# ANGEL: A Synthetic Model for Airline Network Generation Emphasizing Layers

Marzena Fügenschuh, Ralucca Gera, and Andrea Tagarelli

***Abstract*—Air transportation networks develop independently based on the needs of carriers, economic and political factors, as well as their interactions. A deeper study of these networks can provide insights into the growing process and the characteristics of this multi-billion dollar industry. In this regard, reliable network-formation models are required. Generating synthetic models for airline transportation networks is a difficult and current endeavor. The subnetworks induced by airline companies lead naturally to a multilayer structure. Upon a profound analysis of the benchmark European Air Transportation Network (EATN), we propose a novel approach, called Airline Network Generation Emphasizing Layers (ANGEL), to create synthetic multilayer networks mimicking the two patterns typical in air transportation networks, i.e., hub-and-spoke and point-to-point structures. Moreover, we introduce new statistics to analyze the EATN and to validate the synthetic network model, which can be applied to study multilayer networks in general.**

***Index Terms*—synthetic multilayer network generation, air transportation systems, multiplex network analysis**

## I. Introduction

The global transportation system is a dynamic and intricate network, rapidly linking cities and thus presenting the most connected physical network seen to date. In 2010, the global air transportation network transported 2.4 billion passengers and 43 million tons of cargo, it has been responsible for 32 million jobs, 2% of global carbon emissions and 545 billion dollars in revenue [1]. Optimizing travel through this network to efficiently transport goods and people, as well as analyzing its resilience to disruption, is highly desirable based on the real-world limitations of airports, aircraft, financial and personnel resources, and the unpredictability of passengers, weather and natural disasters. In order to effectively study the real-world development of this complex network, a methodology for creating synthetic networks comparable in size and behavior is needed.

The natural development of air networks is difficult to model because of the lack of historical data and their multilayered nature. Here, each airline company is captured by one layer, which is itself a network, and independently creates routes based on market analysis for profit, competitor routes, and available resources and destinations. However, each airport develops separately driven by the municipalities they service. Since each airline attempts to gain routes to increase market share and revenue, the response to these market forces is the basis for the growing multilayer network. Based on our analysis of a benchmark airline network, our goal in this work is to develop a method for generating layers of airline networks and their representation under a multilayer network model.

M. Fügenschuh is with the Beuth University of Applied Sciences, Berlin, Germany. Email: fuegen@beuth-hochschule.de.

R. Gera is with the Dept. of Applied Mathematics at Naval Postgraduate School, Monterey, CA. Email: rgera@nps.edu.

A. Tagarelli is with the Dept. of Computer Engineering, Modeling, Electronics and Systems Engineering (DIMES), University of Calabria, Rende (CS), Italy. Email: tagarelli@dimes.unical.it.

### A. Related work

*Multilayer networks* add an extra dimension to the single-layer networks [2], [3], [4]. Their structural properties have been analyzed on the ensemble of layers, as studied in [4], as well as their spectral properties [5], [6]. Most of the existing works have traditionally studied the relationships between the layers rather than within the layers; particularly, it has been shown that different layers can influence the community structure based on communities between layers [7], overlapping communities between layers [8], and redundancy of pairwise connections across layers [9]. Given the complexity of multilayer network systems, *generative models for synthetic networks* have been studied as a tool to create artificial (fictitious) system models with capability of representing characteristic features of actual networks. This becomes of paramount importance for modeling, simulating, tuning and validating the system dynamics in a controlled environment. Single-layer network models introduced in the last decade have been reviewed in [10], whereas synthetic multilayer networks are reviewed in [4]. The general trends to model growing multiplex networks are based on preferential attachment [11], [12], [13], [14], statistical models [15], [16], [17], ensembles of multiplex networks [18], [19] such as multilayer exponential random graphs [20], [21], [22], [23], and multilayer stochastic block models [24], [25].

*Air transportation systems* represent a prototypical example for a multilayer network. A corpus of studies based on network analysis provides insights into their topological structure and dynamics. [26] investigates the world-wide airport network to understand its anomalous centrality distribution. Using $k$-core decomposition, [27] analyses the structural properties of the Chinese Airline Network as a multiplex consisting of three layers; the core layer includes airports of provincial capital cities and supports most flight flow, while the bridge layer connects the core layer with a periphery layer comprising more remote areas with no direct flight between them. [28] proposes a theoretical framework of causality test to investigate interaction and propagation patterns of delay between airport systems. [29] focuses on multimodal aspects relating to temporal characteristics of the UK public transport system for a given period; each transportation mode is associated with a different layer in a multilayer network framework.

The *European Air Transportation Network* (EATN) has received particular attention in studying its vulnerability to disruption by spacial hazards [30]. The EATN has also been used to generate synthetic networks given by a scale-free structure based on preferential attachment to investigate connections of most connected cities versus central cities [31], or the network flow in multilayer networks [32] or to observe how the aggregated network is affected by the emergence of a single layer [33]. In [34], the multilayer approach is combined with the scale-free structure of the network to generate a model imitating the EATN. This approach, based on an enhanced preferential attachment, has been further enhanced in [35] to enforce a higher diversity in the node counts of the layers.

### B. Our contributions

In this work, we present **A***irline* **N***etworks* **G***eneration* **E***mphasizing* **L***ayers* (*ANGEL*), a novel approach to construct synthetic multilayer networks, especially but not only air transportation networks. We pay particular attention to the structure of the layers and their overlap. In contrast to the preferential attachment approach in [34] or [35], where the multiplex grows edge by edge, in ANGEL the network is built layer by layer on a-priori determined overlapping subsets of nodes. In the creation process of each layer, we distinguish between layers mimicking hub-and-spoke and point-to-point structures, the two patterns typical for air transportation networks. In both approaches, we take the spatial location of the nodes into consideration. Although ANGEL is originally designed to mimic an air transportation network, the flexibility of our model enables a straightforward generalization to create diverse multilayer networks of arbitrary sizes that certainly are of interest to the network science community in general. We follow the air transportation network representation proposed in [33], where an edge in a layer corresponds to a flight offered by the particular airline being modeled by the layer. Nodes represent airports and can show up in several layers. They are captured by connections offered by a particular airline.

The paper is structured as follows. In Section II, we analyze the EATN data and introduce metrics to investigate the multilayer structure of this network. The modular construction of the ANGEL methodology is elaborated in Section III, where we define methods for creating single layers mimicking hub-and-spoke patterns (Section III-A) and point-to-point patterns (Section III-B), and for integrating the layers into a multiplex network (Section III-C). We generalize ANGEL to a model for an arbitrary multilayer network in Section IV. In Section V, we validate the ANGEL model against the EATN. Section VI provides concluding remarks and pointers to future research.

## II. Analysis of the EATN

This section is divided into three parts. In Section II-A, we present a selection of graph statistics on the 37 distinct layers of the EATN. Since the majority of the layers is distinguished by a hub-and-spoke formation, in Section II-B we seek to formulate a definition to indicate hubs among the other nodes in the layer. In Section II-C, we investigate the multiplex as a composition of the layers. An important aspect of our study is the subnetwork of the EATN induced by hubs.

