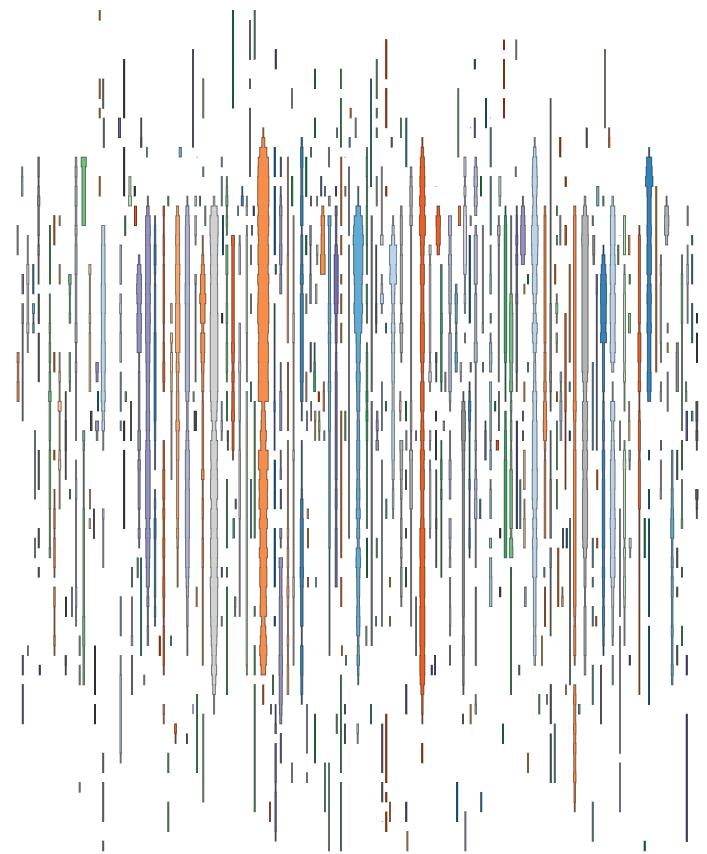


# PhD Thesis

# Complexity in Social Data

Towards mapping and understanding complex phenomena in big social systems, using data science.

**Author:**

Ulf Aslak

Supervisor: David Dreyer Lassen

Advisor: Sune Lehmann

Copenhagen Center for Social Data Science  
Faculty of Social Sciences, University of Copenhagen  
September 2, 2019

## Preface

This thesis was submitted as a final requirement for obtaining the my PhD in Social Data Science. My research was carried out from July 2016 to September 2019 during my employment as a PhD student in the Center for Social Data Science (SODAS) at the University of Copenhagen (UCPH), as well as during a research visit from May to August of 2018 to the Brockmann Lab at the Robert Koch Institute in Berlin. David Dreyer Lassen (Professor, Department of Economics, UCPH) was my main supervisor throughout my PhD employment, and my main scientific advisor was Sune Lehmann (Professor, DTU Compute). I would like to thank both Sune and David for their tremendous support. Throughout my project, I was given ample freedom to pursue my academic interests, visit seminars and conferences abroad, and undertake personally motivating teaching activities. Although I was trained in a different discipline than him, David showed immense trust in my ability to deliver. Likewise, Sune understood and respected my desire to explore different topics, yet kept me on track the whole time. Furthermore, I thank my collaborators for all the energy and time they put into our work together: Sune Lehmann, Martin Rosvall, Søren F. V. Nielsen, Morten Mørup, Dirk Brockmann, Benjamin Maier, Christopher Monk, Robert Arlinghaus, David Kofoed Wind, and Laura Alessandretti. Many thanks also go out to my friends and labmates at SODAS whose company greatly stimulated my development as a scientist. Finally, thanks Vivi and Aslak for giving meaning to it all.

## **Abstract**

It is now possible to accurately measure human behavior and understand it at large scales. Smartphones, social media sites and markets deliver a massive stream of data, that can be tapped into to understand previously unknown social phenomena. One of the things we are discovering is that human social systems are highly complex, displaying many of the hallmarks of complex systems, such as large scale self-organization and reoccurring patterns. At the same time, they are extremely chaotic, making it near impossible to accurately simulate or predict their behavior. Computational social science, or social data science, has therefore emerged as an interdisciplinary field of social scientists turned data scientists and vice versa, with the ambition to answer fundamental questions about human behavior. Operating within this field, this thesis explores complex phenomena in social and behavioral data. It spans a wide range of topics within different fields, from neuroscience to animal ecology, but is connected throughout by the idea of complexity.

## Resume

Det er nu muligt, med stor præcision, at måle menneskelig adfærd og derved forstå den på stor skala. Smartphones, sociale medier og markeder producerer en massiv strøm af data, som til en hvis grad kan indsamles for at forstå førhen ukendte fænomener. En af de ting vi er ved at opdage er at menneskers sociale systemer er svært komplekse, og udtrykker flere af de vigtigste kendetegn ved komplekse systemer, såsom selvorganisering på stor skala og mønstre der gentager sig selv. Samtidig, er disse systemer også ekstremt kaotiske, hvilket gør det nærmest umuligt at simulere eller forudsige deres udvikling med præcision. Beregningsmæssig socialvidenskab, eller social datavidenskab, er således opstået som et interdisciplinært felt af socialvidenskabsfolk som arbejder med datavidenskab og vice versa, med ambitionen om at besvare fundationale spørgsmål omkring menneskelig adfærd. Denne afhandling udforsker komplekse fænomener i social data og adfærdsdata, og opererer altså i dette spænd. Den dækker en bred vifte af emner fra forskellige felter, fra hjernehjerneforskning til økologi, men er knyttet sammen af konceptet kompleksitet.

## Table of Contents

<b>Preface</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Resume</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Patterns in temporal networks</b>	<b>3</b>
Paper: <i>Constrained information flows reveal intermittent communities</i> . . . . .	4
Paper: <i>Temporally intermittent communities in brain fMRI correlation networks</i> . . . . .	14
Paper: <i>Rhythm of relationships in a social fish over the course of a full year in the wild</i> . . .	26
<b>3 Scaling in human mobility</b>	<b>57</b>
Paper: <i>The Scales of Human Mobility</i> . . . . .	58
<b>4 Educational peer review</b>	<b>76</b>
Paper: <i>Quantifying Feedback - Insights Into Peer Assessment Data</i> . . . . .	77
Paper: <i>Optimal allocation of reviewers for peer feedback</i> . . . . .	87
<b>5 Software projects</b>	<b>95</b>
Paper: <i>Netwulf: Interactive visualization of networks in Python</i> . . . . .	96
Paper: <i>Infostop: Python package for detecting stop locations in mobility data</i> . . . . .	99
<b>References</b>	<b>101</b>



*There's no love in a carbon atom, no hurricane in a water molecule, no financial collapse in a dollar bill.*

- Peter Dodds

## 1 Introduction

A fascinating observation can be made in almost every part of nature: when a large collection of components interact on small scales, global structures and behaviors emerge, without instruction, on large scales. This phenomenon is called *complexity* and systems that display it are called *complex systems*. Take, for example, an ant colony. Studied in isolation, interactions between ant and ant, or ant and environment, are rule based and to a great extent predictable. Yet if researchers from an advanced alien civilization were provided a small sample of black garden ants to study their individual and social behavior, they would struggle to predict the astounding structures that self-organize when a million ants colonize. But this very property—that behavior at the interaction scale reveals nothing about behavior at the population scale—is, in fact, a defining trait of complex systems [1].

In many systems understanding complexity has led to some important discoveries about natural phenomena. With modern computing technology it is in some cases possible to simulate large scale behavior from small scale interactions [2], allowing us to reverse engineer the few rules that drive some complex behavior. From agent based simulations, for example, researchers have managed to show that bird flocking, which is so elegantly executed that one may suspect birds were capable of telepathy and thought transfer, is in fact the large scale consequence of each bird trying to stay near the crowd, avoid collision and align with its neighbors [3]. Similarly, locust swarms are now understood to occur because locusts cannibalize on each other after depleting local food stores, forcing each to adopt the same survival strategy that involves fleeing from those approaching and chasing those ahead [4, 5].

Unfortunately, some of the most interesting systems have too many and too complicated interaction rules to make simulation a feasible research approach (yet). Most social systems among animals, and in particular humans, rely on complicated and ever-evolving sets of rules, which can be understood in part but unlikely in their entirety at any point in time. Other tools must, here, be taken into use and entire research fields now focus their attention on understanding societies and social behavior at scale, by applying *data science* methods that have their origins in natural science fields such as physics, applied mathematics and signal processing. This rising interest in *computational social science*—or what we may as well call *social data science* as it draws on ideas and tools from data science to modernize the field of social science—has driven important breakthroughs in understanding our own species. Large scale patterns we never had the slightest understanding of, or did not even know

existed, are now well understood allowing us to make better policies and take effective precautions to avoid now known consequences or mitigate unavoidable ones. In particular, we have refined our understanding of cities [6, 7], migration [8, 9], disease and misinformation spreading [10–13], social group dynamics [14, 15], and even recently critical issues such as how social media advertisement leads to discrimination and unfair distribution of opportunity [16].

These breakthroughs are enabled by a recent explosion in the possible yield of data from devices like smartphones and personal computers, as well as data from online social media platforms, that accurately describe intimate social interactions and individual behaviors. Suddenly, social scientists need not administer surveys to learn about the self-reported social network at an institution; they can simply ask permission to install software on every individual’s smartphone and measure that network directly [17–19]. Similarly, ecologists can tag large animal populations with mobility sensor and learn about general foraging and social patterns that reoccur reliably from the data they automatically collect [20]. And in many other cases, data simply does not need to be collected—it is already there.

This thesis, but more precisely the papers that comprise it, are in a wide sense connected through the concepts of social data science and complex systems. As a natural scientist turned social scientist, I have sought to undertake research in which I could answer interesting questions about human (and animal) behavior using computational skills acquired in my training as a physicist and computer scientist. Whether the subject matter is the social behavior of *Cyprinus carpio* fish or patterns in human brain measurement data, the scientific thinking behind is strongly tied to complexity science and the methods used all belong to the data science toolbox, whether it be network analysis, machine learning, statistics, data visualization, etc.. Throughout, I have kept a strong focus on reproducibility and sought to share much of my code in open source format. Overall, the work presented in thesis contributes to the complexity science and social data science literature on four fronts, each given its own section. In the first section, I present three papers that deepen our understanding of mesoscopic structures in high-resolution temporal networks. In the second section, follows a working paper that describes an important discovery facilitated by human mobility data at a previously unseen scale, which fundamentally disagrees with the current view on human mobility dynamics in the complexity science literature. In the third section, I present two conference articles that analyze the behavior of users in an online educational peer-review platform, and probe the limits of fairness in algorithmic reviewer-reviewee matching. In the fourth and final section, I contribute two open source software packages for visualizing networks and detecting places in mobility trajectories, respectively.

## 2 Patterns in temporal networks

Here follows three papers. The first and second papers are published in the peer-reviewed journals *Physical Review E* and *Applied Network Science*, respectively, while the third is recently submitted to the *Journal of Animal Ecology*.

1. The paper *Constrained information flows in temporal networks reveal intermittent communities* presents a novel method, Neighborhood Flow Coupling (NFC), for detecting intermittent communities in temporal networks. Aside from rigorous testing we show that NFC is extremely well suited for detecting social gatherings in evolving face-to-face networks of humans, and visualize these communities in an interactive format.
2. The paper *Temporally intermittent communities in brain fMRI correlation networks* applies NFC to a temporal fMRI correlation network and discovers two modules that correspond well to so-called *functional networks* known from the literature.
3. In the paper *Rhythm of relationships in a social fish over the course of a full year in the wild*, we analyze the social and individual behavior of a carp population ( $N = 36$ ) in a small lake. The fish carried very high-resolution location sensors and using the resulting data we could reconstruct their social contact network to build an understanding of how seasons and the circadian rhythm modulate their social behavior.

The papers are relatively different in subject matter, but connect through their shared methodology.

## Constrained information flows in temporal networks reveal intermittent communities

Ulf Aslak<sup>\*</sup>

*Centre for Social Data Science, University of Copenhagen, DK-1353 København K, Denmark  
and DTU Compute, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

Martin Rosvall<sup>†</sup>

*Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden*

Sune Lehmann<sup>‡</sup>

*DTU Compute, Technical University of Denmark, DK-2800 Kgs. Lyngby;  
Niels Bohr Institute, University of Copenhagen, DK-2100 København Ø, Denmark;  
and Department of Sociology, University of Copenhagen, DK-1353 København K, Denmark*



(Received 25 November 2017; published 22 June 2018)

Many real-world networks represent dynamic systems with interactions that change over time, often in uncoordinated ways and at irregular intervals. For example, university students connect in intermittent groups that repeatedly form and dissolve based on multiple factors, including their lectures, interests, and friends. Such dynamic systems can be represented as multilayer networks where each layer represents a snapshot of the temporal network. In this representation, it is crucial that the links between layers accurately capture real dependencies between those layers. Often, however, these dependencies are unknown. Therefore, current methods connect layers based on simplistic assumptions that do not capture node-level layer dependencies. For example, connecting every node to itself in other layers with the same weight can wipe out dependencies between intermittent groups, making it difficult or even impossible to identify them. In this paper, we present a principled approach to estimating node-level layer dependencies based on the network structure within each layer. We implement our node-level coupling method in the community detection framework Infomap and demonstrate its performance compared to current methods on synthetic and real temporal networks. We show that our approach more effectively constrains information inside multilayer communities so that Infomap can better recover planted groups in multilayer benchmark networks that represent multiple modes with different groups and better identify intermittent communities in real temporal contact networks. These results suggest that node-level layer coupling can improve the modeling of information spreading in temporal networks and better capture intermittent community structure.

DOI: [10.1103/PhysRevE.97.062312](https://doi.org/10.1103/PhysRevE.97.062312)

### I. INTRODUCTION

Temporal network representations of dynamic complex systems allow researchers to describe changing interaction patterns. Increasingly, high-resolution interaction data require methods that can simplify and highlight important temporal network structures. An important category of such structures is highly interconnected groups of nodes, so-called communities. If the nodes represent individuals who alternate between various roles in social temporal networks, the network's communities will repeatedly form and dissolve at multiple temporal scales in an intermittent way. A simple approach to identify intermittent communities is to first separate a temporal network into a sequence of static snapshots, that is, a multilayer network [1,2], then independently cluster each layer, and finally match the communities across the layers to find the temporal communities [3–8]. Other approaches, including three-way matrix

factorization [9], time-node graphs [10], and stochastic block models [9,11,12], can directly cluster the multilayer network but are also unable to incorporate explicit dependencies between layers. To take into account such interdependencies, some methods cluster multilayer networks using interlayer links that represent specific causal or correlational dependencies between the layers [13–16]. However, explicit interlayer dependencies are often not available to researchers. Moreover, current approaches for estimating such dependencies by, for example, comparing independently inferred community structure between layers [17], using stochastic block modeling [18], or applying link prediction through cross-validation [19], consider only dependencies between entire layers. In contrast, real systems with multiple and asynchronous recurrent events generate dependencies between layers with varying strength within layers. By ignoring such node-level dependencies, current methods wash out important dependencies in multilayer networks with intermittent communities at multiple temporal scales.

In this paper, we present a flow-based method that first couples node pairs in different layers based on the similarity between their network neighborhood flow patterns and

\*ulfaslak@gmail.com

<sup>†</sup>martin@email.edu

<sup>‡</sup>sljo@dtu.dk

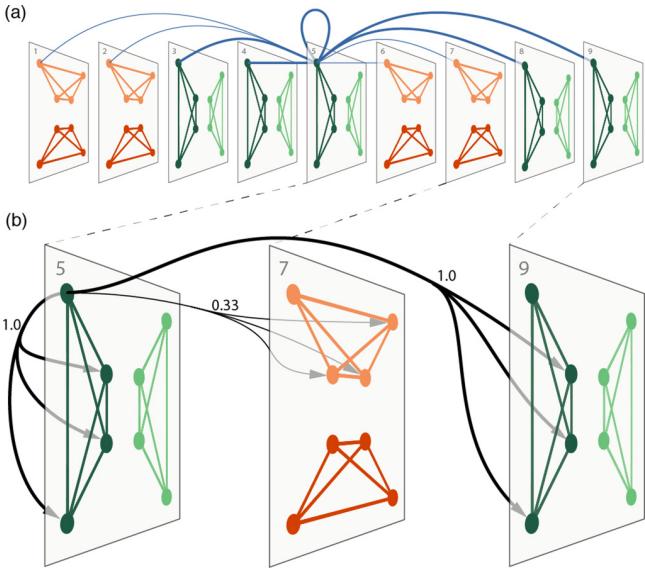


FIG. 1. Neighborhood flow coupling between a state node and its sibling states in a multilayer network. (a) Interlayer coupling,  $D_i^{\alpha\beta}$ , from the top left state node in layer 5 to all state nodes of the same physical node. State nodes with more similar intralayer outlinks couple more strongly, indicated by the stroke width. (b) Interlayer links,  $P_{ij}^{\alpha\beta}$ , are directed and connect a state node to neighbors of state nodes of the same physical node, with weight proportional to the coupling strength and the intralayer link weight of the neighbor node [Eq. (1)].

then—based on the network structure within layers combined with these node-level interlayer dependencies—identifies temporal communities in the resulting multilayer network. For a single node, nonoverlapping neighborhoods are not coupled and identical neighborhoods are maximally coupled. In a social network, this neighborhood flow coupling captures how individuals typically share similar information in similar social contexts. In this sense, neighborhood flow coupling models causal dependencies across time. Finally, we adapt the flow-based community detection algorithm Infomap [20,21] to make use of this information. We demonstrate the usefulness of neighborhood flow coupling for multilayer community detection on benchmark networks. Additionally, we reveal and visualize the temporal evolution of intermittent communities in two temporal human contact networks [22,23]. While our method targets intermittent communities in temporal contact networks represented by multilayer networks, it nevertheless outperforms other methods in standard benchmark tests on multilayer networks.

## II. METHODS

In complex networks, groups of nodes in which flow is contained for a long time provide a useful notion of communities [20,24]. Such communities also can provide straightforward generalizations to multilayer networks [15]. We represent multilayer networks using *physical* nodes and *state* nodes. Physical nodes represent system components, while state nodes, one for each physical node and layer, represent constraints on flows (see Fig. 1). Accordingly, we consider multilayer communities

to be groups of state nodes that capture flows for a significantly long time. In this way, assigning a physical node's state nodes to different communities naturally results in overlapping communities.

In real-world temporal networks, communities often form and dissolve multiple times with a shorter presence than absence [8]. From the perspective of the entire network, these intermittent communities are often asynchronous in the sense that each community forms and dissolves independently in time relative to other communities. Examples of intermittent communities include group voting trends in the U.S. Senate [13], time-dependent sets of correlated financial assets [16], social cores in contact networks [8], and modules of coherently active brain areas [25]. Because nodes in these intermittent communities are not able to share information across their long absent times, since nodes are unlikely to be connected, current methods for identifying communities with long flow-persistence times cannot effectively capture the potential for information transfer. A causal dependency across time requires that interlayer link strengths represent the degree to which information is likely to flow between state nodes in adjacent as well as distant layers. Some existing methods indeed evaluate dependencies between layers, but they do it by coupling entire layers [17–19]. There are two important drawbacks to this approach. First, coupling a physical node's state nodes across all layers generates a large number of links, resulting in computational challenges. For example, in a network with  $n$  nodes in  $t$  layers with average degree  $\langle k \rangle$ , we need  $\langle k \rangle t^2 n$  links in addition to the within-layer links in order to represent connections between state nodes. Second, for large networks with many time slices and intermittent and asynchronous communities, the uniform interlayer links can also “dilute” community boundaries and aggregate distinct communities (we will discuss this point in detail below). To counter the drawbacks of uniform linking, we propose interlayer dependencies at the node level. By forming state-node-specific interlayer links, neighborhood flow coupling generates high-resolution yet sparse multilayer networks that can capture intermittent communities.

### A. Neighborhood flow coupling

The goal of our flow-based approach is to enable interlayer coupling based on the local structural properties of the multilayer network. Each layer's intralayer link structure represents the constraints on network flows at a given time or state of the system. Specifically, we model the network flows in each layer using a random walker that moves from state node to state node guided by the outgoing intralayer links. Because the links represent where flows can move, similar outgoing intralayer link flows in two state nodes of a physical node suggest that the state nodes represent similar states of the physical node. In a social setting, for example, the same group of people may meet again and take up where they left off last time they met. More precisely, the more similar the within-layer flow patterns are, the less the constraints would change and the less information would be lost if the two state nodes were lumped together. We use this information loss measure to couple layers: The less information that is lost if the state nodes were combined, the stronger the interlayer coupling between the state nodes.

Neighborhood flow coupling based on the information loss from merging state nodes is captured by the Jensen-Shannon divergence. In detail, for neighborhood flow coupling between physical node  $i$ 's state nodes in layers  $\alpha$  and  $\beta$ , the state nodes' normalized intralayer outlinks  $\mathbf{P}_i^\alpha$  and  $\mathbf{P}_i^\beta$  give their coupling strength  $D_i^{\alpha\beta}$ ,

$$D_i^{\alpha\beta} = 1 - \text{JSD}(\mathbf{P}_i^\alpha, \mathbf{P}_i^\beta) \quad (1)$$

$$= 1 - H\left(\frac{1}{2}\mathbf{P}_i^\alpha + \frac{1}{2}\mathbf{P}_i^\beta\right) + \frac{1}{2}H(\mathbf{P}_i^\alpha) + \frac{1}{2}H(\mathbf{P}_i^\beta), \quad (2)$$

where  $\text{JSD}(\cdot, \cdot)$  is the Jensen-Shannon divergence and  $H(\cdot)$  is the Shannon entropy. In a multilayer network with neighborhood flow coupling, a random walker moves from state node to state node within a layer guided by the intralayer links and, at rate  $r$ , makes a transition to any layer, including the currently visited layer, proportional to the intralayer link similarity between the state nodes [see Fig. 1(a)]. We include interlayer links to the same layer because they allow for generalizations with complete layer information at rate  $1 - r$  and no layer information when the layer constraints are relaxed at rate  $r$ , as if the layers were aggregated.

Neighborhood flow coupling disregards the temporal ordering of layers. However, for longer timescales or depending on the research question at hand, layer coupling that depends on temporal distance can be implemented. For example, Eq. (1) can be scaled by a factor that depends on the temporal distance between layers.

In any case, intralayer *links* connect state nodes to their neighbors within the same layer, and interlayer *coupling* connects state nodes of the same physical node in different layers. For example, take a random walker at a state node of physical node  $i$  in layer  $\alpha$ ,  $(i, \alpha)$  for short. With probability  $1 - r$  it remains in the same layer and moves to state node  $(j, \alpha)$  with probability proportional to the intralayer link weight  $W_{ij}^\alpha$ . With the remaining probability  $r$  it relaxes the layer constraint, switches to any layer  $\beta$  proportional to the interlayer coupling strength  $D_i^{\alpha\beta}$ , and moves to state node  $(j, \beta)$  proportional to the intralayer link weight  $W_{ij}^\beta$ . Consequently, with intralayer out-strength  $s_i^\beta = \sum_j W_{ij}^\beta$  and interlayer out-strength  $S_i^\alpha = \sum_\beta D_i^{\alpha\beta}$  of state node  $(i, \alpha)$ , the transition probabilities as a function of  $r$  are

$$P_{ij}^{\alpha\beta}(r) = (1 - r)\frac{W_{ij}^\beta}{s_i^\beta}\delta_{\alpha\beta} + r\frac{D_i^{\alpha\beta}}{S_i^\alpha}\frac{W_{ij}^\beta}{s_i^\beta}, \quad (3)$$

where  $\delta_{\alpha\beta}$  is the Kronecker delta function. Therefore, relaxing the layer constraint means that the random walker loses memory of which layer it is currently visiting and instead follows the outgoing links of any state node of the same physical node. With uniform interlayer coupling, relaxing the layer constraint corresponds to a step on the fully aggregated network. However, neighborhood flow coupling takes advantage of higher-order information in the multilayer network that enables longer persistence times in intermittent communities [see Fig. 1(b)].

## B. Neighborhood flow coupling and the map equation

To apply neighborhood flow coupling in the context of community detection, we use the map equation framework for multilayer networks [15,21]. For our purposes, the map equation framework comes with two advantages. First, the map equation is flow-based and directly integrates state-node-specific interlayer flows, as it balances intralayer and interlayer flows by relaxing the intralayer constraints with an interlayer relax rate. Second, the map equation naturally clusters coupled state nodes with similar intralayer links in the same community, as it assigns state nodes of the same physical node and community to the same codeword to capture the fact that they represent the same physical object. Therefore, the flow-based and information-theoretic nature of the map equation is a good fit with neighborhood flow coupling.

In detail, for a two-level modular description of flows from node to node in  $m$  communities, one *index codebook* contains the community-enter codewords and  $m$  *module codebooks* contain the node-visit and community-exit codewords within modules. Each codebook's average codeword length is given by the Shannon entropy of their rates of use,  $\mathcal{Q}$  for enter codewords with total rate of use  $q_\curvearrowright$ , and  $\mathcal{P}_j$  for codewords in community  $j$  with total rate of use  $p_{j\curvearrowright}$ . For node partition  $\mathbf{M}$ , the map equation therefore takes the form

$$L(\mathbf{M}) = q_\curvearrowright H(\mathcal{Q}) + \sum_{j=1}^m p_{j\curvearrowright} H(\mathcal{P}_j). \quad (4)$$

Applied to a possibly weighted and directed network, Infomap searches for the node partition  $\mathbf{M}$  that minimizes the map equation and reveals the most modular regularities in the network flows.

The map equation remains the same for multilayer networks, with one important generalization: when state nodes of the same physical node are assigned to the same community, they are assigned a common code word derived from their total visit rate. This coding scheme captures the very essence of multilayer networks, that all state nodes of the same physical node represent the same physical object [15].

We have implemented the neighborhood flow coupling in the Infomap software package available at [26]. Neighborhood flow coupling is activated with the flag `--multilayer-js-relax-rate`. For memory efficiency or for encoding temporal ordering of layers, interlayer links can be thresholded based on the Jensen-Shannon divergence and temporal distance between layers with the flags `--multilayer-js-relax-limit` and `--multilayer-relax-limit`, respectively.

Neighborhood flow coupling is useful beyond Infomap and can be used with other community detection frameworks, for example, multilayer modularity optimization [13]. In general, the high density of interlayer links between similar layers will make it easier to identify intermittent communities. Moreover, the basic principle of neighborhood flow coupling extends beyond community detection and can be useful for capturing spreading processes in multilayer networks when interlayer coupling information is absent.

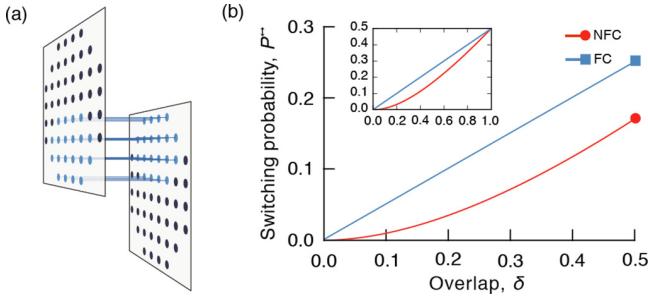


FIG. 2. Coupling strength as a function of overlap. (a) A conceptual illustration of two identical fully connected communities of size  $N = 52$  that reside in adjacent layers and overlap by  $\delta N = 20$  nodes. Overlapping nodes are in blue, and couplings inside each layer are omitted for the purpose of illustration. (b) Analytical layer switching probability for full and neighborhood flow coupling as functions of overlap computed for  $r = 1$  and  $N \gg 1$ . The inset shows the full range from 0 to 1.

### III. RESULTS

We first validate the performance of Infomap with neighborhood flow coupling on benchmark networks with multilayer structure. Then we identify temporal communities in two face-to-face contact networks.

#### A. Performance tests on benchmark networks

We compare neighborhood flow coupling with other interlayer coupling schemes on three types of multilayer benchmark networks to test each method's ability to handle overlapping community structure, recover intermittent communities in increasingly sparse multilayer networks, and retain flows within intermittent communities. We compare neighborhood flow coupling (NFC) with full coupling (FC), adjacent coupling (AC), and no coupling (NC). Full coupling with uniform coupling across layers and no coupling with only the intrinsic coupling from the multilayer coding scheme are extreme cases of neighborhood flow coupling, when the structural similarity in Eq. (1) is either 1 or 0 across all state nodes of the same physical node [15]. Adjacent coupling with uniform coupling strength to the nearest layers is an appealing method for gradually changing communities but cannot capture intermittent communities. These alternative coupling methods provide references to compare and contrast the results of neighborhood flow coupling.

##### 1. Community overlap

In real networks, such as face-to-face networks, communities are rarely completely nonoverlapping but share some members. Therefore, we investigate how neighborhood flow coupling handles overlap compared to full coupling. We begin by considering the simplest possible example: two identical, fully connected communities of size  $N$  that overlap by a fraction  $\delta$  [Fig. 2(a)]. In this network, a random walker traversing the network occupies a node  $i$  inside the overlap with probability  $\delta$  and performs a relax step with probability  $r$ . Consequently, the random walker switches layer with probability

$$P^{\leftrightarrow} = \delta r \frac{D_i^{\alpha\beta}}{D_i^{\alpha\beta} + 1}. \quad (5)$$

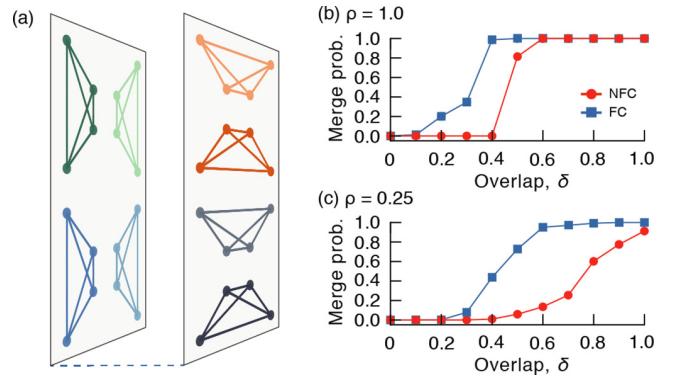


FIG. 3. Full coupling merges communities at lower overlap. (a) Schematic illustration of two layers of cliques with variable overlap. (b) The probability that Infomap with different coupling schemes merges two communities in separate layers as a function of their node overlap. (c) Same as (b) but for sparser networks, in which all but a fraction of  $\rho = 0.25$  links are randomly removed.

A higher  $P^{\leftrightarrow}$  corresponds to stronger coupling between the two communities and increased preference for classifying the two as a single community. For full or adjacent coupling,  $D_i^{\alpha\beta} = 1$  and  $P_{FC}^{\leftrightarrow} = 1/2\delta r$ . For neighborhood flow coupling,  $D_i^{\alpha\beta}$  from (1) for a node in the overlap yields

$$P_{NFC}^{\leftrightarrow} = \delta r \frac{(\delta N - 1)}{(\delta N - 1) + (N - 1)}. \quad (6)$$

That is, the probability of switching layers is the fraction of time steps that a random walker can switch,  $\delta r$ , multiplied by the probability that a relax step will result in a layer switch (which is the number of nodes the walker can reach in the other layer divided by the total number of nodes the walker can reach when inside the overlap). Note that when  $N \gg 1$  the probability of switching layers is only a function of  $\delta$  and  $r$ .

Figure 2(b) shows  $P^{\leftrightarrow}$  as a function of  $\delta$  for full and neighborhood flow coupling at relax rate  $r = 1$  and  $N \gg 1$ . This test shows that the layer switching probability can differ significantly in the important range  $\delta \in [0, 0.5]$ , suggesting that neighborhood flow coupling has a lower tendency to merge overlapping communities compared to both full coupling and adjacent layer coupling.

To compare neighborhood flow coupling and full coupling in a more complex scenario, we measure the threshold of overlap at which the two methods collapse overlapping communities. First, we construct a two-layer network benchmark model with 500 physical nodes partitioned into 50 communities of uniform size 10, where the communities in each layer differ by some number of random edge swaps [Fig. 3(a)]. Then, using Infomap with both coupling schemes on 1000 network realizations, we record the overlap for each pair of communities in different layers and whether or not Infomap merges them [Fig. 3(b)]. We see that full coupling merges communities more aggressively than neighborhood flow coupling, in some cases even when they only overlap by two nodes. Conversely, Infomap with neighborhood flow coupling requires substantial overlap before it merges two distinct communities.

In real-world networks, communities are sometimes sparsely linked internally. Since the neighborhood flow

coupling considers overlap in internal link structure rather than in nodes, partly overlapping communities will merge with lower probability when the communities are sparser. For example, with all but a fraction  $\rho = 0.25$  of the edges randomly removed from each community, the merge probability decreases more for neighborhood flow coupling than full coupling [Fig. 3(c)]. In networks with few layers, the network under study and the research question at hand should determine which coupling method is best.

## 2. Intermittent communities

In networks with many layers, communities may persist over some period of time, then vanish and reemerge again by activating the same subset of nodes with similar within-group link structures. When the goal is to identify such intermittent communities, it is important to avoid inadvertently merging community that are in fact separate. Therefore, we compare how different coupling schemes perform with respect to detecting intermittent communities in increasingly sparse multimode benchmark networks [15].

First, we generate  $T$  independent network layers, which we refer to as *modes*, with the LFR benchmark model [27]. Each mode has 512 nodes, average degree 8, mixing coefficient 0.05, and power-law community-size and degree distributions with exponent 3 (see Appendix C for more details). From each mode we independently sample  $L$  network layers that include links from their mode with probability  $1/L$ . Each multilayer benchmark network thus comprises  $T \times L$  layers, with  $T$  independent sets of  $L$  dependent layers, as schematically illustrated in Fig. 4(a). With increasing  $L$  and sparser communities, the challenge is to detect the communities planted in each mode and distinguish between communities from different modes.

