

Stochastic Count Processes and Hypergraph Databases in the Dynamic Network Analysis of International Visits

Aleksey Domanov¹ and Oleg Semenov²

Abstract

The given study applied dynamic topological network analysis with regime shifts to international visits. Using this quantitative method, we demonstrated, how the degree of connectivity between various countries has been affected by abrupt changes on the international arena — like coronavirus pandemics. High-level officials' visit hypergraphs were compared between the first and the second quarter of 2020. Hypergraph database allowed for modeling multilateral meetings sampled from the GDELT project materials. Transforming them into simplicial and cell complexes allowed for applying a well-established spatial computing method (one of the techniques employed by topological data analysts) to international relations. Topological tools revealed key structural features of pre- and early-Covid networks. Parallel interpersonal channels of communication decreased in number after the pandemic start, and this regime shift was reflected in the number of topological features. A “hole” discovered between France, Turkey and Nigeria suggests emergence of considerable obstacles to communications between their leaders. According to GDELT database, not only did they interrupt direct personal contacts, but also no intermediary head of state personally visited all the three countries, hosted their leaders or met them on a “neutral” territory.

Keywords:

connectivity, Topological Data Analysis, 2-simplex, cell complex, random hypergraph, persistent homology, coronavirus restrictions, clustering, multilateral summits, communication.

¹ Junior Researcher, Sochava Institute of Geography Siberian, Branch Russian Academy of Sciences; Research Fellow, Department of European Integration Research, Institute of Europe, Russian Academy of Sciences domanovaleksey@gmail.com

² Junior Researcher, Sochava Institute of Geography Siberian, Branch Russian Academy of Sciences wetkidereva@gmail.com

Introduction

Fragmenting world politics increases the necessity to provide solid quantitative basis for analyzing the density of interactions between international actors. Being a trigger of major readjustment in global connections, Covid-19 pandemics provided researchers with rich material to understand, what changes in international connectivity resulted from restrictions related to this virus. We suggest approaching this phenomenon using topological network analysis.

Hypothesis 1: global connectivity decreased after the beginning of Covid-19 pandemics.

More specifically, we would like to focus on the subnetwork of connections between 3 following countries.

Hypothesis 2: the connectivity between France, Turkey and Nigeria decreased after the beginning of Covid-19 pandemics.

The following sections describe our approach to network analysis and a sample selected from checked GDELT dataset. Furthermore, we compare two networks by a single global characteristic (representing increased sparsity in international connections around the world after Covid-19 pandemics start). Besides, connections between 3 specific countries and other states are characterized using the same networks with transformed edges.

Stochasticity in visits network dynamics

Connectivity between two states could be reflected in the frequency of friendly actions undertaken by the political elite of one country towards their counterparts abroad - e.g. visits. Two countries could be regarded as connected, if their leaders met (not only on their territories, but also in a neutral location). Therefore via these meetings a network of visits is created, which could have dense connectivity or high sparsity.

Sparse networks could be analyzed within the tradition advanced by Hungarian mathematician Paul Erdős: as a random graph, where edges between vertices appear with a certain probability. In the case of international relations these edges could depict interpersonal ties established or reinforced by politicians – from geographical point of view, they become contiguous in relative space (moreover, they constitute a space in mathematical sense without geodesic coordinates).

Randomness in this graph is not only observed spatially, but also temporally. Politicians' external actions could be generalized as a process including probabilistic events, which contribute to network density. Therefore the neighborhood of every node and the whole graph stochastically evolves – edges are added at variable rate in various periods. If the number of meetings is aggregated by timeslots, it is described as a count process.

This temporal model of contacts between actors theoretically allows for a simultaneous drop in this frequency of negotiations. At the level of the whole network it could result in decreased overall connectivity. This effect could have a long-lasting nature: some inertia from the event, which caused that major reconfiguration, often remains tangible several months or years after this date. Sometimes observers could register that the system transitioned to a new mode of operation – underwent a regime shift. It is worth noticing that not only some values of network nodes could endure changes (e.g., a decrease in the GDP level due to a crisis), but also edge attributes or weights could be altered, and the whole network could be rewired

In practice this dynamics of meeting probability could be witnessed at the start of Covid-19 pandemics, which could accelerate decoupling between key international players. This divergence could arise from inability to eliminate misunderstandings during personal, offline contact (which became more difficult to organize because of health restrictions). Information flow in the network could be hampered, because of disappeared direct connections: although indirect duplicating circuits of interpersonal communication (visits of intermediaries) and low-level meetings could remain, actors could withhold sensitive information for several more months than usual – until it becomes possible to discuss these issues personally in private.

