WAW 2023
18th Workshop on Algorithms and Models for the Web Graph
May 23–26, 2023
https://math.torontomu.ca/waw2023/

Host: **Fields Institute for Research in Mathematical Sciences**, Toronto, Canada.
Location: Fields Institute http://www.fields.utoronto.ca/
222 College Street · Toronto, Ontario · M5T 3J1 · Canada

Steering Committee:

- **Andrei Z. Broder**, Google Research
- **Fan Chung Graham**, University of California, San Diego

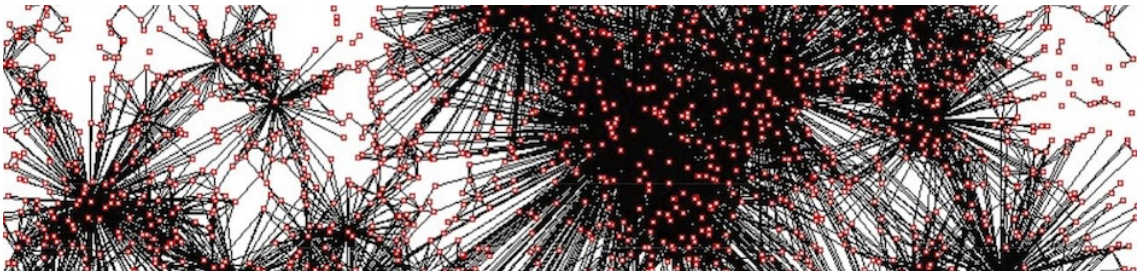
Co-Chairs and Co-organizers:

- **Megan Dewar**, Tutte Institute for Mathematics and Computing
- **Paweł Prałat**, Toronto Metropolitan University
- **Przemysław Szufel**, SGH Warsaw School of Economics
- **François Théberge**, Tutte Institute for Mathematics and Computing
- **Małgorzata Wrzosek**, SGH Warsaw School of Economics
- **Sebastian Zając**, SGH Warsaw School of Economics

1 Introduction

The World Wide Web has become part of our everyday life, and information retrieval and data mining on the Web are now of enormous practical interest. The algorithms supporting these activities combine the view of the Web as a text repository and as a graph, induced in various ways by links among pages, hosts and users.

The aim of the 18th Workshop on Algorithms and Models for the Web Graph (WAW 2023) is to further the understanding of graphs that arise from the Web and various user activities on the Web, and stimulate the development of high-performance algorithms and applications that exploit these graphs. The workshop will also welcome the researchers who are working on graph-theoretic and algorithmic aspects of citation networks, social networks, biological networks, molecular networks, and the Internet.



2 Sponsors

- Polish National Agency for Academic Exchange
  (under the Strategic Partnerships programme, grant number BPI/PST/2021/1/00069/U/00001)
- Google Research
- Springer Lecture Notes in Computer Science (LNCS)
- Tutte Institute for Mathematics and Computing (TIMC)
- Toronto Metropolitan University
- SGH Warsaw School of Economics

3 Schedule

Tuesday, May 23, 2023

- 9:30–10:30, Tutorial #1 — **Cliff Joslyn**
  A Gentle Introduction to Hypergraph Analytics using HyperNetX
- 10:30–11:00, Coffee Break
- 11:00–12:00, Tutorial #2 — **Przemysław Szufel**
  Introduction to Mining Graphs in Julia
- 12:00 – 1:00, Lunch Break
- 1:00 – 1:30, Official opening, a few words about Fields, Tutte, NAWA, TMU, Google
- 1:30 – 2:30, PLENARY TALK — **Remco van der Hofstad**
  It's hard to kill fake news
- 2:30 – 3:00, Coffee Break
- 3:00 – 3:30, **Kalle Alaluusua**, Konstantin Avrachenkov, Vinay Kumar B. R. and Lasse Leskelä.
  Multilayer hypergraph clustering using the aggregate similarity matrix
- 3:30 – 4:00, **Audun Myers**, Cliff Joslyn, Bill Kay, Emilie Purvine, Gregory Roek and Madelyn Shapiro.
  Topological Analysis of Temporal Hypergraphs
- 4:00 – 4:30, Bogumił Kamiński, Paweł Misiorek, Paweł Prałat and **François Théberge**.
  Modularity Based Community Detection in Hypergraphs

Wednesday, May 24, 2023

- 9:30 – 10:30, Tutorial #3 — **Paweł Prałat**
  Graph Embeddings and Their Unsupervised Evaluation
- 10:30 – 11:00, Coffee Break
- 11:00 – 12:00, Tutorial #4 — **Bogumił Kamiński**
  Introduction to Random Walks on Graphs in Julia
- 12:00 – 1:30, Lunch Break
- 1:30 – 2:30, PLENARY TALK — **Yeganeh Ali Mohammadi**
  Local Algorithms to Predict Epidemics on Networks
- 2:30 – 3:00, Coffee Break
- 3:00 – 3:30, **Martijn Gösgens**, Remco van der Hofstad and Nelly Litvak.
  Correcting for Granularity Bias in Modularity-Based Community Detection Methods
- 3:30 – 4:00, **Nicholas Sieger** and Fan Chung.
  A Random Graph Model for Clustering Graphs
- 4:00 – 4:30, Sayan Banerjee, Prabhanka Deka and **Mariana Olvera-Cravioto**.
  PageRank Nibble on the sparse directed stochastic block model

