Unveiling the Features of Network Dynamics: a Data-Oriented Approach

Shibo Lu, Daphne Tuncer
Imperial College London (UK)

Stuart Clayman
University College London (UK)

Abstract—The ability to predict network dynamics is a challenging problem for communication providers as this can have an impact on the performance of the applications running on top of their infrastructure. In this paper we use a data-driven approach to investigate whether observing the variation of the connectivity properties of a network over time can provide insights onto the occurrence of network topological changes. In that direction, we propose a methodology to model various patterns of network dynamics and develop a user-configurable software tool enabling the collection of network connectivity metric data for any type of communication network topology and based on a wide range of dynamics patterns. We use our tool to constitute a dataset of network connectivity metric samples based on 126 real network topologies and 15 different types of dynamics patterns. The analysis performed based on the obtained dataset shows that identifying how connectivity metrics vary has potential for characterizing network topological changes.

I. INTRODUCTION

An important operational problem for any communication or data center network provider is to have in place efficient mechanisms for predicting the performance of the applications running on top of their infrastructure. This is a crucial aspect given that performance indicators directly translate to the quality of experience for the services offered to customers, *i.e.,* how satisfy the customer is with his/her experience of a service, who are likely to unsubscribe in case this does not meet their expectation.

Application and by extension service performance can be affected by various factors, either internal (*e.g.,* device misconfiguration) or external (*e.g.,* unexpected traffic load) to the network's owner. A specific aspect linked to performance concerns the dynamics of the network topology over time. While the topology of a communication network has been regarded for a long time as being relatively static, the emergence of scenarios in which the connectivity structure of the network changes rapidly and autonomously has been challenging this vision [1][2]. Illustrative examples include virtualized infrastructures where new nodes representing virtual machines are added online to the network topology as resources need to be scaled in order to sustain emerging workload [3], or mobile networks where links between nodes change dynamically as mobile users move. Understanding these network dynamics is essential as topological changes can directly impact the performance of an application. The convergence time of the training process of a large-scale machine learning application using distributed stochastic gradient descent to compute the

model's weights can for instance be significantly affected by erratic node behavior and random communication delays [4].

In this paper we propose a data-driven approach to investigate how the connectivity properties of a network are affected by the dynamics of its topology over time. We define these connectivity properties based on a set of metrics traditionally employed in the network graph literature to characterize network topologies. Our objective is to determine whether the observation of the variation of these metrics can be used to identify and hence predict the occurrence of various types of topological changes. Understanding topology dynamics for performance prediction has received a lot of attraction in the mobile and sensor networks domain, *e.g.,* [5][6][7], as well as in the context of self-managed and self-organized networks and services, *e.g.,* [9][8]. Building on top of previous efforts investigating network dynamics, stability and survivability, we develop a methodology to model different patterns of network dynamics, taking into account the different types of topological changes that can alter a network and the frequency at which these changes can manifest. We use the proposed methodology to implement an user-configurable software tool for network dynamics emulation that allows us to collect metric data for a wide range of real communication network topologies. Based on our tool, we constitute a dataset of 71,605 metric-per-topology samples for a total of 126 networks and 15 types of dynamics patterns. The analysis performed on the obtained dataset shows that monitoring the set of metrics that are more likely to be highly influenced by topological changes, as well as the correlation factors between these metrics, can provide insights onto the types of changes a topology undergoes.

The rest of the paper is organized as follows. Sect. II presents the list of graph metrics used to characterize network topologies. Sect. III describes the methodology to emulate network dynamics patterns, as well as the software tool developed to collect network metrics data. Sect. IV analyzes correlation factors between topological metric changes and network dynamics. Finally, Sect. V provides some conclusions and future work directions.

II. NETWORK TOPOLOGY METRICS

The characteristics of a communication network topology can be described based on the properties of the graph representing the connectivity of the network, *i.e.,* nodes and edges. While nodes model various elements of the network, from physical equipment (*e.g.,* router, server) to virtual functions,

edges represent connection between these elements, *e.g.,* at the physical layer or IP-layer links. In this section, we present the list of graph metrics that we use in this work to characterize network topologies. The metrics were chosen so as to account for various aspects of a communication network infrastructure.

A. Basic metrics and node degree distribution properties

We use the number of nodes and the number of edges as the most basic metrics to characterize a network topology [11].

