

Project description

(Source Sans Pro, font size 10)

Funding programme:	Programmes for Project-Related Personal Exchange (PPP) from 2024 with Brazil
Programme objective/s (outcomes) of the funding programme¹:	Outcomes are predefined and must not be changed
Programme objective (outcome) 1:	Junior scientists have gained international research experience and undergone further training at an international level
Programme objective (outcome) 2:	Binational research cooperation has been promoted and can be used as a starting point for future co-operations
Results of the measures/activities (outputs) of the funding programme²:	Outputs are predefined and must not be changed
Result (Output) 1:	Joint research results have been gained
Result (Output) 2:	International joint publications have been created

General information			
Title of the research project	Efficient statistical tools for networks and their applications		
Applying institution	Universität Leipzig		
Organisational unit	Institut für Mathematik und Informatik		
Discipline	Bioinformatik		
Principal investigator (PI)	Peter F. Stadler		
Requested approval period	from: 01.01.2024 to: 31.12.2025	For follow-up application ³ : Previous approval period:	from: 01.01.2021 to: 31.12.2022
Cooperation partner/partner institution (within and outside Germany)	University of São Paulo and Universität Konstanz	André Fujita	Brazil
Parallel funding and/or parallel application in another DAAD programme?	yes <input type="checkbox"/> no <input checked="" type="checkbox"/>		
If yes, please specify?	Please specify		

¹ The project does not necessarily need to aim at achieving all programme objectives (outcomes of the funding programme). 'Funding programme' and 'programme' are used synonymously.

² Only the results of the measures/activities (outputs of the funding programme) which are relevant for the selected programme objectives (outcomes of the funding programme) must be taken into account.

³ Follow-up application: Application for a project which immediately follows on from funding for the previous year in the same funding programme.

Is there already parallel funding and/or a parallel application to another funding body for the same project?	yes <input type="checkbox"/>	no <input checked="" type="checkbox"/>
If yes, please specify?	Please specify	

For follow-up applications: Progress from previous project

1. Please describe the previous project progress (implementation of measures/activities and achievement of objectives) (max. ½ DIN A4-page).

During the first funding phase we were able to complete research on tree-like networks, resulting in an efficient algorithm for estimating eigenvalue counts for this type of networks. This resulted in joint publication in J. Complex Networks. In addition to the DAAD-funded short travel, two PhD students from the Fujita lab in Sao Paulo are currently spending extended research stays in the Stadler lab in Leipzig. Bruno Ilha is working on metabolic networks, and Grover Guzman just arrived in Leipzig. He will continue the joint work on eigenvalues in simple network classes. The DAAD exchange has been very helpful in securing external funding for this purpose. As planned, a network meeting in Sao Paulo has taken place in 2022 with participation of 4 researchers from Leipzig including the PI, and Andre Fujita and collaborators have visited Leipzig in 2022 and in spring 2023. The second scheduled visit of the Leipzig group to Sao Paulo in the current funding period will take place in Nov/Dec 2023. Software development is ongoing. Current focus of the work is on improvements of the theoretical results. Early stage researches, notably Nora Beier, have been closely involved in preparing the annual report for 2022 and this renewal application.

Workshop on Network Statistics:

The workshop took place on September 15th and 16th at Prof. Imre Simon Auditorium at the Institute of Mathematics and Statistics of the University of São Paulo. Approximately 30 people attended this workshop, including (under) graduate students and faculty members from the University of São Paulo, Federal University of ABC, Federal University of Technology - Paraná, Sírío Libanês Hospital, Leipzig University, and the Hamburg University of Technology. We had two main talks - one by Prof. Peter Stadler (German coordinator) and one by Prof. André Fujita (Brazilian coordinator), seven oral presentations, and two group meetings (one each day) to allow all participants to build new and extend existing collaborations.

