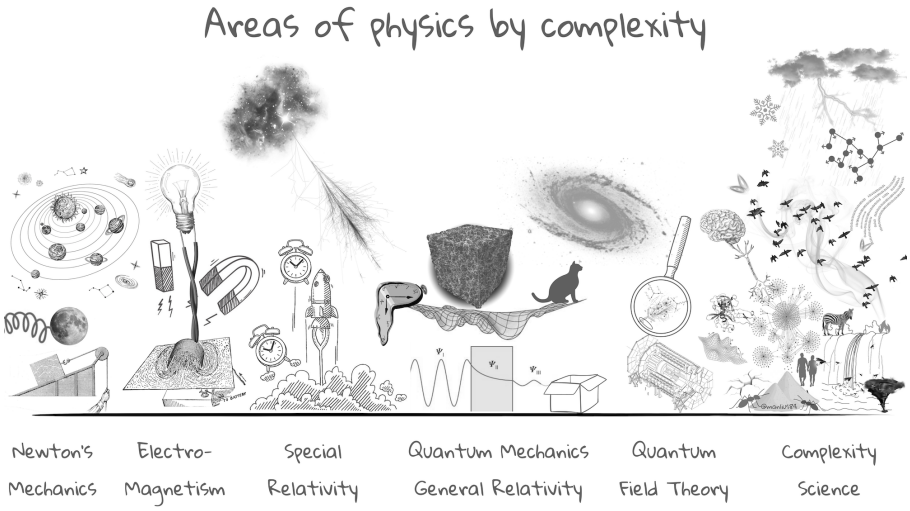
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



Project # X: ....

Salcido-Gutiérrez, Ana Sofía

Last update: June 13, 2025

Contents

1 Motter-Lai Model: Cascade Failure **1**
1.1 Homogeneous vs Heterogeneous networks 1
1.2 Real-World networks 2

2 Language competition dynamics **4**
2.1 Model 5
2.2 Results 5

3 Public transport in large cities worldwide **7**
3.1 The data 7
3.2 Visualization 8
3.3 Topological Properties 8

4 Bibliography **10**

1 | Motter-Lai Model: Cascade Failure

Task leader(s): *Adilson E. Motter, Ying-Cheng Lai*

On 28 of September 2003 Italy suffered one of the biggest blackouts reported in the country. A power line arced near a tree and automatically disabled itself. This power line was one of the main nodes in the entire grid because it carried hydroelectric power from Switzerland to Italy, leaving a large part of Italy without electricity. This is a clear example of a cascade example in real life, where one affected node dismantles the entire power grid network.

By human nature, real-world networks (Internet, power grids, traffic, airlines, etc.) are heterogeneous. This means that the networks have topological features such as clusters, degree-degree correlations, and modularity [9]. This makes the network susceptible to intentional attacks (degree-ranks, hubs, betweeness, etc.), which is a very important aspect of study because it can lead to cascade failures.

A simple model to study this process was proposed by Motter and Lai in 2002 [7]. They proposed that the exchange of information between nodes, the load L_i, is transmitted by the shortest path through the nodes, which means the betweeness of the network. Consequently, the capacity of the node, C_i, is the maximum load it can handle. Then, the capacity is define as:

$$C_i = (1+\alpha)L_i^{initial} \quad (1)$$

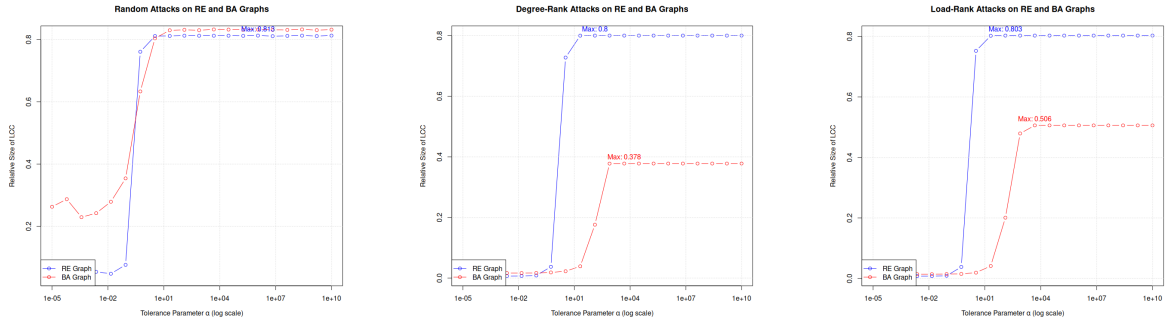
Here we introduce the tolerance parameter, α, which controls how susceptible the network is to dismantle, if alpha is large the network is robustness. From equation (1) we can observe that for large L_i we will reach a failure condition:

$$L_i >= C_i$$

This will make the node fail, creating a redistribution of the loads over the network, and a subsequent failure will occur, until we reach a stable distribution of information or the network stop existing. We can study how the cascade failures acts for homogeneous and heterogeneous networks and which topological aspects are important. Also, we can simulate this process for real-world networks.

1.1 | Homogeneous vs Heterogeneous networks

The first feature to analyze is the cascade effect in homogeneous and heterogeneous networks. For this we compute 20 instances of Erdős–Rényi and Barabási–Albert



(a) Random attacks (b) Degree-rank attacks (c) Load-rank attacks

Figure 1.1: Evolution of the GCC in Erdős–Rényi (blue) and Barabási–Albert (red) networks under different attack strategies by varying values of α.

graphs, each one consisting of 1000 nodes and an average degree of 5. Then we observe how the Giant Connected Component (GCC) is affected for different types of attacks: Random, Degree-Rank and Load-Rank depending on the tolerance parameter.

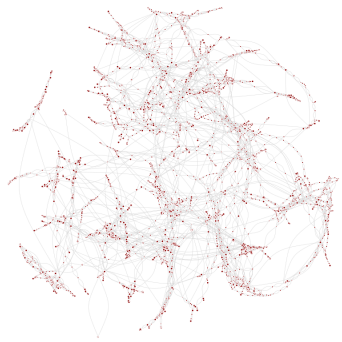
In Figure 1.1, we can observe how networks are affected for different attacks. The first aspect that we observe is the phase transition from a dismantled network to the emergence of a GCC. The principal difference between the models is that for a homogeneous network (blue line), the phase transition occurs faster than for heterogeneous networks (red line), also, we can see that for the Barabási–Albert the degree-rank attack affect in large scale because this networks contain hubs; additionally, the size of the GCC for the homogeneous network remains relatively unaffected by the attacks. This means that homogeneous networks are more robust for cascade failure than the heterogeneous ones.

1.2 | Real-World networks

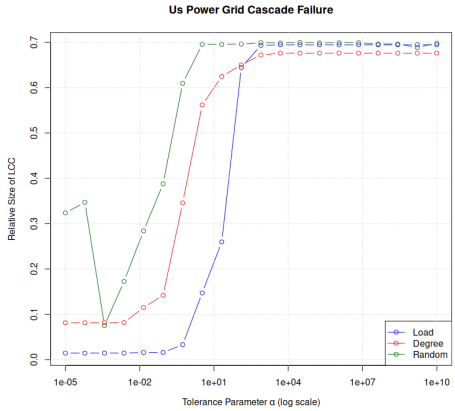
We can study the effect of cascade failure for real-world networks, which, as previously mentioned, are typically heterogeneous. In this case, we study the US power grid which consists of 4941 nodes and an average degree of 5.33, which is very high for spare networks. In Figure 1.2a we observe the network of the US power grid, revealing several modular regions and clusters, confirming the heterogeneous nature of the network.