In addition, we use the properties of the distribution of node degrees, *i.e.,* the distribution of the number of nodes with a given number of direct neighbours, to characterize the importance of each node in the network. In particular, we consider the three following metrics:

- Average node degree: the closer to $n - 1$ the average node degree, with n the total number of nodes, the more connected the network;
- Heterogeneity: variance of node degrees;
- Gini coefficient [12]: used to measure the inequality of node degrees distribution. Unlike the variance, the Gini coefficient is computed using the mean absolute difference between all degree pairs instead of the squared difference deviating from the mean value.

B. Centrality scores

In addition to node degrees, we use different measures of node centrality to characterize the "importance" of a node in the network.

- Closeness centrality: normalized average distance of a node to any other network node[1]. The higher the score, the closer to other nodes [13];
- Betweeness centrality: number of shortest paths passing through a node. The higher the score, the more shortest paths the node lie on [14];
- Katz centrality: as opposed to the node degree that accounts for immediately connected neighbours, the Katz centrality evaluates the connection of a node to all other network nodes using a penalty term [15];
- Eigenvector centrality: Eigenvalues of the adjacency matrix representing the graph of the network [16];
- PageRank centrality: variation of the Katz and Eigenvector centrality [17].

In this paper, we focus more specifically on the influence of the most important nodes and hence use as primary metrics the highest value of each centrality score.

C. Clustering properties

Another set of metrics that we consider in this work concerns the clustering properties of networks, *i.e.,* the tendency for edges to cluster [18]. More specifically, we use the following three main measures[2]:

- Clustering coefficients: probability that a length-2 path is closed (formation of a triangle in the network) [18]. In this work, we use the mean and variance of the distribution of clustering coefficients.
- Number of clusters;
- Max. cluster size: number of nodes in the largest cluster;

D. Distance and resilience properties

We also consider the properties of a network in terms of scale and resilience to events affecting connectivity. We use the diameter, *i.e.,* the longest shortest path between any two nodes, and the average shortest path length, *i.e.,* the mean value of the lengths of all shortest paths in the network, to characterize the scale of the network in terms of distance.

In addition, we use the symmetry ratio [19] and the global efficiency [21] to quantify resilience properties. While Dekker and Colbert show in [19] that the symmetry ratio is highly related to the performance of a network under attack, Latora and Marchiori suggest in [21] to measure the resistance of a network to failures on a small scale using its global efficiency value. We also take into account the algebraic connectivity, *i.e.,* the value of the second largest eigenvalue of the Laplacian matrix defined for the graph representing the network [22].

III. NETWORK TOPOLOGY DYNAMICS

Our objective is to investigate how the topological characteristics of a network is affected by the dynamics of its connectivity over time. In this section, we present the methodology that we designed to emulate various patterns of network dynamics, as well as the software tool that we implemented based on the proposed methodology in order to collect topological metrics data based on a wide range of networks.

A. Methodology

The dynamics of a network is driven by the set of changes affecting its connectivity properties over time. To emulate various patterns of network dynamics, we take into account four aspects: *(i)* the type of events affecting the connectivity of the network; *(ii)* the time at which events occur; *(iii)* the combination of events occurring at a given time; and *(iv)* the sequence of events over time.

Connectivity changes Different events can affect the connectivity properties of a network. In this work, we define four types of events: *(i)* the addition of a node and an edge in the network (ANE)[3]; *(ii)* the addition of an edge (AE); *(iii)* the removal of a node from the network (RN); and *(iv)* the removal of an edge (RE). In practice, these events account for various changes that can alter the topology of a network, *e.g.,* the addition of a new VM in a virtualized infrastructure, the failure of a physical link between two routers, the creation of a new virtual path between two virtual network functions. For each type of events, we define different strategies to decide how the topology needs to be modified (*e.g.,* which node to select for removal). These are presented in Table I.

---

[1]We use Dijkstra shortest path to compute the path between any pair of node and hence determine the distance.

[2]In this work, we use the greedy modularity communities partitioning method proposed by Clauset *et al.* in [20] to compute the clusters.

[3]We consider that all network topologies are connected so that the addition of a node implies the addition of an edge.

TABLE I
Types of network connectivity changes.

