

Forecasting the evolution of transportation networks using machine learning

Weihua Lei,¹ Luiz G. A. Alves,² and Luís A. Nunes Amaral^{1,2,3,*}

⁴ *Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA*

²Department of Chemical and Biological Engineering,

Northwestern University, Evanston, IL 60208, USA

³Northwestern Institute on Complex Systems (NICO),

Northwestern University, Evanston, IL 60208, USA

(Dated: February 23, 2021)

Abstract

11 Transportation networks play a critical role in modern societies due to their importance in human
12 mobility, and the exchange of goods and ideas. On the negative side, transportation networks are
13 also the primary vehicles for the worldwide spread of infections, and for significant emissions of CO_2 .
14 It is precisely due to such emissions, that experts predict the inevitability of future restriction of air
15 transportation. We investigate the connection dynamics of two mature transportation networks:
16 the Brazilian domestic bus transportation network and the US domestic air transportation network.
17 Specifically, we apply a machine learning approach to the challenge of predicting edge removals on
18 a monthly time scale. The complex, competing stakeholders influencing such decisions, make it
19 unsurprising that we cannot fit a model that accurately predicts edge removal for the Brazilian bus
20 network. However, we also find that a machine learning approach yields a model able to accurately
21 predict the edge removal in the US domestic air transportation network. We take advantage of
22 this model to forecast the impact on the US air transportation network of a dramatic reduction in
23 the number of its connections as a result of policies to mitigate the impact of climate change. Our
24 forecast could be useful in planning future infrastructure, such as high-speed rail systems, that
25 could replace lost air connections.

* amaral@northwestern.edu

26 **INTRODUCTION**

27 Transportation networks are critical infrastructures and one of the foundations of mod-
28 ern globalized societies. The air transportation network alone is responsible for the mobility
29 of millions of people every day across the world [1, 2]. However, transportation networks
30 are also responsible, indirectly, for the propagation of diseases such as influenza and, re-
31 cently, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [3–5]. In addition
32 to their role in enabling pandemics, transportation is also a large contributor to greenhouse
33 emissions, accounting for about a quarter of total U.S. greenhouse gas emissions and total
34 global emissions. Among all transportation sectors, the air transportation contributes 9%
35 of emissions in the U.S. and 10.6% globally [6, 7]. Even more concerning, at a time when
36 global emissions must be reduced, emissions from the transportation sector are on the rise.
37 However, as the consequences of climate change become inescapable, it is inevitable that
38 dramatic changes in how transportation networks are organized will occur. Thus, it is cru-
39 cial for policymakers to be able to forecast how transportation networks will evolve in the
40 coming decades.

41 The study of edge dynamics in networked systems has revealed significant insights [8–
42 12]. However, the study of the temporal dynamics of the edges in transportation networks
43 remains underdeveloped. A significant challenge to progress is that the structure of a trans-
44 portation network is the outcome of concurrent actions of businesses motivated by profits,
45 governments motivated by national interests, and historical contingencies.

46 Recently, machine learning (ML) approaches have been successfully applied in the study
47 of human mobility [13], sustainability of transportation structures [14], and the impact of
48 COVID-19 on gasoline demands [15]. Here, we use ML to probe the dynamics of the edges
49 in transportation networks. For modern transportation networks, structural changes are
50 primarily due to the addition and removal of edges; the addition and removal of nodes being
51 much less significant. The addition of edges in networked systems, known as temporal link
52 prediction, has been studied with different machine learning approaches including matrix
53 factorization, probabilistic based models, and embedding based models [16]. The removal of
54 edges, however, has been focused on hypothetical percolation and dismantling of networks
55 where edges either are assumed to be selected for removal at random or targeted accord-
56 ing to their centrality [17, 18]. The mechanisms determining edge removal in real-world

57 transportation networks, thus, remain unexplored.

58 To address this knowledge gap, we investigate the edge dynamics of two large transporta-
59 tion networks: the Brazilian inter-cities bus transportation network (Brazil Bus net) [19, 20]
60 and the US domestic air transportation network (US Air net) [21]. Using machine learning
61 algorithms to classify edges by their structural properties, we find edges retained are sig-
62 nificantly different in their structure from edges removed for both transportation networks.
63 Further, we develop a ML predictive model that enables us to forecast removed edges. We
64 use this model to forecast the effect of a reduction in the number of connections in the US
65 domestic air transportation network.

66 DATA

67 We collected data for the Brazilian inter-city bus transportation network (Brazil Bus
68 net) and the United States domestic air transportation network (US Air net) at a monthly
69 temporal resolution. In the Brazil Bus net, the nodes represent the bus stops in the dataset.
70 An undirected edge e_{ij}^m connects nodes i and j if there is at least one bus route connecting
71 them at some point during month m . We construct an unweighted undirected temporal
72 network $\{G_1 \rightarrow G_2 \rightarrow \dots \rightarrow G_T\}$, where G_m represents the network snapshot constructed
73 with data from month m .

74 In the US Air net, the nodes represent US cities. An edge indicates that at least one
75 airline directly connected the two cities during the monthly observation period (Figure 1A).
76 Figure 1B shows that a significant fraction of existing edges are removed from the network
77 snapshot to snapshot.

78 METHODS

79 Machine learning prediction

80 We formulate the question of how to predict which edges will be removed as a supervised
81 classification problem. We assign to edges in a network snapshot G_m , one of three values:
82 ‘added’, ‘retained’, or ‘removed’. Added edges were not present at the beginning of the
83 monthly observation window but are present at the end. Retained edges are present at the
84 beginning and end of the monthly observation window. Removed edges are present at the