To measure performance, we compute the adjusted mutual information (AMI) [28] between the predicted and true state node labels. We first show that neighborhood flow coupling is less sensitive to variations in the relax rate [Fig. 4(b)]. The no coupling method is independent of the relax rate and serves as a performance baseline. For adjacent coupling, the performance increases with the relax rate because this coupling takes advantage of the ordered layers and completes information in sparse layers. However, when shuffling the layers, this advantage vanishes and the performance drops significantly. Full coupling has a narrow performance optimum, and the performance drops to zero around  $r = 0.7$  when the strong interlayer coupling causes Infomap to label the whole network as one community. Neighborhood flow coupling is more stable and performs best with a relax rate between 0.15 and 0.7 for this type of multilayer network. If not stated otherwise, we use  $r = 0.25$  for all analyses.

For all types of coupling, performance depends on the number of network modes. On single-mode multilayer networks, full coupling achieves the highest score because uniform interlayer coupling maximally aggregates the dependent layers [Fig. 4(d)]. When the number of samples per mode increases, the networks become sparser and the probability of finding high-similarity neighborhood flows decreases. As a result, neighborhood flow coupling converges to no coupling. However, neighborhood flow coupling handles many samples per mode and multiple modes better than any other coupling

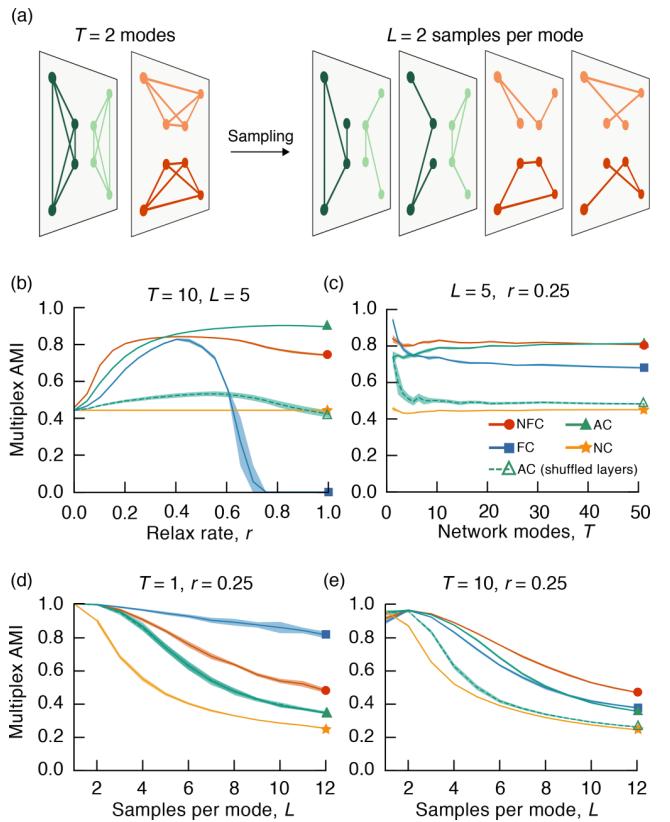


FIG. 4. Neighborhood flow coupling captures intermittent overlapping communities in sparse multilayer networks. (a) Illustration of benchmark network for measuring performance in sparse networks with overlapping intermittent communities in  $T$  network modes and  $L$  sampled layers per mode. (b) Performance of all coupling methods measured by the AMI between the recovered and true partitions as a function of the relax rate  $r$ . Per definition, no coupling does not depend on the relax rate. (c) Performance test for  $L = 5$  layers per mode and increasing number of network modes  $T$ . (d, e) Performance test for fixed number of network modes  $T = 1$  and  $T = 10$ , respectively, and increasing number of increasingly sparse sampled layers  $L$  per network mode. For more than one network mode, neighborhood flow coupling stands out as the best coupling method.

scheme. For example, in the 10-mode multilayer networks with overlapping communities, neighborhood flow coupling performs better than full coupling also for many samples per mode [Fig. 4(e)]. In this case, both coupling schemes perform better for a few samples per mode than only one, because they force interlayer links between spuriously overlapping layers. This behavior suggests that an adaptive relax rate based on the absolute similarity between layers may give even better results. Nevertheless, the performance of neighborhood flow coupling remains stable for much higher numbers of network modes [Fig. 4(c)]. While adjacent coupling performs on par with neighborhood flow coupling in this scenario, its performance relies on the layer order. When we shuffle the layers, adjacent coupling can no longer benefit from similarities between adjacent layers and performs as bad as no coupling. This result highlights that adjacent coupling cannot detect communities with temporal interruptions. Overall, while there is room for

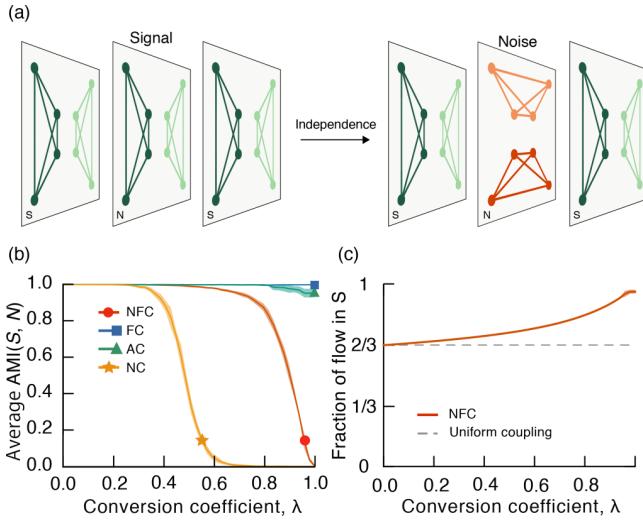


FIG. 5. Neighborhood flow coupling amplifies flow persistence in communities. (a) Schematic benchmark network for testing network flow persistence in planted communities. Two identical signal layers sandwich a noise layer with varying overlap, with the signal layers set by conversion coefficient  $\lambda$ , from identical at  $\lambda = 0$  (left) to independent at  $\lambda = 1$  (right). (b) The average AMI between the identified partitions in the signal and noise layers decreases as they disentangle. (c) The fraction of total network flows captured by nodes in the signal layers increases as the signal and noise layers disentangle for neighborhood flow coupling but is constant at two-thirds of the total flow for the uniform coupling methods. We use relax rate  $r = 1$  for neighborhood flow coupling to maximize interlayer flows and emphasize the effect.

further improvement, neighborhood flow coupling stands out as the best method for detecting intermittent communities.

### 3. Flow persistence

We have developed the neighborhood flow coupling to constrain flows within structurally similar overlapping regions of a network. To explore this feature, we use a multilayer benchmark network model consisting of two identical signal layers with known clusterings at both sides of a noise layer that, to a tunable degree, is more or less independent of the signal layers [Fig. 5(a)]. We generate layers with the same LFR benchmark model as in the previous section. We introduce a tuning parameter  $\lambda \in [0, 1]$  such that the noise layer contains  $n_e(1 - \lambda)$  randomly selected edges from the signal network and likewise  $n_e\lambda$  from another network generated independently following the same procedure. By tuning  $\lambda$  from 0 to 1, we can gradually convert the noise layer from the signal network copy to an independent network [29]. We can now test how well different coupling methods handle interference from the noise layer by measuring the decrease in average adjusted mutual information between the identified signal and noise layer partitions as we increase  $\lambda$ . To emphasize the effects, we use relax rate  $r = 1$ .

Neighborhood flow coupling and no coupling are robust to interference from irrelevant layers. At some level of conversion, noise and signal layer communities should be considered independent of each other, and the AMI between signal and

noise layers should go to zero. No-coupling gives independent labels to the noise layer after 60% conversion, and neighborhood flow coupling gives independent labels after 100% conversion. Full and adjacent coupling suffer from interference with the noise layer even when it is fully converted and thus independent of the signal layers [Fig. 5(b)]. The strong coupling between signal and noise layers for these methods induces interlayer flows in spurious communities. Obviously, the no coupling method is immune to such interference and therefore is unable to pick up actual interlayer coupling in intermittent communities (Fig. 4). In contrast, neighborhood flow coupling is able to both avoid interference from irrelevant structures and pick up information from intermittent communities.

Neighborhood flow coupling can retain flows in intermittent communities. The proportion of flow inside the signal layers explains why neighborhood flow coupling outperforms full and adjacent coupling. In the three-layer example, for any uniform coupling scheme—be it full, adjacent, or no coupling—each layer carries one-third of the total flow, independent of  $\lambda$ . Therefore, two-thirds of the total flow in the signal layers forms a baseline. For neighborhood flow coupling, however, this fraction increases as  $\lambda$  approaches 1 and the signal and noise layers disentangle [Fig. 5(c)]. The adaptive coupling reinforces flows inside the two signal layers together and prevents flows from leaking to the noise layer. As a result, neighborhood flow coupling accentuates structures with long flow persistence times across layers and makes it possible to detect intermittent communities in multilayer networks.

### B. Understanding real-world temporal contact networks

We now apply multiplex Infomap using neighborhood flow coupling, full coupling, adjacent coupling, and no coupling schemes to two empirical temporal contact networks. We represent each data set as a multilayer network and aggregate links over 10-minute intervals in each layer. The first network represents contact events during working hours (approximately 8 a.m. to 6 p.m.) between employees in a workplace environment over two weeks [22]. In this network there are  $n = 92$  physical nodes,  $e = 2.91 \times 10^3$  intralayer links, and  $t = 575$  nonempty layers, and the average intralayer node degree is  $\langle k \rangle = 0.110$ . The second network arises from Bluetooth signal connections between personal smartphones of freshmen university students, also over two weeks [23] ( $n = 636$ ,  $e = 1.27 \times 10^5$ ,  $t = 600$ ,  $\langle k \rangle = 0.665$ ). In the university dataset, links are tracked during a special study period when each student attends the same course every day. The students may meet anytime during the 24 h of the day, but to simplify the comparison to the workplace network, we consider only links that occur during working hours (8 a.m.–4 p.m.). Thus, both networks are cropped to this timeframe, so  $t = 480$ . We start by analyzing the interlayer link structure that neighborhood flow coupling produces. In particular, we are interested in understanding the sparsity of the representations that the method creates, compared to other methods. We then evaluate the performance of Infomap resulting from each coupling scheme by measuring overlap, size, and self-similarity over the time of communities that each method finds. There is no ground truth to measure performance against, so we focus our analysis on showing that neighborhood flow coupling strikes a balance

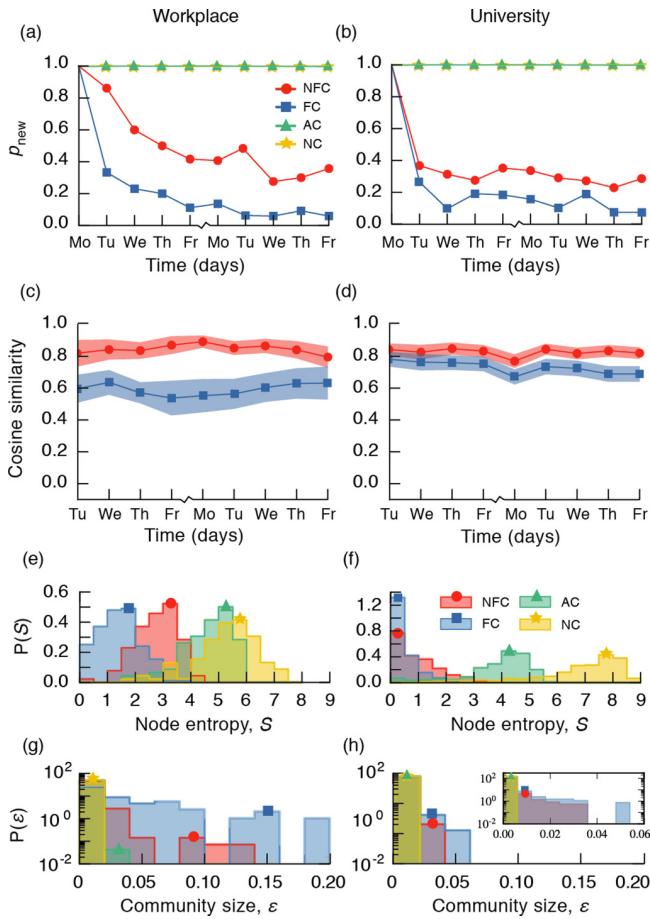


FIG. 6. Properties of communities in real-world networks. Each column corresponds to a separate dataset. (a), (b) The number of new communities on each day,  $p_{\text{new}}$ , in the solutions for each coupling scheme. (c), (d) Average community self-similarity over time. (e), (f) Distribution of node community assignment entropy,  $S$ , as a measure of community overlap. (g), (h) Community size distributions.

between allowing information to flow between all layers—the strength of full coupling—and not mixing unrelated contexts—the strength of no coupling. Finally, we explore each network by visualizing the neighborhood flow coupling community detection solution.

### 1. Neighborhood flow coupling finds communities that are highly self-similar

We know from the literature that the networks under study contain intermittent communities [8,30]. Therefore these networks are useful in order to better understand each different coupling method's ability to capture intermittent community structure. Due to the frequent daily reemerging of communities, a method that couples temporally distant layers should cause the rate of new communities discovered on each day,  $p_{\text{new}}$ , to decline over time. This is indeed the case for full coupling and neighborhood flow coupling [Figs. 6(a)–6(b)]. Full coupling drives  $p_{\text{new}}$  close to zero, which is unrealistic as we should expect some degree of exploration to take place. The reason for this behavior is likely the fact that new communities are merged with previous, slightly overlapping communities.

For neighborhood flow coupling, intermittent communities are appropriately recognized each day, while a significant fraction of new configurations is given new labels.

Knowing that communities are indeed successfully rediscovered each day, we now seek to understand how self-similar intermittent communities are between days of (re)discovery. A good detection algorithm should partition the network such that each reappearance of a community is highly self-similar to its other appearances. We measure the similarity between each temporal community to itself on the most recent previous day as the cosine similarity between the unnormalized 24 h aggregate distributions of member nodes, and plot the similarity distribution over time as their means inside the 95% confidence intervals. It is only relevant to measure self-similarity for full and neighborhood flow coupling, since only those two methods are able to capture reoccurring communities. In both networks, neighborhood flow coupling results in, on average, higher community self-similarity than full coupling does [Figs. 6(c) and 6(d)]. This difference is more pronounced in the university network because the structures are larger. In the case of full coupling, large communities are frequently split into smaller ones that are rarely detected.

### 2. Full coupling solutions tend to merge overlapping communities

We measure the distribution of node entropy,  $P(S)$ , and the distribution of community size,  $P(\epsilon)$ , in each network. We compute the node entropy as  $S = \sum_i c_i \log c_i$  where  $c_i$  is the distribution over time spent in community  $i$  for a given node. Intuitively, if the average node entropy is high, nodes are detected as frequently being in different communities, meaning that communities must overlap on many nodes. Full coupling results in low node entropy and large communities [see Figs. 6(c)–6(e)]. In conjunction with our previous observation that full coupling leads to unrealistically low values of  $p_{\text{new}}$ , this is a strong indication that it causes Infomap to merge communities that overlap in different layers.

For both networks, the  $p_{\text{new}}$  curve for neighborhood flow coupling is similar to full coupling but with more new communities emerging each day. In the workplace dataset, we note that there are almost as many new communities discovered on the second day as there are on the first. We can explain this with the observation that the workplace dataset contains groups that are scheduled to meet every other day, and, as such, we should expect some of those to start on the second day. While neighborhood flow coupling captures this nuance, full coupling does not. In the university network, neighborhood flow coupling identifies fewer and fewer communities as the week progresses, with the exception of Fridays, where relatively more new communities form. This nuance is not captured by full coupling. These results further support the concept that full coupling results in mergers of overlapping communities due to interlayer links that connect them via the nodes they overlap on.

### 3. Visualization of temporal communities

We visualize the temporal expansion and contraction of communities found by Infomap with neighborhood flow layer coupling in each network, as horizontal “strips” of varying height [31]. Figure 7 displays a subset of the communities

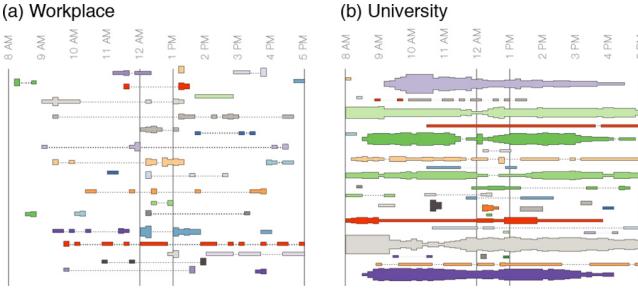


FIG. 7. Temporal communities detected by Infomap with neighborhood flow coupling. Each horizontal track represents a community and its varying height represents the number of active nodes over time. (a) Partition of the workplace network. Height to scale with (b). (b) Partition of the university network. At its tallest point, the largest community (top purple, 10 am) has 22 active members.

discovered in each network (vertical position is arbitrary). There are clear differences between the community structure of the two networks. The university network gives rise to large structures that persist over long periods of time, while the workplace communities are significantly more intermittent, lasting on the order of tens of minutes. Community sizes agree with our insight from Figs. 6(g) and 6(h). In the university network, some are large, corresponding to students attending lectures, some are midsized, corresponding to work-groups and small lectures, and some are small, corresponding to two-to-four-person gatherings. In the workplace network, communities mostly consist of a few people and occasionally are larger around lunch, but never on a scale similar to the university, as we should expect. We provide an interactive version of Fig. 7 [32], which offers further intuition about these networks and the effects of neighborhood flow coupling to the observed structure. With these levels of intermittent communities—here observed in particular for the workplace network but also strongly present in the university network at a daily rate—it is clear that neighborhood flow coupling is a good choice for estimating layer interdependency.

#### IV. CONCLUSION

Our experiments suggest that connecting state nodes across layers in multilayer networks based on the similarity between their network neighborhood flows has multiple benefits over uniform entire-layer coupling approaches. For example, in series of time-stamped face-to-face interaction events represented as multilayer networks, neighborhood flow coupling captures natural constraints on information flows such that flows move freely only within and between similar communities across layers. As a result, Infomap is able to identify intermittent communities with long flow persistence times and recognize spuriously overlapping communities as separate entities. In contrast, existing uniform entire-layer approaches either fail to capture whole communities that are intermittent across temporal layers or collapse spuriously overlapping communities into single communities. Furthermore, we demonstrate that neighborhood flow coupling results in multilayer network representations that are orders of magnitudes sparser in typical real-world networks with corresponding

computational gains. This computational gain allows us to analyze and identify intermittent communities in temporal networks over longer times or higher resolution. Consequently, neighborhood flow coupling opens new avenues for temporal network analysis.

#### ACKNOWLEDGMENTS

We thank Chris Blöcker and Ludvig Bohlin for comments that improved our manuscript. M.R. was supported by the Swedish Research Council, Grant No. 2016-00796. S.L. was funded by the Danish Council for Independent Research, Grant No. 4184-00556a, and the Villum Foundation “High Resolution Networks.”

#### APPENDIX A: INTERLAYER SPARSITY

To couple state nodes, neighborhood flow coupling requires that the network structure around state nodes be similar. This similarity is not required by full coupling, which couples the state nodes of each physical node regardless of local structure. In temporal networks, there are often many state node pairs that have no structural similarity, because they can participate in nonoverlapping communities at different times. With neighborhood flow coupling, this results in network representations with sparse interlayer link structure compared to full coupling, where the interlayer network is always dense. In this Appendix, we investigate the degree to which neighborhood flow coupling reduces the size of the network and thus the memory footprint. We measure the density reduction in relation to the full coupling density, which is always one, and compare to the adjacent coupling density. Furthermore, we analyze how density varies with layer interdependence for neighborhood flow coupling, as we reason that this must be an important factor.

First, we consider sparsity in synthetic networks with independent layers. We define interlayer density,  $S$ , as the ratio of realized to possible interlayer links. Per definition, it is always the case that  $S_{FC} = 1$  and  $S_{NC} = 0$ . If layers are independent, we can derive  $S_{AC} = 2/t$  by dividing the expected number of links from adjacent coupling with the expected number of links from full coupling. For  $t = 600$  (corresponding to two weeks of working hours in 10-minute time bins),  $S_{AC} = 3.3 \times 10^{-3}$ .  $S_{NFC}$  can be approximated as the probability that two state nodes have at least one link in common:

$$S_{NFC}(\langle k \rangle, n) = 1 - P\left(0, \frac{\langle k \rangle^2}{n}\right) = 1 - e^{-(\langle k \rangle^2/n)}, \quad (A1)$$

where  $P(0, \theta)$  is the function value in 0 of a Poisson distribution with average  $\theta$  equal to the expected number of shared links between two state nodes in independent layers  $\langle k \rangle^2/n$ . For a network with similar statistics to the university network ( $n = 636$ ,  $\langle k \rangle = 0.665$ ), Eq. (A1) gives  $S_{NFC} = 7.0 \times 10^{-4}$ . In real temporal networks, however, we observe that the interlayer link structure is more dense because there is significant dependence between layers. The estimations presented here therefore serve only as random network baselines that we can compare with. For the university network, where the possible number of interlayer links is  $7.11 \times 10^7$ , we observe  $S_{NFC} = 0.680$  and  $S_{AC} = 0.005$ , and for the workplace network, where the

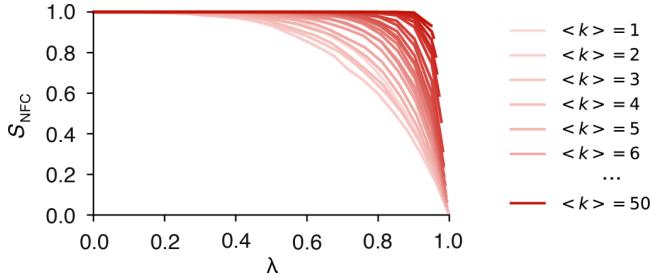


FIG. 8. Interlayer sparsity depends on layer interdependence. The density of interlayer links  $S_{NFC}$  created by neighborhood flow coupling decreases as the layers in a multilayer network become more independent (increasing  $\lambda$ ). The rate at which density decreases depends on the average degree of state nodes  $\langle k \rangle$ . Sparse intralayer link structure, corresponding to low  $\langle k \rangle$ , leads to sparser interlayer link structure.

possible number of interlayer links is  $4.33 \times 10^5$ ,  $S_{NFC} = 0.299$  and  $S_{AC} = 0.012$ . The large increase in  $S$  that we observe for the empirical networks reveals that neighborhood flow coupling is very sensitive to interdependence between layers.

We now test how sensitive interlayer sparsity resulting from neighborhood flow coupling is to layer interdependence, using a simple experiment similar to the approach taken in Sec. III A 3. We create an Erdős-Rényi graph with  $n = 1000$  and variable  $\langle k \rangle$ . We create a two-layer network where both layers are copies of this network, such that the layer independence  $\lambda$ , which we measure as the average Jensen-Shannon divergence across all pairs of state nodes, is zero. We then gradually convert the second layer to an independent network, generated by the same process, using edge swaps, while measuring  $S_{NFC}$  versus  $\lambda$ . When the second layer is fully converted, the two layers are maximally independent and  $\lambda = 1$ . The experiment shows, first, that the relationship between  $S_{NFC}$  and  $\lambda$  is nonlinear. Second, we observe that a sparse intralayer structure (low  $\langle k \rangle$ ) leads to a sparser interlayer link structure, increasingly so when layers are independent (Fig. 8).

Thus neighborhood flow coupling offers significant gains in memory efficiency relative to full coupling, particularly in sparse multilayer networks.

## APPENDIX B: ROBUSTNESS TO RELAX RATE

In the absence of an adaptive relax rate, the problem at hand should decide what relax rate  $r$  to use. In general, for full coupling,  $r$  must be large enough to facilitate flow between layers yet small enough to contain information inside the layer communities. For neighborhood flow coupling, this heuristic does not apply, because interlayer links are established only between structurally similar regions of the network. At the same time,  $r$  still controls the amount of interlayer flow in the network. If  $r = 0$ , information cannot flow between layers, and if  $r = 1$ , important layer information may be diluted.

The optimal relax rate  $r$  should allow Infomap to discover communities that repeat in different layers. To test this criterion, we perform a simple experiment that starts with a multilayer network, selects a random layer, and appends a copy of it to the end of the network. For a range of  $r$  values, we then measure the proportion of nodes in the copied layer to which

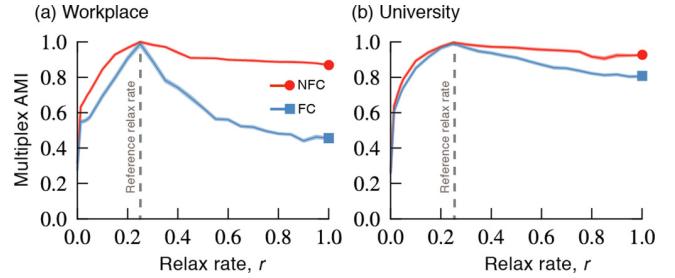


FIG. 9. Neighborhood flow coupling is highly robust to variations in  $r$ . The plots illustrate the similarity of a community detection solution using  $r = 0.25$  with solutions obtained from different relax rates. The high values for neighborhood flow coupling in both networks (a) and (b) demonstrate its high robustness to  $r$  variability compared to full coupling.

Infomap assigns the same label as in the original layer. We perform this test on the university and the workplace networks for neighborhood flow and full coupling, and find that both coupling schemes give perfect labeling of all copied nodes for all values of  $r$  except  $r = 0$ . While this result does not reveal a performance optimum for  $r$ , it shows that the map equation can effectively capture layer interdependences.

The results should not be sensitive to the exact choice of the relax rate. We demonstrate the robustness by clustering a network for a range of relax rates and comparing each solution to the solution for  $r = 0.25$ , with the multiplex AMI as a performance measure. If robustness is high, all solutions should have a high AMI with this reference solution. Performing this test for both networks, we find that neighborhood flow coupling solutions are significantly more robust to varying  $r$  than full coupling solutions. Neighborhood flow coupling is particularly

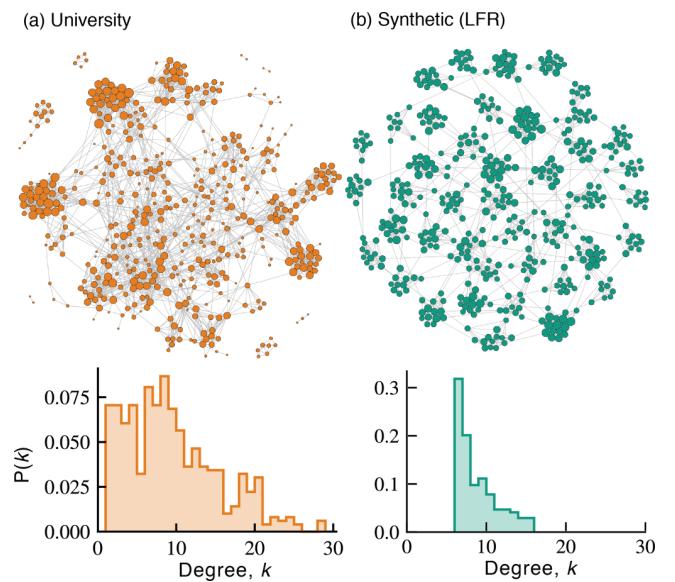


FIG. 10. Real and synthetic networks. (a) Simple network of aggregated links between 8 a.m. and 4 p.m. in the university dataset. The number of nodes is 496 and the mean degree is 9.0. (b) Synthetic LFR network realization with 512 nodes, average degree 8.4, and fitted degree distribution power law exponent 3.8.

robust in the domain  $r > 0.25$ . The similarity decays faster when  $r < 0.25$  and goes to zero for  $r = 0$ , which demonstrates that, while robust to  $r$ , Infomap with neighborhood flow coupling allows for detecting smaller communities. In summary, a broad spectrum of relax rates gives similar solutions for Infomap with neighborhood flow coupling (Fig. 9).

### APPENDIX C: BENCHMARK NETWORKS

For transferring results from benchmark networks to real networks, the benchmark networks should resemble the real networks. However, real networks come in a great variety, and benchmark networks cannot accurately mimic all of them. To find meaningful model parameters for the LFR benchmark networks [27], we consider individual workdays of the university network as aggregated simple graphs (interactions between 8

a.m. and 4 p.m.) and observe that the number of nodes typically lies between 400 and 500 and that the mean degree is in the range 6–12. Figure 10(a) shows an example of one such real network and its degree distribution. We generate synthetic networks with the LFR implementation made available online [33], with input parameters  $N = 512$  (number of nodes),  $k = 8$  (average degree),  $\text{maxk} = 16$  (maximum degree),  $\mu = 0.05$  (mixing),  $t_1 = t_2 = 3$  (degree and community-size power-law distribution exponent), and  $\text{maxc} = 24$  (maximum community size). While there is no guarantee that the statistics of individual resulting networks fully respect the input parameters, we observe that realized degree and power law exponents deviate only marginally (standard deviations 0.1 and 0.8, respectively). Figure 10(b) shows an example of a synthetic network that results from these parameters.

- 
- [1] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *J. Complex Netw.* **2**, 203 (2014).
  - [2] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, *Phys. Rev. X* **3**, 041022 (2013).
  - [3] G. Palla, A.-L. Barabási, and T. Vicsek, *Nature (London)* **446**, 664 (2007).
  - [4] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe, in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, San Jose, CA, USA, 2007), pp. 717–726.
  - [5] A.-K. Pietiläinen and C. Diot, in *Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing* (ACM, Hilton Head, SC, USA, 2012), pp. 165–174.
  - [6] J. Kauffman, A. Kittas, L. Bennett, and S. Tsoka, *PLoS ONE* **9**, e101357 (2014).
  - [7] J. He and D. Chen, *Physica A* **429**, 87 (2015).
  - [8] V. Sekara, A. Stopczynski, and S. Lehmann, *Proc. Natl. Acad. Sci. USA* **113**, 9977 (2016).
  - [9] L. Gauvin, A. Panisson, and C. Cattuto, *PLoS ONE* **9**, e86028 (2014).
  - [10] L. Speidel, T. Takaguchi, and N. Masuda, *Eur. Phys. J. B* **88**, 203 (2015).
  - [11] T. P. Peixoto, *Phys. Rev. E* **92**, 042807 (2015).
  - [12] C. Matias and V. Miele, *J. R. Stat. Soc. B* **79**, 1119 (2016).
  - [13] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Science* **328**, 876 (2010).
  - [14] Y. Chen, V. Kawadia, and R. Urgaonkar, *arXiv:1303.7226* (2013).
  - [15] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, *Phys. Rev. X* **5**, 011027 (2015).
  - [16] M. Bazzi, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison, *Multiscale Model. Simul.* **14**, 1 (2016).
  - [17] D. B. Larremore, A. Clauset, and C. O. Buckee, *PLoS Comput. Biol.* **9**, e1003268 (2013).
  - [18] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, *IEEE Trans. Netw. Sci. Eng.* **3**, 95 (2016).
  - [19] C. De Bacco, E. A. Power, D. B. Larremore, and C. Moore, *Phys. Rev. E* **95**, 042317 (2017).
  - [20] M. Rosvall and C. T. Bergstrom, *Proc. Natl. Acad. Sci. USA* **105**, 1118 (2008).
  - [21] M. Rosvall, D. Axelsson, and C. T. Bergstrom, *Eur. Phys. J.: Spec. Top.* **178**, 13 (2009).
  - [22] M. Génois, C. L. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, and A. Barrat, *Netw. Sci.* **3**, 326 (2015).
  - [23] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann, *PLoS ONE* **9**, e95978 (2014). University contact data in this paper (the Copenhagen Networks Study) was collected with the permission of the Danish Data Protection Agency (Journal number 2012-41-0664).
  - [24] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, *Proc. Natl. Acad. Sci. USA* **107**, 12755 (2010).
  - [25] U. Braun, A. Schäfer, H. Walter, S. Erk, N. Romanczuk-Seiferth, L. Haddad, J. I. Schweiger, O. Grimm, A. Heinz, H. Tost *et al.*, *Proc. Natl. Acad. Sci. USA* **112**, 11678 (2015).
  - [26] Available at [www.mapequation.org](http://www.mapequation.org).
  - [27] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Phys. Rev. E* **78**, 046110 (2008).
  - [28] N. X. Vinh, J. Epps, and J. Bailey, *J. Machine Learn. Res.* **11**, 2837 (2010).
  - [29] R. Aldecoa and I. Marín, *Sci. Rep.* **3**, 2216 (2013).
  - [30] J. Stehlé, N. Voisin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina *et al.*, *PLoS ONE* **6**, e23176 (2011).
  - [31] U. Aslak, How people gather: An interactive visualization approach, <http://www.sciencemag.org/projects/data-stories/winners> (2016).
  - [32] [http://ulfaslak.com/research/temporal\\_communities/](http://ulfaslak.com/research/temporal_communities/).
  - [33] A. Lancichinetti, S. Fortunato, and F. Radicchi, LFR software package, <https://sites.google.com/site/andrealancichinetti> (2014).

RESEARCH

Open Access

# Temporally intermittent communities in brain fMRI correlation networks



Ulf Aslak<sup>1,2\*</sup> Søren F. V. Nielsen<sup>2,4,5</sup>, Morten Mørup<sup>2</sup> and Sune Lehmann<sup>1,2,3</sup>

\*Correspondence:  
[ulfaslak@gmail.com](mailto:ulfaslak@gmail.com)

<sup>1</sup>Centre for Social Data Science,  
University of Copenhagen, 1353  
Copenhagen, Denmark

<sup>2</sup>DTU Compute, Technical  
University of Denmark, 2800 Kgs.  
Lyngby, Denmark

Full list of author information is  
available at the end of the article

## Abstract

It is widely agreed that the human brain is organized as a system of segregated modules that reside in separate regions and, through coordinated integration, support different cognitive functions. Through recent breakthroughs in modeling the activity of the brain, it has been demonstrated that each such module can participate in multiple so-called *functional networks* – networks of brain regions that activate in synchrony during specific types of cognition. If we model the brain as a temporal network, by representing brain regions as nodes and correlations within activity-windows at different times as links, we can formulate the task of finding functional networks as a community detection problem. In spite of this, however, relatively little attention has been given to solving this problem using recently developed techniques for temporal community detection. In this paper, as a proof-of-concept, we apply a novel technique for community detection in temporal networks to a dataset of fMRI measurements from 100 healthy subjects undertaking a working memory task with intermittent fixation (or resting-state) periods. We show that this method recovers two distinct communities that are shared between subjects: one that activates during the fixation period, and another that activates during a period associated with high cognitive load.