After starting attacks on the network, we can observe the phase transition in the cascade failure on 1.2b where the relative size around GCC drop around 0.7. We can also observe the differences in the phase transition for each type of attacks, for random attacks (green line), the network does not suffer a large dismantle in the begging and the phase transition is more soft, for the degree-rank attack (red line) the network does not fully dismantle at low values of α whereas for load-rank attacks, the network is completely dismantled at small α. At the end, the three attacks arrive to the same relative size of the GCC, with the main difference being how and when they reach this state.

Another network to study is the Internet Autonomous System (AS) which contains 6474 nodes and an average degree of 3.8838. The network can be observed in



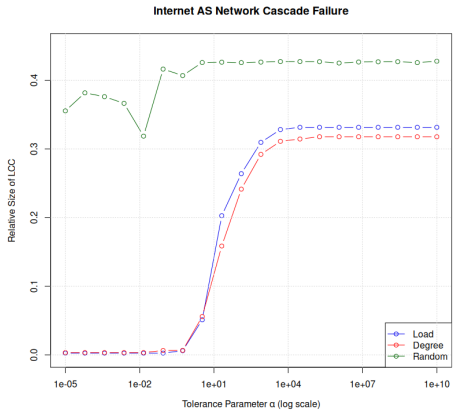
(a) US power grid Network



(b) Cascade Failure

Figure 1.2: Evolution of the GCC for the US power grid with different attacks

(a) Internet AS network



(b) Cascade Failure

Figure 1.3: Evolution of the GCC for the Internet Autonomous System with a different attacks

Figure 1.3. Compared to the US power grid, the Internet AS network is more densely connected, and we can clearly observe large modular groups and clusters. We can observe in Figure 1.3b that for random attacks (green line) the network remains stable. Instead, for the degree-rank (red line) and load-attack (blue line) we can observe the phase transition. We also compare how the relative size of the giant connected component (GCC) is affected. Under random attacks, the GCC stabilizes around 0.4, while for degree-rank and load-rank attacks, it drops to approximately 0.3.

This gives us a better idea of how careful we have to be with the important nodes in the network, because any failure in one of these nodes, will create a cascade effect and can even completely destroy the network. We can also compare how the topological aspect of the network affects the robustness for attacks. This is important even for intentional attacks or a nature effect, as it occurs with the blackout in Italy. With the increasing impact of climate change and stronger storms, it is essential to improve the robustness of critical infrastructure networks to prevent large-scale failures.

2 | Language competition dynamics

Task leader(s): *Daniel M. Abrams, Steven H. Strogatz*

The appearance and disappearance of languages have been in our culture since the beginning of history. Some have evolved by introducing new words, expressions and sounds,while others have been imposed through colonization or emerged as a form of resistance against repression. Currently, around of 7159 are in use, but studies have shown that of this number, 90% can disappear in the current century [5]. This is alarming because the language not only is a way to communicate but it also represent a culture.

We can simulate this phenomenon as a language competition. A simple model was proposed by Abrams and Strogatz [1]. Here we consider that the language is fix (not new words or expressions) and competitive. We consider a highly connected monolingual population with no spatial or social constraints. For simplicity, we consider only two languages, X and Y, and assign them a "popularity" that depends on the number of active speakers and its status.

So we can consider $P_{yx}(x, s)$ as the probability that someone change language from $Y \to X$ in function of x as the population that speaks X and the relative status, s. With this we can model:

$$\frac{dx}{dy} = yP_{yx}(x, s) - xP_{xy}(x, s) \quad (2)$$

With boundary conditions:

$$P_{yx}(0, s) = 0; \qquad P_{yx}(x, 0) = 0$$

Meaning that if X has no speakers or status, no one will speak this language. These equations only have two stable points: $x = 0$ and $x = 1$ leading only with extinction and survival. The limitation of this model is that it do not accept a bilingual scenario. For this, we adopt the dynamics propose by [2] where now we consider 3 agents: x agents that speaks X, y agents that speaks Y and xy agents that are bilingual. The probabilities of changing the language are now:

$$P_{i\ x\to xy} = \frac{1}{2}\sigma_i^y \quad P_{i\ y\to xy} = \frac{1}{2}\sigma_i^x \quad (3)$$

$$P_{i\ xy\to y} = \frac{1}{2}(1 - \sigma_i^x) \quad P_{i\ xy\to x} = \frac{1}{2}(1 - \sigma_i^y) \quad (4)$$