Thursday, May 25, 2023

- 9:30 – 10:00, **Łukasz Kraiński**.
  Scalable Embedding-based Graph Generator
- 10:00 – 10:30, **Tomasz Olczak**.
  Parallel algorithm for sampling large configuration model graphs in Julia
- 10:30 – 11:00, Coffee Break
- 11:00 – 11:20, Bogumił Kamiński, Pawel Pralat, François Théberge and **Sebastian Zajac**.
  Outlier detection with community structure on graphs
- 11:20 – 11:40, **Agata Skorupka**.
  Detection of anomalies in digital markets using graph data on the example of cryptocurrency markets and information
- 11:40 – 12:00, **Samin Aref**, Hriday Chheda and Mahdi Mostajabdaveh.
  The Bayan Algorithm: A Branch-and-Cut Method for Accurate Clustering of Networked Data Through Exact and Approximate Maximization of Modularity
- 12:00 – 1:30, Lunch Break
- 1:30 – 2:30, PLENARY TALK — **Claire Donnat**
  Graphs, Networks, and Estimation: Statistical and Machine Learning Perspectives
- 2:30 – 3:00, Coffee Break
- 3:00 – 3:30, Colin Cooper, Tomasz Radzik and **Nan Kang**.
  A simple model of influence
- 3:30 – 4:00, Peter Gracar, **Lukas Lüchtrath** and Christian Mönch.
  The emergence of a giant component in one-dimensional inhomogeneous networks with long-range effects
- 4:00 – 4:30, Anthony Bonato and **Ketan Chaudhary**.
  The Iterated Local Transitivity model for tournaments

Friday, May 26, 2023

- 9:30 – 10:00, **Julian Samaroo** and Przemysław Szufel.
  Distributed Computing over Graphs with Dagger.jl
- 10:00 – 10:30, **Carlo Lucibello**.
  GraphNeuralNetworks.jl: a geometric deep learning library for the Julia programming language
- 10:30 – 11:00, Coffee Break
- 11:00 – 11:30, **Przemysław Szufel** and Julian Samaroo.
  Distributed computing and unsupervised learning methods for multi-criteria area attractiveness assessment in multi-layered spatial networks based on OpenStreetMap data
- 11:30 – 12:00, **Claudio Moroni** and **Pietro Monticone**.
  Multilayer Network Science in Julia with MultilayerGraphs.jl
- 12:00 – 1:30, Lunch Break
- 1:30 – 2:30, PLENARY TALK — **Ernesto Estrada**
  Circum-Euclidean geometry of networks
- 2:30 – 3:00, Coffee Break

- 3:00 – 3:30, **Ashkan Dehghan**, Kinga Siuta, Agata Skorupka, Andrei Betlen, David Miller, Bogumil Kaminski and Pawel Pralat.
  Unsupervised Framework for Evaluating Structural Node Embeddings of Graphs
- 3:30 – 4:00, **Rouzbeh Hasheminezhad**, August Bøgh Rønberg and Ulrik Brandes.
  The Myth of the Robust-Yet-Fragile Nature of Scale-Free Networks: An Empirical Analysis
- 4:00 – 4:30, **Michal Dvořák**, Dušan Knop and Šimon Schierreich.
  Establishing Herd Immunity is Hard Even in Simple Geometric Networks

4 Program Committee

- **Konstantin Avrachenkov**, INRIA
- **Mindaugas Bloznelis**, Vilnius University
- **Paolo Boldi**, University of Milano
- **Anthony Bonato**, Toronto Metropolitan University
- **Ulrik Brandes**, ETH Zürich
- **Fan Chung Graham**, UC San Diego
- **Collin Cooper**, King's College London
- **Andrzej Dudek**, Western Michigan University
- **Alan Frieze**, Carnegie Mellon University
- **Jeannette Janssen**, Dalhousie University
- **Cliff Joslyn**, Pacific Northwest National Laboratory
- **Bogumil Kaminski**, SGH Warsaw School of Economics
- **Ravi Kumar**, Google
- **Lasse Leskela**, Aalto University
- **Nelly Litvak**, University of Twente
- **Oliver Mason**, NUI Maynooth
- **Pawel Misiorek**, Poznan University of Technology
- **Dieter Mitsche**, Universite de Nice Sophia-Antipolis
- **Peter Morters**, University of Cologne
- **Tobias Mueller**, Groningen University
- **Mariana Olvera-Cravioto**, University of North Carolina at Chapel Hill
- **Pan Peng**, University of Science and Technology of China
- **Xavier Perez-Gimenez**, University of Nebraska-Lincoln
- **Pawel Pralat**, Toronto Metropolitan University
- **Katarzyna Rybarczyk**, Adam Mickiewicz University
- **Vittorio Scarano**, University of Salerno
- **Przemyslaw Szufel**, SGH Warsaw School of Economics
- **François Théberge**, Tutte Institute for Mathematics and Computing
- **Yana Volkovich**, Microsoft
- **Nan Ye**, The University of Queensland
- **Stephen Young**, Pacific Northwest National Laboratory

5 Keynote Talks

5.1 Local Algorithms to Predict Epidemics on Networks

Yeganeh Alimohammadi, Stanford University

Abstract: People's interaction networks play a critical role in epidemics. However, precise mapping of the network structure is often expensive or even impossible. I will show that it is unnecessary to map the entire network. Instead, contact tracing a few samples from the population is enough to estimate an outbreak's likelihood and size. I will present a nonparametric estimator based on the contact tracing results and give theoretical guarantees on the estimator's accuracy for a large class of networks. Finally, I will present two examples of real-world applications of using the network structures for epidemic control: school reopening and placing vaccine sites.