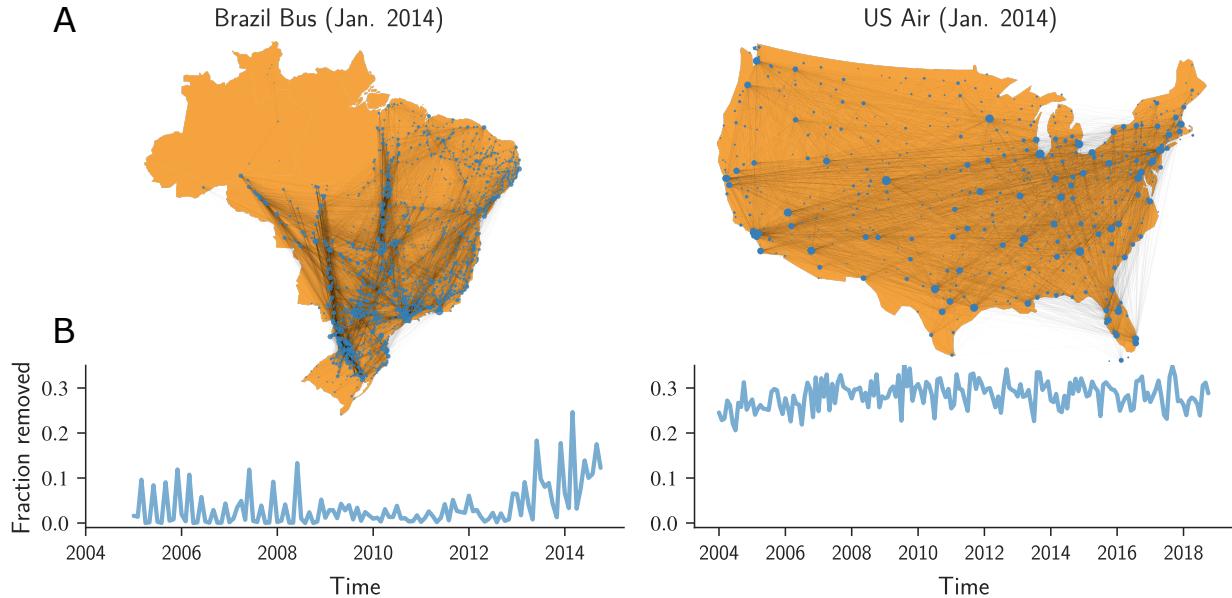


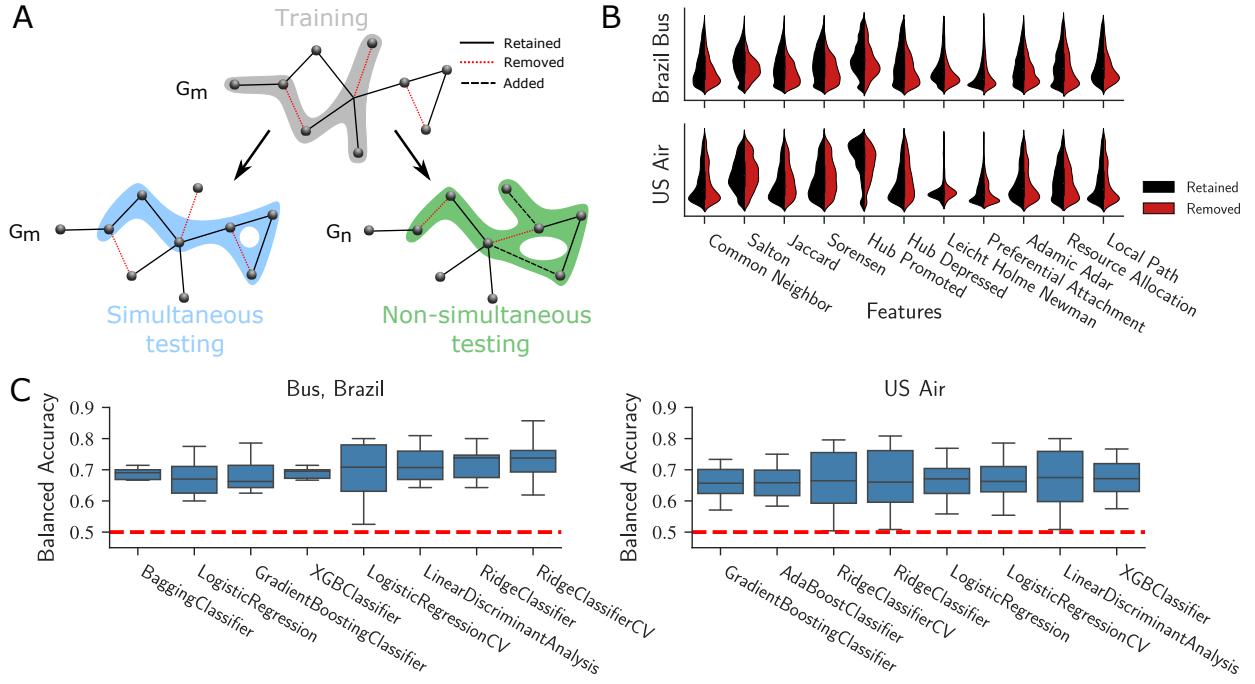
FIG. 1. Edge-dynamics in two countrywide transportation networks. (A) January 2014 snapshots of the Brazilian inter-city bus transportation network and the United States domestic air transportation network (only mainland shown). The size of each circle is proportional to the network degree of the city at that location. (B) Fraction of edges removed during month-long time windows for the two networks.

beginning but not at the end of the monthly observation window. To achieve the goal of predicting removed edges, we only need to consider edges already present in G_m . Those edges can only be ‘retained’ or ‘removed’. The features of an edge can be extracted from the network structure and represented as a feature vector \mathbf{X}_{ij}^m . Thus, we can write the probability of an edge e_{ij}^m being removed as:

$$\text{Prob}(e_{ij}^m = \text{removed}) = f(\mathbf{X}_{ij}^m). \quad (1)$$

To test our hypothesis that those removed edges have different structural features than those retained, we select 70% of retained and removed edges in a selected snapshot for inclusion in the training set. As illustrated in Figure 2A, if a model is trained with G_m , one can perform two different tests depending on whether the testing edges are selected from the same snapshot as the training edges. In the simultaneous test, we test on edges from the same snapshot as the training edges. In the non-simultaneous test, we test on edges from snapshots come after the training snapshot. This test evaluates the similarity of removal

97 dynamics for different snapshots.