Where σ_i are the local densities. The transition probability is given by

$$P_{i\ x\to y} = \frac{1}{2}\sigma_i^y \quad P_{i\ y\to x} = \frac{1}{2}\sigma_i^x$$

2.1 | Model

To model the dynamics of this phenomenon we use a small-world or Watts-Strogatz network, where the nodes represent the persons, and the edges represent the change of languge. To model the dynamics of the network we can use a Kuramoto model, this model offers the dynamics to study the synchronization behavior in some phenomena as an harmonic oscillator. To study the dynamics of language competition, we observe that its behavior exhibits a diffusion-like and peer-influence [6]. This helps us to transform the Kuramoto model to include these features. The diffusive process is computed as:

$$\frac{dx_i}{dt} = \sum_i A_{ij}(x_j - x_i)$$

So, the dynamics on the network will be

$$\frac{dx_i}{dt} = P_{yx}(x_i, s) - P_{xy}(x_i, s) + \sigma \sum_j A_{ij}(x_j - x_i)$$

Where

$$P_{yx} = (1-x)^a(1-s) \qquad P_{xy} = x^a s$$

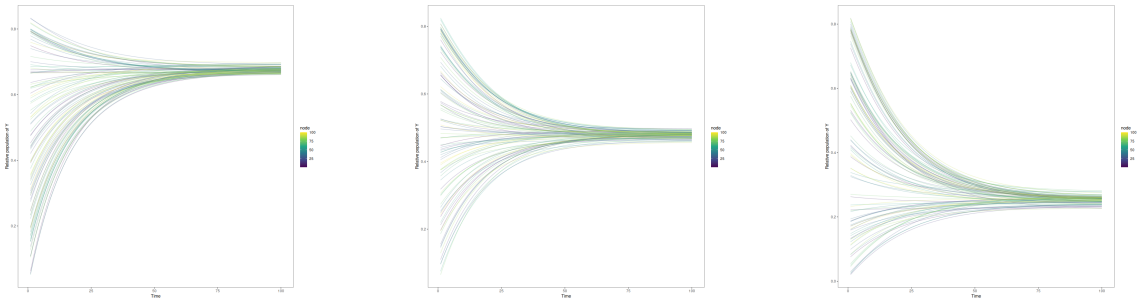
And a is a parameter which have been constant around different studies, with a value of 1.3 ± 0.25[1]

We can simulate this with students of Physics of Data. The majority of the population speaks Italian, but the classes are in English, so this language have a big status. We suppose that only these two languages are spoken and fixed. We study the behavior of the change of Italian (Y language) to English (X language) with different status. For this we consider a Wattz-Strogatz model with 100 nodes and a wiring probability of 0.3. The model is simulated 100 times.