Publications:

1. Guzman, G.E.C., Stadler, P.F., and Fujita, A. (2022). Efficient eigenvalue counts for tree-like networks. Journal of Complex Networks 10, cnac040. 10.1093/comnet/cnac040.
2. Guzman, G.E.C., Stadler, P.F., and Fujita, A. (2021). Efficient Laplacian spectral density computations for networks with arbitrary degree distributions. Network Science 9, 312–327. 10.1017/nws.2021.10.

Project objectives, detailed project description and reference to results logic

1. Please state your project objectives (outcomes), which must be consistent with the programme objectives (outcomes) mentioned above and describe the specialised content of the project. Explanation should refer to the results logic: which specific project results (outputs or results of the

measures/activities) contribute to achieve these project objectives (outcomes)⁴.

2. Touch upon the relevance of your project and ensure that you address all selection criteria in the programme description, which are listed here again:
 - (1) Relationship of the project to the programme objectives (as per the impact analysis structure) and results-oriented planning by using indicators that meet the SMART criteria.⁵
 - (2) Please explain why and how long the planned stay abroad is necessary (respective details to be explained under "Measures/Activity planning")
 - (3) The quality of the project (clarity of project objectives and methods) and scientific relevance of the project (topical nature of the subject matter and the project's degree of innovativeness).
 - (4) Appropriate involvement of junior scientists
 - (5) Transfer of knowledge between the groups of researchers, added scientific value (subject-specific, institutional, interdisciplinary) created through the cooperation for both groups of researchers, scientific and, if applicable, industrial usability of the project results
 - (6) Feasibility of the research project (in particular: financial backing, preliminary work and further plans, adequate travel planning abroad - also with regard to the duration of the stay), project-relevant competence of both research groups, complementarity of the research groups in the joint project (methodologically, content-related, facilities, etc.)
3. Please describe potential risks in relation to the success of the project and explain how you plan to face them.

Please note:

The project objectives (outcomes) and intended results of the project's measures/activities (outputs) must be in line with the results-oriented project planning depicted in the project planning summary.

Outcomes

This proposal aims to consolidate the relationship between all partner groups. New collaboration opportunities will be identified, especially in regard to junior scientist's research network expansion. We will specifically initiate the development of computationally efficient statistical tools to analyze extensive empirical networks from a spectral distribution as well as a cycle-base angle.

Outputs

1. In person meetings in Brazil and Germany will enable all involved scientists to engage in topic specific discussions as well as get firsthand impressions of the work and living environments for future collaborations and research stays. We plan to organize talks every year and two short courses/workshops to also allow other students and researchers to participate. The PROBAL funding for the partner group has been granted for the current funding period, It is imperative therefore that to extend the DAAD funding to enable the continuation of the success cooperation.
2. The methods developed here will impact several fields of science. As aforementioned, random networks are ubiquitous. Thus, they will be helpful to analyze chemical compounds, social interactions, metabolic pathways, neural networks, and the internet. We expect that the works generated in this proposal will have a high impact, given the widespread interest in random networks. Thus, we expect that joint publications will strengthen the scientific footprint for all participating researches, especially young scientists, which will be advantageous for funding applications of their own.
3. All developed algorithms will be implemented as reference software packages e.g., in R or Python libraries. Collaboration in that regard will help junior scientists to develop a set of skills, from collaborative software development via open source platforms like GitHub to planning skills for software and project development as well as distribution and maintenance of open source software.
4. As in the first funding period junior scientists will be closely involved with the composition of the reports and the organization of the joint meetings. The project has already lead to additional longer-term scientific visits by members of the Fujita lab in Germany. This will form the basis for initiatives for

⁴ For the definitions of 'Outcomes' and 'Outputs', please refer to the 'Guide to Results-oriented Monitoring'.

⁵ See 'Guide to Results-oriented Project Planning and Monitoring', Chapter 2.

further funding.