TABLE I
Statistics on the layers and the multiplex (M) of the EATN.

| *layer id* | *node count* | *edge count* | *density* | *transitivity* | *avg degree* | *std degree* | *avg short. path* | *std short. path* | *avg clust. coeff.* | *std clust. coeff.* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 35 | 35 | 0.059 | 0.006 | 2.0 | 5.32 | 1.99 | 0.34 | 0.06 | 0.23 |
| 2 | 35 | 34 | 0.057 | 0 | 1.94 | 5.5 | 1.94 | 0.23 | 0.0 | 0.0 |
| 3 | 36 | 63 | 0.1 | 0.148 | 3.5 | 4.76 | 2.19 | 0.68 | 0.21 | 0.27 |
| 4 | 36 | 37 | 0.059 | 0 | 2.06 | 3.79 | 2.44 | 0.64 | 0.0 | 0.0 |
| 5 | 37 | 43 | 0.065 | 0.038 | 2.32 | 5.54 | 1.98 | 0.34 | 0.14 | 0.32 |
| 6 | 38 | 67 | 0.095 | 0.165 | 3.53 | 5.59 | 2.14 | 0.56 | 0.43 | 0.47 |
| 7 | 38 | 53 | 0.075 | 0.103 | 2.79 | 4.92 | 2.19 | 0.55 | 0.35 | 0.45 |
| 8 | 40 | 57 | 0.073 | 0.044 | 2.85 | 4.7 | 2.43 | 0.88 | 0.24 | 0.4 |
| 9 | 42 | 53 | 0.062 | 0.045 | 2.52 | 6.14 | 1.97 | 0.31 | 0.27 | 0.44 |
| 10 | 42 | 42 | 0.049 | 0.004 | 2.0 | 5.94 | 2.0 | 0.31 | 0.05 | 0.21 |
| 11 | 42 | 41 | 0.048 | 0 | 1.95 | 6.1 | 1.95 | 0.21 | 0.0 | 0.0 |
| 12 | 43 | 99 | 0.11 | 0.188 | 4.6 | 5.21 | 2.28 | 0.7 | 0.24 | 0.31 |
| 13 | 44 | 43 | 0.045 | 0 | 1.95 | 6.26 | 1.95 | 0.21 | 0.0 | 0.0 |
| 14 | 44 | 55 | 0.058 | 0.041 | 2.5 | 5.98 | 2.03 | 0.38 | 0.2 | 0.36 |
| 15 | 44 | 67 | 0.071 | 0.074 | 3.05 | 5.84 | 2.05 | 0.45 | 0.34 | 0.43 |
| 16 | 45 | 92 | 0.093 | 0.022 | 4.09 | 3.48 | 2.66 | 0.88 | 0.01 | 0.03 |
| 17 | 45 | 58 | 0.059 | 0.028 | 2.58 | 5.88 | 2.11 | 0.53 | 0.07 | 0.16 |
| 18 | 45 | 58 | 0.059 | 0.06 | 2.58 | 5.43 | 2.19 | 0.52 | 0.27 | 0.42 |
| 19 | 45 | 45 | 0.045 | 0.004 | 2.0 | 5.89 | 2.08 | 0.41 | 0.04 | 0.21 |
| 20 | 45 | 90 | 0.091 | 0.34 | 4.0 | 2.54 | 3.35 | 1.43 | 0.5 | 0.36 |
| 21 | 48 | 60 | 0.053 | 0.034 | 2.5 | 6.63 | 1.95 | 0.22 | 0.23 | 0.4 |
| 22 | 48 | 69 | 0.061 | 0.034 | 2.88 | 4.91 | 2.51 | 0.86 | 0.08 | 0.2 |
| 23 | 51 | 93 | 0.073 | 0.105 | 3.65 | 7.06 | 1.97 | 0.34 | 0.4 | 0.42 |
| 24 | 52 | 87 | 0.066 | 0.126 | 3.35 | 6.02 | 2.21 | 0.56 | 0.37 | 0.44 |
| 25 | 53 | 73 | 0.053 | 0.104 | 2.75 | 4.63 | 2.55 | 0.72 | 0.17 | 0.32 |
| 26 | 53 | 61 | 0.044 | 0.023 | 2.3 | 4.6 | 2.56 | 0.73 | 0.11 | 0.29 |
| 27 | 59 | 69 | 0.04 | 0.007 | 2.34 | 6.38 | 2.19 | 0.5 | 0.04 | 0.18 |
| 28 | 63 | 62 | 0.032 | 0 | 1.97 | 7.62 | 1.97 | 0.18 | 0.0 | 0.0 |
| 29 | 65 | 66 | 0.032 | 0 | 2.03 | 5.63 | 2.76 | 1.04 | 0.0 | 0.0 |
| 30 | 66 | 110 | 0.051 | 0.107 | 3.33 | 6.62 | 2.31 | 0.59 | 0.4 | 0.46 |
| 31 | 67 | 72 | 0.033 | 0.009 | 2.15 | 7.65 | 2.02 | 0.29 | 0.12 | 0.31 |
| 32 | 75 | 184 | 0.066 | 0.187 | 4.91 | 7.15 | 2.37 | 0.65 | 0.38 | 0.4 |
| 33 | 86 | 118 | 0.032 | 0.026 | 2.74 | 9.05 | 2.02 | 0.28 | 0.26 | 0.4 |
| 34 | 94 | 180 | 0.041 | 0.144 | 3.83 | 4.03 | 3.31 | 1.26 | 0.14 | 0.26 |
| 35 | 99 | 307 | 0.063 | 0.176 | 6.2 | 9.0 | 2.34 | 0.69 | 0.26 | 0.31 |
| 36 | 106 | 244 | 0.044 | 0.1 | 4.6 | 11.14 | 2.16 | 0.49 | 0.55 | 0.47 |
| 37 | 128 | 601 | 0.074 | 0.215 | 9.39 | 11.55 | 2.25 | 0.61 | 0.35 | 0.31 |
| min | 35 | 34 | 0.032 | 0 | 1.94 | 2.54 | 1.94 | 0.18 | 0 | 0 |
| max | 128 | 601 | 0.11 | 0.34 | 9.39 | 11.55 | 3.35 | 1.43 | 0.55 | 0.47 |
| $\mu$ | 54.97 | 96.97 | 0.06 | 0.07 | 3.13 | 6.07 | 2.25 | 0.56 | 0.20 | 0.28 |
| $\sigma$ | 22.04 | 103.00 | 0.02 | 0.08 | 1.45 | 1.85 | 0.34 | 0.29 | 0.16 | 0.16 |
| M | 417 | 3588 | 0.041 | 0.304* | 17.21 | 27.78 | 2.76 | 0.8 | 0.42* | 0.33* |

* These values are calculated for the multiplex downgraded to a simple graph

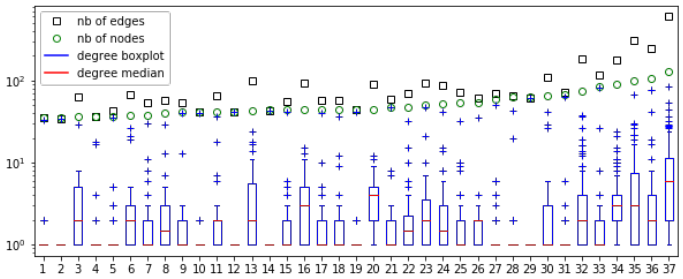


Fig. 1. Node counts, edge counts and degree boxplots per EATN layer.

### A. Layer description

Table I shows a number of statistics for all layers of the EATN. The 37 layers are sorted by the increasing number of nodes. The subsequent four rows report global statistics over the layers, *i.e.*, minimum, maximum, mean and standard deviation. Finally, the last row starting with 'M' corresponds to the *multiplex*, which is a multigraph spanned on all edges belonging to the 37 layers. Figure 1 shows node and edge

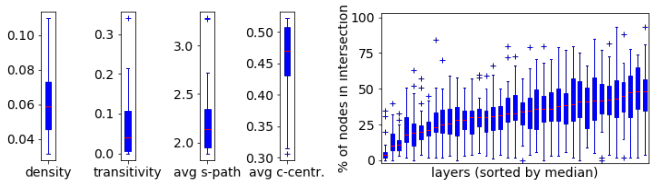


Fig. 2. Boxplots with layer statistics displayed in Table I (left). Boxplots with percentage of node-set overlap per layer in the EATN (right).

counts as well as boxplots with node degrees per layer. Again, layers are sorted increasingly by the node-set size. The boxplots in Fig. 2 consolidate values on the density, the transitivity, the average shortest path and the average clustering coefficient per layer presented in Table I. The right plot in Fig. 2 also shows the node-set overlap of layers. For each layer $L_i$, we compute the overlap percentage of each remaining layer $L_j$ with layer $L_i$ as $p_{ij} = \frac{|V_{L_i} \cap V_{L_j}|}{|V_{L_j}|} \cdot 100\%$, where $V_L$ corresponds to the node set of layer $L$. One boxplot merges $p_{ij}$-values, $1 \leq j \leq 37$, $i \neq j$, for layer $i$. The ordering on the $x$-axis is based on the median of the boxplots.

