Linear Time Trend		✓		✓		✓		✓
Pre-treatment Mean	0.009	0.009	0.082	0.082	0.035	0.035	0.012	0.012
<i>N</i>	41,412	41,412	41,412	41,412	41,412	41,412	41,412	41,412

Notes: Standard errors clustered at the neighbor-area-governorate level in parentheses. In even columns a Unit id-Knesset linear trend is included. *p<0.1, **p<0.05, ***p<0.01.

B Methodology for LLM-Assisted Classification

The following prompt was fed to ChatGPT 4o-mini and used to classify the violent events in the ACLED dataset:

This is the short description of an event in the West Bank:

{ Violent event textual description }

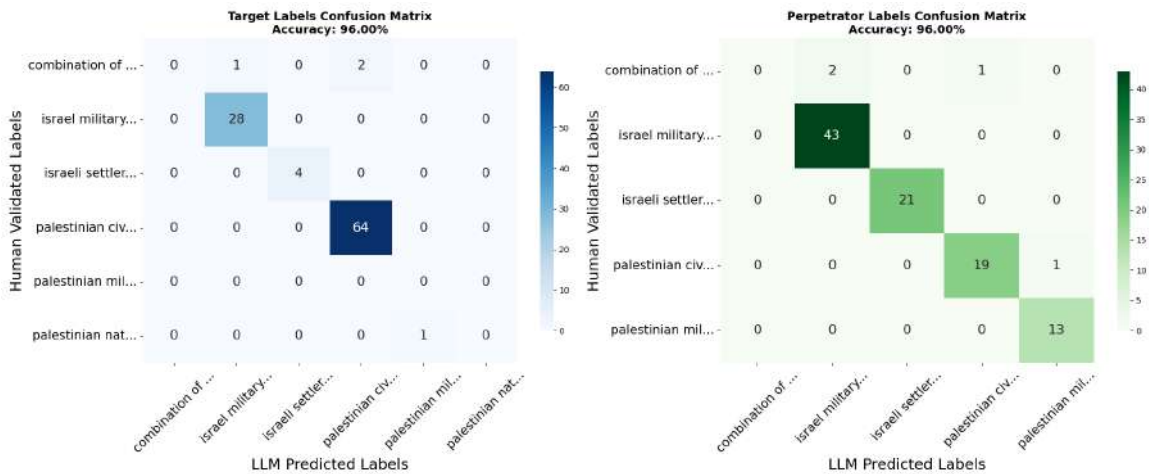
I would like you to provide me the following information:

1. Who was the main perpetrator of the event? Answer Palestinian Civilians, Israeli Settlers/Civilians, Palestinian Militias, Israeli Military, Palestinian National Authority or a combination of those (specify which) or Other.
2. What was the main target of the event? Answer Palestinian Civilians, Israeli Settlers/Civilians, Palestinian Militias, Israeli Military, Palestinian National Authority, a combination of those (specify which) or Other.
3. How many Palestinian Civilians casualties (i.e. killed or injured) in the event? Answer with a number.
4. How many Israeli Settlers/Civilians casualties (i.e. killed or injured) in the event? Answer with a number.
5. How many Palestinian Militias casualties (i.e. killed or injured) in the event? Answer with a number.
6. How many Israeli Military casualties (i.e. killed or injured) in the event? Answer with a number.

Answer the questions in the following format:

1. Perpetrator: Answer
2. Target: Answer
3. Palestinian Civilians: Answer
4. Israeli Settlers/Civilians: Answer
5. Palestinian Militias: Answer

Figure B.33: Confusion Matrices for Perpetrators and Targets



Notes: Confusion matrices for the classification of perpetrators and targets. Only cases in which the classification did not include “Combination” or “Unknown” are considered.

The output was then processed to extract the relevant information. Manual checks were conducted to ensure the accuracy of the classification. As for data on casualties, the output was revised to ensure consistency and to transform textual answers into numerical values –e.g. “several” was transformed into 3, “dozen(s)” into 12(24). When no relevant information was reported the number of casualties was set to missing.

To validate the results, I randomly sampled 100 events and manually classified them. The output of the validation is displayed in Figure B.33. It is important to note that labels are considerably more difficult to assign –even for a human coder –when the perpetrator or target is “Combination” of different actors. Therefore, I only consider cases in which the classification did not include “Combination” or “Unknown”. The accuracy of the classification is 96% for perpetrators and targets. Only in three cases the classification was wrong for perpetrators and targets, as it should have been “Combination”, since it referred to events occurring during a riot, where both Palestinians and Israeli forces were involved. In another instance, the Palestinian National Authority was wrongly classified as Palestinian Militias. Overall, the classification appears to be very accurate and reliable.

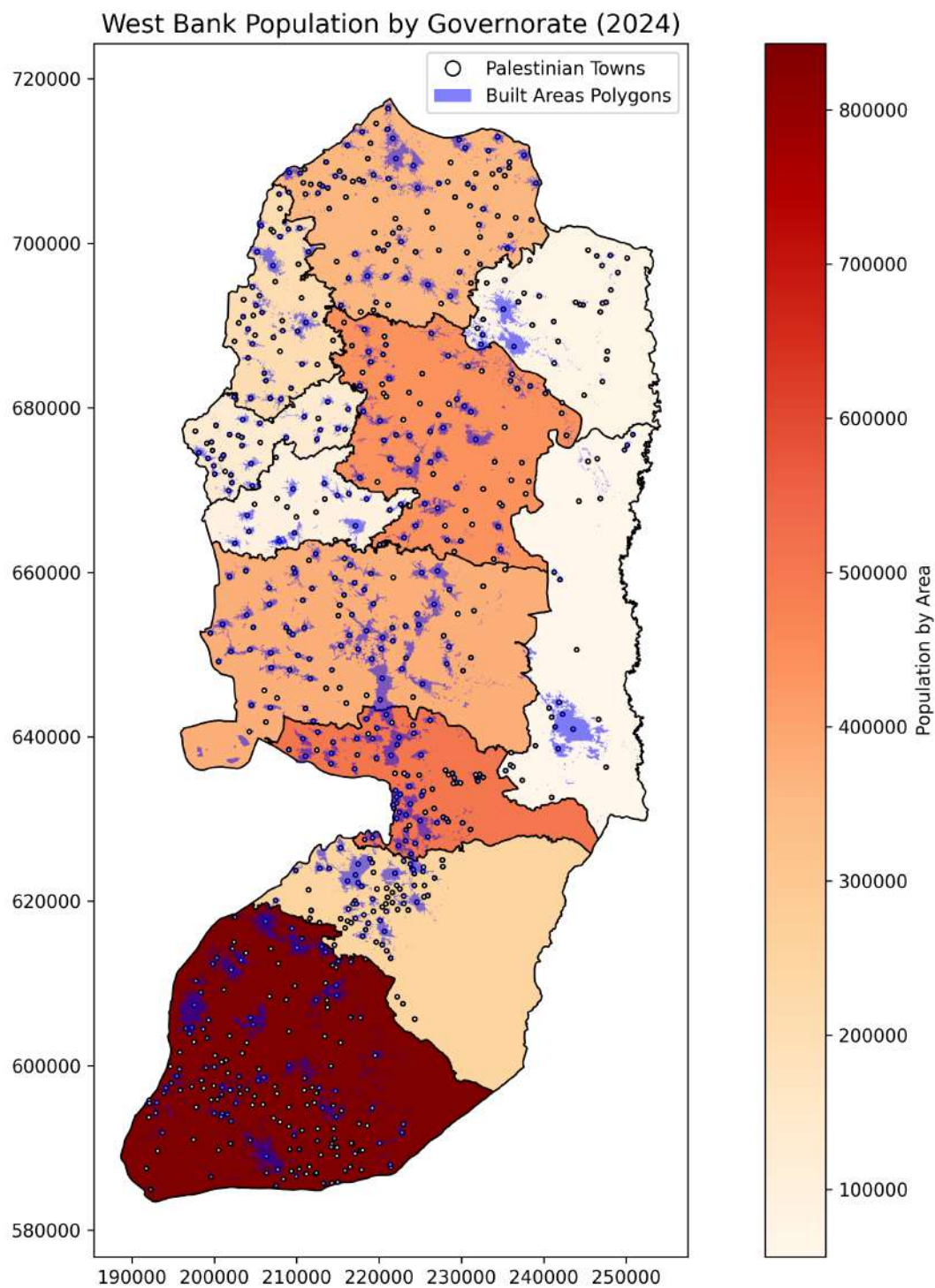
C Estimating Population of Palestinian Towns

In order to estimate the population of Palestinian towns, I first obtain the aggregated population of the eleven West Bank governorates from the [Palestinian Central Bureau of Statistics \(2025\)](#) for the years 2015 to 2023. Then I obtain the population of each town and village with the following procedure:

1. I obtain the list of Palestinian towns and villages, along with built-up areas from the B'Tselem database ([B'Tselem, 2024](#)).
2. Around each Palestinian town –i.e., the dots in Figure [C.34](#), I create a buffer area of 1km.
3. I then calculate the portion of the buffer area that overlaps with the built-up area of the town, i.e., the blue polygons in Figure [C.34](#).
4. By doing this I am able to obtain weights for each town, normalize them to sum to 1, and then multiply them by the population of the governorate in which the town is located.

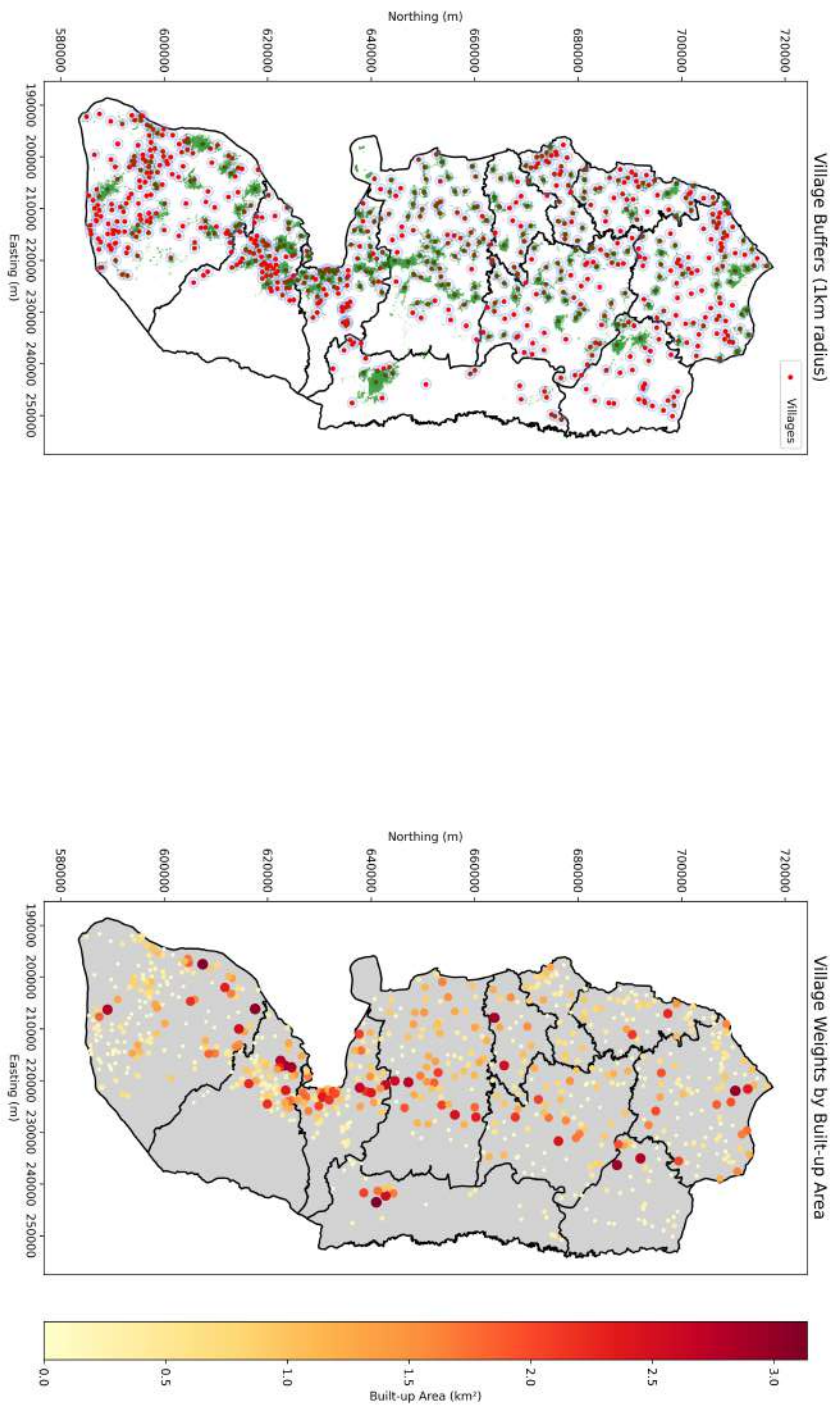
I then use a log-linear interpolation of the yearly data points to obtain the population for each month. Figures [C.35](#) and [C.36](#) graphically illustrate the procedure.

Figure C.34: Population of West Bank Governorates, Towns and Built-up Areas



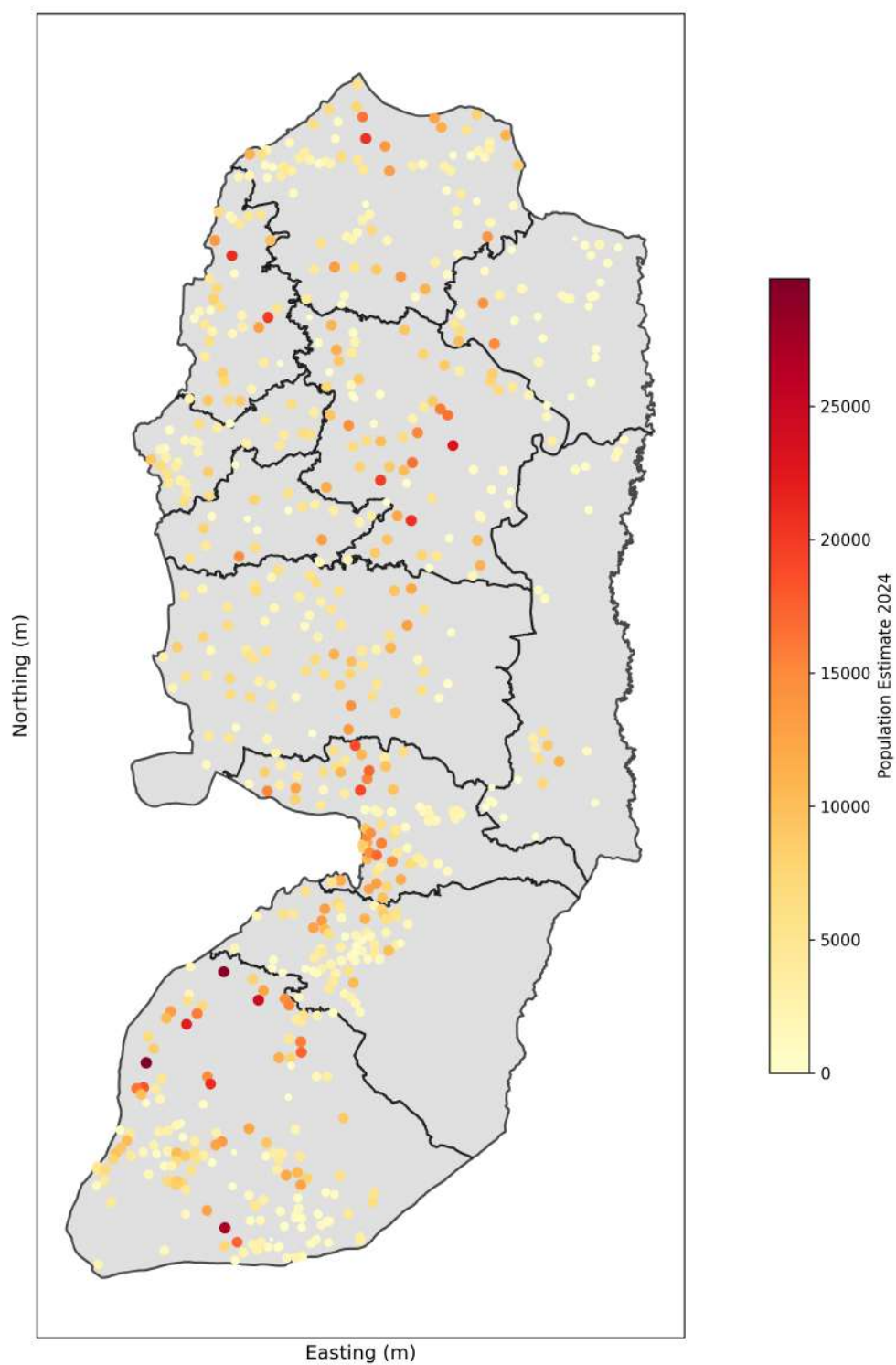
Notes: Population of the West Bank governorates. Source: Palestinian Central Bureau of Statistics. The dots represent the Palestinian towns and villages, while the blue polygons represent the built-up areas of the towns (Source [B'Tselem \(2024\)](#)).

Figure C.35: Population of West Bank Governorates: Buffer Areas and Weights



Notes: On the left panel the buffer areas are shown, while on the right panel the non-normalized weights for each town are displayed.

Figure C.36: Estimated Population of West Bank



Notes: Population estimates for each town.

Chapter 3

Mafia Network: a Machine-Learning Approach

Joint work with Monica Billio and Marco Di Cataldo

Abstract: This paper develops a novel early-warning system to predict firm confiscation due to mafia infiltration in Italy. We construct a dynamic, time-resolved corporate ownership network by tracing the first- and second-degree connections of all Italian firms confiscated between 2000 and 2024. Using this unique dataset, we train a series of machine learning models, combining firm-level financial data with a comprehensive set of network topology measures to forecast firm confiscation up to three years in advance. Our out-of-sample results demonstrate high predictive power, with the combined model achieving an Area Under the Curve (AUC) score of up to 0.975. We find that network features are crucial for achieving high recall, while firm-level financials provide higher precision. The integrated model effectively balances this trade-off, outperforming models that rely on either data source alone. A feature-contribution analysis using SHAP values reveals that firm size is the single most important predictor, with smaller and financially distressed firms facing higher confiscation risk. Network topology features also emerge as the second most important predictors. Taken together, our findings underscore the value of integrating ownership network intelligence with firm fundamentals to create a powerful and interpretable tool for risk-based supervision and enforcement against organized crime.

JEL Classification: L1, K14, L25, C25, C38

Keywords: mafia, organized crime, legal economy, network, machine learning, forecasting, firms, Italy.

3.1 Introduction

Over the past few decades, organized crime groups (OCGs) have pushed well beyond their traditional spheres of illicit activity, making inroads into legitimate businesses and economies across the globe. A recent Europol report underscores this escalating trend, revealing that around 86% of the EU’s most dangerous criminal networks infiltrate lawful enterprises to bolster, conceal, and enable a broad range of illegal operations, as well as to launder their proceeds (Europol, 2024). The scope and complexity of this phenomenon have increasingly drawn the attention of economic scholars. Researchers in the field are examining how mafia *infiltration*, or the presence of OCGs in legitimate markets, can alter economic incentives, undermine fair competition and produce distortions in the broader economic environment (Pinotti, 2015; Di Cataldo and Mastrorocco, 2022; Le Moglie and Sorrenti, 2022). In light of this, it is of vital importance to understand the strategic behaviour of criminals in forming such distortions (Mastrobuoni, 2020; Daniele and Dipoppa, 2023), as well as to investigate the extent to which anti-mafia legislation can be effective in tackling this issue (Calamunci and Drago, 2020; Chircop et al., 2023; Boeri et al., 2024).

While the infiltration of legal economies by OCGs is drawing increasing scholarly attention (Mirenda et al., 2022; Arellano-Bover et al., 2024; Le Moglie and Sorrenti, 2022), research has largely focused on ex-post analysis of the effects of OCG infiltration. This paper breaks new ground by developing an early-warning system that is the first to integrate dynamic corporate ownership networks for the purpose of predicting firm confiscation due to OCG infiltration. By applying machine learning to a unique dataset combining network topology with firm-level data, we create a framework that successfully forecasts confiscation up to three years in advance. Our analysis reveals the distinct predictive roles of network versus financial data and identifies the key firm characteristics associated with infiltration risk. Most importantly, we show the critical importance of integrating both feature sets to build a robust predictive model.

It is important to clarify our predictive target: confiscation, an unambiguous and judicially confirmed signal of OCG infiltration. The broader phenomenon of infiltration itself is latent, as authorities may not detect every compromised firm; thus, not all infiltrated firms are ultimately confiscated. By definition, however, a confiscation event serves as definitive confirmation that a firm was, in fact, infiltrated. Although this focus on judicially-verified outcomes restricts our analysis to a subset of all infiltrated firms,

it is a necessary trade-off. This approach provides the analytical certainty required to train and validate our predictive models. Building on this framework, our paper makes four primary contributions.

Our first contribution is the creation of a novel, dynamic ownership network for Italy. To ensure our analysis is anchored on a ground truth free of false positives, we build our network starting from judicially-confirmed confiscated firms rather than those merely suspected of infiltration. From these seed firms, we reconstruct their corporate environment by linking them to their direct and second-order owners and subsidiaries –encompassing both firms (*Persone Giuridiche*) and individuals (*Persone Fisiche*). The result is a directed, weighted, and dynamic graph of corporate control, with edges representing ownership stakes timestamped at the daily level. After integrating this network information with comprehensive firm-level financials, we obtain a rich, attributed graph that serves as the foundation for our predictive modelling. The unique granularity of this dataset enables our core methodological shift: instead of an ex-post analysis of consequences, we leverage the network’s evolution to test if the changing topology of corporate ownership is a powerful leading indicator of confiscation and, consequently, infiltration risk.

Our second key contribution stems from our predictive modelling, which sheds light on crime infiltration with practical implications for law enforcement authorities. Using a supervised machine learning framework, we develop an early-warning system designed to predict the likelihood of a firm’s future confiscation due to mafia infiltration. Specifically, our analysis reveals a stark precision-recall trade-off between the different feature sets. Models trained exclusively on firm-level financial data tend to be highly precise; that is, when they flag a firm as high-risk, they are very likely to be correct, i.e. up to ≈ 0.99 . However, their critical weakness is low recall, ≈ 0.67 , meaning they fail to identify a large portion of the firms that are actually infiltrated, allowing many potential targets to go undetected. On the other hand, models relying solely on network topology data are comprehensive, achieving high recall by flagging a broad set of suspicious companies, i.e. up to ≈ 0.94 . The significant drawback is lower precision, i.e. ≈ 0.72 , generating a substantial number of false positives and consequently making it less accurate. Crucially, our integrated model, which combines both data sources, successfully resolves this operational trade-off. It achieves a powerful and effective balance, maintaining the high precision needed to avoid false alarms while retaining the high recall necessary to ensure comprehensive detection of potential targets, i.e. an F1-score of up to ≈ 0.91 with

precision ≈ 0.97 and recall ≈ 0.86 . From a law enforcement standpoint, this balanced outcome is vital, as it provides a practical tool that allows for the efficient allocation of scarce investigative resources to a list of the highest-risk firms that is both accurate and complete. This adds a significant contribution to the literature on the *efficiency-security* trade-off, in the context of criminal networks and anti-mafia enforcement (Morselli, 2008; Lu et al., 2025).

Third, we run a feature-contribution analysis, where we complement the model’s performance results by identifying the specific characteristics linked to a higher risk of confiscation. To achieve this, we employ SHapley Additive exPlanations, or SHAP, values, a method that allows us to precisely quantify the impact of each feature on the model’s output for any given firm (Lundberg et al., 2020). This granular analysis reveals a clear hierarchy of predictive factors. The findings show that firm size, measured by the number of employees, is the single most influential predictor. The relationship is unambiguous: smaller and financially distressed firms consistently face the highest confiscation risk. Following size, financial indicators of distress—such as low profitability and poor liquidity—emerge as another crucial set of predictors. This empirical finding aligns with previous research, which posits that smaller, financially vulnerable firms are not only prime targets for criminal infiltration but are also more likely to be discovered by authorities, as they are often directly involved in illicit activities (Le Moglie and Sorrenti, 2022; Arellano-Bover et al., 2024). We also find that the models learn complex, non-linear patterns from network topology that defy simple interpretation, suggesting a promising area for future research. Taken together, our findings highlight the value of combining ownership-network intelligence with firm fundamentals to produce timely and interpretable early-warning signals, offering a practical basis for risk-based supervision and prioritized audits.

Fourth, to address the fundamentally latent nature of infiltration—an activity that often remains hidden until long after it has distorted firms and markets—we develop a representation-based approach that extracts its structural signals directly from the ownership network. We train a Temporal Graph Network (TGN) (Rossi et al., 2020) on the dynamics of link creation, updates, and removal, producing time-varying ownership

embeddings that encode each firm’s evolving structural position.¹ These representations provide a rigorous, data-driven proxy for latent infiltration: firms later confiscated already lie closer in embedding space to previously confiscated firms, and the embeddings alone achieve near-perfect discrimination in out-of-sample tests. Starting from this latent structure, we plan to both identify when infiltration is likely to emerge and use this learned latent representation to study how OCG infiltration shapes and impacts the broader corporate and economic environment.

Crucially, it is important to emphasize that the methodological approach in this paper is fundamentally observational. Our machine learning models, while successful in their predictive capacity, establish powerful statistical correlations between firm characteristics, network structures, and future confiscation events. We make no claims of causality; the findings are descriptive and predictive, not a statement that these features cause infiltration or confiscation. A key avenue for future development, which we plan to explore, involves leveraging the latent infiltration risk measure derived from our TGN embeddings as a new tool to move beyond prediction and better understand the causal impact of OCG infiltration on economic outcomes, confronting our results with those of the existing literature. Importantly, the framework we propose is meant to complement and support the work of investigative and judicial authorities, not to replace it; any insights generated by our model should be interpreted as an additional analytical aid rather than a standalone diagnostic tool.

Related Literature A growing body of research examines how OCGs embed themselves within the legal economy and reshape incentives, firm dynamics, and public resource allocation. Many recent studies attempt to trace the firm-level footprints of infiltration. Using administrative and balance-sheet data, [Mirenda et al. \(2022\)](#) show that targeted firms exhibit anomalous revenue growth unaccompanied by proportional input expansion, followed by deteriorating fundamentals—patterns consistent with rent extraction and money laundering rather than productivity gains. In a broader perspective, [Pinotti \(2015\)](#) document sizeable aggregate costs of organized crime in Southern Italy, while [Le Moglie and Sorrenti \(2022\)](#) suggest that OCG entry tends to occur where firms are smaller and financially weaker, pointing to systematic selection into vulner-

¹This is a conceptually similar approach to recent advancements in NLP and computer vision in using unstructured data for measurement problems, e.g. ([Dell, 2025](#); [Zheng et al., 2023](#); [Ash and Hansen, 2023](#); [Gentzkow et al., 2019](#)). However, to our knowledge, the application of learned latent representations from unstructured relational data, particularly in a dynamic network setting, remains previously unexplored in economics. More on this in Section 3.6.

ability. Other research focuses on the strategic motives behind infiltration, modelling criminal infiltration as an investment decision of OCGs (Arellano-Bover et al., 2024). The study posits that OCGs may establish small-sized firms for criminal activities, infiltrate medium-sized firms post creation, while they infiltrate large firms for clean financial returns and political connections (Arellano-Bover et al., 2024).

A complementary literature uses networks to understand how criminal organizations penetrate and exploit corporate structures. Research on enforcement and asset recovery documents how interventions propagate through economic networks: Calamunci and Drago (2020) study real-time firm outcomes under court-appointed management of confiscated firms. De Luca et al. (2025) looks at changes in firms' ownership structure during the Covid-19 pandemic, finding that prolonged state-mandated business closures restrict changes in ownership, but the effect is weaker in areas with strong mafia presence. The authors interpret this as OCGs infiltrating firms in financial distress (De Luca et al., 2025). Although our aim is fundamentally different from these studies, we expand on their approach by building a time-dynamic network of infiltrated and non-infiltrated firms, in order to systematically test whether the evolving structure of corporate ownership can serve as a leading indicator of infiltration risk.

Since crime is fundamentally a social phenomenon, economic theory has increasingly turned to social network analysis as a powerful framework for understanding and combating it (Lindquist and Zenou, 2019; Ballester et al., 2004). The very architecture of a criminal network is therefore theoretically important, as it can either amplify or dampen criminal activity. This approach allows us to understand the structural roles of nodes—criminals in this case, and help to further characterize and understand criminal activity (Lindquist and Zenou, 2019; Ballester et al., 2004). Moreover, from a statistical modeling perspective, other research frames criminal network structure as the outcome of a fundamental efficiency-security trade-off. This view posits that organizations must create enough connections to operate efficiently while remaining sparse enough to ensure security and avoid detection (Lu et al., 2025). This tension shapes their architecture, leading to unique features such as groups of redundant criminals and a mix of overt and deliberately covert ties (Lu et al., 2025). Our study builds on this foundation by shifting the focus from the internal social dynamics of criminal-only networks to the economic interactions of OCGs within the legal economy. We consider a mixed environment where criminal actors and legitimate firms interact, allowing us to identify the patterns that signal a firm's risk of confiscation—whether it was originally created by an OCG or was

a legitimate enterprise later preyed upon.

Recent advances in predictive modelling underscore the feasibility of anticipating illicit behaviour with administrative data. For example, [Ash et al. \(2025\)](#) develop a machine-learning framework to forecast corruption in Brazilian municipalities, while [Mastrobuoni \(2020\)](#) show how digital information systems enhance police productivity. Applying these predictive techniques to OCG infiltration, [Cariello et al. \(2024\)](#) use a machine learning approach on a unique dataset of Italian firms to detect potentially infiltrated companies. These results collectively motivate our approach, which also focuses on ex-ante risk assessment by forecasting confiscation risk. Our study also expands on this prior work by integrating not only firm fundamentals but also a rich set of network topology measures. While our firm-level only model achieves performance comparable to that reported by [Cariello et al. \(2024\)](#), our integrated model with network features demonstrates substantially superior predictive power, highlighting the significant value added by incorporating network topology.

The existing literature provides a strong foundation by identifying the economic footprints of infiltrated firms, exploring the network effects of anti-mafia interventions, and demonstrating the power of machine learning for forecasting illicit activities. However, these streams of research have largely evolved in parallel. A significant gap remains in leveraging the dynamic evolution of corporate ownership networks as a primary tool for ex-ante risk assessment. Our study bridges this divide by uniting these three approaches. Our network approach systematically tests whether changing topological features can serve as a leading indicator of confiscation. In doing so, we shift the analytical focus from identifying the consequences of infiltration to building a predictive framework that can anticipate it, providing a new lens through which to view and combat organized crime’s encroachment into the legal economy.