100 **FIG. 2. Performance of machine learning models for predicting retained and removed**
101 **edges in a snapshot of a transportation network. (A)** Illustration the process for creating
102 training and testing sets in a transportation network. Edges shown in black are retained during
103 the observation window, whereas edges shown in red are removed. We select a fraction of edges
104 for inclusion in the training set (grey-shaded subgraph in the figure). We identify a simultaneous
105 testing set (blue-shaded subgraph) by considering all edges that were present at month G_m but
106 were not included in the training set. Additionally, we identify a non-simultaneous testing set
107 (green-shade subgraph includes edges in dashed bold line that are added after the training graph)
108 by considering all edges that exist at snapshot G_n where $n > m$, and that were not included in the
109 training set. **(B)** We calculate the edge values for a broad range of network features and compare
110 the distribution of said features for retained and removed set of edges in the training set. For
111 clarity, we normalized all features to the same scale. It is visually apparent for most features that
112 the distributions for retained and removed edges are different. **(C)** We evaluate the performance
113 of 27 common supervised classification algorithms using ten-fold cross-validation. We show box
114 plots of the estimated model balanced accuracies for the Brazil Bus net and the US Air net. We
115 order the algorithms by their average performance for each network. Only the top 8 algorithms are
116 shown here and the performance of all algorithms is shown in the supplementary material. This
117 subset of algorithms has consistent performance, around 0.65 for the Brazil Bus net and the US
118 Air net. In the following, we focus on one of those algorithms, XGBClassifier, because it has a
119 relatively low variance in performance among top-ranking algorithms.

120 **Features**

121 There are numerous features that could be used to characterize an edge in a network. We
122 focus on a subset of possible features widely used in the link prediction literature (Table I).
123 Most are local properties. Therefore, one can perform prediction without full knowledge of
124 the network. Additionally, it has lower computational complexity (with the complexity of
125 local path index being the largest: Nk^2) and is easy to scale up for large networks [22].

126 To illustrate the differences of those features between retained and removed edges, we
127 use data from the January 2014 snapshot for both transportation networks and present the
128 distributions of those 11 features for both retained and removed edges. We compared the
129 feature samples of retained and removed edges using the Kolmogorov-Smirnov statistics, a
130 test for the null hypothesis that two samples are drawn from the same continuous distribu-
131 tion. Because we do multiple sample comparisons, we also used Bonferroni corrections on
132 the significance level, i.e. $p\text{-value} = 0.05/11$, where 11 is the number of pair comparisons.
133 The test can reject the hypothesis that the feature samples of retained and removed edges
134 come from the same distribution, with $p\text{-value} < 3 \times 10^{-4}$ for all pair comparisons.

135 **Model selection and training**

136 In order to select a classification model, we performed stratified 10-fold cross-validation
137 on the balanced training set with 27 widely used classification algorithms available in *scikit-*
138 *learn* Python library [24] and in the *eXtreme Gradient Boost* package [25].

139 We calculate balanced accuracy, the average of the true positive and the true negative
140 rate, to compare classification performance. For 8 of the 27 we obtain good, stable perfor-
141 mance (Figure 2C). The results suggest that those algorithms have consistent and similar
142 predict accuracies ranging from 0.6 to 0.8. We select a single algorithm with stable and
143 satisfying performances, XGBClassifier, for the following analyses.

TABLE I. Considered features: Γ_i refers to the set of neighbors of node i . k_i is the degree of node i [23].

Feature	Definition	Description
Common Neighbors	$ \Gamma_i \cap \Gamma_j $	The number of common neighbors of nodes i and j
Salton Index	$\frac{ \Gamma_i \cap \Gamma_j }{\sqrt{k_i \times k_j}}$	The number of common neighbors normalized by geometric average degree of both nodes
Jaccard Index	$\frac{ \Gamma_i \cap \Gamma_j }{ \Gamma_i \cup \Gamma_j }$	The number of common neighbors normalized by the union of neighbors of both nodes
Sørensen Index	$\frac{2 \Gamma_i \cap \Gamma_j }{k_i + k_j}$	The number of common neighbors normalized by the average degree of the two nodes
Hub Promoted Index	$\frac{ \Gamma_i \cap \Gamma_j }{\min(k_i, k_j)}$	The number of common neighbors normalized by the smaller degree of the two nodes
Hub Depressed Index	$\frac{ \Gamma_i \cap \Gamma_j }{\max(k_i, k_j)}$	The number of common neighbors normalized by the larger degree of the two nodes
Leicht-Holme-Newman Index	$\frac{ \Gamma_i \cap \Gamma_j }{k_i \times k_j}$	The number of common neighbors normalized by the product of degrees of the two nodes
Preferential Attachment Index	$k_i \times k_j$	The product of the degrees of the two nodes
Adamic-Adar Index	$\sum_{z \in \Gamma_i \cap \Gamma_j} \frac{1}{\log k_z}$	The number of common neighbors with each of them normalized by the logarithm of their degree
Resource Allocation Index	$\sum_{z \in \Gamma_i \cap \Gamma_j} \frac{1}{k_z}$	The number of common neighbors with each of them normalized by their degree
Local Path Index	$S_{ij,2} + \epsilon S_{ij,3}$	The first term represents the number of paths of length equal to 2 between the node i and j . The second term is the number of paths of length equal to 3 between the node i and j damped by parameter ϵ . We set $\epsilon = 0.01$.

144 **RESULTS**

145 **Prediction**

146 For the Brazil Bus net, the balanced accuracies using the XGBClassifier in simultaneous
 147 tests have an average of 0.7. For the US Air net, XGBClassifier produces an average balanced
 148 accuracy of 0.68. The results of the simultaneous tests suggest that with the machine learning
 149 approach we can differentiate the retained edges from the removed edges in a given network
 150 snapshot using their structural features.