**Keywords:** Community detection, Temporal networks, fMRI, Network neuroscience

## Introduction

Global brain activity exhibits a surprising level of organization in both space and time. The consilience of evidence from studies using various imaging technologies and computational methods establishes the existence of so-called “functional networks”: sets of brain regions that activate in synchrony and support a vast repertoire of brain functions (Gusnard and Raichle 2001; Fox et al. 2005; Park and Friston 2013). There are many well studied methods for detecting these functional networks, involving non-invasive imaging techniques such as functional magnetic resonance imaging (fMRI), electro- and magnetoencephalography (EEG and MEG). Central to these methods is the inference of synchrony between different brain regions, commonly gauged by measuring mutual information or correlation strength between time-series representing activity in separate regions. The simplest approach is to query the similarity between regions-of-interest (ROI) (Biswal et al. 1995; Lowe et al. 1998; Greicius et al. 2003; Fox et al. 2005), which are typically parcels of a predetermined segmentation of the brain into distinct regions, called a *parcellation* or an *atlas* (Jenkinson et al. 2012; Brodmann 1909; Fan et al. 2016).

A key feature of ROI-based methods is that they require a predefined parcellation of the brain as input. There are multiple limitations due to this strategy: (1) Predefined parcellations may not fit individual subject data well due to processing characteristics and population mismatch. (2) There currently exists many different, but mutually inconsistent, atlases (Bohland et al. 2009). (3) A researcher must make a design decision about which parcellation to use, a choice which can severely impact the results. To remedy these limitations, various unsupervised parcellation methods exist, based on – but not limited to – mixture models,  $k$ -means clustering, hierarchical clustering, spectral clustering, principal component analysis (PCA) and independent component analysis (ICA) (Thirion et al. 2014).

We employ ICA to obtain an unsupervised representation of the brains functional activity across a set of components. ICA is often used as a dimensionality reduction method before dynamic functional connectivity analysis (Allen et al. 2014; Vidaurre et al. 2017), investigating repeating patterns of time-resolved connectivity between independent components. Furthermore, ICA has previously been used to recover spatial functional networks and is used in many applications (Kiviniemi et al. 2003; McKeown et al. 1998; Beckmann et al. 2005), where some attention is given to identifying temporally fluctuating potentially overlapping functional networks (Smith et al. 2012). The domain of network neuroscience has exploited techniques from network science to uncover structure in connectomics data (Bassett and Sporns 2017). In the context of dynamical functional connectivity analysis of fMRI data, the multiplex modularity framework of Ref. (Mucha et al. 2010) has previously been used to characterize modular structure in ROI based time-resolved dynamic functional connectivity (Bassett et al. 2011; Bassett et al. 2013; Khambhati et al. 2018). However, an important limitation of this method – and to our knowledge all other methods for community detection in temporal networks – is the assumption that dependencies between individual nodes across time can be modeled as the dependence between entire layers. Real systems with multiple asynchronous concurrent events must have varying dependencies between the same layers, and by ignoring these node-level dependencies important temporal dynamics are washed out.

In this paper, we apply the novel *neighborhood flow coupling* (NFC) method (Aslak et al. 2018) for temporal community detection to a temporal correlation network of independent components (ICs) inferred with ICA on fMRI data (Barch et al. 2013). The temporal network is represented as a sequence of static networks, or *layers*, that share nodes, but have varying links depending on the inter-IC correlations that the links represent. NFC, in contrast to other methods, inter-couples layers in this temporal network based on link structure at the node-level. Specifically, the time-states of each node are coupled proportional to the similarity of their neighborhoods. This creates a large connected network which we can partition into modules, or *communities*, using any community detection algorithm. Here, we use Infomap (Edler and Rosvall 2014; Rosvall et al. 2009). This framework enables us to capture both the communities, which may correspond to functional networks in the brain, and their very fine-grained temporal evolution.

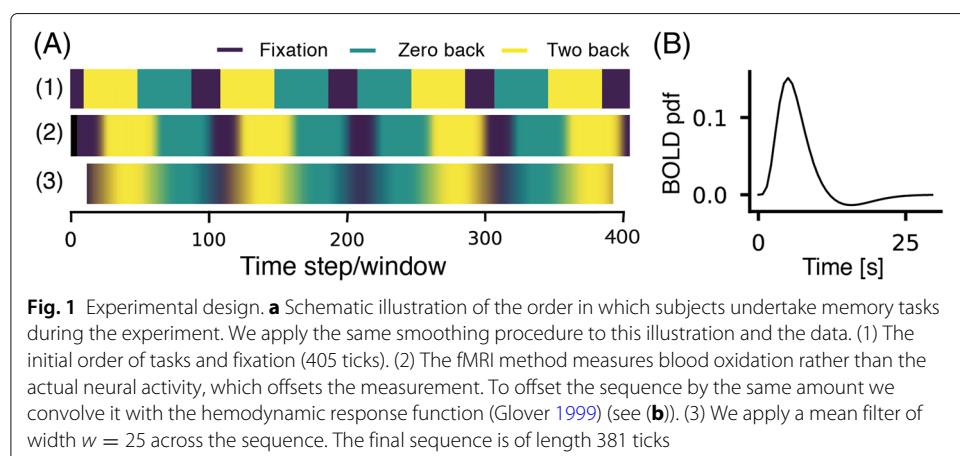
## Results

We use a dataset from the Human Connectome Project (HCP) (Barch et al. 2013), with fMRI measurements from 100 subjects undertaking an  $n$ -back working memory task.

The task consists of three experimental blocks, *2-back*, *0-back* and *fixation*, with varying cognitive load. In the *2-back* condition, subjects are presented with images and are asked to press a button if it matches the image presented two instances back. In the (easier) *0-back* condition, they are asked to press a button when a target image is presented. During the fixation periods, no instructions are given and subjects lie passively with eyes open in a resting-state. For each subject and experimental block, we can measure a performance accuracy as the fraction of correct button presses. The duration of an experiment is just under 5 min, and measurements are taken in 405 ticks with frequency 0.72 s.

The dataset describes activity across hundreds of thousands of voxels in each subject. To reduce noise as well as the network size defined in terms of pairwise correlations we rely on techniques for describing the brain signal using fewer components. We use independent component analysis (ICA) as a method for dimensionality reduction. Using ICA we produce a set of 50 independent components (ICs) each described as a spatial map consistent over subjects (Additional file 1: Figure S1), and an accompanying time-course of component activity. ICs are under few spatial constraints and have a whole-brain spatial distribution. In practice, however, ICs are often peaked around a smaller localized region and components rarely overlap.

We represent component activity as a temporal network, by making the abstraction that ICs of the brain are physical nodes  $i$ , and correlations between their time-courses within a window are links at times  $\alpha$  such that all intra-layer links are represented by adjacency matrices  $W_{ij}^\alpha$ . We use a sliding window of size  $w = 25$  (each time-step is 0.72 s). Thus, the 405 original time-ticks in the data results in a temporal networks of 381 layers. Figure 1 gives a schematic overview of the experimental design and the processing procedures. Note that we use a fairly short window for correlating time-courses compared to what is recommended in the resting-state dynamic functional connectivity literature (Leonardi and Van De Ville 2015; Zalesky and Breakspear 2015). This is necessary, as the window-duration must remain shorter than the time-scale of the two tasks and fixation periods, since, if the window exceeds this time-scale, we cannot compare the temporal dynamics of communities to the experiment design. There is, however, evidence from task-based fMRI studies showing that shorter window-lengths have the necessary stability to detect changes in ongoing cognition (Gonzalez-Castillo et al. 2015). The *inter-layer* adjacency matrix  $D_i^{\alpha\beta}$  that connects states at times  $\alpha$



and  $\beta$  of each physical node  $i$  is inferred with neighborhood flow coupling (NFC) (Aslak et al. 2018). In brief, NFC couples temporal states of each physical node by a strength that is proportional to the neighborhood similarities of each state-pair. If at two separate times a node has similar neighborhoods, there will be a strong link between those two time-states of the node. Consequently, time-intermittent communities (communities that repeatedly disappear and re-appear) will have many strong links between them because, at each recurrence, the member nodes have similar neighborhoods. Thus, inferring the inter-layer adjacency matrix  $D_i^{\alpha\beta}$  using NFC creates a temporal network that, when used as input to a community detection algorithm, can reveal intermittent communities.

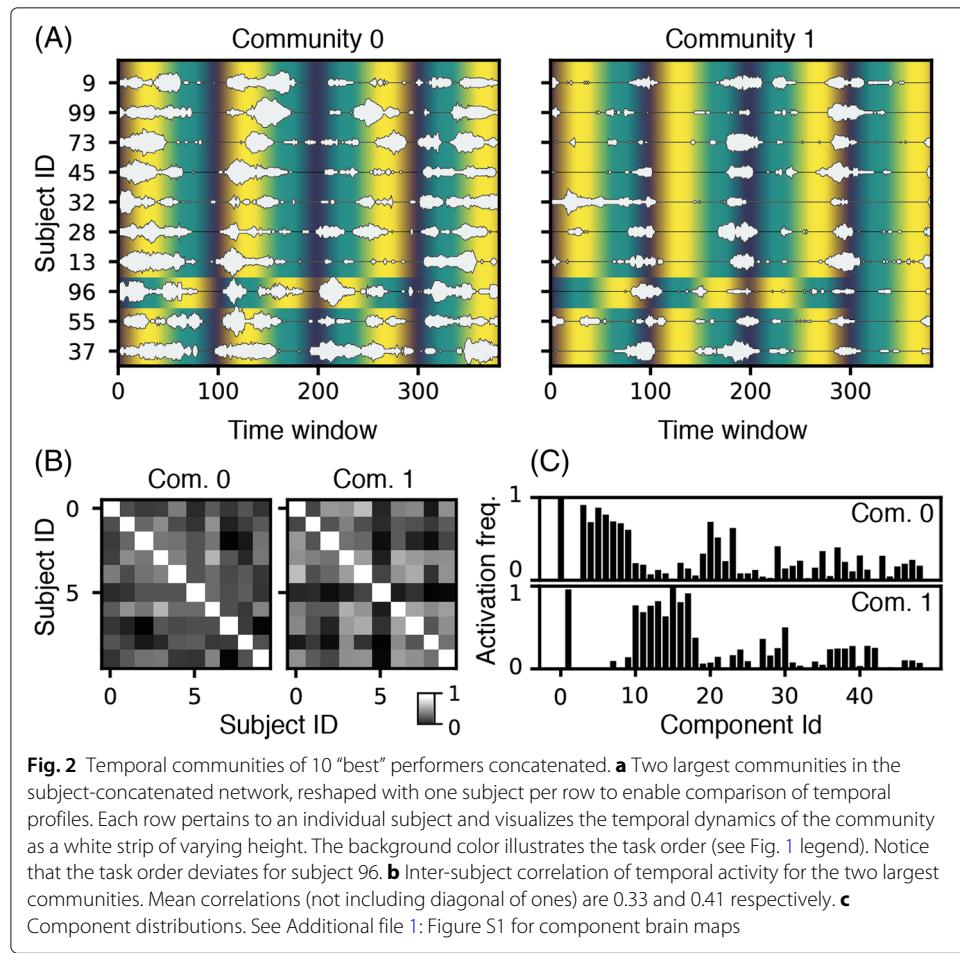
We use the Infomap algorithm for community detection (Rosvall et al. 2009) which, in brief, partitions a network into modules that have maximal *flow* within them. Infomap minimizes a cost function called the Map Equation, which is an analytical lower bound on the per-step description length of a random walker on the network given a partition of nodes into modules. On an abstract level, it considers the network as substrate for *information* flow, which by analogy suits our current application well. We use the software implementation of Infomap which has built-in support for NFC (Edler and Rosvall 2014).

#### We find communities that are shared across subjects

One way that we can search for communities which—if they exist—are guaranteed to be universal across subjects, is to time-concatenate the networks of multiple subjects to build a larger temporal network and then find a single community detection solution. This strategy enables a functional network which activates in multiple different subjects, to be detected as a single community.

We start by performing this analysis for the ten subjects with the highest 2-back task performance accuracy, and find two highly distinct communities (Fig. 2). One community (here, *community 1*) appears regularly during fixation for nearly all subjects. Another, larger, community (*community 0*) seems to activate during task performance and turn off completely during fixation for most subjects. Communities 0 and 1 are, while not entirely disjoint, distinctly different, both in their component distributions and in their temporal activity series, where they exclude each other for many subjects (Fig. 2c). The community detection method does find more communities, but none that are very large and regular across subjects or otherwise distinguishable from noise. In Additional file 2: Figure S2 we display communities 2 and 3 to support this insight. Furthermore, since Infomap indexes communities by descending size, it is highly unlikely that any of the remaining smaller communities contain signal.

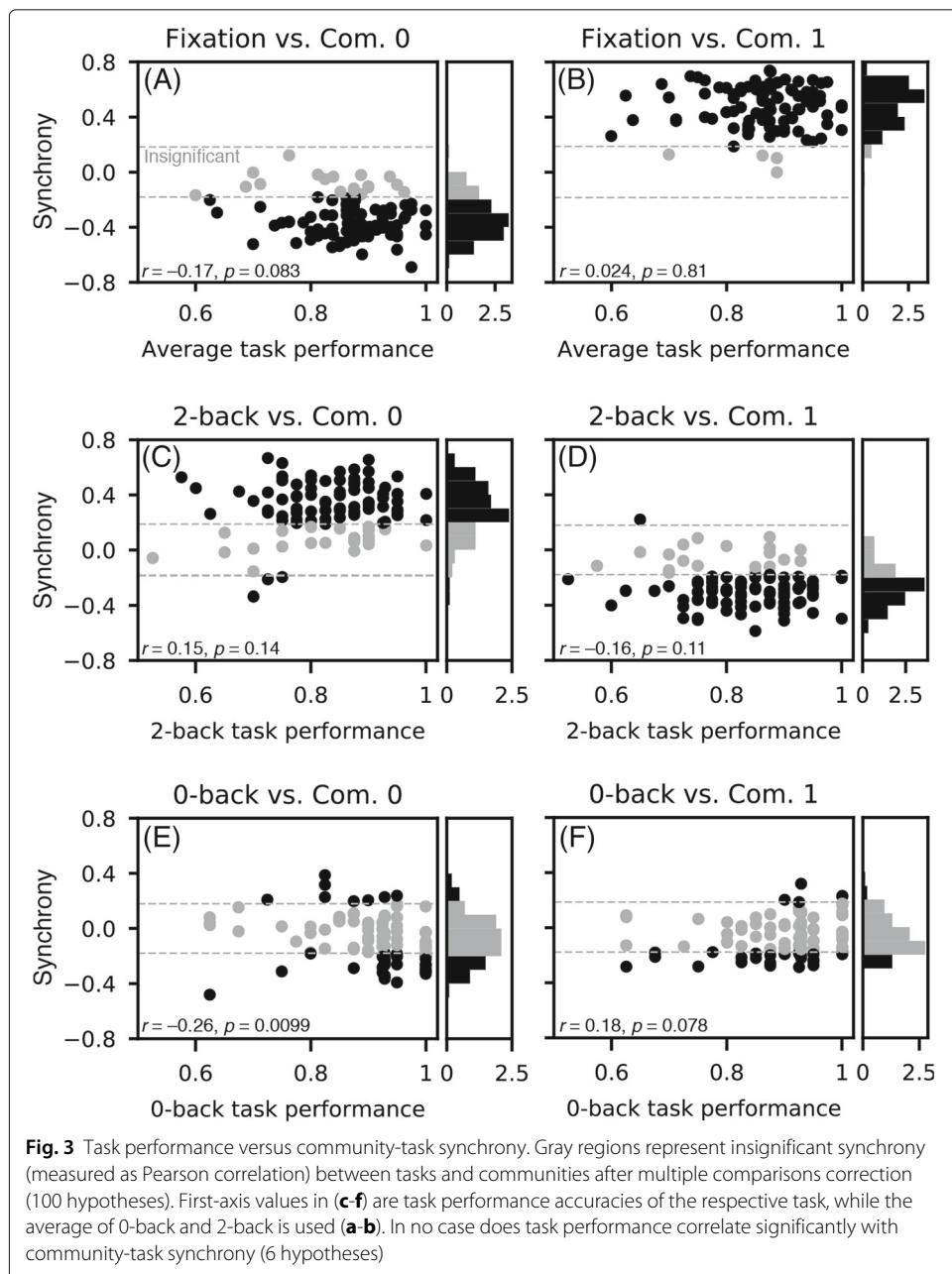
Next, we perform the same analysis on the ten subjects with the lowest 2-back task accuracy, and find that both communities 0 and 1 are recovered with largely the same component profiles (Additional file 3: Figure S3). For community 0 it appears that the time activity series is less modulated by the course of the experiment. Similarly, we perform the analysis on all 100 subjects and recover again communities 0 and 1 with almost identical component distributions (Additional file 2: Figure S2). This repeated detection of communities 0 and 1 strongly suggests that, for the chosen parameter set (see “Materials and methods” section for discussion on parameters), these communities are non-random and prominent in our population.



### Task-synchrony reveals two prominent functional networks

From visual inspection of Fig. 2a, it is clear that the activity patterns of communities 0 and 1 are modulated by the experimental tasks that subjects undertake. We now quantify this correlation. Using the community detection solution obtained from all 100 subjects together (Additional file 2: Figure S2), we gauge the level of synchrony between communities and tasks. For a given subject, we measure synchrony between a community and a task as the Pearson correlation between the community’s activity level over time and a vector representing when a task is being undertaken (specifically, the proportion of a given task inside each time-window, see Fig. 1a3). We plot the distribution of community-task synchrony within the population for each combination of task (fixation, 2-back and 0-back) and community (0 and 1). We also scatter-plot synchrony values against the task performance accuracies to investigate if they co-vary. Synchronies that are insignificant after multiple comparisons correction (100 hypotheses) are rendered in gray.

We observe that, for almost all subjects, community 1 is in strong positive synchrony with the fixation period, while community 0 is in strong negative synchrony with the fixation period (Fig. 3a-b). The exact opposite is observed for the 2-back task, where we find that most subjects activate community 0 and shut down community 1 (Fig. 3c-d). Interestingly, there is no such trend in the synchrony distributions for the 0-back task,



**Fig. 3** Task performance versus community-task synchrony. Gray regions represent insignificant synchrony (measured as Pearson correlation) between tasks and communities after multiple comparisons correction (100 hypotheses). First-axis values in (c-f) are task performance accuracies of the respective task, while the average of 0-back and 2-back is used (a-b). In no case does task performance correlate significantly with community-task synchrony (6 hypotheses)

revealing that community 0 is not a general task-related community, but strictly a 2-back-task related community.

We plot the community-task synchrony values for each subject against their task performance accuracy score (recall: *fraction of correct presses*) and measure the Pearson correlation between these – since it is reasonable to hypothesize that if a community represents brain activity related to a specific task, then the performance accuracy of that task should modulate the dynamics of the community. However, we measure no significant correlations between task performance accuracy and community-task synchrony. Finally, as a way to correct for individual subject ability, we test the relation between performance gain (2-back performance divided by 0-back performance) and synchrony, yet find that it is not significant.

In reviewing the core components of communities 0 and 1, we found a good correspondence when matching them against brain domains that are known to activate during working memory and default mode function, respectively. Both community 0 (working memory) and community 1 (default mode) have components that together cover the regions known to activate in their associated systems. At the same time they also have a number of core components that are not associated with these, especially ones located in the occipital lobe. We perform the matching using the *Neurosynth* software (Yarkoni et al. 2011; Yarkoni 2011). In Additional file 4: Figure S4, we visualize the core components that best match active regions in their associated activation maps, next to the relevant z-axis slices of those maps.

### Conclusion & discussion

By modeling the dynamic dependencies between brain regions in fMRI data, using a temporal network, and inferring inter-temporal dependencies in that network using the NFC principle, we are able to detect two salient and distinctly characteristic communities. The communities are shared across subjects, and for most subjects the communities are strongly modulated by the experimental conditions of the task. The first community primarily activates during the 2-back condition – and importantly not the 0-back condition – while the second activates during the fixation periods.

We do not find any significant correlations between the subjects' task performance accuracies and the extent to which they activate certain communities in synchrony with specific tasks. This is something we might have expected to be the case as it is reasonable to hypothesize that when a task is being performed poorly, the associated community activates less in synchrony with that task. A possible explanation for this is that the variation in task performance accuracy is small (almost everyone gets at least 60% right), and that even performing above some of the lowest scores requires significant engagement.

The overall method we present entails two steps, first constructing a temporal network and second clustering it. Together this yields five important hyper-parameters which impact the results: In the first step we (1) fix the *edge density* of the temporal network, and increasing this yields denser networks which in turn encode more of the weak component correlations as links. (2) We choose a *threshold for component activation*. Lowering this allows links between nodes representing less activated components. (3) We choose the *length of the window* inside which we measure correlations between components over time. Shortening this window increases the time-resolution but destabilizes the link dynamics of the temporal network. Note also, that we could have chosen a different measure of similarity rather than the Pearson correlation. In the second step, when performing community detection using NFC with Infomap, we must (4) choose a *relax rate* which encodes how freely information can flow between layers of time. Lowering the relax rate yields communities that are more constrained in time. (5) Finally, we set a *neighborhood similarity threshold*, below which inter-temporal links connecting states of each node are removed. Choosing a high threshold restricts information to only flow between temporal states that are highly similar, which yields communities with very well-defined component profiles.

In this proof-of-concept work, we do not set model parameters randomly, but instead sample a large number of solutions for different parameters and choose the set that,

based on our qualitative assessment, gives the best results. However, it is clear that future work should include a more rigorous hyper-parameter tuning protocol. There are two challenges in this regard, the most severe being to choose a cost function since community detection is an unsupervised method and thus difficult to quantitatively evaluate. Second, each evaluation step is slow (minutes) and with five hyper-parameters the search space is large, thus it may be necessary to employ specialized gradient-free techniques for sampling parameters. Our approach, with the chosen parameters, only recovers two communities that are repeatedly detected across different runs and population sizes, i.e. 10 and 100 subjects. It is reasonable to suppose that a better approach could render the modular structure of the temporal network in finer detail. In the current paper, it was, however, not our intention to give the best solution, but merely to demonstrate that a solution with this approach is feasible.

Finally, although computationally impractical, we would ideally want to perform the current analysis entirely without the initial dimensionality reduction step, as it puts a hard bound on the granularity of our results. Future work building on this paper should, therefore, consider using more fine-grained techniques for segmenting the brain into parcels or components.

## Materials and methods

### Dataset

We obtain the dataset from the Human Connectome Project (HCP) (Barch et al. 2013). It originally comprises fMRI measurements for 1200 subjects with various tasks and resting-state data that all are repeated twice (in two separate phase encodings). In this paper, we only use the second run of the first 100 subjects. We analyze the *n*-back task designed originally to study working memory. Subjects are presented with images and are asked to press a button if it matches the image presented two instances back (2-back condition). Furthermore, they are asked in another experiment block to press when a target image is presented (0-back). Finally, there are periods of rest with no explicit task (fixation). The duration of an experiment is just under 5 min. For each subject, measurements are taken in 405 ticks with frequency 0.72 s.

### ICA

We apply a group independent component analysis (ICA) (Calhoun et al. 2001; Varoquaux et al. 2010) to extract 50 spatially independent components using the nilearn Python library (Abraham et al. 2014). In group (spatial) ICA, subjects data are concatenated in time such that we end up with a matrix of size #voxels  $\times$  #timepoints  $\cdot$  #subjects  $\cdot$  #runs. This is then first reduced using PCA to 50 components, and the spatial maps rotated according to some independence criterion to make them as spatially independent as possible. In our case, we used the cubic-approximation to the negative entropy, which is the default in nilearn. The ICA algorithm is run on the first 100 subjects in the data repository using both runs of the experiment. Preprocessing includes spatial smoothing (FWHM 6 mm) and high-pass filtering at 0.008 Hz. The ICA algorithm (fastICA (Hyvärinen 1999)) is restarted 10 times and the best solution is chosen using the cost-function as a measure-of-fit. We finally rank the components by their temporal consistency over subjects. This is done by calculating for each component separately the temporal correlation between all pairs of subjects, then taking the average and ranking components in descending

order. One component is removed due to a large overlap with regions associated with cerebrospinal fluid.

Component activations are first standardized such that the distribution of activities in each component over time was centered at 0 with standard deviation 1, then rescaled to [0, 1].

### Temporal networks

We define a link between two components to be the correlation between their activity within a window of time,  $w$ . While comparing at each instant of time is meaningful for creating temporal networks from time-series in other datasets, fMRI measurements are too noisy for this. As discussed in the “[Results](#)” section, we choose the window-duration  $w = 25$ , and the window then slides one tick at a time, thus for  $T = 405$  timesteps the resulting number of time-windows after this procedure is  $381 = 405 + 1 - 25$ . When high-pass filtering the fMRI signal at 0.008 Hz this time-window is short enough to induce spurious correlations between components (Leonardi and Van De Ville [2015](#); Zalesky and Breakspear [2015](#)), and we, therefore, threshold away components that are not activated slightly above the mean (specifically 0.6). Moreover, for each subject we choose the top 1% links with highest positive correlation coefficient out of all possible links in the network, such that link density is consistent across subjects. Since we concatenate subjects in time and perform community detection on this new larger network, a constant density is necessary as subject variation would otherwise be interpreted by the model as temporal variation.

### Community detection

We use Infomap with the Neighborhood Flow Coupling (NFC) method (Aslak et al. [2018](#)) for community detection in our temporal networks. NFC infers dependencies across time between states of each node based on their neighborhood similarity and embeds inter-temporal links in the network proportional to these. Infomap minimizes the per-step description length of a random walker on the network, lower-bound by the *Map Equation*, by finding modules and submodules that capture large amounts of flow (Rosvall et al. [2009](#)). This method takes two important hyperparameters: First, one has to choose the *relax rate* (we use 0.5), which describes how frequently the walker can transition between layer. A high relax rate leads to large amounts of flow across time, which gives rise to large communities that span many time-layers. With this dataset, we find that all nodes are classified as being in the same community when the relax rate is too high. Second, one has to choose a similarity threshold for connecting sibling-nodes across time, above which interlayer links may be created (we use 0.75). A low threshold allows two states of a node, residing in different temporal layers, to have an interlayer link that connects them even when they only have weakly similar neighborhoods. Reversely, a high threshold sets a high requirement for neighborhood similarity for interlayer links to be created. In this data, a too high threshold yields communities that are only detected when the few nodes that make them up are linked. These communities in turn share many nodes but have different labels. Low thresholds create the opposite problem, which is that only few communities with very broad component profiles are detected.

In this work, we do not search for parameters in any rigorous way, but simply sample a large number of solutions and choose a corresponding parameter set that gives

meaningful results. There are two reasons for this. First, we cannot easily define a good objective function for an optimization program, since community detection is an unsupervised problem. Second, the primary purpose of this study is to demonstrate the feasibility of using the described temporal community detection framework on fMRI data. To this end, we do not strictly need *the* optimal parameter set, but only a set of parameters that serves our point.

### Statistical methods

Across all statistical tests, a significance level of 0.05 is used. We use the Bonferroni method for all multiple comparisons corrections.

### Additional files

**Additional file 1:** All ICA components. Note that IC 2 is removed from the network because it is an artifact (see "Materials and methods" section). Yellow and blue squares aside component labels indicate how frequently they participate in community 0 and 1, respectively. (PNG 751 kb)

**Additional file 2:** Temporal activity comparison for all 100 subjects. Identical to Fig. 2, but with a temporal network constructed from all 100 subjects. Subjects on the second axis are sorted by 2-back performance accuracy (higher on axis indicates worse performance). Refer to Fig. 1 for explanation on color coding. **a** Mean inter-subject correlations are 0.24, 0.39, 0.028, and 0.046, respectively. (PNG 1071 kb)

**Additional file 3:** Temporal communities of 10 "worst" performers concatenated. Identical to Fig. 2, but with a temporal network constructed from the ten subjects with lowest 2-back performance accuracy. Refer to Fig. 1 for explanation on color coding. **b** Mean inter-subject correlations are 0.13 and 0.38, respectively. (PNG 293 kb)

**Additional file 4:** Neuroanatomical interpretation. We match brain regions to anatomical regions using the *Neurosynth* (Yarkoni et al. 2011; Yarkoni 2011) software which, given a keyword, synthesizes neuroimaging data from the relevant literature into an activation map. Communities 0 and 1 contain components that correspond well to the synthesized activation maps for keywords "working memory" and "default mode", respectively. **a** Community 0, component 6 is situated in the dorsolateral prefrontal cortex, its component 7 is distributed both in the parietal lobe and the intraparietal sulcus, and there is furthermore some activity in the anterior insula through component 21. **b** Community 1 component 11 is centered on the posterior cingulate cortex, 13 and 15 are both in the medial prefrontal cortex region and 16 is located in the angular gyrus. Each set of components cover quite well the regions that together make up their associated activation maps. At the same time, the communities also include core components that are not typically active in the associated systems. Notably, there are many active components situated in the occipital lobe which do not map to regions commonly associated with either working memory or default mode function. See Additional file 1: Figure S1 for component maps. (PNG 899 kb)

### Acknowledgements

Not applicable.

### Authors' contributions

UA performed network analysis, produced and edited figures and was a major contributor in writing the manuscript. SFVN analysed fMRI data and extracted ICA components. MM and SL were engaged as senior advisors, contributing domain knowledge and ideas that shaped the methodology. All authors engaged in refining the manuscript, and read and approved the final version.

### Funding

S.L. was funded by the Danish Council for Independent Research, Grant No. 4184-00556a, and the Villum Foundation "High Resolution Networks." The Human Connectome Project was funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

### Availability of data and materials

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Data available at Ref. (The Human Connectome Project 2009). Infomap and the implementation of NFC is available at Ref. (Edler and Rosvall 2014).

### Competing interests

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Centre for Social Data Science, University of Copenhagen, 1353 Copenhagen, Denmark. <sup>2</sup>DTU Compute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. <sup>3</sup>Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark. <sup>4</sup>Mental Health Services - Capital Region of Denmark, Psychiatric Centre Copenhagen, Copenhagen University Hospital, Rigshospitalet, 2100 Copenhagen, Denmark. <sup>5</sup>Department of Psychology, University of Copenhagen, 1353 Copenhagen, Denmark.

