Research Statement 2021

Mikko Kivelä
Assistant Professor
Department of Computer Science
Aalto University School of Science
mikko.kivela@aalto.fi

26 March 2021

My research focuses on the type of complex systems which can be represented as networks, *i.e.* systems consisting of elementary parts with non-trivial interactions that together display emergent properties. On the one hand, network science offers a rich tapestry of theoretical problems, which are in the intersection of fields such as computer science, graph theory, and dynamical systems. On the other hand, networked systems are ubiquitous in nature and society, and network theory is a highly interdisciplinary field. As an example, my research has so far included collaborations with researchers from fields such as psychology, population ecology, engineering, medicine, political science, communication, finance, and bioinformatics. Several grand challenges and burning questions we are currently facing are inherently questions about networks. As detailed in this statement, I am addressing a selected set of such questions in my research.

The field of network science is reaching a mature age, and this comes with challenges and opportunities. Since the revolution in network science around 20 years ago [36], the field has been extremely successful explaining various phenomena and fundamental concepts in a wide array of systems from societies to brain and cellular biology. Most of this progress has been achieved by using the same theories and tools across the seemingly different systems. While many fundamental questions in networked systems could be answered by mapping them into simple network abstractions (simple graphs), the field is moving more and more towards a fine-grained understanding of these systems. This not only requires working more closely with domain experts but also new tools and abstractions that can retain some of the general nature of network science. Here my work has been at the international forefront: I have developed a representation called *multilayer networks* which is now routinely used in almost every field where the traditional network analysis has also been successful. In addition, I have pioneering work on temporal networks, i.e., networks that change in time. Both multilayer networks and temporal networks are still growing areas of research, with several problems in application areas ready to be tackled with the new tools and theory, but at the same time several theoretical challenges that need to be addressed. My research is at the crossing of these two aspects.

I have divided this statement into two main sections: *foundations of network science* describes my work in the more theoretical aspects of network science. This includes topics at the core of my expertise, such as multilayer networks or temporal networks. The section *application areas* describes the various areas of research where my work has made an impact and I am either currently working on or have medium-term research plans for. Here I have chosen the most interesting and important applications I want to work on. This includes areas that have a potential for high impact and are of wide interest, such as climate change, social polarization, and disease spreading. The main motivation for me to work with them are the theoretical challenges they offer: A problem that is both critically important and unsolved has most likely escaped being fully understood because its theoretical underpinnings are not understood yet.

1 Foundations of network science

1.1 Multilayer networks

As network science has matured and as ever more complicated data has emerged, it has become increasingly important to develop tools to analyze more complicated structures. For example, many systems that were typically initially studied as simple graphs are now often represented as networks with multiple types of connections or interdependent networks. This has allowed deeper and more realistic analyses of complex networked systems, but it has simultaneously introduced mathematical constructions, jargon, and methodology that are specific to research in each type of system. My work on the development of the concept of *multilayer networks* [15] has the aim of unifying the aforementioned disparate language (and disparate notation) and to bring together the different generalized network concepts that included layered graphical structures.

Multilayer networks have the potential of unifying tools and theories related to many types of generalized networks. The downside is that these general tools and theories can become very abstract and complicated with many degrees of freedom. In the literature, this has meant that scholars in application areas have more difficulties using such methodology, and there have been several ad-hoc generalizations of various network concepts. My aim is to do grounded research starting from the first principles. For example, in my previous work, I generalized walks in multiplex networks in order to then generalize clustering coefficients [9]. I have further generalized the concept of graph isomorphisms, which is a very fundamental notion behind many other methods, for multilayer networks [19]. Currently, my research group is working on applying this method for graphlet analysis methods [28] (which is equivalent to counting automorphism orbits in a graph) and motif analysis methods [23]. Both of these concepts will have immediate use-cases in application areas as explained later. In the medium-term, my plan is to develop theory and methods on a related concept of edit distances [10, 11] in multilayer networks. This will have immediate applications on anonymization [30] and in alleviating the problems related to the combinatorial explosion in multilayer subgraphs.