2.2 | Results

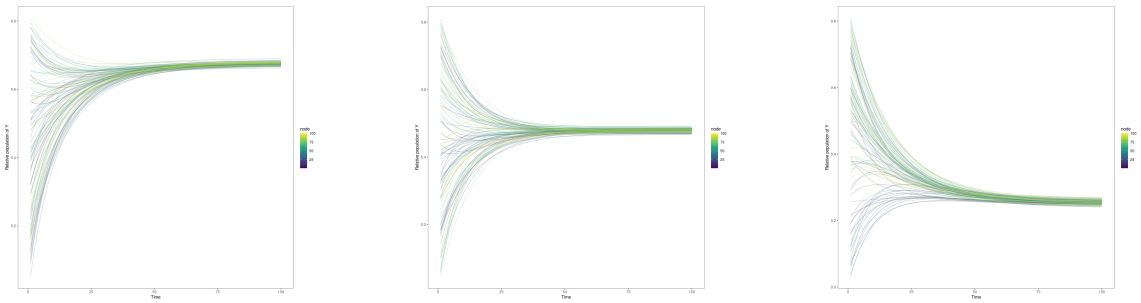
In Figure 2.1 we explain the observe of the language competition of the model. Here we model for 3 different English status. Additionally, the initial size of speaker is random between values of 0 to 1. In Figure 2.1a we use an English status of 0.2, here the population of Italian speaker remains around 0.8 meaning that for small population of Italian speaker, this increase. Instead, if we increase the the english status, in Figure 2.1b we observe that we remain in almost half-half partition in the speakers population and for a high english status, $s = 0.8$, we observe how the large Italian population decreases, Figure 2.1c.

We can also study how the strength of the coupling affects this dynamics, for the previous model we use a small coupling factor of 0.1, we can think of this parameter as the sensibility of the population to learn/change of languages, where a strong strength means that the people are more accessible to this change. Now we can study how the strength of the coupling affects the dynamics, in Figure 2.2 we use a strength of 0.8, as we can observe we arrive to the same ratio of the population, but the difference is in the time to arrive to these steady-states.



(a) Small status $s_E = 0.2$ (b) Med Status $s_E = 0.5$ (c) High Status $s_E = 0.8$

Figure 2.1: Evolution of the relative size of Italian population for different status of English.



(a) Small status $s_E = 0.2$ (b) Med Status $s_E = 0.5$ (c) High Status $s_E = 0.8$

Figure 2.2: Evolution of the relative size of Italian population for different status of English with a strength coupling.

For a small status with low strength, we reach the steady state around a 75 epoch; instead, for a strong strength the same steady state is reached around the 50 epoch. We also observe how more initial states reach the same value of relative population (we have less variance). We can compare the result with different networks models to see if this influences the behavior of the dynamics. In Figure 2.3a we compare a Barabási-Albert with the Small-World, Figure 2.3b, used in the project. For both we use the same average degree equal to 3. Both reach the same steady-state, but the dynamics are different.

(a) Barabási-Albert Model (b) Small World Model

Figure 2.3: Language Dynamics for different networks models for a status of 0.5

3 | Public transport in large cities worldwide

In large cities we have multiple modes of transportation, from buses to metro systems. Understanding the distribution of these transportation modes is essential; with this analysis, we can identify where most people move, which lines are the most used, etc. having this information is crucial for optimization of routes, planning new lines or add different transportation modes. The analysis of this behavior can be done from a complex network perspective.

In this project, we use open data provided by CityLines [4], a portal that offers access to multimodal public transport data for over 200 cities worldwide. Our objective is to generate two CSV files: one for the nodes, containing the fields nodeID, nodeLabel, latitude, longitude, mode, and year; and one for the edges, containing nodeID_from, nodeID_to, mode, line, and year.

With these data, we can reconstruct urban transportation networks and analyze their structural properties. This includes studying topological characteristics such as betweenness centrality, closeness, degree distribution, network robustness against failures or attacks, and overall homogeneity, among others.

3.1 | The data

The data consist on 8 files that can be JSON or CSV that are: cities, lines, stations, trams, means of transportation, lines of stations, lines of trams and systems. The last one do not contribute with important information. We can think in this files as a relational database, this can help us to understand the structure of the data and construct the network. As a general structure, we can observe of Figure 3.1 how are the relations.

Additionally to systems file, we do not have a way to connect the trams file with the rest of the network directly because the only feature that they share is city id. We use lines files to create a "bridge" between the tables of cities, means of transportation and stations. After this we can analyses the dataset and clean up. In station file we have the column of "closure" which by default is 99999 meaning that the station still working, so we remove the stations that are not more in use. Additionally, we remove the columns that are not necessary as it can be color, url_name, country, state, etc. of the rest of the files. So we can leave only the information to construct our network.