Received: 21 February 2019 Accepted: 23 July 2019

Published online: 23 August 2019

**References**

- Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G (2014) Machine learning for neuroimaging with scikit-learn. *Front Neuroinformatics* 8:14
- Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD (2014) Tracking whole-brain connectivity dynamics in the resting state. *Cereb Cortex* 24(3):663–676
- Aslak U, Rosvall M, Lehmann S (2018) Constrained information flows in temporal networks reveal intermittent communities. *Phys Rev E* 97(6):062312
- Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C, et al. (2013) Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80:169–189
- Bassett DS, Porter MA, Wymbs NF, Grafton ST, Carlson JM, Mucha PJ (2013) Robust detection of dynamic community structure in networks. *Chaos Interdiscip J Nonlinear Sci* 23(1):013142
- Bassett DS, Sporns O (2017) Network neuroscience. *Nat Neurosci* 20(3):353
- Bassett DS, Wymbs NF, Porter MA, Mucha PJ, Carlson JM, Grafton ST (2011) Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci* 108(18):7641–7646
- Beckmann CF, DeLuca M, Devlin JT, Smith SM (2005) Investigations into resting-state connectivity using independent component analysis. *Philos Trans R Soc Lond B Biol Sci* 360(1457):1001–1013
- Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 34(4):537–541
- Bokland JW, Bokil H, Allen CB, Mitra PP (2009) The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PloS ONE* 4(9):7200
- Brodmann K (1909) Vergleichende Lokalisationslehre der Grosshirnrinde in Ihren Prinzipien Dargestellt Auf Grund des Zellenbaues. Verlag von Johann Ambrosius Barth, Leipzig
- Calhoun VD, Adali T, Pearlson GD, Pekar J (2001) A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* 14(3):140–151
- Edler D, Rosvall M (2014) The Infomap Software Package. <http://mapequation.org>. Accessed Apr 2019
- Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, et al. (2016) The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb Cortex* 26(8):3508–3526
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci* 102(27):9673–9678
- Glover GH (1999) Deconvolution of impulse response in event-related bold fMRI. *Neuroimage* 9(4):416–429
- Gonzalez-Castillo J, Hoy CW, Handwerker DA, Robinson ME, Buchanan LC, Saad ZS, Bandettini PA (2015) Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proc Natl Acad Sci* 112(28):8762–8767
- Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci* 100(1):253–258
- Gusnard DA, Raichle ME (2001) Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci* 2(10):685
- Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10(3):626–634
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) Fsl. *Neuroimage* 62(2):782–790
- Khambhati AN, Mattar MG, Wymbs NF, Grafton ST, Bassett DS (2018) Beyond modularity: Fine-scale mechanisms and rules for brain network reconfiguration. *Neuroimage* 166:385–399
- Kiviniemi V, Kantola J-H, Jauhainen J, Hyvärinen A, Tervonen O (2003) Independent component analysis of nondeterministic fMRI signal sources. *Neuroimage* 19(2):253–260
- Leonardi N, Van De Ville D (2015) On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage* 104:430–436
- Lowe M, Mock B, Sorenson J (1998) Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *Neuroimage* 7(2):119–132
- McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ (1998) Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp* 6(3):160–188
- Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328(5980):876–878
- Park H-J, Friston K (2013) Structural and functional brain networks: from connections to cognition. *Science* 342(6158):1238411
- Rosvall M, Axelsson D, Bergstrom CT (2009) The map equation. *Eur Phys J Spec Top* 178(1):13–23
- Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, et al. (2012) Temporally-independent functional modes of spontaneous brain activity. *Proc Natl Acad Sci* 109(8):3131–3136
- The Human Connectome Project (2009). <http://www.humanconnectomeproject.org/>. Accessed Apr 2019
- Thirion B, Varoquaux G, Dohmatob E, Poline J-B (2014) Which fMRI clustering gives good brain parcellations? *Front Neurosci* 8:167

- Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium W-MH, et al. (2013) The wu-minn human connectome project: an overview. *NeuroImage* 80:62–79
- Varoquaux G, Sadaghiani S, Pinel P, Kleinschmidt A, Poline J-B, Thirion B (2010) A group model for stable multi-subject ICA on fMRI datasets. *NeuroImage* 51(1):288–299
- Vidaurre D, Smith SM, Woolrich MW (2017) Brain network dynamics are hierarchically organized in time. *Proc Natl Acad Sci* 114(48):12827–12832
- Yarkoni T (2011) Neurosynth. <http://neurosynth.org>. Accessed Apr 2019
- Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* 8(8):665
- Zalesky A, Breakspear M (2015) Towards a statistical test for functional connectivity dynamics. *NeuroImage* 114:466–470

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---

1                   **Rhythm of relationships in a social fish over the course of a full year in the wild**

2                   **Short Title: Dynamics of Carp Social Networks**

3                   **Ulf Aslak<sup>1</sup>, Christopher T. Monk<sup>2</sup>, Dirk Brockmann<sup>3,4</sup>, Robert Arlinghaus<sup>2,5,6</sup>**

4                   <sup>1</sup>*Centre for Social Data Science, University of Copenhagen, DK-1353 København K and DTU Compute,*  
5                   *Technical University of Denmark, DK-2800 Kgs. Lyngby.* <sup>2</sup>*Department of Biology and Ecology of Fishes,*  
6                   *Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin,*  
7                   *Germany.* <sup>3</sup>*Robert Koch-Institute, Nordufer 20, D-13353 Berlin, Germany.* <sup>4</sup>*Institute for Theoretical*  
8                   *Biology and Integrative Research Institute for the Life Sciences, Humboldt Universität zu Berlin, Berlin,*  
9                   *Germany.* <sup>5</sup>*Faculty of Life Sciences and Integrative Research Institute for the Transformation of*  
10                  *Human-Environmental Systems, Humboldt-Universität zu Berlin, Invalidenstrasse 42, 10115*  
11                  *Berlin, Germany.* <sup>6</sup>*Division of Integrative Fisheries Management, Department of Crop and Animal*  
12                  *Sciences, Faculty of Life Science, Humboldt-Universität zu Berlin, Invalidenstrasse 42, 10115 Berlin,*  
13                  *Germany.*

14                  **Corresponding author:** UA

15                  **Contact:** [ulfaslak@gmail.com](mailto:ulfaslak@gmail.com)

16                  **Keywords:** acoustic telemetry, common carp, fish behavior, social networks, reality mining

17                  **Author Contributions:** UA, CTM, DB & RA conceived the ideas and developed the methodology,  
18                  CTM and RA collected the data, UA and DB analysed the data, UA, CTM and RA led the writing  
19                  and all authors contributed critically to the drafts and gave final approval to the data

20

21     **Abstract**

- 22       1. Animals are expected to adjust their social behaviour to cope with challenges in their  
23       environment. Therefore, for fish populations, in temperate regions with seasonal and daily  
24       environmental oscillations, characteristic rhythms of social relationships should be pronounced.  
25       To date, most research concerning fish social networks and biorhythms has occurred in artificial  
26       laboratory environments or over confined temporal scales of days to weeks. By contrast, little is  
27       known about the social networks of wild, freely roaming fish, including how seasonal and  
28       diurnal rhythms modulate social networks over the course of a full year. The advent of high-  
29       resolution acoustic telemetry enables us to quantify detailed social interactions in the wild over  
30       time-scales sufficient to examine seasonal rhythms at whole-ecosystems scales.
- 31       2. Our objective was to explore the rhythms of social interactions in a social fish population at  
32       various time-scales over one full year in the wild by examining high-resolution snapshots of  
33       dynamic social network.
- 34       3. To that end, we tracked the behaviour of 36 adult common carp, *Cyprinus carpio*, in a 25 ha lake  
35       and constructed temporal social networks among individuals across various time-scales, where  
36       social interactions were defined by proximity. We compared the network structure to a  
37       temporally shuffled null model to examine the importance of social attraction, and checked for  
38       persistent characteristic groups (“friendships”) over time.
- 39       4. The clustering within the carp social network tended to be more pronounced during daytime  
40       than nighttime throughout the year. Social attraction, particularly during daytime, was a key  
41       driver for interactions. Shoaling behavior substantially increased during daytime in the  
42       wintertime, whereas in summer carp interacted less frequently, but the interaction duration  
43       increased. Characteristic groups were more common in the summer months and during  
44       nighttime, where the social memory of carp lasted up to two weeks.
- 45       5. We conclude that social relationships of carp change diurnally and seasonally. These patterns  
46       were likely driven by predator avoidance, and seasonal shifts in lake temperature, visibility,  
47       forage availability and the presence and location of anoxic zones. The techniques we employed  
48       can be applied generally to high-resolution biotelemetry data to reveal social structures across  
49       other fish species at ecologically realistic scales.

50      **Introduction**

51           Animals are faced with continual exogenous oscillations of their environment and endogenous  
52        oscillations of their own physiology, over various timescales (Hastings, 2010). For example, all animals  
53        must cope with daily oscillations driven by the Earth's rotation around its axis, ~30 day oscillations  
54        driven by the lunar cycle and yearly oscillations driven by the Earth's elliptical orbit around the sun.  
55           Animals must also cope with internal oscillations driven by, for example, a heartbeat at very short time  
56        scales, or a reproductive cycle at longer, maybe even seasonal time scales. Furthermore, to feed and  
57        survive animals must also track the responses of their predators and prey to environmental oscillations  
58        (Vandermeer, 2004). Importantly, in ectotherms such as fish, ecosystem metabolism (e.g., productivity  
59        of resources and thermal environment) in the temperate zone reacts strongly to seasonal and daily  
60        changes in light and temperature, which causes periodic variation in the availability of food and the  
61        distribution of habitats (Hunt, Jardine, Hamilton, & Bunn, 2012; Stæhr & Jensen, 2007). Accordingly,  
62        animals must constantly respond to oscillations in their biotic environment throughout the food web as  
63        resource requirements and availability, danger and shelter, and reproductive opportunities oscillate  
64        across various frequencies to find food and shelter and avoid mortality.

65           Animals have adopted a number of physiological and behavioural strategies to cope with  
66        periodic fluctuations of the environment over time. The circadian clock, for example, is an inner  
67        oscillator synchronized with solar time that appears universally across taxa (Bernard, Gonze, Čajavec,  
68        Herzel, & Kramer, 2007; Dunlap, 1999; Edgar et al., 2012); it influences metabolism (Kohsaka & Bass,  
69        2007), hormones (Leatherland & McKeown, 1973) and ultimately behavior (Naylor, 1988). To survive  
70        resource fluctuations over yearly time-scales organisms may employ strategies like hibernation or other  
71        forms of metabolic depression (Ruf & Geiser, 2015). For most animals, and for ectotherms in particular,  
72        temperature is a critical resource they must adapt to (Magnuson, Crowder, & Medvick, 1979). In fishes,  
73        all biological processes are influenced by exogenously triggered temperature, including enzyme activity,  
74        metabolism, digestion, and feeding rate, leading to a strict dependency on warm waters to grow and  
75        reproduce (Conover & Present, 2016; Shultz, Reynolds, & Conover, 1996). In addition to modifying their  
76        physiology, animals also have the option to alter their behaviour in response to environmental changes  
77        (Holland, Brill, Chang, Sibert, & Fournier, 1992), which may take the form of migrating to more  
78        favourable habitats (Somveille, Rodrigues, & Manica, 2015), and importantly, animals may also change  
79        their response to other conspecifics by becoming more or less social (Monk et al., 2018).

80           There are costs and benefits to both pro-social and anti-social behaviour, which depend on an  
81        organism's environment and resource requirements (Monk et al., 2018; Snijders, Kurvers, Krause,  
82        Ramnarine, & Krause, 2018; Wiens, 1976). Group living can allow for increased predator avoidance  
83        (Foster & Treherne, 1981; Landeau & Terborgh, 1986; Pulliam, 1973), faster ability to find rare or mobile  
84        resources (Hills et al., 2015; Magurran & Higham, 1988; Pitcher, Magurran, & Winfield, 1982), increased  
85        ability to hunt large prey, and better conservation of resources (Gilbert, Robertson, Le Maho, Naito, &  
86        Ancel, 2006). However, living in groups also comes at the cost of sharing resources among conspecifics  
87        (Bertram, 1978) or increased transmission of parasites and pathogens (Côté, Poulin, & Zealand, 1985).  
88        Under certain conditions it is therefore, better to be solitary and defend a territory (J. Brown, 1968;  
89        Bryant & Grant, 1995) or behave nomadically (Eklöv, 1992). Hence, as the abiotic and biotic

90 environments as well as the internal physiological state of an animal oscillate we expect to observe  
91 periodic patterns in social behaviour across various time-scales. The behavioural reactions that occur  
92 most likely depend on the evolutionary adaptions of particular species, modified by local environmental  
93 conditions. For example, killer whales, *Orcinus orca*, increase their sociality with increasing resource  
94 abundance (Franks et al., 2012), while chacma baboons, *Papio hamadryas ursinus*, increase their  
95 sociality when resources become scarce (Henzi, Lusseau, Weingrill, Van Schaik, & Barrett, 2009). The  
96 environment-species interaction also dictates among population differences in social responses to the  
97 environment within the same species. For example banded killifish, *Fundulus diaphanus*, form larger  
98 groups when a predator is detected, but reduce their group size when food is available (Hoare, Couzin,  
99 Godin, & Krause, 2004).

100 To date, there have been few explorations into the temporal dynamics of animal social networks  
101 (Blonder, Wey, Dornhaus, James, & Sih, 2012; Pinter-Wollman et al., 2014), and none have been able to  
102 explore the dynamics across the full spectra of timescales expected within a full growing season in the  
103 temperate zone. The lack of long-term (i.e., multiple months), high-resolution animal social network  
104 data in the wild is a result of the immense challenge in collecting it at sufficient spatial-temporal  
105 resolutions (Krause et al., 2013). Most long-term datasets are generated through consistent periodic  
106 visual observations of interactions among identifiable animals (Henzi et al., 2009; Wittemyer, Douglas-  
107 Hamilton, & Getz, 2005). There is a particular lack of long-term social network data for fish because  
108 making long-term underwater observations has not been possible until recently. Initial work in the wild  
109 has however shown that certain fish species tend to be detected in groups of characteristic individuals  
110 (Hay & McKinnell, 2002; Klimley & Holloway, 1999; Ward et al., 2002; Wilson et al., 2014), but this  
111 pattern is not universal across species (Helfman, 1984). The lack of long term studies of fish in the wild is  
112 problematic because fish have been used widely as valuable model organisms to study social  
113 behaviours, such as shoaling (Wilson, Croft, & Krause, 2014), but most research in social behavior among  
114 fishes has occurred over short time periods of a few weeks, and often in non-naturalistic laboratory  
115 environments (Wilson et al., 2014). Evidence from non-human animal studies shows that social behavior  
116 is usually more dependent of ecology than taxonomy (Lefebvre, Palameta, & Hatch, 1996), raising  
117 doubts as to whether our lab-scale understanding is transferable to the wild (Sutter & Arlinghaus, et. al.,  
118 2012; Niemelä & Dingemanse, 2014). Consequently, very little is known with certainty about the  
119 relationships and social lives of wild-living fishes in large populations over long time-scales of multiple  
120 months in the wild.

121 Most species of fish, even cannibalistic ones, must engage in social relationships for at least a  
122 portion of their life, for example during mating and shoaling for predator avoidance (Pitcher, 1986;  
123 Shaw, 1978). To increase fitness, many species socialize to establish hierarchies, exchange information  
124 and avoid predators (Seppälä, Karvonen, & Valtonen, 2008; Suboski & Templeton, 1989). Research in  
125 social learning has demonstrated a number of advanced social behaviors. For example, it has been  
126 shown that fish can learn escape routes from each other (C. Brown & Laland, 2002), infer social  
127 hierarchies by observing fights alone (Grosenick, Clement, & Fernald, 2007), recognize individual  
128 conspecifics (Griffiths & Magurran, 1997), and when given conflicting social cues tend to value public  
129 information over immediate social information (Coolen, Ward, Hart, & Laland, 2005). More complex

130 social behaviors such as cooperation, establishment of partnerships and reciprocation of behaviors that  
131 require risk taking (Croft et al., 2006; Granroth-Wilding & Magurran, 2013; Milinski, Kulling, Kettler, &  
132 Bern, 1990), indicate that fish have well developed social cognition capabilities (Bshary, Gingins, & Vail,  
133 2014).

134 Today, with advanced computational methods and modern tracking technology, such as high-  
135 resolution, precise, acoustic telemetry it is possible to collect snapshots of social interactions at several  
136 second frequencies over years in the wild (Baktoft, Zajicek, Klefoth, Svendsen, & Jacobsen, 2015; Guzzo  
137 et al., 2018; Krause et al., 2013), both underwater (Lennox et al., 2017) and in terrestrial environments  
138 (Wilmers et al., 2015). A suite of techniques are now available for constructing and analyzing social  
139 networks generated from acoustic data (Blonder et al., 2012; Finn et al., 2010; Jacoby & Freeman, 2016;  
140 Mourier, Brown, & Planes, 2017). With a full-year dataset containing high-resolution mobility traces of  
141 36 common carp (*Cyprinus carpio*) in a small lake, we for the first time explore the “rhythm of  
142 relationships” in nature at a whole ecosystem scale for a fish. We ask a basic question: how do seasons  
143 and daytime modulate the social behavior of a fish species described as highly social from laboratory  
144 contexts (Huntingford et al. 2010)?

145 Wild common carp are long-lived, omnivorous, typically benthivorous, warmwater cyprinids,  
146 native to eastern Europe and Asia, but the species has been domesticated for aquaculture purposes as  
147 early as 2000 years ago and globally introduced into the wild where it forms feral populations (Balon,  
148 2004). Carp have permanently established widely across the freshwater ecosystems around the globe  
149 (Howes, 1991; Parameswaran, Alikunhi, & Sukumaran, 1972; Vilizzi, 2012). They constitute a key  
150 fisheries resource for both commercial (FAO, 2018) and recreational fisheries (Arlinghaus & Mehner,  
151 2003). Yet, appreciation of carp is not global. The species is also considered a pest in certain regions  
152 where it is originally non-native, such as North America (Bajer, Sullivan, & Sorensen, 2009), Australia  
153 (Taylor, Tracey, Hartmann, & Patil, 2012) and parts of Europe, such as Spain (Benito, Benejam, Zamora,  
154 & García-Berthou, 2015) as they can disturb aquatic macrophytes, leading to loss of water clarity and  
155 increased nutrient concentrations (Bajer et al., 2016). Accordingly, learning about the social behaviour  
156 of carp could help to improve both carp fisheries management (Klefoth, Skov, Kuparinen, & Arlinghaus,  
157 2017), and eradication techniques (Bajer, Chizinski, & Sorensen, 2011).

158 The behaviour of carp is variable, both among populations (Hennen & Weber, 2014; Benito et  
159 al., 2015; Weber, Brown & Willis, 2016) and among individuals (Monk & Arlinghaus, 2017; Pollux, 2017).  
160 Increasing levels of domestication in stocked fish are known to increase the carp’s boldness (Klefoth,  
161 Skov, Krause, & Arlinghaus, 2012) and foraging activity and ingestion rates (Klefoth, Pieterek, &  
162 Arlinghaus, 2013). Carp are also known to be a social species, frequently found in groups (Bajer et al.,  
163 2011; Johnsen & Hasler, 1977; Osborne, Ling, Hicks, & Tempero, 2009), and can learn by social  
164 facilitation (Zion, Barki, Grinshpon, Rosenfeld, & Karplus, 2007). Carp are generally thought to occupy  
165 littoral habitats during the spring and summer and move to deeper waters to overwinter in larger  
166 groups (Armstrong et al., 2016; Johnsen & Hasler, 1977; Jones & Stuart, 2009; Penne & Pierce, 2008).  
167 Carp are also known to show marked diurnal behavioural patterns in the spring and summer months.  
168 For example, non-native carp that became established in a reservoir in the Ebro catchment in Spain  
169 were relatively inactive in deep hypoxic waters at nighttime, possibly refuging from nocturnal predators

170 (in particular catfish, *Silurus glanis*), but became more active in shallow waters during the daytime  
171 (Benito et al., 2015). By contrast, in other populations, carp have been observed to increase food  
172 consumption in the nighttime (Bajer, Lim, Travalline, Miller, & Sorensen, 2010; Proske, 1972), strongly  
173 indicating that diel patterns of behavior will vary with local ecological conditions.

174 Our objective was to explore the social relationships of a population of carp over one full year at  
175 a whole ecosystem scale, and to identify the relevant time-scales of carp social behavior in the wild  
176 under realistic ecological scales. To that end, we recorded three-dimensional positions of carp in a  
177 whole-lake using high resolution acoustic telemetry, and inferred the temporal social network of carp by  
178 logging proximity events. We describe how seasons and daytime influence both mobility and social  
179 behavior, perform statistical tests to assess the degree to which time spent together is explained by  
180 social attraction given that other ecological factors (e.g., local food availability) may drive co-location of  
181 two individuals, and finally study the social network at varying time-scales to measure the persistence of  
182 community structure over time. Having found that there is significant local clustering in the social  
183 network, we investigate whether these groups, or communities, that emerge on short time-scales and  
184 give rise to local clustering, persist over time. By analyzing the persistence of clustering over time we  
185 aim to get at the more fundamental question which is whether there is social memory in the system that  
186 drives the groups of fish to meet repeatedly.

## 187 **Methods**

### 188 **Study site**

189 Kleiner Döllnsee ( $52^{\circ}59'32.100\text{ N}$ ,  $13^{\circ}34'046.500\text{ E}$ ) is a 25 ha lake in northern Brandenburg,  
190 Germany ( $52^{\circ}59'41.9''\text{N}$ ,  $13^{\circ}34'56.4''\text{E}$ ), classified as eutrophic with a total phosphorous concentration  
191 of  $38 \mu\text{g L}^{-1}$  at spring overturn. The average depth during the study period was 4.4 m, while the  
192 maximum depth is 7.8 m and the secchi depth was  $1.97 \pm 0.61\text{ m}$  (mean  $\pm$  standard deviation). Reeds  
193 (*Phragmites australis*) form a belt of growth around the lake. Between May and October the lake  
194 stratifies turning layers below ca. 4 m anoxic (See figures SI1 and SI2). Kleiner Döllnsee hosts 14 fish  
195 species typical for mesotrophic to slightly eutrophic natural lakes in German lowlands (Eckmann, 1995).  
196 Top predators include introduced European catfish (*Silurus glanis*), and native northern pike (*Esox lucius*)  
197 and Eurasian perch (*Perca fluviatilis*). Carp competitors include an abundant populations of large  
198 common bream (*Abramis brama*), tench (*Tinca tinca*), rudd (*Scardinius erythrophthalmus*), roach (*Rutilus*  
199 *rutilus*) and white bream (*Blicca bjoerkna*). Through recent nutrient increases, the submerged  
200 macrophyte coverage has been declining in the lake and is now restricted to near shore locations and  
201 macrophytes taller than 10 cm covered 9.2% of the lake area.

### 202 **Telemetry system**

203 The lake has been equipped with 20 submerged (~2 m) high-resolution acoustic telemetry receivers  
204 (WHS 3050; 200 kHz; Lotek Wireless Inc., Newmarket , Ontario, Canada) distributed at fixed locations  
205 throughout the lake (described in detail in Baktoft et al. 2015). The system allows whole-lake positional  
206 telemetry in 3-D at high spatio-temporal resolution with several position fixes per minute depending on  
207 transmitter burst rates (Baktoft et al. 2015). Location datapoints are estimated by hypertriangulation of

208 ultrasonic signals originating from surgically implanted transmitters. Median location accuracy  
209 throughout the lake is 3.1 m (Baktoft et al., 2015). Macrophytes, known to strongly attenuate acoustic  
210 signals, were scarce during the study period; thus we can reason that most signal loss occurred when  
211 the fish swam among the reed close to shore. However, significant decreases in data yield did occur over  
212 the some periods in warmer summer months, indicating improved telemetry performance during the  
213 cooler periods of the year (Fig. 2A). Average across the year data yield was about 40%. For a full  
214 description of the system and its performance see Baktoft et al., (2015).

## 215 Carp population

216 All carp recorded in the dataset were hatchery born and bred in earthen ponds as is typical in  
217 many European fisheries where carp are stocked after being raised in pond aquaculture. In June 2014,  
218 91 carp with transmitters implanted (0.3% to 2.2% body mass) were released to Kleiner Döllnsee. Due to  
219 tag loss, known to be a prevalent problem in carp tagging (Daniel, Hicks, Ling, & David, 2009; Økland,  
220 Hay, Naesje, Nickandor, & Thorstad, 2003), an additional 24 tagged carp were released in September  
221 2014. Of these 115 carp known to have been introduced to the lake, between 25-36 ( $540 \pm 79$  cm total  
222 length, mean  $\pm$  standard deviation; see Fig SI1 for individual level data) were successfully monitored  
223 throughout all of 2015 for an entire year (Fig. 1). The rest experienced substantial tag loss. Tagging-  
224 induced mortality was extremely low as revealed by recaptures that had lost tags but were alive.

225 For transmitter implantation, carp were anaesthetized using a 9:1 EtOH:clove oil solution added  
226 at  $1\text{mL L}^{-1}$  (Carl Roth, Karlsruhe, Germany). All surgical tools and acoustic telemetry tags were sterilized  
227 with a mixture of tap-water and 7.5% povidone-iodine (PVP; Braunol®; B. Braun, Kronberg, Germany)  
228 before each transmitter implantation. We implanted the transmitters (model MM-M-TP-16-50,  
229 dimension: 16 by 85 mm, wet weight: 21 g; Lotek Wireless, Newmarket, Canada) into the body cavity  
230 (see (Klefth, Kobler, & Arlinghaus, 2008) for procedures), and each fish received 4–5 sutures using PDS-  
231 II adsorbable monofilament suture material and FS-1 3–0 needles (Ethicon, USA). Following recovery  
232 from surgery the fish were immediately released into the study lake. The burst frequency of the  
233 transmitters was five seconds, and the transmitters were equipped with a temperature sensor,  
234 recording once per minute, and a pressure sensor to record depth at all other transmissions.

## 235 Inferring social networks

236 Because we did not directly measure social interactions between fish, we had to infer it using  
237 mobility traces. In the current analysis, we used persistent proximity as a proxy for contact. We  
238 employed a number of post-processing techniques to increase accuracy and data yield to produce the  
239 best reconstruction of the temporal contact network as possible. First, we resampled the location data  
240 from 5 s to 15 s to remove noise and recover potentially lost measurements by applying a 30 s median  
241 filter in 15 s increments across the location trajectory of each fish. We then measured pairwise distance  
242 in each time-bin, resulting in  $\frac{N(N-1)}{2}$  time series of inter-fish distance. Each series would have some  
243 missing values which we linearly imputed by up to 30 minutes. To get the times at which links were on  
244 and off we thresholded the distance series at 10 m. Finally, we applied a filter which removed singleton  
245 contact events and clustered consecutive ones with short breaks (up to 5 minutes). Figure 5 shows two

246 examples of distance time series, how we threshold them and how that results in a link activity time  
247 series.

248 Based on individual temperature and location measurements of the  $N = 36$  carp we computed,  
249 for each fish,  $i$ , on each day the local temperature,  $T_i$ , the distance from shore,  $d_i$ , the Shannon entropy,  
250  $S_i = \sum_m c_m \log c_m$ , where  $c_m$  is the fraction of time spent in lake area  $m$  (when the lake is split into  $10\text{ m} \times 10\text{ m}$  cells),  
251 the velocity,  $v_i$ , and the depth,  $h_i$ . From the inferred social network we measured for each pair,  
252  $n$ , the average interaction duration,  $\tau_n^+$ , the average time between interactions,  $\tau_n^-$ , and the interaction  
253 probability,  $p_n$ . Quantities were estimated separately for daytime and nighttime once per solar cycle and  
254 reported as population averages, denoted by dropping the node/link index.

255 We, furthermore, measured the local clustering coefficient as a population average for each  
256 monthly aggregated social network split into day and night. The average local clustering in a network is  
257 bounded between zero and one, and reflects the tendency for triangles, or triads, to form in the  
258 network (Saramäki, Kivelä, Onnela, Kaski, & Kertesz, 2007). Intuitively, if a node has a high local  
259 clustering coefficient its neighbors are highly interconnected. Triads are indicative of community  
260 structure and informs about social behavior at the group-scale (Wasserman & Faust, 1994; Scott, 2000).

## 261 Null model of social attraction

262 Since our inferred social network builds on the assumption that co-location equates interaction,  
263 it is natural to wonder how much of that interaction is due to “social attraction” – meaning very broadly  
264 that fish go to specific places because there are other fish there – and how much is due to the  
265 environment driving the fishes to visit the same places at the same times. Indeed, it is plausible that the  
266 population is entirely non-social and any co-location is due to similar resource use (e.g., habitat choice).  
267 To assess the impact of social attraction on interaction, we created a shuffled dataset using a null model  
268 that time-shifts the mobility trace of each fish independently by a random number of whole days  
269 between zero and six. In the shuffled dataset, any potential correlation in the location traces of two  
270 individuals due to their social attraction was broken (Spiegel, Leu, Sih, & Bull, 2016). The data with  
271 individually shifted mobility tracks then modeled a mobility pattern where each pair swims entirely  
272 independent of each other, effectively breaking location dependencies that may have existed due to  
273 social attraction. Hence, we got an estimate of the background level of that statistic due to habitual  
274 space use. This was, furthermore, a very strong null model since some fraction of fish pairs would be  
275 shifted the same amount (statistically:  $7 (1/7)^2 = 14.3\%$ ), thus not removing all inter-pair dependency.  
276 Effect sizes would therefore be slightly underestimated. We then measured and compared the raw  
277 number of interactions as well as the average local clustering coefficient in the real and shuffled data.  
278 We used this null model because, intuitively, if carp were truly non-social and only interacted when they  
279 happened to use the same areas simultaneously, randomizing the data in this fashion would likely yield  
280 the same amount of co-location events. The only assumption that this null model makes is that key  
281 resources which drive mobility do not fluctuate significantly on the scale of days. See (Spiegel et al.,  
282 2016) for an in-depth discussion of the null model.

283 Finally, we acknowledge that a stronger statistical approach would have been to produce many  
284 (thousands) of such shuffled datasets with this null model and report average effect sizes as well as  $p$ -  
285 values associated with each effect. For the current dataset this was not computationally feasible, since  
286 we inferred interactions by querying the distance between every pair in every time-step, which  
287 yields  $\frac{N(N-1)}{2} \cdot 2.1 \cdot 10^6$  timesteps  $\approx 1.3$  billion queries (or  $\sim 2$  days of computing time using 56 2.60GHz  
288 processors) per shuffled dataset.

## 289 Timescale Analysis to Identify Communities

290 We conducted a computational experiment where we measured how the number of  
291 communities changed when we incrementally split each monthly social network into multiple shorter  
292 aggregates. Specifically, we first aggregated all interactions in a four week window within a given month  
293 and weighted links by the number of interactions between two fish that exceeded the background level  
294 of interaction (number of interactions in the null data). We then broke this network into two two-week  
295 networks each mapping the interactions in their given window. We continue breaking up the networks  
296 into an increasing number of temporally shorter networks until the window size was one hour and the  
297 number of networks was  $24 \text{ h/day} \cdot 28 \text{ days} = 672$ . In each iteration we ran the community detection  
298 algorithm Infomap (Rosvall, Axelsson, & Bergstrom, 2009) on the networks and recorded the number of  
299 communities that had three or more members. In the results, we report the average and standard error  
300 of the mean across slices for each aggregation level.

## 301 Results

### 302 Behavioral trends during day and night across the year

#### 303 Behaviour was variable across the season, while strong shoaling was a daytime winter phenomenon

304 We observed a high variation across most individual and social behaviors across the year. This  
305 variation was largely due to the fish shoaling in deep waters at the center of the lake during daytime in  
306 colder months. We saw differences in swimming speed (Jan. avg.: 1.03 m/s, June avg.: 0.51 m/s) depth  
307 (Jan. avg.: 5.15 m, June avg.: 1.48 m) and distance from shore (Jan. avg.: 103.38 m, June avg.: 40.22 m).  
308 Moreover, we found that indicators of social interactions in the colder months were elevated, such as  
309 time spent together (Jan. avg.: 12.22 minutes, June avg.: 9.32 minutes), and interaction probability (Jan.  
310 avg.: 15.40%, June avg.: 4.20%). In colder months there was also a great difference between social  
311 indicators during day and night, which we interpret as a strong signal that shoaling is indeed a daytime  
312 phenomenon in this population. Figure 2 contains a full summary of these results. In warmer months  
313 starting late March, as the lake stratified (Figure SI2), deep waters turned anoxic forcing the carp closer  
314 to the surface. We observed already in late February that interaction probability and duration of  
315 interaction during daytime decreased, while time between interactions increased overall. This means  
316 that shoaling ceased before deeper water layers turned anoxic in spring and over the summer. Shoaling  
317 in daytime reappeared gradually during the fall and peaked again in December.

#### 318 Behavioral differences between day and night were seasonal

319        Across all the behavioral and social indicators we studied, we only observed high variation over  
320 daytime during the winter, but not in the other seasons. Comparing the time-series of  $p$  (interaction  
321 probability) for January with July, it is clear that social interactions were a periodic function of daytime  
322 only in the winter, whereas they were far more sporadic in the summer (Fig. 2D). A notable exception,  
323 though not directly an indicator of behavior, was data yield. Here, we observed the opposite: in winter,  
324 roughly the same amount of location measurements were successful across the solar cycle, whereas in  
325 summer, the fish were significantly easier to detect (higher yield) at night (Fig. 2A). This is surprising  
326 because we simultaneously observed the carp to swim closer to the shore during the night (lowered  $d$ ,  
327 Fig. 2B), where we would expect more signal attenuation due to reed growth, suggesting that the carp  
328 actively swam in denser vegetation during daytime in the summer, moving to the sublittoral areas  
329 during night where detectability increased.

330        **In summer, interactions are less frequent but more persistent**

331        In warmer months, when the deep zones of the lake become anoxic and the food was  
332 concentrated in the littoral zone, the carp had little to gain foraging-wise from swimming at the lake  
333 centre or in the deep water in the lake centre. Instead, they resided in the shallow waters alongside the  
334 shore where they could seek protection from predators among the food-rich reeds without investing  
335 much energy in mobility (lowered  $d$ ,  $S$ , and  $v$ , Fig. 2B). This introduced a technical inconvenience as reed  
336 attenuated acoustic signals, which caused the data yield to drop substantially during summer (Fig. 2A).  
337 Despite this, we can report the somewhat surprising result that during the summer when carp  
338 associated strongly with the vegetated littoral zone, they interacted less often (lowered  $p$ , Fig. 2C) but  
339 also spend less time apart between interactions (lowered  $\tau^-$ , Fig. 2C). Additionally,  $\tau^+$ , which we should  
340 expect to drop across the summer due signal attenuation, appeared, although noisy, stable throughout  
341 the year. This suggests that inside the reed, fish were interacting in small persistent groups.

342        **Social attraction or co-use of suitable habitats?**

343        **Social attraction was a key driver of social interactions**

344        We found that across the year, there were one to six times more interactions in the real data  
345 than in the shuffled data (Fig. 3B). This is strong evidence that social attraction was, in most months of  
346 the year, a key driving mechanism for proximity interaction of carp. The effect sizes were greatest at the  
347 start of the year in winter and decreased towards autumn. The overall number of interactions in the real  
348 data dropped as well (Fig. 3A, note that y-axis is log-scaled). We also observed that effect sizes were  
349 larger during daytime than at night (Fig. 3B). This indicates that nightly mobility (or *stationarity*) took  
350 place in the same locations over at least a week and was less driven by where other fish spend their  
351 nights.

352        **Elevated local clustering revealed tight knit community structure**

353        We found that clustering was higher during daytime than at nighttime in most months (Fig. 3).  
354 Clustering varied over the year, where we observed a decline in clustering from winter through summer  
355 and then an increase from autumn to winter, where by December, clustering had returned to the same  
356 level as in January, in spite of the number of interactions being many times smaller. Nighttime clustering

357 also varied periodically in synchrony with the seasons. We measured the effect size of clustering by  
358 comparing to our null model, and observed that the effect sizes decreased over the year, to the point  
359 where we could not confidently state that any of the observed local clustering was due to social  
360 attraction. At the same time, however, the local clustering was just as high at the end of the year as it  
361 was in the beginning, leading us to reason that over the year the fish grew more habitual, visiting the  
362 same few locations every day at regular times.

### 363 Network structure at varying time-scales

#### 364 Social memory varied over the year

365 If the carp population had no collective social memory they would not form clusters, but instead  
366 mix randomly, and any emerging communities in the interaction network would be due to simultaneous  
367 space use. We studied the interaction networks that emerged when aggregating over time-windows of  
368 different duration (subtracting the corresponding random network produced by the null model), and  
369 measured the number of communities with three or more members as a proxy for the use of social  
370 memory in interactions (Fig. 4). We reason that when it is possible to aggregate over a long time window  
371 and still obtain a number of communities, there is a high degree of social memory in the population.  
372 Furthermore, the longest aggregation that does not wash out communities is a good estimate of the  
373 time-scale of social memory. We found that social memory was significantly more time-persistent in the  
374 summer where, for example, the number of detected communities at night peeked at the two week  
375 aggregation for months May and August. Reversely, in winter months like January and December, the  
376 network mixed on very short time-scales. This is not surprising, as we know this mixing happened due to  
377 shoaling, however, since the fish only shoaled during daytime we were surprised to find that that this  
378 mixing to a large extent persisted into the night.