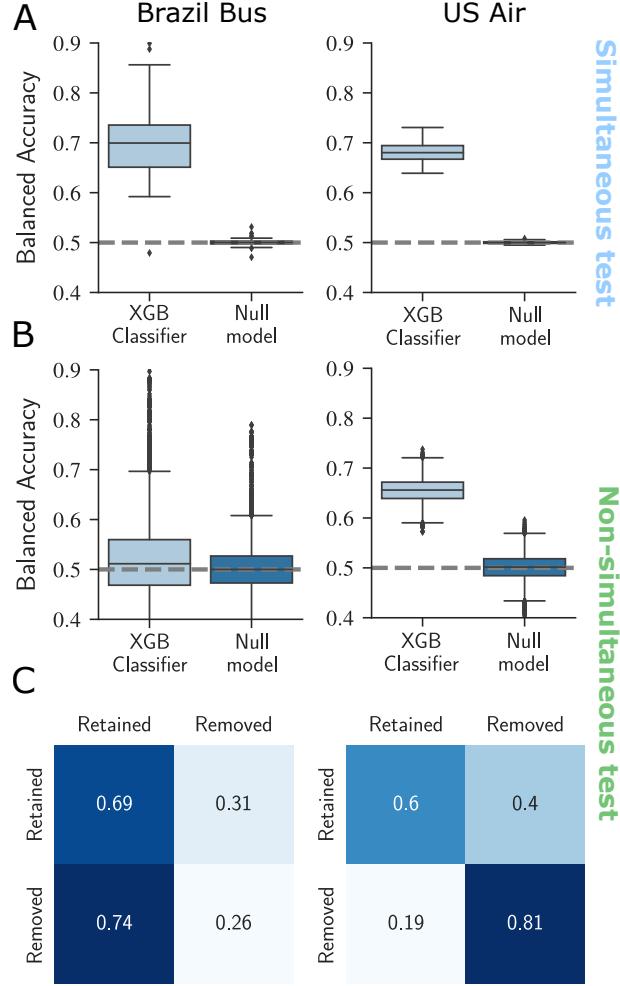


FIG. 3. **Comparison of XGBClassifier model performance against an appropriate null model.** The box-plots show the balanced accuracies for (A) simultaneous tests and (B) non-simultaneous tests on all time steps for bus transportation network and air transportation network. (C) Confusion matrices for one of all snapshots of the Brazil Bus net and the US Air net. Even though the model of the Brazil Bus net is able to perform well on the simultaneous test, its performance on the non-simultaneous test is poor.

151 A more general and useful test is to use a model trained on a known snapshot to predict
 152 edge removals in latter snapshots. Surprisingly, the prediction of the XGBClassifier for the
 153 non-simultaneous tests in the Brazil Bus net are not better than those of the null model.
 154 (Figure 3B). In the Supplementary Information there is a description of the null model.
 155 However, for the US Air net, the model gives an average balanced accuracy of 0.66. In
 156 Figure 3C, we show the confusion matrix of the model predictions from non-simultaneous
 157 tests. For the Brazil bus net, the model captures only 26% of true removals. For the US Air
 158 net, 81% of the removed edges are captured correctly.

159 **Model interpretation**

160 Next, we investigate why our predictions fail for the Brazil Bus net in non-simultaneous
161 tests. To this end, we use the SHapley Additive exPlanations (SHAP) values [26–28]. Fig-
162 ure 4A shows the features importance as well as how their values affect the outputs of the
163 model for the simultaneous test in the Brazil Bus net. Our analysis reveals that edges with
164 low local path index values are more likely to be removed in the Brazil Bus net. However,
165 the ranking of features importance changes from snapshot to snapshot (Figure 4C). Features
166 like local path index are important in certain snapshot but not in others. This variability
167 explains why non-simultaneous predictions fail for the Brazil Bus net.

168 In Figure 4A, the SHAP values summary of the US Air net reveals that the most impor-
169 tant feature for predicting edges removal is resource allocation index. Edges with low value
170 of resource allocation index are more likely to be removed in the air transportation network.
171 The ranking of features importance remain quite stable for different time snapshots, resource
172 allocation index and hub promoted index being consistently the most important features in
173 determining which edges are removed (Figure 4B). This implies that the US Air net has
174 consistent removal dynamics over time.

175 **Forecasting changes to the US Air net**

176 Armed with a model that can predict edge removals, we simulate the effect of air travel
177 restrictions aiming to mitigate climate change. We use the model trained on a known
178 snapshot to predict the probability that a given edge is removed and remove it according to
179 that probability. We take the December 2018 snapshot of the US Air net as the initial state
180 of the network. In each simulation, we assume to remove R_f fractions of edges exponentially
181 and asymptotically

$$\delta N_m = -\gamma(N_m - N_f), \quad (2)$$

182 where m is the number of months from the start of the simulation, N_m is the number of
183 edges in the current snapshot, $N_f = N_0(1 - R_f)$ is the asymptotic number of edges, and γ
184 controls the edge removal rate.

185 Figure 5A shows an ensemble of 30 simulations starting on December 2018 and removing
186 ($R_f = 2/3, 4/5$) of edges at two different rates ($\gamma = 0.02, 0.04$). Based on the predicted edge