### B. Hubs

Most of EATN layers have a simple *hub-and-spoke* structure, especially those representing smaller national airlines. However, there are few layers, usually expanded by a commercial airline, which follow the *point-to-point* strategy. The term hub-and-spoke refers to a model of development of the airline network consisting of a relatively small number of central airports, *i.e.*, the hubs, through which most of the flights are routed. Spokes are the routes that planes take out of the hub airport. Major airlines have multiple hubs. However, the bigger an airline is in terms of the number of links, the less planar is the structure of its network. It grows into a *mixture* of a hub-and-spoke and point-to-point patterns [36]. A substantial step in the replication of a network of a single airline is to decide which structure and how strictly it should develop.

In the analysis of EATN layers, we seek a simple and significant criterion for a node to be a hub. One approach to identify candidates for hubs could be to find nodes of a high *degree centrality*, or of a high *betweenness centrality*. Given a layer as a simple graph $G_L = (V_L, E_L)$, $V_L$ being the set of nodes and $E_L$ the set of edges, in addition to the two above measures, we consider the $s$-metric defined in [37] as $s_m(G_L) = \sum_{(u,v)\in E_L} deg(u)deg(v)$, where $deg(u)$ is the degree of node $u$. The *s-metric* measures the extent to which a graph has a hub-like core as it is maximized when high-degree nodes are connected to other high-degree nodes. To break it down on the node level, we define for each $u \in V_L$

$$s_m(u) = \frac{\sum_{v\in N(u)} deg(u)deg(v)}{|E_L|^2},$$

where $N(u)$ is the set of neighbours of node $u$.

Generally, nodes achieving high values for any of the above measures would be tagged as hubs. To figure out the exact limit value, we proceed as follows. We iterate over the layers and the rates $0.1, 0.2, \ldots, 0.9$, and we tag a node as a hub if its corresponding measures value exceeds the rate. Moreover,

TABLE II
STATISTICS FOR A HUB CRITERION.

| rate ($r$) | degree centr. $\sigma(d_c, r)$ | betweenness centr. $\sigma(b_c, r)$ | $s$-metric $\sigma(s_m, r)$ |
|---|---|---|---|
| 0.1 | 65.69 | 5.22 | 35.28 |
| 0.2 | 5.36 | 1.14 | 2.28 |
| **0.3** | 0.44 | 0.78 | **0.14** |
| 0.4 | 0.47 | 1.03 | 0.58 |
| 0.5 | 0.83 | 1.31 | 0.97 |
| 0.6 | 1.19 | 1.58 | 1.31 |
| 0.7 | 1.69 | 1.61 | 1.61 |
| 0.8 | 2.19 | 1.94 | 2.14 |
| 0.9 | 2.36 | 2.36 | 2.39 |

we collect in Table II the standard deviation of normalized degree centrality ($d_c$), normalized betweenness centrality ($b_c$), and $s$-metric ($s_m$),[1] w.r.t. the number of hubs in the EATN layers,

$$\sigma(\mathrm{m}, r) = \frac{\sum_L (h_L - h_L(\mathrm{m}, r))^2}{\#layers - 1},$$

where $h_L$ is the number of hubs in the EATN layer $L$ and $h_L(\mathrm{m}, r)$ is the hub count resulting from the value of measure $\mathrm{m}$ (*i.e.*, $d_c, b_c, s_m$) compared to the rate $r$.

As it can be noted from Table II, the rate $0.3$ appears to be the limit value for all statistics, *i.e.*, the value associated with the smallest standard deviation. Also, the minimum standard deviation corresponds to the $s$-metric. This leads us to a definition of hub, referred to in the remainder of this work.

*Definition 2.1: A* **hub** *is a node having s-metric greater than 0.3. Consequently, a layer has a hub-and-spoke structure, if it contains at least one hub.*

In the upper plot of Fig. 3, we display hub numbers per layer resulting from each statistic, and compare them with the EATN reference. The ticks on the $x$-axis correspond to layers sorted by the increasing number of hubs in the reference layers. On the $y$-axis, for each layer, we plot the number of nodes for which the values of the degree centrality (green triangles), the betweenness centrality (blue squares), and the $s$-metric (black diamonds) exceeded the limit value $0.3$, and the reference hubs count (red circles). Layers with the number of hubs equal to zero are considered point-to-point structured.

Exploring the EATN layers, we also found out that the hub count can be roughly linked with the ratio $r_L = \frac{|V_L|}{|E_L|}$. This relation is shown in the bottom plot of Fig. 3. Here, all 37 EATN layers are sorted increasingly by $r_L$ and the $x$-axis is labeled by $r_L$-values. On the $y$-axis, a dot matches the number of hubs in the corresponding layer. As it can be noted, the highest hub counts belong to layers with $r_L$-value in the range $(0.4, 0.7)$ and the majority of layers with the ratio greater than $0.8$ contains only one hub. This dependency between hub count and node-to-edge-count ratio is helpful to forecast how many hubs will evolve on a layer with a given node and edge counts. In fact, this is leveraged by Algorithm 2 to assign a number of hubs to a layer. Only layers with $r_L$ value close to 1, and usually those having exactly one hub, evolve a clear hub-and-spoke structure. Most of the layers build a mixture of the hub-and-spoke and point-to-point patterns and have the

[1] All considered measures range within [0,1].

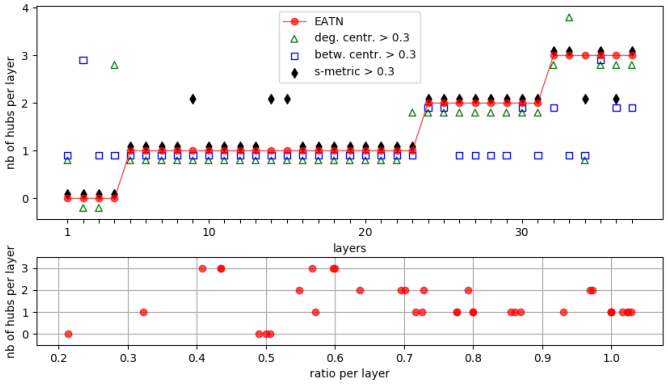


Fig. 3. Number of hubs per layer in the EATN: actual, according to the degree and betweenness centrality, and $s$-metric value (top). Number of hubs per layer in the EATN with respect to the ratio $r_L = |V_L|/|E_L|$ (bottom).

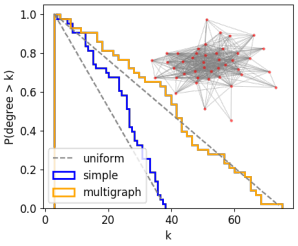


Fig. 4. The hub-subnetwork of the EATN with the degree histograms and uniform fit.

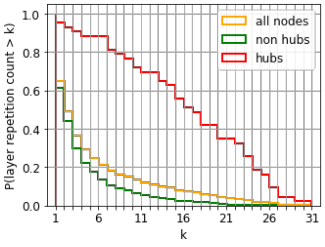


Fig. 5. A cumulative histogram of the layer repetition count per node in the EATN multiplex.

$r_L$ value in the range $[0.5, 0.9]$. Examples of the layouts are shown in the top row of Fig. 9.

### C. Multiplex description

The EATN multiplex network consists of 417 non-zero degree nodes and $3,588$ edges; after discarding the edge multiplicity, the number of edges is $2,953$. Further statistics are shown in the last line of Table I. In this and the following sections, unless otherwise specified, the probability distribution functions we will present are complementary cumulative.