Project description

Unlike deterministic graphs, empirical networks are stochastic, either by the underlying processes that generate them or the measurement procedures. For example, brain networks are different even among healthy individuals. Thus, many typical properties used to characterize graphs do not apply to large empirical networks. The reason is that they are not robust against the insertion or deletion of a small number of vertices or edges. Therefore, we need measures that quantify how close a graph is to exhibit a specific property, rather than the strict notion of isomorphism, which we rarely, if ever, attain.

Furthermore, empirical networks are usually massive. For example, it is estimated that the brain is composed of approximately 100 billion neurons. Thus, we cannot use current statistical approaches to analyze big data. The main reason is that we need to calculate the graph's spectrum, which is computationally expensive. Suppose a network is composed of n nodes. Then, the computational cost of naïve approaches, such as the diagonalization method, is $O(n^3)$. The message passing approach of Cantwell and Newman (2019) for the normalized Laplacian constituted a major advance. Still, it requires computing matrix inversions and matrix-vector multiplications, which are computationally expensive.

We know that the spectrum has codified structural characteristics of the network. For example, by analyzing the Laplacian spectrum, we can obtain its diameter (Chung et al., 1989), the number of spanning trees (Bollobás, 1998), vertex covers (Chen and Jost, 2012), Kemeny's constant (Pan et al., 2018), and chromatic number (Sun and Das, 2020). However, we do not know the contribution of a node to a network's spectral distribution. In other words, although we identify differences in the networks' spectra, we cannot associate these differences with the networks' structures. Therefore, we cannot interpret them. In the first funding phase we in particular obtained results for tree-like networks.

The partners will continue to work in parallel on complementary measures to quantify how close a graph is to exhibit a specific property. In the second project phase the project will in particular benefit from synergies with the MATOMICS project which produces detailed metabolic networks that can be reused in the context of the project. The added focus on metabolic networks also matches Bruno Ilha's interests, a member of Fujita's lab who spends most of 2023 in the Stadler lab in Leipzig. A common approach is determining graph invariants such as centrality measures (van den Heuvel and Sporns, 2013). However, graphs generated by different models may present similar centralities. Conversely, graphs generated by the same set of parameters may present a vastly different centrality measure. Thus, the analyses of empirical networks using methods grounded on deterministic graph theory seem to be inappropriate.

One potential solution is to assume that graphs are generated by probabilistic processes and then develop statistical methods, which will be the focus of the Brazilian partners. Statistical approaches for random graphs are new, with few reports in the literature (Asta and Shalizi, 2015; Ginestet et al., 2017; Tang et al., 2017; Cerqueira et al., 2017; Ghoshdastidar et al., 2017; Schieber et al., 2017; Kolaczyk et al., 2019). One of the reasons that graphs are challenging to study from a statistical viewpoint is that graphs are objects composed of vertices and edges, i.e., they are not numbers.

The group of Fujita has therefore experimented on the analysis of the graph spectrum, which "codifies" information about the graph structure (Takahashi et al., 2012; Fujita et al., 2017a, 2019), and developed a concept of correlation between vectors of graphs (Fujita et al., 2017b) which showed to be helpful to better understand new biological mechanisms, identify biomarkers, and find differences between controls and patients. The German side under Prof. Stadler has ample experience with the analysis of graph theoretical problems (Brian Davies et al., 2001; Gu et al., 2016; Hellmuth et al., 2009; Fritz et al., 2020) and will meanwhile focus on cycles. Cycles encapsulate semi-local information in a graph. Cycle bases provide well-defined, manageable cycle sets that can be computed efficiently. The length distribution of cycle sets such as the relevant cycles, i.e., those that are contained in at least one minimum cycle basis can be computed efficiently even without enumerating the sometimes

exponentially large cycle sets. We therefore plan to use cycle distributions as complementary source of information. To this end we make use of the connection with so-called cavity methods explored by Mark Newman and collaborators in recent years (Cantwell & Newman, 2019). The corresponding message-passing algorithms are exact on graphs without long "primitive cycles". Since very little is known on this graph class, we will characterize these graphs and their generalizations. To this end, we will in particular investigate the relationships between Laplacian eigenvalues and cycle distribution and explore to what extent and which graph classes they can be used for alternative classification tasks. Properties of cycle bases also characterize planarity and potentially other embedding properties. The latter is likely of particular relevance to application in brain-networks and other networks that are embedded into low-dimensional (Euclidean) spaces. As proposed in the original application, we plan for a second round of funding to systematically investigate the constraints of embeddings on the cycle distributions and the presence of cycles with special properties, such as long primitive cycles.