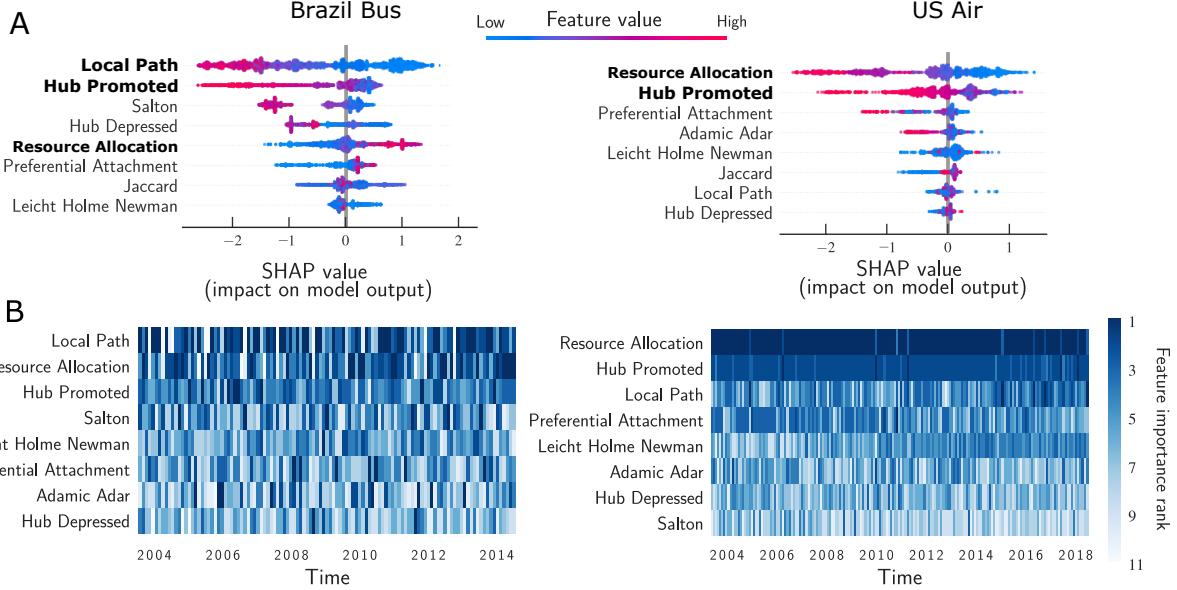


FIG. 4. Resource allocation has the largest predictive power for whether edges will be removed or retained, and it is consistent in different snapshots for the US Air net. We use SHAP values to quantify the importance of features and how they affect the prediction accuracy. The feature importance is ranked by the sum of SHAP values from top to bottom. For every feature, each point represents a data observation and the color shows its corresponding value. The impact on model output is shown along the x-axis value, positive values corresponding to the removed edges and negative values corresponding to the retained edges. **(A)** The summary of SHAP values for the simultaneous test of the January 2005 Brazil Bus net and the January 2004 US Air net. In the Brazil Bus net, the important features are local path index and hub promoted index. Edges with low local path and hub promoted index are more likely to push the model output to positive values (removed edges). Features such as common neighbor, Adamic Adar index, and Sørensen index have little influences of the model output. In the January 2004 US Air net, resource allocation index and hub promoted index have the largest influences on the model output. Edges with low resource allocation index and low hub promoted index are more likely to be removed. The rest of the features have less effects on model outputs. **(B)** Rankings of features of all snapshots using SHAP values for the Brazil Bus net and the US Air net. The darker the more important a feature is in that snapshot. The rankings changing over time means the removal dynamics are not stable over time. The ranking of features importance is quite stable over time for the US Air net.

removals, we project the estimated carbon emissions relative to the emissions in 2018 (see Figure 5B and the Supplementary Information for details of the estimations) [29]. Figure 5C and D show a snapshot from one of the simulation with edges connecting Chicago, IL on December 2030. It shows that edges with airports in larger cities are still retained in the network while edges like Chicago, IL to Appleton, WI have been removed. To better quantify the likelihood of removal, we calculate the average time an edge is retained and rank edges

193 from the shortest survival time to the longest survival time (Figure 5E). In Figure 5F, we also
194 compare the survival time with the inter-city flow predicted by the demographic gravitation
195 “law” ($F_{ij} = N_i N_j / r^2$) [30]. Here, N_i , N_j are the populations of two connected cities obtained
196 from the US 2010 census [31], and r is the distance (in miles) between them. A plausible
197 hypothesis is that connections with high predicted inter-city flows would have larger survival
198 times. This is indeed what we observe. A question then arises: could the inter-city flow
199 have better predicted the edges to be removed than the machine learning approach pursued
200 here? We test this possibility and find that our machine learning approach has a better
201 performance compared to the removal model based on the the demographic gravitation
202 “law” (see the Supplementary Information for details).

203 DISCUSSION

204 We show that edge removal processes in transportation networks are not random and
205 that accurate prediction, based on local network structures, is possible. Even though those
206 features are able to differentiate edges removed and retained, a model trained in a time
207 snapshot is not able to correctly predict removed edges in different time snapshots for the
208 Brazil Bus net. While in the US Air net, the non-simultaneous tests show that the features
209 of the edges removed are similar in all snapshots.

210 For the US Air net, we find that resource allocation index and hub promoted index
211 consistently have the largest predictive power. This largest predictive power of resource
212 allocation comes from the underlining resource allocation process that explains the nonlin-
213 ear correlation between transportation capacity and connectivity of each airport [22, 23].
214 However, the number of structural features tested in our work is limited. The predictive
215 power could possibly be improved by including additional features. Indeed, we tested the
216 impact of global features like edge betweenness centrality, but did not see any improvement
217 in predictive power.

218 A more fruitful research direction, however, may be to incorporate the multi-layer network
219 aspect of transportation networks and using machine learning approaches to explore the
220 interplay between the removal of edges in one network (e.g. air transportation network) and
221 the growth of edges in another transportation network (e.g. high-speed rail systems).

