Final project report to the NKFIH Grant No. 123957, “Measurement-based analysis of operational paths in complex networks” (hungarian title: “A természet által használt útvonalak tulajdonságainak vizsgálata valós hálózatokban”) under the FK-17 funding scheme

Research objectives of the project

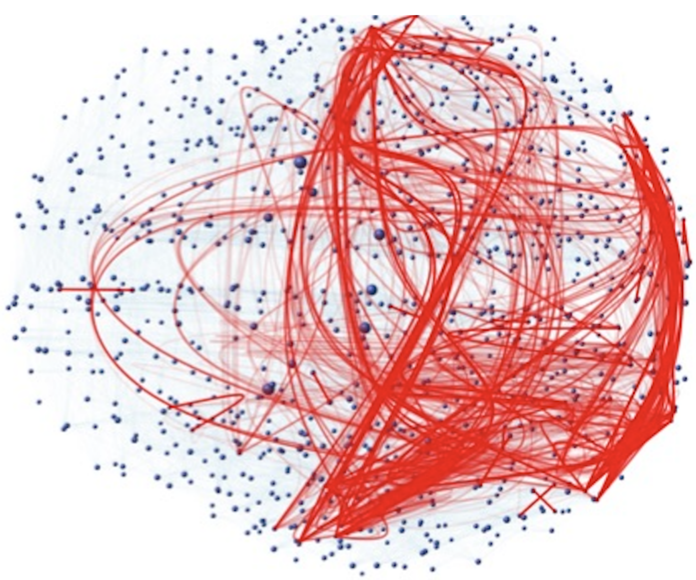
Due to the results of network science, the fact that real networks are similar to each other in many topological characteristics is well known today. Most of them (including many biological, social, transportation, and organizational networks) are small worlds, have scale-free degree distribution and large clustering confirmed by thousands of real measurements. We argue that the system of operational paths used over these networks is at least that important as topology. The reaction of cells to stress, the stability of airport networks, or the behavior of the Internet in case of massive workloads are all fundamentally determined by the operational paths the signals/passengers/packets go through. Despite its clear importance, the volume of measurement-based studies targeting the characterization of operational paths is quite small. NKFIH project No. 123957 was aimed to fill this gap and take a step towards the correct interpretation of paths used in real systems. From the analysis of operational paths in existing and newly created datasets, we planned to infer the path selection mechanism used by nature when picking a path. We also planned to propose synthetic routing algorithms which can recover the basic statistical properties of operational paths in networks.

The main question of our project was whether the operational paths used in real networks resemble each other similarly as their topology does. Our main hypothesis was that there are common rules of paths selection that nature uses when picking a path despite the fact that entities that the various real network builds upon are completely different. Social networks are built among humans, the Internet consists of Autonomous Systems, protein networks contain proteins, but the common ground of these networks is self-organization. If self-organization has driven these networks in such a way that their topology developed extremely similarly, then there is hope that the operational paths over these networks are pretty similar. We planned to first address the most basic statistic, i.e. the length of the paths. Many papers in the literature (sometimes implicitly) simply assume that operational paths are the ones with the shortest length. We had good reasons to believe that nature does not always use the shortest path. For example, it is well-known among Internet practitioners that on the Internet a non-negligible portion of the paths is longer than their counterparts. We expected that we will find something similar in other real networks. Secondly, we planed to characterize the rules of path selection used in nature based on ground truth measurements. These rules may enable us to predict the behavior of networks in the presence of specific workloads.

Research results

I’m really proud of my research team as we managed to fulfill all our plans and objectives to an appropriate extent. In what follows we highlight our main contributions in three major domains: data acquisition, data sharing, and scientific results.

Data acquisition In the first phase of the project we worked out know-hows for extracting operational paths from networks in diverse areas of life. We have reconstructed the Internet AS topology from the measurements of the Center for Applied Internet Data Analysis (CAIDA) and managed to obtain millions of Internet paths suitable for our research. Our query robot written for the Rome to Rio flight ticket booking system managed to collect more than 13K possible flight paths. Using these paths and the OpenFlights database we have also reconstructed an up-to-date map of the world’s flight system. Due to the fruitful cooperation with our colleagues at EPFL and Indiana University Bloomington a contemporary map of the human connectome and around 400K brain paths have been collected into our brain database (see Figure below). Finally, our smartphone word-morph game called “Fit-fat-cat” collected more than 30K human paths over the word maze of the three-letter English words. These paths are footprints of how people solve navigation tasks in a complex word network. All of our datasets have been transformed into a common format and prepared for comparative path analysis. In the last year of the project, we managed to get access to a much larger dataset which is slightly similar to our fit-fat-cat data. This dataset collects human paths over Wikipedia using the smartphone game called WikiGame. The processing of this dataset has not been completed within this project but will be continued in the future.