Studying the dynamic brain interaction network is vital to understand the brain's role in behavior. In the last decade, we witnessed the introduction of many methods to measure the presence or absence of dynamic interaction between brain parts. Nevertheless, little progress has been made on developing strategies to interpret and rigorously test the characteristics of the entire inferred brain interaction networks. We will develop rigorous statistical methods to compare ultra-high-dimensional networks and expect that these methods will allow us to correctly interpret the results of massive brain interaction networks obtained with state-of-art methods.

Our algorithms will allow us to identify genes or brain regions associated with diseases, abnormal connectivity structures, and changes over time, space, and subjects which will potentially lead to the development of drugs for treatment, biomarkers for diagnosis and prognosis, and a better understanding of the biological mechanisms.

Furthermore, our algorithms will be helpful in computer science, engineering, physics, and chemistry. E.g., for network feature extraction (Newman, 2018), low-rank approximation (Le et al., 2016; Luo et al., 2018), spectral clustering, and community detection (Newman, 2006). It also has applications in the dynamical systems theory (Porter and Gleeson, 2014), including structural phase transitions, such as percolation (Bollobás et al., 2010), localization (Martin et al., 2014), and detectability (Nadakuditi and Newman, 2012).

For all goals proposed in this project, we already supervise Ph.D. students and post-docs /post-doc candidates for the internships in Brazil/Germany, which will be tightly integrated in the development and application process. Having been able to secure external funding sources, this exchange also includes extended research stays for Bruno Iha and Grover Guzman in Leipzig in 2023/24.

We plan to send Ph.D. students and post-docs in both years of the second funding period to maintain constant communication. PIs will interact mostly via videoconference over the year and visit once a year. PIs will discuss manuscript and other proposals design during the scientific missions every year. We also plan talks in every Ph.D., post-doc, PIs visit. We will organize short courses/workshops and invite students/researchers of other universities to participate remotely (via videoconference) in the second and fourth years. We further plan to submit a proposal to the Research Group Linkage Programme (<https://bit.ly/3f1zS4k>) of the Alexander von Humboldt Foundation for further interaction between Brazilian and German groups. Dr. Fujita is an Alexander von Humboldt Fellow as such he satisfies the minimum requirement to submit a proposal to this call.

Available infrastructure

Prof. Fujita is the Brazilian coordinator. He has a fully equipped IT laboratory composed of dozens of high-performance workstations and computer servers. Together with the Interdisciplinary Center for Bioinformatics, the Stadler group at Leipzig University has sufficient computing power for all high performance-computing tasks associated with the proposed research. In addition, the group has access to the High-Performance Computer Center in Dresden and the de.NBI cloud, maintained by the German Network for Bioinformatics Infrastructure. Furthermore, Dr. Thomas Gatter, a staff scientist at

the Stadler Lab will contribute to this project and collaborate with local and Brazilian scientists. His involvement in the project will be funded from other sources and does not require DAAD funding.

Takahashi's lab is in the Brain Institute. The institute has a state-of-art primate facility, a primate surgery room, level 2 bio-security rooms, two-photon microscopy, molecular biology, and viral core facilities. The Brain Institute also has access to a supercomputer. Takahashi's lab has access to a fully trained veterinarian, animal welfare specialist, and husbandry team.