As for quantitative characteristic of the “regime shift” in focus, meetings network could become more sparse. Its density could decrease (following the number of edges) and more messages could be delivered along bridges – intermediaries instead of direct meetings. The pandemics start could have the disconnection effect depicted on Fig. 1: the solid, black circuit between 3 partners could be replaced by indirect communications (dotted contour).

Long-lasting consequences of Covid-19 restrictions make reasons to attempt modeling this dynamic as a “regime shift”: substantial change of system parameters from one period to another. This temporal approach resembles difference in differences research design, widely used for spatial analysis – inside geographic regression discontinuity algorithms.

For that purpose the stochastic processes related to political actors were sampled in order to conduct dynamic comparison between pre-Covid and early-Covid periods. It is worth noticing, that this “shift” has not been sudden (after 11 March 2020, when the World Health Organization declared Covid-19 a pandemic, some countries didn’t introduce full restrictions, and their presidents continued meeting each other). Therefore our network was divided at the end of March 2020 – into two windows by 3 months (equal periods in order to avoid unbalanced sampling) for the first quarter 2020 and the second quarter 2020.

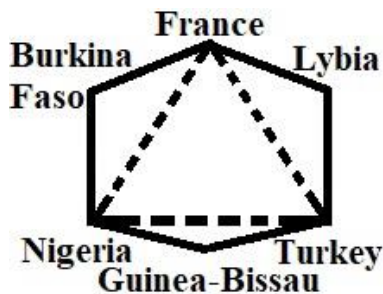


Fig. 1. Personal communication channels in parallel to disappeared (dotted) connections between French, Turkish and Nigerian leaders.

Hypergraph database, simplicial and cell complexes

A dataset, which could be divided into pre- and early-Covid-19 periods, has been collected by filtering and checking information from the Global Database of Events, Language and Tone (GDELT project) with the help of our colleagues I. Ivlev, V. Lazereg and V. Botvinkin. This is a catalog of media materials, which underwent speech recognition and natural language processing procedures [The GDELT Project].

Our database consisted of individual meetings sampled according to some requirements. In particular, we were interested only in contacts between heads of state, which occurred in person (phone calls have also been registered, but not analyzed below).

Information about multiplicity of participants in some summits (multilateral conferences) should have been accurately reflected in our dataset: otherwise we could lose interesting patterns related to cumulative qualities of communities and its members – e.g. 3 leaders’ readiness to adjust their schedules for a trilateral meeting rather than having 3 separate discussions. In order to preserve this multilateral data, it was stored in hypergraph format.

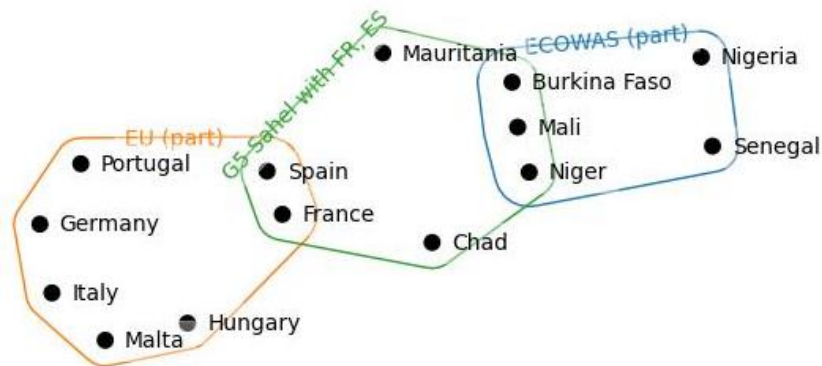


Fig. 2. Fragments of overlapping hyperedges corresponding to summer 2020 multilateral conferences.

Hypergraphs improve upon graph structures in the following way: they allow for modeling connections between more than two vertices (in our case, multilateral conferences including more than two state leaders). As hyperedges are not limited to relations between two vertices (Fig. 2), hypergraphs could demonstrate that, for example, strong connectivity could exist between NATO governments, judging by their decision to conduct a roundtable discussion even if Turkish and Greek presidents would prefer not meeting (bilaterally) some year because of their military tensions.

One could also represent the same multilateral meetings as generalized clique complexes - simplicial and cell complexes. These are topological structures which could have groups of vertices belonging to the same hyperedges – n -ary polyhedra. For instance, 2-simplices for trilateral meetings, other types of cells – for summits with more than 3 participants.