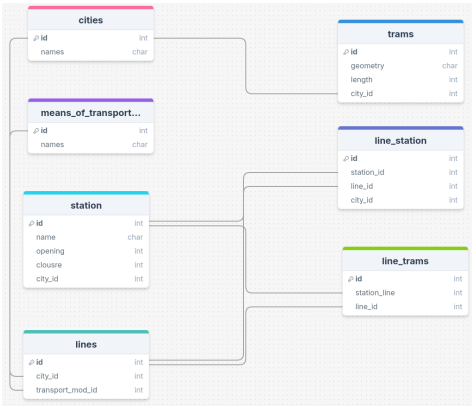


Figure 3.1: General graphic of the relations in the dataset

As we can see in 3.1, all the databases are related to the city id, so we can take the unique values and make a loop over this to separate for each city. We use stations as our nodes, for this we use the geometry columns that give us the geographical points on where are the stations, for this we can use the stringr or sf [8] library. The complicate part comes at the moment to generate the edges because we do not have a file that tell us how is the order of the stations, so we have two options: try to order the stations one follow by the other in the number row, or try to match the latitude and longitude of the station points with the start and end of the trams points.

3.2 | Visualization

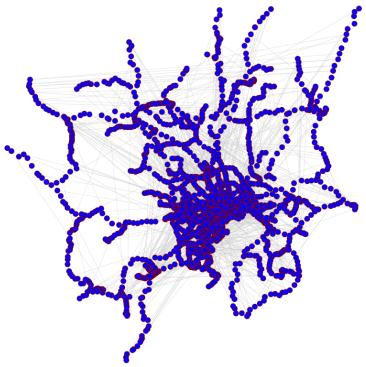
Now we have the files of nodes and edges done, with this we can simply call them and use the latitude and longitude as our layout. This is the simple way to create a network using igraph. But having this information can lead us to use interactive libraries as is Leaflet [3]. This package uses the coordinates obtained and overlapped in a real map. This would help us to confirm the fidelity of the network obtained.

In figure 3.2a we can observe the network for Tokyo, which is one of the cities with more data, using the order of the rows as the edges. As we can see, it looks a bit messy because we get connections from the distant points to the center of the network.

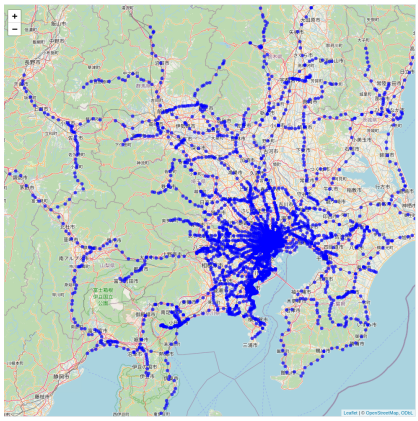
3.3 | Topological Properties

After computing the network for all available cities, we can study in general how the topological properties of public transport networks are. Our first property of interest is the average degree, in Figure 3.3a we observe an histogram of the average degree for the cities in the dataset, we see that the most frequent values is around 2 which correspond to a spare network meaning that almost all the stations only connects, in average, with 2 more stations. This also means that the nodes are in a tree-like structure, which corresponds to the expected value because the stations are arranged sequentially.

In Figure 3.3b we have an histogram of the assortative values of the network. This value give us information about the degree-degree correlations where values of -1 means that the network is dissortative, no correlations, and if the value is 1 the network