Theory and methods related to multilayer networks are far from being complete. My aim is to continue working in this area. In addition, I am looking into openings with related concepts such as hypergraphs. My theoretical work is strongly driven by problems arising in applications, and I aim to continue with this route. Outstanding problems in application areas are often outstanding because of some gap in the theoretical understanding, and this is a way to identify both practically important problems and theoretically interesting ones. Another clear way forward is doing technology transfer between areas. For example, I plan to work on randomized reference models for multilayer networks, and by doing so continue the work I started within the domain of temporal networks [12].

1.2 Temporal networks

Temporal networks, i.e., networks where nodes and links are only active at some specific time instances, can be seen as a special case of multilayer networks. However, such multilayer analysis would disregard the general tools and ideas related to analyzing temporal data. In fact, working with temporal networks is often at the intersection of analyzing point processes and graphs.

My early work was part of the foundations of the field of temporal networks. In this work, we showed that it is crucial to consider realistic dynamics of link activations when studying processes relying on the connectivity of the networks (e.g., disease spreading) [14, 18]. We demonstrated how the burstiness of human behavior affects such processes, which is in contrast to previous literature which was built on the assumption that social connections follow the Poisson process. I also created a systematic way of analyzing temporal networks using reference models and hierarchies they induce [18]. These results later lead to a body of literature applying the same framework to other dynamics on temporal networks.

After taking a break from temporal networks to focus on the development of the multilayer network framework, I returned to the field by addressing problems that had emerged. I developed a statistical method to correct for previously unnoticed sampling bias that had partly caused some debates in the literature about the nature of the local dynamics observed in temporal networks [20]. We also developed transformation to turn the problem of quantifying burstiness in time series into a problem of analyzing the structure of a tree [13]. Finally, I was recently part of a multi-year effort from an international team of experts to review and create a coherent framework on temporal reference models [12], which I initially conceived around 10 years ago [14, 18]. My main role was to create a formal theory for such reference model, which can be used also outside of temporal networks to any random reference models of any discrete structures [12].

My future goals for temporal networks are more ambitious. A major theoretical and practical problem in temporal networks is that there is no theory of connectivity. In normal graphs, for example, disease spreading is intimately related to connectivity and the phenomena can be understood using tools of percolation theory for graphs [24, 25]. With my research team, our aim is to develop such a theory of connectivity for temporal networks [17], by mapping the problem to *directed percolation* in the physics literature. To this end, we have recently developed a new algorithm for computing connectivity which solves problems related to the scaling of the computations and has made some computations in our empirical networks that previously would have taken thousands of years to complete with the previous methods to run in a few hours [2]. This theoretical work is promising, but there remain some open problems and risks. If successful we can open up a completely new and more principled way of investigating temporal networks and many processes taking place on them, such as spreading (e.g., for disease models), diffusion (e.g., in social networks), and error tolerance (e.g., in public transport). This would mean that we would not be bound by so much by simulation results but could understand the underlying phenomena in these temporal systems similar to the way we understand spreading phenomena on static networks [25].

2 Application areas

I will next list application areas where I have current active projects that have the potential of being expanded to larger branches of research in the future. Note that this is not a comprehensive list of possible future topics. My focus application areas are likely to change somewhat even in the medium term depending on the opportunities and problems at hand.

2.1 Social networks

Social networks have been the main application area in my work since the beginning of my career [33, 34, 14, 26, 18, 9, 20, 29, 6, 1, 8, 30, 37, 35, 31]. I plan to continue within this wide research line both by completing smaller projects and by investigating larger lines of research. Below I describe two such research lines. For smaller projects, a good example is a project that I am currently working on together with the group of Prof. Eero Hyvönen on large historical correspondence data set, where the aim is to compare historical social networks to the modern ones.

2.1.1 Political communication and polarization

Political polarization is an increasingly difficult problem for societies. It is inherently a problem of social networks becoming clustered due to various mechanisms acting on creating and maintaining the structure. For example, our recent work shows the innate tendency for individuals to connect to similar people (for example in terms of beliefs, race, or gender) is amplified in social networks due to a mechanism known as triadic closure [1].