In this approach, still a fundamental challenge remains: infiltration itself is latent and never directly observed. Empirical strategies in the existing literature have relied on partial or proxy indicators—such as abnormal financial behavior ([Mirenda et al., 2022](#)), ownership transfers involving persons of interest identified by authorities ([Cariello et al., 2024](#)), or more structural approaches modelling OCG investment decisions ([Arellano-Bover et al., 2024](#)). While these approaches provide valuable insights, they remain limited to observable manifestations of infiltration ex post. By contrast, we also propose a new framework that infers infiltration risk from the latent dynamics of the ownership network. Using temporal graph networks ([Rossi et al., 2020](#)), we learn firm-level embeddings

that summarize the evolution of ownership relations over time, allowing us to capture subtle and unobserved structural regularities—such as recurrent patterns of control, indirect links, and bursts of ownership turnover—that may precede formal detection or confiscation. This approach enables a shift from static, attribute-based classification toward a dynamic, representation-learning perspective on latent criminal infiltration. In future expansions of this work, we aim to validate this framework against existing past contribution the OCG literature, to assess its effectiveness in identifying infiltrated firms beyond those that are judicially confirmed.

In conclusion, this representation-learning framework creates a novel bridge between (i) AI-based measurement of unstructured economic data, (ii) network analysis of firm behavior, and (iii) the economics of organized crime. By introducing embeddings from temporal graph neural networks—a method previously unexplored in economics—we open up a promising avenue for future empirical applications. This methodology offers a new way to represent and analyze dynamic, unstructured network data, holding potential insights for diverse fields from micro to macro and finance. Within the context of this paper, we apply graph representation learning to our unique, large-scale dynamic ownership network rooted in judicially confirmed confiscations. Building on this, we propose a new, statistically rigorous approach to identify latent infiltration risk ex-ante. This advancement is particularly relevant to the state-of-the-art in assessing OCG infiltration, offering a new framework that moves beyond static proxies to capture the dynamic, relational patterns of illicit economic activity.

3.2 Data

This section will briefly describe all the data sources and how they have been used for the study.

Confiscated Firms Data Data on confiscated firms has been provided by the National Agency for Seized and Confiscated Mafia Assets *Agenzia Nazionale per i beni sequestrati e confiscati alla criminalità organizzata* (ANBSC) which reports detailed data on all firms ever confiscated for illicit activities in relation to OCG. This dataset comprises over 6,200 firms confiscated between 2000 to 2024 and includes key variables such as the percentage of confiscated assets, the date of confiscation, the economic sector in which the firm operates, and precise geo-referenced data, including the firm’s address.

Furthermore, the data documents the complete judicial history of the confiscation process, from the initial seizure of assets by authorities –*sequestro*, to the final legal outcome, be it the confirmation of the confiscation through multiple judicial instances or its eventual revocation. Crucially, in contrast to prior research (e.g., [Chircop et al., 2023](#)), our dataset provides an unprecedentedly complete picture of this phenomenon at a national scale, with the number of confiscated firms being an order of magnitude larger.

For the remainder of this paper, the confiscation date refers to the initial judicial seizure of proprietary assets by authorities, which marks the beginning of legal proceedings against a firm suspected of mafia involvement. Figure [A.1](#) in the Appendix displays the number of confiscations over time. We can observe an increase in confiscations in the years 2009 to 2017 and a sharp decline in later years. This is probably due to the lack of processing of more recent events by ANBSC. Figure [A.2](#) displays the spatial distribution of confiscated firms at the municipality-level activity. We can immediately see that there is a high concentration of confiscated entities in Rome and Milan, followed by Naples and Palermo. Overall, there seems to be a higher concentration of confiscated firms in the South of Italy, which is consistent with past literature on mafia presence in Italy ([Arellano-Bover et al., 2024](#); [Pinotti, 2015](#)).

Ownership Network We also collected firm-level data from *InfoCamere*, the Italian company responsible for managing the official business registry. To analyze the economic environment of the confiscated firms, we reconstructed their ownership networks. The network nodes represent either firms (*Persone Giuridiche*, PG) or individuals (*Persone Fisiche*, PF), and the directed edges represent ownership stakes, timestamped at the day level. We mapped the network surrounding each confiscated firm up to two degrees of separation considering *all* ownership stakes² to define a connection. The reconstruction followed a two-step process:

1. **First-Degree Connections:** For each confiscated firm, we first identified its immediate connections. We compiled a list of all its direct shareholders (those who owned it) and all of its direct subsidiaries (those it owned) over a 20-year period, covering the 10 years before and after the confiscation.
2. **Second-Degree Connections:** For each upstream shareholder, we included all other firms in their portfolio, ensuring these connections fell within the 20-year time

²i.e. $> 0\%$

window of the original confiscated firm. Similarly, for each downstream subsidiary, we included all of its owners and any other firms it controlled, again respecting the same temporal window. In cases where a second-degree entity was connected to multiple confiscated firms, we retained only the link associated with the earliest confiscation date to provide a unique temporal anchor for our analysis.

This methodology captures a detailed view of the local ownership structure, revealing both the investors and assets linked to the confiscated entities. The resulting network is a weighted—i.e., ownership percentage—, directed—i.e., ownership direction—heterogeneous³—dynamic graph, spanning from 1993 to 2024.⁴

This network-based approach represents a fundamentally different way to measure the degree of OCG involvement in the legal economy. Rather than attempting to identify potentially infiltrated firms from the entire population (Mirenda et al., 2022; Arellano-Bover et al., 2024), our methodology begins with a solid core of firms whose involvement in criminal activity is judicially confirmed through confiscation. From this reliable anchor of definitively infiltrated entities, we then expand outward, building a targeted network of all connected firms and individuals up to two degrees of separation. The resulting dataset provides a unique view into the immediate economic environment surrounding confirmed OCG activity, enabling us to analyze the firms within this high-risk ecosystem to find predictive signals of infiltration.

Firm-Level Data From *InfoCamere*, we also collect comprehensive details on company registrations, firm characteristics —i.e. number of employees, municipality, industry, and so forth —economic and financial performance as well as balance-sheet data at the yearly level. The two main sources are the business registry (*registro*), with quarterly data covering years from 2001 to 2024, and firm-level company financials (*bilancio*), with yearly data covering the period 2005 to 2024. By integrating the ANBSC confiscation records with the Infocamere ownership and firm-level datasets, we construct a rich, attributed network of yearly snapshots. Figure A.3 displays the distribution across sectors of our sample of firms. We can observe that most confiscated entities operate in construction, real estate, wholesale and manufacturing, with fewer confiscations occurring in

³as it includes both people and firms. In this context people can only be *source* nodes, i.e. they cannot be owned by any other node

⁴Ideally, this network would be benchmarked against a random sample of the Italian firm population ownership network. However, access to the comprehensive ownership data required to construct such a baseline is presently restricted.

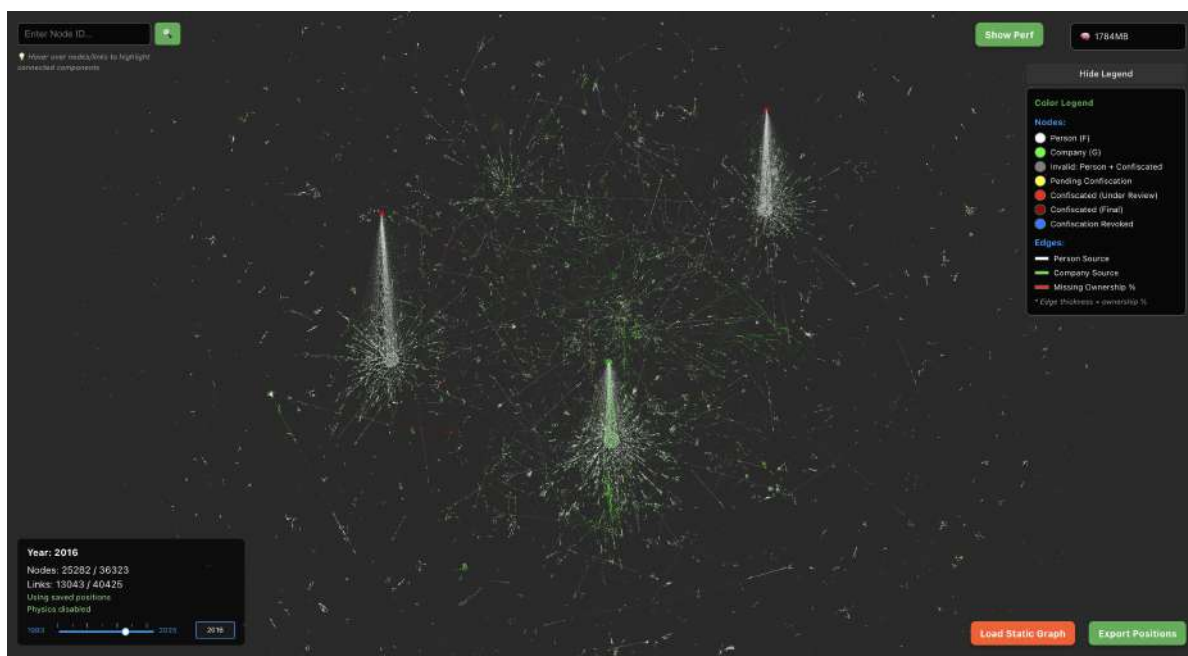
the remaining sectors. Table A.2 in the Appendix provides a detailed list of all firm-level variables used in our analysis, categorized into firm characteristics, financial indicators, balance sheet assets, balance sheet liabilities and equity, and income statement items.

The Network By integrating the confiscation records, ownership data, and firm-level financials, we construct the final, rich, attributed network used in our analysis. To explore this complex dataset, we developed an interactive 3D rendering of the graph.⁵ A key feature of this tool is its dynamic capability; a time slider allows for navigation through yearly snapshots, making it possible to visualize how ownership structures evolve over time. In addition to this dynamic view, the rendering can display the complete static network, which aggregates all connections across the entire period. The static version of the graph counts 40,425 edges and 36,323 nodes, of which 11,202 PFs and 25,121 PGs. Figure A.4a displays the static version of the graph, Figure A.4b displays the static graph’s main component and Table 3.1 displays the static and yearly networks’ descriptive statistics. In both Figures A.4a and A.4b, as well as in Table 3.1, we can observe that there is a giant component accounting for 60% of the nodes and 70% of the edges in the static version of the network. Interestingly, this feature disappears when we look at the yearly snapshots, where the largest component accounts for only 10-15% of the nodes and edges. This stark difference between the long-term aggregate and short-term snapshots may hint at strategic infiltration patterns, such as the use of temporary or rapidly changing ownership structures to evade detection. While a full analysis of this network dynamic is beyond the scope of this paper, it represents a compelling area for future research.

Figures 3.1 and A.5 provide some visualizations of the network’s yearly snapshots. Figure 3.1 illustrates the network’s structure in 2016, which is dominated by three prominent hubs. These hubs are all part of a single, large, weakly connected component that, as detailed in the Appendix (Figure A.5d), encompasses nearly 6,000 nodes, just under 7,000 edges, and over 150 confiscated firms. A closer look at the hubs reveals a striking contrast among their most central nodes: two are confiscated micro-firms—i.e. less than 10 employees (indicated by red nodes), while the third is anchored by a large, non-confiscated financial institution (a green node). Panels A.5a to A.5c display other recurrent network structures. Panel A.5a shows a star-like structure centered on a con-

⁵More on this on Section B in the Appendix

Figure 3.1: Graph Snapshot for 2016



Notes: The figure shows the rendered network for 2016. The legend is displayed on the top right. The 3 most central nodes have been highlighted. Note that each of these hub nodes belong to the same weakly connected component, see [A.5d](#)

fiscated firm, with many direct connections to both individuals and other firms. Here edges are red to indicate that information on the ownership percentage is missing. Panel [A.5b](#) shows a more complex structure, with multiple interconnected firms and individuals, suggesting a web of relationships that could facilitate illicit activities. Panel [A.5c](#) displays a dense cluster of non-confiscated firms with a person (white node) at the center, indicating a potentially influential individual within this network.

It is also important to note that this final dataset yields a non-random network centered on confiscated firms, which presents a distinct challenge for our classification task. One might expect mafia-related firms to have a unique network topology that simplifies identification, although a descriptive analysis suggests this is not the case. As shown in Figures [A.6](#), [A.7](#), [A.8](#) and Table [A.1](#) the distributions of node degrees and centralities show no clear separation across firm types and PFs. It could also be that this sample might actually cluster together similar firms, making the classification task much harder. Still, a definitive assessment would require a more comprehensive graph that also maps the ownership structures of a random sample of firms with no known links to criminal activity.

Table 3.1: Descriptive statistics - Directed Mafia Network

Metric	1993	1998	2003	2008	2013	2015	2017	2019	2021	2023	Whole
Network characteristics:											
Nodes	7,957	11,384	17,819	25,411	25,577	25,366	24,726	23,097	21,340	19,218	36,323
Links	326	3,860	9,271	18,341	13,485	12,686	11,967	9,159	6,988	4,742	40,425
Average Degree	0.041	0.339	0.520	0.722	0.527	0.500	0.484	0.397	0.327	0.247	1.113
Companies (G nodes)	7,757	9,978	14,057	18,189	19,370	19,218	18,881	18,262	17,394	16,197	25,121
Persons (F nodes)	200	1,406	3,762	7,222	6,207	6,148	5,845	4,835	3,946	3,021	11,202
Prop. Companies	0.975	0.876	0.789	0.716	0.757	0.758	0.764	0.791	0.815	0.843	0.692
Confiscated Companies	3,353	3,712	4,380	5,346	5,914	5,957	5,867	5,791	5,676	5,416	6,284
Prop. Confiscated	0.432	0.372	0.312	0.294	0.305	0.310	0.311	0.317	0.326	0.334	0.250
Avg. Confiscation Quota	92.167	91.823	90.909	90.093	90.173	90.246	90.316	90.473	90.566	90.677	90.357
N. of components	7,672 (7631)	8,502 (7517)	10,254 (8544)	10,000 (7779)	13,696 (12074)	14,045 (12694)	14,047 (12762)	14,749 (13930)	14,903 (14355)	14,780 (14472)	4,513 (4576)
Average Clustering Coefficient	0.000 (0.000)	0.002 (0.000)	0.002 (0.000)	0.004 (0.000)	0.003 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.001 (0.000)	0.001 (0.000)	0.009 (0.000)
Overall Clustering Coefficient	0.014 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	0.005 (0.000)	0.004 (0.000)	0.005 (0.000)	0.005 (0.000)	0.004 (0.000)	0.005 (0.000)	0.004 (0.000)
Largest Component characteristics:											
Nodes	18	123	825	4,472	2,615	4,100	5,610	3,340	2,732	2,267	22,071
Links	26	385	1,388	5,728	3,070	4,708	6,442	3,759	3,018	2,449	28,948
Average Degree	1.444	3.130	1.682	1.281	1.174	1.148	1.148	1.125	1.105	1.080	1.312
Companies (G nodes)	13	106	654	2,811	1,486	1,946	2,772	1,334	958	609	15,222
Persons (F nodes)	5	17	171	1,661	1,129	2,154	2,838	2,006	1,774	1,658	6,849
Prop. Companies	0.722	0.862	0.793	0.629	0.568	0.475	0.494	0.399	0.351	0.269	0.690
Confiscated Companies	10	1	42	166	77	87	146	113	124	110	1,013
Prop. Confiscated	0.769	0.009	0.064	0.059	0.052	0.045	0.053	0.085	0.129	0.181	0.067
Fraction of Nodes	0.002 (0.001)	0.011 (0.003)	0.046 (0.068)	0.176 (0.544)	0.102 (0.092)	0.162 (0.030)	0.227 (0.019)	0.145 (0.003)	0.128 (0.002)	0.118 (0.001)	0.608 (0.849)
Average Clustering Coefficient	0.046 (0.000)	0.000 (0.000)	0.011 (0.000)	0.008 (0.000)	0.007 (0.000)	0.006 (0.000)	0.006 (0.000)	0.004 (0.000)	0.002 (0.000)	0.002 (0.000)	0.011 (0.000)
Overall Clustering Coefficient	0.100 (0.000)	0.000 (0.000)	0.001 (0.000)	0.003 (0.000)	0.006 (0.000)	0.006 (0.000)	0.005 (0.000)	0.005 (0.000)	0.004 (0.000)	0.004 (0.000)	0.004 (0.000)

Notes: The table reports network characteristics for the mafia network for selected years. Number in parenthesis are from simulated Erdős-Renyi random networks with the same number of nodes, links and density as the observed network. Time intervals have been chosen to be every 5 years for older snapshots and every two years for more recent ones. The last column refers to the whole network, i.e. the one including all links and nodes from 1993 to 2023.

3.3 Methodology

Our goal is to build a model that estimates the probability, p_i , that a firm i will be confiscated because of OCG infiltration. This prediction is based on the firm's historical data, including its financial fundamentals and network signals, which we collect into a feature set, $\mathbf{X}_i$. The machine learning task is to train a function, f , that takes the firm's

data as input and outputs an estimated confiscation risk:

$$p_i = f(\mathbf{X}_i).$$

We then test this model’s early-warning capabilities by evaluating its performance at different forecast horizons, from the year of the event ($h = 0$) up to three years in advance ($h = 3$). This forecasting window is particularly relevant from a policy perspective, as the large majority of judicial inquiries begin one to two years before the initial asset seizure, i.e. *sequestro*, making our model a potentially valuable tool for early-stage investigations.⁶

3.3.1 Data pre-processing

We assemble a firm-year panel from the dynamic ownership graphs’ node attributes and we expand it with network information. Centrality measures and neighbourhood counts⁷ are computed on yearly directed and undirected graphs—both unweighted and weighted by the ownership stake—and merged back to the nodes and construct a firm-level count of adjacent individuals (PFs). Also, we remove PF nodes, as cannot be confiscated entities.

Because our goal is to forecast confiscation either in the year it occurs (t^c) or in a specified number of years prior ($t^c - h$), every observation must be aligned to a *reference year* representing the forecast horizon. For confiscated firms, this reference year is observed directly. For non-confiscated firms, no such date exists, so we assign a synthetic reference year via propensity-score matching. The reason for this is to make the comparison as honest as possible and prevent the model from learning a spurious temporal rule. If the reference years for all non-confiscated firms were clustered at the end of the sample period, the model could make artificially easy predictions based on the time-period alone, rather than by identifying the underlying financial and network signals present in the years prior.

Specifically, we estimate a scalar proximity score from observable firm characteristics and select the nearest confiscated “twin” for each non-confiscated firm. The matching variables include industry sector (NACE 2-digit level), firm size and geographic location. The choice of these covariates is motivated by their relative stability over time. Unlike

⁶This typical one-to-two-year timeline for inquiries preceding a *sequestro* was confirmed in personal communication with ANBSC officials and data.

⁷i.e. number of direct PFs connections

financial performance metrics which can be highly volatile, a firm’s industry, location, and general size category tend to change infrequently, making them robust variables for identifying a structurally comparable “twin”. The reference year for the non-confiscated firm is then set equal to that of its matched confiscated twin. It is important to stress that this matching procedure is in no way related to recovering treatment effects of being confiscated, it is just needed to obtain a synthetic confiscation year for non-confiscated firms. We also consider an alternative methodology, where we sample confiscation dates for non-confiscated firms from the same empirical distribution function of observed confiscation years. Figure A.9 displays the result of both methodologies.

Then, for a given forecast horizon h and any firm i , we set

$$t_{i,h}^c = t_i^c - h,$$

drop observations without sufficient history, and keep only years $t \leq t_{i,h}^c$. This produces a longitudinal history for every firm that ends at its (shifted) reference year. This enables us to build a dataset that includes both firm-level data⁸ and network measures⁹ for each firm, and their lags. We linearly impute missing values and standardize all numerical features. For the categorical variables we consider two lags (i.e. $K = 2$) and for continuous variables, five lags (i.e. $L = 5$). So for each firm i we obtain the following row-vector of features to feed to the classifier:

$$\mathbf{X}_{i,h} = \left[\underbrace{\mathbf{x}_{i,t_h^c-1}, \dots, \mathbf{x}_{i,t_h^c-L}, \mathbf{c}_{i,t_h^c-1}, \dots, \mathbf{c}_{i,t_h^c-K}}_{\text{Firm level data}}, \underbrace{\text{PN}_{i,t_h^c-1}, \dots, \text{PN}_{i,t_h^c-L}, \text{NC}_{i,t_h^c-1}, \dots, \text{NC}_{i,t_h^c-L}}_{\text{Network data}} \right],$$

where $\mathbf{x}_{i,t}$ is the vector of firm-level financial variables at time t , $\mathbf{c}_{i,t}$ is the vector of categorical firm-level variables at time t , $\text{PN}_{i,t}$ is the number of PFs connected to firm i at time t and $\text{NC}_{i,t}$ are the network centrality measures of firm i at time t .

Finally, we label each observation with a binary target variable, $y_{i,h}$, which equals 1 if the firm is confiscated in year $t_{i,h}^c$ and 0 otherwise. This results in a balanced panel dataset with one firm per row, where each row contains the historical features up to

⁸All variables listed on Table A.2.

⁹All network measures listed on Table A.1 and the count of PFs direct neighbours.

the reference year and the corresponding confiscation label. In order to understand the importance of network features, we run the same classification task with three different feature sets: *Firm-level* data only, *Network* data only, *Firm-level and Network* data. The resulting matrices, with number of observations and average confiscation rate is displayed on Table 3.2. As expected, there is a natural decline in the number of observations as the forecast horizon increases, as well as a mild class imbalance –i.e. confiscated firms are a minority of the total sample –that becomes more pronounced at longer horizons. The next subsection will describe how we address this imbalance in our classification task.

Table 3.2: Descriptive Statistics by Forecast Horizon

	$h = 0$	$h = 1$	$h = 2$	$h = 3$
Observations	17,967	17,484	16,767	16,204
Mean Confiscation	0.2739	0.2697	0.2562	0.2426

Notes: The table reports descriptive statistics for the classification dataset across different forecast horizons. $h = 0$ corresponds to the contemporaneous dataset, i.e. $t_h^c = t^c$, while $h > 0$ indicates the number of years ahead for the prediction task. The number of observations decreases for higher forecast horizons due to data availability constraints.

3.3.2 Classification

To build our predictive model, we employ a supervised machine learning framework designed to distinguish between confiscated and non-confiscated firms. Similarly to [Ash et al. \(2025\)](#), we test four distinct classification algorithms: Logistic Regression, Random Forest ([Breiman, 2001](#)) and two gradient boosting algorithms, LightGBM ([Ke et al., 2017](#)) and XGBoost ([Chen and Guestrin, 2016](#)). These models were chosen to span a range of complexity and flexibility. We start with Logistic Regression, which assumes a linear relationship between the predictors and the outcome, providing a strong baseline.¹⁰ We then evaluate more powerful tree-based ensemble models—Random Forest, LightGBM, and XGBoost—which relax this linearity assumption and are adept at capturing complex, non-linear relationships and interactions within the data¹¹. The performance

¹⁰To be precise, while the sigmoid function means the model’s probability output is a non-linear function of the features, Logistic Regression is considered a linear classifier because its decision boundary is linear. That is, it separates the classes in the feature space using a hyperplane (e.g., a line in two dimensions).

¹¹see Section C for a more in-depth discussion.

of these four models is systematically compared across the three distinct feature sets for each different forecast horizon.

Our experimental design follows a rigorous pipeline for data preparation, training, and evaluation. Since confiscations occur in different points in time, arranging the data in a random train/test split could bias the results, i.e. temporal *data leakage*.¹² To address this, we employ a chronological split, training our models on all firm-year observations up to a specific cut-off date. We chose 2018 as cutoff date, to provide sufficient training examples to the algorithms –i.e. $t_i^c \leq 2018 - h$, but we also consider 2016 as a cutoff date as a robustness check. We then evaluate the models’ performance exclusively on observations from subsequent years. The train-test split is displayed on Table A.9. Moreover, to address the inherent class imbalance—where non-confiscated firms outnumber confiscated ones—we incorporate the Synthetic Minority Over-sampling TEchnique (SMOTE) (Chawla et al., 2002). Applied only to the training data, SMOTE generates new synthetic examples of the minority (confiscated) class, creating a balanced dataset that prevents the classifiers from developing a bias towards the majority class.

The final stage involves training the models and evaluating their out-of-sample performance. For each combination of model and feature set, we perform hyperparameter tuning using a randomized search with 5-fold cross-validation and 10 random draws of the parameter space. This process systematically explores different model configurations to identify the one that maximizes the ROC-AUC, a key metric that evaluates a model’s ability to discriminate between positive and negative classes across all possible classification thresholds. Once the best-performing model is identified for each scenario, its predictive power is conclusively assessed on the unseen test set. We report a comprehensive suite of metrics: Accuracy (the overall correct predictions), Precision (the proportion of correctly identified confiscated firms among all firms flagged as such), Recall (the proportion of actual confiscated firms that were correctly identified), the F1-score (the harmonic mean of Precision and Recall), and the ROC-AUC score. This entire procedure is independently repeated for each forecast horizon, from $h = 0$ to $h = 3$, allowing us to evaluate the model’s early-warning capabilities at predicting confiscation up to three years in advance.

¹²Although there might not be data leakage of the confiscation of firm i on the data concerning firm j , the aim of this classification exercise is to make the problem as close as possible as the one that law enforcement authorities have to deal with.

3.4 Results

This section will evaluate the model performance across the three feature sets and then will provide a more detailed breakdown of the feature contributions for predictions.

3.4.1 Main Results

Table 3.3 reports the results for forecasting confiscation in the year it occurs ($h = 0$), while Table 3.4 evaluates the models' early-warning capability one year in advance ($h = 1$). The findings are presented across three distinct feature sets—Firm-Level Only, Network Only, and the combined Firm + Network—to isolate the predictive contribution of each data source. As we evaluate these results, it is crucial to remember that our model predicts *confiscation*, not *infiltration* directly. While a confiscation event provides a certain label for an infiltrated firm, the absence of confiscation does not guarantee the absence of infiltration. This implies that some of the model's false positives—firms predicted to be at high risk that are not in our confiscated sample—could represent successful identifications of firms that are truly infiltrated but remain undetected by law enforcement. Consequently, the measured accuracy of a model displaying high recall may be biased downwards.

A primary finding, consistent across all models and time horizons, is the substantial performance gain achieved by combining firm-level and network features (Panel C in both tables). This combined effect is evident across all metrics. For instance, in the contemporaneous prediction task (Table 3.3), the XGBoost model's AUC score increases from 0.904 (firm-level only) and 0.796 (network only) to an exceptional 0.975 when both feature sets are used. Similarly, the F1-score, which balances precision and recall, jumps from 0.783 to 0.911, indicating a far more reliable classification. In the appendix Tables A.3 to A.4 show the performance results on the test-set for longer forecast horizons, where we can observe a similar pattern as in Table 3.3 and 3.4. Most importantly, model performance is similar for different forecast horizons, signalling that the model is robust over time and effective as an early-warning system, capable of identifying at-risk firms several years in advance. Except for a very low recall score, the Firm-Level Only model achieves performance levels that are comparable to, or worse than, those reported in Cariello et al. (2024), suggesting that our non-random sample of Italian firms is suitable for meaningful comparison with earlier work. This is not the case for the Firm

Table 3.3: Model Performance Comparison Across Feature Sets, $h = 0$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.830	0.809	0.799	0.793
Precision	0.950	0.991	0.985	0.979
Recall	0.742	0.674	0.660	0.652
F1	0.833	0.802	0.791	0.783
AUC	0.893	0.902	0.908	0.904
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.539	0.765	0.758	0.761
Precision	0.795	0.725	0.723	0.723
Recall	0.266	0.949	0.938	0.947
F1	0.398	0.822	0.816	0.820
AUC	0.632	0.841	0.858	0.796
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.871	0.905	0.888	0.904
Precision	0.958	0.976	0.935	0.971
Recall	0.811	0.855	0.865	0.857
F1	0.878	0.912	0.899	0.911
AUC	0.958	0.962	0.960	0.975

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table 3.4: Model Performance Comparison Across Feature Sets, $h = 1$

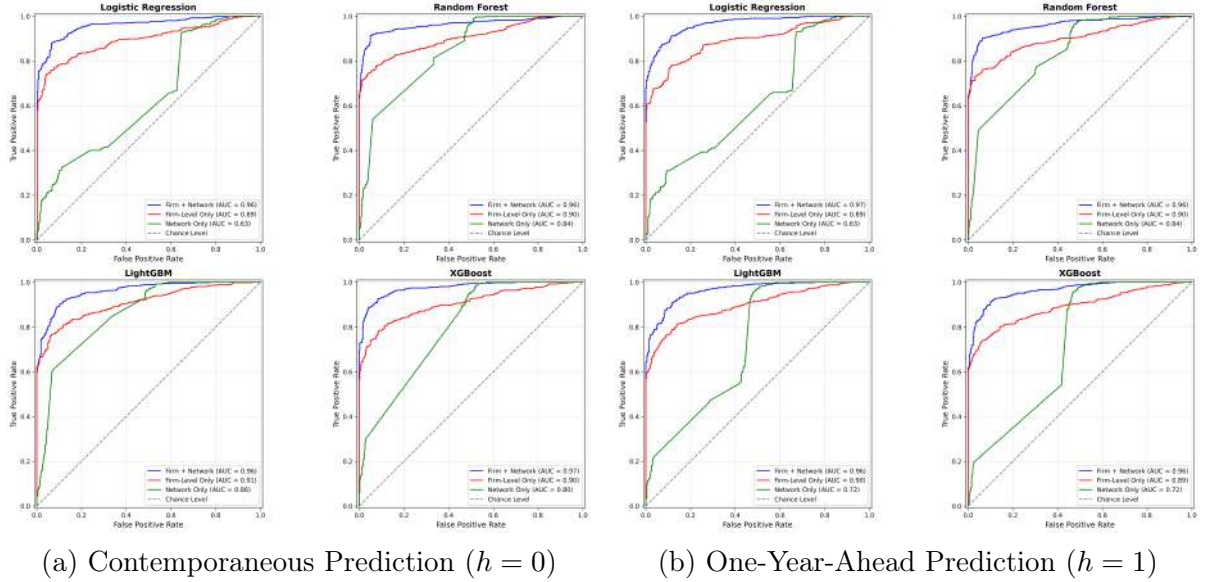
<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.796	0.806	0.796	0.795
Precision	0.924	0.988	0.968	0.982
Recall	0.695	0.663	0.659	0.648
F1	0.793	0.794	0.785	0.781
AUC	0.890	0.900	0.896	0.890
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.543	0.759	0.765	0.769
Precision	0.786	0.723	0.715	0.721
Recall	0.258	0.928	0.971	0.961
F1	0.389	0.812	0.823	0.824
AUC	0.628	0.835	0.721	0.718
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.873	0.889	0.893	0.901
Precision	0.967	0.977	0.898	0.943
Recall	0.802	0.822	0.914	0.877
F1	0.877	0.893	0.906	0.909
AUC	0.966	0.961	0.963	0.963

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

and Network model, which consistently outperforms the results in [Cariello et al. \(2024\)](#), highlighting the significant value added by incorporating network topology.