Bio: Yeganeh is a fifth-year Ph.D. student at Stanford University, where she is advised by Amin Saberi. Her research interests are algorithm design and operations research with an emphasis on applications. In particular, she studies the theoretical grounds of network models of practical importance, mainly focusing on 1) studying epidemics on networks, 2) designing efficient sampling algorithms from large networks, and 3) network optimization.

5.2 Graphs, Networks, and Estimation: Statistical and Machine Learning Perspectives

Claire Donnat, University of Chicago

Abstract: The recent years have witnessed a surge in the amount of available structured data, typically modelled as a network capturing the relationships between different entities. This structure can exist "horizontally" across features, or "vertically" across observations, and can be leveraged to considerably improve estimation — two aspects that we propose exploring throughout this talk.

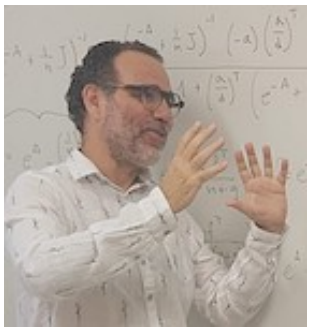
In the first part of this talk, we will concentrate on a classical statistical approach that utilizes graph structure for improved parameter estimation. We will employ network structure as a regularizer by introducing a new combined ℓ_1 and ℓ_2 penalty, known as the Generalized Elastic Net, for regression problems where feature vectors correspond to vertices of a given graph, and the true signal is assumed to be smooth or piecewise constant with respect to this graph. We will explore how this setting can be used to derive upper bounds for prediction and estimation errors under the assumption of a correlated Gaussian design.

In the second part of this talk, we will explore the assumption that observations are situated on a graph, an area recently investigated through the application of Graph Neural Networks (GNNs) within the Machine Learning community. Despite the proliferation of GNN methods across various tasks and applications, the effects of their individual components (e.g. activation function, convolution operator, etc) on their performance remain ambiguous. I will describe some recent work that attempts to understand the effect of the convolution operator used to aggregate information over entire neighborhoods on the geometry of the GNN embedding space.

Bio: Claire Donnat is an Assistant Professor in the Statistics Department at the University of Chicago. She focuses on the analysis and the development of methods for data with graph structure using both statistical and ML viewpoints. She completed her PhD in Statistics in 2020 at Stanford University, where she was supervised by Professor Susan Holmes and worked with Prof. Jure Leskovec. Prior to Stanford, she studied Applied Mathematics at Ecole Polytechnique (France), where she received her M.S and B.S equivalent.

5.3 Circum-Euclidean geometry of networks

Ernesto Estrada, Institute for Cross-Disciplinary Physics and Complex Systems



Abstract: Real-world networks are neither regular nor random, a fact elegantly explained by mechanisms such as the Watts-Strogatz or the Barabasi-Albert models. Both mechanisms naturally create shortcuts and hubs, which enhance network's navigability. They also tend to be overused during geodesic navigational processes, making the networks fragile against jamming. Why, then, networks with complex topologies are ubiquitous? I first will define here a measure of "communicability" between pairs of nodes in a network which is based on all weighted walks connecting both nodes. It converges to matrix functions of the adjacency matrix. Then, I will define a measure of the proximity between nodes in terms of their goodness of communication and prove that it is a circum-Euclidean distance between nodes in the network. Using this measure I will propose a geometrization of the graph such that we can navigate it through "shortest communicability paths". Then, I will prove that every circum-Euclidean distance matrix is the resistance matrix of a weighted graph. Continuing with the problem of network navigavility I will show how network models entropically generate network bypasses: alternative routes to shortest paths which are topologically longer but easier to navigate. I develop a mathematical theory that elucidates the emergence and consolidation of network bypasses and measures their navigability gain. Finally, I will apply this theory to a wide range of real-world networks and find that they sustain complexity by different amounts of network bypasses.

Bio: Ernesto Estrada is full research professor of the Spanish National Research Council at the Institute of Cross-Disciplinary Physics and Complex Systems (IFISC) in Palma de Mallorca, Spain. He is a SIAM Fellow, member of the Academia Europaea, Fellow of the Institute of Mathematics and Applications (FIMA), and of the Latin American Academy of Sciences. He has received several international distinctions including the "Wolfson Research Merit Award" of the Royal Society of London. Estrada has published more than 230 papers which have received more than 17,000 citations, published two textbooks with Oxford University Press and Edited one book with Springer. His main field of research is the mathematics of networks and their applications. He is well known for the "Estrada index" of a graph, the concepts of "network communicability", "d-path Laplacians", the use of fractional calculus on networks, as well as topics of algebraic network theory. He is a frequent speaker at international conferences on these topics.