#### 379 Night communities were numerous and had longer time-scales

380 In nine out of 12 months, the number of communities during the night was higher than in  
381 daytime. This is not surprising because the fish interacted much more during the day, which caused  
382 mixing between existing groups on shorter time-scales. Curiously, however, the measured time-scales  
383 (aggregation time where number of communities peaks and then drops) of nightly communities were in  
384 many cases longer than those of day communities.

### 385 Discussion

386 We observed that the social behaviour of a population of carp was highly dynamic across daily  
387 and seasonal scales, where the clustering and social memory was oscillating based on interactions  
388 between diurnal and seasonal rhythms. Thus, inferring social network information from time snapshots  
389 of data in selected weeks or months can lead to biased conclusions. Our analysis also showed that the  
390 groups of carp were indeed aggregating because of social attraction and the carp were likely to spend  
391 time in characteristic groups with the same individuals for up to two weeks particularly during the  
392 nighttime in the summer, indicative of pronounced social memory. We finally observed higher clustering  
393 during the daytime than the nighttime, and surprisingly large shoals of actively moving carp in the

394 wintertime during daytime, which as discussed below is likely a response to environmental conditions in  
395 the lake and the origin of the fish. Methodologically, our analysis can serve as a template for future  
396 investigations into the causes and consequences of social behaviour in the wild as more high-resolution  
397 movement data in the wild begins to emerge. Our findings can also help inform both sampling design  
398 (Bajer & Sorensen, 2012; Muška et al., 2018) or removal strategies (Bajer et al., 2011; Carl, Weber &  
399 Brown, 2016) for common carp in the wild.

400 We found highly variable social behaviour of our carp population in the wild over the course of a  
401 full year. The carp aggregated in somewhat deeper zones in wintertime in agreement with previous  
402 natural history knowledge of the species (Armstrong et al., 2016; Bajer et al., 2011; Bauer & Schlott,  
403 2004; Gusar, 1989; Johnsen & Hasler, 1977; Penne & Pierce, 2008; Taylor et al., 2012), however, these  
404 zones did not encompass the deepest points of the lake. Previous work in other systems with low  
405 resolution telemetry has reported carp were largely sedentary during the wintertime with only localized  
406 activity (Bauer & Schlott, 2004; Johnsen & Hasler, 1977; Jurajda et al., 2016; Penne & Pierce, 2008), in  
407 stark contrast to the high levels of daytime activity we observed during winter in Kleiner Döllnsee.  
408 Elevated levels of overwintering activity of carp are usually attributed to stressors, such as low oxygen  
409 (Bauer & Schlott, 2004), movement of humans on the ice surface (Johnsen & Hasler, 1977), or predators  
410 (Adámek, Sukop, Moreno Rendón, & Kouřil, 2003). In our study, hypoxic conditions was an unlikely  
411 stressor as Kleiner Döllnsee is well oxygenated in the wintertime. Moreover, the fish were not manually  
412 tracked eliminating human on-ice activity as a possible disturbance. Therefore, reasons for the  
413 surprisingly active shoaling behaviour are likely related to “hypothetical” predator avoidance behavior  
414 shown by the introduced carp in their new, unfamiliar environment particularly during daytime in  
415 response to visual predators. Shoaling is a common response to predation risk (Pitcher, 1986) as it offers  
416 increased predator detection probability (Godin, Classon, & Abrahams, 1988) and dilution of risk  
417 (Queiroz & Magurran, 2005), and shoaling can confuse predators (Krakauer, 1995). The carp we tracked  
418 were still likely responsive to predation risk despite the fact that they were large enough to escape the  
419 threat of predation from most predators in the lake (Gaeta et al., 2018). Possible predators include  
420 European catfish (Carol, Benejam, Benito, & García-Berthou, 2009), great cormorants (*Phalacrocorax*  
421 *carbo*) (Adámek, Kucerova & Roche, 1999) or mustelids such as the European otter (*Lutra lutra*)  
422 (Adámek et al., 2003; Britton, Pegg, Shepherd, & Toms, 2006). Indeed, carp in a reservoir were observed  
423 to spend their nights in deep hypoxic waters, which was also speculated to be a response to a  
424 population of wels catfish present in the reservoir, which hunt nocturnally (Benito et al., 2015). We  
425 found that the carp reduced their shoaling behaviour in the second winter in our study, after they were  
426 more familiar with the (rather low predation) risks of the novel environment, which lends further  
427 support to the idea that daytime winter shoaling in the first winter after introduction to their new  
428 environment was likely a predator-avoidance response driven by unfamiliarity with possible predation  
429 threats during daytime in winter. We expect that carp should behave cautiously in a novel lake  
430 environment, despite a lack of strong predation risk due to their larger size (Lorenzen, 2000). The anti-  
431 predator behaviour could become fixed from early life experience or from the evolutionary past  
432 (Blumstein, 2006; Magurran, 1990; Swaney, Cabrera-Álvarez, & Reader, 2015). When compared to  
433 behaviour in a controlled laboratory environment conducted in large tanks, carp have been observed to  
434 behave highly cautiously after introduction to a predator free semi-natural pond environment of

435 comparable size, where the fish reduced their visits to open feeding sites (Klefeth et al., 2012). In the  
436 case of the experiment by Klefeth et al. (2012) water from a nearby lake was flowing through the semi-  
437 natural ponds, and therefore the carp were likely exposed to chemical cues signalling that predators  
438 could be present. Hence, it is reasonable that the fish tracked in our experiment, which were recently  
439 introduced to the lake, could be behaving cautiously, despite a low actual predation risk. After learning  
440 about the true predation threat in the lake was likely low (in fact no otters were recently seen in the  
441 lake and cormorant predation is low too), the carp likely behaved less cautiously in the second winter  
442 and reduced their daytime social behavior substantially.

443 The carp were more likely to shoal in the daytime than the nighttime throughout most of the  
444 year. As the daytime clustering corresponded with offshore movements, into riskier habitat outside  
445 plant refuges, the daytime clustering may partially be explained as a response to visual predators such  
446 as cormorants (White et al., 2008) or pike (Eklöv, 1992) that actively hunt during daytime in the study  
447 lake (Kobler et al., 2008). Diurnal migrations, in particular diurnal vertical migrations, are also a well  
448 known behavioural response displayed by many smaller bodied fish species, macroinvertebrates and  
449 zooplankton, where individuals balance bioenergetic efficiency, foraging opportunities and predation  
450 risk, by sheltering in deeper water and foraging in shallower waters when predators are less active  
451 (Mehner, 2012). Some cyprinids have also been observed to migrate horizontally on a daily scale  
452 (Kubečka, 1993; Nakayama et al., 2018), as we observed the carp doing in our study. Similar to the carp,  
453 common bream (Schulz & Berg, 1987) and freshwater drum, *Aplodinotus grunniens* (Rypel & Mitchell,  
454 2007), have been observed to move from littoral habitats during the night to pelagic habitats during the  
455 daytime. Further a whole fish assemblage in a Czech reservoir aggregated in the pelagic during the  
456 daytime and spread out in the littoral during the nighttime (Muška et al., 2013, 2018). Other small  
457 cyprinid species have been observed to migrate horizontally in the opposite direction to our tracked  
458 carp, sheltering in the littoral during the daytime and foraging in the pelagic at nighttime (Haertel &  
459 Eckmann, 2002; Nakayama et al., 2018). The ultimate mechanism for diurnal horizontal migrations and  
460 in which direction they occur in these species is not known, but it is expected to also relate to a tradeoff  
461 between resource availability and predation risk (Rypel & Mitchell, 2007; Schulz & Berg, 1987; Shoup,  
462 Boswell, & Wahl, 2014). Hence, in the summertime the carp may be foraging in the littoral habitat  
463 during the night-time, avoiding predation risk from the pelagic catfish (Benito et al., 2015; Carol,  
464 Zamora, & García-Berthou, 2007) and foraging in more pelagic habitat during the daytime, while forming  
465 shoals to reduce perceived predation risk from pike or otters.

466 We found longer-lasting and smaller groups of fish during the summertime, in particular during  
467 the nighttime. Our comparisons to a null model of behaviour (Spiegel et al., 2016) indicated that these  
468 clusters were not driven by attraction to similar locations at similar times, but were truly a result of  
469 attraction to the individual carp. We found that the carp during the summertime had a pronounced  
470 social memory, showing preferences to interact with certain individuals for up to two weeks.  
471 Importantly, two weeks is longer than it took the twelve days for guppies, *Poecilia reticulata*, to learn  
472 and retain the identity of conspecifics (Griffiths & Magurran, 1997); hence it is likely that the carp were  
473 able to remember the identities of their conspecifics. Furthermore, although the carp spent their days  
474 mixing in larger groups, they tended to spend the nights together in smaller characteristic groups. Carp

475      anglers have long noted that carp have “friends”, as many have observed that after catching one specific  
476      individual carp a second individual is predictably captured (Hearn, 2000). In juvenile fish preferred  
477      interactions have can be based on kinship (Piyapong et al., 2011); however, kinship is typically not the  
478      case in adult fish (Croft et al., 2012; Russell, Kelley, Graves, & Magurran, 2004), and therefore it is likely  
479      that the groups we observed were not based on kinship. In the summer months, Kleiner Döllnsee is  
480      productive and resource-rich; however, the spatiotemporal distribution of food is patchy across the  
481      lake. In such patchy environment, carp should benefit from information sharing (Monk et al., 2018) to  
482      find resources faster. Indeed, when feed bags were introduced into a lake the whole population of carp  
483      were able to find a feed bag within four nights, much faster than possible by individual private searching  
484      (Bajer et al., 2010), suggesting social learning and other forms of communication (e.g., chemical  
485      communication through excretion; Brönmark, & Hansson, 2000). Therefore we suggest that the small  
486      groups of carp we found in the summer may provide them with valuable information regarding resource  
487      distribution in the lake. Carp are also quick to learn from trained demonstrators (Karplus et al., 2007;  
488      Klefoth et al., 2012; Zion et al., 2007), and can retain socially learned information for at least one year  
489      as, population-wide catch rates were found to decline for an entire year after angling a pond despite  
490      most individuals in the population never having been captured (Beukema, 1970).

491           Recognizing specific individuals by social memory may provide additional foraging benefits to  
492      the carp in our study. Familiarity in general is known to provide fitness benefits (Seppä, Laurila,  
493      Peuhkuri, Piironen, & Lower, 2011), in particular via the increased foraging success through directed  
494      social learning as fish may learn better from familiar individuals (Swaney et al., 2015). Hence, knowing  
495      and following the most productive foragers in the population should be beneficial, especially in the  
496      context of producer scrounger dynamics, where some individuals may be generally better at finding new  
497      food sources through private information, while other individuals tend to follow those individuals to find  
498      food (Caraco & Giraldeau, 1991). As well, carp are known to show consistent inter-individual differences  
499      in foraging rate (Pollux, 2017) and individual variation in diet (Mehner et al., 2018) and taste preferences  
500      (Kasumyan, 2000). Consequently, certain individual fish may have better information regarding certain  
501      food sources within the patchy resource distribution in a natural lake, providing fitness benefits to  
502      sociality during resource-rich environmental conditions in the warmer periods of the year.

503

#### 504      Conclusion

505           We found that carp adjust their social behaviour following several timescales of oscillation,  
506      where yearly and daily oscillations very likely respond to variation in perceived predation risk and  
507      resource availability, moderated by the benefits of social interactions. Despite a low realized threat of  
508      predation in our study lake given the large size of the tagged carp, the fish displayed cautious behaviour  
509      after introduction into a novel environment, particularly during the day and in winter by revealing strong  
510      tendencies of shoaling. Further, the carp displayed pronounced social interactions based on social  
511      attraction with the community organization being non-random and based on a social memory. These  
512      findings strongly indicated that carp are able to recognize one another and to take advantage of  
513      familiarity during productive phases in warmer months of the year. To our knowledge, our work is

514 among the first year long high-resolution analyses of animal social networks in the wild. Our analysis  
515 may serve as a methodological template for future analyses into the rhythm of relationships in other  
516 species and taxa.

517 **Acknowledgements**

518 This work was supported by a Leibniz Community (“B-types”, SAW-2013-IGB-2) grant received  
519 by RA. This work was also supported by Strategic Grants by Princeton University and Humboldt-  
520 Universität zu Berlin on “Princeton-Humboldt Centre for the Reality Mining of Animal Systems,” the  
521 ‘Cooperation and Collective Cognition Network’. We would like to thank Andreas Mühlbradt, Alexander  
522 Türck, Jan Hallermann, Jacob Weinrautner, Jonathan Nickl, Bernard Chéret and many other technicians  
523 and students for help in the field and processing the data. The experiments were approved through  
524 animal care permit (2347-21-2014) granted by the Ministry of Environment, Health and Consumer  
525 Protection Brandenburg, according to the German Animal Protection Act.

526 **References**

- 527 Adámek, Z., Sukop, I., Moreno Rendón, P., & Kouřil, J. (2003). Food competition between 2 +  
528 tench (*Tinca tinca* L.), common carp (*Cyprinus carpio* L.) and bigmouth buffalo (*Ictiobus*  
529 *cyprinellus* Val.) in pond polyculture. *Journal of Applied Ichthyology*, 19(3), 165–169.
- 530 Adámek, Z., Kucerova, M., & Roche, K. (1999) The role of common carp, *Cyprinus carpio*, in the  
531 diet of piscivorous predators-cormorant (*Phalacrocorax carbo*) and otter (*Lutra lutra*).  
532 *Bulletin VURH Vodnany (Czech Republic)*, 35(4), 185-193.
- 533 Arlinghaus, R., & Mehner, T. (2003). Socio-economic characterisation of specialised common  
534 carp (*Cyprinus carpio* L.) anglers in Germany, and implications for inland fisheries  
535 management and eutrophication control. *Fisheries Research*, 61(1), 19–33.
- 536 Armstrong, D. E., Hennen, M. J., Saunders, C. P., Brandenburger, T., Brown, M. L., & Neumann,  
537 C. (2016). Modeling common carp under-ice movement using hierarchical Markov  
538 simulation. *Ecological Modelling*, 334, 44–50.
- 539 Bajer, P. G., Beck, M. W., Cross, T. K., Koch, J. D., Bartodziej, W. M., & Sorensen, P. W. (2016).  
540 Biological invasion by a benthivorous fish reduced the cover and species richness of  
541 aquatic plants in most lakes of a large North American ecoregion. *Global Change Biology*,  
542 22(12), 3937–3947.
- 543 Bajer, P. G., Chizinski, C. J., & Sorensen, P. W. (2011). Using the Judas technique to locate and  
544 remove wintertime aggregations of invasive common carp. *Fisheries Management and*  
545 *Ecology*, 18(6), 497–505.
- 546 Bajer, P. G., Lim, H., Travallie, M. J., Miller, B. D., & Sorensen, P. W. (2010). Cognitive aspects of  
547 food searching behavior in free-ranging wild Common Carp. *Environmental Biology of*  
548 *Fishes*, 88(3), 295–300.

- 549 Bajer, P. G., & Sorensen, P. W. (2012). Using boat electrofishing to estimate the abundance of  
550 invasive common carp in small Midwestern lakes. *North American Journal of Fisheries  
551 Management*, 32(5), 817–822.
- 552 Bajer, P. G., Sullivan, G., & Sorensen, P. W. (2009). Effects of a rapidly increasing population of  
553 common carp on vegetative cover and waterfowl in a recently restored Midwestern  
554 shallow lake. *Hydrobiologia*, 632(1), 235–245.
- 555 Baktoft, H., Zajicek, P., Klefoth, T., Svendsen, J. C., & Jacobsen, L. (2015). Performance  
556 assessment of two whole-lake acoustic positional telemetry systems - is reality mining of  
557 free-ranging aquatic animals technologically possible? *PLoS ONE*, 10(5), e0126534.
- 558 Balon, E. K. (2004). About the oldest domesticates among fishes. *Journal of Fish Biology*, 65, 1–  
559 27.
- 560 Bauer, C., & Schlott, G. (2004). Overwintering of farmed common carp (*Cyprinus carpio* L.) in  
561 the ponds of a central European aquaculture facility - Measurement of activity by radio  
562 telemetry. *Aquaculture*, 241(1–4), 301–317.
- 563 Benito, J., Benejam, L., Zamora, L., & García-Berthou, E. (2015). Diel cycle and effects of water  
564 flow on activity and use of depth by common carp. *Transactions of the American Fisheries  
565 Society*, 144(3), 491–501.
- 566 Bernard, S., Gonze, D., Čajavec, B., Herzl, H., & Kramer, A. (2007). Synchronization-induced  
567 rhythmicity of circadian oscillators in the suprachiasmatic nucleus. *PLoS Computational  
568 Biology*, 3(4), 667–679.
- 569 Bertram, B. C. R. (1978). Living in groups: predators and prey. *Behavioural Ecology: An  
570 Evolutionary Approach*, 221–248.
- 571 Beukema, J. J. (1970). Angling experiment with carp (*Cyprinus carpio* L.), II. Decreasing  
572 catchability through one-trial learning. *Netherlands Journal of Zoology*, 20, 81–92.
- 573 Blonder, B., Wey, T. W., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and  
574 network analysis. *Methods in Ecology and Evolution*, 3(6), 958–972.
- 575 Blumstein, D. T. (2006). The multipredator hypothesis and the evolutionary persistence of  
576 antipredator behavior. *Ethology*, 112(3), 209–217.
- 577 Britton, J. R., Pegg, J., Shepherd, J. S., & Toms, S. (2006). Revealing the prey items of the otter  
578 *Lutra lutra* in South West England using stomach contents analysis. *FOLIA ZOOLOGICA-  
579 PRAHA-*, 55(2), 167.
- 580 Brönmark, C., & Hansson, L. (2000). Chemical communication in aquatic systems: and  
581 introduction. *Oikos*, 88(1), 103–109.

- 582 Brown, C., & Laland, K. N. (2002). Social learning of a novel avoidance task in the guppy:  
583 Conformity and social release. *Behaviour*, 64(1), 41–47.
- 584 Brown, J. (1968). The evolution of diversity in avian territorial systems. *The Wilson Bulletin*,  
585 76(2), 160–169.
- 586 Bryant, M. J., & Grant, J. W. A. (1995). Resource defence, monopolization and variation of  
587 fitness in groups of female Japanese medaka depend on the synchrony of food arrival.  
588 *Animal Behaviour*, 49(6), 1469–1479.
- 589 Bshary, R., Gingins, S., & Vail, A. L. (2014). Social cognition in fishes. *Trends in Cognitive  
590 Sciences*, 18(9), 465–471.
- 591 Caraco, T., & Giraldeau, L.-A. (1991). Social foraging: Producing and scrounging in a stochastic  
592 environment. *Journal of Theoretical Biology*, 153(4), 559–583.
- 593 Carl, D. D., Weber, M. J., & Brown, M. L. (2016). An evaluation of attractants to increase catch  
594 rates and deplete age-0 common carp in shallow South Dakota Lakes. *North American  
595 Journal of Fisheries Management*, 36(3), 506–513.
- 596 Carol, J., Zamora, L., & García-Berthou, E. (2007). Preliminary telemetry data on the movement  
597 patterns and habitat use of European catfish (*Silurus glanis*) in a reservoir of the River  
598 Ebro, Spain. *Ecology of Freshwater Fish*, 16(3), 450–456.
- 599 Carol, J., Benejam, L., Benito, J., & García-Berthou, E. (2009). Growth and diet of European  
600 catfish (*Silurus glanis*) in early and late invasion stages. *Fundamental and Applied  
601 Limnology / Archiv Für Hydrobiologie*, 174(4), 317–328.
- 602 Conover, D. O., & Present, T. M. C. (2016). Countergradient variation in growth rate :  
603 compensation for length of the growing season among Atlantic silversides from different  
604 latitudes. *Oecologia*, 83(3), 316–324.
- 605 Coolen, I., Ward, A. J. W., Hart, P. J. B., & Laland, K. N. (2005). Foraging nine-spined sticklebacks  
606 prefer to rely on public information over simpler social cues. *Behavioral Ecology*, 16(5),  
607 865–870.
- 608 Côté, I. M., Poulin, R., & Zealand, N. (1985). Parasitism and group size in social animals : a meta-  
609 analysis. *Behavioral Ecology*, 6(1), 159–165.
- 610 Croft, D. P., James, R., Thomas, P. O. R., Hathaway, C., Mawdsley, D., Laland, K. N., & Krause, J.  
611 (2006). Social structure and co-operative interactions in a wild population of guppies  
612 (*Poecilia reticulata*). *Behavioral Ecology and Sociobiology*, 59(5), 644–650.
- 613 Croft, D. P., Hamilton, P. B., Darden, S. K., Jacoby, D. M. P., James, R., Bettaney, E. M., & Tyler, C.  
614 R. (2012). The role of relatedness in structuring the social network of a wild guppy  
615 population. *Oecologia*, 170(4), 955–963.

- 616 Daniel, A. J., Hicks, B. J., Ling, N., & David, B. O. (2009). Acoustic and radio-transmitter retention  
617 in common carp (*Cyprinus carpio*) in New Zealand. *Marine and Freshwater Research*, 60(4),  
618 328–333.
- 619 Dunlap, J. C. (1999). Molecular Bases for Circadian Clocks. *Cell*, 96, 271–290.
- 620 Eckmann, R. (1995). Fish species richness in lakes of the northeastern lowlands in Germany.  
621 *Ecology of Freshwater Fish*, 4(2), 62–69.
- 622 Edgar, R. S., Green, E. W., Zhao, Y., Ooijen, G. Van, Qin, X., Xu, Y., ... Akhilesh, B. (2012). Europe  
623 PMC Funders Group Peroxiredoxins are conserved markers of circadian rhythms. *Nature*,  
624 485(7399), 459–464.
- 625 Eklöv, P. (1992). Group foraging versus solitary foraging efficiency in piscivorous predators: the  
626 perch, *Perca fluviatilis*, and pike, *Esox lucius*, patterns. *Animal Behaviour*, 44, 313–326.
- 627 FAO. (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable  
628 development goals*. Rome.
- 629 Finn, J. T., Brownscombe, J. W., Haak, C. R., Cooke, S. J., Cormier, R., Gagne, T., & Danylchuk, A.  
630 J. (2010). Applying network methods to acoustic telemetry data: Modeling the movements  
631 of tropical marine fishes. *Ecological Modelling*, 293(January), 139–149.
- 632 Foster, W. A., & Treherne, J. E. (1981). Evidence for the dilution effect in the selfish herd from  
633 fish predation on a marine insect. *Nature*, 293(5832), 466.
- 634 Franks, D. W., Croft, D. P., van Ginneken, A., Morrell, L. J., Balcomb, K. C., Foster, E. A., &  
635 Parsons, K. M. (2012). Social network correlates of food availability in an endangered  
636 population of killer whales, *Orcinus orca*. *Animal Behaviour*, 83(3), 731–736.
- 637 Gaeta, J. W., Ahrenstorff, T. D., Diana, J. S., Fetzer, W. W., Jones, T. S., Lawson, Z. J., ... Vander  
638 Zanden, J. M. (2018). Go big or ...don't? A field-based diet evaluation of freshwater  
639 piscivore and prey fish size relationships. *Plos One*, 13(3), e094092.
- 640 Gilbert, C., Robertson, G., Le Maho, Y., Naito, Y., & Ancel, A. (2006). Huddling behavior in  
641 emperor penguins: dynamics of huddling. *Physiology & Behavior*, 88(4–5), 479–488.
- 642 Godin, J.-G. J., Classon, L. J., & Abrahams, M. V. (1988). Group vigilance and shoal size in a small  
643 characin fish. *Behaviour*, 104(1/2), 29–40.
- 644 Granroth-Wilding, H. M. V., & Magurran, A. E. (2013). Asymmetry in pay-off predicts how  
645 familiar individuals respond to one another. *Biology Letters*, 9(3).
- 646 Griffiths, S. W., & Magurran, A. E. (1997). Familiarity in schooling fish: how long does it take to  
647 acquire? *Animal Behaviour*, 53(5), 945–949.

- 648 Grosenick, L., Clement, T. S., & Fernald, R. D. (2007). Fish can infer social rank by observation  
649 alone. *Nature*, 445(7126), 429.
- 650 Gusar, A. G. (1989). The results of ultrasonic telemetry of the carp, *Cyprinus carpio*, in a  
651 wintering pond during the winter period. *Fol. Zool.*, 38, 87–95.
- 652 Guzzo, M. M., Van Leeuwen, T. E., Hollins, J., Koeck, B., Newton, M., Webber, D. M., ... Killen, S.  
653 S. (2018). Field testing a novel high residence positioning system for monitoring the fine-  
654 scale movements of aquatic organisms. *Methods in Ecology and Evolution*, 9(6), 1478–  
655 1488.
- 656 Haertel, S.S., & Eckmann, R. (2002). Diel diet shift of roach and its implications for the  
657 estimation of daily rations. *Journal of Fish Biology*, 60, 876-892.
- 658 Hastings, A. (2010). Timescales, dynamics, and ecological understanding. *Ecology*, 91(12), 3471–  
659 3480.
- 660 Hay, D. E., & McKinnell, S. M. (2002). Tagging along: association among individual Pacific herring  
661 (*Clupea pallasi*) revealed by tagging. *Canadian Journal of Fisheries and Aquatic Sciences*,  
662 59(12), 1960–1968.
- 663 Hearn, T. (2000). *In Pursuit of the Largest*. Bountyhunter Publication. Hampshire, p 318.
- 664 Helfman, G. S. (1984). School fidelity in fishes: The yellow perch pattern. *Animal Behaviour*,  
665 32(3), 663–672.
- 666 Hennen, M.J. & Brown, M.L. (2014). Movement and spatial distribution of common carp in a  
667 South Dakota glacial lake system: Implications for management and removal. *North  
668 American Journal of Fisheries Management*. 34(6), 1270-1281.
- 669 Henzi, S. P., Lusseau, D., Weingrill, T., Van Schaik, C. P., & Barrett, L. (2009). Cyclicity in the  
670 structure of female baboon social networks. *Behavioral Ecology and Sociobiology*, 63(7),  
671 1015–1021.
- 672 Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., Bateson, M., ... Wolfe, J. W. (2015).  
673 Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*,  
674 19(1), 46–54.
- 675 Hoare, D. J., Couzin, I. D., Godin, J. G. J., & Krause, J. (2004). Context-dependent group size  
676 choice in fish. *Animal Behaviour*, 67(1), 155–164.
- 677 Holland, K. N., Brill, R. W., Chang, R. K. C., Sibert, J. R., & Fournier, D. A. (1992). Physiological  
678 and behavioural thermoregulation in bigeye tuna (*Thunnus obesus*). *Nature*, 358(6385),  
679 410.
- 680 Howes, G. J. (1991). Systematics and biogeography: an overview. In *Cyprinid fishes* (pp. 1–33).

- 681 Springer.
- 682 Hunt, R. J., Jardine, T. D., Hamilton, S. K., & Bunn, S. E. (2012). Temporal and spatial variation in  
683 ecosystem metabolism and food web carbon transfer in a wet-dry tropical river.  
684 *Freshwater Biology*, 57(3), 435–450.
- 685 Jacoby, D. M. P., & Freeman, R. (2016). Emerging network-based tools in movement ecology.  
686 *Trends in Ecology and Evolution*, 31(4), 301–314.
- 687 Johnsen, P. B., & Hasler, A. D. (1977). Winter aggregations of carp (*Cyprinus carpio*) as revealed  
688 by ultrasonic tracking. *Transactions of the American Fisheries Society*, 106(6), 556–559.
- 689 Jones, M. J., & Stuart, I. G. (2009). Lateral movement of common carp (*Cyprinus carpio* L.) in a  
690 large lowland river and floodplain. *Ecology of Freshwater Fish*, 18(1), 72–82.
- 691 Jurajda, P., Adámek, Z., Roche, K., Mrkvová, M., Štarhová, D., Prášek, V., & Zukal, J. (2016). Carp  
692 feeding activity and habitat utilisation in relation to supplementary feeding in a semi-  
693 intensive aquaculture pond. *Aquaculture International*, 24(6), 1627–1640.
- 694 Karplus, I., Zion, B., Rosenfeld, L., Grinshpun, Y., Slosman, T., Goshen, Z., & Barki, A. (2007).  
695 Social facilitation of learning in mixed-species schools of common carp *Cyprinus carpio* L.  
696 and Nile tilapia *Oreochromis niloticus* (L.). *Journal of Fish Biology*, 71(4), 1023–1034.
- 697 Kasumyan, A. O. (2000). Individual variability of taste preferences and of behavioural response  
698 to taste stimuli in carp *Cyprinus carpio*. *Journal of Ichthyology*, 40(8), 661–669.
- 699 Klefoth, T., Kobler, A., & Arlinghaus, R. (2008). The impact of catch-and-release angling on  
700 short-term behaviour and habitat choice of northern pike (*Esox lucius* L.). *Hydrobiologia*,  
701 601, 99–110.
- 702 Klefoth, T., Pieterik, T., & Arlinghaus, R. (2013). Impacts of domestication on angling  
703 vulnerability of common carp, *Cyprinus carpio*: The role of learning, foraging behaviour and  
704 food preferences. *Fisheries Management and Ecology*, 20(2–3), 174–186.
- 705 Klefoth, T., Skov, C., Krause, J., & Arlinghaus, R. (2012). The role of ecological context and  
706 predation risk-stimuli in revealing the true picture about the genetic basis of boldness  
707 evolution in fish. *Behavioral Ecology and Sociobiology*, 66(4), 547–559.
- 708 Klefoth, T., Skov, C., Kuparinen, A., & Arlinghaus, R. (2017). Towards a mechanistic  
709 understanding of vulnerability to hook-and-line fishing: boldness as the basic target of  
710 angling-induced selection. *Evolutionary Applications*, 1–13.
- 711 Klimley, A. P., & Holloway, C. F. (1999). School fidelity and homing synchronicity of yellowfin  
712 tuna, *Thunnus albacares*. *Marine Biology*, 133, 307–317.
- 713 Kobler, A., Klefoth, T., Wolter, C., Fredrich, R., & Arlinghaus R. (2008). Contrasting movement

- 714 and habitat choice between summer and winter in a small lake. *Hydrobiologia*, 601, 17–27.
- 715 Kohsaka, A., & Bass, J. (2007). A sense of time: how molecular clocks organize metabolism.  
716 *Trends in Endocrinology & Metabolism*, 18(1), 4–11.
- 717 Krakauer, D. C. (1995). Groups confuse predators by exploiting perceptual bottlenecks: a  
718 connectionist model of the confusion effect. *Behavioral Ecology and Sociobiology*, 36(6),  
719 421–429.
- 720 Krause, J., Krause, S., Arlinghaus, R., Psorakis, I., Roberts, S., & Rutz, C. (2013). Reality mining of  
721 animal social systems. *Trends in Ecology and Evolution*, 28(9), 541–551.
- 722 Kubečka, J. (1993). Night inshore migration and capture of adult fish by shore seining.  
723 *Aquaculture Research*, 24(5), 685–689.
- 724 Landeau, L., & Terborgh, J. (1986). Oddity and the ‘confusion effect’ in predation. *Animal  
725 Behaviour*, 34(5), 1372–1380.
- 726 Leatherland, J. F., & McKeown, B. A. (1973). Circadian rhythm in the plasma levels of prolactin  
727 in goldfish, *Carassius auratus* L. *Biological Rhythm Research*, 4(2), 137–143.
- 728 Lefebvre, L., Palameta, B., & Hatch, K. K. (1996). Is group-living associated with social learning?  
729 A comparative test of a gregarious and a territorial columbid. *Behaviour*, 133(3/4), 241–  
730 261.
- 731 Sutter, D. A., Suski, C. D., Philipp, D. P., Klefoth, T., Wahl, D. H., Kersten, P., ... & Arlinghaus, R.  
732 (2012). Recreational fishing selectively captures individuals with the highest fitness  
733 potential. *Proceedings of the National Academy of Sciences*, 109(51), 20960–20965.
- 734 Niemelä, P. T., & Dingemanse, N. J. (2014). Artificial environments and the study of ‘adaptive’  
735 personalities. *Trends in Ecology & Evolution*, 29(5), 245–247.
- 736 Lennox, R. J., Aarestrup, K., Cooke, S. J., Cowley, P. D., Deng, Z. D., Fisk, A. T., ... Young, N.  
737 (2017). Envisioning the future of aquatic animal tracking: technology, science, and  
738 application. *BioScience*, 67(10), 884–896.
- 739 Lorenzen, K. (2000). Allometry of natural mortality as a basis for assessing optimal release size  
740 in fish-stocking programmes. *Canadian Journal of Fisheries and Aquatic Sciences*, 57(12),  
741 2374–2381.
- 742 Magnuson, J. J., Crowder, L. B., & Medvick, P. A. (1979). Temperature as an ecological resource.  
743 *American Zoologist*, 19, 331–343.
- 744 Magurran, A. E. (1990). The inheritance and development of minnow anti-predator behaviour.  
745 *Animal Behaviour*, 39(5), 834–842.