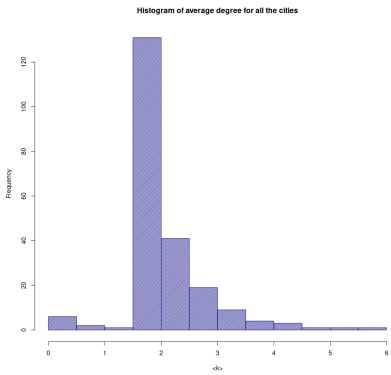


(a) Tokyo network

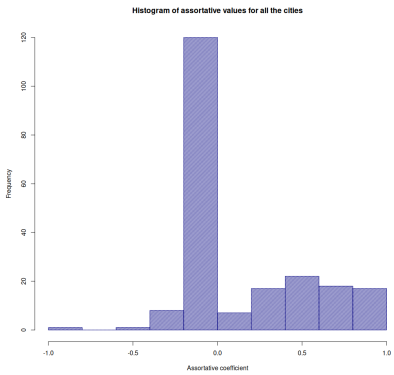


(b) Overlap of the stations with a real-map

Figure 3.2: Comparison of the network obtained for Tokyo transport and its overlap in a real-world map



(a) Histogram of average degree over all the cities



(b) Histogram of Assortative coefficients over all the cities

Figure 3.3: Histograms of the values of average degree (a) and assortative values (b) for the cities in the dataset

have degree-degree correlation. A value around 0 suggests no significant correlation, or random mixing. In our dataset, most cities show assortativity values close to zero, indicating that their public transport networks do not exhibit strong degree–degree correlations and are effectively uncorrelated. This confirms that the connection by row number is wrong, and we need to find an approach to match the geographical points in the tram files and station files to have the correct connection between stations.

4 | Bibliography

[1] Daniel M. Abrams and Steven H. Strogatz. Modelling the dynamics of language death. *Nature*, 424(6951):900, 2003. doi: 10.1038/424900a. URL https://doi.org/10.1038/424900a.

[2] XAVIER CASTELLÓ, VÍCTOR M. EGUÍLUZ, MAXI SAN MIGUEL, LUCÍA LOUREIRO-PORTO, RIITTA TOIVONEN, J. SARAMÄKI, and K. KASKI. *MODELLING LANGUAGE COMPETITION: BILINGUALISM AND COMPLEX SOCIAL NETWORKS*, pages 59–66. doi: 10.1142/9789812776129_0008. URL https://www.worldscientific.com/doi/abs/10.1142/9789812776129_0008.

[3] Joe Cheng, Barret Schloerke, Bhaskar Karambelkar, and Yihui Xie. *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*, 2025. URL https://rstudio.github.io/leaflet/. R package version 2.2.2.9000, https://github.com/rstudio/leaflet.

[4] CityLines. Citylines: Public transport network data, 2025. URL https://www.citylines.co/.

[5] Endangered Languages Project. Endangered languages: Why so important?, n.d. URL https://www.endangeredlanguages.com/about_importance/. Accessed: 2025-06-02.

[6] Jinyun Ke, Gong Tao, and S-Y William, Wang. Language change and social networks. *Communications in Computational Physics*, 3(4):935–949, 2008. ISSN 1991-7120. doi: https://doi.org/2008-CiCP-7882. URL https://global-sci.com/article/81229/language-change-and-social-networks.

[7] Adilson E. Motter and Ying-Cheng Lai. Cascade-based attacks on complex networks. *Phys. Rev. E*, 66:065102, Dec 2002. doi: 10.1103/PhysRevE.66.065102. URL https://link.aps.org/doi/10.1103/PhysRevE.66.065102.

[8] Edzer Pebesma and Roger Bivand. *Spatial Data Science: With applications in R*. Chapman and Hall/CRC, 2023. doi: 10.1201/9780429459016. URL https://r-spatial.org/book/.

[9] Lucas D Valdez, Louis Shekhtman, Cristian E La Rocca, Xin Zhang, Sergey V Buldyrev, Paul A Trunfio, Lidia A Braunstein, and Shlomo Havlin. Cascading failures in complex networks. *Journal of Complex Networks*, 8(2), April 2020. ISSN 2051-1329. doi: 10.1093/comnet/cnaa013. URL http://dx.doi.org/10.1093/comnet/cnaa013.