While the combined feature set is clearly superior, a comparison of the models using isolated feature sets reveals their complementary predictive roles. Models trained exclusively on firm-level data (Panel A) consistently achieve higher precision, suggesting that financial indicators, when triggered, are strong signals of risk. Conversely, models relying solely on network data (Panel B) demonstrate markedly higher recall. For example, at $h = 1$, the Random Forest, LightGBM, and XGBoost models achieve recall scores of 0.928, 0.971, and 0.961 respectively with network data, compared to scores of 0.663, 0.659, and 0.648 with firm-level data. This suggests that ownership network topology is exceptionally effective at flagging the broader set of potentially infiltrated firms. This pattern is also evident in Tables A.3 to A.4, where recall is highest using only network data.

Figure 3.2: Model Performance ROC Curves by Forecast Horizon



Notes: ROC-AUC Curves for each classification model across three feature sets. The left panel (a) shows performance for contemporaneous predictions ($h = 0$), while the right panel (b) shows performance for one-year-ahead predictions ($h = 1$). The blue line represents the combined Network and Firm-level feature set, the green line represents only network data, and the red line represents only firm-level data. All the plots refer to the best performing hyperparameter configuration.

These divergent outcomes highlight the trade-off with significant strategic implications for law enforcement. A model relying solely on financial indicators, while precise, operates with lower recall, risking that many infiltrated firms go undetected. Conversely,

a network-only approach, despite its excellent recall, would generate a high volume of false positives, hampering the efficient allocation of investigative resources. The superior performance of the combined model demonstrates its ability to navigate this trade-off effectively. By integrating both data sources, our framework achieves an efficient balance, maintaining high recall without unduly sacrificing precision. This approach is crucial for operational effectiveness, as it enables authorities to cast a wide yet targeted net, ensuring that potential threats are not overlooked while simultaneously optimizing the deployment of limited enforcement efforts. This is also clearly shown in the confusion matrices displayed in Figures A.10 and A.11: the off-diagonal elements –i.e. Type I and Type II errors, are clearly minimized with the combined model.

Figure 3.2a and Figure 3.7a visualize these findings by plotting the ROC curves for the contemporaneous ($h = 0$) and one-year-ahead ($h = 1$) predictions, respectively. Across all four models, the ROC curve for the combined Firm + Network feature set (blue line) consistently dominates, positioned closest to the ideal top-left corner. In statistical terms, the curve for the combined model stochastically dominates the others, meaning it is unequivocally superior at distinguishing between classes regardless of the chosen classification threshold. This visually confirms its superior discriminative power, achieving a high True Positive Rate (recall) for any given False Positive Rate.

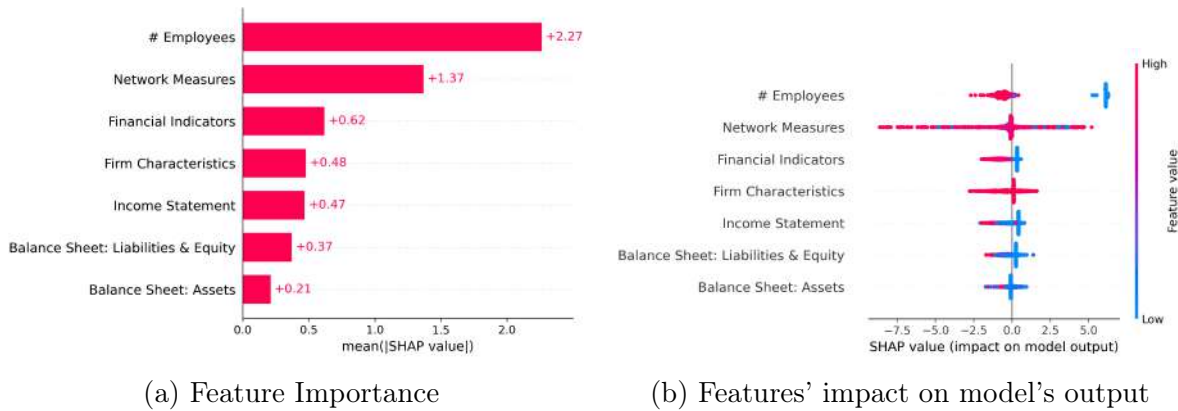
3.4.2 SHAP Values

To move beyond aggregate performance metrics and understand which characteristics drive confiscation risk, we conduct a feature contribution analysis using SHAP (SHapley Additive exPlanations) values. Originating from cooperative game theory, SHAP provides a unified framework for interpreting model predictions by assigning each feature an importance value for a particular prediction (Lundberg et al., 2020). For a given observation, each feature’s SHAP value quantifies its contribution to pushing the model’s output from the baseline (the average prediction over the training set) to its final value. Crucially, this analysis does not explain the real-world causes of confiscation; rather, it reveals the internal logic the model uses to make its predictions. This approach allows for both local interpretability—explaining why an individual firm received its risk score—and global interpretability, as we can aggregate the values across the entire dataset to assess the overall impact and directionality of each feature. For the remainder of this subsection, we will be focussing on the best performing models, i.e.

XGBoost and LightGBM. Figures A.16 and A.17 in the Appendix report the results for the latter. For more details on SHAP, see Section D in the Appendix.

The magnitude of a SHAP value has a direct quantitative interpretation based on its additive property on the model’s output scale, which for these classifiers is log-odds, i.e. $\log\left(\frac{p}{1-p}\right)$. For instance, a SHAP value of 0.5 for a given feature indicates that this feature increases the log-odds of confiscation by 0.5, holding all other features constant. This translates to an increase in the odds of confiscation by a factor of $e^{0.5} \approx 1.65$, or a 65% increase in odds, assuming all other factors remain unchanged. Conversely, a negative SHAP value would indicate a decrease in the log-odds, thus reducing the odds of confiscation.

Figure 3.3: XGBoost SHAP Values by Feature, $h = 0$



Notes: On panel (a): SHAP Value for each aggregated category of feature. The x-axis the mean absolute SHAP Value, i.e. average impact of each feature group on the magnitude of the model’s prediction. Panel (b): each point on the plot corresponds to a single observation for a given feature. The position on the x-axis is the SHAP value, representing the impact of that feature on the model’s prediction for that observation, if $x > 0$ observation is more likely to be confiscated. The color indicates feature value, from low (blue) to high (red).

Figure 3.3 presents the mean absolute SHAP values for each group of features, revealing a clear and consistent hierarchy of feature importance. The quantitative impact of the top predictors is substantial; the number of employees is the single most influential feature set, with an average absolute SHAP value of 2.27, which translates to a change in the odds of confiscation by a factor of approximately 9.7 ($e^{2.27}$). The network measures follow as the second most important group, with a value of 1.37, corresponding to a change in the odds by a factor of nearly 4 ($e^{1.37}$).

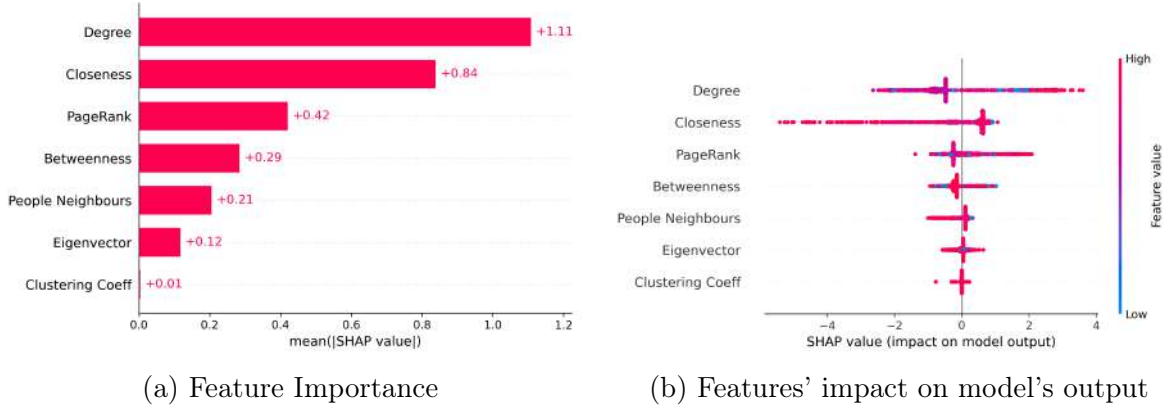
On the right panel, Figure 3.3b provides a more granular view of how individual features within these groups influence the model’s predictions. Each point represents a

single firm’s observation for a specific feature, with its position on the x-axis indicating the SHAP value (the feature’s contribution to the log-odds of confiscation) and its color representing the actual feature value (from low in blue to high in red). The figure shows a clear separation: smaller firms consistently contribute positively to a confiscation prediction, whereas larger firms do the opposite. This empirical finding resonates with the work of [Arellano-Bover et al. \(2024\)](#), where they argue that smaller firms are more likely to be directly involved in criminal activities, thus making their discovery by authorities more probable.

The emergence of network measures as the second most important feature set is also a key finding. This result strengthens our conclusions, as it suggests the model’s high performance is not merely an artifact of our sampling strategy; if the network’s construction around confiscated firms were creating an artificially easy task, network features would likely have been the single dominant predictor, which is not what we observe. Following firm size and network metrics, financial indicators emerge as another crucial set of predictors. The beeswarm plot again provides insight, demonstrating that firms exhibiting signs of financial distress—such as low profitability (e.g., a low ROE) and poor liquidity (e.g., cash and cash equivalents)—have a higher probability of being confiscated. This observation is consistent with the mechanism described by [Le Moglie and Sorrenti \(2022\)](#), whereby firms might turn to investments from organized crime groups (OCGs) when legitimate avenues for borrowing are exhausted.

From Figures 3.3 and 3.4 we can see that network topology measures are also important predictors. We see that Degree (both node’s in and out degree and degree centrality) changes the odds by a factor of 3, followed by Closeness which prompts a change in the odds of a factor of 2, and all the other measures. However, both from 3.3b and 3.4b there is no clear pattern pushing the probability on either side given feature values. The mix of blue and red dots in both Figures suggests that the model has learned nuanced and non-linear relationships, without a clear separation. We can only observe the distinct hierarchy emerging from Figure 3.4a. The relevant SHAP plots for the remaining forecast horizons, i.e. $h > 0$, are reported in Figures A.14 and A.15. For every event horizon, the results are largely left unchanged, with similar patterns in both feature importance and values, although minor differences can be observed for the LightGBM model (see A.16 and A.17). In sum, although a large impact on the odds from a given feature does not, in itself, guarantee better performance, the outstanding accuracy of our model on the test set validates its decision-making process. Therefore, we can conclude that the

Figure 3.4: XGBoost SHAP Values by Network Feature, $h = 0$



Notes: On panel (a): SHAP Value for each aggregated category of feature. The x-axis shows the mean absolute SHAP Value, i.e. average impact of each feature group on the model's prediction. Panel (b): each point on the plot corresponds to a single observation for a given feature. The position on the x-axis is the SHAP value, representing the impact of that feature on the model's prediction for that observation, if $x > 0$ observation is more likely to be confiscated. The color indicates feature value, from low (blue) to high (red).

features identified by SHAP as most influential are indeed reliable and powerful signals of confiscation risk.

3.5 Robustness

Alternative Matching To ensure that our results are not driven by the specific matching strategy employed to assign synthetic confiscation years to non-confiscated firms, we re-evaluate our models using an alternative approach. Instead of propensity-score matching, we randomly sample confiscation years for non-confiscated firms from the empirical distribution of observed confiscation years, as discussed in Section 3.3. This method maintains the temporal distribution of confiscation events without relying on firm characteristics for matching. The results, presented in Tables A.5 to A.8, confirm the robustness of our main findings. The combined feature set continues to outperform individual sets, and the complementary roles of network and firm-level data remain evident.

Removing Lags from the Features Another question concerning our modeling approach is the role of historical data. To assess this, we compare the performance of our models with lagged features (Table 3.3) against models trained only on contemporaneous

data, i.e. Table A.9. We mostly consider the F1 and AUC score as they provide a comprehensive picture of a model’s classification ability. This is because they collectively assess two critical aspects of performance: the model’s underlying power to distinguish between positive and negative cases (AUC) and its practical success at balancing the trade-off between finding all true cases and avoiding false alarms (F1-score).

The inclusion of lags consistently improves performance, particularly for the more complex tree-based models. For instance, the XGBoost model’s AUC score in the combined feature set (Panel C) increases from 0.945 to 0.975 with the addition of lags, and its F1-score rises from 0.880 to 0.911. While the largest performance gains are observed in the Network Only feature set (Panel B), the difference is smallest in the Firm-Level Only models (Panel A), suggesting that current-year financial distress is already a very powerful stand-alone predictor. This pattern can be observed also for longer prediction horizons, i.e. Tables A.10 to A.12

Interestingly, the simpler Logistic Regression model does not benefit from the added complexity of lagged features. This is likely because its rigid linear decision boundary struggles to effectively map the high-dimensional space created by the historical variables to an accurate confiscation probability. Without the complexity of lags, the model can find a clearer linear separation between the classes, resulting in a more stable performance. All in all, these results suggest that more complex models, particularly gradient boosting algorithms, are better suited to capturing the subtle temporal dynamics and non-linear patterns present in the data, leveraging historical information to achieve a higher degree of predictive accuracy.

Alternative Cutoff Date To further validate the robustness of our findings, we re-evaluate our models using an earlier cutoff date of 2016 for the train-test split, as opposed to the original 2018. This adjustment allows us to test the models’ performance with a longer out-of-sample period, thereby assessing their stability and generalizability over time. Figure A.18 illustrates the distribution of actual and synthetic confiscation dates under this new cutoff. The results, presented in Tables A.13 to A.16, reaffirm the robustness of our main conclusions: the combined feature set continues to deliver superior performance across all metrics. Notably, while the integrated model remains superior, the distinct precision-recall trade-off observed between the firm-level and network-only feature sets in our main analysis is less evident with this earlier cutoff date. The network data feature set still displays higher recall, but now the combined model also achieves

the highest precision.

3.6 Going Forward: Graph Representation Learning for Infiltration Risk

Because infiltration is fundamentally a latent process—rarely directly observed and often detected only after its effects have materialized—we require a method capable of extracting its hidden relational signatures from the data. Our goal is therefore to build a representation of firms’ behavior that captures these unobservable dynamics as they unfold within the ownership network. This section describes the methodology used to generate firm-level embeddings from the dynamic ownership network. The objective of this approach is to obtain low-dimensional representations that summarize each firm’s evolving position and behavior within the network, while preserving the temporal and structural complexity of ownership relations. To do so, we employ a Temporal Graph Network (TGN) (Rossi et al., 2020), a neural architecture designed for dynamic graphs in which both the links and node attributes evolve over time. TGNs are particularly well-suited to this setting, as they process ownership changes in continuous time and use the exact timestamps of these events to model how information, control, and influence propagate through the network. In turn, the resulting embeddings offer a principled tool to examine when infiltration is likely to have occurred, by capturing relational behaviors that precede and accompany the emergence of OCG influence.

Our main objective is to build a TGN architecture that can learn *ownership embeddings*—low-dimensional vectors that encode each firm’s structural role, relational history, and exposure to other entities in the ownership network. For the model to learn meaningful representations, we supervise it on the three fundamental behaviors that fully characterize edge dynamics in ownership data: (i) *link creation* (a new ownership relation appears), (ii) *link update* (an existing ownership share changes), and (iii) *link removal* (an ownership relation disappears). Each observed event at time t is a tuple $(u, v, t, \text{type}, y_t)$, involving a source node u , i.e. investing firm or person, a target firm node d , the event type $\in \{\text{creation, update, deletion}\}$ and when applicable the ownership stake or its change from the most recent decision.

This design choice is central to our methodological approach. A model trained to predict confiscation would tend to internalize the statistical regularities of law enforcement

and judicial detection—learning *where confiscations occur*, not *how ownership structures evolve*. In contrast, our multi-head temporal architecture forces the network to learn the fundamental generative rules of the corporate control system itself. Through the joint supervision on link creation, update, and removal, the model must infer the latent mechanisms that govern *who* invests in *whom*, *under what temporal and structural conditions*, and *how* these relationships intensify or dissolve over time. In doing so, it learns to represent the endogenous dynamics of capital ties and firm interdependence—dynamics that precede and often underlie legal outcomes such as confiscation. The resulting ownership embeddings encode each entity’s evolving relational identity: how it participates in the flow of ownership, accumulates or disperses control, and interacts with other actors in its economic environment. Once trained, these embeddings serve as general-purpose descriptors for downstream analyses, including econometric modeling of infiltration risk, early-warning systems for anomalous ownership behavior, and structural studies of firm evolution.

Data Processing The dataset combines two sources of temporal information: (i) daily-stamped ownership relations between firms (Persone Giuridiche, PG) and individuals (Persone Fisiche, PF), and (ii) yearly financial statements for firms. To make these heterogeneous data compatible with a temporal graph model, we convert them into a single chronologically ordered stream of timestamped events.

We first define the universe of nodes by including all entities that appear in the ownership network and excluding isolates that never participate in a link. Each ownership relation—defined by a start and end date—is represented as two distinct events: a creation event at its start and a deletion event at its termination.¹³ This decomposition allows the model to track the life cycle of ownership ties. To stabilize the input features, we apply a $\log(1 + x)$ transformation to ownership shares and share capital, and we convert historical values from Italian Lira to Euros for temporal comparability.¹⁴ Firm-level financial statements are treated as node update events. Each firm generates one such event per year, represented as a self-loop (an edge connecting the node to itself) timestamped on December 31st of the reporting year. These self-loops inject updated firm-specific information—financial indicators and judicial status—into the model. Firms

¹³When an end date for a specific pair correspond to a start date for the same pair, we consider this a node update if something changes, i.e. stake or total equity of destination node, otherwise we consider it a normal continuation of the relation, i.e. no event

¹⁴1 euro $\approx$ 1936.27 Italian Lira.

that are confiscated undergo a discrete feature change at the moment of seizure, when their status variable switches from “not confiscated” to the relevant judicial stage, and their “confiscated share” variable updates from 0% to the seized percentage. Individuals (PFs), lacking such information, never generate node events and evolve purely through their relational activity. This asymmetry is handled with a firm-level mask ensuring that node features are updated only for firms.

Finally, we merge the link and node events into a single temporal stream. A unified message vector includes all event features, together with a binary flag indicating whether the event is a node event (1) or a link event (0). Link events can be of three types, creation, update, or deletion. When multiple events share the same timestamp, we impose a deterministic order: node deletions first, followed by link updates and then link creations. The resulting sequence is split into training, validation, and test sets using chronological cutoffs—training from 1993 to end of 2015, validation from end of 2015 to beginning of 2018 and testing from 2019 to 2024. Figure A.19 in the Appendix displays the number of total events, i.e. node and edge events combined, while Figure A.20 displays the number of edge events only, broken down by type. Creations dominate the event stream, followed by updates and then deletions. Table A.18 breaks down the events by the data splits.

Architecture and Training The Temporal Graph Network (TGN) (Rossi et al., 2020) models the ownership network as a continuous stream of timestamped events, each reflecting a structural or attribute-level change in the system, as well as a message passing framework that maintains a dynamic memory state for each node, evolving over time to capture historical interactions. Every firm or individual is represented by a node equipped with a dynamic memory state $m_u(t)$ —or *hidden state*, that summarizes its past interactions and evolving structural role. As the network evolves, these memory states are iteratively updated, enabling the model to capture both short-term fluctuations and long-term relational dependencies.

When an ownership event occurs between nodes u and v at time t with $u \neq v$, the model processes it following the operational flow of our TGN architecture.¹⁵ At the beginning, the model performs a memory read, retrieving the states $m_u(t^-)$ and $m_v(t^-)$ that summarize all information available up to—but not including—the present event.

¹⁵see Figure E.26 for a graphical representation of the model’s architecture.

For example, if u has recently accumulated several participations, this pattern is encoded in $m_u(t^-)$ and influences how the model interprets a new ownership transfer at time t .

Using these memory states, the encoder reconstructs the event’s temporal context by sampling each node’s historical neighborhood $\mathcal{N}_u(t)$ and $\mathcal{N}_v(t)$. This neighborhood consists of past interaction partners together with the times at which those interactions occurred. For instance, if v changed ownership multiple times in the preceding months, the sampler retrieves the firms involved in those transactions along with their timestamps, allowing the encoder to recognize that v belongs to a rapidly rotating corporate structure. For each neighbor $v' \in \mathcal{N}_u(t)$ with past interaction time $t' < t$, the model collects the neighbor’s memory $m_{v'}(t')$ and the relative time encoding $\Delta t_{uv'} = t - t'$, which expresses how far in the past the interaction occurred. The temporal encoding follows the functional time encoder introduced by Xu et al. (2020), which maps elapsed time Δt into a continuous vector using learnable periodic projections of the form $\cos(W \Delta t + b)$, where W and b are trainable frequency and phase parameters.¹⁶ This representation captures temporal recency and multi-scale decay patterns while remaining invariant to absolute timestamps.

These elements are passed to a Transformer-based graph convolution (**TransformerConv**), which performs multi-head attention¹⁷ over the sampled temporal neighbors. The mechanism weights each neighbor by structural relevance and temporal proximity—recent or strategically similar neighbors receive higher attention than distant or unrelated ones. In this way, if u has historically acquired firms of a similar type to v , the model will highlight those neighbors in the attention computation. The resulting embeddings $z_u(t)$ and $z_v(t)$ integrate each node’s prior memory with the temporal patterns embedded in its historical neighborhood.

Three prediction heads operate on these dynamic embeddings. The **Link Prediction Head** estimates whether an ownership relation $u \rightarrow v$ exists at time t and is trained with binary cross-entropy using typed negative samples.¹⁸ The **Level Regression Head** predicts the log-ownership share s_{uv} for active links using an ℓ_1 loss, while the **Delta Regression Head** predicts the change in log-ownership share Δs_{uv} using a Huber loss.

¹⁶In the original formulation by Xu et al. (2020), the time encoder uses both sine and cosine components. In our implementation, we follow the PyG TGN Memory module, which adopts a simplified formulation

¹⁷In our architecture we employ two attention heads

¹⁸For each observed edge (u, v, t) , ten negatives are generated by pairing u with destinations v' that have never been connected to u up to time t .

Creation events primarily supervise the link-prediction head, ownership updates activate both regression heads, and removals supervise only the delta head.

Task contributions are balanced by two coefficients, γ and η , regulating respectively the relative weight of regression versus link prediction and the trade-off between level and delta regression. We set $\eta = 0.5$, while γ is computed dynamically per batch:

$$\gamma = \min\left(\frac{n_{\text{upd}} + n_{\text{del}}}{n_{\text{cre}} + n_{\text{upd}} + n_{\text{del}}}, 0.5\right),$$

where $n_{\text{cre}}, n_{\text{upd}}, n_{\text{del}}$ denote counts of creation, update, and deletion events. This increases the influence of regression tasks when modifications or terminations are more common, while guaranteeing that link prediction remains dominant when new connections drive the dynamics.

The full training objective is a convex combination of supervised components, augmented with two auxiliary regularizers:

$$\mathcal{L}_{\text{total}} = (1 - \gamma) \mathcal{L}_{\text{link}} + \gamma[(1 - \eta) \mathcal{L}_{\text{level}} + \eta \mathcal{L}_{\Delta}] + \lambda_{\text{NC}} \mathcal{L}_{\text{Barlow}} + \lambda_{\text{TMP}} \mathcal{L}_{\text{temporal}}.$$

Here, $\mathcal{L}_{\text{Barlow}}$ penalizes redundancy between embedding dimensions (Zbontar et al., 2021), and $\mathcal{L}_{\text{temporal}}$ enforces smoothness in node trajectories across time (Trivedi et al., 2019). Their weights are set to $\lambda_{\text{NC}} = 0.1$ and $\lambda_{\text{TMP}} = 0.05$. This formulation enables the network to jointly learn the structure, intensity, and evolution of ownership relations while maintaining stable, temporally coherent embeddings.

Only after predictions and losses are computed does the model construct the message associated with the current event. This delayed construction is intentional: it prevents information from propagating within a single mini-batch and ensures strict temporal causality. The message encodes (i) the current memories of u and v , (ii) the event attributes such as the ownership share or its temporal change, and (iii) the relative time encoding capturing the elapsed time since each node’s last update. This message is routed by a *Block Message Module* dedicated to edge events and is then used to update the memories of u and v via a GRU transition, preparing them for all subsequent events in the stream.

Self-loop events ($u = v$), corresponding to PG firm-level attribute updates, follow a simplified path. They do not trigger neighborhood sampling or attention; instead, they directly refresh the node’s memory through a GRU update. For example, a balance-sheet

update modifies only $m_u(t)$ without introducing any structural signal. This separation ensures that the model distinguishes between changes in the ownership graph and updates to internal firm attributes.

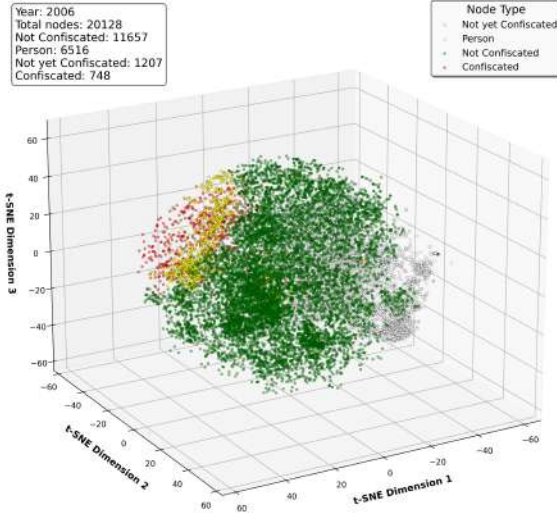
Training proceeds for 30 epochs using the Adam optimizer, supplemented with gradient clipping and adaptive learning-rate scheduling. Model parameters are updated after each batch, and node memories are carried forward across batches to preserve temporal continuity. To prevent memory drift, firm-level attributes are periodically refreshed via self-loop (PG) events, ensuring that structural and attribute dynamics remain synchronized. Evaluation follows the same event-driven procedure. Link prediction performance is assessed using the Area Under the ROC Curve (AUC) and Average Precision (AP), while the regression head performances are evaluated with Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). All metrics are computed in a strictly time-respecting fashion, with test events occurring after all training and validation intervals. After convergence, we extract the time-indexed embeddings $z_t(\cdot)$ from the final memory state, obtaining compact representations that summarize each firm’s evolving relational behavior and exposure within the ownership network. Figure E.26 illustrates the full event flow, message routing, and update cycle.

The results in Table A.19 show that both TGN variants achieve strong link-prediction performance (AUC ≈ 0.99 , AP ≈ 0.89), indicating that the model captures key temporal and structural regularities of ownership formation. For the regression tasks, absolute stake levels (creations+updates) exhibit MAE and RMSE in the 0.64–0.94 range on the log scale, which is consistent with the large dispersion and skewness of ownership shares—from minority holdings to full control. Delta regression (updates+deletions) performs more modestly, with $\text{MAE}_\Delta \approx 0.72$ and $\text{RMSE}_\Delta \approx 1.22$, reflecting the sparsity and discontinuity of stake changes, where many events correspond to little variation and non-zero deltas often involve large divestments—see Figure A.21. It is important to note that these predictive metrics serve only as indirect indicators of representation quality, since the purpose of this evaluation is not to deploy the model as an operational predictor. Instead, our goal is to verify that the learned embeddings capture the latent structure and temporal dependencies of the ownership network. Given that, the results suggest the model has learned meaningful representations. Henceforth, we will be only using node embeddings from the trained TGN as the low-dimensional, time-varying inputs to validate that the learned latent representations are indeed capturing signals of mafia infiltration.

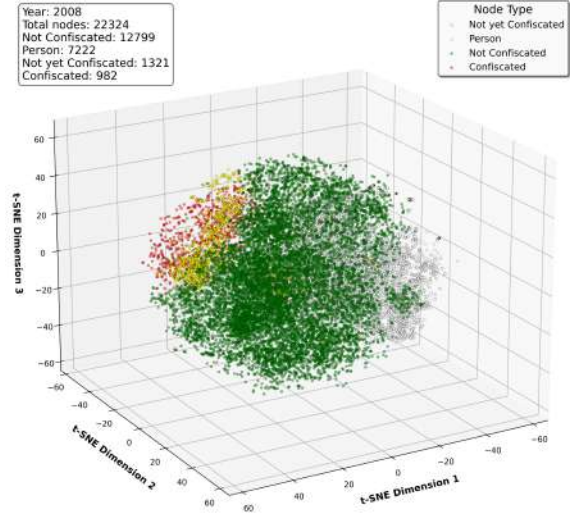
Embedding Extraction After training, we extract yearly snapshots of node embeddings to serve as features for downstream tasks. A naive extraction approach, such as querying the model at arbitrary dates, risks data leakage. To ensure strict temporal consistency, we employ a robust extraction procedure. We first reset the trained model’s temporal state and then stream the *entire* chronological sequence of events through it in evaluation mode. This process updates the internal memory of each node exactly as it would have evolved during training, but without gradient updates. We define a grid of calendar timestamps corresponding to January 1st for each year. As the event stream’s processed time crosses one of these boundaries, we pause and query the model for the current state of all node embeddings (or their memory vectors) at that precise reference time. This method guarantees that the embedding for any given year T strictly reflects information from events occurring *before* January 1st of year T , preventing any leakage from future events. The extracted base embeddings have 64 dimensions.¹⁹ These yearly snapshots form the basis for all subsequent analyses of infiltration risk and structural dynamics within the ownership network.

Figure 3.5 presents a t-SNE visualization of the 64-dimensional base embeddings for all nodes across four time points: 2006, 2008, 2010, and 2012. The plots reveal a strong and persistent structural separation learned by the model. A clear distinction emerges between individuals (Persone Fisiche, white circles), which form a large, distinct cluster at the bottom, and firms (Persone Giuridiche, green and red dots), which group together in the upper mass. This demonstrates the model’s ability to differentiate node types based purely on their relational and attribute event patterns. Furthermore, within the firm cluster, nodes are not randomly mixed. Confiscated firms (red) show a tendency to form localized sub-clusters, distinct from the non-confiscated firms (green), implying the model captures shared relational signatures. Beyond this static separation, the comparison across the four snapshots highlights a dynamic aspect. There appears to be a gradual drift in the embedding space over time, with the main cluster of firms seemingly moving from the north-west region in 2010 towards the east and south-east by 2016. We will quantitatively validate all these visual findings in the following sections to further understand the specific economic and structural features these embeddings are capturing.