- 746 Magurran, A. E., & Higham, A. (1988). Information transfer across fish shoals under predator  
747 threat. *Ethology*, 78(2), 153–158.
- 748 Mehner, T. (2012). Diel vertical migration of freshwater fishes - proximate triggers, ultimate  
749 causes and research perspectives. *Freshwater Biology*, 57(7), 1342–1359.
- 750 Mehner, T., Rapp, T., Monk, C. T., Beck, M. E., Trudeau, A., Kiljunen, M., ... Arlinghaus, R. (2018).  
751 Feeding aquatic ecosystems: whole-lake experimental addition of angler's ground bait  
752 strongly affects omnivorous fish despite low contribution to lake carbon budget.  
753 *Ecosystems*, 22(2), 346–362.
- 754 Milinski, M., Kulling, D., Kettler, R., & Bern, U. (1990). Tit for tat: sticklebacks (*Gasterosteus*  
755 *aculeatus*) “trusting” a cooperating partner. *Behavioral Ecology*, 1(7), 7–11.
- 756 Monk, C. T., & Arlinghaus, R. (2017). Encountering a bait is necessary but insufficient to explain  
757 individual variability in vulnerability to angling in two freshwater benthivorous fish in the  
758 wild. *PLoS ONE*, 12(3), 1–25.
- 759 Monk, C. T., Barbier, M., Romanczuk, P., Watson, J. R., Alós, J., Nakayama, S., ... Arlinghaus, R.  
760 (2018). How ecology shapes exploitation: a framework to predict the behavioural response  
761 of human and animal foragers along exploration–exploitation trade-offs. *Ecology Letters*,  
762 21(6), 779–793.
- 763 Mourier, J., Brown, C., & Planes, S. (2017). Learning and robustness to catch-and-release fishing  
764 in a shark social network. *Biology Letters*, 13, 20160824.
- 765 Muška, M., Tušer, M., Frouzová, J., Draštík, V., Čech, M., Jůza, T., ... Kubečka, J. (2013). To  
766 migrate, or not to migrate: Partial diel horizontal migration of fish in a temperate  
767 freshwater reservoir. *Hydrobiologia*, 707(1), 17–28.
- 768 Muška, M., Tušer, M., Frouzová, J., Mrkvíčka, T., Ricard, D., Sedá, J., ... Kubečka, J. (2018). Real-  
769 Time distribution of pelagic fish: Combining hydroacoustics, GIS and spatial modelling at a  
770 fine spatial scale. *Scientific Reports*, 8(1), 1–11.
- 771 Nakayama, S., Doering-Arjes, P., Linzmaier, S., Briege, J., Klefoth, T., Pieterik, T., & Arlinghaus R.  
772 (2018). Fine-scale movement ecology of a freshwater top predator Eurasian perch (*Perca*  
773 *fluviatilis*), in response to the abiotic environment over the course of a year. *Ecology and*  
774 *Freshwater Fish*, 27(3), 798–812.
- 775 Naylor, E. (1988). Clock-controlled behaviour in intertidal animals. In *Behavioral adaptation to*  
776 *intertidal life* (pp. 1–14). Springer.
- 777 Økland, F., Hay, C. J., Naesje, T. F., Nickandor, N., & Thorstad, E. B. (2003). Learning from  
778 unsuccessful radio tagging of common carp in a Namibian reservoir. *Journal of Fish*  
779 *Biology*, 62(3), 735–739.

- 780 Osborne, M. W., Ling, N., Hicks, B. J., & Tempero, G. W. (2009). Movement, social cohesion and  
781 site fidelity in adult koi carp, *Cyprinus carpio*. *Fisheries Management and Ecology*, 16(3),  
782 169–176.
- 783 Parameswaran, S., Alikunhi, K. H., & Sukumaran, K. K. (1972). Observations on the maturation,  
784 fecundity and breeding of the common carp, *Cyprinus carpio* Linnaeus. *Indian Journal of  
785 Fisheries*, 19(1 & 2), 110–124.
- 786 Penne, C. R., & Pierce, C. L. (2008). Seasonal distribution, aggregation, and habitat selection of  
787 common carp in Clear Lake, Iowa. *Transactions of the American Fisheries Society*, 137(4),  
788 1050–1062.
- 789 Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., De Silva, S., ...  
790 McDonald, D. B. (2014). The dynamics of animal social networks: Analytical, conceptual,  
791 and theoretical advances. *Behavioral Ecology*, 25(2), 242–255.
- 792 Pitcher, T. J. (1986). Functions of shoaling behaviour in teleosts. In *The behaviour of teleost  
793 fishes* (pp. 294–337). Springer.
- 794 Pitcher, T. J., Magurran, A. E., & Winfield, I. J. (1982). Fish in larger shoals find food faster.  
795 *Behavioral Ecology and Sociobiology*, 10(2), 149–151.
- 796 Piaypong, C., Butlin, R. K., Faria, J. J., Scruton, K. J., Wang, J., & Krause, J. (2011). Kin assortment  
797 in juvenile shoals in wild guppy populations. *Heredity*, 106(5), 749.
- 798 Pollux, B. J. A. (2017). Consistent individual differences in seed disperser quality in a seed-eating  
799 fish. *Oecologia*, 183, 81–91.
- 800 Proske, C. (1972). Das Futteraufnahmeverhalten einsömmeriger Karpfen unter den  
801 Bedingungen der Intensivhaltung. *Bayerisches Landwirtschaftliches Jahrbuch*, 49, 899–908.
- 802 Pulliam, H. R. (1973). On the advantages of flocking. *Journal of Theoretical Biology*, 38, 419–  
803 422.
- 804 Queiroz, H., & Magurran, A. E. (2005). Safety in numbers? Shoaling behaviour of the Amazonian  
805 red-bellied piranha. *Biology Letters*, 1(2), 155–157.
- 806 Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *European Physical  
807 Journal: Special Topics*, 178(1), 13–23.
- 808 Ruf, T., & Geiser, F. (2015). Daily torpor and hibernation in birds and mammals. *Biological  
809 Reviews*, 90(3), 891–926.
- 810 Rypel, A. L., & Mitchell, J. B. (2007). Summer nocturnal patterns in freshwater drum  
811 (*Aplodinotus grunniens*). *The American Midland Naturalist*, 157(1), 230–234.

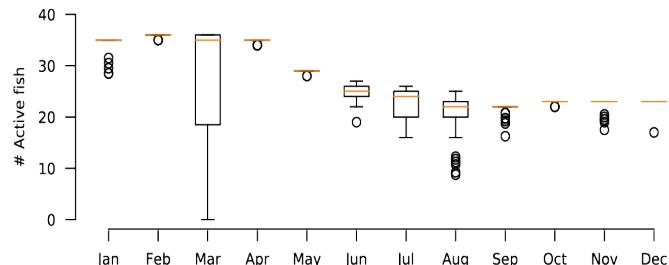
- 812 Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., & Kertesz, J. (2007). Generalizations of the  
813 clustering coefficient to weighted complex networks. *Physical Review E*, 75(2), 27105.
- 814 Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8).  
815 Cambridge university press.
- 816 Scott, A. J. (2000). *The cultural economy of cities: essays on the geography of image-producing*  
817 *industries*. Sage.
- 818 Schulz, U., & Berg, R. (1987). The migration of ultrasonic-tagged bream, *Abramis brama* (L), in  
819 Lake Constance (Bodensee-Untersee). *Journal of Fish Biology*, 31(3), 409–414.
- 820 Seppä, T., Laurila, A., Peuhkuri, N., Piironen, J., & Lower, N. (2011). Early familiarity has fitness  
821 consequences for Arctic char (*Salvelinus alpinus*) juveniles. *Canadian Journal of Fisheries*  
822 and *Aquatic Sciences*, 58(7), 1380–1385.
- 823 Seppälä, O., Karvonen, A., & Valtonen, E. T. (2008). Shoaling behaviour of fish under parasitism  
824 and predation risk. *Animal Behaviour*, 75(1), 145–150.
- 825 Shaw, E. (1978). Schooling fishes: the school, a truly egalitarian form of organization in which all  
826 members of the group are alike in influence, offers substantial benefits to its participants.  
827 *American Scientist*, 66(2), 166–175.
- 828 Shoup, D. E., Boswell, K. M., & Wahl, D. H. (2014). Diel littoral-pelagic movements by juvenile  
829 bluegills in a small lake. *Transactions of the American Fisheries Society*, 143(3), 796–801.
- 830 Shultz, E. T., Reynolds, K. E., & Conover, D. O. (1996). Countergradient variation in growth  
831 among newly hatched *Fundulus Heteroclitus*: Geographic differences revealed by common-  
832 environment experiments. *Functional Ecology*, 10(3), 366–374.
- 833 Snijders, L., Kurvers, R. H. J. M., Krause, S., Ramnarine, I. W., & Krause, J. (2018). Individual-and  
834 population-level drivers of consistent foraging success across environments. *Nature Ecology & Evolution*, 2(10), 1610.
- 836 Somveille, M., Rodrigues, A. S. L., & Manica, A. (2015). Why do birds migrate? A  
837 macroecological perspective. *Global Ecology and Biogeography*, 24(6), 664–674.
- 838 Spiegel, O., Leu, S. T., Sih, A., & Bull, C. M. (2016). Socially interacting or indifferent neighbours?  
839 Randomization of movement paths to tease apart social preference and spatial  
840 constraints. *Methods in Ecology and Evolution*, 7(8), 971–979.
- 841 Staehr, P. A., & Jensen, K. S. (2007). Temporal dynamics and regulation of lake metabolism.  
842 *Limnology and Oceanography*, 52(1), 108–120.
- 843 Suboski, M. D., & Templeton, J. J. (1989). Life skills training for hatchery fish: social learning and  
844 survival. *Fisheries Research*, 7(4), 343–352.

- 845 Swaney, W. T., Cabrera-Álvarez, M. J., & Reader, S. M. (2015). Behavioural responses of feral  
846 and domestic guppies (*Poecilia reticulata*) to predators and their cues. *Behavioural*  
847 *Processes*, 118, 42–46.
- 848 Russell, S.T., L. Kelley, J., A. Graves, J., & E. Magurran, A. (2004). Kin structure and shoal  
849 composition dynamics in the guppy, *Poecilia reticulata*. *Oikos*, 106(3), 520–526.
- 850 Taylor, A. H., Tracey, S. R., Hartmann, K., & Patil, J. G. (2012). Exploiting seasonal habitat use of  
851 the common carp, *Cyprinus carpio*, in a lacustrine system for management and  
852 eradication. *Marine and Freshwater Research*, 63(7), 587.
- 853 Vandermeer, J. (2004). Coupled oscillations in food webs: Balancing competition and mutualism  
854 in simple ecological models. *The American Naturalist*, 163(6), 857–867.
- 855 Vilizzi, L. (2012). The common carp, *Cyprinus carpio*, in the Mediterranean region: Origin,  
856 distribution, economic benefits, impacts and management. *Fisheries Management and*  
857 *Ecology*, 19(2), 93–110.
- 858 Ward, A. J. W., Botham, M. S., Hoare, D. J., James, R., Broom, M., Godin, J. G. J., & Krause, J.  
859 (2002). Association patterns and shoal fidelity in the three-spined stickleback. *Proceedings*  
860 *of the Royal Society B: Biological Sciences*, 269(1508), 2451–2455.
- 861 Weber, M. J., Brown, M. L., & Willis, D. W. (2010). Spatial variability of common carp  
862 populations in relation to lake morphology and physicochemical parameters in the upper  
863 Midwest United States. *Ecology of Freshwater Fish*, 19(4), 555–565.
- 864 White, C.R., Butler, P.J., Grémillet, D., & Martin, G.R. (2008). Behavioural strategies of  
865 cormorants (Phalacrocoracidae) foraging under challenging light conditions. *Ibis*, 150, 231–  
866 239.
- 867 Wiens, J. A. (1976). Population responses to patchy environments. *Annual Review of Ecology*  
868 *and Systematics*, 7, 81–120.
- 869 Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E., & Yovovich, V. (2015). The  
870 golden age of bio-logging: how animal-borne sensors are advancing the frontiers of  
871 ecology. *Ecology*, 96(7), 1741–1753.
- 872 Wilson, A. D. M., Croft, D. P., & Krause, J. (2014). Social networks in elasmobranchs and teleost  
873 fishes. *Fish and Fisheries*, 15(4), 676–689.
- 874 Wilson, A. D. M., Krause, S., James, R., Croft, D. P., Ramnarine, I. W., Borner, K. K., ... Krause, J.  
875 (2014). Dynamic social networks in guppies (*Poecilia reticulata*). *Behavioral Ecology and*  
876 *Sociobiology*, 68(6), 915–925.
- 877 Wittemyer, G., Douglas-Hamilton, I., & Getz, W. M. (2005). The socioecology of elephants:  
878 Analysis of the processes creating multitiered social structures. *Animal Behaviour*, 69(6),

- 879            1357–1371.
- 880    Zion, B., Barki, A., Grinshpon, J., Rosenfeld, L., & Karplus, I. (2007). Social facilitation of acoustic  
881            training in the common carp *Cyprinus carpio* (L.). *Behaviour*, 144(6), 611–630.

882    **Tables and Figures**

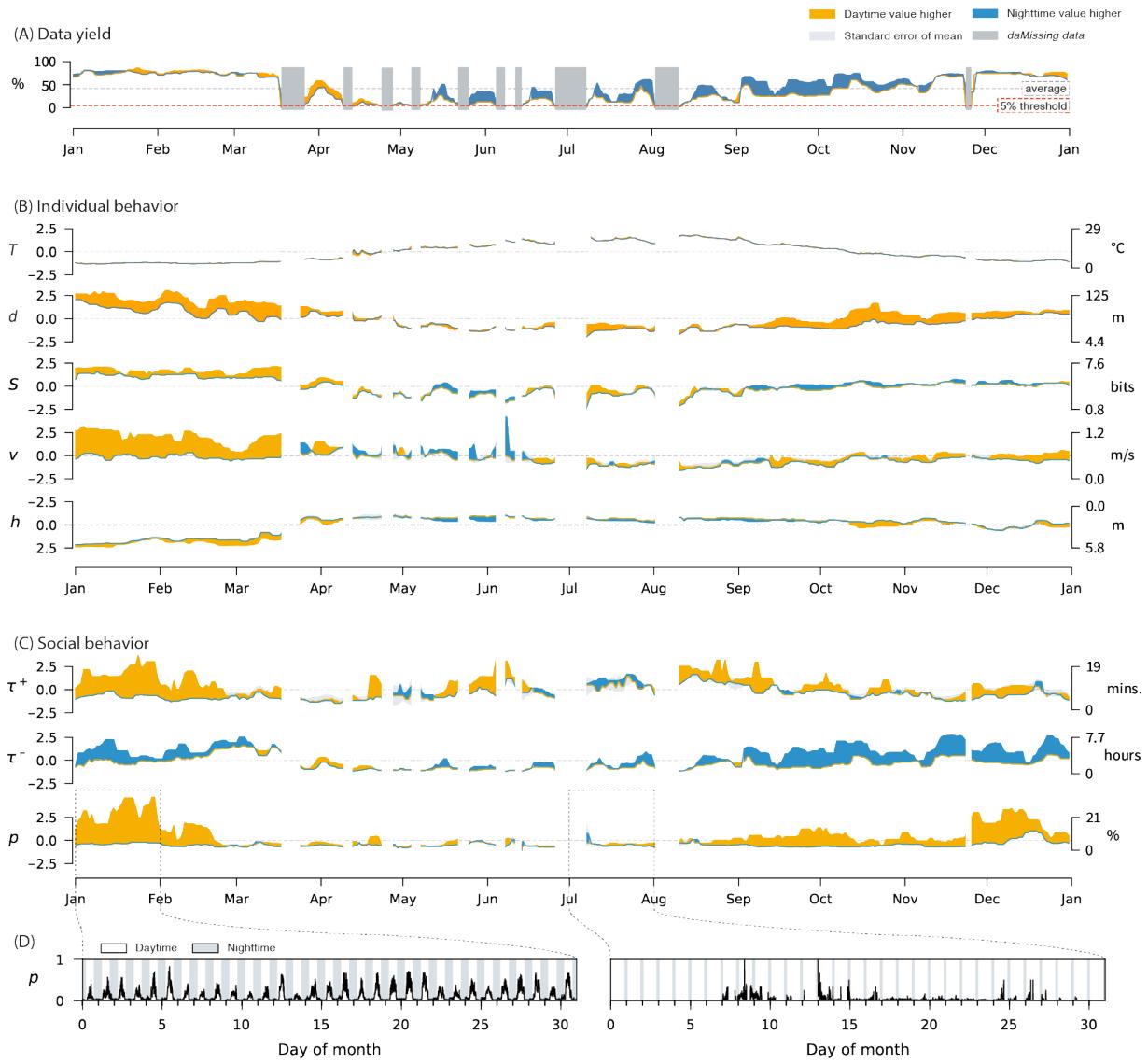
883



884

885

886    FIG. 1. Distribution of number of fish with successfully recorded mobility on each day, across  
887    the year. At the beginning of the year 36 fish are active and at the end of the year 23 are active.

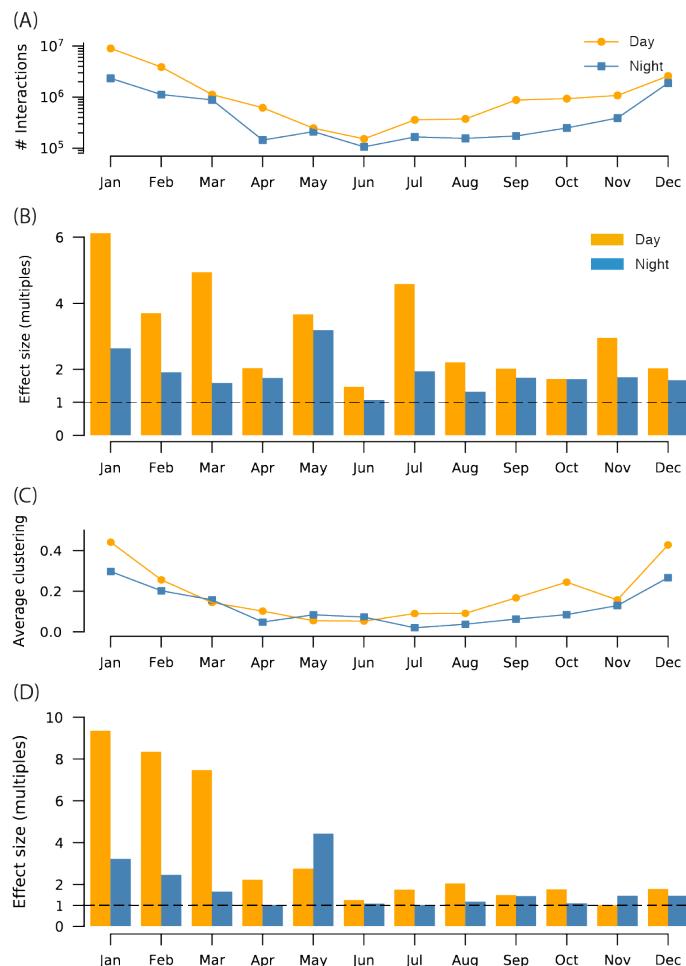


888

889

890 **FIG. 2. Seasons and daytime modulate individual and social behavior.** (A) Percentage  
891 successful location measurements across the year. Using a 5% threshold we label the periods  
892 marked in grey as missing data. Notice how these periods are missing in panels (B) and (C). (B)  
893 Individual behavioral indicators computed from location data, including temperature,  $T$ ,  
894 distance from shore,  $d$ , spatial entropy,  $S$ , velocity,  $v$ , and depth,  $h$ . All values in panels (B) and  
895 (C) are population averages. Deviations are represented as the average standard error of the  
896 mean of day and night, which rarely exceeds the absolute difference between day and night.  
897 The left set of axes are standardized values representing standard deviations from the mean,  
898 and the right give the actual values. (C) Social behavioral indicators measured at the level of  
899 pairs. We measure social activity in terms of the typical interaction duration,  $+$ , time between  
900 interactions,  $-$ , and interaction probability,  $p$ . (D) Example of how  $p$  varies rhythmically  
901 throughout January and July.

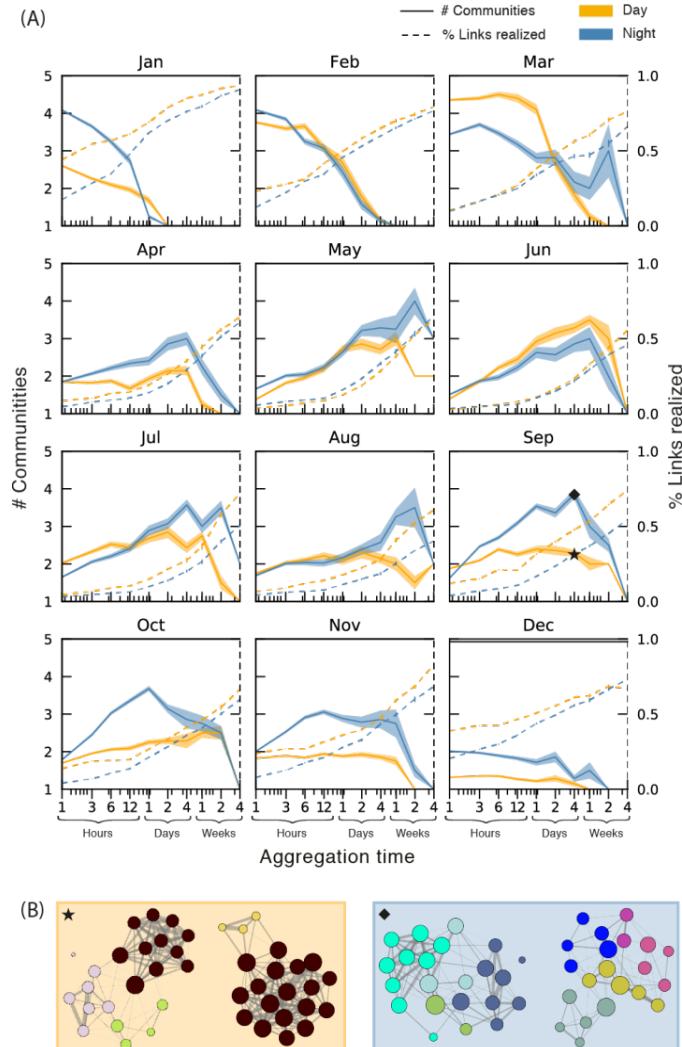
902



903

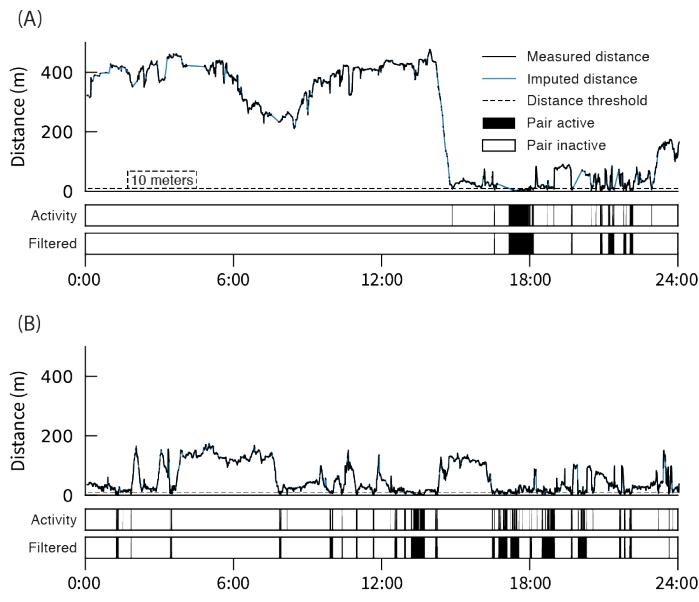
904

905 **FIG. 3. Effect sizes compared to shuffled data.** (A) Number of interactions observed during day  
 906 and night in the real data over the year. Note the y-axis is log-scaled. There are many times  
 907 more interactions during daytime, and a decrease in total number of interactions observed over  
 908 the year. (B) The effect size of number of interactions in the real data as compared to the  
 909 shuffled data (count of real divided by count of shuffled). The dashed line signifies a one-to-one  
 910 ratio and therefore no effect. (C) Average network clustering coefficient. (D) Effect size of  
 911 clustering.  
 912



913  
914  
915 FIG. 4. Number of communities at different levels of aggregation. (A) Number of communities  
916 (3 or more members) and percentage of realized links as a function of aggregation time for  
917 each month. Each has its own y-axis. Shaded regions surrounding the solid lines are standard  
918 errors. The x-axis is log-scaled. (B) Four example networks, two for day and two for night, with  
919 communities detected by the Infomap algorithm [18]. Each is an aggregate over four days in  
920 September, the references marked with a star and a diamond on the corresponding curves in  
921 (A).  
922

923  
924



925 FIG. 5. **Inferring proximity network with distance threshold.** (A) and (B) show how distance  
926 between two different pairs of fish vary throughout the same day. The dashed line at *Distance* =  
927 10 m is the threshold we enforce to get the activity series shown below, which we filter to  
928 produce the bottom series.

### 3 Scaling in human mobility

The following paper, *The Scales of Human Mobility*, presents a model which takes a mobility trajectory as input and returns a hierarchical partition of space into nested containers. When fitted to an individual's mobility trajectory, at each level, these containers tend to match our understanding of common geographical containers such as *building*, *city*, ..., *continents*. We analyze the properties of containers for ~700 000 individuals and find that there are strong commonalities within large populations implying that the levels actually correspond to characteristic *scales* of human mobility. In a practical sense this finding is intuitive: the smallest characteristic container size corresponds to mobility within buildings, the slightly larger one corresponds to trips taken within neighborhoods around visited buildings, and so on. At the same time it violates years of research within complexity science, arguing that human mobility is *scale-free*, i.e. that it has no characteristic scales [21–24]; although this goes against a widely acknowledged body of literature within traditional geography [25–28]. With the following working paper we try to resolve this paradox, and show that human mobility does indeed have characteristic scales. We plan to submit the paper to a high-impact journal and for this the model requires further validation, however, in its current form the paper should communicate with enough clarity the most important contributions that it brings.

# The Scales of Human Mobility

Ulf Aslak<sup>1</sup>, Laura Alessandretti<sup>1</sup> & Sune Lehmann<sup>1,2</sup>

September 2, 2019

## Abstract

There is a paradox at the heart of our current understanding of human mobility. On one hand, there is a highly influential stream of literature driven by analyses of massive empirical data sets which finds that human movements show no evidence of characteristic spatial scales. Here, human mobility is described as ‘scale-free’.<sup>1–4</sup> On the other hand, in geography, the concept of *scale*, referring to meaningful levels of description from individual buildings through neighborhoods, cities, regions, and countries, is central for the description of various aspects of human behaviour such as socio-economic interactions, or political and cultural dynamics.<sup>5–9</sup> Here, we resolve this apparent contradiction by showing that human mobility does indeed contain meaningful scales, corresponding to spatial containers restricting mobility behavior. The scale-free results arise from aggregating displacements across containers. We present a simple model, which given a person’s trajectory, infers their neighborhoods, cities, and so on, as well as the sizes of these geographical containers. We find that the containers characterizing the trajectories of more than 700 000 individuals do indeed have typical sizes. We show that our description dramatically improves on the state-of-the-art in modeling, and leads to better prediction of mobility patterns.

It is nearly impossible to underestimate the importance of a solid foundation for our understanding of human mobility. Hundreds of millions of individuals spend billions of collective hours commuting every day. Goods and food are transported through a global network using shared infrastructure. And as policy-makers and officials develop plans for ever-growing cities, a key component for the design is an understanding of

the mechanisms driving human mobility. Understanding mobility patterns also help us mitigate epidemic spreading, assist in crisis management, prepare for dramatic shifts in modes of transporation, and many other cases. For this reason, resolving the paradox of scale-free distributions of displacements<sup>1–4</sup> in empirical mobility traces is crucial, as this paradox is currently in many ways separating the large-scale data driven mobility research from traditional geography.<sup>5–9</sup>

Our mental representation of physical space has a hierarchical structure.<sup>10–12</sup> We describe space referring to *places*,<sup>13–15</sup> meaningful spatial entities of different scales, from rooms and buildings – via neighborhoods, cities, and states – to nations and continents<sup>16–18</sup> that are naturally organized in a nested structure.<sup>8,19</sup> Further, there is conjecture, which states that spatial scales also shape the way we move in such a way that there are fundamental differences between forms of moving from place to place at different scales, from moving within a building to traveling across the globe.<sup>20,21</sup>

Empirical research enabled by the availability of human trajectory data, however, has found no evidence for characteristic spatial scales in human mobility.<sup>1–4,22–25</sup> On the contrary, studies have shown that the distribution of displacement lengths  $\Delta r$  travelled by an individual has a power-law tail  $P(\Delta r) \sim \Delta r^{-\beta}$  over several orders of magnitude, where typically  $1 \leq \beta \leq 2$ .<sup>26</sup> Power-law distributions are also called *scale free*, precisely because they are the only mathematical distribution to have no associated typical scale<sup>27</sup> (see Supplementary Information, Section 1).

## Resolving the paradox

So the question becomes: How is it possible that our intuitive conception of space is clearly hierarchical and characterized by typical scales, while a variety of empirical datasets, ranging from displacements of dollar bills<sup>1</sup> or cell-tower data<sup>2</sup> to public transportation systems, and GPS data<sup>28–30</sup> all suggest that human mobility is scale free?

To explain this apparent contradiction, we propose that each typical scale of human mobility corresponds to a ‘container’ of a certain mobility behavior. These containers (neighborhoods, cities, countries, and so on) have typical sizes, see Figure 1 and roughly correspond to the notion of ‘places’ in geography.<sup>13–15</sup> The observed power-law data arises, when we aggregate data, merging mobility *within* containers with mobility that transports a person *between* containers. Specifically, it is well known that

mixtures of normal (or lognormal) distributions with different variances can generate power laws<sup>31,32</sup>.

More specifically, we assume that for each individual, physical space is organized as a nested structure of containers, where at each hierarchical level, containers have comparable sizes. These comparable sizes are given, in part by the concrete organization of our built environment: rooms, buildings, neighbourhoods, cities and countries, etc,<sup>13–15</sup> see Figure 1.

We now make the second assumption that this structure affects how individuals move. Specifically, we make the mild assumption that the amount of time spent in the same container depends on its hierarchical level. This assumption is supported by the literature, showing that e.g. transitions between buildings are more frequent than transitions between countries.<sup>20,21</sup> This assumption implies that trips with a specific purpose, such as going for a coffee, fetching groceries, or a trip to the movies occur at certain frequencies.<sup>33</sup> Now, due to the empirical clustering of resources and activities in space (e.g. at least one toilet within a building, one grocery store within a neighborhood, one cinema within a city), it is reasonable to assume that there is a typical trip-size for each trip purpose.

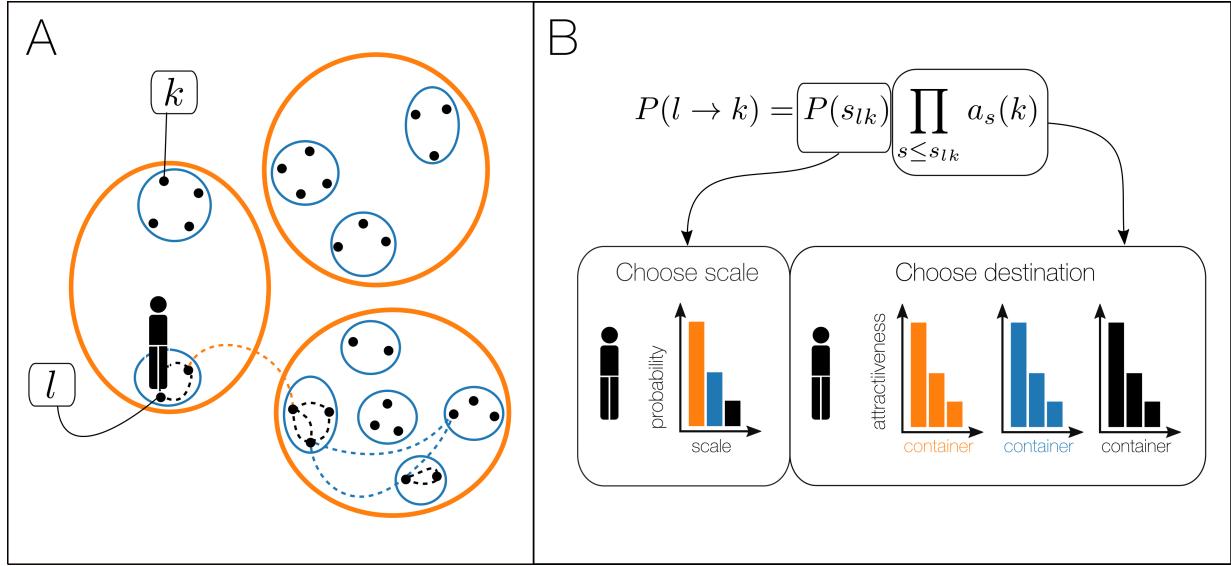
### A simple model identifies containers

We now describe a simple model which is able to estimate the size of a person containers from their empirical mobility patterns, based on these two assumptions, see Figure 2. We model physical space as a hierarchy of  $S$  levels, ordered from the smallest to the largest (e.g. buildings to countries). At any level  $s$ , space is partitioned into compact containers, with a characteristic size. For  $s < S$ , each container is fully included within a single parent container (for example each neighbourhood is part of a single city). Hence, each geographical location  $k$  can be identified as a sequence of containers,  $k = (k_1, \dots, k_s, \dots, k_S)$ , where place  $k_s$  is included in  $k_{s+1}$ .

Next, consistent with most models of human mobility,<sup>3,30,34</sup> each container  $k_s$  is characterized by its attractiveness  $a(k_s)$ , such that the total attractiveness of containers within a parent container is equal to 1. We define the *level-distance*  $d(l, k)$  between locations  $l$  and  $k$  as the highest index at which the two sequences of containers describing  $l$  and  $k$  differ (e.g. two locations are at neighborhood distance if they are in different neighborhoods).



**Figure 1: The hierarchical structures of containers: examples.** A-C) The containers of an individual living in a small danish town, characterized by the town container size (blue, A), the district containers (orange, A), the region containers (green, B), and the neighbouring-countries containers (red, C). D-F) Those of a Copenhagen citizen, characterized by the neighbourhood container size (blue, D), the city containers (orange, D), the urban agglomeration containers (green, E) and the region container (red, F). A) Individuals move between locations (black dots), where each location  $l$  is contained within a container at level  $s = 1$  (blue circles) and at level  $s = 2$  (orange circles). B) The probability of transitioning between two locations  $l$  and  $k$  is given by the probability of traveling at level-distance  $d(l, k)$  times the probability of choosing location  $k$ .