Dr. El Hady is affiliated with Universität Konstanz and the Max Planck Institute (MPI) of Animal Behavior. The MPI has one of the most advanced facilities to study animal behavior in the world. It is equipped with virtual reality arenas where researchers can change the environment in real-time. We can record animal behavior using multiple sensors (high time-of-flight cameras, ultrasound microphones) simultaneously. As part of the proposed research we will apply the methodologies developed throughout to data sets generated by these partner groups.

References

- ALEX, P, et al. Partitioning sparse matrices with eigenvectors of graphs, **SIAM Journal on matrix analysis and applications**, 11, p. 430 – 452, 1990.
- ASTA, DM, and SHALIZI, CR. Geometric network comparisons, **Proceedings of the Thirty- First Conference on Uncertainty in Artificial Intelligence. AUAI Press**, p. 102-110, 2015.
- BOLLOBÁS, B. Random graphs: Modern graph theory, **Springer**, p. 215 – 252, 1998.
- BOLLOBÁS, B, et al. Percolation on dense graph sequences, **The Annals of Probability**, 38, p. 150 – 183, 2010.
- BOOTH, TE. Power iteration method for the several largest eigenvalues and eigenfunctions, **Nuclear science and engineering**, 154, p. 48 – 62, 2006.
- BRIANDAVIES, E, et al. Discrete nodal domains theorems, **Linear Algebra and its Applications**, 336, p. 51 - 60, 2001
- CAN, ML, et al. Optimization via low-rank approximation for community detection in networks, **The Annals of Statistics**, 44, p. 373 – 400, 2016.
- CANTWELL, GT and NEWMAN MEJ. Message passing on networks with loops, **Proceedings of the National Academy of Sciences**, 116, p. 23398 - 23403, 2019.
- CERQUEIRA, A, et al. A test of hypotheses for random graph distributions built from EEG data, **IEEE Transactions on Network Science and Engineering**, 4, p. 75-82, 2017.
- CHEN, H and JOST, J. Minimum vertex covers and the spectrum of the normalized Laplacian on trees, **Linear algebra and its applications**, 437, 1089 - 1101, 2012.
- CHUNG, FRK. Diameters and eigenvalues, **Journal of the American Mathematical Society**, 2, p. 187 – 196, 1989.
- CHUNG, F, et al. The spectra of random graphs with given expected degrees, **Internet Mathematics**, 1, p. 257 – 275, 2004.
- FA-YUEH, W. Theory of resistor networks: the two-point resistance, **Journal of Physics A: Mathematical and General**, 37, p. 6653, 2004.
- FRITZ, A, et al. Cograph editing: Merging modules is equivalent to editing P_4 s, **The Art of Discrete and Applied Mathematics**, 3, 2, 2020.
- FUJITA, A, et al. A statistical method to distinguish functional brain networks, **Frontiers in Neuroscience**, 11, 1, 2017a.
- FUJITA, A, et al. Correlation between graphs with an application to brain network analysis, **Computational Statistics & Data Analysis**, 109, p. 76 – 92, 2017b.
- FUJITA, A, et al. A semi-parametric statistical test to compare complex networks, **Journal of Complex Networks**, 8, p. cnz028, 2020.
- GANDER, W. Algorithms for the QR decomposition, **Research Report**, 80, p. 1251 – 1268, 1980.
- GHOSHDASTIDAR, D, et al. Two-sample tests for large random graphs using network statistics, **Conference on Learning Theory. PMLR**, 2017.
- GINESTET, CE, et al. Hypothesis testing for network data in functional neuroimaging, **Ann. Appl. Stat.**, 11, p. 725-750, 2017.
- GU, J, et al. Spectral classes of regular, random, and empirical graphs, **Linear Algebra and its Applications**. 489, p. 30 - 49, 2016.
- HAIYAN, C and FUJI Z. Resistance distance and the normalized Laplacian spectrum, **Discrete Applied Mathematics**, 155, p. 654 – 661, 2007.
- HAO, C and JÜRGEN, J. Minimum vertex covers and the spectrum of the normalized Laplacian on trees, **Linear algebra and its applications**, 437, p. 1089 – 1101, 2012.
- HELLMUTH, M, et al. Approximate graph products, **European Journal of Combinatorics**, 30, p. 1119 - 1133, 2009.
- HUIMIN, L, et al. Computational drug repositioning using low-rank matrix approximation and randomized algorithms, **Bioinformatics**, 34, p. 1904 – 1912, 2018.
- KOLACZYK, ED, et al. Averages of unlabeled networks: Geometric characterization and asymptotic behavior, **Annals of Statistics**, 48, p. 514-538, 2020.
- LE, CM, et al. Optimization via low-rank approximation for community detection in networks, **The Annals of Statistics**, 44, p. 373 – 400, 2016.
- LUO, H, et al. Computational drug repositioning using low-rank matrix approximation and randomized algorithms, **Bioinformatics**, 34, 1904 – 1912, 2018.
- MARTIN, T, et al. Localization and centrality in networks, **Physical Review E**, 90, p. 052808, 2014.
- MASON, P, and GLEESON, JP. Dynamical systems on networks: A tutorial, **arXiv preprint arXiv:1403.7663**, 2014.
- METZ, FL, et al. Spectra of sparse regular graphs with loops, **Physical Review E**, 84, p. 055101, 2011.