5.4 It's hard to kill fake news

Remco van der Hofstad, Eindhoven University of Technology

Abstract: Empirical findings have shown that many real-world networks are scale-free, in the sense that there is a high variability in the number of connections of the elements of the networks. Spurred by these empirical findings, models have been proposed for such networks. In this talk, we investigate the spread of fake news on them.

We assume that news starts spreading from a source using a first-passage percolation rumour spread dynamics. The source later realises that the news is in fact wrong. After this realisation, it starts spreading the correct news. We make the (optimistic) assumption that a vertex, once having heard the correct version of the news item, will only spread the correct information. As such, we are modelling misinformation rather than fake news, with fake news being able to sustain on a network even longer.

Our results show that in many settings, even when the correct news spreads faster, the incorrect news is likely to reach a large part of the network. We distinguish between the incorrect news weakly surviving, meaning that it reaching a growing number of vertices, and strong survival, where the incorrect news reaches a positive proportion of the vertices. We give explicit criteria for the incorrect news to weakly and strongly survive on the configuration model, which is one of the most popular networks models. Interestingly, while from the definition, it is obvious that strong survival implies weak survival, from the analytical conditions this is highly non-trivial.

This lecture is based on joint work with Seva Shneer, and builds on earlier work with Gerard Hooghiemstra and Shankar Bhamidi.

Bio: Remco van der Hofstad received his PhD at the University of Utrecht in 1997, under the supervision of Frank den Hollander and Richard Gill. Since then, he worked at McMaster University in Hamilton, Canada, and Delft University of Technology. Since 2005, he is full professor in probability at Eindhoven University of Technology. Remco is further acting scientific director of Eurandom, and jointly with Frank den Hollander he is responsible for the 'Random Spatial Structures' Program at Eurandom.

Remco received the Prix Henri Poincare 2003 jointly with Gordon Slade, the Rollo Davidson Prize 2007, and is a laureate of the 'Innovative Research VIDI Scheme' 2003 and 'Innovative Research VICI Scheme' 2008. He is also one of the 11 co-applicants of the Gravitation program NETWORKS (see https://www.thenetworkcenter.nl/ for more information). In 2018, Remco was elected in the Royal Academy of Science and Arts (KNAW), where he currently is the chair of the Mathematics Section and member of the Board Natural and Technical Sciences.

Remco works on the mathematical foundations of networks, on statistical mechanics and on applications of probability in various other scientific disciplines. He has authored 2 books, on Random Graphs and High-dimensional Percolation, and some 190 articles.

Remco is editor in chief of the 'Network Pages', an interactive website by the networks community for everyone interested in networks (see https://www.networkpages.nl/ for more information). Remco is contact person for the research area Grip on Complexity of the Institute for Complex Molecular Systems. He is also the chair of the Board of Trustees of the Applied Probability Trust, and member of the Steering Committee of the Dutch NetSci chapter. Remco was spokesman for the Dutch Mathematics platform (2013-2019) 'Platform Wiskunde Nederland' (see http://www.platformwiskunde.nl/ for more information).

6 Tutorials

6.1 A Gentle Introduction to Hypergraph Analytics using HyperNetX

Presenter: **Cliff Joslyn**

Abstract: The Pacific Northwest National Laboratory in the United States has been supporting a concerted research effort in complex systems modeling incorporating a distinct perspective on complex networks. We cast complex systems as hypergraphs, that is, generalized graphs supporting multiple vertices per (hyper)edge. While all graphs are hypergraphs, proper hypergraphs can more faithfully represent complex interactions in systems, also reflecting their inherent multidimensional topological structure. We will first introduce the core formalism of hypergraph modeling, including duality, topological properties, and the lifting of traditional network science metrics into the hypergraph context. We will demonstrate equivalent representations integrating networks with relational structures, ordered structures, and finite topologies. Throughout, we will use HyperNetX (HNX, `https://pnnl.github.io/HyperNetX`), PNNL's hypergraph modeling platform, to concretely demonstrate these ideas.

6.2 Introduction to Mining Graphs in Julia

Presenter: **Przemysław Szufel**

Abstract: This is an introductory tutorial for people interested in the Julia programming language and its tools for analyzing graphs. No previous knowledge of Julia is required. However, it is assumed that participants have general programming experience and some experience with any other graph analysis library, such as Python's NetworkX. The tutorial will start with a concise introduction to Julia and its algebraic capabilities. Subsequently, the Graphs.jl (`https://github.com/JuliaGraphs/Graphs.jl`) library will be presented, including its options for graph generation, analysis, and visualization.

6.3 Graph Embeddings and Their Unsupervised Evaluation

Presenter: **Paweł Prałat**

Abstract: Users on social networks such as Twitter interact with and are influenced by each other without much knowledge of the identity behind each user. This anonymity has created a perfect environment for bot and hostile accounts to influence the network by mimicking real-user behaviour. To combat this, research into designing algorithms and datasets for identifying bot users has gained significant attention. Since bots can create content that is indistinguishable from human-generated text (think of GPT-3), the hope is to investigate network structure around bots in order to identify them. Indeed, the goal of many machine learning applications (not only bot detection algorithms) is to make predictions or discover new patterns using graph-structured data as feature information. In order to extract useful structural information from graphs, one might want to try to embed it in a geometric space by assigning coordinates to each node such that nearby nodes are more likely to share an edge than those far from each other. We will discuss such techniques and present a tool to select good embeddings.