We begin with the analysis of the subnetwork induced by hubs, according to Def. 2.1. The hub-subnetwork has 54 nodes, out of which only 11 are duplicates belonging exactly to two layers. Counting edge multiplicities, the multigraph induced by the 43 unique hubs has 851 edges, a density of 0.942 (and 516 edges, a density of 0.571 if discarding duplicate edges), and a high transitivity of 0.733. Figure 4 shows the degree histograms for both the simple (blue line) and the multigraph (orange line), and their distribution fittings. According to the Kolmogorov-Smirnov (KS) test, the degree distribution of the multigraph as well as of the simple graph fit the uniform distributions $\mathcal{U}(2.99, 72.00)$ and $\mathcal{U}(2.99, 35.38)$, respectively, with the $p$-value greater than 0.5 in both cases.[2]

In the EATN, a small number of nodes (11) are distinguished as hubs in different layers, but many hubs repeat in other layers (even as one-degree nodes). Therefore, to capture an essential indicator of the inter-layer structure of the multiplex, we counted the number of layers every node appears in. Figure 5 shows the cumulative histogram of the *layer repetition count* values for all nodes in the EATN multiplex (yellow line). These counts are broken down by hubs (red line) and non-hubs (green line). We observe that most of the non-hub nodes repeat in just a few layers as opposed to hubs tending to have a high layer overlap. In fact, the non-hub distribution has a tailed shape, that we fit to both power-law and exponential probability distributions. The fitting to the power-law probability density function,

$$\mathrm{PDF}_p(x, \alpha, x_m) = \frac{\alpha - 1}{x_m} \left( \frac{x}{x_m} \right)^{\alpha}, \tag{1}$$

yields the parameters $\alpha = 2.039$ and $x_m = 1$. When fitted to the exponential probability density function,

$$\mathrm{PDF}_e(x, l, s) = \frac{1}{s} e^{(-\frac{x-l}{s})}, \tag{2}$$

we obtain $s = 3.88$ and $l = 1.0$. Corresponding plots are shown in Fig. 6 (with log-scaled $y$-axis and non-cumulative distribution plot, in the inset). With the $p$-value 0.424 returned by the KS test, we find out a good fit of the power-law distribution. The test applied for the exponential distribution does not confirm the apparently good fit due to the big skip caused by the large count of nodes appearing in exactly one layer. This corresponds to a maximum positive deviation of the sample cumulative distribution from the hypothesized curve equal to $D^+ = 0.3871$; nonetheless, the maximum negative deviation is $D^- = 0.0488$, with $p$-value 0.169.

A simple but meaningful indicator of the diversity of the layers composing a multiplex is the shape of the distribution of their node and edge counts. Cumulative histograms along with fitted power-law and exponential distribution functions are shown in Fig. 7. The outcomes of the fittings are as follows. For node counts, we obtained $\mathrm{PDF}_p(x, 4.231, 35)$ and $\mathrm{PDF}_e(x, 34.99, 19.75)$. For edge counts, the fitted probability functions are $\mathrm{PDF}_p(x, 2.87, 38)$ and $\mathrm{PDF}_e(x, 33.99, 62.86)$. According to the KS test, both power-law distributions can be highly trusted with $p$-value close to 1 for both node and edge counts. The exponential distributions deliver $p$-values 0.4326 and 0.4301, respectively.

As we shall describe next, the outcomes of the above analysis of the EATN layers will have effect on the definition of our proposed network generation model.

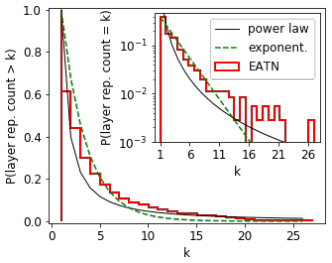


Fig. 6. Distribution fittings to layer repetition counts per non-hubs in the EATN multiplex.

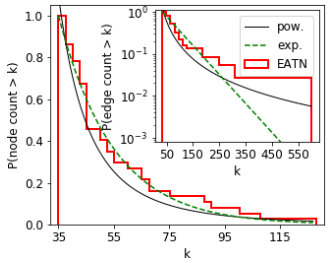


Fig. 7. Node (frame) and edge (inset) counts of the EATN layers with distribution fitting.

[2]Small $p$-values (usually $< 0.05$) will indicate that the test rejected the hypothesis that the original data could have been drawn from the fitted distribution.

**Algorithm 1:** ANGEL hub-and-spoke layer

**Input:** layer node count $n$, layer edge count $m$
(a) Assign the number of hubs to the layer
(b) Assign nodes to grid points and compute the minimum spanning tree
(c) Contract the tree till hubs evolve
(d) Select nodes for bridges - linked to two hubs
(e) Add remaining edges
**Output:** A synthetic layer with $n$ nodes and $m$ edges

**Algorithm 2:** Assigning the number of hubs to a layer

**Input:** layer node count $n$, layer edge count $m$
$r_L = \frac{n}{m}$, $n_h = 1$
**if** $r_L < 0.95$ **then**
  **foreach** $1 \le i \le \lfloor 10(1 - r_L) \rfloor$ **do**
    **if** rand$(0, 1) < 1 - r_L$ **then**
      $n_h = n_h + 1$
**Output:** $n_h$, number of hubs
rand$(0, 1)$ is a random number within the range $(0, 1)$.

## III. The ANGEL Model

In this section, we establish a methodology to simulate the discussed structure of networks of the EATN type. We focus on the multilayer structure of the EATN to reproduce its inner formation of the layers as well as its inter-layer composition. Our approach is first to distribute the nodes of the multiplex among the layers, enforcing their overlapping. Then, each layer attaches edges separately but simultaneously contributes the links to the whole multiplex due to nodes shared with other layers. In Section III-A and III-B, we propose algorithms to create a layer with a hub-and-spoke and point-to-point formation, respectively. Consequently, in Section III-C, we introduce a methodology to compose the layers into a multiplex.

### A. Generating a hub-and-spoke-structured layer

Given $n$ nodes and $m$ edges, our method for creating a hub-and-spoke layer consists of the following major steps, which are summarized in Algorithm 1.

(a) Intuitively, and as we previously observed in Fig. 3 (bottom plot), the more edges and the fewer nodes a layer has the more hubs are needed. Algorithm 2 bases on this observation. Layers with $r_L$ close to 1.0 obtain exactly one hub. The lower the ratio, the higher probability that further hubs will be assigned to the layer.

(b) Our generative model is designed to replicate a transportation network. In networks of this type, hubs usually are nodes with a central geographical location. To incorporate the spatial aspect in our model, we randomly distribute nodes of a layer on a rectangular grid of the size $k \times k$ with $k = 3\lfloor\sqrt{n}\rfloor$. The grid points have discrete coordinates in the range $[0, k]$. Next, the minimum spanning tree, say $T$, is computed with respect to the Euclidean distance between the nodes.

(c) The purpose of this step is to identify nodes in $T$ to be the hubs of the layer. For that, we iteratively reconnect low degree nodes with nodes close to the spatial center of the tree. This contraction of the tree is performed in three steps:

**Algorithm 3:** Contraction with respect to degree $d$

**Input:** a tree $T = (V_T, E_T)$, a degree value $d$
**foreach** $u \in V_T$ with $deg(u) = d$ **do**
  $L = \{s \in V_T \,|\, (u, s) \in E_T\}$, sort $L$ by decreasing degree
  $v = L.\text{pop}(0)$
  **foreach** $s \in L$ **do**
    remove edge $(u, s)$ from $E_T$ and append edge $(v, s)$
**Output:** $T = (V_T, \overline{E_T})$

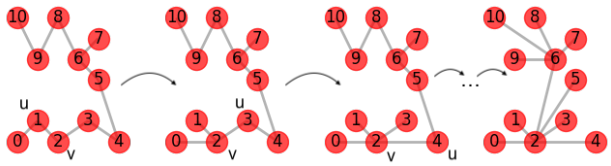


Fig. 8. Contraction of a tree with respect to degree 2.

(c.1) Contract $T$ w.r.t. degree 2 as outlined in Algorithm 3, which works as follows: for every node $u$ of degree $d$, the highest degree neighbour, say $v$, is selected; all other neighbours of $u$ are disconnected from it and linked to $v$ (cf. Fig. 8).

(c.2) Discard all leaf-edges from $T$. Contract $T$ to degree 2. Add all previously discarded edges.

(c.3) Contract $T$ w.r.t. degree $d = 2, 3, \ldots$ until as many hubs are evolved as many assigned to the layer in step (a). Note that, at this stage, we have a very sparse graph, and hence Definition 2.1 cannot be applied to identify hubs. Therefore, a node is tagged as a hub if all its three quantities (degree, betweenness, and the $s$-metric) exceed the value 0.1. Once the hubs are tagged, we make sure that all leaves are connected to the closest hub.

(d) Bridges, *i.e.*, nodes linked to two hubs, are built in this step. The obvious candidates are nodes located relatively close to both hubs. Therefore we create a list of all possible edges, not yet in the tree, with one end being a hub. The list is sorted increasingly by the distance. Next, we randomly set a number $\overline{m}$ which amounts between 10 and 50% of the predetermined layer edge count. The first $\overline{m}$ edges from the list are added to the tree.