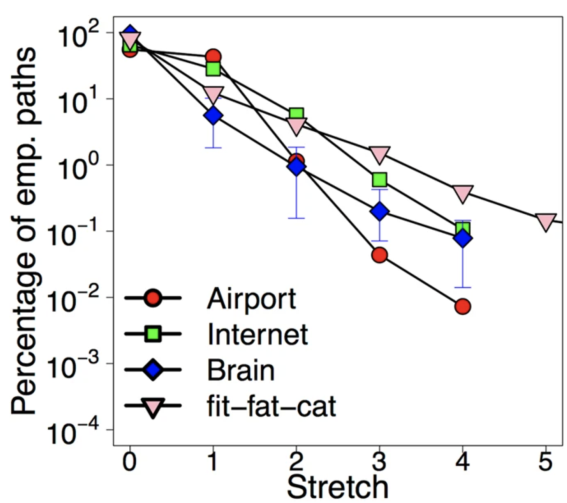
Our approximation of operational paths in the human brain.

Data sharing in Open Access

Our dataset recording human navigational paths in the network of three-letter English words have been accepted for open access publication in Nature Scientific Data (Kőrösi, A. et al. A dataset on human navigation strategies in foreign networked systems. Sci. Data 5:180037 doi: 10.1038/sdata.2018.37 (2018) IF: 5.862). The data is publicly available from its data repository hosted at OSF (https://osf.io/jtyvd/). The dataset has already been used by other researchers. Mareike Bockholt and her team used our open data for a systematic analysis of network centrality measures, which are clearly affected by the structure of operational paths.

Scientific results

Using our datasets we found an interesting result that the operational paths in various networks are significantly stretched compared to their shortest counterparts. We have found that path stretch follows a very similar distribution (see figure below) regardless of the system producing the paths. Based on this finding we suspected that path in real systems does not necessarily coincide with the shortest possible path in the network. There is an average of 10-30 percent of stretch in all our datasets.



Stretch distribution of operational paths in our networks

The deeper analysis of the paths in our datasets revealed the following high-level picture about operational paths. First, we have seen that operational paths can be slightly longer than the shortest paths. Although in some particular cases, the stretch can be more than four steps, the operational paths are only around 10-30% longer on average. Thus our first finding is that nature prefers the usage of short but not necessarily shortest paths. Secondly, we have seen that real networks seem to have an internal logic or an internal hierarchy, which the operational paths follow in the majority of cases. This simply means that operational paths go first upwards in the hierarchy and then downwards and paths cannot contain down-up jumps. Finally, paths avoid stepping upwards in the hierarchy if possible, meaning that if there are both downstream and upstream paths available, then the downstream one will be picked with high probability.

Thus our findings hint at the operation of the “prefer short paths”, “prefer regular paths” (i.e. follow the hierarchy) and “prefer downstream” path selection rules. We showed that these rules are not equally important and there are clear relative priorities among the identified rules. We have identified a reasonable prioritization among those components, which allows us to set up a synthetic path selection ruleset imitating nature’s path picking process. This ruleset serves as the base of our developed synthetic routing algorithm. The ``prefer shortest path” rule can only have lower priority than the “prefer regular path” and the “prefer downstream” otherwise we would not experience stretch at all, which would contradict our measurements. Since “prefer downstream” implies the “prefer regular path” rule, the only reasonable choice is to: prefer regular paths at first, then prefer downstream if there is a downstream path and from the remaining paths, prefer the short paths. Our interesting finding is that the length of the paths is just the third thing on this checklist.