6.4 Random Walks on Graphs in Julia

Presenter: **Bogumił Kamiński**

Abstract: In this tutorial we will consider a random walk on a graph, where we move from one node to one of its neighbors uniformly at random. The tutorial is organized in the following parts:

- checking is the probability of visiting a node and the average degree of visited node using stochastic simulation;
- verification of obtained results analytically and numerically;
- modification of the random walk rule so that the long-run probability of visiting a node in the graph is uniform (and checking it using stochastic simulation, numerically, and analytically).

For the tutorial we will use the GitHub Social Network dataset (`https://snap.stanford.edu/data/github-social.html`).

7 Talks Associated with Proceeding Papers

7.1 Multilayer hypergraph clustering using the aggregate similarity matrix

Authors: **Kalle Alaluusua**, Konstantin Avrachenkov, Vinay Kumar B. R. and Lasse Leskelä

Abstract: We consider the community recovery problem on a multilayer variant of the hypergraph stochastic block model (HSBM). Each layer is associated with an independent realization of a d-uniform HSBM on N vertices. Given the similarity matrix containing the aggregated number of hyperedges incident to each pair of vertices, the goal is to obtain a partition of the N vertices into disjoint communities. In this work, we investigate a semidefinite programming (SDP) approach and obtain information–theoretic conditions on the model parameters that guarantee exact recovery both in the assortative and the disassortative cases.

7.2 Topological Analysis of Temporal Hypergraphs

Authors: **Audun Myers**, Cliff Joslyn, Bill Kay, Emilie Purvine, Gregory Roek and Madelyn Shapiro

Abstract: In this work we study the topological properties of temporal hypergraphs. Hypergraphs provide a higher dimensional generalization of a graph that is capable of capturing multi-way connections. As such, they have become an integral part of network science. A common use of hypergraphs is to model events as hyperedges in which the event can involve many elements as nodes. This provides a more complete picture of the event, which is not limited by the standard dyadic connections of a graph. However, a common attribution to events is temporal information as an interval for when the event occurred. Consequently, a temporal hypergraph is born, which accurately captures both the temporal information of events and their multi-way connections. Common tools for studying these temporal hypergraphs typically capture changes in the underlying dynamics with summary statistics of snapshots sampled in a sliding window procedure. However, these tools do not characterize the evolution of hypergraph structure over time, nor do they provide insight on persistent components which are influential to the underlying system. To alleviate this need, we leverage zigzag persistence from the field of Topological Data Analysis (TDA) to study the change in topological structure of time-evolving hypergraphs. We apply our pipeline to both a cyber security and social network dataset and show how the topological structure of their temporal hypergraphs change and can be used to understand the underlying dynamics.

7.3 Modularity Based Community Detection in Hypergraphs

Authors: Bogumił Kamiński, Paweł Misiorek, Paweł Prałat and **François Théberge**

Abstract: In this paper, we make a significant step toward designing a scalable community detection algorithm using hypergraph modularity function. The main obstacle with adjusting the initial stage of the classical Louvain algorithm is dealt via carefully adjusted linear combination of the graph modularity function of the corresponding two-section graph and the desired hypergraph modularity function. It remains to properly tune the algorithm and design a mechanism to adjust the weights in the modularity function (in an unsupervised way), depending on how often nodes in one community share hyperedges with nodes from other communities. It will be done in the journal version of this paper.

7.4 Correcting for Granularity Bias in Modularity-Based Community Detection Methods

Authors: **Martijn Gösgens**, Remco van der Hofstad and Nelly Litvak

Abstract: Maximizing modularity is currently the most widely-used community detection method in applications. Modularity comes with a parameter that indirectly controls the granularity of the resulting clustering. Moreover, one can choose this parameter in such a way that modularity maximization becomes equivalent to maximizing the likelihood of a stochastic block model. Thus, this method is statistically justified, while at the same time, it is known to have a bias towards fine-grained clusterings. In this work,

we introduce a heuristic to correct for this bias. This heuristic is based on prior work where modularity is described in geometric terms. This has led to a broad generalization of modularity-based community detection methods, and the heuristic presented in this paper applies to each of them. We justify the heuristic by describing a relation between several distances that we observe to hold in many instances. We prove that, assuming the validity of this relation, our heuristic leads to a clustering of the same granularity as the ground-truth clustering. We compare our heuristic to likelihood-based community detection methods on several synthetic graphs and show that our method indeed results in clusterings with granularity closer to the granularity of the ground-truth clustering. Moreover, our heuristic often outperforms likelihood maximization in terms of similarity to the ground-truth clustering.

7.5 A Random Graph Model for Clustering Graphs

Authors: **Nicholas Sieger** and Fan Chung

Abstract: We introduce a random graph model for clustering graphs with a given degree sequence. Unlike previous random graph models, we incorporate clustering effects into the model without any geometric conditions. We show that random clustering graphs can yield graphs with a power-law expected degree sequence, small diameter, and any desired clustering coefficient. Our results follow from a general theorem on subgraph counts which may be of independent interest.