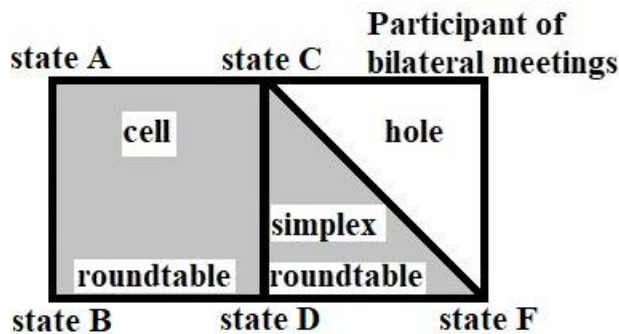


Fig. 3. A network with cell complex.

This simplicial representation is used (Aktas, Akbas, Fatmaoui, 2019) to focus on topological invariants of a network. A topological characteristic could be given to the whole configuration of edges, in particular one could compute the number of holes between the international links. This procedure is based on the following theoretical ground.

The topological data analysis (TDA)

This database was analyzed with an emphasis on its topology. One of the methods suitable for this task is the topological data analysis (TDA). It is reputed for adopting a holistic perspective on samples: this is made possible by approaching the space constituted by individual elements.

The TDA regards space as a set of objects with connections between them, and connections could be defined in abstract terms: any item could be considered close to another according to the researcher's choice (following the "axiom of choice" in set theory), even if they are located 1000 km from each other. These relaxed assumptions about neighborhood allow for clarifying, how individual units (organized in triplets, polyhedra and other simplices) structure this space. The TDA formalizes this global relationship structure, and due to this operationalization a new information about it could be extracted from particular datasets.

It is worth noticing, that the TDA is a multifunctional tool: it is applied to cases with established metric distances, or when contiguity between countries is not expressed in km. Topological perspective presupposes abstracting the network structure – analyzing it regardless of the vector space, in which it is embedded (e.g. constituted by geophysical locations of nodes). The TDA attempts at revealing fundamental properties inherent in a dataset's shape, looks for phenomena, which are present at different measurement scales.

The existence of these fundamental properties provides a solid methodological ground for comparison between different spaces. For instance, in the realm of world economy researchers could be interested in finding similarities between foreign direct investment network and regional trade blocs: whether the latter tends to structure the former ties. Topology formalizes resemblance between these graphs by describing similar ways, in which nodes are wired together in both cases. The TDA also ensures drawing analogies

between large-scale and small networks, because sets with uneven number of objects are approached in the same manner during the computation process.

Due to its focus on the dataset shape, the TDA is particularly instrumental for dynamic comparisons. Indeed, various configurations of the same network at different time could be inspected for major structural changes resulting from external influences or internal tensions, while disregarding minor distinctions: e.g. related to a change in distances between objects, if their relative locations remained the same – their order has been preserved. Therefore the TDA appears to be useful to study the consequences of the Covid-19 pandemics (regarded as an operator, which reshaped international networks), and properties of a new early- or post-Covid configurations could be compared against the parameters of the same system before this major shift.

Moreover, deformation research motivated topologists to specify the properties in question and focus on phenomena remaining stable under special types of operations – homeomorphisms. The network features, which are robust against space deformations are called topological invariants (because they proved to be permutation-invariant). Most parameters, suggested by mathematicians since Leonhard Euler in 18th century, were aimed at quantifying the shape of continuous, infinite spaces, especially 3-dimensional spheres. Numerous invariants were defined for vector spaces as corollaries of theorems proved on manifolds, where most points did not represent individual isolated objects. Nevertheless, some of theoretical topological studies are considered applicable not only to smooth surfaces, but also to relations between discrete entities (for instance, systems visualized as graphs). Besides, some achievements in the field of continuous spaces could be transposed to discrete or combinatorial geometry, and further applied to data science and statistics.

In particular, many topological studies of continuous and discrete spaces share the same topic: connectivity between points or individual objects. Expanding on observations made by colleagues studying connection density and graph sparsity, topologists emphasize, that connectivity or ruptures could be deformation-invariant: some places could remain inaccessible after network configuration had changed due to a mathematical operation or an event. Consequently, specific evaluation procedures were elaborated in order to quantify and compare spaces by their interrelatedness.