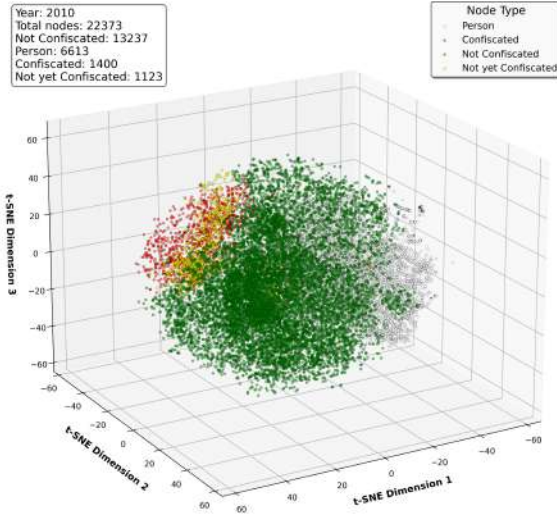
¹⁹For models using the two-tower architecture, we additionally save the separate 64-dimensional projections for source and destination roles, i.e. a total of $64 \times 3 = 192$ dimensions.



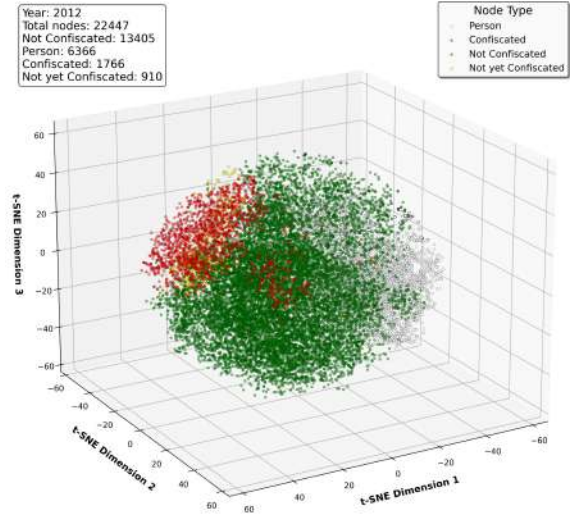
(a) Embeddings for 2006



(b) Embeddings for 2008



(c) Embeddings for 2010



(d) Embeddings for 2012

Figure 3.5: Embeddings Visualization (2006-2012)

Notes: Embeddings for different years capture the evolving relationships and structures within the ownership network. Red nodes represent already confiscated firms, yellow nodes represent firms confiscated in the future, green nodes represent non-confiscated firms and white nodes people. t-SNE is used for dimensionality reduction in this 3D visualization.

3.7 Embedding Analysis

This section will analyze the properties of the learned ownership embeddings and assess their effectiveness in capturing signals of mafia infiltration. We will first conduct a

descriptive analysis to understand how confiscated firms are positioned within the embedding space compared to non-confiscated firms. We will also explore the relationship between network topology and learned representations.

The “Geometry” of Infiltration To probe the resulting embedding space, we investigate whether firms that *will be* confiscated in the future already reside in a region of the embedding space “close” to firms that *have already* been confiscated. We conduct this analysis across multiple time snapshots (e.g., $t \in \{2000, 2005\}$).²⁰ For each year t , we first define a “seed set” C_t containing all firms judicially confiscated on or before that year. We then analyze all other firms not in C_t . For each of these firms i (which are either not-yet-confiscated or will never be), we compute its proximity to the confiscated seed set. This proximity is measured using the *cosine similarity* between its embedding vector z_i and the vector z_j for every firm $j \in C_t$:

$$S_{ij} = \frac{z_i \cdot z_j}{\|z_i\| \|z_j\|}$$

A score of 1 indicates identical representations (maximum proximity). To create a robust “infiltration proximity” score for firm i , we identify its top 5 most similar confiscated firms from the set C_t . We then calculate the average cosine similarity to these 5 firms. This score for firm i at time t is:

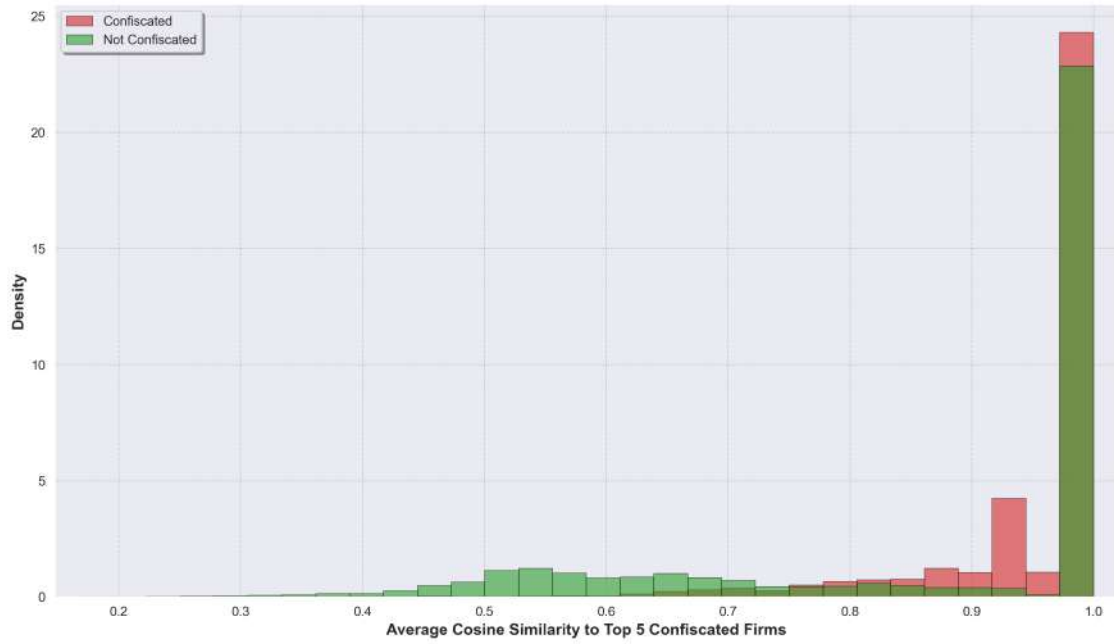
$$\text{AvgSim}_i(t) = \frac{1}{5} \sum_{j \in \text{Top5}(i, C_t)} S_{ij}$$

This procedure yields a single measure for each firm i at time t , representing its average proximity to the 5-nearest-confiscated-firms in the embedding space. We then pool these scores across the different snapshot years and plot their distributions, splitting the firms into two distinct groups: not confiscated firms and firms that will be confiscated in the future (at any point in time $t' > t$).

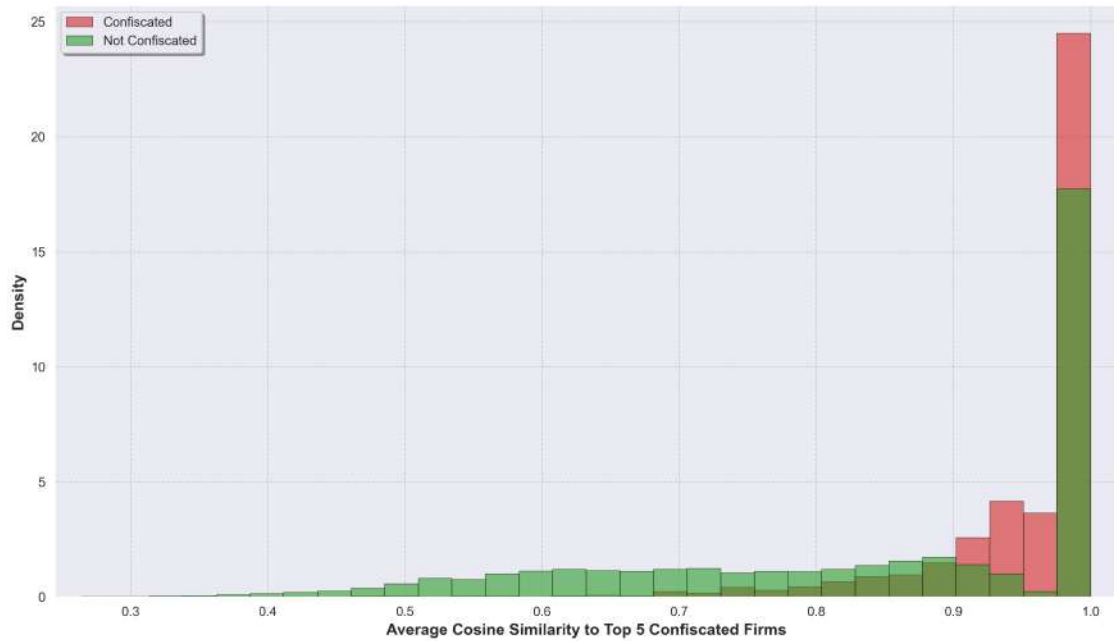
Figure 3.6 plots these two distributions, comparing the average similarity for firms that will be confiscated against those that will not. The results reveal a striking concentration for firms destined for confiscation. These firms are almost exclusively located at the upper bound of the similarity scale, indicating that they occupy regions of the embedding space nearly identical to already confiscated firms. This pattern holds consistently

²⁰2010 and 2015 in the Appendix A.22

Figure 3.6: Histogram of Average Cosine Similarity to Nearest Confiscated Firms



(a) Confiscation year: 2000



(b) Confiscation year: 2005

Notes: Distribution of Average Cosine Similarity to Nearest Confiscated Firms for the years 2000 and 2005. Green: never confiscated, Red: will be confiscated in the future

across both years analyzed (2000 and 2005), confirming that the learned embeddings effectively capture latent structural features associated with mafia infiltration risk. Consequently, firms that will be confiscated exhibit relational signatures in their ownership patterns that align them perfectly with known infiltrated entities, often well before any legal action occurs.

Interestingly, firms that will never be confiscated exhibit a highly polarized, bimodal distribution. We observe two sharply distinct subpopulations: a broad cluster of firms that remains structurally distant from confiscated entities, and a second, prominent cluster that is structurally indistinguishable from the confiscated set, massing near a similarity score of 1. This latter group represents a significant set of firms that, despite having no judicial record of confiscation, share the exact same relational signatures as known infiltrated entities, suggesting they may be at high risk of undetected infiltration. This strong polarization persists in 2010 and 2015, as shown in Figure A.22 in the Appendix, where the separation between the “safe” economy and this high-risk cluster remains clearly defined over time.

Forecasting Confiscation with Embeddings We finally validate the practical utility of the learned embeddings in predicting future confiscations. We repeat the experimental setup from Section 3.3, in a slightly different setting. First, we exclude nodes that never participate in a link, restricting the sample compared to Table 3.2,²¹ the new table is reported in Appendix Table A.20. Second we employ an earlier train/test split, i.e. 2016.²² This is done in order to limit the concern of data leakage in the generated embeddings: the TGN training data stops towards the end of 2015, while subsequent years are used for validation and testing.²³ Therefore, to ensure that the embeddings used as features do not contain information from the test period, we must restrict the test set to years after 2016. Thus, while the model saw post-2016 events to maintain the temporal continuity of node states, its parameters were never exposed to future information during training. While this setup mitigates indirect leakage, we acknowledge that concerns may not be fully eliminated. This exercise is therefore primarily intended as a proof of concept to showcase that the resulting embedding geometry is extremely

²¹These nodes will not have embeddings

²²See Figure A.23.

²³That means that years 2016 to 2024 are not used to update the TGN’s parameters—i.e. they are used in evaluation mode only, without performing and gradient updates: data starting from 2016 has no effect on the TGN’s estimated parameters.

powerful in spotting confirmed cases of infiltration.

We then rerun the same classification experiments as in Section 3.3, comparing three feature sets: (1) only firm-level data; (2) firm-level data and network features; and (3) only TGN embeddings. Note that the results from feature sets (1) and (2) are expected to differ: the sample now is smaller and ignores firms that never had ownership links.²⁴ In addition, the share of confiscated firms is significantly lower (from 25% to 15%), making the classification task more challenging. Moreover, to make the exercise even more challenging, we reduce the prediction horizon to 1, 2, 3 and 4 years ahead— $h > 0$.

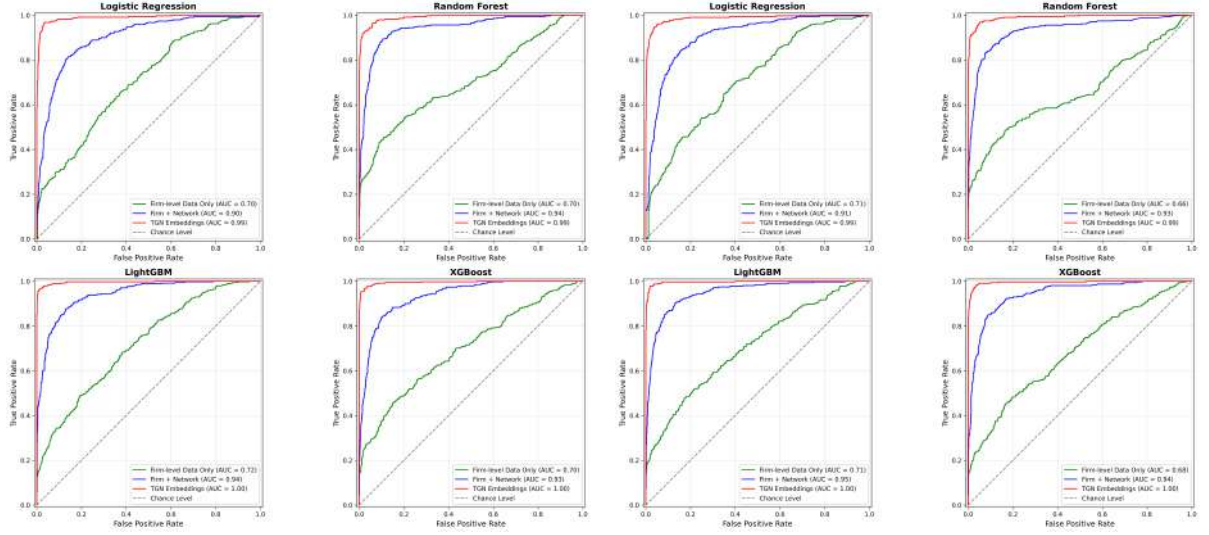
Results are reported in Tables A.21–A.24 along with the ROC curves in Figure 3.7 below and Figure A.24 in the Appendix. Two clear patterns emerge. First, embeddings consistently outperform all other feature sets across every forecast horizon and classifier. The models trained on TGN embeddings achieve near-perfect discrimination, with AUC scores approaching 1.0 even at longer horizons. This margin of improvement—from approximately 0.93 with firm and network data to nearly 0.99 with embeddings—shows that the TGN captures latent relational and temporal dependencies that traditional network statistics fail to represent.

Second, there is no substantial performance gain when moving from simple to more complex classifiers. Logistic Regression performs almost as well as advanced ensemble methods such as LightGBM and XGBoost, indicating that the embedding space is already well-structured and linearly separable. This high separability also suggests potential for interpretable clustering and visualization of infiltrated firms in the latent space. Remarkably, the predictive accuracy of the embedding-based models remains virtually unchanged across forecast horizons, revealing the temporal robustness of the learned representations. This implies that the TGN embeddings encode stable and persistent structural signatures of infiltration rather than short-term correlations.

Moreover, compared to previous studies attempting to forecast confiscations using firm-level financial data alone, our firm-only models exhibit noticeably lower performance (Cariello et al., 2024). This difference suggests that the sample used here—focused exclusively on firms within judicially confirmed confiscation networks—represents a genuinely more challenging and heterogeneous classification problem, in which traditional financial indicators provide weaker predictive signals. Together, these findings demonstrate that the TGN has effectively internalized the complex non-linear dynamics of ownership and

²⁴see Table A.20

Figure 3.7: Model Performance ROC Curves by Forecast Horizon



(a) One-Year-Ahead Prediction ($h = 1$)

(b) Two-Years-Ahead Prediction ($h = 2$)

Notes: ROC-AUC Curves for each classification model across three feature sets. The left panel (a) shows performance for 1-step-ahead predictions ($h = 1$), while the right panel (b) shows performance for two-steps-ahead predictions ($h = 2$). The blue line represents the combined Network and Firm-level feature set, the green line represents only firm-level data, and the red line represents only TGN embeddings. All the plots refer to the best performing hyperparameter configuration.

control, producing embeddings that not only organize firms in a discriminative latent space but also generalize across time. The resulting representations capture enduring topological and temporal patterns of infiltration, confirming their strong potential as general-purpose features for downstream risk prediction tasks.

3.8 Conclusion

The infiltration of legitimate economies by organized crime groups represents a critical and evolving threat. This paper introduces a novel early-warning system designed to predict firm confiscation due to mafia involvement in Italy. By constructing a unique, time-resolved corporate ownership network and integrating it with firm-level financial data, we develop a machine-learning framework capable of identifying at-risk firms up to three years in advance with a high degree of accuracy.

Our findings yield three key contributions. First, we demonstrate that a hybrid approach, combining network topology with firm fundamentals, is substantially more powerful than relying on either data source alone, achieving an out-of-sample AUC

score of up to 0.975. Second, we uncover the complementary roles of these data sources: network features are paramount for ensuring high recall (capturing a wide net of potential targets), while financial indicators provide the precision needed to pinpoint high-risk cases. Third, our feature-contribution analysis confirms existing theories that smaller, financially vulnerable firms are prime targets for infiltration. It also establishes the pivotal importance of network structure as an important indicator of risk, providing an empirical basis for a new dimension of risk-based supervision.

From a policy perspective, our framework provides a potential tool to strengthen early-warning systems for OCG infiltration, offering an additional way to signal firms at higher risk and support supervisory efforts. We acknowledge, however, that our dataset is centered on confiscated firms and their connected ownership networks, which limits external validity. Our primary contribution should therefore be understood as highlighting the importance of ownership-network information for improving prediction, while recognizing that applying such models in real-world classification tasks would require broader data access and validation.

Another contribution of this paper is the introduction of a novel, representation-based methodology for measuring unstructured relational data within economics. We apply a Temporal Graph Network (TGN) to our dynamic ownership graph, moving beyond traditional network metrics to generate dynamic “ownership embeddings” that encode the complex, evolving relational signatures of each firm. We validate this new measure by demonstrating its ability to capture latent signals of infiltration risk. In an out-of-sample forecasting exercise, used here as a proof-of-concept, these embeddings alone achieve near-perfect discrimination (AUC scores approaching 1.0). This confirms that the TGN effectively maps firms into a latent space where infiltration risk is structurally encoded. This finding strongly suggests that these embeddings can be leveraged in future empirical applications as a novel, data-driven measure to spot and analyze OCG infiltration, providing a new tool for studying the economics of crime. Future work could extend further the use of embeddings to better understand when and how infiltration occurs and to explore its economic consequences—such as its effects on competition, productivity, or firm performance—and to develop theoretical or empirical frameworks that connect the predictive findings to broader questions of market efficiency and institutional integrity.

Furthermore, we emphasize that this study is fundamentally observational. Our machine learning models are designed to maximize predictive power by identifying robust correlations between firm characteristics, network structures, and confiscation events.

We do not claim to isolate causal mechanisms. The results should be interpreted as a descriptive profiling of risk rather than evidence of a causal link between specific network features and infiltration. Nevertheless, establishing these predictive patterns is a necessary step, providing a validated, data-driven foundation for future research to investigate the causal impacts of organized crime on economic outcomes. Finally, we contextualize the broader implications of this study. It is important to emphasize that such a tool is intended to complement, not replace, investigative or judicial processes. Furthermore, while this work focuses on risk identification, the methodology offers a foundation for causal analysis.

In conclusion, this study demonstrates that the architecture of corporate control contains deep structural signals of illicit activity. By uniting advanced representation learning with network science, we not only offer a new framework for understanding criminal infiltration but also open a promising methodological avenue for analyzing complex, dynamic economic relationships.

Bibliography

- Arellano-Bover, J., De Simoni, M., Guiso, L., Macchiavello, R., Marchetti, D. J., and Prem, M. (2024). Mafias and firms.
- Ash, E., Galletta, S., and Giommoni, T. (2025). A machine learning approach to analyze and support anticorruption policy. *American Economic Journal: Economic Policy*, 17(2):162–193.
- Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15:659–688.
- Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2004). Who’s who in crime network. wanted the key player. Technical report, IUI Working Paper.
- Boeri, F., Di Cataldo, M., and Pietrostefani, E. (2024). Localized effects of confiscated and re-allocated real estate mafia assets. *Journal of Economic Geography*, 24(2):219–240.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Calamunci, F. and Drago, F. (2020). The economic impact of organized crime infiltration in the legal economy: Evidence from the judicial administration of organized crime firms. *Italian Economic Journal*, 6(2):275–297.
- Cariello, P., De Simoni, M., and Iezzi, S. (2024). A machine learning approach for the detection of firms infiltrated by organised crime in Italy. *Quaderni dell’antiriciclaggio d’Italia, Unità di Informazione Finanziaria per l’Italia, Roma, Banca*.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.
- Chircop, J., Fabrizi, M., Malaspina, P., and Parbonetti, A. (2023). Anti-mafia police actions, criminal firms, and peer firm tax avoidance. *Journal of Accounting Research*, 61(1):243–277.
- Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 191–198. ACM.

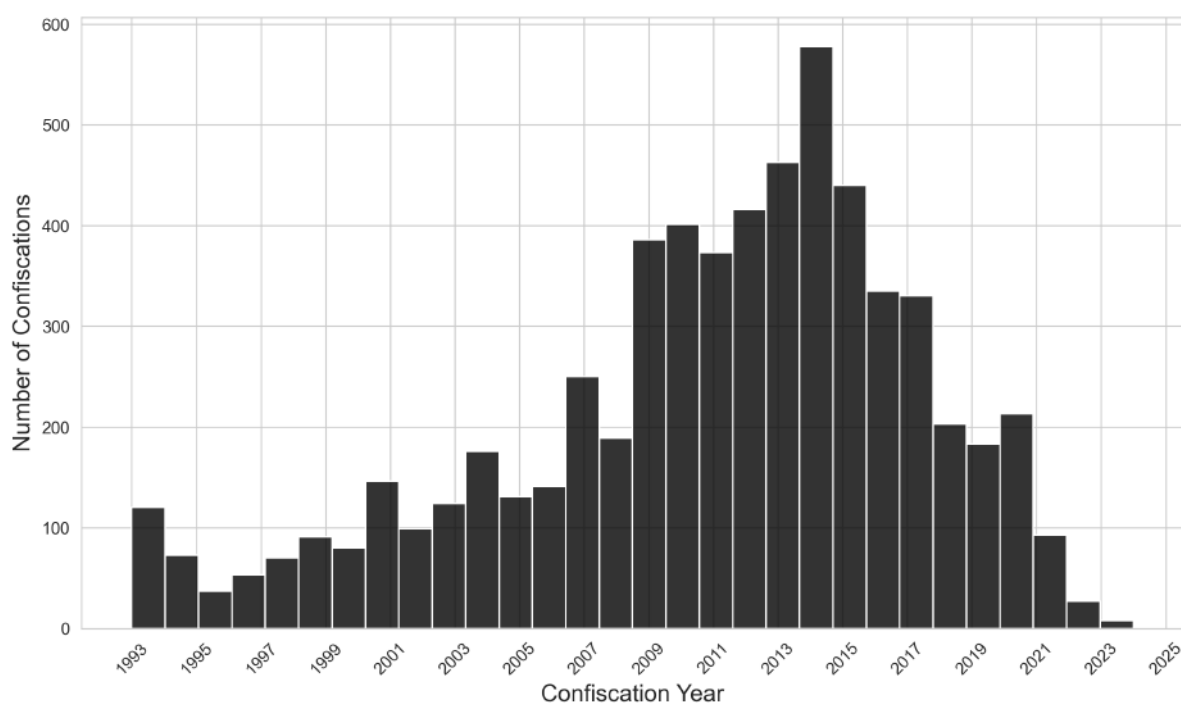
- Daniele, G. and Dipoppa, G. (2023). Fighting organized crime by targeting their revenue: Screening, mafias, and public funds. *The Journal of Law, Economics, and Organization*, 39(3):722–746.
- De Luca, R., Greco, R., Immordino, G., and Oliviero, T. (2025). Mafia infiltration and ownership dynamics in italian companies during covid-19. Technical report, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.
- Dell, M. (2025). Deep learning for economists. *Journal of Economic Literature*, 63(1):5–58.
- Di Cataldo, M. and Mastrorocco, N. (2022). Organized crime, captured politicians, and the allocation of public resources. *The Journal of Law, Economics, and Organization*, 38(3):774–839.
- Europol (2024). Leveraging legitimacy: How the eu’s most threatening criminal networks abuse legal business structures. Technical report, Publications Office of the European Union. ISBN 978-92-95236-92-9, DOI: 10.2813/7731594, QL-01-24-014-EN-N.
- Fey, M. and Lenssen, J. E. (2019). Fast graph representation learning with pytorch geometric.
- Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Le Moglie, M. and Sorrenti, G. (2022). Revealing “mafia inc.”? financial crisis, organized crime, and the birth of new enterprises. *Review of Economics and Statistics*, 104(1):142–156.
- Lindquist, M. J. and Zenou, Y. (2019). Crime and networks: Ten policy lessons. *Oxford Review of Economic Policy*, 35(4):746–771.
- Lu, C., Durante, D., and Friel, N. (2025). Zero-inflated stochastic block modelling of efficiency-security trade-offs in weighted criminal networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, page qnaf029.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67.
- Mastrobuoni, G. (2020). Crime is terribly revealing: Information technology and police productivity. *The Review of Economic Studies*, 87(6):2727–2753.

- Mirenda, L., Mocetti, S., and Rizzica, L. (2022). The economic effects of mafia: Firm level evidence. *American Economic Review*, 112(8):2748–2773.
- Morselli, C. (2008). The efficiency–security trade-off. In *Inside Criminal Networks*, pages 1–9. Springer.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library.
- Pinotti, P. (2015). The economic costs of organised crime: Evidence from southern Italy. *The Economic Journal*, 125(586):F203–F232.
- Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., and Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*.
- Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., and Sun, Y. (2021). Masked label prediction: Unified message passing model for semi-supervised classification.
- Trivedi, R., Farajtabar, M., Biswal, P., and Zha, H. (2019). Dyrep: Learning representations over dynamic graphs. In *International Conference on Learning Representations (ICLR)*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., and Achan, K. (2020). Inductive representation learning on temporal graphs. In *International Conference on Learning Representations (ICLR)*.
- Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Dévlin, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 12310–12320. PMLR.
- Zheng, Y., Xu, Z., and Xiao, A. (2023). Deep learning in economics: a systematic and critical review. *Artificial Intelligence Review*, 56(9):9497–9539.

A Appendix

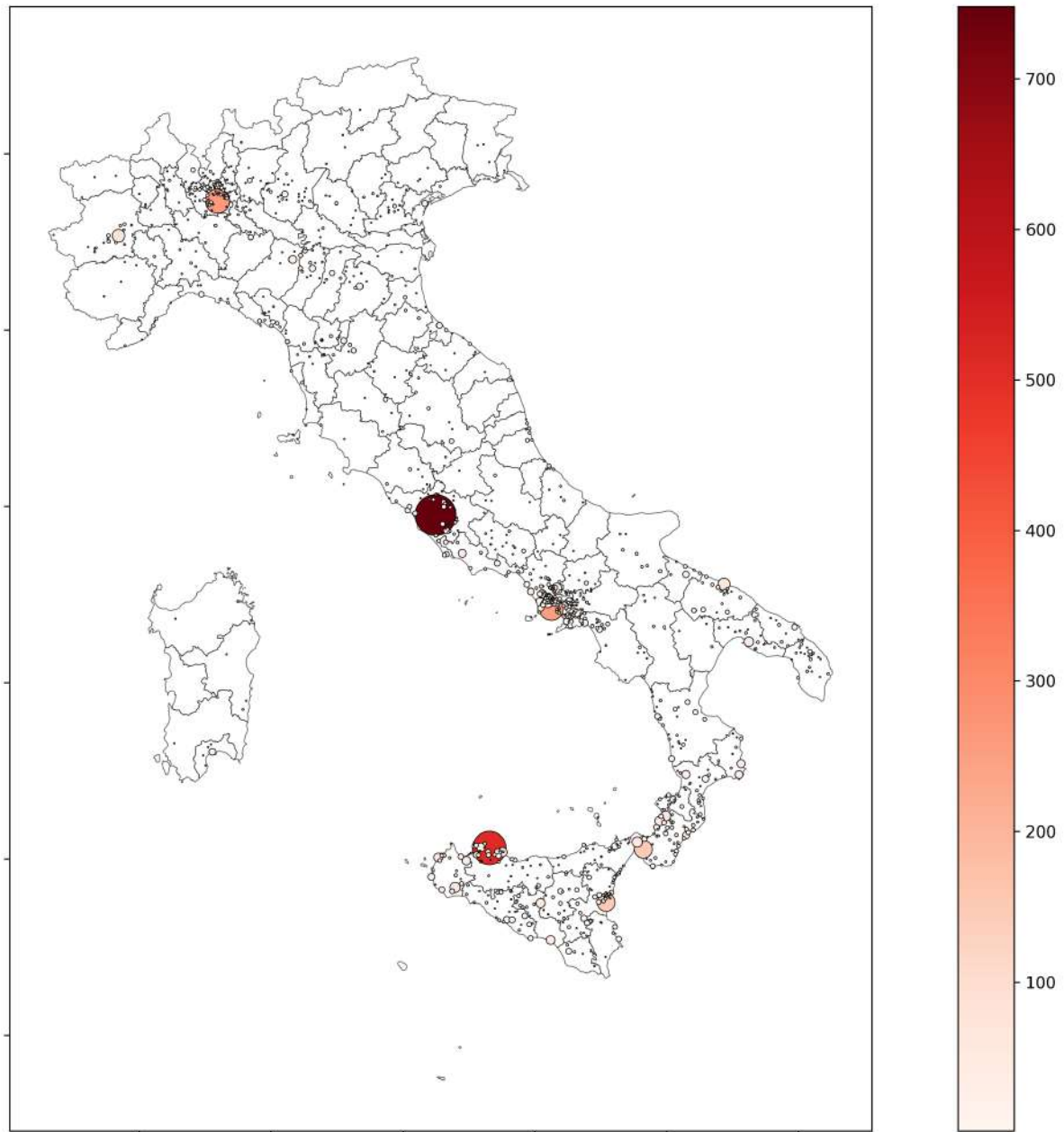
A.1 Figures

Figure A.1: Distribution of Confiscations over time



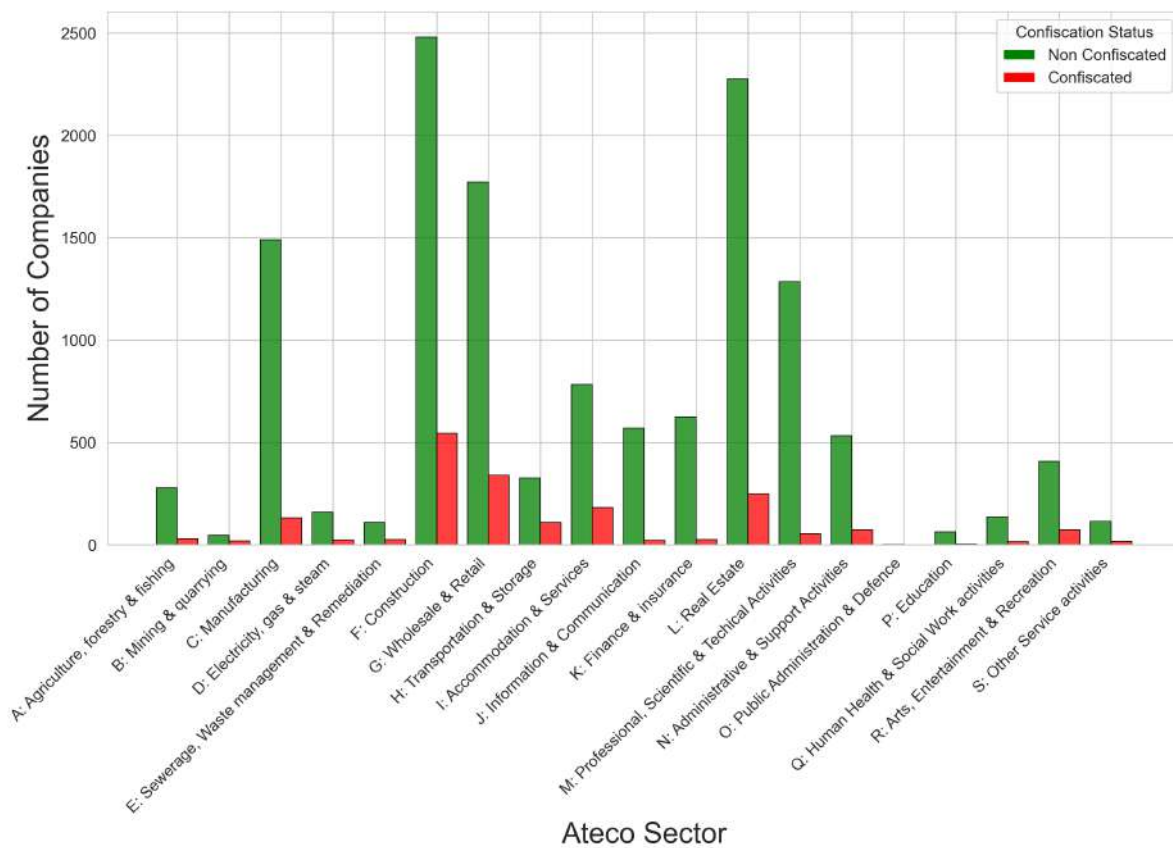
Notes: The figure shows the distribution of confiscation dates over time as reported by ANBSC. The confiscation date indicates the beginning of legal proceedings against the firm.