Event ID	Event Type	Modification Strategy
RN1	Remove Node	Random
RN2	Remove Node	With Highest degree
RN3	Remove Node	With Lowest degree
RN4	Remove Node	Based on network SIR epidemic model [23]
RN5	Remove Node	Based on Breadth First Search model[24]
RE1	Remove Edge	Random
RE2	Remove Edge	Connecting highest degree nodes
RE3	Remove Edge	Connecting lowest degrees nodes
RE4	Remove Edge	Between lowest and highest degrees nodes
AE1	Add Edge	Random
AE2	Add Edge	Between highest centrality nodes
AE3	Add Edge	Between lowest centrality nodes
ANE1	Add Node & Edge	Random
ANE2	Add Node & Edge	Connect to highest centrality node
ANE3	Add Node & Edge	Connect to lowest centrality node

Events occurrence Events can occur at different moments in time. We model time as a discrete variable t and assume that events are independent[4], *i.e.,* the type of events occurring at time t is independent from the type of events occurring at time $t+1$. In this work, we use two different approaches to model the occurrence of events: *(i)* as a periodic process, *i.e.,* events affecting the connectivity of the network are triggered at every $t+\delta$ where δ represents the periodicity of the events; and *(ii)* as a random process, *i.e.,* events are triggered at random time.

Events combination In this paper, we assume that only one type of events can occur at any time t. The effect of combined events can however easily be modelled by setting the value of the frequency of event occurrence to 0.

Sequence of events We use four different models to emulate various sequences of events over time: *(i)* repetitive mode, *i.e.,* the same event is repeatedly triggered; *(ii)* mixed mode, *(i.e.,)* a random event, selected between all types reported in Table I, is triggered; *(iii)* SIR epidemic mode, *i.e.,* events are triggered according to the SIR epidemic model [25]; and *(iv)* the Breadth First Search mode, *i.e.,* events are triggered according to the Breadth First Search model [24]. While the repetitive mode enables us to investigate the effect of specific types of events, we use the mixed mode to emulate the effect of a random succession of events on the topological properties of the network. In addition, we use the SIR epidemic model and the Breadth First Search model as these constitute reference approaches for the analysis of dynamics in the complex network literature.

B. Software tool and dataset

We implemented the proposed methodology in a software tool for network dynamics pattern emulation that we developed in Python. The functionality of the tool is depicted in Fig. 1. It involves two main functions: *dynamics pattern generator* (DPG) and *topological metric computation* (TMC). The DPG

---

[4]The tool can easily be extended to integrate any type of event dependency modelling component
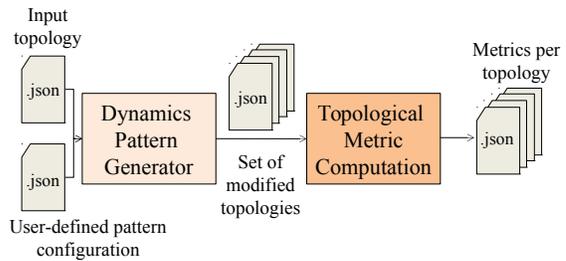
Fig. 1. Functionality of the network dynamics pattern emulation tool.

function is responsible for computing, based on an input network topology and a set of user-defined pattern configurations, a sequence of events affecting the connectivity of the network. Changes are applied sequentially and all resulting modified topologies are returned in output. These are further passed to the TMC function responsible for computing for each topology the values of all metrics described in Section II. In our tool, all configurations are encoded using the .json format (*i.e.,* network topologies, dynamics pattern parameters and metrics).

To obtain a dataset of topological metrics, we use a set of 126 networks extracted from the Internet Topology Zoo repository [10] as inputs to our tool. The selected networks vary in size (from 6 to 200 nodes), structure, as well as types of communication infrastructures they represent (backbone, customer and/or transit[5]). For each network, we applied 15 different types of dynamics patterns based on the models described in Section III-A. This gave us a total of 71,605 metric-per-topology samples. The obtained dataset is available at: https://www.dropbox.com/s/qvc178mawbz78s5/ConnectivityMetricDataset.zip?dl=0

In the next section, we used the obtained dataset to analyze the correlation factors between topological metric changes and network dynamics.

IV. Results Analysis

Correlation factors can be investigated from different perspectives, *i.e.,* across topologies, across patterns of network dynamics and across metrics. In this paper, we focus on how events affect specific types of metrics and to which extent pairs of metrics tend be correlated under specific patterns.

Most influenced metrics To quantify the effect of a network dynamics pattern on a given topological metric, we determine the normalized changes of the value of the metric across the sequence of events defining that pattern. We further apply linear least squares fitting to the resulting profile of changes and evaluate the absolute value of the coefficient of regression of the obtained fitting curve.

We define *the most influenced metrics* (MIMs) for a network topology as the top five metrics with the largest absolute coefficient of regression value across all dynamics patterns. We observe that, for some patterns, the ranking of the MIMs is not influenced by the network topology, *i.e.,* part or all of the metrics in the top-five list are shared by the majority of the

---

[5]The terminology refers to the one used in [10].