(e) Finally, to maintain the hub-and-spoke structure of the layer, the remaining amount of edges is attached in such a way that low degree nodes but leaves are preferably linked with high degree nodes but hubs. If there are still leftover edges, the preferential attachment method in Algorithm 4 is performed.

### B. Generating a point-to-point-structured layer

The routine to replicate a layer according to the point-to-point strategy is outlined in Algorithm 5 and described next.

(a) We start with assigning nodes to points on a grid and compute a minimum spanning tree (cf. Algorithm 1, step (b)).

(b) While adding further edges, we try to avoid too long and too short distances. Therefore, we consider a list $\overline{E}$ of all potential edges of the layer not yet in the tree and sort it by lengths. Next, we sample random numbers with respect to the normal probability distribution with $\mu$ and $\sigma$ calculated

**Algorithm 4:** Preferential attachment

**Input:** a graph $G = (V, E)$, number of edges $m$ to add
**foreach** $k = 1 \dots m$ **do**
  $P = \left\{ \left( u, \frac{\deg(u)}{\sum_{v \in V} \deg(v)} \right) \mid u \in V \right\}$
  sample randomly nodes $s, t \in V$, $s \neq t$ with respect to the probability distribution $P$
  **while** $(s, t) \in E$ **do**
    repeat the above sampling
  add $(s, t)$ to the set of edges $E$
**Output:** $G = (V, \overline{E})$

**Algorithm 5:** ANGEL point-to-point layer

**Input:** layer node count $n$, layer edge count $m$
(a) Assign nodes to grid points and compute a minimum spanning tree $T = (V_T, E_T)$
(b) Sort $\overline{E} = \{(u, v) \mid u, v \in V_T,\ (u, v) \notin E_T\}$ by edge len., select randomly $\overline{m}$, $1 \leq \overline{m} \leq m - |E_T|$, pick up $\overline{m}$ edges with lengths close to the average.
(c) Add remaining edges using preferential attachment.
**Output:** A synthetic layer with $n$ nodes and $m$ edges

on the lengths of edges in $\overline{E}$ and select edges with lengths close to these random values. The number of edges attached in this manner is randomly chosen from the range $(1, m - m_T)$, where $m$ is the predetermined edge-set size of the replicated layer and $m_T$ is the edge count of the minimum spanning tree calculated in step (a).

(c) The remaining edges are added according to a preferential attachment (cf. Algorithm 4).

Figure 9 presents four examples of layer replicas generated according to Algorithms 1 and 5. In the first column to the left examples of point-to-point-layers are shown, and hub-and-spoke-layers in the remaining three. The ratio $r_L$ and the number of hubs according to Def. 2.1 are displayed above each layer.

### C. Generating an ANGEL multiplex

The algorithms for layer creation we presented in the previous section, are substantial steps in the methodology to generate a whole multiplex. As outlined in Algorithm 6, the procedure runs in four main steps. First, nodes (distributed on a grid) are assigned to layers. Next, the hubs are distinguished within the layers and connected across the layers. After that, edges are supplemented to the layers. Finally, the multiplex emerges as a multigraph on the union of the layers.

A crucial step in our methodology is to establish a subnetwork based on all hubs within the multiplex, i. e. to connect hubs from different layers. This has to be done before all edges are added to the layers.

All input data for Algorithm 6 is introduced in Table III. To account for the arbitrary size of the network as well as its multilayer structure, we use a probability distribution on node counts per layer, $P_{nodeL}$, and edge counts per layer, $P_{edgeL}$. Furthermore, to control the node overlap across the layers (Fig. 2), we apply a fitted distribution of the layer repetition count per node, $P_{layerN}$.

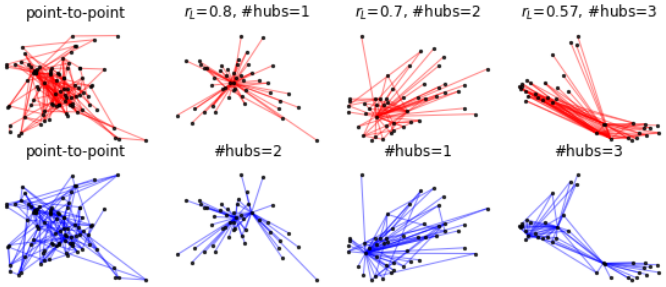


Fig. 9. EATN layers (red, top) versus their ANGEL replicas (blue, bottom). Node positions correspond to their geographical location.

TABLE III
NOTATION USED IN SECTION III-C

| Notation | Meaning |
|---|---|
| $\ell$ | number of layers, $\ell \in \mathbb{N}$ |
| $p$ | number of layers with the point-to-point structure, $p \leq \ell$ |
| $n$, $m$ | $n \in \mathbb{N}$, number of nodes, $m \in \mathbb{N}$, number of edges |
| $P_{layerN}$ | the probability distribution of the layer repetition count per node, used for the random selection of the number of layers each node appears in |
| $P_{nodeL}$ | the probability distribution of the node count per layer, used for the random layer selection as nodes get assigned to layers |
| $P_{edgeL}$ | the probability distribution of the edge count per layer, used for the random edge-set size assignment to layers |

Algorithm 6 works in detail as described next. Foremost, a list of $\ell$ simple graphs is created and $p$ of them is tagged randomly to evolve the point-to-point structure. To incorporate the spatial aspect of the mimicked multiplex, all $n$ nodes of the synthetic multiplex are randomly distributed on a grid in the same manner as nodes of a single layer in step (b) of Algorithm 1.

Nodes: A multiplex with $n$ isolated nodes is initialized. To each node $u$ we assign its layer repetition count $r_u$ by sampling a number in $\{1, 2, \dots, \ell\}$ with probability weights $P_{layerN}$. Next, each node $u$ is added to $r_u$ different layers selected randomly according to the probability distribution $P_{nodeL}$.

Hubs: This step is performed only for layers of the hub-and-spoke structure. For each of these layers Algorithm 1.(a)-(c) is called. At this stage, hubs are identified and connected to a clique within the layer. Next, hubs are connected across the layers. According to Fig. 4 the node degrees of the hub-subnetwork of the EATN are nearly uniformly distributed. Thus we mimic this subnetwork by using the configuration model with respect to a degree sequence sampled uniformly between 1 and $n_H$, where $n_H$ is the total hub count in the multiplex. Self-loops are removed.

Edges: $\ell$ numbers are sampled randomly with respect to the probability distribution $P_{edgeL}$. Each number corresponds to one layer's edge-set size. Large values are assigned to layers having a large node-set size. If a layer is tagged as point-to-point, we call Algorithm 5, and Algorithm 1.(d)-(e) otherwise.

Multiplex: The multiplex emerges as a multigraph obtained as the union of all nodes, discounting repetition, and all the layers' edges, allowing the repetition from different layers.

**Algorithm 6:** The ANGEL multiplex creation method

**Input:** $\ell$, $n$, $m$, $p$, $P_{layerN}$, $P_{nodeL}$, $P_{edgeL}$
Initialize:
$P$ - a list of $p$ randomly selected different numbers in the range $1 \dots \ell$ indicating point-to-point layers
$L_1, \dots, L_\ell$ - empty graphs representing layers
$H$ - an empty list to store hubs
map each node $1, \dots, n$ to a unique, randomly selected point on the grid $3\lfloor\sqrt{n}\rfloor \times 3\lfloor\sqrt{n}\rfloor$
Nodes:
**foreach** $u = 1$ **to** $n$ **do**
  sample layer repetition count, $r_u$, applying $P_{layerN}$
  with respect to weights $P_{nodeL}$ select randomly $r_u$ different layers from $L_1, \dots, L_\ell$ to place $u$ in
Hubs:
**foreach** $i = 1$ **to** $\ell$, $i \notin P$ **do**
  call Algorithm 1.(a)-(c)
  make a clique on all hubs in $L_i$
  add all hubs of layer $L_i$ to $H$
create a multigraph on nodes in $H$ using configuration model
Edges:
sample $\ell$ numbers w.r.t. the probability distribution $P_{edgeL}$
each value indicates a layer edge count
assign high values to layers with high node counts
**foreach** $i = 1$ **to** $\ell$ **do**
  **if** $i \in P$ call Algorithm 5 **else** call Algorithm 1.(d)-(e) **end**
Multiplex:
create a multigraph on the layers $L_1, \dots, L_\ell$
**Output:** The layers $L_1, \dots, L_\ell$ and the multiplex