Figure A.2: Geographical Distribution of Confiscations



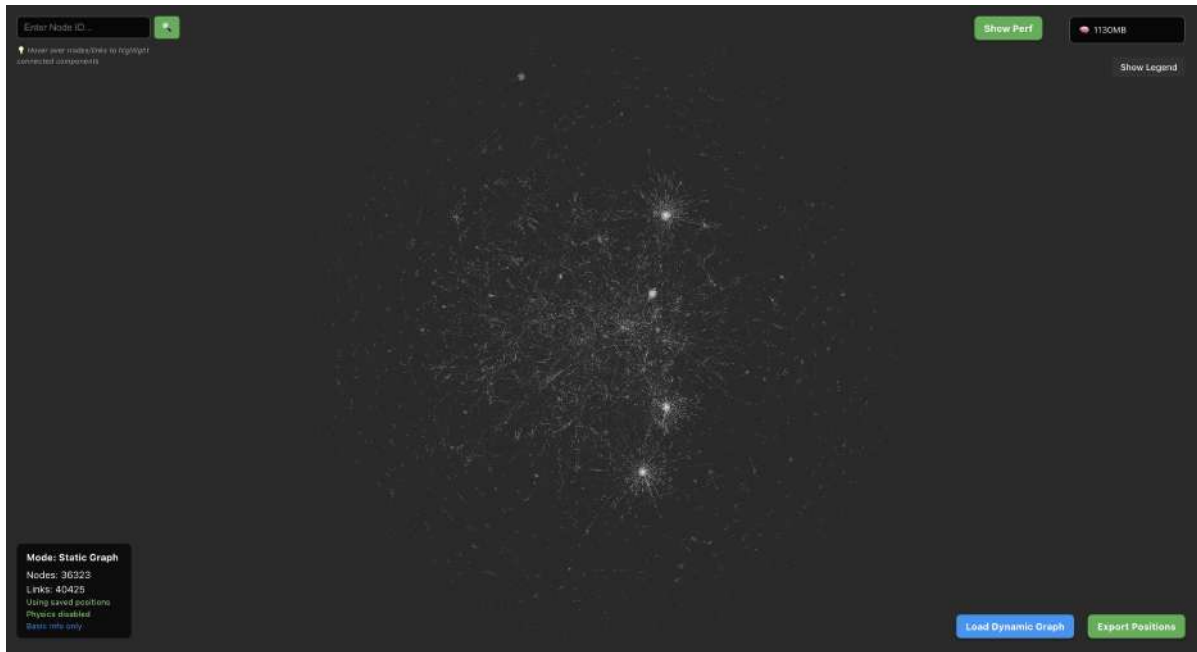
Notes: The figure shows the municipality-level geographic distribution of confiscations over Italy as reported by ANBSC. The size of the circle indicates the number of firms.

Figure A.3: Sector Distribution of Confiscated and Non-Confiscated Firms

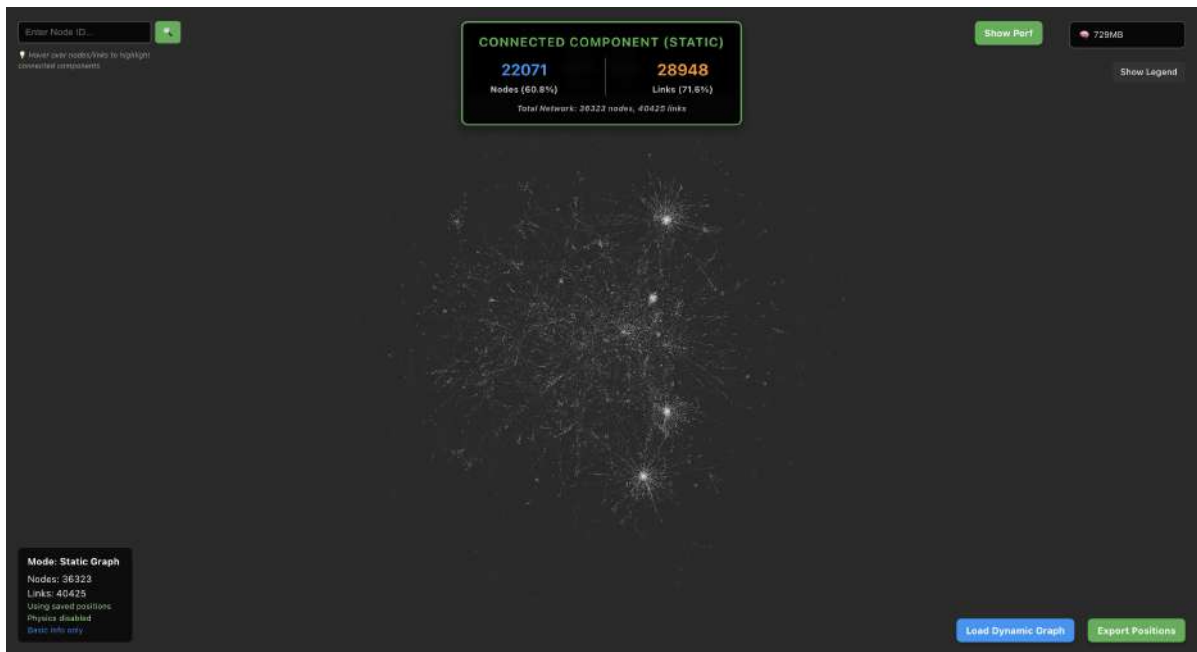


Notes: The figure shows the sectoral distribution of confiscated and non-confiscated firms, following the NACE classification.

Figure A.4: Static Graph



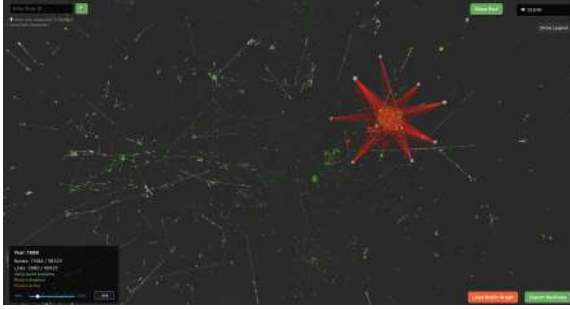
(a) All Nodes



(b) Main Component

Notes: Panel (a) shows the static graph, including all nodes and edges over time. Panel (b) highlights nodes belonging to the main component. No node coloring for the static graph.

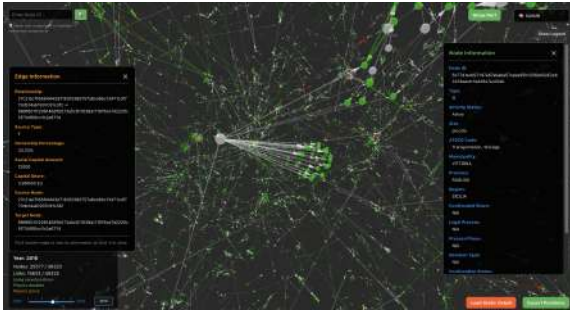
Figure A.5: Network Snapshots over Time



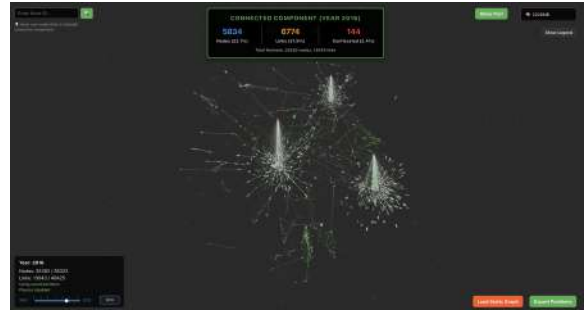
(a) Network Snapshot: 1998



(b) Network Snapshot: 2001



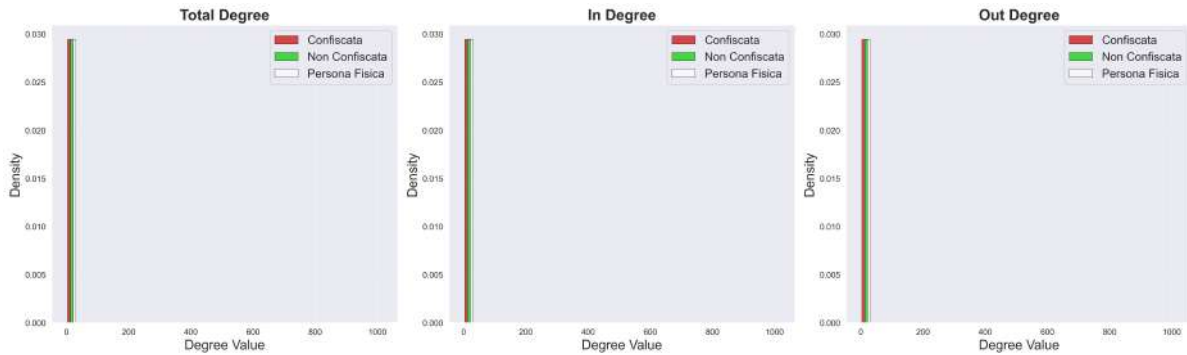
(c) Network Snapshot: 2010



(d) Network Snapshot: 2016

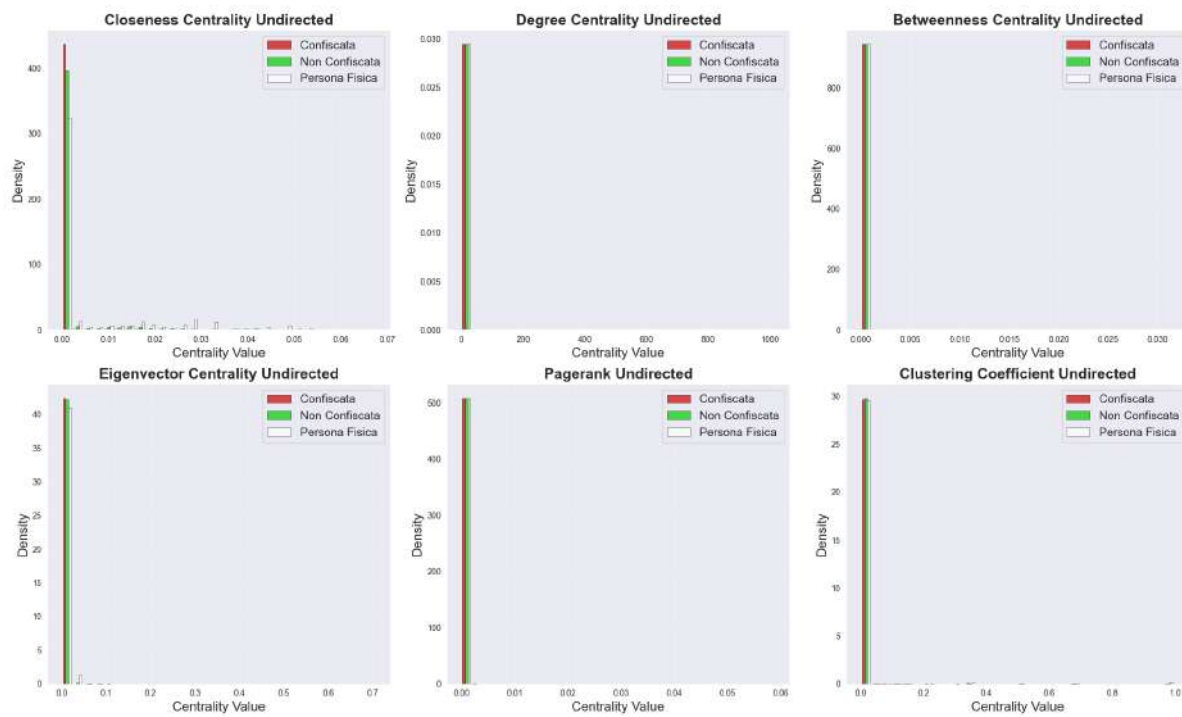
Notes: Different snapshots of the dynamic graph, showcasing visualization features. In panel A.5c the right banner displays node attributes and the left banner edge attributes. Panel A.5d displays the main connected component with the three hub nodes

Figure A.6: Degree Distributions



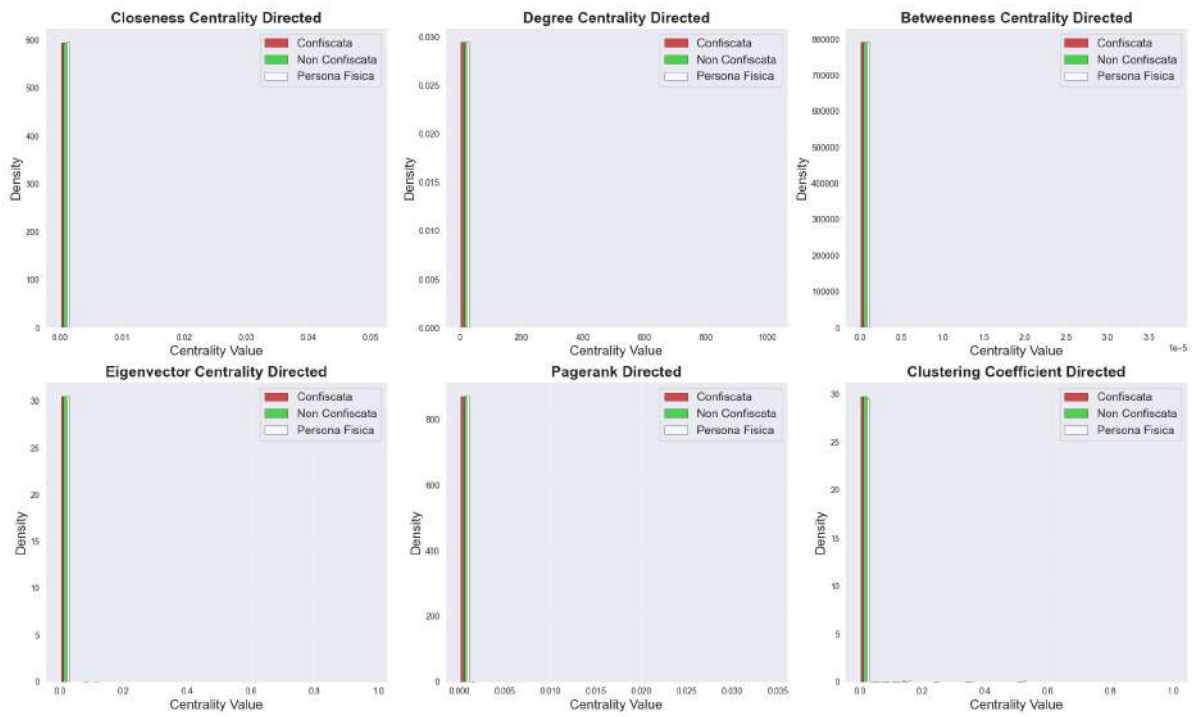
Notes: Yearly degree distributions by node type.

Figure A.7: Undirected Centralities



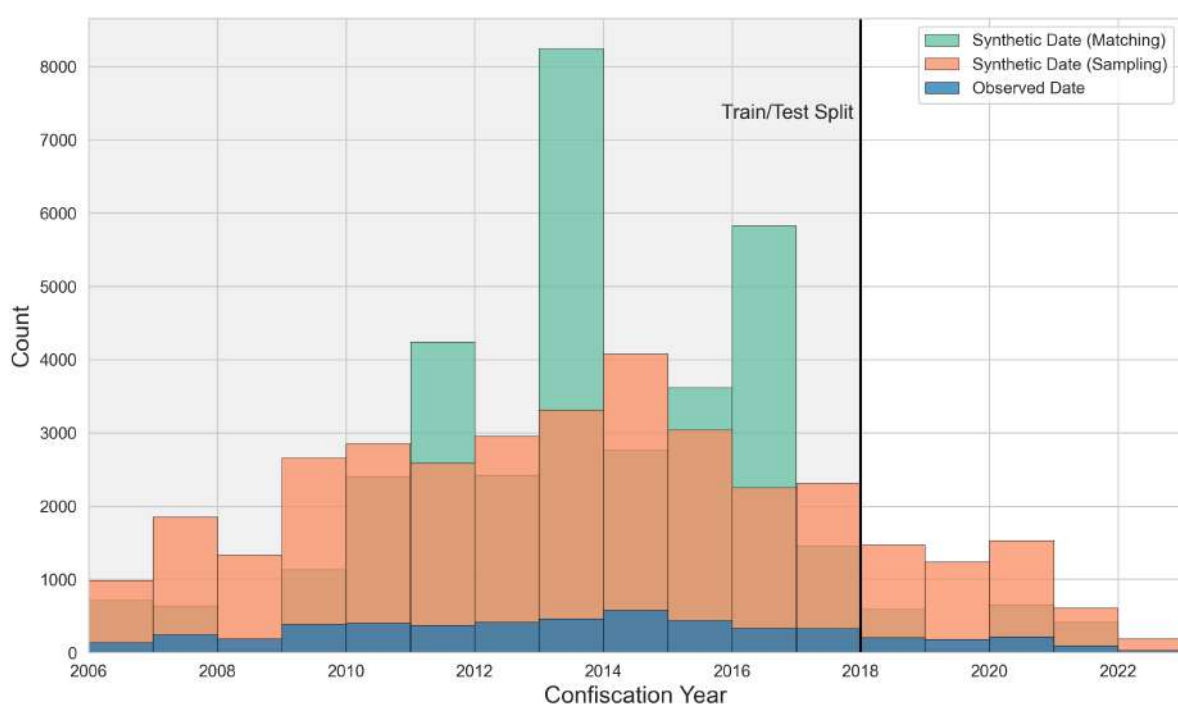
Notes: Yearly Undirected centralities by node type.

Figure A.8: Directed Centralities



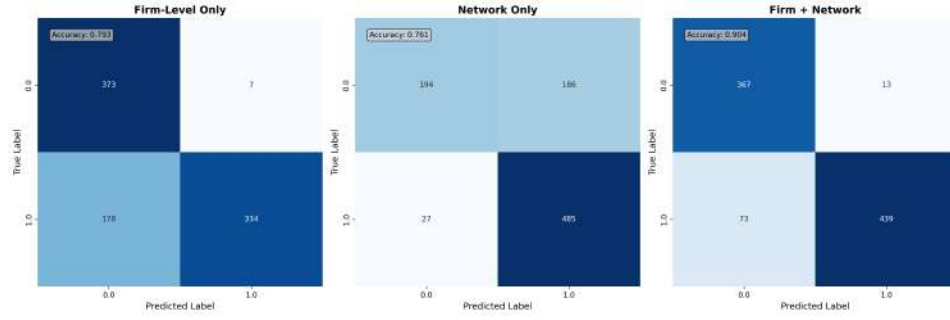
Notes: Yearly directed centralities by node type.

Figure A.9: Actual and Synthetic Confiscation Dates

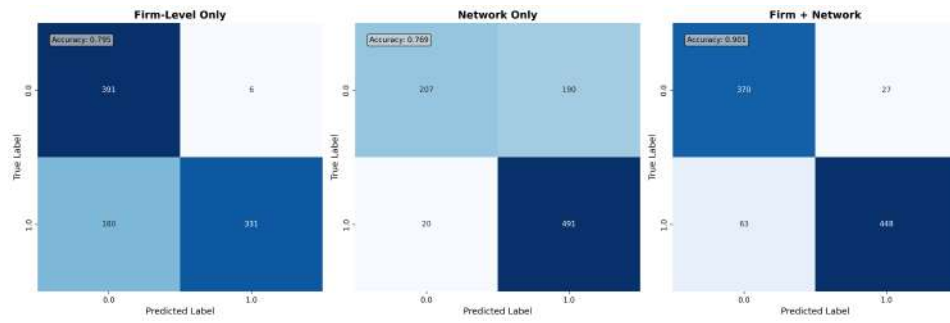


Notes: Output of the matching procedure for inferring “confiscation” dates for non-confiscated firms. Both matching and sampling methodologies for creating synthetic confiscation dates are displayed.

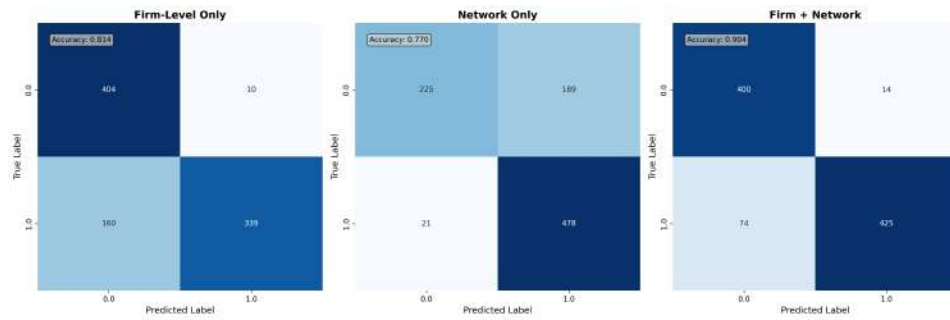
Figure A.10: XGBoost, Confusion Matrices



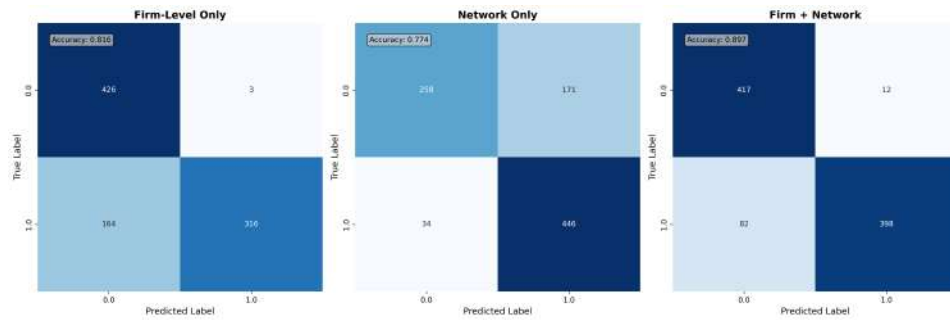
(a) $h = 0$



(b) $h = 1$



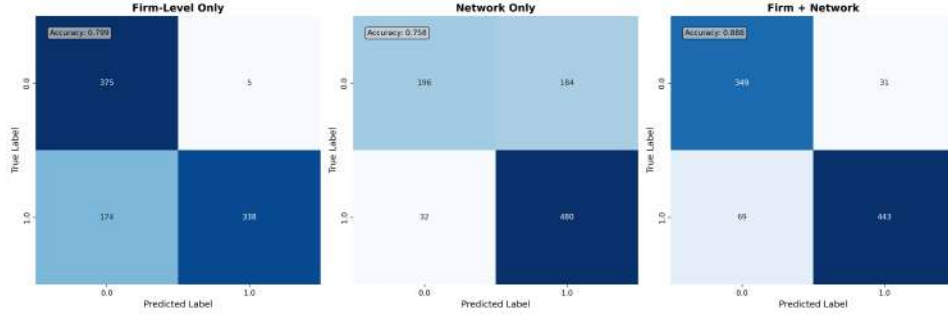
(c) $h = 2$



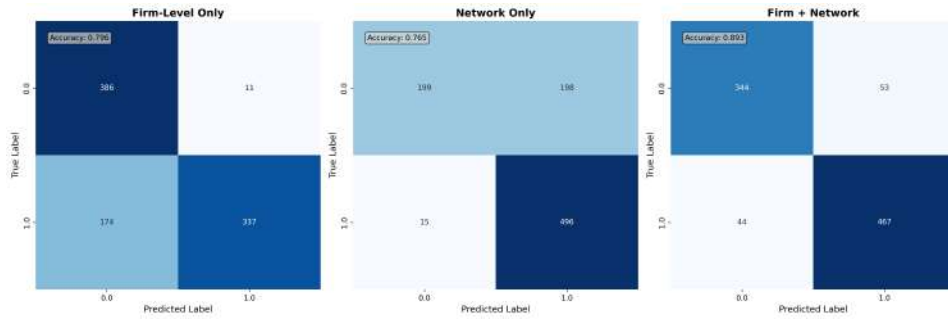
(d) $h = 3$

Notes: Distribution of correct predictions, along with Type I and Type II Errors for XGBoost model across different feature sets.