7.6 PageRank Nibble on the sparse directed stochastic block model

Authors: Sayan Banerjee, **Prabhanka Deka** and Mariana Olvera-Cravioto

Abstract: We present new results on community recovery based on the PageRank Nibble algorithm on a sparse directed stochastic block model (dSBM). Our results are based on a characterization of the local weak limit of the dSBM and the limiting PageRank distribution. This characterization allows us to estimate the probability of misclassification for any given connection kernel and any given number of seeds (vertices whose community label is known). The fact that PageRank is a local algorithm that can be efficiently computed in both a distributed and asynchronous fashion, makes it an appealing method for identifying members of a given community in very large networks where the identity of some vertices is known.

7.7 A simple model of influence

Authors: Colin Cooper, Tomasz Radzik and **Nan Kang**

Abstract: We propose a simple model of influence in a network, based on edge density. In the model vertices (people) follow the opinion of the group they belong to. The opinion percolates down from an active vertex, the influencer, at the head of the group. Groups can merge, based on interactions between influencers (i.e., interactions along 'active edges' of the network), so that the number of opinions is reduced. Eventually no active edges remain, and the groups and their opinions become static.

Our analysis is for $G(n,m)$ as m increases from zero to $N = \binom{n}{2}$. Initially every vertex is active, and finally G is a clique, and with only one active vertex. For $m \le N/\omega$, where $\omega = \omega(n)$ grows to infinity, but arbitrarily slowly, we prove that the number of active vertices $a(m)$ is concentrated and we give w.h.p. results for this quantity. For larger values of m our results give an upper bound on $E\,a(m)$.

We make an equivalent analysis for the same network when there are two types of influencers. Independent ones as described above, and stubborn vertices (dictators) who accept followers, but never follow. This leads to a reduction in the number of independent influencers as the network density increases. In the deterministic approximation (obtained by solving the deterministic recurrence corresponding to the formula for the expected change in one step), when $m = cN$, a single stubborn vertex reduces the number of influencers by a factor of $\sqrt{1-c}$, i.e., from $a(m)$ to $(\sqrt{1-c})\,a(m)$. If the number of stubborn vertices tends to infinity slowly with n, then no independent influencers remain, even if $m = N/\omega$.

Finally we analyse the size of the largest influence group which is of order $(n/k)\log k$ when there are k active vertices, and remark that in the limit the size distribution of groups is equivalent to a continuous stick

breaking process.

7.8 The emergence of a giant component in one-dimensional inhomogeneous networks with long-range effects

Authors: Peter Gracar, **Lukas Lüchtrath** and Christian Mönch

Abstract: We study the weight-dependent random connection model, a class of sparse graphs featuring many real-world properties such as heavy-tailed degree distributions and clustering. We introduce a coefficient, δ, measuring the effect of the degree-distribution on the occurrence of long edges. We identify a sharp phase transition in δ for the existence of a giant component in dimension $d = 1$.

7.9 The Iterated Local Transitivity model for tournaments

Authors: Anthony Bonato and **Ketan Chaudhary**

Abstract: A key generative principle within social and other complex networks is transitivity, where friends of friends are more likely friends. We propose a new model for highly dense complex networks based on transitivity, called the Iterated Local Transitivity Tournament (or ILTT) model. In ILTT and a dual version of the model, we iteratively apply the principle of transitivity to form new tournaments. The resulting models generate tournaments with small average distances as observed in real-world complex networks. We explore properties of small subtournaments or motifs in the ILTT model and study its graph-theoretic properties, such as Hamilton cycles, spectral properties, and domination numbers. We finish with a set of open problems and the next steps for the ILTT model.

7.10 Unsupervised Framework for Evaluating Structural Node Embeddings of Graphs

Authors: **Ashkan Dehghan**, Kinga Siuta, Agata Skorupka, Andrei Betlen, David Miller, Bogumił Kamiński and Paweł Prałat

Abstract: An embedding is a mapping from a set of nodes of a network into a real vector space. Embeddings can have various aims like capturing the underlying graph topology and structure, node-to-node relationship, or other relevant information about the graph, its subgraphs or nodes themselves. A practical challenge with using embeddings is that there are many available variants to choose from. Selecting a small set of most promising embeddings from the long list of possible options for a given task is challenging and often requires domain expertise. Embeddings can be categorized into two main types: classical embeddings and structural embeddings. Classical embeddings focus on learning both local and global proximity of nodes, while structural embeddings learn information specifically about the local structure of nodes' neighbourhood. For classical node embeddings there exists a framework which helps data scientists to identify (in an unsupervised way) a few embeddings that are worth further investigation. Unfortunately, no such framework exists for structural embeddings. In this paper we propose a framework for unsupervised ranking of structural graph embeddings. The proposed framework, apart from assigning an aggregate quality score for a structural embedding, additionally gives a data scientist insights into properties of this embedding. It produces information which predefined node features the embedding learns, how well it learns them, and which dimensions in the embedded space represent the predefined node features. Using this information the user gets a level of explainability to an otherwise complex black-box embedding algorithm.