222 As daily global CO_2 emissions temporally decreased by 17 percent during the ongoing

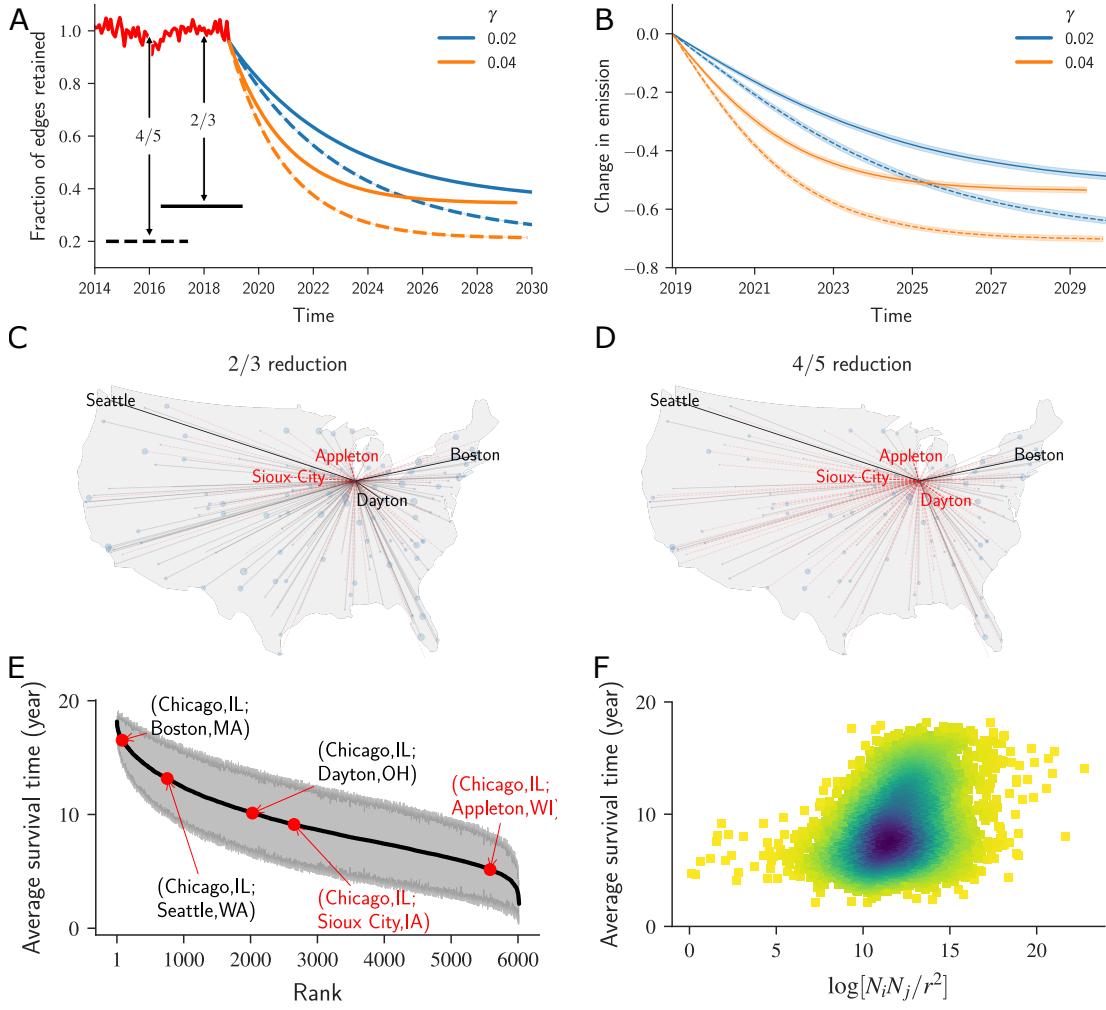


FIG. 5. Forecasting the effect of possible travel restrictions on the US Air net. Starting with the most recent network in our data set (December 2018), we removed edges at a constant rate. **(A)** Fraction of edges retained in the network for different scenarios with different removal rate controlled by γ . **(B)** We estimate the change in CO_2 emissions relative to the emissions in 2018. The curves show the averages of estimations and the shaded areas are the 95% confident intervals. To illustrate the trajectory of edge removal, we show the remaining edges connecting Chicago, IL in December 2030 for $\gamma = 0.02$ **(C)** $R_f = 2/3$ and **(D)** $R_f = 4/5$ reduction. The edges have been removed are shown in red dashed line and edges retained are shown in black. **(E)** Survival time of edges in the removal process calculated from an ensemble of 30 simulations. The edges are ranked by the average of their survival time. The grey envelope shows the 95% confident intervals. Noticed that important edges like (Chicago, IL – Boston, MA) have a greater survival time, while edges like (Chicago, IL – Appleton, WI) are more likely to be removed in the early stages of the simulation. **(F)** Edges predicted with longer average survival time have on average higher values of demographic gravitation flow, which has been proposed to approximate the human mobility between two regions. The correlation coefficient between the average survival time and the demographic gravitation flow is $\rho = 0.33$ (p -value $< 10^{-16}$). However, there is greater variation in the data, suggesting that our approach better captures the dynamics of the system.

223 COVID-19 pandemic, some have suggested continuing restrictions on air travel to help coun-
224 tries achieve the goals of the Paris Climate Agreement [28]. In addition to the environmental
225 benefits, companies have learned that some business travels can, in fact, be replaced by low-
226 cost teleconferencing. Reductions in business travel are expected even in a post pandemic
227 world.

- 228 [1] Guimera, R., Mossa, S., Turtschi, A. & Amaral, L. N. The worldwide air transportation
229 network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of*
230 *the National Academy of Sciences* **102**, 7794–7799 (2005).
- 231 [2] Barbosa, H. *et al.* Human mobility: Models and applications. *Physics Reports* **734**, 1 – 74
232 (2018).
- 233 [3] Colizza, V., Barrat, A., Barthélémy, M. & Vespignani, A. The role of the airline transportation
234 network in the prediction and predictability of global epidemics. *Proceedings of the National*
235 *Academy of Sciences* **103**, 2015–2020 (2006).
- 236 [4] Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus
237 (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
- 238 [5] Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent
239 confirmed COVID-19 cases in China. *Science* **368**, 742–746 (2020).
- 240 [6] EPA. Fast Facts on Transportation Greenhouse Gas Emissions (2019). URL <https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions>.
- 241 [7] IPCC. AR5 climate change 2014: Mitigation of climate change - Chapter 8 "Transport"
242 (2014). URL <https://www.ipcc.ch/report/ar5/wg3/>.
- 243 [8] Holme, P. & Saramaki, J. Temporal networks. *Physics Reports* **519**, 97 – 125 (2012).
- 244 [9] Scholtes, I. *et al.* Causality-driven slow-down and speed-up of diffusion in non-markovian
245 temporal networks. *Nature Communications* **5** (2014).
- 246 [10] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes
247 in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- 248 [11] Koher, A., Lentz, H. H. K., Gleeson, J. P. & Hövel, P. Contact-based model for epidemic
249 spreading on temporal networks. *Physical Review X* **9**, 031017 (2019).
- 250 [12] Berthier, E., Porter, M. A. & Daniels, K. E. Forecasting failure locations in 2-dimensional