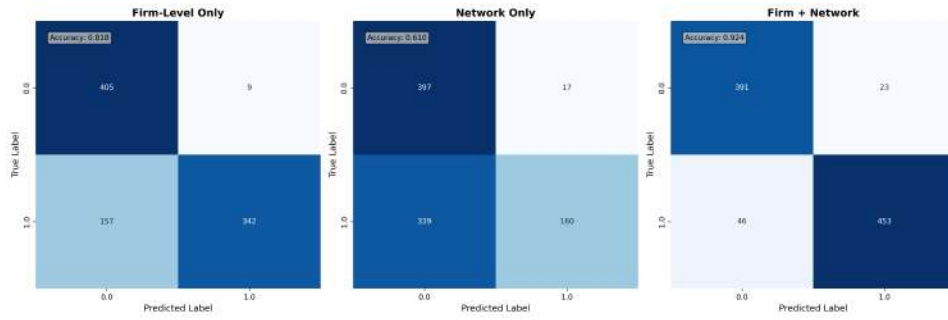
Figure A.11: LightGBM, Confusion Matrices



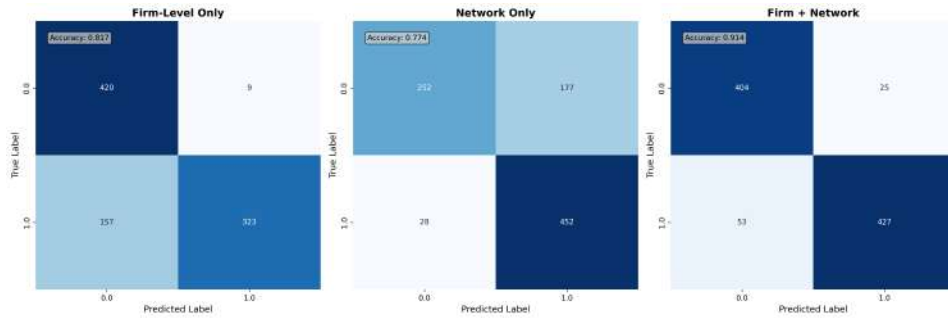
(a) $h = 0$



(b) $h = 1$



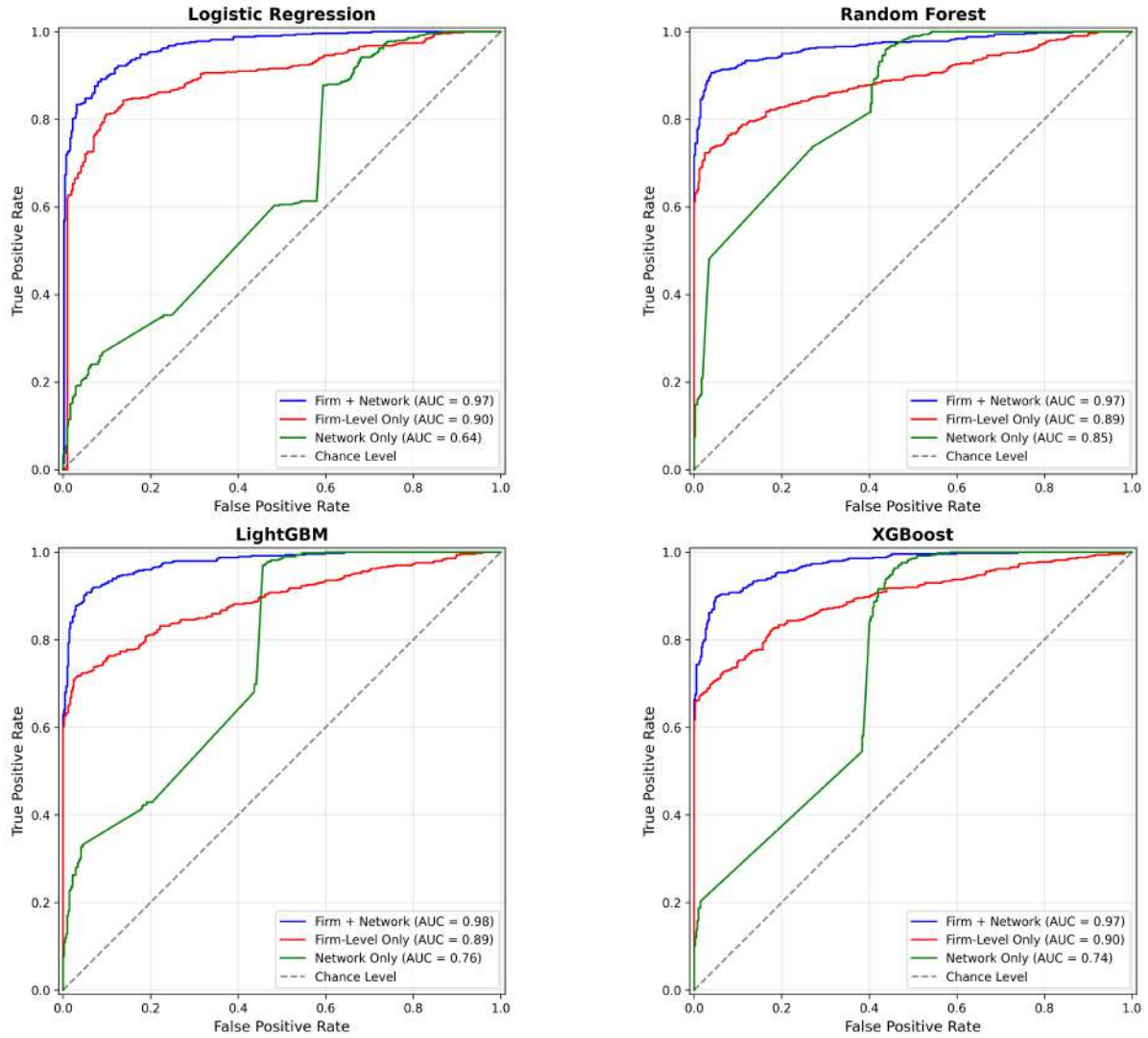
(c) $h = 2$



(d) $h = 3$

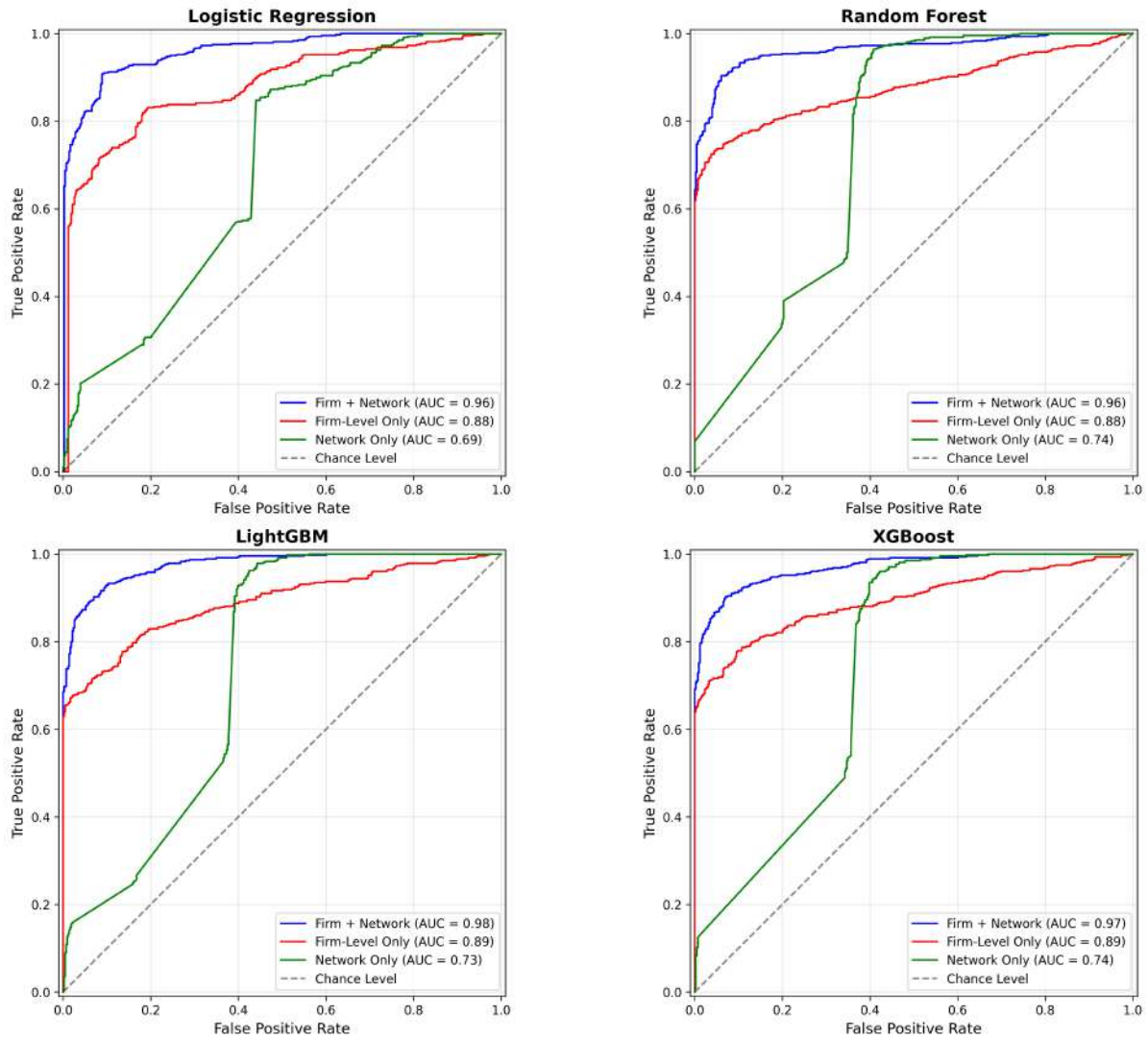
Notes: Distribution of correct predictions, along with Type I and Type II Errors for XGBoost model across different feature sets.

Figure A.12: Model Performance, $h = 2$



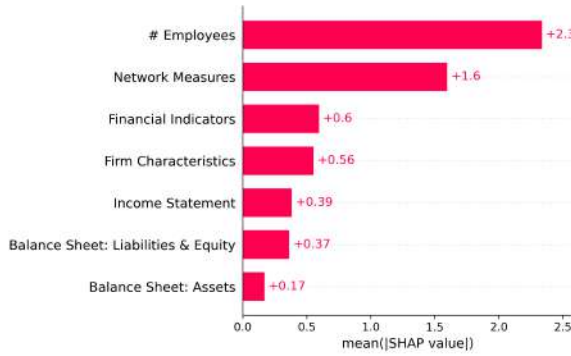
Notes: ROC-AUC Curves across the three different feature sets for each classification model. The blue line is for the Network and Firm-level data feature set, the green line for only network data and the red for firm-level data only.

Figure A.13: Model Performance, $h = 3$

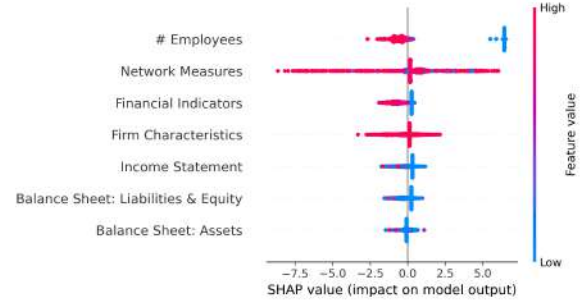


Notes: ROC-AUC Curves across the three different feature sets for each classification model. The blue line is for the Network and Firm-level data feature set, the green line for only network data and the red for firm-level data only.

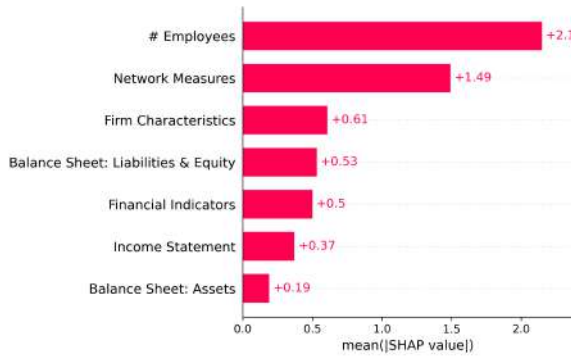
Figure A.14: XGBoost SHAP Values by Feature, $h \in \{1, 2, 3\}$



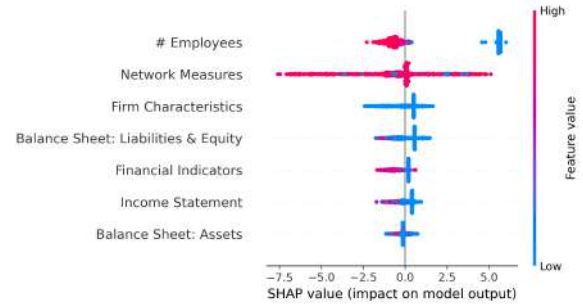
(a) Feature Importance, $h = 1$



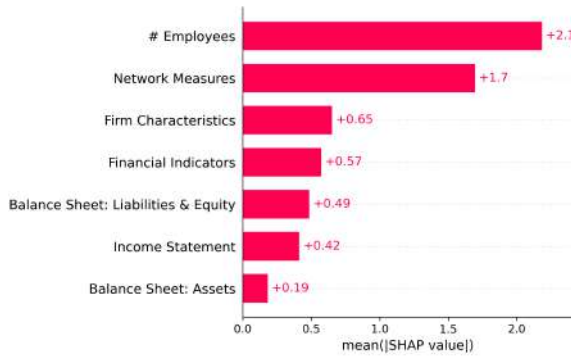
(b) Features' impact on model's output, $h = 1$



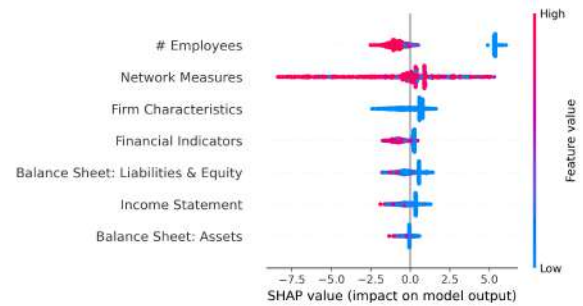
(c) Feature Importance, $h = 2$



(d) Features' impact on model's output, $h = 2$



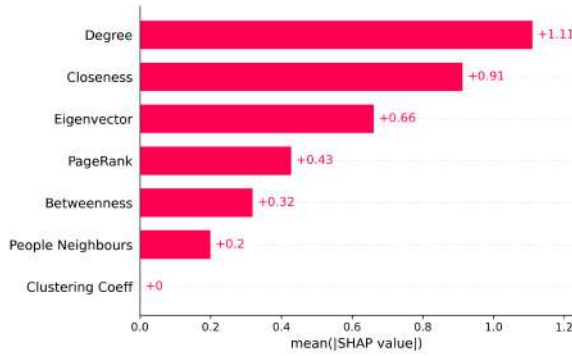
(e) Feature Importance, $h = 3$



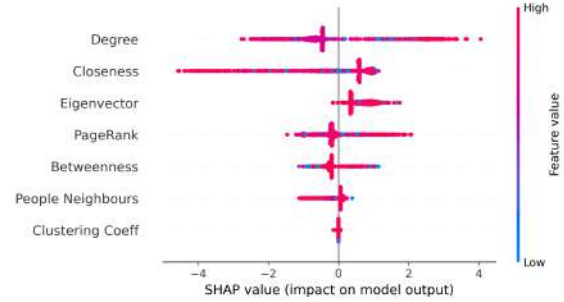
(f) Features' impact on model's output, $h = 3$

Notes: On the left column: SHAP Value for each aggregated category of feature. The x-axis the mean absolute SHAP Value, i.e. average impact of each feature group on the magnitude of the model's prediction. On the right column: each point on the plot corresponds to a single observation for a given feature. The position on the x-axis is the SHAP value, representing the impact of that feature on the model's prediction for that observation, if $x > 0$, observation is more likely to be confiscated. The color indicates feature value, from low (blue) to high (red).

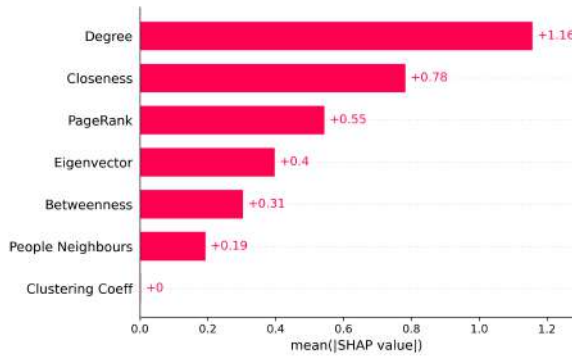
Figure A.15: XGBoost SHAP Values by Network Feature, $h \in \{1, 2, 3\}$



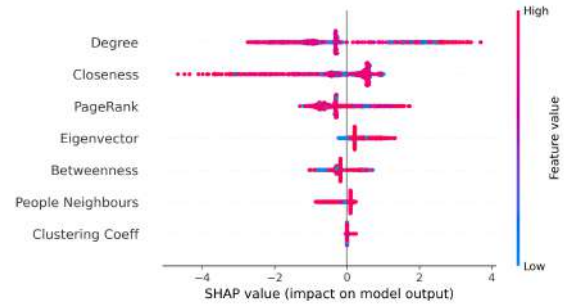
(a) Feature Importance, $h = 1$



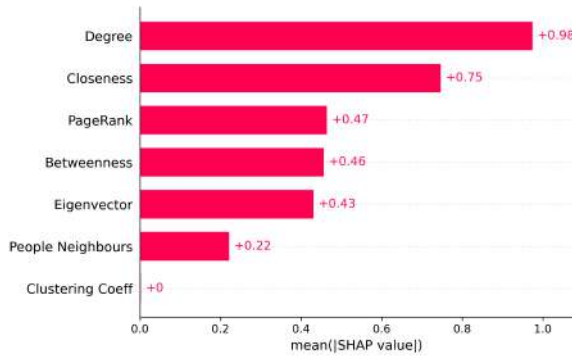
(b) Features' impact on model's output, $h = 1$



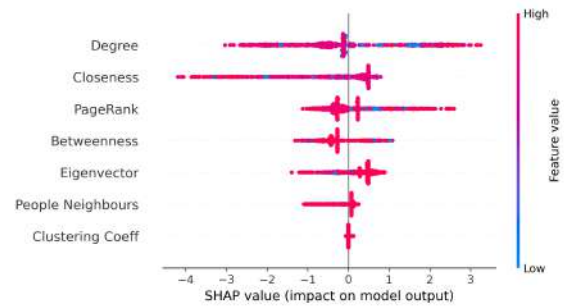
(c) Feature Importance, $h = 2$



(d) Features' impact on model's output, $h = 2$



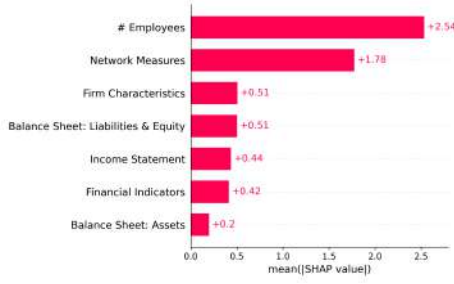
(e) Feature Importance, $h = 3$



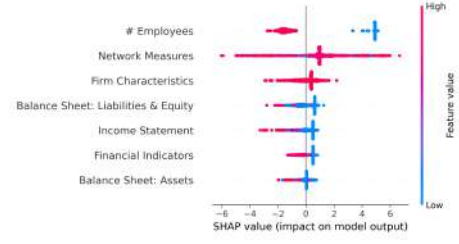
(f) Features' impact on model's output, $h = 3$

Notes: On the left column: SHAP Value for each aggregated category of feature. The x-axis the mean absolute SHAP Value, i.e. average impact of each feature group on the magnitude of the model's prediction. On the right column: each point on the plot corresponds to a single observation for a given feature. The position on the x-axis is the SHAP value, representing the impact of that feature on the model's prediction for that observation, if $x > 0$ observation is more likely to be confiscated. The color indicates feature value, from low (blue) to high (red).

Figure A.16: LightGBM SHAP Values by Feature, $h \in \{0, 1, 2, 3\}$



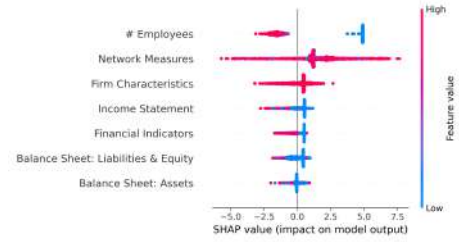
(a) Feature Importance, $h = 0$



(b) Features' impact on model's output, $h = 0$



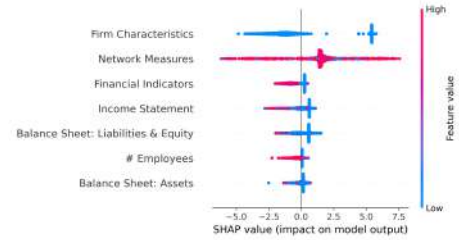
(c) Feature Importance, $h = 1$



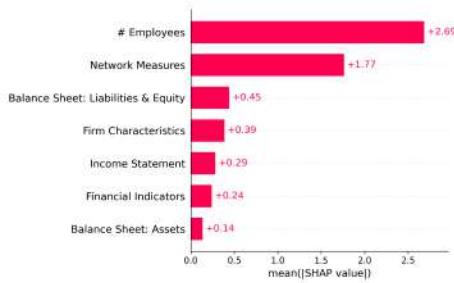
(d) Features' impact on model's output, $h = 1$



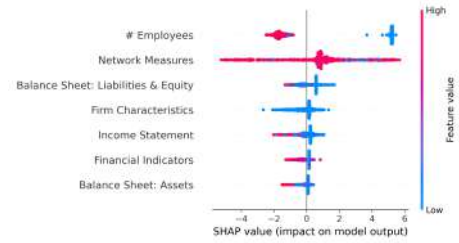
(e) Feature Importance, $h = 2$



(f) Features' impact on model's output, $h = 2$



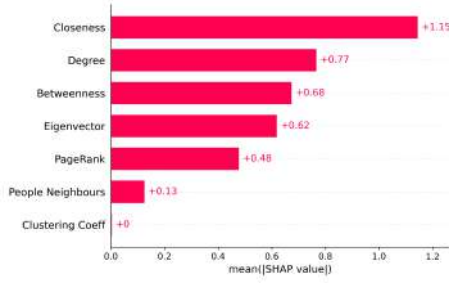
(g) Feature Importance, $h = 3$



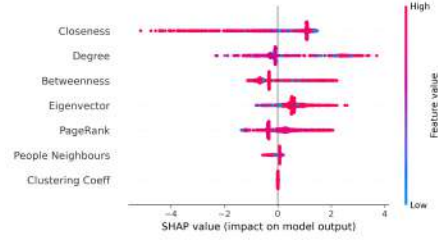
(h) Features' impact on model's output, $h = 3$

Notes: On the left column: SHAP Value for each aggregated category of feature. The x-axis the mean absolute SHAP Value, i.e. average impact of each feature group on the magnitude of the model's prediction. On the right column: each point on the plot corresponds to a single observation for a given feature. The position on the x-axis is the SHAP value, representing the impact of that feature on the model's prediction for that observation, if $x > 0$ observation is more likely to be confiscated. The color indicates feature value, from low (blue) to high (red).

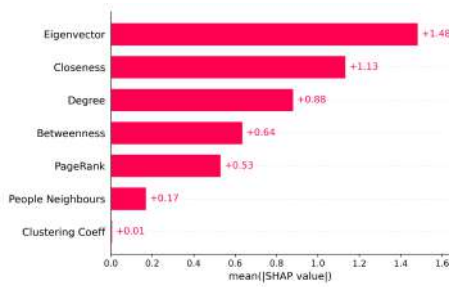
Figure A.17: LightGBM SHAP Values by Network Feature, $h \in \{0, 1, 2, 3\}$



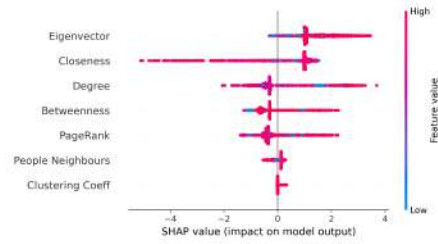
(a) Feature Importance, $h = 0$



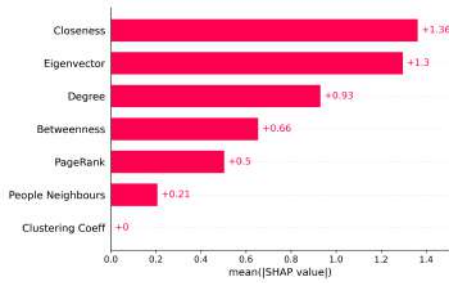
(b) Features' impact on model's output, $h = 0$



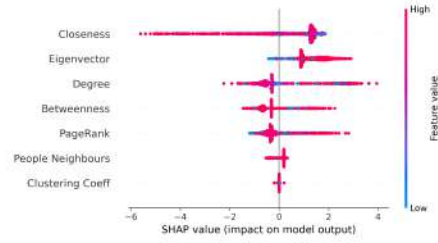
(c) Feature Importance, $h = 1$



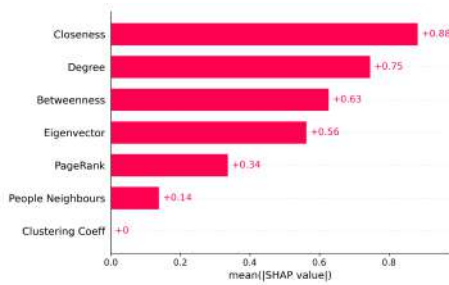
(d) Features' impact on model's output, $h = 1$



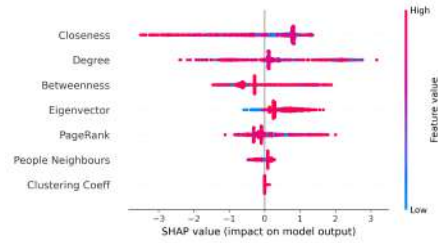
(e) Feature Importance, $h = 2$



(f) Features' impact on model's output, $h = 2$



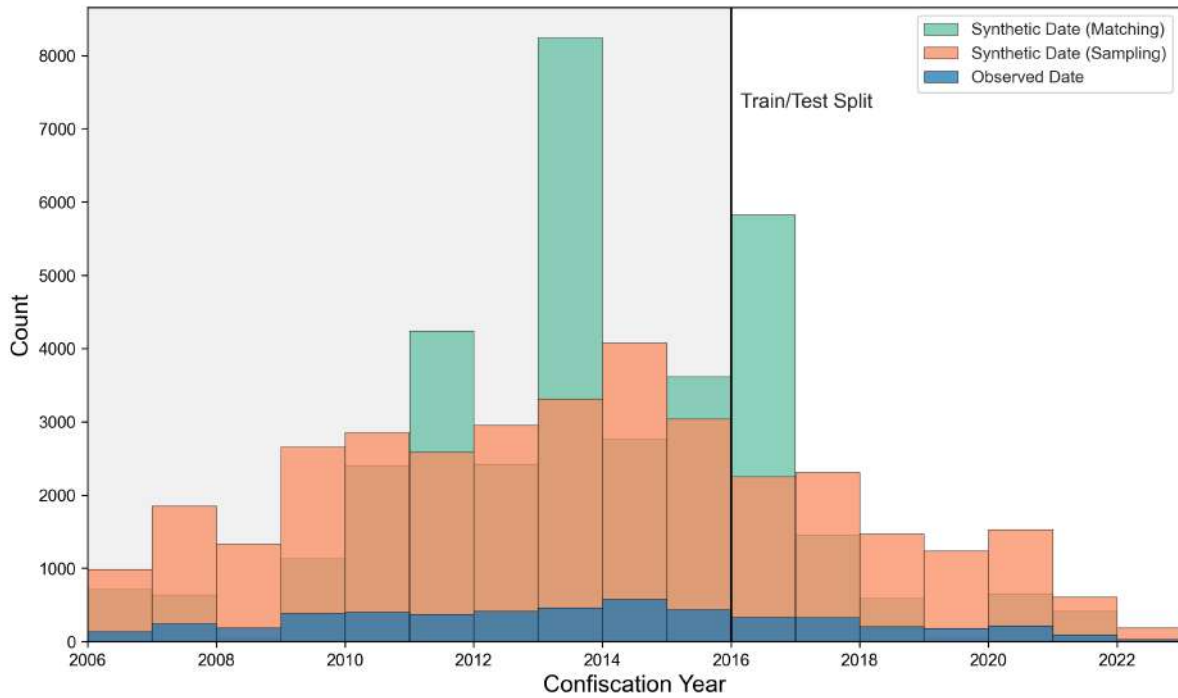
(g) Feature Importance, $h = 3$



(h) Features' impact on model's output, $h = 3$

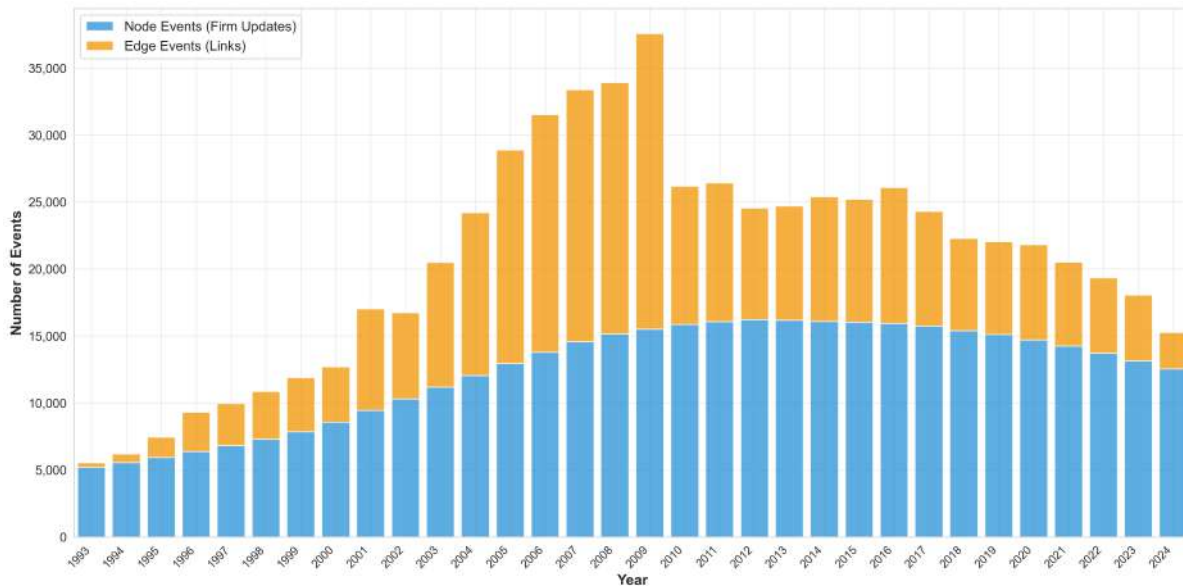
Notes: On the left column: SHAP Value for each aggregated category of feature. The x-axis the mean absolute SHAP Value, i.e. average impact of each feature group on the magnitude of the model's prediction. On the right column: each point on the plot corresponds to a single observation for a given feature. The position on the x-axis is the SHAP value, representing the impact of that feature on the model's prediction for that observation, if $x > 0$ observation is more likely to be confiscated. The color indicates feature value, from low (blue) to high (red).

Figure A.18: Earlier Cut-off Date: Actual and Synthetic Confiscation Dates



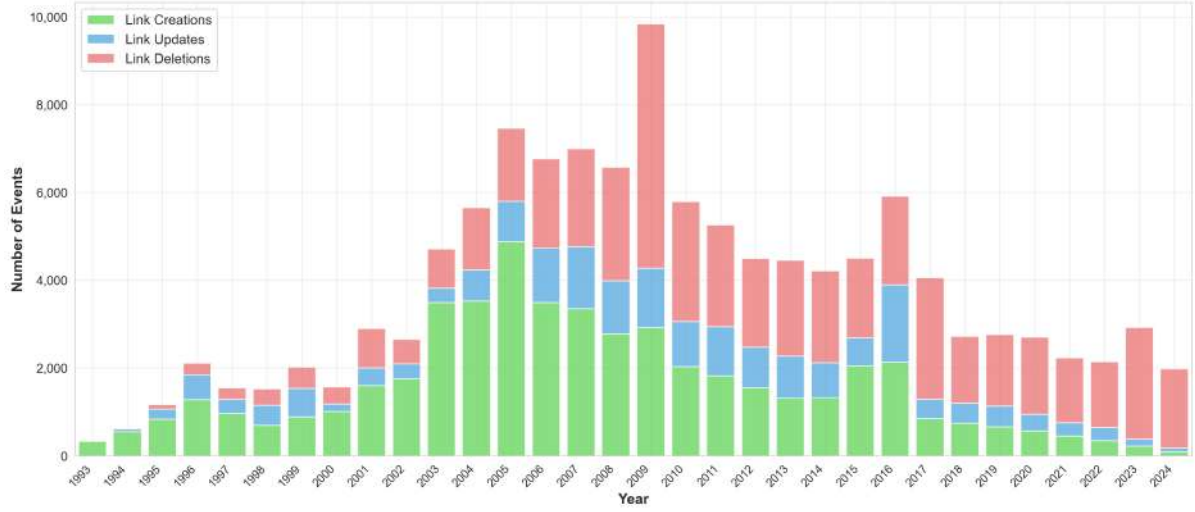
Notes: Output of the matching procedure for inferring “confiscation” dates for non-confiscated firms. Both matching and sampling methodologies for creating synthetic confiscation dates are displayed. Here the train/test cut-off date is 2016.

Figure A.19: Network Events Over Time



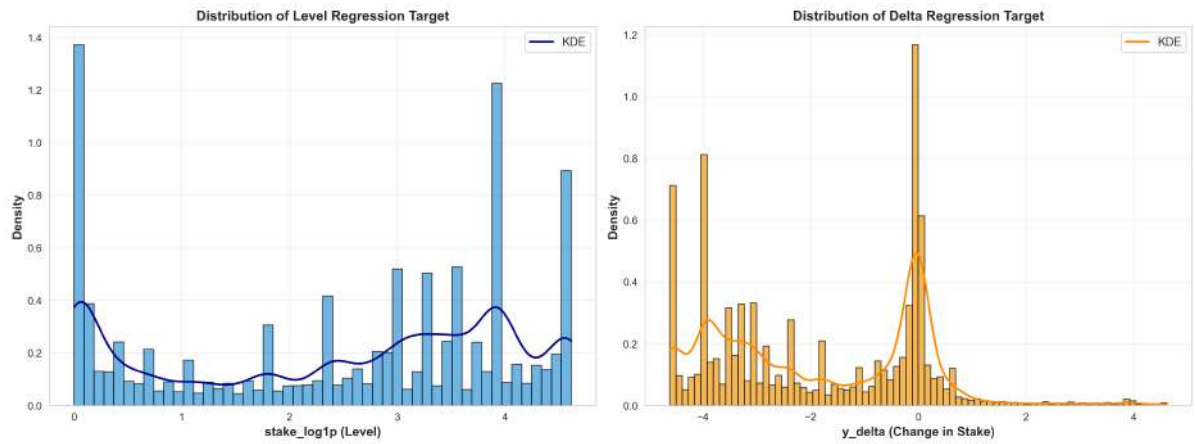
Notes: The figure illustrates the number of events per year, considering both edge events, i.e. *interaction events* in blue and node events, i.e. *self-loop events* in orange.