7.11 The Myth of the Robust-Yet-Fragile Nature of Scale-Free Networks: An Empirical Analysis

Authors: **Rouzbeh Hasheminezhad**, August Bøgh Rønberg and Ulrik Brandes

Abstract: In addition to their defining skewed degree distribution, the class of scale-free networks are generally described as robust-yet-fragile. This description suggests that, compared to random graphs of the same size, scale-free networks are more robust against random failures but more vulnerable to targeted attacks. Here, we report on experiments on a comprehensive collection of networks across different domains that assess the empirical prevalence of scale-free networks fitting this description. We find that robust-yet-fragile networks are a distinct minority, even among those networks that come closest to being classified as scale-free.

7.12 Establishing Herd Immunity is Hard Even in Simple Geometric Networks

Authors: **Michal Dvořák**, Dušan Knop and Šimon Schierreich

Abstract: We study the following model of disease spread in a social network. In the beginning, all individuals are either *infected* or *healthy*. Next, in discrete rounds, the disease spreads in the network from infected to healthy individuals such that a healthy individual gets infected if and only if a sufficient number of its direct neighbours are already infected. We represent the social network as a graph. Inspired by the real-world restrictions in the current epidemic, especially by social and physical distancing requirements, we restrict ourselves to networks that can be represented as geometric intersection graphs. We show that finding a minimal vertex set of initially infected individuals to spread the disease in the whole network is computationally hard, already on unit disk graphs. Hence, to provide some algorithmic results, we focus ourselves on simpler geometric graph families, such as interval graphs and grid graphs.

8 Talks Associated with Abstracts

8.1 Multilayer Network Science in Julia with MultilayerGraphs.jl

Authors: **Claudio Moroni** and **Pietro Monticone**

Abstract: MultilayerGraphs.jl is a Julia package for the creation, manipulation and analysis of multilayer graphs, which have been adopted to model a wide range of complex systems from bio-chemical to socio-technical networks. The package has been integrated with the JuliaGraphs and the JuliaDynamics ecosystems and features an implementation that maps a standard integer-labelled vertex representation to a more user-friendly framework exporting all the objects a practitioner would expect such as nodes, vertices, layers, interlayers, etc. In our talk we will briefly introduce the theory and applications of multilayer graphs and showcase some of the main features of the current version of the package through a quick tutorial including: how to install the package; how to define layers and interlayers with a variety of constructors and underlying graphs; how to construct a directed multilayer graph with those layers and interlayers; how to add nodes, vertices and edges to the multilayer graph; how to compute some standard multilayer metrics.

8.2 Parallel algorithm for sampling large configuration model graphs in Julia

Authors: **Tomasz Olczak**

Abstract: A parallel implementation of the configuration model for sampling sparse graphs with large number of vertices and edges is proposed. We discuss parallelization strategy and optimized memory access patterns. We compare performance against state-of-the-art benchmarks. We show a multi-threaded implementation in Julia to produce a symmetric adjacency matrix of a graph with one billion edges in approximately 30 seconds on a commodity four-core CPU.

8.3 GraphNeuralNetworks.jl: a geometric deep learning library for the Julia programming language

Authors: **Carlo Lucibello**

Abstract: Machine learning on graphs has been rapidly advancing in recent years, thanks to the use of graph neural network architectures. In this talk, I will describe the basic ingredients of graph deep learning and highlight some applications in chemistry, biology, and combinatorial optimization. Then I will introduce GraphNeuralNetworks.jl, a Julia library that exposes primitive operators and common layers for creating rich and performant models and train them on cpu and gpu.

8.4 Distributed computing and unsupervised learning methods for multi-criteria area attractiveness assessment in multi-layered spatial networks based on OpenStreetMap data

Authors: **Przemysław Szufel** and Julian Samaroo

Abstract: Openly accessible maps such as OpenStreetMap (OSM, `https://wiki.openstreetmap.org/`) contains vast sources of information about environment, road system, businesses, public services as well as touristic attractions. This data is represented as spatial graphs, where each node represents some geographic location denoted by latitude and longitude. Nodes are grouped into ways that represent linear features on the ground. Nodes and ways are further grouped into relations that defined logical relationships such as a bus route or a university campus. The goal of this research is to develop an approach along with a set of tools to analyze OSM multi-layered graph structure to automatically discover multi-criteria measures of location's attractiveness across the entire geographic locations. This includes approaches for collecting the OSM data into usable multi-layered graph format as well as using data clustering, community detection based on graph modularity for automated measuring and classification of geographic regions. The abundance of

data requires developing methods for grouping graphs at various layers (map based transportation system, education leisure as well as relations present in other connected sources such as Wikipedia). The OSM data have a significant size and hence another major challenge is the processing such graphs at scale — for this project we have used Julia language together with the Dagger.jl library to achieve the required scalability in a distributed computing