MIROSLAV, F. Algebraic connectivity of graphs, **Czechoslovak Mathematical Journal**, 23, p. 298 – 305, 1973.

NADAKUDITI, RR and NEWMAN, MEJ. Graph spectra and the detectability of community structure in networks, **Physical Review Letters**, 108, p. 188701, 2012.

NADAKUDITI, RR and NEWMAN, MEJ. Spectra of random graphs with arbitrary expected degrees, **Physical Review E**, 87, p. 012803, 2013.

NEWMAN, MEJ. Modularity and community structure in networks, **Proceedings of the National Academy of Sciences**, 103, p. 8577 – 8582, 2006.

NEWMAN, M. Networks. **Oxford university press**, 2018.

NEWMAN, MEJ. Spectra of networks containing short loops, **Physical Review E**, 100, p. 012314, 2019.

NEWMAN, MEJ, et al. Spectra of random networks with arbitrary degrees, **Physical Review E**, 99, p. 042309, 2019.

PAN, Y, et al. On the normalized Laplacians with some classical parameters involving graph transformations, **Linear and Multilinear Algebra**, p. 1 - 23, 2018.

PHILLIP, B. Power and centrality: A family of measures, **American journal of sociology**, 92, p. 1170 – 1182, 1987.

PORTER, MA and GLEESON, JP. Dynamical systems on networks: A tutorial, **arXiv preprint arXiv:1403.7663**, 2014.

POTHEN, A, et al. Partitioning sparse matrices with eigenvectors of graphs, **SIAM Journal on matrix analysis and applications**, 11, 430 – 452, 1990.

ROGERS, T, et al. Cavity approach to the spectral density of sparse symmetric random matrices, **Physical Review E**, 78, p. 031116, 2008.

SEMERJIAN G and CUGLIANDOLO, LF. Sparse random matrices: the eigenvalue spectrum revisited, **Journal of Physics A: Mathematical and General**, 35, p. 4837, 2002.

SHAOWEI, S, KINKAR CH, D. Normalized Laplacian eigenvalues with chromatic number and independence number of graphs, **Linear and Multilinear Algebra**, 68, p. 63 – 80, 2020.

SIN, S et al. Parallel clustered low-rank approximation of graphs and its application to link prediction, **International Workshop on Languages and Compilers for Parallel Computing**, Springer, p. 76 – 95, 2012.

SLEIJPEN, GLG and VAN DER VORST HA. A Jacobi–Davidson iteration method for linear eigenvalue problems, **SIAM Review**, 42, p. 267 – 293, 2000.

SUN, S and DAS, KC. Normalized Laplacian eigenvalues with chromatic number and independence number of graphs, **Linear and Multilinear Algebra**, 68, 63 – 80, 2020.