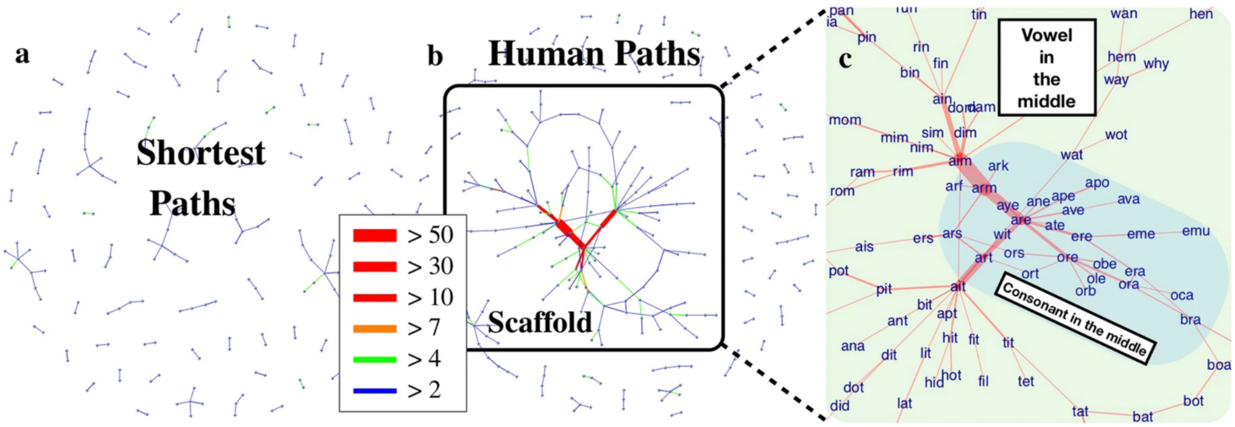
We have checked via computer simulations, how close our argument about path selection above comes to real operational paths. To do this, we defined our synthetic path selection ruleset precisely and compare the selected paths to the shortest and operational ones. We defined our synthetic path selection ruleset to:

- Rule 1 Use regular paths only
- Rule 2 Pick the downstream paths if available
- Rule 3 From the paths remaining after Rule 1 and Rule 2, pick the shortest ones
- Rule 4 If there are still multiple paths remaining break ties randomly

It turned out that the above simple path selection rule set gives a very realistic path stretch, close to the stretch computed for the real empirical paths. However, since we explicitly prohibit the use of non-regular paths, all of them are regular, unlike real operational paths. So, what our simple path selection method cannot reproduce is that operational paths sometimes violate the “prefer regular path” and the “prefer downstream” rules, although in only a minority of cases. However, the slight randomization of the centrality values of the nodes fixes this. Why would someone randomize the centrality values of the nodes? In the case of large and dynamic networks (like the Internet or a social network), we cannot even reconstruct an up-to-date map of the network. Therefore, it doesn’t sound reasonable to suppose that any real person or entity representing a node in the network would know the exact structure, i.e., the correct hierarchy of such networks. So, the randomization can be interpreted as simulating the case in which nodes have an approximate picture of the network and therefore their position in the hierarchy is known only with some random error. In that case, the nodes can only have an approximate picture of the internal logic or the hierarchy of the network. This small modification recovers both the stretch and the level of regularity exhibited by the operational paths.

Our findings are suitable for e.g. gaining an estimation of the traffic situation in a large city during busy hours. Monday mornings are always great stress for the road network and the public transportation system, as everybody goes to work roughly at the same time. On the one hand, due to stretch, empirical paths impose a larger average load on the network nodes, on the other, the load is even more concentrated in the inner parts of the network hierarchy (i.e., on nodes located more centrally in the network). The above-presented path selection ruleset seems to explain the load footprint of real paths better, than resting simply on the assumption that people always using the shortest paths. In the context of the public transportation example, the synthetic policy allows us to better estimate the mass of people at various stations and the possible length of lines at the ticket offices compared to the approximations based on pure shortest paths. Our findings regarding the hierarchical structure of paths have been published in Nature Scientific Reports (Csoma A, Korösi A, Rétvári G, Heszberger Z, Bíró J, Slíz M, Avena-Koenigsberger A, Griffa A, Hagmann P, Gulyás A: Routes Obey Hierarchy in Complex Networks, SCI REP 7: (1) Paper 7412., 2017, IF: 4.609).