- 252 disordered lattices. *Proceedings of the National Academy of Sciences* **116**, 16742–16749 (2019).
- 253 [13] Bassolas, A. *et al.* Hierarchical organization of urban mobility and its connection with city
254 livability. *Nature Communications* **10**, 4817 (2019).
- 255 [14] Asensio, O. I. *et al.* Real-time data from mobile platforms to evaluate sustainable transporta-
256 tion infrastructure. *Nature Sustainability* **3**, 463–471 (2020).
- 257 [15] Ou, S. *et al.* Machine learning model to project the impact of covid-19 on us motor gasoline
258 demand. *Nature Energy* **5**, 666–673 (2020).
- 259 [16] Divakaran, A. & Mohan, A. Temporal link prediction: A survey. *New Generation Computing*
260 **38**, 213–258 (2020).
- 261 [17] Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and
262 fragility: Percolation on random graphs. *Physical Review Letters* **85**, 5468–5471 (2000).
- 263 [18] D’Souza, R. M. & Nagler, J. Anomalous critical and supercritical phenomena in explosive
264 percolation. *Nature Physics* **11**, 531–538 (2015).
- 265 [19] Agency, N. L. T. National Land Transport Agency-ANTT 2017 Statistics and Road Studies-
266 Operational Data. URL <http://antt.gov.br>.
- 267 [20] Alves, L. G. A., Aleta, A., Rodrigues, F. A., Moreno, Y. & Amaral, L. A. N. Centrality
268 anomalies in complex networks as a result of model over-simplification. *New Journal of Physics*
269 **22**, 013043 (2020).
- 270 [21] BTS. BTS-Transtats (2018). URL <https://www.transtats.bts.gov/TableInfo.asp>.
- 271 [22] Li, W. & Cai, X. Statistical analysis of airport network of China. *Phys. Rev. E* **69**, 046106
272 (2004).
- 273 [23] Zhou, T., Lü, L. & Zhang, Y.-C. Predicting missing links via local information. *The European*
274 *Physical Journal B* **71**, 623–630 (2009).
- 275 [24] Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning*
276 *Research* **12**, 2825–2830 (2011).
- 277 [25] Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the*
278 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*,
279 785–794 (ACM, 2016).
- 280 [26] Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In Guyon,
281 I. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 30, 4765–4774 (Curran
282 Associates, Inc., 2017).

- 283 [27] Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypox-
284 aemia during surgery. *Nature Biomedical Engineering* **2**, 749–760 (2018).
- 285 [28] Le Quere, C. *et al.* Temporary reduction in daily global CO_2 emissions during the COVID-19
286 forced confinement. *Nature Climate Change* **10**, 647–653 (2020).
- 287 [29] Xinyi Sola Zheng, B. G. & Rutherford, D. U.S. domestic airline file efficiency
288 ranking, 2017-2018. Tech. Rep. (2019). URL <https://theicct.org/publications/us-domestic-airline-fuel-efficiency-ranking-2017-18>.
- 290 [30] Barthelemy, M. The statistical physics of cities. *Nature Reviews Physics* **1**, 406–415 (2019).
- 291 [31] United States Census Bureau. City and Town Population Totals: 2010-2019 (2019). URL
292 <https://www.census.gov/>.
- 293 [32] Lundberg, S. M. *et al.* From local explanations to global understanding with explainable ai
294 for trees. *Nature Machine Intelligence* **2**, 56–67 (2020).

295 **SUPPLEMENTARY INFORMATION**

296 **Data**

297 The Brazilian inter-cities bus data is collected from the Brazilian National Land Trans-
298 portation Agency (ANTT) [19]. The dataset contains all inter-cities bus transportation from
299 January 2005 to December 2014 with monthly resolution. It is represented as a temporal
300 unweighted undirected network, where nodes are individual bus stops and edges are whether
301 there are bus routes between the two bus stops within that monthly snapshot. The network
302 has 120 snapshots with about 1734 nodes and 18781 edges on average.

303 We obtained United States domestic air transportation data from the Bureau of Trans-
304 portation Statistics (BTS) [21]. The data is in the period from January 2004 to December
305 2018. Using the same approach we used in Brazil inter-cities bus data, each snapshot of
306 the network is constructed from data of the corresponding month. The nodes are airports
307 and an edge in a snapshot represents there is at least one airline connected the two airports
308 in that month. The networks have 192 snapshots with about 819 nodes and 6547 edges in
309 average.

310 **Class balancing**

311 Most machine learning classification algorithms favor majority class in an imbalanced
312 dataset. In two transportation networks we study, the number of removed edges is much
313 smaller than the edges that are retained. To eliminate this issue in our highly imbalanced
314 data, we balanced the training data by keeping the same number of the majority class data
315 samples (retained edges) as the minority class data samples (removed edges) using random
316 under-sampling method.

317 **Performance of widely used classification algorithms**

318 We performed stratified 10-fold cross-validation on the balanced training set with 27
319 widely used classification algorithms available in scikit-learn python library [24] and in the
320 eXtreme Gradient Boost package [25]. We calculate balanced accuracy, the average of the
321 true positive and the true negative rate, to compare classification performance (Figure S2).