TAKAHASHI, DY, et al. Discriminating different classes of biological networks by analyzing the graphs spectra distribution, **PLoS ONE**, 7, p. e49949, 2012.

TANG, M, et al. A nonparametric two-sample hypothesis testing problem for random graphs, **Bernoulli**, 23.3, p. 1599-1630, 2017.

VAN DEN HEUVEL, MP and SPORNS, O. Network hubs in the human brain, **Trends in Cognitive Sciences**, 17, p. 683 – 696, 2013.

YINGUI, P, JIANPING, L, SHUCHAO, L, WENJUN, L. On the normalized Laplacians with some classical parameters involving graph transformations, **Linear and Multilinear Algebra**, p. 1 – 23, 2018.

Measures/activities planning

Description of the measures/activities

Please describe the planned measures/activities (also see the category 'Measures/activities eligible for funding' in the call for proposals).

If necessary, insert new rows in the table for further planned measures/activities.

Note:

Measures/activities must be in line with the results-oriented project planning depicted in the project planning summary.

When describing the measures, you should also indicate, which research group will perform which work items and where this will take place. Please indicate which methods will be used.

Title of measure/activity 1:	First visit in Germany
Description:	In a first meeting of all involved research group we will give all junior scientist the opportunity to present their current work and plans for the project at hand to their peers. PostDocs and PIs will discuss the presented workplans and give valuable feedback. In this meeting we will also focus on the curation of data that are to be used for the development and testing of algorithms development in the course of this project.

Place and time frame	Germany: 01/2024
Title of measure/activity 2:	Travel of the German group to Brazil
Description:	During our meeting in Brazil all junior scientists will present their results and the state of their progress to a broader audience. We will consolidate results, discuss future directions and start work on manuscripts. We will also use this as an opportunity to broaden the collaborations in the area of network science and graph theory.
Place and time frame	Brazil: 11/2024
Title of measure/activity 3:	Second visit in Germany
Description:	Shortly after our visit in Brazil we will again meet in Germany to finalize work on manuscripts and begin work on the application for a second round of funding from the DAAD. Again, junior scientist will get the opportunity to present their results and discuss future approaches, this time to an audience from the German science network. Furthermore, we will work on the first project report.
Place and time frame	Germany: 01/2025
Title of measure/activity 4:	Second visit in Brazil
Description:	During the second and last visit of the German side to Brazil in the first funding period, we will focus on the finalization of manuscripts and discuss potential follow up projects as well as funding possibilities. Junior Scientist will present their (final) results in an international setting. The meeting will also serve to coordinate the preparation of the final report.
Place and time frame	Brazil: 11/2025

Sustainable impact and perspectives

Please explain how sustainable impact beyond the planned funding period may be achieved as well as potential perspectives for the project. Risk factors, responsibilities, and possible strategies to address these, should be outlined, too. (*max. ¼ DIN A4-Page*)

All participants of the proposed project will strengthen their scientific networks and improve their outreach. Junior scientist will have gained valuable insights in international project planning and execution, proposal and manuscript drafting and collaboration. All software and mathematical concepts developed in the course of this project will be made available via manuscripts and hosting platforms like GitHub to the interested public.

Preliminary results from the first funding period strongly suggest that the scientific goals related to the methodological parts of the proposal are indeed achievable. Moreover, given the already very successful first two years of the project, we do not expect risk factors for the success of the full project, although adjustments regarding the availability and quality of network data might still become necessary. Currently available public network data are sufficient, however, to achieve all major project goals. The participating project leads take responsibility for publication of results and software and tight integration of Junior scientist in all project related matters, including the preparation and writing of drafts of publications and reports.

Planned international mobility of the research groups

Please enter the planned stays of the researchers at the respective partner institution abroad during the funding period in the table below chronologically.