For taking a closer look at the structure of the hierarchy that the operational paths follow, we have analyzed the fit-fat-cat dataset more carefully. This dataset contains the results of navigation tasks performed by human subjects in the network of three-letter English words. We have recorded on average 40.9 timely-ordered paths from 259 subjects and more than 200 paths from 9 subjects which makes the analysis of individual human navigation patterns possible. Our study (published in Nature Scientific Reports: Gulyás, A., Bíró, J., Rétvári, G. et al. The role of detours in individual human navigation patterns of complex networks. Sci Rep 10, 1098 (2020), and in a short form at the Dynamics of the brain conference organized by FENS) shows that human subjects frequently apply detours, even in the long run. Our main finding is that these detours are the consequences of how an individual interprets a complex networked system on its own level. We showed that people tend to build up a significantly simpler representation of the word-morph network in the form of a hierarchy in their minds. These hierarchies are then used as scaffolds to the paths in the network (see figure below). As a result, real paths will be somewhat longer than the shortest alternatives, but the detours will be characteristic to the individual taking them, as no two may abstract away the same hierarchy of the network. Although our study concentrates on a networked system, the underlying problem of human navigation in the word-morph network seems even more interesting in the light that current explanations of physical navigation tend to apply models considering the graph-like abstraction of the surrounding physical environment. In fact, there is an ongoing debate about if we build a detailed cognitive map or a much simpler cognitive graph of the possible physical choice points inside our heads. Furthermore, recent studies reported major correlations between navigation and the learning skills of humans while others go even further and investigate the possibility that navigation in cognitive spaces may lie in the core of any form of organized knowledge and thinking. The word-morph network is a special mixed system over which navigation relies strongly on domain-general mechanisms since both spatial, manifested in the Hamming distance between words, and cognitive dimensions, i.e. the function and the meaning of the words, contribute to the formation paths. Thus, the identification of individual scaffolds guiding human navigation in the word-morph network may contribute to a better understanding of how humans structure, encode, and navigate through cognitive spaces.



Comparison of shortest and human paths in the word network

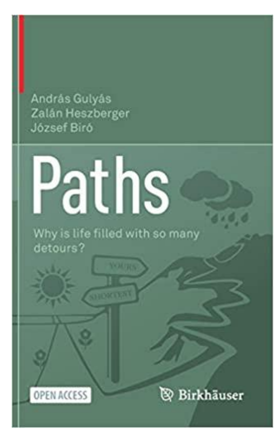
Our preliminary analysis of the newly accessed WikiGame dataset confirms the presence of scaffolds, that human subjects rely on when navigating, in much larger networks too.

My team has also achieved results regarding the robust navigational core in the human brain (published in the highly-rated conference of Complex Networks 2018 in Cambridge). Our extended abstract and poster there presented results about a minimal navigational structure (a core) that is necessary to support maximal navigation between arbitrary regions of the human brain. We have shown that this core is not just there, but it seems to be highly robust as being identified regardless of the resolution of the MRI data used for capturing the structural map of the human brain.

Education

The methodologies and results of the project are presented to the students at the Faculty of Electrical Engineering and Informatics of Budapest University of Technology and Economics in various forms. The complex network aspect of the Internet is covered by lectures in multiple classes (Modeling Seminar for Engineers, Building and Operation of Networks, Internet Ecosystem, and its Evolution), while the students can dig out the properties of the human paths themselves in a laboratory experiment (Infocommunications laboratory).

We have also assembled a popular science book aimed at popularizing our research results. We think that the way humans find paths in a complex network of choices has significant consequences for how we think and how we live our lives. In this respect, there may be an interest in our findings not just from the networking research community but from a wider audience too. Our book is published at Springer-Nature-Birkhäuser (Gulyás A., Heszberger Z., Biró J. (2021) Paths: Why is life filled with so many detours?). This book is intended for readers with no background in



network science. They are introduced to our results through several easy-to-interpret stories. Besides the popularization of our results, the book also revolves around the possible philosophical outreach of our findings.

Media

Our research work has been covered in several media releases too:
https://phys.org/wire-news/266309356/towards-the-science-of-paths-in-networks.html
https://phys.org/news/2018-03-dataset-human-strategies-foreign-networked.html
https://researchdata.springernature.com/users/82878-andras-gulyas/posts/29920-dataset-about-human-navigation
http://www.bme.hu/node/4910

András Gulyás
Principal Investigator