TABLE II
SHARED MOST INFLUENCED METRICS.

Event ID	Metrics
RN2	Number of nodes
RN3	Number of nodes & number of edges
ANE1	Heterogeneity
ANE2	Maximum cluster size
ANE3	Average shortest path & number of nodes

TABLE III
PERCENTAGE OF THE HIGHLY CORRELATED METRIC PAIRS IN EACH
RECURRENCE CATEGORY.

Event ID	Strongly recurrent	Recurrent	Not strongly recurrent
RN1	0.52%	2.63%	96.84%
RN2	1.58%	3.16%	95.26%
RN3	0.52%	2.11%	97.37%
RN4	0%	1.05%	98.95%
RN5	0%	1.05%	98.95%
RE1	1.58%	1.05%	97.37%
RE2	0.52%	1.05%	98.42%
RE3	2.63%	2.11%	95.26%
AE1	0.52%	3.68%	95.79%
AE2	1.05%	6.84%	92.11%
AE3	8.64%	5.79%	87.37%
ANE1	2.11%	10.53%	87.37%
ANE2	8.42%	9.47%	82.11%
ANE3	5.79%	6.32%	87.89%

input networks. Table II shows the cases in which more than 80% of the networks have shared MIMs. As can be observed, the type of MIMs that are common between networks is affected by the type of topological changes.

Highly correlated metric pairs In addition, we analyze for each network the correlation factor between pairs of metrics as the network topology undergoes a sequence of changes. The total number of pairs of metrics is 190. We use the Pearson correlation coefficient to determine the degree of correlation between any two metrics m_1 and m_2 under a series of n changes. It is expressed as follows:

$$\rho_{m_1,m_2} = \frac{\sum_{i=1}^{n}(m_{1i}-\overline{m}_1)(m_{2i}-\overline{m}_2)}{\sqrt{\sum_{i=1}^{n}(m_{1i}-\overline{m}_1)^2 \sum_{i=1}^{n}(m_{2i}-\overline{m}_2)^2}}$$

We define as *the highly correlated metric pairs* (HCMPs) the pairs of metrics for which the correlation coefficient is larger than 0.95. To investigate whether network topologies have commonalities in terms of their HCMPs, we classify metric pairs in three categories: *1)* strongly recurrent HCMPs, *2)* recurrent HCMPs, and *3)* not strongly recurrent HCMPs. Category 1 corresponds to the case where more than 80% of the input networks share the same HCMPs. Category 2 concerns the case where the same HCMPs are shared by 50% to 79% of the networks, and category 3, when this happens for less than 50% of the networks. Table III shows the percentage of HCMPs falling under each category per event type.

In general, only subsets of HCMPs are shared by a majority of networks, *i.e.,* fall in category 1. The size of the shared subsets also depends on the type of topological changes. For instance, no strongly recurrent HCMPs are recorded when networks depict dynamics patterns based on event types RN4 and RN5 (*i.e.,* SIR Epidemic and Breadth First search mode, respectively) while more than 8% of HCMPs are recurrent to the majority of networks when the dynamics follow a repetitive mode and are driven by AE3 or ANE2 type of topological changes. In terms of metrics, the global efficiency, the maximum of the centrality scores, the mean and variance of the distribution of clustering coefficients, and the mean and variance of the distribution of node degrees can be noted to be the main constituents of highly recurrent HCMPs.

MIMs and HCMPs correlation The observations obtained for both the MIMs and HCMPs show that results differ based on the patterns of network dynamics (type of events and

mode). This suggests that the analysis of both components can provide some insights onto the topological changes undergone by a network. Given that the MIMs and HCMPs both "measure" the effect of changes on the network properties, we investigate whether a relationship exists between them. In particular, we determine for each network, the percentage of MIMs that are part of HCMPs.

Among all 8,190 most influenced metrics calculated for all 126 input networks and 15 dynamics patterns, 4,787 (58.45%) of them appear in the highly correlated metric pairs. This shows that, in general, the MIMs are more likely to be part of HCMPs. However, a more quantitative hypothesis test is required to mathematically determine the relationship between them. This is a direction of future research.

V. CONCLUSIONS

This paper describes a methodology to model patterns of network dynamics and presents an emulation software tool to collect network connectivity metric data for any type of communication network topologies. The analysis of the dataset collected using our tool for 126 networks and 15 types of dynamics patterns shows that observing the mostly influenced metrics and the highly correlated metric pairs can provide insights onto the types of changes a topology undergoes. Next step will investigate how to apply Convolutional Neural Network on network adjacency matrices to predict the mostly influenced metrics and the highly correlated metric pairs of any network topology. More generally, our long term objective is to investigate the dynamics and stability characteristics of a networking environment within and across dimensions (from the infrastructure to management tasks and service operations) and develop novel frameworks for the prediction of network and service performance.