**Algorithm 7:** The MGEL algorithm

**Input:** $\ell$, $n$, $m$, $P_{layerN}$, $P_{nodeL}$, $P_{edgeL}$
Initialize:
$L_1, \dots, L_\ell$ - list of empty graphs representing layers
$M$ - multiplex with $n$ isolated nodes
Assign nodes to layers:
**foreach** $u$ *in* $M$ **do**
  sample layer repetition count, $r_u$, from the $P_{layerN}$
  use $P_{nodeL}$ to select $r_u$ different layers from $L_1, \dots, L_\ell$ to place $u$ in
Assign edge-set sizes to layers:
use $P_{edgeL}$ to assign number of edges to $L_1, \dots, L_\ell$
Create layers:
**foreach** $i = 1$ **to** $\ell$ **do**
  call a layer creation procedure for $L_i$
  add all edges from $L_i$ to $M$
**Output:** The layers $L_1, \dots, L_\ell$ and the multiplex $M$

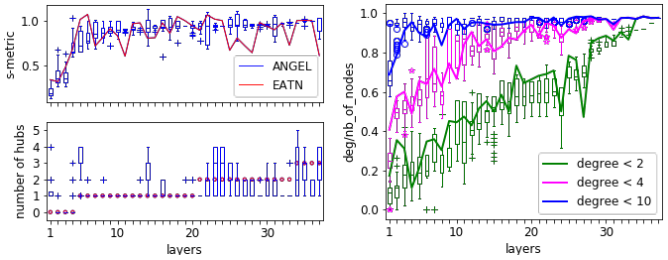


Fig. 10. Boxplots of $s$-metric values (top left), hub counts per layer (bottom left), and normalized counts of three representative degree groups (right) of 100 ANGEL layer-replicas versus EATN references.

## IV. A Generalization of the ANGEL Model

ANGEL is designed to mimic an air transportation network. Nonetheless, a generalization of Algorithm 6 can in principle be used to generate arbitrary multilayer networks. The input of the model allows not only a free selection of the network's global values like node, edge and layer count. With arbitrarily chosen probabilities $P_{nodeL}$ and $P_{edgeL}$, one can control the node and edge counts of the layers. The same applies to the weights $P_{layerN}$, which are meant to stimulate the overlap among the layers. The creation methods of single layers, Algorithm 1 and 5, are called as plug-ins within Algorithm 6. This also enables other layer creation methods (e.g., the preferential attachment method); the same applies to the method for building the hub-network, which can be even disregarded. These generalizations lead to the *Multiplex Generation Emphasizing Layers* (MGEL) model, summarized in Algorithm 7.

## V. Analysis of the ANGEL model

We organize the analysis of ANGEL into two main parts: layers and the multiplex.

### A. ANGEL layers

To validate the layer replication methods we create 100 replicas of each of the 37 EATN layers. The number of nodes and edges of the reference layer is taken as the input for the 100 replicas. The decision about which algorithm to call—a hub-and-spoke or point-to point layer—is made as follows. In Section II-B we pointed out that the $s$-metric is a helpful measure to identify graphs with a hubs-spoke structure. The higher the value, the more apparent the hub-and-spoke formation. Based on experimental expertise, if the reference layer has the normalized $s$-metric[3] at least $0.5$ we call Algorithm 1 otherwise Algorithm 5. The boxplots in the left top of Fig. 10 consolidate the $s$-metric values of the 100 copies. EATN values are connected by the solid red line. Layers are depicted on the $x$-axis and ordered by the median of the boxplots. The three first layers (with the lowest $s$-metric value) are of the type point-to-point. As expected, hub-and-spoke structured layers exhibit high values of the $s$-metric even if they are not always coherent with the EATN reference. The standard deviation values—computed for each boxplot with respect to the corresponding EATN value—vary between $0$ and $0.37$ and have an average of $0.07$. The bottom left plot in Fig. 10 shows the performance of the prediction method, how many hubs should span a layer, formulated in Algorithm 2. On the $y$-axis we capture the hub counts in the layers, and layers are here sorted according to the increasing number of hub counts in the EATN references. The standard deviation varies between $0$ and $13.89$, with average of $2.35$. As one can observe, the volatility of predicted hub numbers increases if there is more than one hub in the references.

The right plot in Fig. 10 shows a normalized count of three representative node degree groups. The $y$-values reflect the normalized count of vertices of degree less than $t$, $t \in \{2, 4, 10\}$. The actual counts are divided by the number

[3]The $s$-metric is normalized by dividing it by squared number of edges in the considered graph.

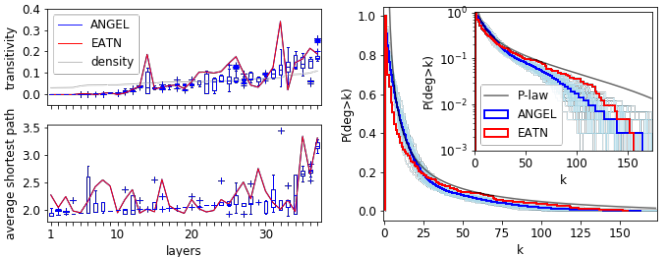


Fig. 11. Boxplots with clustering coefficients (top left) and average shortest path lengths (bottom left) of 100 ANGEL layer replicas versus EATN references. The cumulative degree histograms of 100 ANGEL multiplex replicas versus the EATN reference, with logarithmic scale on $y$-axis in the inset (right).

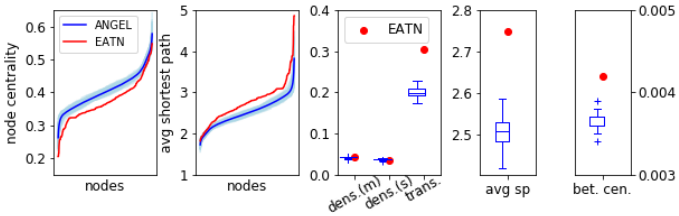


Fig. 12. Statistics of 100 ANGEL multiplex replicas versus the EATN.

of vertices in the particular layer. Each boxplot per layer merges the values of the 100 replicas. The EATN values are represented by the solid line. Layers are depicted on the $x$-axis and are ordered by the median of the green group. Except for a few outliers, most layers evolve in a similar manner as their original counterparts with respect to the degree distribution.

In the left chart of Fig. 11, we display boxplots with the transitivity (top) and the average shortest path lengths (bottom) of the 100 samples along with the respective values of EATN layers (connected as usual with a red line). In both charts, the layers ($x$-axis) are ordered by the median of the boxplots. The transitivity values close to zero correspond to the star-shaped layers, which our method perfectly imitates. The maximal standard deviation over all boxplots is 0.14, which hints that ANGEL may reproduce the typically low-clustered structures of the airline subnetworks. Note that the density of EATN layers (grey line in the upper plot) is perfectly imitated by all ANGEL replicas. Also, the synthetic layers tend to have shorter path lengths than the originals. Nevertheless, the standard deviation is 0.23 on average, which confirms short path lengths typical for hub-and-spoke structured graphs.

### B. ANGEL multiplex

In this section we compare the ANGEL multiplex generated using Algorithm 6 with the EATN. For our tests, we sampled 100 synthetic multilayer networks, and set $\ell = 37$, $p = 4$, $n = 417$, $m = 3588$ as extracted from the EATN. The analysis of the EATN urged to apply $\mathrm{PDF}_e(x, 1, 3.88)$ for $P_{layerN}$ and $\mathrm{PDF}_e(x, 40, 20)$, and $\mathrm{PDF}_e(x, 40, 60)$ for $P_{nodeL}$ and $P_{edgeL}$, respectively (cf. Figs. 6 and 7).

In the right plot of Fig. 11, we compare the degree distributions of the multiplex networks. The reference EATN is represented by a red line, whereas the thick blue line corresponds to the average over all the 100 degree distribution histograms, plotted in light blue.[4] The inset shows the curves with a logarithmic scale on $y$-axis. The degrees of the reference vary between 1 and 156. The maximum degree over all replicas is 283, though the average over the highest degree stays close to 165. As one can observe, all replica histograms along with their average follow the EATN shape.