8.5 Detection of anomalies in digital markets using graph data on the example of cryptocurrency markets and information

Authors: **Agata Skorupka**

Abstract: Digital markets are becoming one of most often used economic channels in developed economies. The digital market is defined as a market, where the subject and the object of the transaction rely on on-line presence (instead of physical one), which allows for significant benefits for both producers and consumers. Examples of reduced costs include search, transaction and verification costs. On the other hand, digital markets and lowering entry barriers open up the possibility for abuses in new forms (hereinafter referred to as anomalies). Examples of such anomalies may include providing false data, failure to fulfill the transaction, forging identity, disseminating false opinions, using fake accounts. As fake accounts, as well as text generation (meaning accounts' input, e.g. posts in social media) are nowadays easy to automate thanks to advancements in technology, which means that fake accounts are becoming bots or even botnets, they can be more and more prevalent and influential. Presented research utilizes graph structure of data stemming from digital markets, where a user is represented by a vertex, whereas a transaction - by an edge, and focuses on using graph data in order to detect anomalies. In particular, presented research explores and compares efficiency of algorithms based on graph embeddings, vertices' statistics, vertices' statistics enriched in information regarding neighbors, as well as compression algorithms, to propose a three-step procedure useful for anomaly detection in graph networks. The analysis is conducted on the example of cryptocurrency and information market graph (a.k.a. social network).

8.6 The Bayan Algorithm: A Branch-and-Cut Method for Accurate Clustering of Networked Data Through Exact and Approximate Maximization of Modularity.

Authors: **Samin Aref**, Hriday Chheda and Mahdi Mostajabdaveh

Abstract: Community detection is a fundamental problem in network science with extensive applications in various fields. The most commonly used methods are the algorithms designed to maximize modularity over different partitions of the network nodes. Using 80 real and random networks from a wide range of contexts, we investigate the extent to which current heuristic modularity maximization algorithms succeed in returning maximum-modularity (optimal) partitions. We compare eight existing heuristic algorithms against an exact integer programming method that globally maximizes modularity. The average modularity-based heuristic algorithm returns optimal partitions for only 16.9% of the 80 graphs considered. Additionally, results on adjusted mutual information reveal substantial dissimilarity between the sub-optimal partitions and any optimal partition of the networks in our experiments. More importantly, our results show that near-optimal partitions are often disproportionately dissimilar to any optimal partition. Taken together, our analysis points to a crucial limitation of commonly used modularity-based heuristics for discovering communities: they rarely produce an optimal partition or a partition resembling an optimal partition. If modularity is to be used for detecting communities in small and mid-sized networks, exact or approximate optimization algorithms are recommendable for a more methodologically sound usage of modularity within its applicability limits. (`https://arxiv.org/abs/2302.14698`)

We also propose a specialized algorithm, Bayan, which returns partitions with a guarantee of either optimality or proximity to an optimal partition. At the core of the Bayan algorithm is a branch-and-cut scheme that solves an integer programming formulation of the problem to optimality or approximate it within a factor. We compare Bayan's accuracy and stability with 21 other algorithms in retrieving ground-truth communities in synthetic benchmarks (LFR) and node labels in real networks. Bayan is several times faster

than open-source and commercial solvers for modularity maximization making it capable of finding optimal partitions for instances that cannot be optimized by any other existing method. Overall, our assessments point to Bayan as a suitable choice for exact maximization of modularity in networks with up to 3000 edges (in their largest connected component) and approximating maximum modularity in larger networks on ordinary computers. A Python implementation of the Bayan algorithm (bayanpy) is publicly available through the package installer for Python (pip). (`https://arxiv.org/abs/2209.04562`)

8.7 Outlier detection with community structure on graphs.

Authors: Bogumił Kamiński, Pawel Pralat, François Théberge and **Sebastian Zajac**

Abstract: The community structure of networks is one of the essential properties of empirical graphs, for example representing social networks. There are multiple algorithms proposed for community detection. One of the most used unsupervised methods for identifying communities in graphs is maximising the modularity function. In the presentation, we will discuss the results of modularity function modification and how it can be used for outlier detection on artificially generated and real-life datasets. Also, we will talk about possible extensions of the approach to treating small communities.

8.8 Distributed Computing over Graphs with Dagger.jl

Authors: **Julian Samaroo** and Przemysław Szufel

Abstract: Dagger.jl is a Julia package which utilitizes a DAG as its core computational currency, and builds many powerful and extensible features on top of this abstraction. Key among them is the ability to perform heterogeneous computing; a given node in a Dagger DAG can be executed on any thread, of any Julia process, on any server within the compute cluster. Additionally, nodes can execute on the CPU, GPU, or (in the future) on other accelerator hardware like TPUs or FPGAs. Like Julia's own native Tasks, the nodes in a DAG execute asynchronously, and the results produced by a node can be fetched at any time.

Dagger is aiming to take over the distributed computing space in Julia with an approach that is easy to use, scalable, and flexible. By removing the barriers to writing distributed, heterogeneous programs, Dagger hopes to make distributed computing accessible to all, and provides an easy way to write programs that scale from laptops to supercomputers.

In this talk, I'll discuss how Dagger works at its core, and how the internal DAG representation is beneficial for implementing many of Dagger's useful features. I'll also show some examples of how graph representations of code can allow us to build powerful abstractions and provide many opportunities for performance optimizations.

8.9 Scalable Embedding-based Graph Generator

Authors: **Łukasz Kraiński**

Abstract: The talk introduces an Embedding-based Graph Generator (EGG) that can produce a synthetic graph when given a graph and its embedding as input. The EGG algorithm focuses on scalability and can generate networks with hundreds of thousands of nodes while maintaining quality comparable to state-of-the-art techniques.