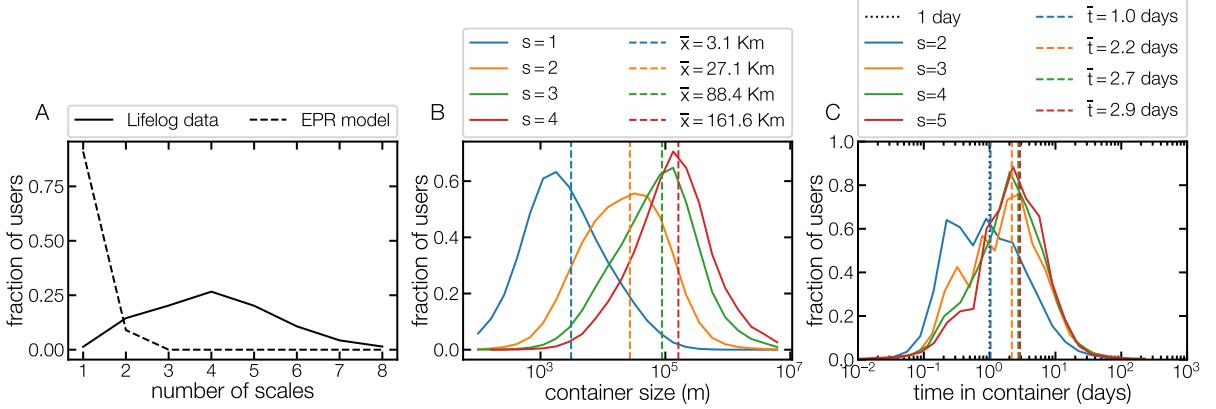
**Figure 2: Schematic representation of the ‘container model’.** A) Individuals move between locations (black dots), where each location  $l_1$  is contained within a container at level  $s = 2$  (blue circles) and at level  $s = 3$  (orange circles). B) The probability of transitioning between two locations  $l$  and  $k$  is given by the probability of traveling at level-distance  $d(l, k)$  times the probability of choosing location  $k$ .

hoods, but in the same city). We model this hierarchical structure to be unique for each mobility trace.

Based on the assumption that the amount of time spent in a container depends upon its place in the hierarchy, we design a model of trajectories where the probability of transitioning from location  $l$  to location  $k$  depends on the level-distance between them. For an agent located in  $l$ , we model the probability of moving to  $k$  as the product of two terms:  $P(l \rightarrow k) = p_{d(l,k)} \prod_{s \leq d(l,k)} a(k_s)$ . The first term,  $p_{d(l,k)}$ , is the probability of traveling at level-distance  $d(l, k)$ , which follows a multinomial distribution. The second term  $\prod_{s \leq d(l,k)} a(k_s)$  is the probability of choosing a specific location  $k$  at that level-distance, where  $a(k_s)$  is the attractiveness of a container including  $k$ .

## Results: Empirical findings

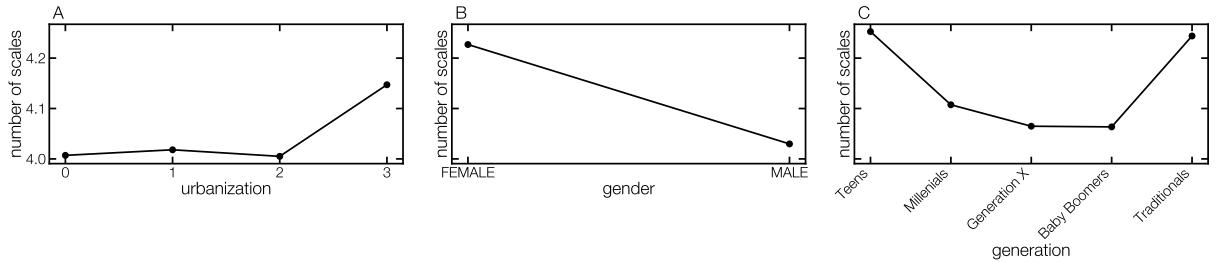
We fit our model to the individual mobility traces of  $\sim 700,000$  individuals from two different datasets (see Methods, Data description), using maximum likelihood estimation (see Methods, Likelihood optimization). For each individual, the fitting procedure outputs the most likely hierarchical spatial structure, along with attractiveness of con-



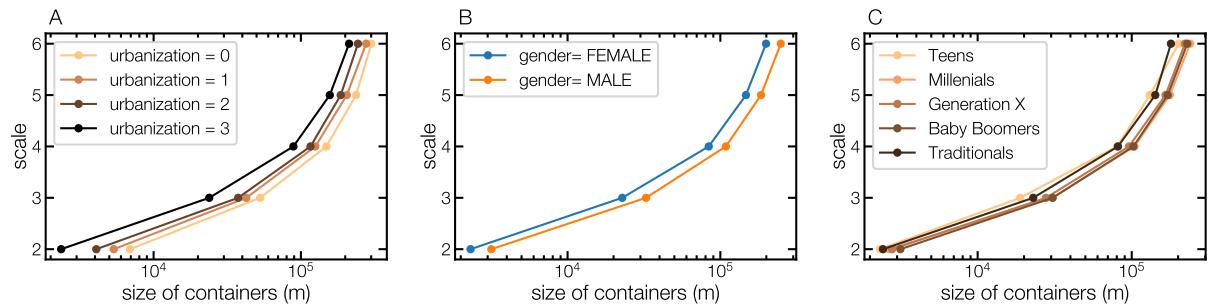
**Figure 3: The scales of human mobility.** (A) Fraction of individuals with a given number of scales for the Lifelog data population (filled line) and for the synthetic data generated using the ‘Exploration and Preferential Return’ (EPR) model (dashed line). (B) Distribution of containers size across users. Results are shown for containers at different hierarchical levels. (C) Distribution of median time spent in the same container across users, for containers at different hierarchical levels.

tainers and probabilities of traveling at given level-distance. We find that empirical individual mobility traces are characterized, on average, by four hierarchical levels (see Fig. 3A). In contrast, synthetic traces generated by the state-of-the-art ‘Exploration and preferential return’ model<sup>3,28,30</sup> are best described by a single hierarchical level (see Fig. 3A). We find that the size of containers characterizing human data – defined as the maximum distance between two locations in the same container at a given level – are well described by lognormal distribution across the population under study. The parameters  $\mu$  of the lognormal distributions at different hierarchical levels ( $\mu_1 = 3.1\text{ km}$ ,  $\mu_2 = 27.1\text{ km}$ ,  $\mu_3 = 88.4\text{ km}$  and  $\mu_4 = 161.6\text{ km}$ ) characterize the typical ‘scales’ of human mobility (see Fig. 3B). We find that there are characteristic staying times associated to spatial scales, since the median time spent within the same container at any given level is homogeneously distributed across the population (see Fig. 3C).

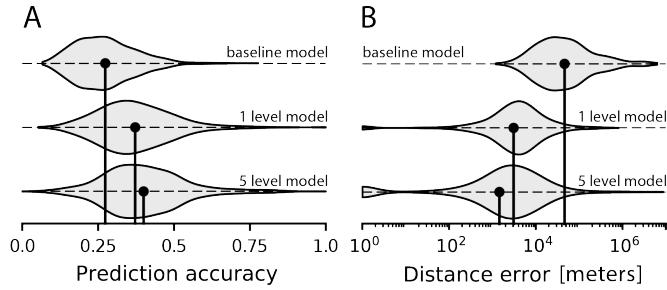
The differences between individuals can be partially explained resorting to socioemographics factors including rural/urban divide, gender and age. We find that the number of levels is significantly higher for urban individuals, compared to non-urban (see Fig. 4A) and for women compared to men (see Fig. 4B). Further, we find that the typical size of containers is significantly larger for rural compared to urban individuals (see Fig. 5A) and for men compared to women (see Fig. 5B). Interestingly, the number of scales exhibits a minimum for individuals in working age (see Fig. 5C).



**Figure 4: The number of scales.** Average number of scales for users with different urbanization level (A), gender (B) and generations (C). Errorbars (standard deviation of the mean) are too small to be visible. In panel C, 'Teens' refers to 18 years old users, 'Millenials' to 18-35 years old users, 'Generation X' to 35-50 years olds, 'Baby Boomers' to 50-70 years olds and 'Traditionals' to users more than 70 years old.



**Figure 5: The size of containers.** Average size of containers at different hierarchical levels for users with different urbanization level (A), gender (B) and generations (C). Errorbars (standard deviation of the mean) are too small to be visible. In panel C, 'Teens' refers to 18 years old users, 'Millenials' to 18-35 years old users, 'Generation X' to 35-50 years olds, 'Baby Boomers' to 50-70 years olds and 'Traditionals' to users more than 70 years old.



**Figure 6: Comparing performance of 1- and 5-level models for mobility prediction.** Performance distributions over CNS users measured as A) prediction accuracy and B) distance accuracy. Vertical lines track the mean values of each distribution. The baseline model predicts the second most frequent location when the user is in the most frequent location and the most frequent location when the user is anywhere else.

### Results: Validation by prediction

To assess the amount of predictive power in our model’s hierarchical description of space, we train two recurrent neural networks models with identical architecture for next place prediction in trajectories, see Methods, Model validation for implementation details. Both models are general and learn from the trajectories of an 800+ sized population of university students.<sup>35</sup> The first model is trained on place trajectories  $T = \{k_1(0), \dots, k_1(i), \dots, k_1(n_T)\}$ , i.e. a 1-level description of space, and the second model is trained on container sequence trajectories  $T = \{k(0), \dots, k(i), \dots, k(n_T)\}$  up to five levels, where the hierarchy of containers for each individual is fitted by our model. When comparing how well the two models are able to predict unseen trajectories, we find a significant increase in prediction accuracy for the model trained on trajectories described in five levels (baseline: 0.27; 1-level: 0.37; 5-level: 0.40), see Figure 6. Moreover, when measuring the model performance as the distance between the predicted location and the actual location, the five level model also give significantly better results (baseline: 46 159 m; 1-level: 3 083 m; 5-level: 1 451 m). The increased distance accuracy suggests that the model is able to use the hierarchical description of containers to relate places to one another. This implies that when the second model predicts the wrong place, it is nevertheless likely to get its higher level containers right.

## Summary

In this paper, we proposed a model in which human mobility is characterized by a hierarchical structure of spatial containers, corresponding to the notion of ‘places’ in geography. Under this model, the observed power-law data arises by merging mobility *within* containers with mobility that transports a person *between* containers. Fitting the model to trajectories collected in two different datasets, we found that the containers characterizing individuals have typical sizes, the ‘scales’ of human mobility. We showed that the presented model improves on the state-of-the-art in modeling, and leads to better prediction of mobility patterns.

## Methods

### Data description and pre-processing.

Our analyses are based on two mobile phone datasets collecting high-resolution human trajectories.

The *Copenhagen Network Study* (CNS) data was collected as part of an experiment that took place between September 2013 and September 2015.<sup>35</sup> The experiment involved 851 Technical University of Denmark students ( $\sim 22\%$  female and  $\sim 78\%$  male), typically aged between 19 and 21 years old. Participants position over time was estimated from GPS data sampled every 1 – 2 minutes. The location estimation error was below 50m in 95% of the cases. Data collection was approved by the Danish Data Protection Agency. All participants provided informed consent .

The *Lifelog* dataset consists of anonymized GPS location data for  $\sim 5\,000\,000$  users of the Lifelog app between 2017 and 2019. We selected  $\sim 700\,000$  users with at least one year of data and whose position is known, every day, at least 50% of the time. Lifelog users are geolocalized across the world and are aged between 18 and 80 years old, with an average age of 36 years. About one-third of users are female. Data are not collected with a fixed time interval. Instead, the app receives updates when there is a change in the motion-state of the device (if the accelerometer registers a change). Location estimation error is below 100 m for 93% of data points. Data collection for the Sony dataset has been approved by the Sony Mobile Logging Board, and informed consent was obtained

for all study participants according to the Sony Mobile Application Terms of Service and the Sony Mobile Privacy Policy.

We pre-processed both trajectory datasets to obtain stop-locations using the Infostop algorithm<sup>36</sup>. We used different algorithm parameters to account for differences in data collection: `r1= 30 m`, `r2= 30 m`, `min_staying_time= 10 minutes`, `max_time_between= 24 hours` (Lifelog dataset), .

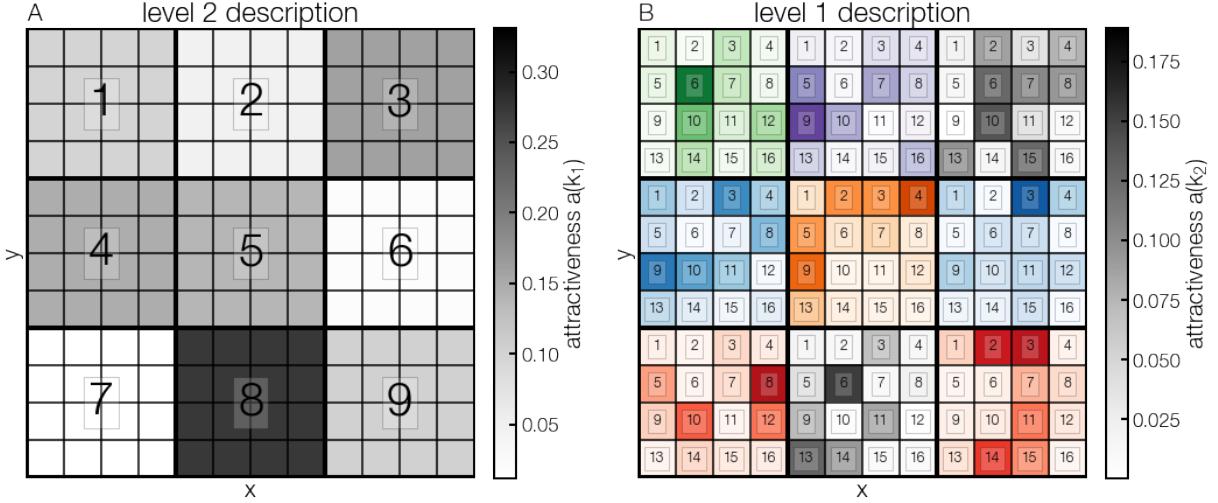
## The generative model.

### The generative model: Nested grid representation.

We develop an  $S - levels$  hierarchical partitioning of a two-dimensional square space. Levels are ordered from the most fine grained  $s = 1$  to the most coarse grained  $s = S$ . At each level  $s$ , space is divided into  $N_s$  containers  $k_s$  of equal size, such that each of them includes the same number of containers  $k_{s-1}$  at a lower level. This representation is effectively a nested grid, that we define as  $G = [N_S, \dots, N_1]$  (see Fig. 7) . At level  $s = 1$ , there are  $N_1$  containers, corresponding to all locations an individual can occupy. Each of these locations,  $k$ , can be uniquely identified, either by its 2-dimensional coordinates  $(x, y)$ , or as a sequence of containers  $k = (k_1, \dots, k_s, \dots, k_S)$ . At each level  $s$ , all containers  $k_s$  are characterized by their attractiveness  $a(k_s)$ , extracted from a power law distribution with exponent  $\beta_s$ . The distribution is normalized within each container, such that  $\sum_{k_s \in k_{s+1}} a(k_s) = 1$ , for all choices of  $k_{s+1}$ . Here, the notation  $k_s \in k_{s+1}$  stands for containers at level  $s$  within the same container at level  $s + 1$  (see Fig. 7). We define the collection of containers attractiveness as  $a$ .

### The generative model: Generation of traces.

We simulate individual traces as follows. We fix the vector of parameters  $p = \{p_1, \dots, p_s, \dots, p_S\}$  (that sums to 1), defining the probability of traveling between two locations at *level-distance*  $s$ . Here, by *level-distance*  $d(k, j)$  between two locations  $k$  and  $j$ , we mean the number of non-common containers in their sequence of containers description. We initialize a user in location  $k = (k_1, \dots, k_s, \dots, k_S)$ . At each displacement, the user chooses the level  $s^*$  at which to travel based on a multinoulli trial experiment with probabilities  $\{p_1, \dots, p_s, \dots, p_S\}$ . Then, she chooses a destination container at level  $s^*, j_{s^*}$ , with probability



**Figure 7: Hierarchy of levels in the generative model.** Example with  $S = 2$  scales,  $N_1 = 144$  containers at scale  $s = 1$ ,  $N_2 = 9$  containers at scale  $s = 2$ . (A) At scale  $s = 2$ , each of the  $k_2$  containers has a different attractiveness (see colorbar) extracted from a power-law distribution, such that the total attractiveness is  $\sum_{k_1} a(k_1) = 1$ . (B) At scale  $s = 1$ , attractiveness is normalized within each of the 9  $s = 2$  containers, such that  $\sum_{k_1 \in k_2} a(k_1) = 1$ , for all choices of  $k_2$ .

proportional to its attractiveness  $a(j_{s^*})$ . Note that the destination container  $j_{s^*}$  must be different from the origin  $k_{s^*}$ . In the same manner, she chooses a container  $j_{s^*-1} \in j_{s^*}$  at level  $s^* - 1$ , and so on, until the container at level  $s = 1$  is selected. A synthetic trajectory is produced by iterating the procedure above to generate a sequence of displacements.

### The generative model: Likelihood of parameters.

Given a hierarchical partitioning of space  $G$ , a collection of containers' attractiveness  $\mathbf{a}$  and a vector  $\mathbf{p}$ , the probability of a displacement  $k \rightarrow j$  between two cells  $k$  and  $j$  can be written as:

$$P(k \rightarrow j | G, \mathbf{a}, \mathbf{p}) = p_{d(k,j)} \frac{a(j_{d(k,j)})}{1 - a(k_{d(k,j)})} \prod_{s=1}^{d(k,j)-1} a(j_s)$$

The first term,  $p_{d(l,k)}$ , is the probability of traveling at *level-distance*  $d(l,k)$ . The second term  $\frac{a(j_{d(k,j)})}{1 - a(k_{d(k,j)})}$  is the probability of choosing container  $j_{d(k,j)}$  at *level-distance*  $d(k,j)$ , where  $k_{d(k,j)}$  can not be selected. The third term  $\prod_{s=1}^{d(k,j)-1} a(j_s)$  is the probability of picking all containers  $j_s$ , with  $s < d(k,j)$ . Thus, the likelihood of a trajectory of locations  $T =$

$\{k(0), \dots, k(i), \dots k(n_T)\}$ , with length  $n_T$ , can be written as:

$$L(T \mid \mathbf{G}, \mathbf{a}, \mathbf{p}) = \prod_{i=0}^{n_T-1} P(k(i-1) \rightarrow k(i) \mid \mathbf{G}, \mathbf{a}, \mathbf{p})$$

## Likelihood optimization.

### Synthetic traces

Given a synthetic trajectory  $T = \{k(0), \dots, k(i), \dots k(n_T)\}$ , we can estimate the hierarchical partitioning  $\mathbf{G}$ , the collection of containers attractiveness  $\mathbf{a}$ , and the probabilities  $\mathbf{p}$  that generated the trace, via maximum likelihood estimation. Namely, we can explore the space of these parameters, to find a combination of the three maximizing the likelihood. This task is expensive, and we approach the problem as follows.

First, one should note that, when the hierarchical partitioning of space  $\mathbf{G}$  is fixed, and for large trajectory length  $n_t$ ,  $\mathbf{p}$  can be estimated fairly easily. In fact, for  $n_t \rightarrow \infty$ , the probability of travelling at a given *level-distance*,  $p_s$ , is simply the fraction of displacements between locations at *level-distance*  $s$ :

$$\frac{|\{k(i) \rightarrow k(i+1) \mid d(k(i), k(i+1)) = s\}|}{n_t - 1} \xrightarrow{n_t \rightarrow \infty} p_s$$

Likewise, for  $n_t \rightarrow \infty$ , the attractiveness of containers can be estimated as the fraction of times a given container is chosen, among all containers at the same level.

Hence, for long enough traces, we can estimate the maximum likelihood by exploring the space of possible hierarchical partitions  $\mathbf{G}$  only. For example, for a space of  $12 \times 12 = 144$  locations (see Fig.7), the set of possible hierarchies include: a 1-level grid with  $\mathbf{G} = [144]$ , three 2-levels grids with  $\mathbf{G} = [4, 144]$ ,  $\mathbf{G} = [36, 144]$ ,  $\mathbf{G} = [9, 144]$  (this is the configuration in Fig. 7), or  $\mathbf{G} = [16, 144]$ , and three 3-levels grids with  $\mathbf{G} = [4, 16, 144]$ ,  $\mathbf{G} = [9, 36, 144]$ , or  $\mathbf{G} = [4, 36, 144]$ .

In principle, for a given sequence of displacement within a square  $12 \times 12$ , one can compute the trace likelihood for all the configurations above, then find the one configuration  $\mathbf{G}^*$  that maximises the likelihood. This procedure, however, is expensive when the space considered is much larger than in this example. For this reason, we design the

*one-level-at-the-time* optimization algorithm, that reduces substantially the set of configurations to try, and results in an optimal solution.

The algorithm works as follows. We aim at finding the partitioning  $G^*$  corresponding to the maximum likelihood  $L^*$  for a given sequence of locations  $T$ . We start assuming that the best partitioning has 1-level, so we assign  $G^* = [N_1]$  (there is only one 1-level partitioning), and we compute the corresponding likelihood  $L^* = L(T \mid [N_1])$ . Then, we compute the likelihoods for all choices of 2-level partitions. If none of them is larger than  $L^*$ , the best partition has 1-level only. Otherwise, we assign  $G^* = [N_2, N_1]$ , where  $[N_2, N_1]$  is the partition with maximum likelihood, and we update  $L^*$ . We continue by choosing, among all possible 3-level hierarchical partitions, only those including both  $N_1$  and  $N_2$ . From the example above, if the selected 2-level partition is  $G^* = [4, 144]$ , we only test the 3-level partitions  $G^* = [4, 16, 144]$  and  $G^* = [4, 36, 144]$ , while if it is  $G^* = [36, 144]$ , we test the 3-level partitions  $G^* = [9, 36, 144]$  and  $G^* = [4, 36, 144]$ . This process is iterated for increasing number of levels, until either the likelihood does not improve or one has reached the maximum number of possible levels (in the example above, that is 3). Note that, to avoid overfitting, at each iteration we compute a distribution of likelihood, by bootstrapping 1000 random samples of 300 displacements. We consider that there is an increase in likelihood if both the total likelihood increases, and the distribution of bootstrapped values of likelihoods is significantly different from the best distribution under Kolomogorov-Smirnov test (at significance level  $\alpha = 0.01$ ). The idea here, is that we can proceed with ‘local’ optimizations (one level at the time), to reach the global optimum.

## Real traces

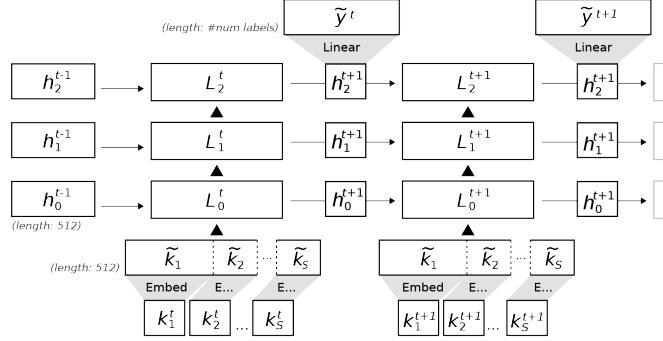
In many aspects, the maximum likelihood method described above can be applied to find the model parameters characterizing any sequence of locations, including synthetic and real trajectories. A nested grid, however, is not suitable to describe the space humans move in. Thus, we adapt the method above to fit human traces. Rather than choosing the best hierarchical partitioning among nested grids, we choose it among the partitioning generated by complete-linkage clustering of an individual’s locations.<sup>37</sup> Given a set of  $N$  different locations, the complete linkage clustering generates a set of  $N$  possible partitions of locations. For any of these partitions, locations are grouped into compact containers with similar size. Different partitions have different typical con-

tainer sizes, the two extreme cases being each location in a separate container and all locations in the same container. Since the partitions are organized in a nested fashion, we can directly apply the methodology used for synthetic traces, where we add one scale at the time until the likelihood can not be further improved.

## Model validation

### Next place prediction

We assess to what extent the hierarchical description of space provided by our proposed model aids the performance of a recurrent neural network for next location prediction. We use a three layer gated recurrent unit (GRU) with 512 hidden units, regularized with constant weight decay at  $10^{-5}$ . The model is trained for 5000 epochs with early stopping, a chunk length of 100 and a batch size of 10. We train a general model that learns from all CNS users in such a way that each training batch contains subtrajectories from different users. Mixing data from different users in each training batch has a regularizing effect since every gradient descent step thus optimizes model performance for multiple users at the same time making it harder to overfit to single trajectories. For each user the first 95% of their full trajectory is used for training, and the remaining 5% is used for testing. During prediction, the input vector  $k = \{k_1, \dots, k_s\}$  is converted in the first layer to a concatenation of random embedding vectors for each container-level,  $\tilde{k}$ . We store a separate embedding for each level such that  $k_s$  and  $k_{s+1}$  are embedded as different vectors. The embedded input and hidden vectors from previous time-step,  $h_l^{t-1}$ , are fed into three GRU layers  $L_l^t$ , where the top hidden vector,  $h_2^t$ , is matrix multiplied with a weight matrix,  $W \in \mathbb{R}^{925 \times 512}$ , yielding an output vector  $\tilde{y}^t$  which encodes one value for each of the 925 possible place indices. Places are for each user indexed by decreasing frequency and reused between users, but do not correspond to the same places for different users. Still, the model can learn population-wide patterns at the  $s = 1$  level likely because frequency encodes semantics to some extent, e.g. your home and work are usually your most visited places. We use a the cross entropy to estimate the cost of each prediction. When estimating accuracy we take the index of the largest value to be the predicted place. The baseline model which we compare with is a simple algorithm that always guesses the most frequent place unless the user is already there, in which case it guesses for the second most frequent place.



**Figure 8: Prediction model architecture.** Example with  $S$ -level input. Container-indices get embedded into higher dimensional vectors and then concatenated together and fed into a three GRU layers. The output is transformed into a vector that stores a value for every possible place.

## References

- [1] Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462 (2006).
- [2] Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *nature* **453**, 779 (2008).
- [3] Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nature Physics* **6**, 818 (2010).
- [4] Deville, P. *et al.* Scaling identity connects human mobility and social interactions. *Proceedings of the National Academy of Sciences* **113**, 7047–7052 (2016).
- [5] Sheppard, E. & McMaster, R. B. Scale and geographic inquiry. *Scale and Geographic Inquiry: Nature, society and method*. Malden: Blackwell 1–22 (2004).
- [6] Moore, A. Rethinking scale as a geographical category: from analysis to practice. *Progress in human geography* **32**, 203–225 (2008).
- [7] Sayre, N. F. Ecological and geographical scale: parallels and potential for integration. *Progress in human geography* **29**, 276–290 (2005).
- [8] Paasi, A. Place and region: looking through the prism of scale. *Progress in human geography* **28**, 536–546 (2004).

- [9] Marston, S. A. The social construction of scale. *Progress in human geography* **24**, 219–242 (2000).
- [10] Hirtle, S. C. & Jonides, J. Evidence of hierarchies in cognitive maps. *Memory & cognition* **13**, 208–217 (1985).
- [11] McNamara, T. P., Hardy, J. K. & Hirtle, S. C. Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 211 (1989).
- [12] Tomko, M., Winter, S. & Claramunt, C. Experiential hierarchies of streets. *Computers, Environment and Urban Systems* **32**, 41–52 (2008).
- [13] Tuan, Y.-F. Place: an experiential perspective. *Geographical review* 151–165 (1975).
- [14] Gustafson, P. Roots and routes: Exploring the relationship between place attachment and mobility. *Environment and behavior* **33**, 667–686 (2001).
- [15] Canter, D. The facets of place. In *Toward the Integration of Theory, Methods, Research, and Utilization*, 109–147 (Springer, 1997).
- [16] Low, S. M. & Altman, I. Place attachment. In *Place attachment*, 1–12 (Springer, 1992).
- [17] Rubinstein, R. I. & Parmelee, P. A. Attachment to place and the representation of the life course by the elderly. In *Place attachment*, 139–163 (Springer, 1992).
- [18] Lewicka, M. What makes neighborhood different from home and city? effects of place scale on place attachment. *Journal of environmental psychology* **30**, 35–51 (2010).
- [19] Sheppard, E. The spaces and times of globalization: Place, scale, networks, and positionality. *Economic geography* **78**, 307–330 (2002).
- [20] Cresswell, T. *On the move: Mobility in the modern western world* (Taylor & Francis, 2006).
- [21] Cresswell, T. Mobilities i: catching up. *Progress in human geography* **35**, 550–558 (2011).
- [22] Hu, Y., Zhang, J., Huan, D. & Di, Z. Toward a general understanding of the scaling laws in human and animal mobility. *EPL (Europhysics Letters)* **96**, 38006 (2011).

- [23] Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *PLoS one* **7**, e37027 (2012).
- [24] Zhao, Z.-D., Huang, Z.-G., Huang, L., Liu, H. & Lai, Y.-C. Scaling and correlation of human movements in cyberspace and physical space. *Physical Review E* **90**, 050802 (2014).
- [25] Rhee, I. *et al.* On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking (TON)* **19**, 630–643 (2011).
- [26] Alessandretti, L., Sapiezynski, P., Lehmann, S. & Baronchelli, A. Multi-scale spatio-temporal analysis of human mobility. *PLoS one* **12**, e0171686 (2017).
- [27] Newman, M. E. Power laws, pareto distributions and zipf's law. *Contemporary physics* **46**, 323–351 (2005).
- [28] Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S. & Baronchelli, A. Evidence for a conserved quantity in human mobility. *Nature human behaviour* **2**, 485 (2018).
- [29] Gallotti, R., Bazzani, A., Rambaldi, S. & Barthelemy, M. A stochastic model of randomly accelerated walkers for human mobility. *Nature communications* **7**, 12600 (2016).
- [30] Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nature communications* **6**, 8166 (2015).
- [31] Zhao, K., Musolesi, M., Hui, P., Rao, W. & Tarkoma, S. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Scientific reports* **5**, 9136 (2015).
- [32] Gheorghiu, S. & Coppens, M.-O. Heterogeneity explains features of anomalous thermodynamics and statistics. *Proceedings of the National Academy of Sciences* **101**, 15852–15856 (2004).
- [33] Krizek, K. J. Neighborhood services, trip purpose, and tour-based travel. *Transportation* **30**, 387–410 (2003).
- [34] Barbosa, H. *et al.* Human mobility: Models and applications. *Physics Reports* **734**, 1–74 (2018).

- [35] Stopczynski, A. *et al.* Measuring large-scale social networks with high resolution. *PloS one* **9**, e95978 (2014).
- [36] Alessandretti, L. & Ulf, A. Infostop: Python package for detecting stop locations in mobility data (2018).
- [37] Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967).

## Methods

**Data description and pre-processing.**

**Estimation of model parameters via maximum likelihood.**

## 4 Educational peer review

The following two papers is joint work together with *Peergrade*, a company which delivers an online software service that facilitates student peer reviews. Students use Peergrade to turn in assignments, and through their system proceed to anonymously review each other's papers based on a set of teacher-defined criteria. After having received feedback from multiple students (often three or more), students rate the quality of the feedback they received. These interactions yield lots of behavioral data, enabling research that is relevant to teachers and educational planners, as well as the wider e-Learning community. The papers are published in proceedings of the *12th International Conference on e-Learning* and *16th European Conference on e-Learning*, respectively. The first paper was awarded *best paper* after presentation at the corresponding conference.

1. The paper *Quantifying Feedback – Insights Into Peer Assessment Data* probes the question: are high-performing students better peer reviewers? To our surprise we find that they are not, which has the important implication that low-performing students are just as good reviewers as good students. While this result is obviously important to validate student peer-reviews as a legitimate method for grading papers, it also gives fundamental insight into how students learn; that making many mistakes on an assignment still leads to learning, enough so that the student can review others' work.
2. The paper *Optimal allocation of reviewers for peer feedback* assesses probes the limits of *fairness* in algorithmic allocation of reviewers to papers. Because some students are careful reviewers while others are less so, a fair allocation entails every paper getting close to the same number of reviews from both good and less good reviewers. Review session play out within a time-window, and the abilities of each reviewer as well as the quality of the placed reviews are updated continuously, therefore, multiple algorithms can be tested. Through simulation, we find that random pre-allocation of reviewers lead to very unfair review quality distribution across papers, while a simple *online* algorithm that updates the expected ability of each reviewer after each of their reviews and then allocates reviewers according to need, performs almost as well as the theoretical best.

# Quantifying Feedback – Insights Into Peer Assessment Data

David Kofoed Wind<sup>1</sup>, Ulf Aslak Jensen<sup>2</sup>

<sup>1</sup>Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>University of Copenhagen, Copenhagen, Denmark

[dawi@dtu.dk](mailto:dawi@dtu.dk)

[uaj@econ.ku.dk](mailto:uaj@econ.ku.dk)

## 1. Abstract

The act of producing content - for example in forms of written reports - is one of the most used methods for teaching and learning all the way from primary school to university. It is a learning tool which helps students relate their theories to practice. Getting relevant and helpful feedback on this work is important to ensure a good learning experience for the students. Providing this feedback is often a time-consuming job for the teacher. An effective way to learn is to teach others, and similarly give feedback on work done by others. One way to approach a combined solution to the above challenges, is to use peer assessment in the classroom which as a learning method has become more and more popular. In this paper we look at data collected using the web-based peer assessment system Peergrade. The dataset consists of over 350 courses at more than 20 educational institutions and with a total of more than 10,000 students. The students have together made more than 100,000 peer-evaluations of work by other students, and these evaluations together contain more than 10,000,000 words of text feedback. A key problem when using peer assessment is to ensure high quality feedback between peers. Feedback here can be a combination of quantitative / summative feedback (numerical) and qualitative / formative feedback (text). A lot of work has been done on validating and ensuring quality of quantitative feedback. We propose a way to let students evaluate the quality of the feedback they receive to obtain a quality measure for the feedback. We investigate this measure of feedback quality, which biases are present and what trends can be observed across the dataset. Using our measure of feedback quality, we investigate how it relates to various factors like the length of the feedback text, the number of spelling mistakes, how positive it is and measures of the student's report-writing skills.