[4]The average curve is calculated as follows: The list of values per each replica is sorted increasingly. Values of all replicas are added per position and the per-position average is taken. This procedure also applies through other statistics shown in this section.

TABLE IV
STATISTICS ON THE ANGEL MULTIPLEX REPLICAS.

| | ANGEL | | EATN |
|---|---|---|---|
| | $min$ | $max$ | |
| *max degree* | 103 | 283 | 156 |
| *min av. short. path (per node)* | 1.5566 | 1.8746 | 1.8216 |
| *max av. short. path (per node)* | 3.1108 | 4.4771 | 4.8602 |
| *min node centrality (per node)* | 0.2228 | 0.3206 | 0.2052 |
| *max node centrality (per node)* | 0.5321 | 0.6408 | 0.5476 |
| *density multigraph* | 0.0371 | 0.0422 | 0.0417 |
| *density simple graph* | 0.0327 | 0.0396 | 0.0343 |
| *transitivity* | 0.1734 | 0.2287 | 0.3035 |
| *average shortest path* | 2.4189 | 2.5865 | 2.7483 |
| *betweenness centrality* | 0.0034 | 0.0039 | 0.0042 |

In Fig. 12, we present a selection of statistics for 100 ANGEL multiplex replica. From the left, per multiplex the average shortest path length per node is calculated and the values sorted increasingly are connected by a line. Next, the average node centrality is analogously calculated. According to the plots, ANGEL multiplex networks have on average shorter paths but higher node centrality. As one can read from the remaining plots in Fig. 12, the density of the multiplex—with (dens. (m)) or without (dens. (s)) multiple edges—is reproduced by ANGEL. The remaining statistics on transitivity, average shortest path, and betweenness centrality in ANGEL multiplexes stay generally below their EATN counterparts, though the difference is not striking. For a better overview, in Table IV we list the minimum and maximum values over all 100 multiplex replicas per statistic.

The similarity of the node and edge counts in ANGEL replicas and in the EATN (left plot in Fig. 13) is not surprising, since it results from application of the fitted probability distributions to sample these sizes in the ANGEL model. The same applies for the layer repetition count per node (right plot in Fig. 13). However, in the ANGEL multiplex there are nodes present in all layers, and these are usually hubs (as consequently identified by the application of Def. 2.1). In the EATN, the maximum layer-repetition count amounts to 27. These discrepancies are due to the fact that the assignment of the nodes to hubs is the least controlled in the model. As for that matter, the performance of this statistic is appealing.

Statistics shown in Fig. 14 consider the subnetwork induced by hubs. The range of degrees in the ANGEL model exceeds the EATN's range on both sides. There are isolated nodes in the ANGEL, which is not the case in the EATN. Also, ANGEL hubs tend to have higher degrees, which is coupled with the higher layer repetition counts observed in the previous statistic. Analyzing the density and the transitivity of the subnetwork, we observe a very good approximation of the ANGEL values

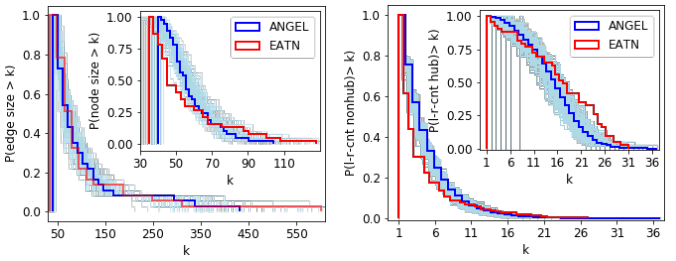


Fig. 13. Layer node and edge count distributions (left) and layer repetition counts per node (right) of 100 ANGEL multiplex replicas versus the EATN.

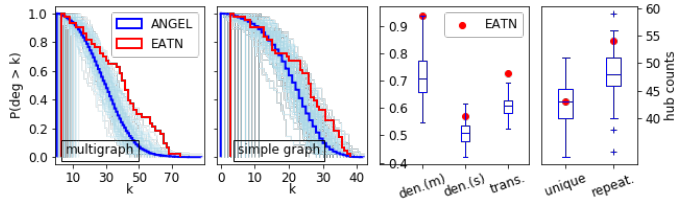


Fig. 14. Statistics on the hub subnetwork on 100 ANGEL multiplex-replicas versus the EATN reference. From the left: degree distribution and density with and without multiple edges, transitivity, unique and repeated hub counts in the multiplex.

towards the EATN. As the last boxplots to the right show, we obtain a very good coherence in the hub counts. This is an implicit outcome of Algorithm 2, which runs at the beginning of the layer creation. During the layer formation process, further nodes may evolve to hubs, or nodes assigned to be hubs may not qualify to be a hub in the end.

### C. Mesoscale structure analysis

In a further stage of evaluation, we share some insights into the mesoscale structure of the ANGEL layers and multiplex compared to the reference EATN. First, we investigated structural similarity aspects. Figure 15 shows the cosine similarity per node-pair (i.e., the number of common neighbours of the two nodes divided by the geometric mean of the two nodes' degrees) of a randomly selected ANGEL-multiplex replica (lower-left) versus the EATN (upper-right). Nodes, on $x$- and $y$-axis, are sorted by the degree. According to the plot, in both cases, the fraction of common neighbours increases with the growing degree of the nodes in both cases - the darkest area can be observed in the lower left corner. Apparently, hubs share their neighbours. However, ANGEL's transition is smoother in comparison to the EATN, which features a partially discrete texture.

Figure 16 summarizes a comparison between the $k$-core decomposition obtained on the EATN compared to ANGEL layers and multiplex. We generated 100 ANGEL-replicas per layer and corresponding multiplexes, then we calculated the *coreness* of each node in a layer-replica and in the multiplex, and finally evaluated the correlation with the corresponding node's coreness in the EATN. Results are displayed in the boxplots of Fig. 16 (boxplots corresponding to the layers are sorted by the median value.) Recall that the coreness of a node is the largest value $k$ of a $k$-core containing that node, where a $k$-core is a maximal subgraph that contains nodes of degree at least $k$. It is noticeable that most of the ANGEL

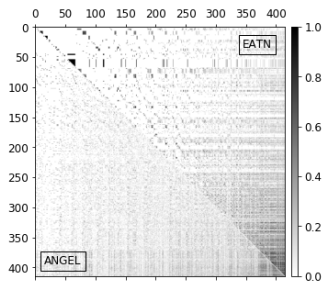


Fig. 15. Cosine similarity per node pair in an ANGEL multiplex (lower-left) in comparison to the EATN (upper-right).

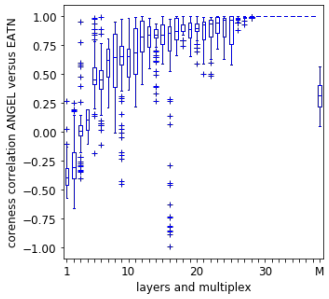


Fig. 16. Boxplots of coreness correlation over 100 ANGEL replicas per layer, and the ANGEL multiplex (M), versus the EATN.

layers' boxplots correspond to moderately or highly positive correlation with EATN references. In fact, by averaging over all layers and multiplex, the mean of correlation versus EATN is 0.73 (with standard deviation 0.11).

Due to space limitations of this article, we shall not delve into any deeper into this mesoscale analysis. Nonetheless, we would like to mention main qualitative findings in terms of community structures that were detected in the EATN layers and the ANGEL counterparts. By considering communities identified by either betweenness-based or modularity-optimization-based methods as case in point, we found out that the layer graphs of EATN against the ANGEL ones often have similar community-structures, with comparable number of communities especially in the case of star-shaped (*i.e.*, hub-and-spoke) structures. Details can be found in Sect. II of the Supplemental Material.