Reformulating a problem in topological terms and representing a dataset as a simplicial or cell complex or a hypergraph makes it possible to discover “handles” and “holes” besides them. Having defined connections of higher order (e.g. hyperedges “ECOWAS” and “G5 Sahel” are linked with 3 common points), we could go beyond traditional graph measures

like the minimum cut coefficient. Namely, the TDA is used to find points of multidimensional (and therefore intensive) communication between nodes – along handles –, and locate parts of a network lacking important links – where they are replaced by holes. The number of holes in a dataset could give an idea about the extent, to which the real network resembles an ideal fully-connected, dense graph. Researchers of dynamics could also count holes in order to evaluate, how many of these features disappeared after transformation (e.g. after network reshaping by the Covid pandemics): planes equivalent to “filled holes” taken into account in one of the earliest coefficients developed in topology – Euler characteristic/number.

Two TDA approaches to visits network dynamics

This quantitative method consists in looking for connected components including all the possible combinations of vertices (e.g. 2-simplices – trilateral meetings). Failure to find high-rank simplices could presuppose the existence of other topological features – holes. A hole between several meetings – (hyper-)edges of a network – could visualize parallel channels of communication. International configurations having these duplicating liaison circuits could have relatively high connectivity, whereas without this intermediaries (intermediate vertices) governments could have limited ways to exchange opinions). E.g. German government could get information about Latvian approach to the pandemic by using two channels:

- indirect (during the bilateral meeting with Kersti Kaljulaid, Estonian president on 30 June 2020 about the results of Kaljulaid's recent visit to Latvia of 25 June 2020);
- and direct (during personal meeting with Krišjānis Kariņš, Latvian prime minister on 18 July 2020) - to check the information gathered from Estonia.

Having compared the pre-Covid and early-Covid datasets, one could register a loss in the number of holes-circuits. Only 181 holes remained after the pandemics started, which could reflect a decrease in connectivity, i.e. confirm hypothesis 1. In particular, there is no evidence of any personal connection between leaders of any South-East Asian state with the rest of the world (whereas prior to Covid-19 these countries were fully integrated into global communications).

Topological data analysis could also be useful to confirm the emergence of communication barriers between particular actors. For this task the data should be transformed: these target nodes are separated from the rest, because researchers are

interested in the absence or presence of contacts with these special nodes. Nevertheless, binary arrays (one for every source node) are inspected for the same topological feature – hole. In this case its existence could mean that some obstacles prevents all the countries from reaching all these target states.

As France, Turkey and Nigeria had been selected as target nodes, the dataset was recalculated into evidence about combinations of target countries, which had been contacted by other governments during the studied periods. An edge was drawn on Fig. 1, if, for example, US president contacted only two countries out of the given triple, hyperedge or 2-simplex – if he met leaders of all the three target states (not necessarily at a quadrilateral meeting – not only simultaneously between France, Turkey, Nigeria and the US at the same table).

We were mostly interested in the following configurations: if authorities from this country had met only one pair of those leaders (so an edge appeared on Fig. 1, for example between Turkey and Nigeria), but had not reached the remaining third country (in this case, France). E.g. during the summer of 2020 Lybian government contacted France and Turkey, but not Nigeria. This fact could suggest some obstacle preventing Lybians from reaching Nigeria, although Libya is located on the same continent, unlike France and Turkey.

Unlike the above-mentioned processing of unchanged database, this technique would demonstrate the existence of a single hole after considering all the countries, which could theoretically contact France, Turkey and Nigeria. Another calculation method is used: all the countries’ “French-Turkish”, “Turkish-Nigerian” and “French-Nigerian” edges, as well as “French-Turkish-Nigerian” hyperedges or 1-simplices are pooled together on the same triangle (Fig. 1) and the computer finds out, whether this triangle remains empty or this hole is filled. Theoretically, one could fail to construct this triangle: one or several edges along its boundary could be absent if no president visited or hosted, e.g. French *and* Turkish leader during this period.

This approach has been established a decade ago to deal with accessibility problems - to check if actors are able to reach some objects (Ghrist et al., 2012; general principles of this approach were later applied to spatial computing: Derenick, Speranzon, Ghrist, 2013). To our knowledge, this approach has not been used in political geography yet.

Technically, this method presupposes computing persistent homology, which are often visualized using persistence diagrams. These homological structures appear (are “born”) during automatic step-by-step reconstruction of the cell complex – at recorded stages of the layout *process*. Many features are cancelled out after further steps (e.g. a hole could be

covered because a “French-Turkish-Nigerian” simplex has been found – like the middle point on Fig. 4), but some remain for infinite time (“persist” till the upper line on the diagram).