German project participant to perform stay abroad	Academic status	Research task to be performed	Duration in days	Planned date of stay (MM/YYYY)
Peter F. Stadler	Professor	Review of progress; Discussion of further directions	21	11/2024
Peter F. Stadler	Professor	Review of progress; Manuscript writing	21	11/2025
Jörg Fallmann	PostDoc	Discussion of results; Adaptation of workflows	21	11/2024
Jörg Fallmann	PostDoc	Manuscript and follow-up grant proposal finalization	21	11/2025
Bruno Schmidt	Doktorand	Presentation of results; work on manuscripts; exchange with students from Brazil	28	11/2024
Bruno Schmidt	Doktorand	Presentation of results; work on manuscripts; exchange with students from Brazil	28	11/2025
Nora Beier	Doktorand	Presentation of results; work on manuscripts; exchange with students from Brazil	28	11/2024
Nora Beier	Doktorand	Presentation of results; work on manuscripts; exchange with students from Brazil	28	11/2025

Project participant of the partner to perform stay in Germany	Academic status	Research task to be performed	Duration in days	Planned date of stay (MM/YYYY)
André Fujita	Associate Professor	Workshop and proposal writing to be submitted to AvH	10	04/2024
André Fujita	Associate Professor	Manuscript writing	10	04/2025
Daniel Y. Takahashi	Assistant Professor	Dr. El Hady's data analysis and discussions	10	04/2024
Daniela Bizinelli	Graduate student	Develop algorithm described in goal 1	180	04/2024
Jaqueline Yu Ting Wang	Graduate student	Develop algorithm described in goal 2	180	06/2024
Leonardo Sanches	Graduate student	Analyze Dr. El Hady's data	180	04/2024
	Postdoc	Develop algorithm described	270	04/2024

Caio Matheus Prates Batalha Faria		in goal 1		
Diego Trindade de Souza	Postdoc	Develop algorithm described in goal 2	270	04/2024

Further programme specific information

Structure of the research group

Briefly describe the structure of your research group and the criteria for selecting the project participants.

Dr. Fujita coordinates a FAPESP thematic project in network statistics, including dozens of graduate students and postdocs to whom we will provide training and internship. Thus, this proposal complements the FAPESP thematic project. The Stadler Lab in Leipzig has worked on several aspects of graph theory. While Fujita's team is specialized in statistics and will focus on spectral analysis, the Stadler group will tackle the problem from a cycle-base angle, thus both teams are complementing each other's work. Furthermore, two neuroscience teams, one in each country will be involved in the project, providing the groups with datasets for development and testing of developed algorithms. Our proposal ranges from theoretical/methodology development to application in neuroscience. Thus, this proposal comprises two groups of researchers, one of mathematics/computer science and one of neuroscience. Each group is composed of two labs. Mathematics/computer science: Dr. Stadler's and Dr. Fujita's labs. Neuroscience: Dr. El Hady's and Dr. Takahashi's labs. We based the participants selection criteria on the fitness for our problems treated in this proposal. Thus, participants should have background in at least one of the following areas: mathematics, theoretical computer science, statistics, neuroscience.

Roles in the project

Please explain what each project participant – inside and outside Germany – is responsible for within the project as a whole.

Peter F. Stadler (German coordinator) and André Fujita (Brazilian coordinator) will supervise the development of methods/algorithms and analysis of empirical data.
Daniel Y. Takahashi (Brazilian collaborator) and Ahmed El Hady (German collaborator) will provide the biological data to be analyzed and help with the interpretation of results.
Dr. Jörg Fallmann will supervise students on the German side and contribute to planning, manuscript writing and follow-up application while extending his own research network.
Nora Beier and Bruno Schmidt will work on the project in direct collaboration with our partners from Brazil, are responsible for internal progress reports and will help with manuscript and follow-up application preparation. Daniela Bizinelli, Jaqueline Yu Ting Wang, Leonardo Sanches, Caio Matheus Prates Batalha Faria and Diego Trindade de Souza will collaborate with the German side on data integration, analysis, algorithm development and implementation and also work on manuscript and follow-up application drafts.