As polarization is a pressing issue, it is also a rapidly developing research area. As with my other research my aim is to produce methodologically well-grounded research rather than publishing fast results on an emerging topic. For example, we recently published an article showing that all current measures of polarization in networks can produce very high scores even for random networks and developed a method for fixing this problem [31].

Most models of polarization in social networks (including recent ones [7]), are based on a simplifying assumption that polarization and echo chambers are created in a single social network where a belief and the network topology interact. This is in contrast to empirical findings of separate networks being formed based on each controversial topic [8], and the literature that indicate that polarization only becomes problematic when multiple opinion clusters are aligned [22, 3]. My research group is currently in the process of developing theory, models, and data analysis methods for multilayer network polarization to address these issues.

Beyond theoretical work, I have analyzed polarization within an Academy of Finland consortium project "Echo Chambers, Experts, and Activists: Networks of Mediated Political Communication". Here our aim is to understand multilayer networks of political communication via social media, traditional media, and direct communication between the various actors. This work has already yielded several publications using data on climate change discussions and the Finnish political landscape as testbeds [8, 37, 31], and we are currently expanding this to other countries and temporal analysis. Further, the work plan includes a novel network model explaining our empirical findings.

Related to this project, I ran a small on monitoring social media during the Finnish parliamentary and EU elections of 2019, which was commissioned by the Ministry of Education, The Security Committee, and the Prime Minister's Office [32]. I am currently in discussion regarding social media monitoring and action plan during the local elections in 2021. Even though there are several scientific challenges related to this project, I see it more as a service to the local society rather than a research project, in which I usually aim to have a global impact.

2.2 Investor networks

Investor networks are bipartite networks of individuals and assets they invest in. One can analyze these networks directly [5], or build networks of investors or assets. These networks can further be analyzed as multilayer networks [4] or temporal networks [29]. This work builds on a unique data set of all investments in the Helsinki stock exchange recorded since the beginning of the 90s, which amounts to a huge number of individual interactions. Our main aim is to transfer the techniques and knowledge of analyzing human behavior from large data sets that the field of computational social science has on behavioral economics, which is still mostly based on more traditional methods. To our surprise, in our initial work, we have seen that many of the laws governing human social behavior, for example, related to attention, can also be found in how people invest in stocks [5]. These similarities (and differences) in how people manage their *social capital* to how they manage their *monetary capital* is at the forefront of this research direction. We are also looking into overall networks of the flow of money between assets in stock markets and how they change in different market conditions.

2.3 Network neuroscience

The brain is one of the archetypical complex systems where the complexity does not arise so much from the individual units (neurons) but from the network connecting them. Network neuroscience has rapidly taken off within the last ten years. This new field is still struggling with fundamental notions such as how to define nodes in a consistent way in networks built out of fMRI data [21]. The main problem is that conventional network analysis assumes that nodes are static entities, whereas in brains the different areas keep changing their contents depending on the task they are performing. Our aim is to tackle these problems by using multilayer networks that have more flexibility on how networked data is presented. More precisely, we will let the nodes dynamically change following the data, and link the dynamic nodes across the layers that represent time indicating the similarity of the nodes in previous layers. This data-driven approach will give us a completely new kind of view of the coarse-grained functionality of brains and removes major problems related to forcing the data into a framework that is not suitable for it. Our long-term goal is to fundamentally change how network neuroscience is done currently and make it shift towards multilayer network methods.

2.4 Ecological networks

Ecosystems are driven by networks of individuals and populations interacting at many levels. Within the emerging field of network ecology, I have previously worked with marine biologists and contributed by creating tools for them to analyze ecological networks [16]. Within this field, it has become clear that ecosystems are often ideally represented as multilayer networks [27]. Our aim with my marine ecologist collaborators is to initiate projects to collect and analyze data on multilayer marine ecology networks. This is still an emerging field within ecology with good prospects for future research, and now is a good time to enter this research area.