Figure A.20: Edge Events over Time



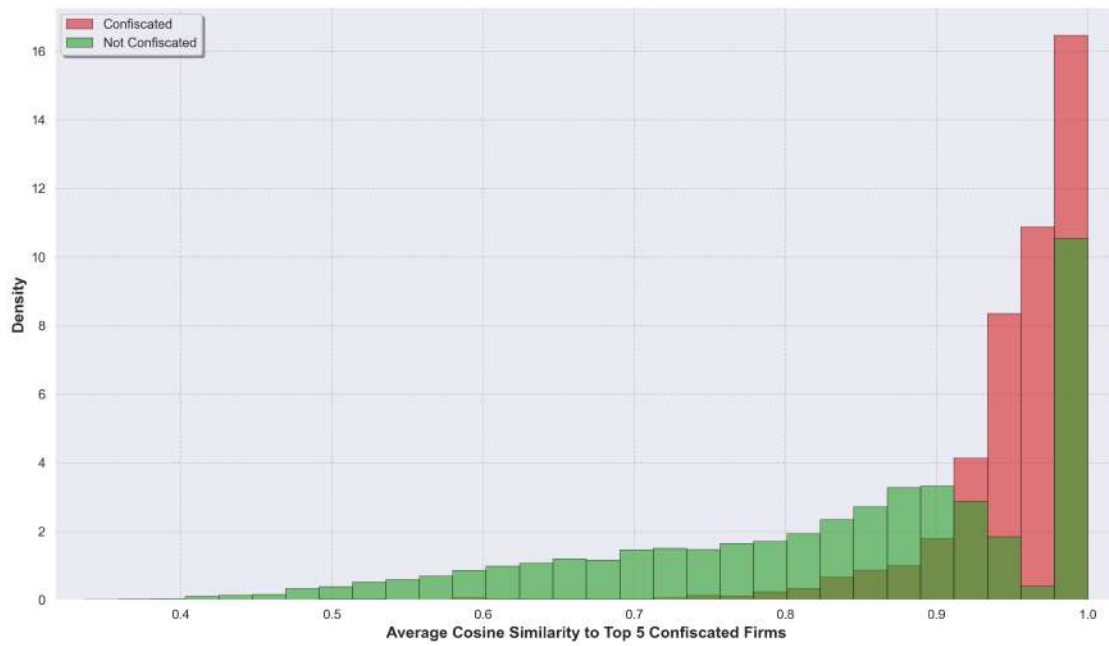
Notes: The figure shows the breakdown of edge events by type, distinguishing between link creation (in green), link deletion (in red).

Figure A.21: Distribution Ownership Stake Target Variables

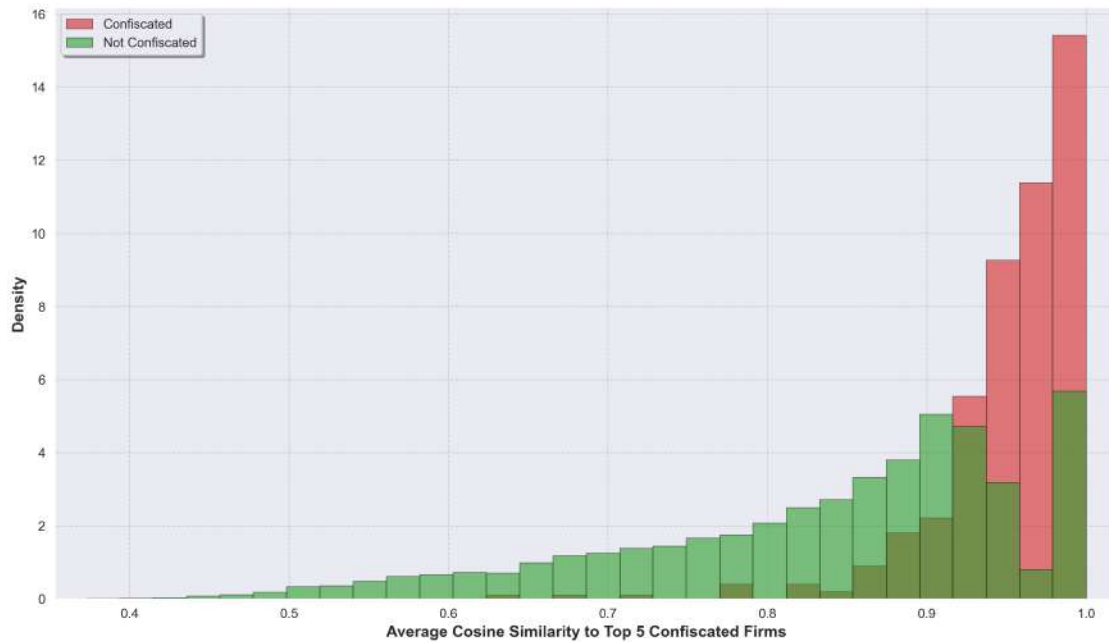


Notes: The figure shows the distribution of continuous target variables of the TGN's regression heads: (log of) ownership stake and Δ ownership stake. The solid line indicates the KDE estimate.

Figure A.22: Histogram of Average Cosine Similarity to Nearest Confiscated Firms



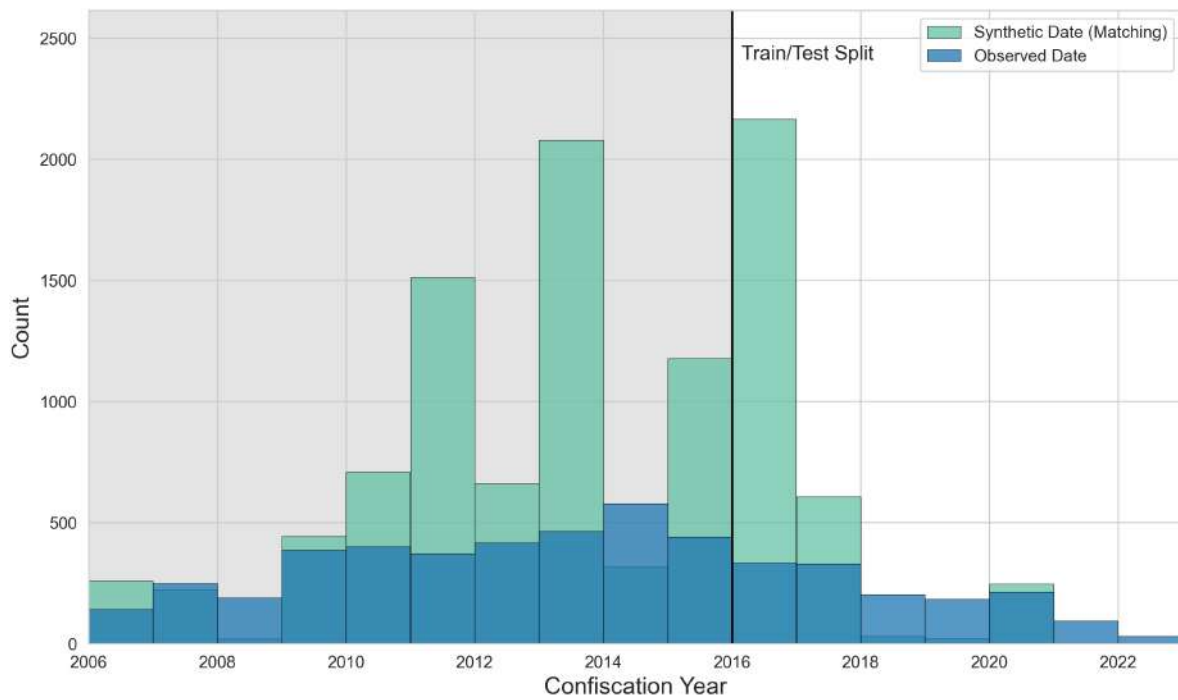
(a) Confiscation year: 2010



(b) Confiscation year: 2010

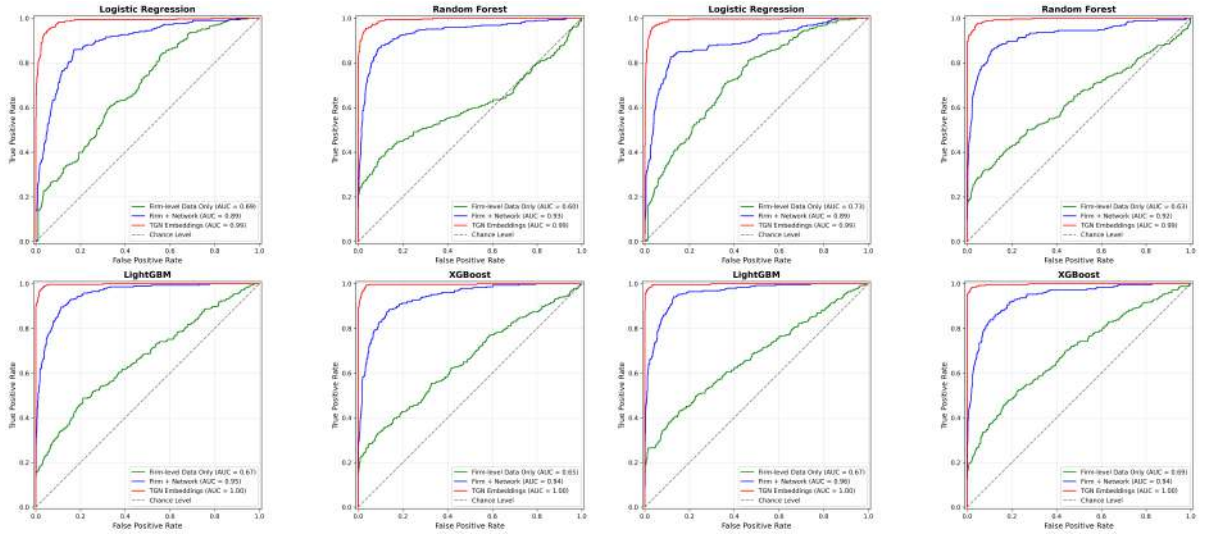
Notes: Distribution of Average Cosine Similarity to Nearest Confiscated Firms for the years 2010 and 2015. Green: never confiscated, Red: will be confiscated in the future

Figure A.23: Matching Strategy for Synthetic Reference Year Assignment



Notes: The figure depicts the results from the matching strategy used to assign synthetic reference years for confiscation events to non-confiscated firms (green). The blue bars indicate confiscation events for confiscated firms (observed)

Figure A.24: Model Performance ROC Curves by Forecast Horizon



(a) Contemporaneous Prediction ($h = 3$)

(b) One-Year-Ahead Prediction ($h = 4$)

Notes: ROC-AUC Curves for each classification model across three feature sets. The left panel (a) shows performance for 1-step-ahead predictions ($h = 1$), while the right panel (b) shows performance for two-steps-ahead predictions ($h = 2$). The blue line represents the combined Network and Firm-level feature set, the green line represents only firm-level data, and the red line represents only TGN embeddings. All the plots refer to the best performing hyperparameter configuration.

A.2 Tables

Table A.1: Centrality Measures Descriptive Statistics by Network Type and Node Category

Centrality Measure	Yearly Networks			Static Network		
	People	Non-Confiscated	Confiscated	People	Non-Confiscated	Confiscated
Degree Centrality (Directed)	1.9903 (2.4802)	0.6577 (2.9199)	0.6511 (9.9793)	3.1355 (4.5168)	1.6868 (7.7052)	2.2088 (19.5877)
Degree Centrality (Undirected)	1.9903 (2.4802)	0.6574 (2.9186)	0.6508 (9.9791)	3.1355 (4.5168)	1.6861 (7.7032)	2.2059 (19.5851)
PageRank (Directed)	0.0000 (0.0000)	0.0001 (0.0001)	0.0001 (0.0003)	0.0000 (0.0000)	0.0000 (0.0001)	0.0000 (0.0003)
PageRank (Undirected)	0.0001 (0.0001)	0.0000 (0.0001)	0.0000 (0.0004)	0.0000 (0.0000)	0.0000 (0.0001)	0.0000 (0.0002)
Eigenvector Centrality (Directed)	0.0000 (0.0000)	0.0003 (0.0084)	0.0002 (0.0068)	0.0000 (0.0000)	0.0005 (0.0062)	0.0004 (0.0060)
Eigenvector Centrality (Undirected)	0.0027 (0.0104)	0.0003 (0.0049)	0.0001 (0.0074)	0.0021 (0.0061)	0.0001 (0.0013)	0.0001 (0.0089)
Closeness Centrality (Directed)	0.0000 (0.0000)	0.0000 (0.0002)	0.0000 (0.0005)	0.0000 (0.0000)	0.0001 (0.0006)	0.0001 (0.0009)
Closeness Centrality (Undirected)	0.0064 (0.0123)	0.0020 (0.0066)	0.0003 (0.0019)	0.0428 (0.0369)	0.0490 (0.0303)	0.0094 (0.0221)
Clustering Coefficient (Directed)	0.0032 (0.0348)	0.0019 (0.0287)	0.0023 (0.0288)	0.0070 (0.0469)	0.0103 (0.0656)	0.0075 (0.0444)
Clustering Coefficient (Undirected)	0.0065 (0.0690)	0.0038 (0.0543)	0.0046 (0.0558)	0.0140 (0.0935)	0.0198 (0.1230)	0.0145 (0.0835)
Betweenness Centrality (Directed)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Betweenness Centrality (Undirected)	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0002)	0.0001 (0.0007)	0.0001 (0.0012)	0.0002 (0.0025)

Notes: Mean values of centrality measures with standard deviations in parentheses. For yearly networks, statistics are computed by first averaging across years for each node, then by node category. Static networks show direct statistics by node category.

Table A.2: List of Firm-level Variables from *Infocamere* by Category

Firm Characteristics	Financial Indicators	Balance Sheet: Assets	Balance Sheet: Liabilities & Equity	Income Statement
size	i_valpro	aa_crediti_verso_soci	ba_totale_patrimonio_netto	ea_totale_valore_della_produzione
statute	i_valagg	ab_totale_immobilizzazioni	ba1_capitale_sociale	ea01_ricavi_vendite_e_prestazioni
comune	i_marope	ab1_totale_immobilizzazioni_immateriali	ba2_riserva_da_sovrapprezzo	ea02_variazioni_rimanenze_prodotti
provincia	i_risnope	ab2_totale_immobilizzazioni_materiali	ba3_riserva_di_rivaltizzazione	ea03_variazione_lavori_in_corso_sul_ordinazione
regione	i_risngca	ab3_totale_immobilizzazioni_finanziarie	ba4_riserva_legale	ea04_incrementi_di_immobilizz_per_lav_interni
n_addetti_tot	i_risnam	ac_attivo_circolante	ba5_riserva_statutaria	ea05_altri_ricavi
status_attivita	i_rismet	ac1_totale_rimanenze	ba6_riserva_azioni_proprie	eb_totale_costi_della_produzione
c_s_attivita	i_immob	ac2_totale_crediti	ba7_altre_riserve	eb06_acquisti_di_materie
c_attivita	i_rimane	ac3_totale_attivita_finanziarie	ba8_utile_riserva	eb07_servizi
c_ateco_bene	i_crediti	ac4_disponibilita_liquide	ba9_utile_perdita_a_nuovo	eb08_godimento_di_beni_di_terzi
gruppo_impresa	i_displiq	ad_ratei_e_risconti	ba9_utile_perdita_di_esercizio	eb09_costi_per_il_personale
ateco_letter	i_totatt	at_totale_attivo	bb_totale_fondi_rischi	eb10_ammortamenti_e svalutazioni
industry	i_patrnet		bb1_fondo_di_quiescenza	eb11_variazione_rimanenze_materie_prime
	i_passecons		bb2_fondo_imposte	eb12_accantonamenti_per_rischi
	i_passecorr		bb3_altri_fondi	eb13_altri_accantonamenti
	i_totpass		bc_trattamento_di_fine_rapporto	eb14_oneri_diversi_di_gestione
	i_roe		bd_totale_debiti	eb15_altri_proventi_e_oneri_finanziari
	i_roic		bd1_obbligazioni	ec15_proventi_da_partecipazioni
	i_indfin		bd10_debiti_verso_imprese_collegate	ec16_altri_proventi_finanziari
	i_copium		bd11_debiti_verso_controllanti	ec17_interessi_e_altri_oneri_finanziari
	i_liqum		bd2_obbligazioni_convertibili	ed18_rivalutazioni
			bd4_totale_debiti_vs_banche	ed19_svalutazioni
			bd5_totale_debiti_vs_altri_finanziatori	ec_totale_proventi_e_oneri_straordinari
			bd6_acconti_anticipi_da_clienti	ef00_risultato_prima_delle_imposte
			bd7_totale_debiti_vs_fornitori	ef22_imposte_sul_reddito
			bd9_debiti_verso_imprese_controllate	ef23_utili_perdita_di_esercizio
			bdx_totale_debiti_commerciali	
			bdy_totale_debiti_residui	
			bdz_a_totale_debiti_entro_esercizio	
			bdz_b_totale_debiti_oltre_esercizio	
			be_ratei_e_risconti	
			bt_totale_passivo	

Notes: The table reports the firm-level features used for the predictive modeling, categorized into five groups: Firm Characteristics, Financial Indicators, Balance Sheet Assets, Balance Sheet Liabilities & Equity, and Income Statement. Each variable is represented by its code as per the *Infocamere* database.

Table A.3: Model Performance Comparison Across Feature Sets, $h = 2$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.807	0.805	0.818	0.814
Precision	0.952	0.988	0.974	0.971
Recall	0.681	0.651	0.685	0.679
F1	0.794	0.785	0.805	0.800
AUC	0.901	0.891	0.889	0.897
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.553	0.774	0.610	0.770
Precision	0.786	0.724	0.904	0.717
Recall	0.251	0.948	0.321	0.958
F1	0.380	0.821	0.473	0.820
AUC	0.640	0.847	0.758	0.744
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.883	0.907	0.924	0.904
Precision	0.971	0.981	0.952	0.968
Recall	0.810	0.846	0.908	0.852
F1	0.883	0.909	0.929	0.906
AUC	0.966	0.968	0.976	0.972

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.4: Model Performance Comparison Across Feature Sets, $h = 3$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.806	0.807	0.817	0.816
Precision	0.906	0.990	0.973	0.991
Recall	0.706	0.642	0.673	0.658
F1	0.794	0.779	0.796	0.791
AUC	0.878	0.877	0.891	0.894
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.673	0.777	0.774	0.774
Precision	0.635	0.727	0.719	0.723
Recall	0.898	0.925	0.942	0.929
F1	0.744	0.814	0.815	0.813
AUC	0.688	0.743	0.732	0.743
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.872	0.882	0.914	0.897
Precision	0.957	0.963	0.945	0.971
Recall	0.794	0.808	0.890	0.829
F1	0.868	0.879	0.916	0.894
AUC	0.960	0.965	0.976	0.970

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.5: Model Performance Comparison Across Feature Sets (Sampled Years), $h = 0$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.816	0.871	0.864	0.869
Precision	0.646	0.824	0.777	0.812
Recall	0.742	0.681	0.715	0.687
F1	0.691	0.745	0.744	0.745
AUC	0.877	0.889	0.902	0.902
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.429	0.617	0.616	0.613
Precision	0.322	0.417	0.415	0.415
Recall	0.956	0.953	0.941	0.964
F1	0.482	0.580	0.576	0.580
AUC	0.582	0.808	0.766	0.767
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.880	0.925	0.930	0.930
Precision	0.765	0.902	0.895	0.912
Recall	0.818	0.820	0.848	0.829
F1	0.791	0.859	0.871	0.868
AUC	0.937	0.957	0.972	0.972

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. The synthetic reference year for non-confiscated firms has been obtained sampling from the EDF of confiscation years, see [A.9](#). Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.6: Model Performance Comparison Across Feature Sets (Sampled Years), $h = 1$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.830	0.868	0.868	0.856
Precision	0.668	0.813	0.805	0.755
Recall	0.744	0.670	0.681	0.696
F1	0.704	0.735	0.738	0.724
AUC	0.879	0.879	0.893	0.896
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.458	0.647	0.646	0.646
Precision	0.320	0.433	0.431	0.431
Recall	0.879	0.966	0.951	0.958
F1	0.469	0.598	0.593	0.595
AUC	0.597	0.807	0.775	0.785
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.881	0.928	0.928	0.926
Precision	0.765	0.914	0.912	0.902
Recall	0.812	0.812	0.814	0.818
F1	0.788	0.860	0.860	0.858
AUC	0.937	0.958	0.969	0.969

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. The synthetic reference year for non-confiscated firms has been obtained sampling from the EDF of confiscation years, see [A.9](#). Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.7: Model Performance Comparison Across Feature Sets (Sampled Years), $h = 2$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.835	0.873	0.869	0.873
Precision	0.669	0.814	0.798	0.798
Recall	0.748	0.673	0.677	0.699
F1	0.706	0.737	0.733	0.745
AUC	0.875	0.876	0.892	0.892
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.493	0.671	0.671	0.665
Precision	0.331	0.444	0.443	0.439
Recall	0.893	0.959	0.947	0.951
F1	0.483	0.607	0.604	0.601
AUC	0.623	0.817	0.790	0.790
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.892	0.931	0.929	0.928
Precision	0.783	0.924	0.901	0.907
Recall	0.818	0.808	0.821	0.812
F1	0.800	0.862	0.859	0.857
AUC	0.938	0.959	0.968	0.970

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. The synthetic reference year for non-confiscated firms has been obtained sampling from the EDF of confiscation years, see [A.9](#). Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.8: Model Performance Comparison Across Feature Sets (Sampled Years), $h = 3$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.863	0.872	0.875	0.874
Precision	0.748	0.811	0.810	0.825
Recall	0.712	0.662	0.678	0.656
F1	0.730	0.729	0.738	0.731
AUC	0.886	0.878	0.891	0.885
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.493	0.697	0.696	0.697
Precision	0.333	0.460	0.459	0.460
Recall	0.950	0.965	0.961	0.959
F1	0.493	0.623	0.622	0.621
AUC	0.670	0.831	0.833	0.815
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.892	0.929	0.933	0.930
Precision	0.768	0.913	0.894	0.906
Recall	0.837	0.802	0.843	0.817
F1	0.801	0.854	0.868	0.859
AUC	0.946	0.959	0.970	0.970

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. The synthetic reference year for non-confiscated firms has been obtained sampling from the EDF of confiscation years, see [A.9](#). Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.9: Model Performance Comparison Across Feature Sets (No Lag Features), $h = 0$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.795	0.794	0.811	0.815
Precision	0.898	0.994	0.965	0.968
Recall	0.725	0.645	0.695	0.701
F1	0.802	0.782	0.808	0.813
AUC	0.874	0.886	0.892	0.898
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.655	0.677	0.668	0.669
Precision	0.637	0.645	0.639	0.643
Recall	0.928	0.971	0.971	0.951
F1	0.755	0.775	0.771	0.768
AUC	0.587	0.735	0.693	0.657
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.871	0.883	0.854	0.868
Precision	0.954	0.972	0.875	0.919
Recall	0.814	0.820	0.871	0.844
F1	0.879	0.890	0.873	0.880
AUC	0.951	0.953	0.935	0.945

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. All lagged features have been excluded from the feature sets. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.10: Model Performance Comparison Across Feature Sets (No Lag Features), $h = 1$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.796	0.794	0.792	0.809
Precision	0.894	0.991	0.965	0.967
Recall	0.724	0.640	0.654	0.685
F1	0.800	0.778	0.779	0.802
AUC	0.876	0.878	0.889	0.892
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.556	0.683	0.682	0.680
Precision	0.814	0.644	0.646	0.642
Recall	0.274	0.977	0.963	0.973
F1	0.410	0.776	0.773	0.774
AUC	0.596	0.754	0.580	0.698
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.868	0.896	0.869	0.851
Precision	0.960	0.969	0.860	0.884
Recall	0.798	0.843	0.916	0.847
F1	0.872	0.902	0.887	0.865
AUC	0.951	0.949	0.947	0.938

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. All lagged features have been excluded from the feature sets. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.11: Model Performance Comparison Across Feature Sets (No Lag Features), $h = 2$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.805	0.818	0.807	0.820
Precision	0.902	1.000	0.974	0.956
Recall	0.721	0.667	0.665	0.703
F1	0.802	0.800	0.790	0.811
AUC	0.880	0.901	0.894	0.897
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.521	0.679	0.690	0.689
Precision	0.552	0.638	0.642	0.642
Recall	0.655	0.952	0.980	0.972
F1	0.599	0.764	0.776	0.774
AUC	0.607	0.610	0.622	0.621
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.888	0.897	0.904	0.887
Precision	0.963	0.976	0.960	0.965
Recall	0.828	0.832	0.860	0.824
F1	0.890	0.898	0.907	0.889
AUC	0.955	0.958	0.968	0.963

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. All lagged features have been excluded from the feature sets. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.12: Model Performance Comparison Across Feature Sets (No Lag Features), $h = 3$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.809	0.815	0.809	0.825
Precision	0.882	0.997	0.966	0.979
Recall	0.735	0.652	0.660	0.683
F1	0.802	0.788	0.785	0.805
AUC	0.877	0.882	0.890	0.892
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.672	0.693	0.690	0.697
Precision	0.624	0.641	0.639	0.643
Recall	0.952	0.950	0.946	0.958
F1	0.754	0.766	0.763	0.770
AUC	0.627	0.612	0.631	0.619
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.893	0.910	0.877	0.888
Precision	0.961	0.974	0.911	0.950
Recall	0.831	0.852	0.850	0.831
F1	0.892	0.909	0.879	0.887
AUC	0.953	0.958	0.949	0.956

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. All lagged features have been excluded from the feature sets. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.13: Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 0$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.761	0.813	0.832	0.814
Precision	0.724	0.883	0.919	0.877
Recall	0.794	0.694	0.702	0.702
F1	0.757	0.777	0.796	0.780
AUC	0.871	0.878	0.891	0.893
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.574	0.738	0.714	0.708
Precision	0.623	0.668	0.647	0.643
Recall	0.228	0.875	0.859	0.844
F1	0.334	0.758	0.738	0.730
AUC	0.620	0.805	0.792	0.788
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.850	0.890	0.901	0.906
Precision	0.850	0.912	0.945	0.944
Recall	0.826	0.848	0.838	0.850
F1	0.838	0.879	0.888	0.895
AUC	0.934	0.960	0.967	0.971

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.14: Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 1$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.805	0.828	0.820	0.825
Precision	0.817	0.908	0.870	0.896
Recall	0.742	0.697	0.717	0.702
F1	0.778	0.789	0.786	0.787
AUC	0.885	0.872	0.881	0.885
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.580	0.675	0.719	0.708
Precision	0.635	0.655	0.647	0.692
Recall	0.206	0.624	0.858	0.661
F1	0.311	0.639	0.738	0.676
AUC	0.640	0.802	0.753	0.797
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.854	0.907	0.896	0.890
Precision	0.858	0.965	0.934	0.945
Recall	0.818	0.829	0.832	0.808
F1	0.838	0.892	0.880	0.872
AUC	0.933	0.963	0.960	0.962

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.15: Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 2$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.780	0.845	0.835	0.847
Precision	0.750	0.894	0.886	0.908
Recall	0.762	0.741	0.725	0.731
F1	0.756	0.810	0.798	0.810
AUC	0.877	0.865	0.885	0.892
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.601	0.744	0.742	0.743
Precision	0.643	0.655	0.651	0.653
Recall	0.241	0.903	0.911	0.910
F1	0.351	0.759	0.759	0.760
AUC	0.694	0.799	0.793	0.800
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.877	0.904	0.901	0.906
Precision	0.875	0.948	0.926	0.955
Recall	0.846	0.832	0.847	0.828
F1	0.860	0.886	0.885	0.887
AUC	0.941	0.959	0.969	0.972

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC). All the metrics refer to the best performing hyperparameter configuration identified via cross-validation on the training set.

Table A.16: Robustness (2016 Train-Test Split) - Model Performance Comparison Across Feature Sets, $h = 3$

<i>Panel A: Firm-Level Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.790	0.859	0.849	0.856
Precision	0.752	0.942	0.908	0.930
Recall	0.766	0.717	0.723	0.720
F1	0.759	0.814	0.805	0.812
AUC	0.872	0.863	0.885	0.889
<i>Panel B: Network Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.628	0.772	0.771	0.762
Precision	0.540	0.678	0.673	0.670
Recall	0.940	0.899	0.916	0.884
F1	0.686	0.773	0.775	0.762
AUC	0.668	0.847	0.842	0.800
<i>Panel C: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.887	0.887	0.905	0.907
Precision	0.879	0.873	0.893	0.917
Recall	0.855	0.865	0.886	0.863
F1	0.867	0.869	0.890	0.889
AUC	0.942	0.954	0.973	0.969

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric. Panel A uses only firm-level features, Panel B uses only network features, and Panel C combines both feature sets. The evaluation metrics include Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC).

Table A.17: Edge Event Breakdown and Link Duration Statistics

Metric	Value
Total edge events	120,800
Link creations	50,685
Link updates	20,199
Link deletions	49,916
Creation/Deletion ratio	1.02
Updates/Creations ratio	0.40
Links with complete lifecycle	49,910
Average link duration (days)	1859.07
Std deviation (days)	1808.20
Average link duration (years)	5.09
Std deviation (years)	4.95

Notes: Link creations represent new connections established in the network. Link updates indicate modifications to existing connections (e.g., changes in relationship type or strength). Link deletions occur when connections are dissolved or removed from the network. Link duration is calculated only for links with both creation and deletion events, representing the complete lifecycle from establishment to dissolution. Duration statistics are computed in both days and years, where 1 year = 365.25 days.

Table A.18: Event Type Distribution Across Data Splits

Event Type	Train	Validation	Test	Total
Node events	280,762	60,915	53,603	395,280
Link creations	44,451	4,379	1,855	50,685
Link updates	15,851	3,135	1,213	20,199
Link deletions	32,865	7,949	9,102	49,916
Total events	373,929	76,378	65,773	516,080

Notes: Node events represent updates to firm-level attributes (e.g., financial data). Link creations are new connections established between firms. Link updates modify existing connections. Link deletions remove connections from the network. The data is split temporally into train, validation, and test sets for model evaluation.

Table A.19: Performance Comparison of TGN Model Variants

Metric	Standard TGN	Two-Tower TGN
AP	0.891 ± 0.001	0.873 ± 0.011
AUC	0.988 ± 0.000	0.987 ± 0.001
<i>Level Regression (creations + updates)</i>		
MAE	0.643 ± 0.004	0.649 ± 0.013
RMSE	0.933 ± 0.007	0.938 ± 0.007
<i>Delta Regression (updates + deletions)</i>		
MAE_Δ	0.727 ± 0.011	0.717 ± 0.008
RMSE_Δ	1.221 ± 0.011	1.215 ± 0.004

Note: Results averaged over 3 random seeds. All models trained for 30 epochs. Performance metrics for binary outcome (edge existence) are AP: Average Precision, AUC: Area Under ROC Curve. Regression for continuous outcome (stake level and change in stake). MAE/RMSE: regression metrics for absolute stake levels (creations+updates). MAE_Δ/RMSE_Δ: regression metrics for stake changes (updates+deletions). Training data: 373,929 events, Test data: 101,059 events.