Therefore our hypothesis 2 would be confirmed, if the topological hole between France, Turkey and Nigeria remained till the end of our Covid computation, but eventually disappear (i.e. filled) in pre-Covid case. Fig. 4 and 5 demonstrate that this structure appeared during processing of both pre-Covid and Covid networks, but remained intact only for the 2nd quarter of 2020: the diagram for the beginning of 2020 shows, that this hole was filled (see the “short-lived” middle point) and at the same time a new topological feature appeared – the simplex which covered that hole (see the highest point).

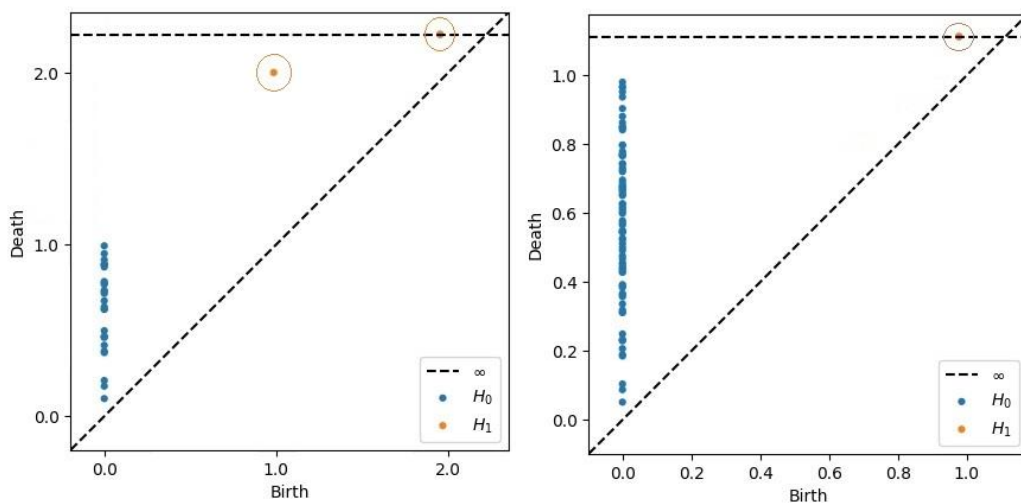


Fig. 4 and 5. Persistence diagrams visualizing homology, simplices and holes in two cell complexes around France, Turkey and Nigeria.

This difference between networks means hampered communication between France, Turkey and Nigeria: not only did their leaders cease direct personal contacts till end summer of 2020, but also no intermediary head of state personally visited all the three countries, hosted their leaders or met them on a “neutral” territory. This failure to maintain trustful connections could prevent governments from exchanging sensitive information required for globalization and interregional cooperation.

Discussion

In the given study hypergraphs of international high-level visits in the 1st and the 2nd quarter of 2020 have been compared using topological data analysis. This quantitative method has been applied to evaluate connections density prior to and after Covid-19 *pandemics* start.

Hypergraph databases and cell complexes have been used in order to model connections between more than two states at the same time (e.g. at a multilateral conference). These topological structures demonstrated, how simplices and holes could give a single dynamic characteristic to the whole international network and take into account disconnection (topological “holes”) between 3 or more entities.

Our conclusions about connectivity decrease (both globally and between France, Turkey and Nigeria) have been derived by interpreting the same algebraic structures (“holes”) in two opposite ways. For the first task of characterizing the *whole* network, holes represented duplicating circuits of communication (so a loss of these intermediated channels could mean fewer opportunities for sensitive data exchange). As for persistent homology analysis after database transformation (computation, whether any country leader established personal connections with all these 3 countries), the same topological feature represented communication barriers between *particular* 3 actors.

Acknowledgement:

The research was funded by the Russian Science Foundation (grant №23-77-10048, <https://rscf.ru/en/project/23-77-10048/>)

References

Aktas M., Akbas E., Fatmaoui A., 2019. "Persistence Homology of Networks: Methods and Applications". *Applied Network Science* 4 (61). doi: 10.1007/s41109-019-0179-3.

Derenick J., Speranzon A., Ghrist R., 2013. "Homological Sensing for Mobile Robot Localization". *IEEE International Conference on Robotics and Automation*. IEEE: Karlsruhe, pp. 572-579, doi: 10.1109/ICRA.2013.6630631, ISSN: 1050-4729.

Ghrist R. et al., 2012. "Topological Landmark-Based Navigation and Mapping", preprint <https://www2.math.upenn.edu/~ghrist/preprints/landmarkvisibility.pdf>. Accessed 25 June 2024.

The GDELT Project. <https://www.gdeltproject.org/>. Accessed 25 June 2024.