REFERENCES

[1] X. Chen, Y. Mao, Z. M. Mao, and J. Van der Merwe, "Declarative configuration management for complex and dynamic networks," in *proc. of the 6th ACM International COnference on emerging Networking EXperiments and Technologies (CoNEXT)*, 2010.

[2] S. Clayman, R.G. Clegg, A. Galis, A. Manzalini, "Stability in Dynamic Networks," in *proc. of Future Network and Mobile Summit (FUNEMS)*, pp. 4-6, July 2012.

[3] E. Hernandez-Valencia, and B. Polonsky, "How will NFV/SDN transform service provider opex?," *IEEE Network*, vol. 29(3), pp. 60–67, 2015.

[4] J. Wang, J., and G. Joshi, "Adaptive communication strategies to achieve the best error-runtime trade-off in local-update SGD," arXiv preprint arXiv:1810.08313, 2018.

[5] J. Monteiro, A. Goldman and A. Ferreira, "Performance Evaluation of Dynamic Networks using an Evolving Graph Combinatorial Model," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, pp. 173-180, 2006.

[6] D. Fischer, D. Basin and T. Engel, "Topology dynamics and routing for predictable mobile networks," in *IEEE International Conference on Network Protocols*, pp. 207-217, 2008.

[7] Q. Guan, F. R. Yu, S. Jiang and G. Wei, "Prediction-Based Topology Control and Routing in Cognitive Radio Mobile Ad Hoc Networks," in *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4443-4452, Nov. 2010.

[8] L. Bennacer, Y. Amirat, A. Chibani, A. Mellouk and L. Ciavaglia, "Self-Diagnosis Technique for Virtual Private Networks Combining Bayesian Networks and Case-Based Reasoning," in *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 354-366, Jan. 2015.

[9] Yu Dai, Lei Yang, and Bin Zhang, "QoS-driven self-healing web service composition based on performance prediction," *Journal of Computer Science and Technology*, vol. 24(2), pp. 250-261, 2009.

[10] S. Knight, et al., "The Internet Topology Zoo," *IEEE Journal on Selected Areas in Communications*, vol. 29(9), pp. 1765-1775, Sept 2011.

[11] D. K. Arrowsmith, R. J. Mondrag, and M. Woolf, "Data traffic, topology and congestion," *in Complex Dynamics in Communication Networks*, pp. 127-157, 2005.

[12] S. Goswami, C. A. Murthy, and Asit K. Das, "Sparsity measure of a network graph: Gini index," *Information Sciences*, vol. 462, pp. 16-39, 2018.

[13] G. Sabidussi, "The Centrality Index of a Graph," *Psychometrika* vol. 31(4), pp 581-603, 1996.

[14] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40(1), pp. 35-41, 1977.

[15] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18(1), pp 39–43, 1953.

[16] M. E. J. Newman, "The mathematics of networks," *The New Palgrave Dictionary of Economics*, pp 1–8, 2016

[17] S. Herrmann, G. Ochoa, and F. Rothlauf, "PageRank centrality for performance prediction: the impact of the local optima network model," *Journal of Heuristics*, vol. 24(3), pp. 243-264, 2018.

[18] H. Yin, A. R. Benson, and J. Leskovec, "Higher-order clustering in networks," arXiv:1704.03913, 2017.

[19] A. H. Dekker, and B. Colbert, "The symmetric ratio of a network," in *proc. of the 2005 Australasian symposium on Theory of computing (CATS'05)*, vol. 41, pp. 13-20, 2005.

[20] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70(6), 2004.

[21] V. Latora, and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87(19), 2001.

[22] W. N. Anderson Jr, and T. D. Morley, "Eigenvalues of the Laplacian of a graph," *Linear and multilinear algebra*, vol. 18(2), pp. 141-145, 1985.

[23] A.-L. Barabási. "Network Science", section 10, pp. 372-434

[24] E. F. Moore, "The shortest path through a maze," in *proc. of the International Symposium on the Theory of Switching*, Harvard University Press, pp. 285–292, 1959.

[25] Z. Tao, F. Zhongqian, and W. Binghong, "Epidemic dynamics on complex networks," *Progress in Natural Science*, vol. 16(5), pp. 452-457, 2006.