Table A.20: Descriptive Statistics by Forecast Horizon

	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Observations	13,182	12,007	11,611	10,675
Mean Confiscation	0.1761	0.1554	0.1422	0.1284

Notes: The table reports descriptive statistics for the classification dataset across different forecast horizons. $h > 0$ indicates the number of years ahead for the prediction task. The number of observations decreases for higher forecast horizons due to data availability constraints.

Table A.21: Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 1$

<i>Panel A: Firm-level Data Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.768	0.764	0.750	0.771
Precision	0.621	0.628	0.604	0.659
Recall	0.442	0.381	0.314	0.381
F1	0.516	0.474	0.413	0.482
AUC	0.726	0.706	0.741	0.738
<i>Panel B: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.849	0.905	0.892	0.883
Precision	0.729	0.850	0.877	0.867
Recall	0.733	0.803	0.714	0.686
F1	0.731	0.826	0.787	0.766
AUC	0.904	0.942	0.947	0.944
<i>Panel C: TGN Embeddings</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.977	0.976	0.975	0.978
Precision	0.934	0.941	0.936	0.944
Recall	0.986	0.975	0.978	0.981
F1	0.959	0.958	0.957	0.962
AUC	0.998	0.998	0.998	0.999

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric.

Table A.22: Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 2$

<i>Panel A: Firm-level Data Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.659	0.771	0.759	0.757
Precision	0.410	0.622	0.591	0.581
Recall	0.592	0.394	0.359	0.367
F1	0.484	0.482	0.446	0.450
AUC	0.711	0.670	0.715	0.706
<i>Panel B: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.871	0.904	0.901	0.883
Precision	0.737	0.838	0.881	0.860
Recall	0.810	0.799	0.735	0.679
F1	0.772	0.818	0.801	0.759
AUC	0.918	0.938	0.953	0.946
<i>Panel C: TGN Embeddings</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.982	0.980	0.981	0.980
Precision	0.944	0.954	0.952	0.952
Recall	0.991	0.974	0.980	0.977
F1	0.967	0.964	0.966	0.964
AUC	0.999	0.998	0.998	0.998

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric.

Table A.23: Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 3$

<i>Panel A: Firm-level Data Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.642	0.776	0.757	0.766
Precision	0.367	0.607	0.513	0.544
Recall	0.617	0.264	0.322	0.338
F1	0.460	0.368	0.395	0.417
AUC	0.681	0.611	0.688	0.662
<i>Panel B: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.864	0.897	0.903	0.889
Precision	0.680	0.870	0.832	0.849
Recall	0.849	0.688	0.762	0.669
F1	0.755	0.768	0.795	0.748
AUC	0.899	0.938	0.947	0.944
<i>Panel C: TGN Embeddings</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.977	0.975	0.975	0.977
Precision	0.917	0.929	0.921	0.925
Recall	0.997	0.971	0.981	0.987
F1	0.955	0.950	0.950	0.955
AUC	0.998	0.997	0.997	0.998

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric.

Table A.24: Model Performance Comparison: Firm-level Data Only vs Firm + Network vs TGN Embeddings, $h = 4$

<i>Panel A: Firm-level Data Only</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.729	0.814	0.795	0.816
Precision	0.436	0.750	0.628	0.788
Recall	0.549	0.304	0.301	0.287
F1	0.486	0.433	0.407	0.421
AUC	0.749	0.647	0.710	0.731
<i>Panel B: Firm + Network</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.862	0.896	0.903	0.886
Precision	0.688	0.876	0.838	0.884
Recall	0.748	0.643	0.724	0.587
F1	0.717	0.742	0.777	0.706
AUC	0.895	0.935	0.954	0.948
<i>Panel C: TGN Embeddings</i>				
	Logistic Regression	Random Forest	LightGBM	XGBoost
Accuracy	0.976	0.971	0.974	0.972
Precision	0.913	0.909	0.915	0.906
Recall	0.993	0.976	0.979	0.983
F1	0.951	0.941	0.946	0.943
AUC	0.998	0.996	0.997	0.997

Notes: The table reports key classification performance metrics for four models across three feature sets using time-series split validation. Bold values indicate the best performing feature set for a given model and metric.

B Network Visualization: Software and Settings

To facilitate the exploration of the complex, high-dimensional network dataset constructed for this paper, we developed a custom interactive 3D visualization tool. The software is built in **React** and utilizes the **react-force-graph-3d** library to render the graph. The tool is designed to be accessible to a non-technical audience while providing deep analytical capabilities. Its primary features are detailed below.

B.1 Interactive Features and Data Exploration

The visualization is designed to be a fully interactive analytical tool, allowing users to explore the network’s structure and attributes in detail.

Dynamic and Static Views The tool supports two primary visualization modes. The *dynamic view* allows users to navigate through time using a slider, displaying a snapshot of the network for any given year. This feature is crucial for observing the evolution of ownership structures, the formation of new clusters, and the changing roles of specific entities over the three-decade period. The *static view* aggregates all nodes and links across the entire period into a single graph, providing a comprehensive overview of the network’s full scale and identifying entities that are central over the long run.

Interactive Information Panels The graph is richly attributed and fully interactive. Clicking on any node (a firm or an individual) or edge (an ownership link) opens a detailed information panel.

- **Node Information:** For a selected firm, the panel displays key attributes such as its industry sector, size, location, and its complete judicial history, including the status and date of confiscation.
- **Edge Information:** For a selected ownership link, the panel details the ownership percentage, the corresponding monetary value of the stake, and the type of entity at the source of the link.

Visual Symbology To convey information intuitively, the rendering uses a clear visual symbology. Node colors distinguish between individuals, non-confiscated firms, and firms at various stages of the confiscation process (e.g., pending, final, or revoked). The thickness of an edge is directly proportional to the percentage of the ownership stake it

represents, making it easy to identify significant control relationships at a glance.

Component Highlighting and Search The tool includes analytical features to aid exploration. A search function allows for the immediate identification and navigation to any specific node in the network. Furthermore, hovering the cursor over a node or link automatically highlights the entire weakly connected component it belongs to, providing an immediate sense of the size and scope of a particular subgroup within the larger network.

B.2 Network Layout and Physics Simulation

The spatial arrangement of the nodes in the 3D space is determined by a force-directed layout algorithm, but it is specifically designed to ensure temporal consistency for tracking changes over time. This physics-based approach places connected entities closer together while pushing apart unconnected ones, which helps to visually reveal the underlying cluster structure of the network.

A Master Layout for Temporal Consistency To make meaningful comparisons between different yearly snapshots, the visualization is anchored on a single, consistent layout. This is achieved by first generating a force-directed layout for the complete *static graph*, which aggregates all nodes and links across the entire time period. The resulting (x, y, z) coordinate for each node is then stored and used as its fixed “master” position throughout the visualization. When a user navigates to a specific year, nodes present in that snapshot are displayed at their master coordinates. This methodology is crucial as it ensures that a node’s position remains constant from one year to the next, preventing the layout from “jumping” and allowing for a clear and intuitive analysis of how the network evolves. For this reason, the computationally expensive live physics simulation is disabled during normal use, though nodes can still be dragged for inspection. In the absence of pre-calculated positions (for instance, when analyzing a new dataset), the tool can run a live, aggressive physics simulation to generate a layout from scratch.

C Methodological Details: Machine-Learning Models

This appendix provides a detailed description of the machine learning models, hyperparameter tuning strategy, and data balancing techniques employed in the classification task.

C.1 Models

Three tree-based ensemble models were selected for their high performance in classification tasks: Random Forest, LightGBM, and XGBoost. These models are implemented within a `scikit-learn` pipeline framework.

Random Forest The Random Forest model is an ensemble learning method that operates by constructing a multitude of decision trees at training time. For classification, the final output is the class selected by the most trees. It mitigates the overfitting tendencies of individual decision trees by employing two key techniques:

1. **Bagging (Bootstrap Aggregating):** Each tree is trained on a different bootstrap sample (random sample with replacement) of the training data.
2. **Feature Randomness:** When splitting a node, each tree considers only a random subset of the features, which decorrelates the trees and further reduces variance.

This combination results in a robust model that generalizes well to new data.

LightGBM LightGBM is a high-performance gradient boosting framework based on decision trees. It belongs to the family of Gradient Boosting Decision Tree (GBDT) algorithms, which build models in a sequential, stage-wise fashion. Each new tree corrects the errors of its predecessor. LightGBM’s key innovations for speed and efficiency are:

1. **Gradient-based One-Side Sampling (GOSS):** To reduce the number of data instances used for computing gradients, GOSS retains all instances with large gradients (i.e., those that are poorly trained) and randomly samples from instances with small gradients.
2. **Exclusive Feature Bundling (EFB):** It bundles mutually exclusive features (i.e., features that rarely take non-zero values simultaneously) to reduce the feature dimensionality.

These techniques allow LightGBM to train significantly faster than other boosting algorithms, especially on large datasets.

XGBoost XGBoost, which stands for Extreme Gradient Boosting, is another highly efficient and scalable implementation of the GBDT algorithm. It is renowned for its performance in machine learning competitions. Key features include:

1. **Regularization:** It incorporates both L1 (Lasso) and L2 (Ridge) regularization terms in its objective function, which penalizes model complexity and helps prevent overfitting.
2. **Sparsity Awareness:** It handles missing values efficiently by learning default directions for them in each tree node.
3. **Parallelization:** It allows for parallel processing during tree construction, significantly speeding up the training process.

The metric used for evaluation is *logarithmic loss*.

C.2 Hyperparameters and Randomized Cross-Validation

To optimize model performance, we tuned key hyperparameters using `scikit-learn`'s randomized search with time-series cross-validation. Instead of exhaustively searching all possible hyperparameter combinations like a grid search, randomized search samples a fixed number of parameter settings (i.e. 10 iterations) from specified probability distributions. This is computationally more efficient and often yields comparable results.

The cross-validation strategy considers 5 splits. This is crucial for our temporal dataset, as it ensures that the training data always precedes the validation data in each fold, mimicking a real-world scenario where the model is trained on past data to predict future outcomes. The performance of each hyperparameter combination is evaluated using the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) score, which is well-suited for imbalanced classification tasks.

The hyperparameter search spaces for each model were defined as follows:

- **Random Forest**

- `n_estimators`: Number of trees in the forest [100, 200, 300].

- `max_depth`: Maximum depth of the trees [10, 20, 30, None].
- `min_samples_split`: Minimum number of samples required to split an internal node [2, 5, 10].

- **LightGBM**

- `n_estimators`: Number of boosting rounds [100, 200, 400].
- `learning_rate`: Step size shrinkage to prevent overfitting [0.01, 0.1, 0.2].
- `num_leaves`: Maximum number of leaves in one tree [20, 31, 40].

- **XGBoost**

- `n_estimators`: Number of boosting rounds [100, 200, 300].
- `learning_rate`: Step size shrinkage [0.01, 0.1, 0.2].
- `max_depth`: Maximum depth of a tree [3, 5, 7].

C.3 Synthetic Minority Over-sampling TEchnique (SMOTE)

Our dataset exhibits a significant class imbalance, where the number of non-confiscated firms (majority class) far exceeds the number of confiscated firms (minority class). To address this, we integrated the Synthetic Minority Over-sampling TEchnique (SMOTE) (Chawla et al., 2002), from the `imblearn` library, into our training pipeline. SMOTE is applied only to the training data within each cross-validation fold to prevent data leakage into the validation set.

The SMOTE algorithm works by creating synthetic samples for the minority class instead of simply duplicating existing ones (Chawla et al., 2002). This helps the classifier learn a more generalized decision boundary. The process is as follows:

1. For each instance $\vec{x}_i$ in the minority class, its k -nearest neighbors from the same class are identified.
2. One of these neighbors, $\vec{x}_{zi}$, is randomly selected.
3. A new synthetic sample, $\vec{x}_{new}$, is generated at a random point along the line segment connecting the original instance and its selected neighbor.

Mathematically, a new synthetic instance is created using the formula:

$$\vec{x}_{new} = \vec{x}_i + \lambda \times (\vec{x}_{zi} - \vec{x}_i)$$

where λ is a random number uniformly distributed in the interval $[0, 1]$. This procedure effectively populates the feature space around existing minority class samples, creating a larger and less sparse decision region for the minority class, which helps the model to better learn its characteristics.

D Model Interpretability with SHAP

To understand the drivers behind the predictions of our best-performing models (LightGBM and XGBoost), we employed SHAP (SHapley Additive exPlanations) (Lundberg et al., 2020). This appendix details the theoretical underpinnings of SHAP and its practical implementation in our analysis. SHAP is a unified approach to explaining the output of any machine learning model, based on the principles of cooperative game theory and their classical Shapley values. The core idea is to treat features as “players” in a game, where the “payout” is the model’s prediction. SHAP values represent the fairest way to distribute this payout among the features, indicating the contribution of each feature to the prediction for a specific instance.

A Shapley value is the average marginal contribution of a feature value across all possible feature combinations (coalitions). It is the only allocation method that satisfies three desirable properties:

1. **Local Accuracy (Efficiency):** The sum of the SHAP values for all features for a given instance is equal to the difference between the model’s prediction for that instance and the base value (the average prediction over the entire dataset). This is expressed as:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i$$

where $f(x)$ is the model’s prediction for instance x , ϕ_0 is the base value, M is the number of features, and ϕ_i is the SHAP value for feature i .

2. **Missingness:** A feature that has no impact on the model’s prediction receives a SHAP value of 0.
3. **Consistency:** If a model changes so that a feature’s marginal contribution increases or stays the same (regardless of other features), its SHAP value will not decrease.

The Shapley value for a feature i is computed by considering every possible subset S of features that does not include feature i . The marginal contribution of adding feature i to the subset S is calculated. This is then averaged over all possible subsets S :

$$\phi_i(f, x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)]$$

Here, F is the full set of features, and $f_x(S)$ is the model prediction for instance x given only the feature values in subset S . Our analysis was implemented using the **shap** library in Python.

E Temporal Graph Network

This section provides a formal description of the Temporal Graph Network (TGN) architecture adopted in this study. Let $\mathcal{G}_t = (V, E_t)$ denote a temporal directed graph where E_t is the set of events observed up to time t . Each event is a quadruple $(u, v, t, \mathbf{x}_{uv}^t)$ representing an interaction between source node u and target node v at time t with associated edge features $\mathbf{x}_{uv}^t$.

Event Routing and Attention-based Message Passing

Each temporal event is a quadruple $(u, v, t, \mathbf{x}_{uv}^t)$. The model first performs *event routing*: if $u = v$ the event is treated as a *self-loop update* and no graph message passing or scoring is performed; if $u \neq v$ the full TGN pipeline is applied.

Memory Read For any node $i \in \{u, v\}$, let

$$\mathbf{h}_i^{t-} = \text{Memory}[i]$$

denote its state immediately prior to event time t .

Case $u \neq v$: Edge Events with Attention

Historical Neighbor Construction For each endpoint $i \in \{u, v\}$ the model retrieves the most recent K temporal neighbors $\mathcal{N}_K(i, t)$. For a past interaction (i, j, t_j) with $t_j < t$, the neighbor message is

$$\mathbf{m}_{i,j}^{\text{hist}} = \left[\mathbf{x}_{ij}^{t_j} \parallel \phi(t - t_j) \parallel \mathbf{h}_j^{t_j-} \right],$$

where ϕ denotes the temporal encoding. A permutation-invariant aggregator produces

$$\mathbf{M}_i^{\text{hist}} = \text{Agg}\{\mathbf{m}_{i,j}^{\text{hist}} : j \in \mathcal{N}_K(i, t)\}.$$

Attention Mechanism The attention module is implemented using the **TransformerConv** operator from PyTorch Geometric (Shi et al., 2021), which provides a graph-based vari-

ant of query-key-value attention.²⁵

For node u , the input representation is

$$\mathbf{s}_u^t = \left[\mathbf{h}_u^{t^-} \parallel \mathbf{M}_u^{\text{hist}} \right],$$

while each neighbor $i \in \mathcal{N}_K(u, t)$ contributes

$$\mathbf{s}_{u,i}^{t_i} = \left[\mathbf{h}_i^{t_i^-} \parallel \mathbf{m}_{u,i}^{\text{hist}} \right].$$

A single-head attention mechanism is used. Queries, keys, and values are computed as

$$\mathbf{q}_u = W_Q \mathbf{s}_u^t, \quad \mathbf{k}_{u,i} = W_{\text{neigh}} \mathbf{s}_{u,i}^{t_i}, \quad \mathbf{v}_{u,i} = W_V \mathbf{s}_{u,i}^{t_i},$$

with $W_Q, W_{\text{neigh}}, W_V$ learnable matrices.

Attention weights are

$$\alpha_{u,i} = \frac{\exp\left(\mathbf{q}_u^\top \mathbf{k}_{u,i} / \sqrt{d}\right)}{\sum_{j \in \mathcal{N}_K(u,t)} \exp\left(\mathbf{q}_u^\top \mathbf{k}_{u,j} / \sqrt{d}\right)}.$$

The neighbor context vector is

$$\mathbf{c}_u = \sum_{i \in \mathcal{N}_K(u,t)} \alpha_{u,i} \mathbf{v}_{u,i}.$$

The embedding of node u at time t combines a self-term and an attention-based neighbor term:

$$\mathbf{z}_u^t = W_{\text{self}} \mathbf{m}_u^{t^-} + W_{\text{neigh}} \mathbf{c}_u.$$

The embedding $\mathbf{z}_v^t$ is computed analogously.

Scoring Architectures and Link Prediction Loss The model supports two alternative scoring architectures for link prediction. The first is a single-tower architecture, in which the ordered concatenation

$$[\mathbf{z}_u^t \parallel \mathbf{z}_v^t]$$

²⁵Unlike the original Transformer self-attention mechanism introduced in Vaswani et al. (2017), **TransformerConv** performs attention over a node’s graph neighborhood rather than over a full sequence and omits components such as multi-layer feed-forward blocks and positional encodings. It can therefore be viewed as a localized and structurally constrained adaptation of Transformer-style attention to graph data.

is passed through a Multi-Layer Perceptron (MLP)²⁶ to produce

$$\hat{y}_{uv}^t = \sigma(\text{MLP}([\mathbf{z}_u^t \parallel \mathbf{z}_v^t])) .$$

Because the MLP receives an *ordered* pair, the resulting score is directional, i.e. $\hat{y}_{uv}^t \neq \hat{y}_{vu}^t$ in general.

The second scoring architecture is a two-tower model (Covington et al., 2016), which computes

$$\hat{y}_{uv}^t = \sigma(f_u(\mathbf{z}_u^t)^\top f_v(\mathbf{z}_v^t)) ,$$

where f_u and f_v are separate feed-forward networks applied to the source embedding and target embedding, respectively. Since the two towers are distinct and the model always evaluates the ordered pair (u, v) , the score is again directional. This factorized formulation enables additional flexibility by allowing asymmetric transformations and gradients for the two endpoints. The benefit of this architecture is that it induces three related but distinct embedding spaces: a shared base embedding space $\mathbf{z}$, and two task-specific transformed spaces, one for nodes acting as sources (e.g., investors) and one for nodes acting as targets (e.g., investees).

In both architectures, the link prediction loss is the binary cross-entropy with negative sampling,

$$\mathcal{L}_{\text{link}} = -\log \hat{y}_{uv}^t - \sum_{(u,n) \in \mathcal{N}^{\text{neg}}} \log(1 - \hat{y}_{un}^t) .$$

Regression Heads Beyond link prediction, the model also produces node-level regression outputs by attaching separate regression heads $g(\cdot)$ to the node embeddings. For each node i with regression target y_i^t (e.g. firm-level economic variables), the model predicts

$$\hat{y}_i^t = g(\mathbf{z}_i^t) ,$$

and the corresponding regression loss is

$$\mathcal{L}_{\text{reg}} = \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \ell(\hat{y}_i^t, y_i^t) ,$$

where ℓ is either the mean absolute error (MAE) or the squared error used to compute

²⁶An MLP, or Multi-Layer Perceptron, is a type of artificial neural network with at least three layers: an input layer, one or more hidden layers, and an output layer.

RMSE, depending on the task. Regression heads share the same embeddings as the link prediction component but do not alter the event memory update mechanism.

Barlow Twins Loss To further regularize the embeddings, we include a Barlow Twins-style redundancy-reduction term. For each node i , we generate two stochastic views of its embedding, $\mathbf{z}_i^{(1),t}$ and $\mathbf{z}_i^{(2),t}$ (e.g. via dropout²⁷), and form the empirical cross-correlation matrix $C \in \mathbb{R}^{d \times d}$ with entries

$$C_{ab} = \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} z_{i,a}^{(1),t} z_{i,b}^{(2),t}.$$

The Barlow Twins loss is

$$\mathcal{L}_{\text{BT}} = \sum_a (1 - C_{aa})^2 + \lambda_{\text{off}} \sum_{a \neq b} C_{ab}^2,$$

which enforces invariance across views on the diagonal and discourages redundancy across dimensions on the off-diagonal.

Temporal Consistency Loss Finally, we impose a temporal smoothness constraint on node embeddings to penalize abrupt changes over time. For nodes observed at consecutive instants, the temporal loss is

$$\mathcal{L}_{\text{temp}} = \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left\| \mathbf{z}_i^t - \mathbf{z}_i^{t-} \right\|_2^2,$$

where $\mathbf{z}_i^{t-}$ denotes the embedding of node i immediately prior to time t (when defined).

Memory Update Each endpoint receives an event-level message, e.g.

$$\mathbf{m}_u^t = \left[\mathbf{x}_{uv}^t \parallel \phi(t - t_u^{\text{last}}) \parallel \mathbf{z}_u^t \right],$$

and the memory is updated via

$$\mathbf{h}_u^t = \text{GRU}\left(\mathbf{h}_u^{t-}, \mathbf{m}_u^t\right), \quad \mathbf{h}_v^t = \text{GRU}\left(\mathbf{h}_v^{t-}, \mathbf{m}_v^t\right).$$

Total Loss The total loss at time t combines link prediction, regression, Barlow Twins,

²⁷Dropout randomly zeroes out a subset of embedding dimensions, producing two stochastically different *views* of the same embedding vector.

and temporal consistency components as

$$\mathcal{L}_t = \mathcal{L}_{\text{link}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}} + \lambda_{\text{BT}} \mathcal{L}_{\text{BT}} + \lambda_{\text{temp}} \mathcal{L}_{\text{temp}},$$

with λ_{reg} , λ_{BT} , and λ_{temp} controlling the relative importance of each auxiliary objective.

Case $u = v$: Self-loop Events

Self-loop events update the memory of the involved node but do not trigger message passing, attention, scoring, or loss computation. For such events, the model forms a simplified update message $\mathbf{m}_u^t$ based solely on the event features and temporal encoding, and applies

$$\mathbf{h}_u^t = \text{GRU}(\mathbf{h}_u^{t-}, \mathbf{m}_u^t).$$

No embedding $\mathbf{z}_u^t$ is generated for prediction, and the event contributes no gradient to the scoring model.

Hyperparameters and Optimization

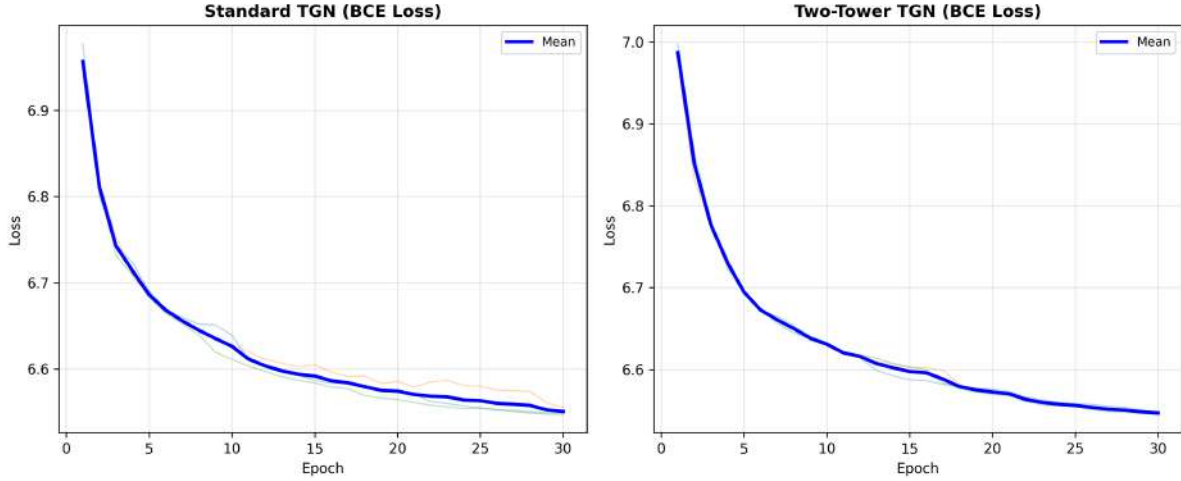
Unless otherwise stated, all experiments use the following hyperparameter configuration for the TGN model and training procedure.

Model architecture.

- Memory / embedding dimension: $d = 64$.
- Time encoding dimension: $d_{\text{time}} = 100$.
- Temporal neighbors per node: $K = 50$ (most recent neighbors).
- GNN layer: **TransformerConv** with
 - number of heads: $H = 2$,
 - dropout rate: 0.2,
 - input and output channels both equal to the memory dimension.
- Negative sampling: “filtered” sampler with 10 negatives per positive edge.

Training and optimization.

Figure E.25: Training Loss Comparison: Standard TGN vs. Two-Tower TGN

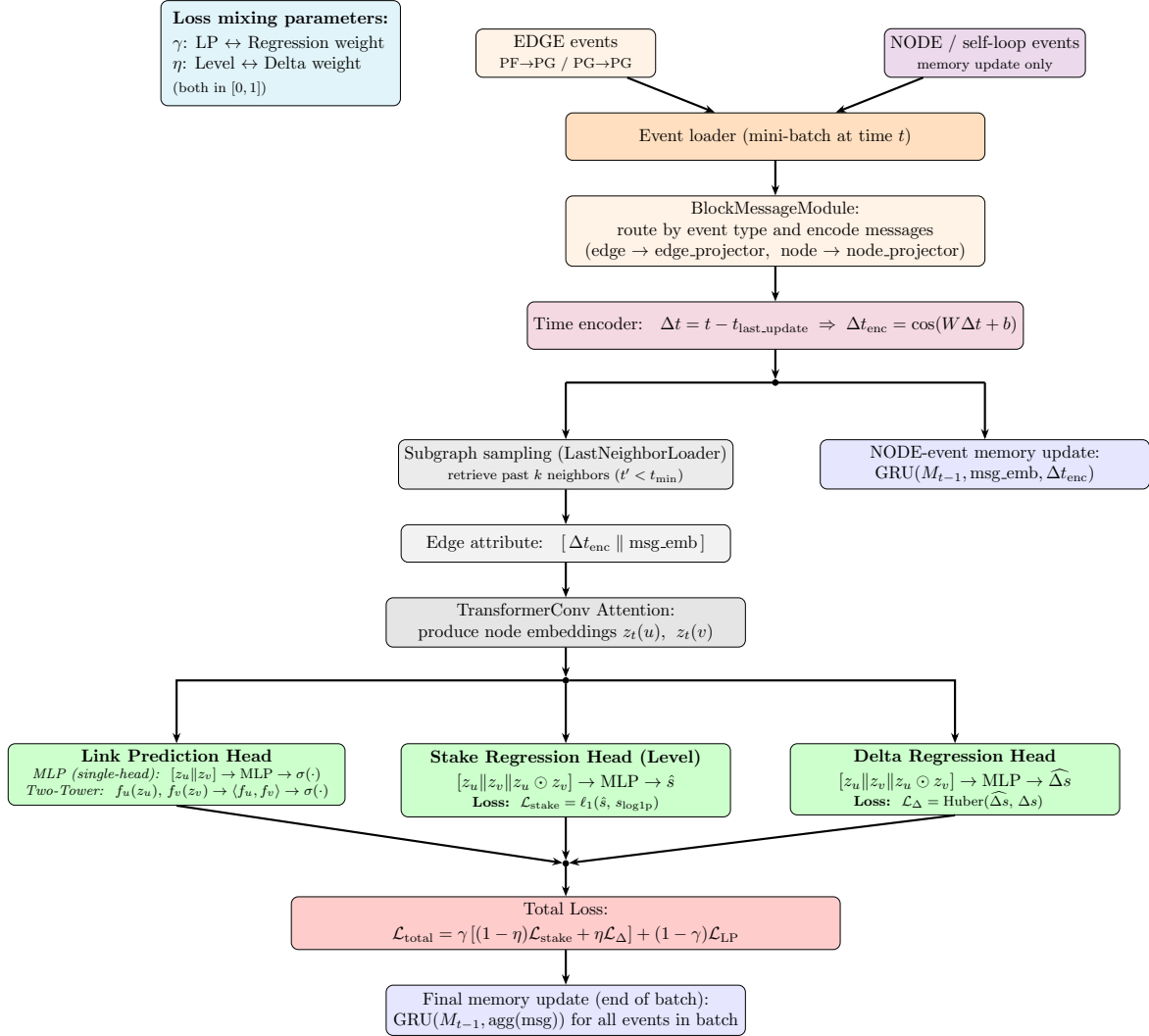


Notes: The figure illustrates the training loss of the two models, standard TGN on the left and Two-Tower TGN on the right over 30 epochs. The solid blue line represents the average across the three different seeds

- Optimizer: Adam with learning rate $\eta_0 = 10^{-3}$ and weight decay 10^{-5} .
- Batch size: 256 events.
- Number of epochs per model: 30.
- Gradient clipping: global ℓ_2 -norm clipped at 1.0.
- Learning rate scheduler: `ReduceLROnPlateau` on validation AP, with
- Regularization weights:
 - node-level Barlow Twins loss: $\lambda_{\text{node-contrast}} = 0.10$,
 - temporal consistency loss: $\lambda_{\text{temporal}} = 0.05$.

The model is implemented using PyTorch (Paszke et al., 2019) and PyTorch Geometric (Fey and Lenssen, 2019). Experiments are executed on a single NVIDIA V100 GPU equipped with 16 GB of memory. End-to-end training of one model takes about 45 minutes, and the architecture comprises roughly 220k trainable parameters. Figure E.26 illustrates the overall TGN architecture, while Figure E.25 compares the training loss of the standard TGN and the Two-Tower TGN over 30 epochs.

Figure E.26: Temporal Graph Network Architecture



Notes: The figure illustrates the architecture of the Temporal Graph Network (TGN) used to generate firm-level embeddings from the dynamic ownership network. The model processes a stream of time-stamped events, computing context-aware embeddings through message passing and attention mechanisms and updating node memories. Node events only contribute to the memory update