## VI. Conclusions and Further Directions

In this work, we presented ANGEL, a novel approach to generate synthetic multilayer networks. We focused on creating a multiplex having majority of the layers with a hub-and-spoke structure to mimic a benchmark airline traffic network, the EATN. We considered this network from the local point of view, broken down by its layers, and from the global point of view as a layer combination within a multiplex. The ANGEL model convinces with its variety of aspects taken into account and the deep concern for the structure layers composing the multiplex. Also, the ANGEL model refrains from known so far preferential attachment approaches and allows to design a framework for an arbitrary random multiplex network. We nonetheless observed that it is extremely difficult to fully simulate the complexity of a multilayer network, especially the inner- and inter-layer formation. Despite a wide portfolio of statistical analysis we performed in this work, there is still need for methods capturing simultaneously the global and local structure of multilayer networks.

Our ANGEL model paves the way for the development of several applications. Some of these are of major interest to airline companies for pursuing a variety of goals, such as organizing the airline routes in a cost-efficient way, simulating and validating hub-and-spoke, point-to-point or mixed strategies, and investigating their system robustness against various kinds of disruptions. A generative model like ours can

also aid to address "indirect" applications. For example, if we consider the millions of people traveling by flight every day, air transportation networks play a key role in the spread of some infections, hence ANGEL would be valuable to model and simulate the diffusion of epidemics through air flights. Another example of alternative application related to the massive use of air transports refers to support for counter-terrorism: modeling a synthetic network, *e.g.*, to simulate potential or preferential routes taken by selected targets would be helpful to unveil important behavioral patterns of those people.

## REFERENCES


[1] A. Cardillo, M. Zanin, J. Gómez-Gardenes, M. Romance, A. J. G. del Amo, and S. Boccaletti, "Modeling the multi-layer nature of the european air transport network: Resilience and passengers re-scheduling under random failures," *The Europ. Phys. J. Special Topics*, vol. 215 (1), pp. 23–33, 2013.

[2] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3 (4), 2013.

[3] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2 (3), pp. 203–271, 2014.

[4] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Phys. Rep.*, vol. 544 (1), pp. 1–122, 2014.

[5] E. Cozzo, G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, "Multilayer networks: metrics and spectral properties," in *Interconn. Networks*, 2016.

[6] P.-Y. Chen and A. O. Hero III, "Multilayer spectral graph clustering via convex layer aggregation," *arXiv preprint 1609.07200*, 2016.

[7] F. Battiston, J. Iacovacci, V. Nicosia, G. Bianconi, and V. Latora, "Emergence of multiplex communities in collaboration networks," *PloS one*, vol. 11 (1), 2016.

[8] C. De Bacco, E. A. Power, D. B. Larremore, and C. Moore, "Community detection, link prediction and layer interdependence in multilayer networks," *arXiv:1701.01369*, 2017.

[9] F. Battiston, V. Nicosia, V. Latora, and M. S. Miguel, "Robust multiculturality emerges from layered social influence," *arXiv preprint arXiv:1606.05641*, 2016.

[10] M. Barthélemy, "Spatial networks," *Ph. Rep.*, vol. 499, pp. 1–101, 2011.

[11] J. Y. Kim and K.-I. Goh, "Coevolution and correlated multiplexity in multiplex networks," *Phys. Rev. Lett.*, vol. 111 (5), 2013.

[12] M. Magnani and L. Rossi, "The ml-model for multi-layer social networks," in *Proc. IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining*, 2011, pp. 5–12.

[13] V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy, "Growing multiplex networks," *Phys. Rev. Lett.*, vol. 111 (5), 2013.

[14] V. Nicosia and V. Latora, "Measuring and modeling correlations in multiplex networks," *Phys. Rev. E*, vol. 92 (3), 2015.

[15] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, "Ranking in interconnected multilayer networks reveals versatile nodes," *Nature comm.*, vol. 6, 2015.

[16] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, "Modeling the dynamical interaction between epidemics on overlay networks," *Phys. Rev. E*, vol. 84 (2), 2011.

[17] B. Min, S. Do Yi, K.-M. Lee, and K.-I. Goh, "Network robustness of multiplex networks with interlayer degree correlations," *Phys. Rev. E*, vol. 89 (4), 2014.

[18] G. Bianconi, "The entropy of randomized network ensembles," *EPL*, vol. 81 (2), 2007.

[19] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p*) models for social networks," *Soc. Networks*, vol. 29 (2), pp. 173–191, 2007.

[20] G. Bianconi, "Statistical mechanics of multiplex networks: Entropy and overlap," *Phys. Rev. E*, vol. 87 (6), 2013.

[21] A. Halu, S. Mukherjee, and G. Bianconi, "Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles," *Phys. Rev. E*, vol. 89 (1), 2014.

[22] G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondragón, and G. Bianconi, "Weighted multiplex networks," *PloS one*, vol. 9 (2), 2014.

[23] P. Wang, G. Robins, P. Pattison, and E. Lazega, "Exponential random graph models for multilevel networks," *Soc. Networks*, vol. 35 (1), pp. 96–115, 2013.

[24] T. Valles-Catala, F. A. Massucci, R. Guimera, and M. Sales-Pardo, "Multilayer stochastic block models reveal the multilayer structure of complex networks," *Phys. Rev. X*, vol. 6 (1), 2016.

[25] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, "Clustering network layers with the strata multilayer stochastic block model," *IEEE Trans. on Net. Sci. and Eng.*, vol. 3 (2), pp. 95–105, 2016.

[26] R. Guimera, S. Mossa, A. Turtschi, and L. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proc. of the Nat. Acad of Sci.*, vol. 102 (22), pp. 7794–7799, 2005.

[27] W.-B. Du, X.-L. Zhou., O. Lordan, Z. Wang, C. Zhao, , and Y.-B. Zhu., "Analysis of the chinese airline network as multi-layer networks," *Transportation Research Part E*, vol. 89, pp. 108–116, 2016.

[28] W.-B. Du, M.-Y. Zhang, Y. Zhang, X.-B. Cao, and J. Zhang, "Delay causality network in air transport systems," *Transportation Research Part E*, vol. 118, pp. 466–476, 2018.

[29] R. Gallotti and M. Barthelemy, "The multilayer temporal network of public transport in great britain," *Sci. Data*, vol. 2, p. 140056, 2015.

[30] D. S. Wilkinson, S.M. and S. Ma, "The vulnerability of the european air traffic network to spatial hazards," *Natural Hazards*, vol. 60 (3), pp. 1027–1036, 2012.

[31] R. Guimer and L. Amaral, "Modeling the world-wide airport network," *The Europ. Phys. J. B*, vol. 38 (2), pp. 381–385, 2004.

[32] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems," *Phys. Rev. X*, vol. 5 (1), 2015.

[33] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, and S. Boccaletti, "Emergence of network features from multiplexity," *Sc. Rep.*, vol. 3, p. 1344, 2013.

[34] P. Basu, R. Sundaram, and M. Dippel, "Multiplex networks: A generative model and algorithmic complexity," in *Proc. IEEE/ACM Int. Conf. on Advances in Soc. Networks Analysis and Mining*, 2015, pp. 456–463.

[35] M. Fügenschuh, R. Gera, and T. Lory, "A synthetic model for multilevel air transportation network," in *Proc. Conf. on OR*, 2017, pp. 347–353.

[36] G. Assenova, "Topological evolution of networks: Case studies in the us airlines and language wikipedias," *Doctor thesis*, 5 2011.

[37] J. D. W. W. L Li, D. Alderson, "Towards a theory of scale-free graphs: Definition, properties, and implications," *Internet Math.*, vol. 2 (4), pp. 431–523, 2005.


**Marzena Fügenschuh** received the PhD degree in mathematics from the Technical University of Darmstadt, Germany, in 2007. She is an associate professor of Beuth University of Applied Sciences, Berlin, Germany. Her research interests include the areas of discrete and combinatorial optimization, especially random heuristics, and recently randomly generated complex networks.

**Ralucca Gera** is the Associate Provost for Graduate Education and Professor of mathematics at the Naval Postgraduate School. She is a researcher in the Center for Cyber Warfare at the Naval Postgraduate School, and in the Network Science Center at United States Military Academy. Her research interests are in graph theory and network science, with applications to the study of the Internet, Cybernetworks and Natural Language Processing.

**Andrea Tagarelli** received the PhD degree in computer and systems engineering from the University of Calabria, Italy, in 2006. He is an associate professor of computer engineering at the University of Calabria, Italy. His research interests include the areas of data mining, machine learning, network analysis, web and semistructured data management, and information retrieval.