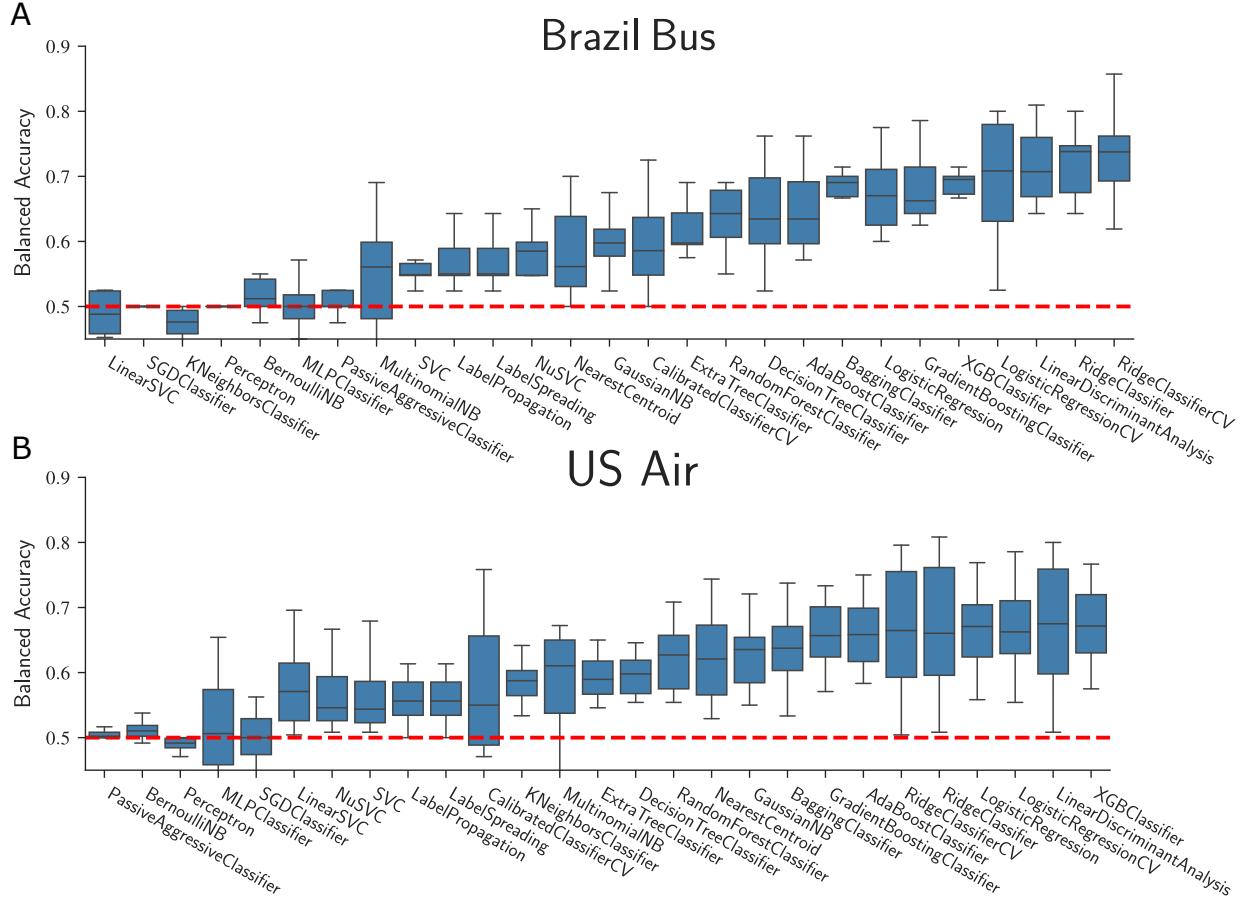


FIG. S1. Performance of 27 widely used classification algorithms. We evaluate the performance of 27 common supervised classification algorithms using ten-fold cross-validation. We show box plots of the estimated model balanced accuracies for **(A)** the Brazil Bus net and **(B)** the US air net.

322 **Null model**

323 The null model is constructed by assuming edges are randomly removed. For each snap-
 324 shot, the same number of edges observed in the data is selected at random and labeled as
 325 removed. Then, the same algorithm (XGBClassifier) is trained and tested on the randomized
 326 data.

327 **Estimate the CO_2 emissions reduction**

328 To estimate the CO_2 emissions from the United States domestic air transportation, we
 329 use the average fuel efficiency of US airlines in 2018. The methodology calculates the CO_2

330 associated to a specific route, as follows:

331 Annual CO_2 emissions (in tons) = 3.16×32.5 (gram fuel per km) \times trip distance (in km)
332 \times number of flights each year $\times 10^{-6}$ (tons per gram).

333 3.16 is the constant representing the number of tonnes of CO_2 produced by burning a
334 tonne of aviation fuel. 32.5 g fuel per km is the average fuel efficiency of the US airlines in
335 2018 [29]. Number of flights are assumed to be the same and calculated from 2018.

336 **SHAP (SHapley Additive exPlanation) values**

337 SHAP values is an unified measure of feature importance [32]. To compute the importance
338 of a feature, it calculated the change in the expected model prediction by withholding that
339 feature. Mathematically, this method retrains model on all subset of features $S \subset F$, where
340 F is the set of all features. Since there are multiple subsets that satisfied $S \subseteq F \setminus \{i\}$. The
341 SHAP values are then the weighted average of all possible changes:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (3)$$

342 Where $f_{S \cup \{i\}}$ is a model trained with feature $S \cup \{i\}$, and f_S is a model trained on S
343 without feature i .

344 **Comparison with edge removals predicted by the demographic gravitation flow**

345 In addition to the comparison with the random null model from Fig. 3, we also computed
346 a model where the edges are removed according to inter-city flow predicted by demographic
347 gravitation “law”. To do so, for each snapshot m , we computed and ranked the edges by
348 the demographic gravitation “law” $F_{ij} = \frac{N_i N_j}{r^2}$, and removed N edges with the smallest flow,
349 where N is equal to the number of edges removed in the real data. We show a confusion
350 matrix for one snapshot of the network and the average accuracy of the demographic gravi-
351 tation “law” in capturing edges removal over the entire period of our data in Figure S2. Our
352 machine learning approach produce much better results than the demographic gravitation
353 “law”.

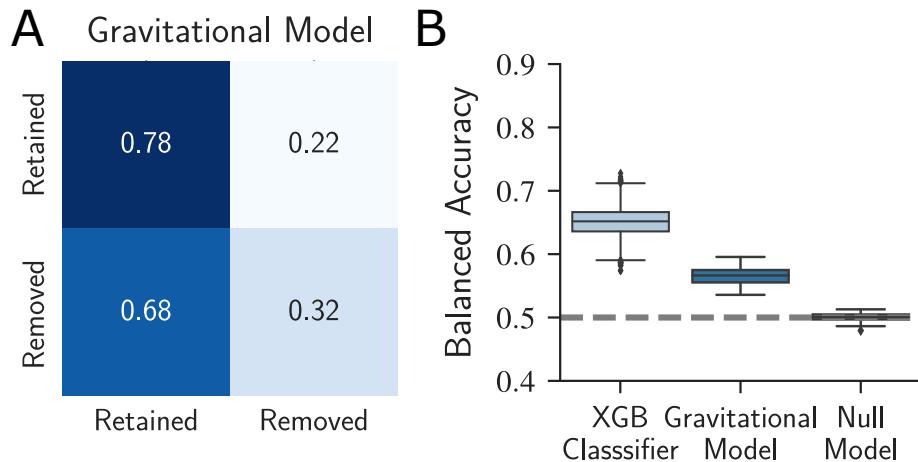


FIG. S2. **Results for removal based on inter-city flow predicted by the demographic gravitation “law”.** (A) Confusion matrix for one of all snapshots of the US Air net. (B) The box-plots show the balanced accuracies for non-simultaneous tests on all time step for the US Air net.