I am the Finnish management committee member in the COST Action Sea-Unicorn (to start at the end of 2020), where one of the goals is to collect multilayer data from sea ecosystems and develop methods to analyze it.

2.5 Disease spreading and COVID-19

My work on temporal networks has been focusing on spreading processes starting from the early work I did on the topic already 10 years ago [14, 18]. In the current situation with COVID-19 large number of scholars in network science are working on disease spreading models, including many in the national modeling teams. Epidemic modeling has taken giant leaps forward within the last year, and networks have turned out to be an important modeling framework, with the more traditional epidemiology community also getting more interested in these models.

I have worked on theoretical spreading models predicting the usefulness of mobile phone applications for tracking infections [6]. The inclusion of applications to epidemic models introduces memory of the infection path to the epidemic models that could have previously been modeled as memoryless branching processes. My work here show that the fraction of app users needs to be very larger in order for them to make any difference, and the ongoing work within my group is elaborating on this by including various types of heterogeneities to the population. We also recently won a 2-year Nordforsk grant with a Nordic consortium for modeling COVID-19. Here our aim is to produce practical modeling solutions related to spatial spreading models and mobile phone operator tracking data (which we have acquired from a major operator in Finland). This type of modeling is done by our collaborators in Norway but not within the Finnish government-supported modeling team of the Finnish Institute for Health and Welfare (THL). Most recently we have been working on the theory of vaccination strategies.

3 Summary

Within my first term as an assistant professor at Aalto University, I have achieved the short-term theoretical goals I set out to have in my research statement 4 years ago. I have adapted the application areas, for example by adding COVID-19 research into my research agenda, and several lines of work have already lead to successes both in terms of funding and publications.

References

[1] Aili Asikainen, Gerardo Iñiguez, Javier Ureña-Carrión, Kimmo Kaski, and Mikko Kivelä. Cumulative effects of triadic closure and homophily in social networks. *Science Advances*, 6(19):eaax7310, 2020.

[2] Arash Badie-Modiri, Márton Karsai, and Mikko Kivelä. Efficient limited-time reachability estimation in temporal networks. *Physical Review E*, 101(5):052303, 2020.

[3] Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408–446, 2008.

[4] Kęstutis Baltakys, Juho Kanniainen, and Frank Emmert-Streib. Multilayer aggregation with statistical validation: Application to investor networks. *Scientific reports*, 8(1):1–12, 2018.

[5] Kęstutis Baltakys, Juho Kanniainen, Jari Saramäki, and Mikko Kivela. Trading signatures: Investor attention allocation in stock markets. *Available at SSRN*, 2020.

[6] Alain Barrat, Ciro Cattuto, Mikko Kivelä, Sune Lehmann, and Jari Saramäki. Effect of manual and digital contact tracing on covid-19 outbreaks: a study on empirical contact data. *medRxiv*, 2020.

[7] Fabian Baumann, Philipp Lorenz-Spreen, Igor M Sokolov, and Michele Starnini. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124(4):048301, 2020.

[8] Ted Hsuan Yun Chen, Ali Salloum, Antti Gronow, Tuomas Ylä-Anttila, and Mikko Kivelä. Polarization of climate politics results from partisan sorting: Evidence from finnish twittersphere. *arXiv:2007.02706*, 2020.

[9] E. Cozzo, M. Kivelä, M. De Domenico, A. Solé, A. Arenas, S. Gómez, M. A. Porter, and Y. Moreno. Structure of triadic relations in multiplex networks. *New J. Phys.*, 17(7):073029, 2015.

[10] Frank Emmert-Streib, Matthias Dehmer, and Yongtang Shi. Fifty years of graph matching, network alignment and network comparison. *Information sciences*, 346:180–197, 2016.

[11] Xinbo Gao, Bing Xiao, Dacheng Tao, and Xuelong Li. A survey of graph edit distance. *Pattern Analysis and applications*, 13(1):113–129, 2010.

[12] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L Vestergaard. Randomized reference models for temporal networks. *arXiv preprint arXiv:1806.04032*, 2018.

[13] Hang-Hyun Jo, Takayuki Hiraoka, and Mikko Kivelä. Burst-tree decomposition of time series reveals the structure of temporal correlations. *Scientific reports*, 10(1):1–11, 2020.

[14] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E*, 83(2):025102, 2011.

[15] M Kivelä, A Arenas, M Barthelemy, J P Gleeson, Y Moreno, and M A Porter. Multilayer networks. *J. Complex Netw.*, 2(3):203 – 271, 2014.

[16] Mikko Kivelä, Sophie Arnaud-Haond, and Jari Saramäki. Edenetworks: A user-friendly software to build and analyze networks in biogeography, ecology and population genetics. *Molecular Ecology Resources*, 2014.

[17] Mikko Kivelä, Jordan Cambe, Jari Saramäki, and Márton Karsai. Mapping temporal-network percolation to weighted, static event graphs. *Scientific reports*, 8(1):1–9, 2018.

[18] Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, Jari Saramäki, and Márton Karsai. Multiscale analysis of spreading in a large communication network. *J. Stat. Mech.*, 2012(03):P03005, 2012.

[19] Mikko Kivelä and Mason A. Porter, 2015. arXiv:1506.00508 [physics.soc-ph].

[20] Mikko Kivelä and Mason A. Porter. Estimating interevent time distributions from finite observation periods in communication networks. *Phys. Rev. E*, 92:052813, 2015.

[21] Onerva Korhonen, Heini Saarimäki, Enrico Glerean, Mikko Sams, and Jari Saramäki. Consistency of regions of interest as nodes of fmri functional brain networks. *Network Neuroscience*, 1(3):254–274, 2017.

[22] Lilliana Mason. "i disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1):128–145, 2015.

[23] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[24] Denis Mollison. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3):283–313, 1977.

[25] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[26] Raj Kumar Pan, Mikko Kivelä, Jari Saramäki, Kimmo Kaski, and János Kertész. Using explosive percolation in analysis of real-world networks. *Phys. Rev. E*, 83(4):046112, 2011.

[27] Shai Pilosof, Mason A Porter, Mercedes Pascual, and Sonia Kéfi. The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1(4):1–9, 2017.

[28] Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, 2007.

[29] Sindhuja Ranganathan, Mikko Kivelä, and Juho Kanniainen. Dynamics of investor spanning trees around dot-com bubble. *PloS one*, 13(6):e0198807, 2018.

[30] Daniele Romanini, Sune Lehmann, and Mikko Kivelä. Privacy and uniqueness of neighborhoods in social networks. *arXiv preprint arXiv:2009.09973*, 2020.

[31] Ali Salloum, Ted Hsuan Yun Chen, and Mikko Kivelä. Separating controversy from noise: Comparison and normalization of structural polarization measures. *arXiv preprint arXiv:2101.07009*, 2021.

[32] Ali Salloum, Tuomas Takko, Markus Peuhkuri, Raimo Kantola, Mikko Kivelä, et al. Botit ja informaatiovaikuttaminen twitterissä suomen eduskunta-ja eu-vaaleissa 2019-elebot-hanke. 2019.

[33] Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and János Kertész. Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E*, 75:027105, 2007.

[34] Riitta Toivonen, Lauri Kovanen, Mikko Kivelä, Jukka-Pekka Onnela, Jari Saramäki, and Kimmo Kaski. A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks*, 31(4):240–254, 2009.

[35] Javier Ureña-Carrion, Jari Saramäki, and Mikko Kivelä. Estimating tie strength in social networks using temporal communication data. *EPJ Data Science*, 9(1):37, 2020.

[36] Alessandro Vespignani. Twenty years of network science, 2018.

[37] Yan Xia, Ted Hsuan Yun Chen, and Mikko Kivelä. Spread of tweets in climate discussions. *arXiv:2010.09801*, 2020.