**Keywords:** peer assessment, peer feedback, feedback, peer review, peer evaluation, peer grading

## 2. Introduction

### 2.1 What is peer assessment?

Peer assessment is the act of letting students partake in the process of evaluation and giving feedback. Peer feedback is a formative variant of this where students discuss the strengths and weaknesses of a specific performance at length and indicate suggestions for further improvement. It is the counterpart of feedback by a teacher. Both are outcomes of formative assessment, also called assessment for learning (Black & Wiliam, 1998; Gielen et al 2010).

### 2.2 Why is it meaningful to use peer assessment?

The use of peer assessment in education comes with a number of benefits for both teachers and students (Tsui & Ng, 2000). Peer feedback can increase pressure on students to perform well by introducing a social pressure. Cole (1991) finds this effect in fourth grade and Pope (2005) in university. An increase in time spent on assignments that are given peer feedback is documented in (Tsui & Ng, 2000; Pope, 2001; Gibbs & Simpson, 2004). When an entire class of students collaborate on evaluating homework and with a good tool in place where the logistics are handled, the teacher can save a large amount of time and students can receive feedback on their work much faster (Sadler, 2006; Searby, 1997). Gibbs et al (2004) writes that getting imperfect feedback immediately might have much more impact than perfect feedback four weeks later. Gielen et al (2010) also talks about the possibility of introducing more feedback into the classroom through "several 'intermediate' peer assessment sessions on draft versions of for instance an essay or report" arguing that this "could answer to this need of regular feedback if teachers are not able or willing to increase its frequency themselves.". When students assess and give feedback to work of other students, they get a deeper understanding of the material, both from seeing alternative solutions to the task they have solved themselves and from reflecting upon the work of others and formulating their thoughts into constructive feedback (Bloom

et al, 1956; Boud, 1989; Sadler, 2006; Liu, 2006). Topping (2003) describes that students find peer feedback more understandable and useful due to other students being ‘on the same wavelength’, and teacher feedback is often misinterpreted by students due to a difference in discourse (Hounsell, 1987; Higgins, 2000; Gibbs et al., 2004; Yang et al., 2006). Giving feedback is also a good way to train the ability to receive and understand feedback (Rust et al., 2003; Bloxham & West, 2004). The use of peer evaluations in a class can even make the classroom friendlier, more cooperative and consequently build a greater sense of shared ownership for the learning process (Sadler & Good, 2006). The process of peer assessment is not limited to educational purposes, and the use of self-evaluation and peer review are an important part of future, adult, professional practice, and test grading is a good way to develop these skills (Boud, 1989).

### **2.3 Why is feedback important?**

Graham et al (2015) did a meta-analysis of true and quasi-experiments with students in grades 1 to 8. They found that feedback to students about writing from adults, peers, self, and computers statistically enhanced writing quality, yielding average weighted effect sizes of 0.87, 0.58, 0.62, and 0.38, respectively. Kingston & Nash (2011) also did a meta-study using 42 independent effect sizes and determined a weighted mean effect size of 0.20. Smith & Gorard (2005) investigate the effect of replacing numerical feedback (grades) with just formative feedback. They split 104 year 7 pupils into 4 groups and find that the group receiving no grades progresses substantially inferior compared to the other groups - pointing towards importance of measurable and summative feedback. Kim (2005) studies how feedback composition affects performance of the receiving student. Feedback was considered constructive when marks and comments for each content criterion were present and supported by a rationale and revision suggestion. There was no significant indication that well-composed feedback leads to a performance increase. Kim (2005) argues that this might be due to a too limited variance in peer feedback quality.

### **2.4 What is good feedback?**

Gielen et al (2010) studied 43 students of Grade 7 in secondary education and showed that receiving ‘justified’ comments in feedback improves performance (however with diminishing effect for students with better pretest results). They also showed that justification was more important than accuracy in comments. Strijbos et al (2010) investigates how the feedback content and competence level of the feedback giver affects efficiency and perception of the feedback. They find that concise general feedback outperforms elaborated specific feedback, and surprisingly that groups with a low competent peer outperformed groups with a high competent peer during a posttest. There was no correlation between the feedback perception and performance.

### **2.5 Previous work**

Davies (2003) investigates using only text feedback from peers for determining the quality of a hand-in. They do this by creating a “feedback index” for mapping qualitative feedback to a numerical score. They find a positive correlation between the marks awarded by students and the numerical value of comments awarded by peers. The method used (constructing and using the feedback index) requires a lot of manual work for the teacher and peers. Rada & Hu (2002) argues that when students spend time on marking their peers, and there exists a pedagogical benefit for the practice, there should be a reward for giving good feedback. Davies (2003) follows up on this by asking the question of what good feedback is.

Prins et al (2006) attempts to identify the style and quality of feedback practices among GPiTs (general practitioners in training). First, in separate sessions, six groups of 8–12 GPiTs and 12 GP-trainers write a qualitative feedback report for a video recording. The feedback receivers were then asked to evaluate the quality of the feedback they received on style and quality, namely use of criteria, nature of the feedback, and writing style. The feedback was evaluated using two rubrics. The first related to the style of the feedback (not the quality). It was based on use of criteria, nature of the feedback and writing style. The second rubric was about the quality of the feedback and scored feedback between 0 and 100. The results of the feedback evaluations was analyzed with principal component analysis and found to have more than one dimension: “The results [...] suggest that feedback style has more than one dimension, and that the first two dimensions more or less resemble the feedback skills identified by Sluijsmans (2002).” Overall, the study showed that GPiTs can determine quality of feedback, and that individual differences in preferences was small, with the majority preferring feedback which is descriptive, reflective and personal.

## 2.6 Data

Our data set consists of peer evaluation data from the platform *Peergrade* ([www.peergrade.io](http://www.peergrade.io)) by Wind et al (2017) from 351 courses run throughout two semester periods in 2016. Each course contains a number of assignments (between 1 and 16 per course). Each course contains a number of teachers and a number of students (between 12 and 315). Across all courses, 8% of peer assessments have feedback scores (a numerical evaluation of the feedback quality). The analysis focuses only on this data, which contains 1600 students, 3159 peer evaluations, 981 hand-ins, 139 assignments, 74 courses and 26 institutions. An assignment in a course usually corresponds to a specific task that the students were asked to solve (eg. write an essay or complete a set of math problems). For each assignment the students in the related course upload a hand-in either alone or with other students in a group. After students have handed in their work, the peer assessment phase starts. In this phase, each student will individually assess and give feedback to a number of hand-ins (usually between 3 and 5) from other students. This assessment is done using a number of evaluation criteria created by the teacher for the specific assignment. The evaluation criteria can be either free-form text questions, boolean (yes/no) questions or numerical/scale questions. The teacher is able to define the questions and the possible answers as they want.

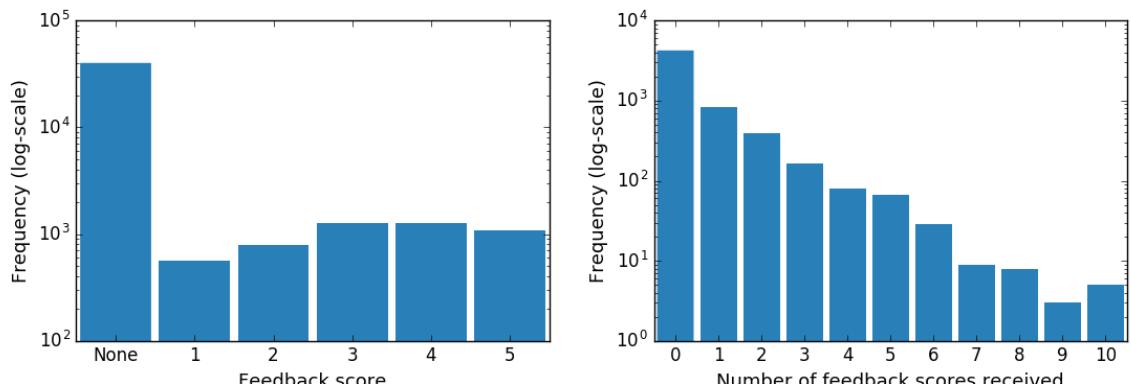
When the peer assessment phase is over the students are able to view the feedback given to them by their peers. They are then able to interact with this feedback in a few ways. Firstly, they can flag feedback for moderation by the teacher, which notifies the teacher and asks them to overwrite the given evaluation if they disagree with the reviewer. Secondly the receiving student is able to rate the quality of the feedback (*feedback score*) on a scale:

- Not at all useful (*corresponds to 0%*)
- Not very useful (*corresponds to 25%*)
- Somewhat useful, although it could have been more elaborate (*corresponds to 50%*)
- Very useful, although minor things could have been better (*corresponds to 75%*)
- Extremely useful, constructive and justified (*corresponds to 100%*)

After the peer assessment session is over, aggregated scores are computed in the system. Each hand-in receives a total *hand-in score* which is the average score assigned to the hand-in by the reviewers (each reviewer's score is in turn an average of their evaluation for each evaluation criteria - not counting text criteria). In addition, each student,  $s$ , receives a total *feedback score* which is the average of feedback scores awarded to  $s$  by the students that received feedback from  $s$ .

## 3. Results

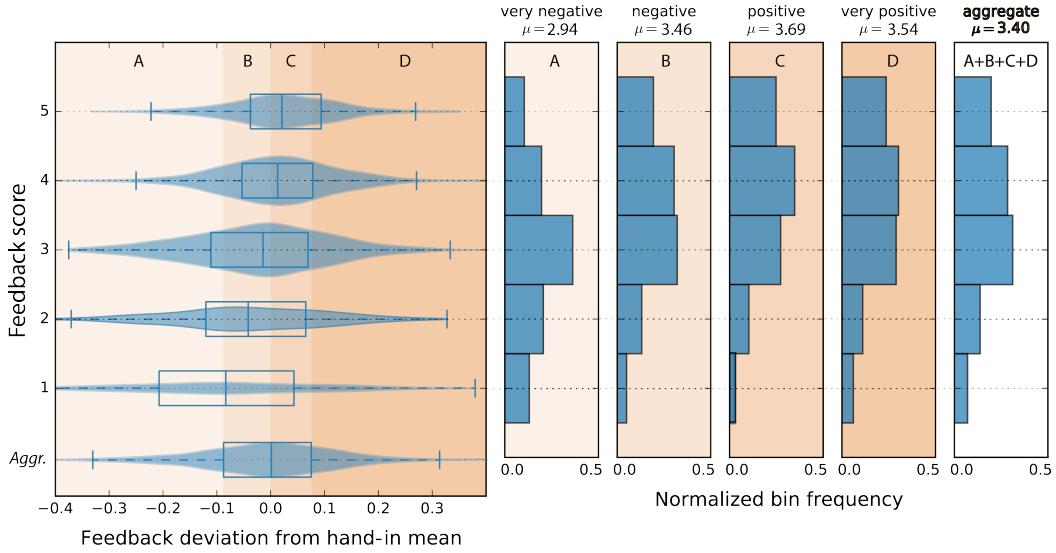
Assessment of work by students is generally a subjective task. Generally, teachers are educated to make good informed choices in these situations, whereas most students are not. In Peergrade, students are asked to review the usefulness of the feedback they receive in order to incentivize giving helpful feedback. Like with peer assessment in general this process of evaluating feedback quality is prone to various biases. Consequently, it is important that the biases involved in the feedback quality rating is well understood.



**Figure 1:** The first histogram shows how the feedback scores are distributed with most feedback not receiving a rating. The second histogram shows the distribution of how many feedback scores each student gets.

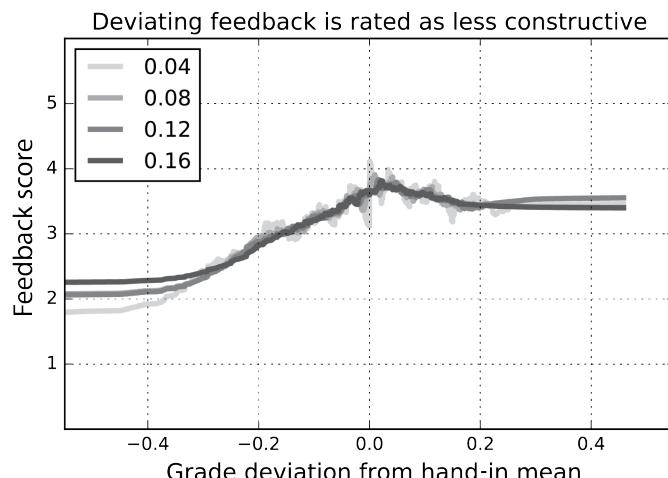
### 3.1 Students consider disagreeing feedback as less helpful

We observe that students tend to award lower feedback scores to feedback that deviates from other feedback they have received on the same hand-in. The histograms in Figure 2 show that the *penalty* for deviating is heavier in the negative direction than in the positive, and that the highest feedback scores are given to feedback that deviates positively by a small amount. A seemingly unintuitive finding is that feedback that is too positive compared to the rest also receives a lower feedback score.



**Figure 2:** There is a tendency for feedback to receive a lower feedback score when it deviates numerically from the mean score of a hand-in. This happens both when the feedback giver is too “harsh” giving the hand-in a lower grade than the other peers did, and when the feedback is too “kind”. The shaded regions A-D, show the ranges between 0%, 25%, 50%, 75% and 100% percentiles of the aggregated distribution (bottom row violin plot). The color gradient increases as feedback moves from very negative to very positive compared to average hand-in grade. The distribution of feedback scores in each region is shown as a histogram, where the background color is consistent with the region shade in the leftmost panel. The rightmost histogram shows the feedback score distribution across all the shaded regions.

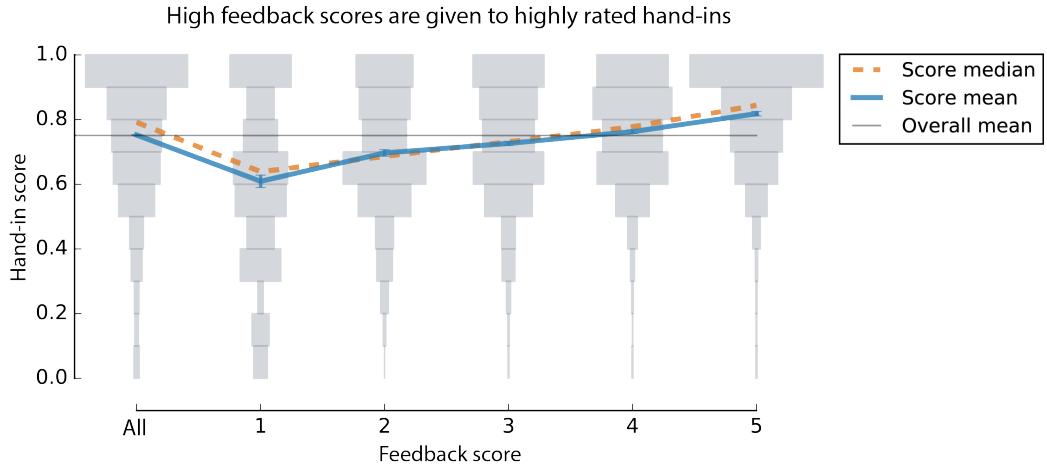
To get a better understanding of exactly how skewed the assignment of feedback scores is, we apply a sliding window filter to the data in order to remove the noise. Figure 3 shows a point-wise window-adjusted feedback score, similar to the mean values shown in Figure 2, but across many more windows. It shows more clearly that there is a bias towards giving higher feedback scores to positively deviating hand-ing graders than to negatively deviating ones.



**Figure 3:** A conceptually similar visualization to that of Figure 2, which shows feedback scores as a function of grade deviation from hand-in mean after application of a sliding window Savitzky-Golay filter (a type of linear regression filter) to every point in the series. The different plots show that the general trend is robust under different window-sizes, ranging from 4% - 16% of the data.

### 3.2 Feedback scores are generally higher on highly rated hand-ins

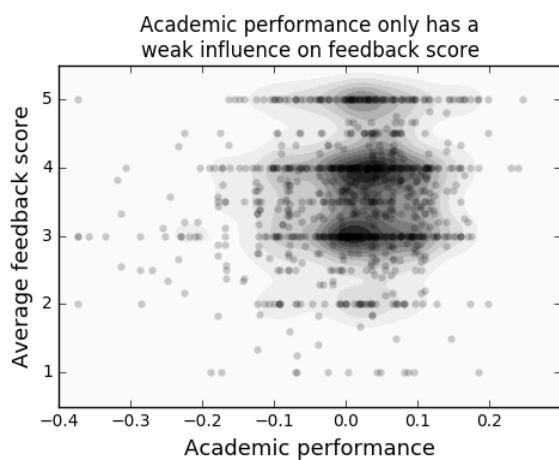
In Figure 4 we show that there is a moderate correlation ( $r = 0.27, p = 3.06E-49$ ) between the feedback score given to a piece of feedback and the overall grade given to the hand-in by the peers. In other words, it shows that peer feedback on good hand-ins receives higher feedback scores and that peer feedback on bad hand-ins receives lower feedback scores.



**Figure 4:** On average, the highest rated feedback is given to the highest rated hand-ins. The horizontal bars give the normalized bin frequencies for each feedback quality grade, the solid and dashed lines intersects with the mean and median respectively.

### 3.3 Academic performance and obtained feedback scores are weakly correlated

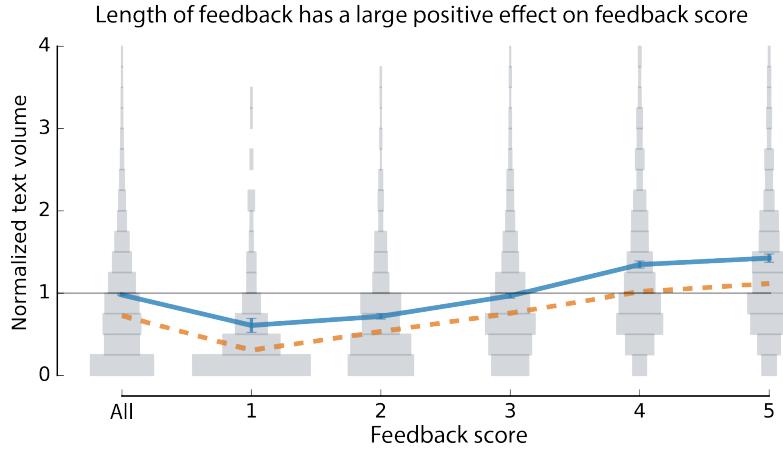
To determine if there is a correlation between the academic performance of a student and their ability to provide helpful feedback to their peers, we plot the aggregated score for each student against their aggregated feedback score in Figure 5. We find a very weak correlation between the academic performance of a student and the perceived quality of the feedback they have given to other peers ( $r=0.11, p=1.64E-3$ ).



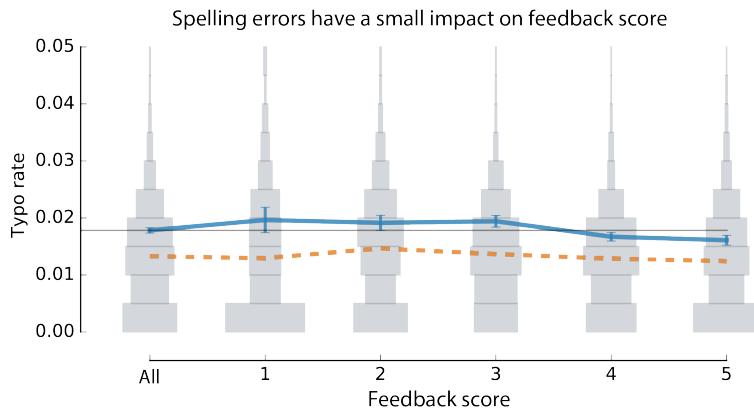
**Figure 5:** There is a weak correlation between the academic performance of students and the feedback scores they obtain on the feedback they provide to their. The horizontal axis shows academic performance measured as the average score across hand-ins relative to the corresponding assignment average scores. The vertical axis shows the average feedback score of feedback they gave to their peers.

### 3.4 High quality feedback is long, positive and without spelling errors

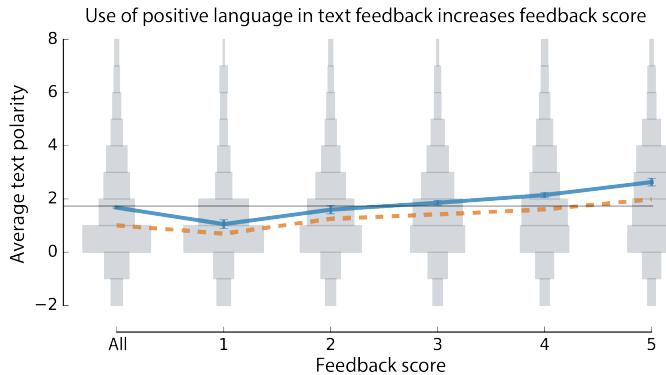
To better understand what makes up high quality feedback, we look at various features of the feedback text and relate it to the feedback score. In Figure 6 we compare the feedback score to the normalized (for the assignment) length of the feedback text and find that there is a 0.24 correlation between the text volume and the feedback quality. In Figure 7 we compare the feedback score to the amount of spelling errors in the feedback text and find that there is a -0.05 correlation between the number of spelling errors (average number of spelling mistakes per word, skipping everything non-english.) and the feedback quality. In Figure 8 we compare the feedback score to the positivity of the words in the feedback text and find that there is a 0.13 correlation between the text polarity and the feedback quality.



**Figure 6:** There is a correlation between the amount of text in the feedback and the perceived quality of the feedback. The horizontal axis shows the feedback score for each peer evaluation. The vertical axis shows the length of the text normalised by the average amount of text given by other peers for the same assignment.



**Figure 7:** There is a very weak correlation between the amount of spelling errors in the feedback text and the perceived quality of the feedback. The horizontal axis shows the feedback score for each peer evaluation. The vertical axis shows the per-word spelling error. Note on the vertical axis that the average typo rate is very low and that the error bars are relatively large. There is a significant difference from the distributions around 1, 2 and 3 to the distributions around 4 and 5, but the effect size is very small.



**Figure 8:** There is a weak correlation between the positivity (polarity) in the feedback text and the feedback score. The horizontal axis shows the feedback score for each peer evaluation. The vertical axis shows the average polarity of text feedback, measured using the AFINN framework by Nielsen (2011).

## 4. Discussion

In this paper we asked students to evaluate the quality of feedback they received from their peers. We used a five-step scale to rank the feedback, which implies that feedback quality is one-dimensional. In Prins et al (2006) they use a larger number of criteria, and using principal component analysis they find that feedback quality is likely to be multi-dimensional. An interesting follow up study would be to ask students to rate the feedback quality using more than one dimension. Prins et al (2006) also take an extra step and let a group of experts rate the feedback quality, to make a comparative study of the validity of the feedback quality evaluations. It would be interesting to make a similar comparison with our dataset. Habeshaw (1993) writes that students “.. are in the best position to know what their difficulties are and to judge what kind of feedback is helpful”. This makes it unclear if comparing of the receiving student to an expert is the right way to ensure validity.

### 4.1 Students consider disagreeing feedback as less helpful

Our results show that students consider feedback which disagrees from the mean as less useful. One potential reason for this is that feedback which is disagreeing with the average is more likely to be wrong, and consequently also less useful. Another potential reason would be that students take “revenge” on their peers if they rate the feedback wrong. If this was the case, we would only expect students to penalize feedback that is too negative. We found that students penalize feedback that is too negative harder than feedback that is too positive. A first probable cause for this is that the average hand-in grade in Peergrade is around 75%, and because of this, it is in most cases mechanically impossible to deviate from the average more in the positive direction than in the negative thus leading to a skew. The second contribution might be that students are more prone to penalize negative feedback because it is perceived as more critical and thus requires more justification. One can propose the argument that if people are naturally biased towards thinking positive of their own work they will agree more with overtly positive graders and therefore require less justification to consider feedback as constructive.

### 4.2 There is a positive correlation between quality of a hand-in and the feedback it receives

We see that hand-ins receiving high scores also receives feedback of high quality and we see the opposite namely that hand-ins with low scores receive feedback of low quality. There are multiple potential reasons for this. One reason could be that it is simply easier to give useful feedback on good work, for example because really poor work might be missing important parts or have no solid content to anchor suggestions for improvement and grade justifications in. Another reason could be that students are more careful and spend more time on evaluating good work. A third potential reason could be related to the revenge / reward patterns described earlier.

### 4.3 There is a weak correlation between academic performance and feedback quality

When looking at all students we find a weak correlation between their academic performance (measured as an average of how good their hand-ins are in relation to their peers) and their feedback quality (the average feedback score of the feedback they have given). One reason for this is that giving helpful feedback is not tied

strongly to the skills needed for doing good in a course. Another reason is the fact that an unknown fraction of the courses in Peergrade do not count feedback quality as part of the final grade, consequently giving less of an incentive to do it well.

#### **4.4 Good feedback is long, positive and with few spelling errors**

Looking at the correlation between the feedback quality and various textual features such as the length of the feedback, the sentiment of the text and the number of spelling errors. We find that good feedback is long, has a positive sentiment and has fewer spelling errors. The positive correlation between length of the feedback and perceived quality can be explained either by the fact that students are likely to perceive long feedback as useful, but also by the fact that only students who take the task of writing feedback seriously write long text, implying an indirect relationship between the length of the feedback and its quality. That a positive sentiment correlates positively with feedback quality can either be due to feedback receivers reacting positively to compliments, or be a result of the fact that good feedback "encourages positive motivational beliefs and self-esteem" (Nicol et al 2006). Finally the negative correlation between the amount of spelling errors in the feedback text and the perceived quality is likely to be a result of the fact that spelling errors in the text makes the feedback giver appear less competent, which influences the receivers view of the validity of the feedback (Strijbos et al 2010).

#### **4.5 Future work**

Based on the discussions in this section it would be great to do a follow up study with the dataset to learn more about the quality of feedback and its relationships to other things. Firstly it would be interesting to extend our measure of feedback quality to be multi-dimensional. Secondly there are other sources of information in Peergrade that we have not looked at, such as "likes" (students are able to mark specific parts of the feedback as useful) and "flags" (where students can mark evaluations they disagree with). Another interesting perspective to take would be to use a quantitative approach to determine which evaluation criteria lead to students writing the most helpful feedback and thus get more insights into the effects of scaffolding through rubric design. Finally it would be interesting to experiment with different ways of teaching students to give useful feedback, and see how it affects the results.

### **5. Conclusion**

The use of peer assessment in education has a number of benefits for students and teachers, including higher learning, increased amounts of feedback, time saving, better classroom culture and other things. One of the most challenging parts of using peer assessment effectively in a classroom is ensuring that the feedback given by the students is useful to the receiving peers. One way to get students to write useful feedback is to incentivise it by giving them a score based on the quality of their feedback. In this paper we propose a way to let students evaluate the quality of the feedback they receive. We have collected data from more than 100,000 peer evaluations by 10,000 students in over 500 courses, making this one of the largest quantitative studies – and widest spanning in terms of subjects and types of courses. Our results show that peers penalize feedback that disagrees with the average feedback, and that good hand-ins get more useful feedback than bad hand-ins. We also explore correlation between textual features of the feedback and its quality, where we find that feedback which is perceived useful is long, positive and has few spelling errors.

### **6. References**

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Bloom, B.S., 1956. Taxonomy of educational objectives: The classification of educational goals .
- Bloxham, S. and West, A. 2004. Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment and Evaluation in Higher Education*, 29: 721–733.
- Boud, D. and Falchikov, N., 1989. Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher education*, 18(5), pp.529-549.
- Cole, D.A., 1991. Change in self-perceived competence as a function of peer and teacher evaluation. *Developmental Psychology*, 27(4), p.682.
- Davies, P., 2003, September. Peer-Assessment: No marks required just feedback? Evaluating the Quality of Computerized Peer-Feedback compared with Computerized Peer-Marking. In Cook, J and McConnell, D (eds),

- Communities of Practice, Research Proceedings of the 10th Association for Learning Technology Conference (ALT-C 2003) (pp. 8-10).
- Gibbs, G. and Simpson, C. 2004. Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1: 3–31.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P. and Struyven, K., 2010. Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), pp.304-315.
- Gielen, S., Tops, L., Dochy, F., Onghena, P. and Smeets, S., 2010. A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British educational research journal*, 36(1), pp.143-162.
- Graham, S., Hebert, M. and Harris, K.R., 2015. Formative assessment and writing. *The Elementary School Journal*, 115(4), pp.523-547.
- Habeshaw, S., Gibbs, G. and Habeshaw, T. (1993) 53 interesting ways to assess your students, Melksham: The Cromwell Press.
- Kim, M., 2016. Effects of the Assessor and Assessee's Roles on Preservice Teachers' Metacognitive Awareness, Performance, and Attitude in a Technology-Related Design Task.
- Higgins, R. 'Be more critical': rethinking assessment feedback. Paper presented at the space British Educational Research Association Conference, Cardiff University.space, September, 7–10, Wales.
- Hounsell, D. 1987. "Essay writing and the quality of feedback". In *Student learning: research in education and cognitive psychology*, Edited by: Richardson, J., Eysenck, M. W. and Piper, D. W. Milton Keynes: Open University Press.
- Kingston, N. and Nash, B., 2011. Formative assessment: A meta-analysis and a call for research. *Educational measurement: Issues and practice*, 30(4), pp.28-37.
- Liu, N.F. and Carless, D., 2006. Peer feedback: the learning element of peer assessment. *Teaching in Higher education*, 11(3), pp.279-290.
- Nielsen, F.Å., 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.
- Nicol, D.J. and Macfarlane-Dick, D., 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), pp.199-218.
- Pope, N. 2001. An examination of the use of peer rating for formative assessment in the context of the theory of consumption values. *Assessment and Evaluation in Higher Education*, 26: 235–246.
- Pope, N. K. L. 2005. The impact of stress in self- and peer assessment. *Assessment and Evaluation in Higher Education*, 30: 51–63.
- Prins, F.J., Sluijsmans, D.M. and Kirschner, P.A., 2006. Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education*, 11(3), pp.289-303.
- Rada, R. and Hu, K., 2002. Patterns in student-student commenting. *IEEE Transactions on Education*, 45(3), pp.262-267.
- Rust, C., Price, M. and O'donovan, B. 2003. Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28: 147–164.
- Sadler, D. R. 1998. Formative assessment: revisiting the territory. *Assessment in Education*, 5: 77–84.
- Sadler, P.M. and Good, E., 2006. The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1), pp.1-31.
- Searby, M. and Ewers, T., 1997. An evaluation of the use of peer assessment in higher education: A case study in the School of Music, Kingston University. *Assessment & Evaluation in Higher Education*, 22(4), pp.371-383.
- Sluijsmans, D. 2002. Student involvement in assessment: the training of peer assessment skills, Heerlen: Open Universiteit Nederland. PhD dissertation.
- Smith, E. and Gorard, S., 2005. 'They don't give us our marks': the role of formative feedback in student progress. *Assessment in Education: Principles, Policy & Practice*, 12(1), pp.21-38.
- Strijbos, J.W., Narciss, S. and Dünnebier, K., 2010. Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?. *Learning and instruction*, 20(4), pp.291-303.
- Topping, K. J. 2003. "Self- and peer assessment in school and university: reliability, validity and utility". In *Optimizing new modes of assessment: in search of qualities and standards*, Edited by: Segers, M., Dochy, F. and Cascallar, E. Dordrecht: Kluwer Academic.
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9, 147e170.
- Wind, David Kofoed, Rasmus Malthe Jørgensen, and Simon Lind Hansen. "Peergrade - Better Feedback For Your Students". Peergrade. N.p., 2017. Web. 6 Jan. 2017.

Yang, M., Badger, R. and Yu, Z. 2006. A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15: 179–200.

# Optimal allocation of reviewers for peer feedback

David Kofoed Wind<sup>1,3</sup>, Ulf Aslak Jensen<sup>2</sup>, Rasmus Malthe Jørgensen<sup>3</sup>, Simon Lind Hansen<sup>3</sup>, Ole Winther<sup>1</sup>

<sup>1</sup>Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>Peergrade, Copenhagen, Denmark

[dawi@dtu.dk](mailto:dawi@dtu.dk)

[uaj@econ.ku.dk](mailto:uaj@econ.ku.dk)

[olwi@dtu.dk](mailto:olwi@dtu.dk)

[malthe@peergrade.io](mailto:malthe@peergrade.io)

[simon@peergrade.io](mailto:simon@peergrade.io)

## 1. Abstract

Peer feedback is the act of letting students give feedback to each other on submitted work. There are multiple reasons to use peer feedback, including students getting more feedback, time saving for teachers and increased learning by letting students reflect on work by others. In order for peer feedback to be effective students should give and receive useful feedback. A key challenge in peer feedback is allocating the feedback givers in a good way. It is important that reviewers are allocated to submissions such that the feedback distribution is fair - meaning that all students receive good feedback.

In this paper we present a novel way to intelligently allocate reviewers for peer feedback. We train a statistical model to infer the quality of feedback based on a dataset of feedback quality evaluations. This dataset contains more than 20,000 reviews where the receiver of the feedback has indicated the quality of the feedback. Using this model together with historical data we calculate the feedback-giving skill of each student and uses that as input to an allocation algorithm that assigns submissions to reviewers, in order to optimize the feedback quality for all students.

We test the performance of our allocation strategy using real data from over 600 peer feedback sessions and simulate the effects of different allocation strategies. By comparing our method with a random allocation algorithm and a “super-informed oracle” algorithm we demonstrate that we are able to allocate reviewers to submissions in such a way that all submissions receive feedback of similar quality and that we are able to significantly outperform simple random allocation of reviewers. Additionally we investigate the effect of pre-allocating reviews in comparison to allocating reviewers live during the review process and show that live-allocation leads to better results. Our method is robust to reviews not being completed and other real-life quirks and improves as more feedback data is collected.

**Keywords:** peer assessment, peer feedback, feedback, peer review, peer evaluation, peer grading, task assignment, reviewer allocation

## 2. Introduction

### 2.1 What is peer assessment?

Peer assessment is the act of letting students partake in the process of evaluation and giving feedback. Peer feedback is a formative variant of this where students discuss the strengths and weaknesses of a specific piece of work at length and indicate suggestions for further improvement. It is the counterpart of feedback by a teacher. Both are outcomes of formative assessment, also called assessment for learning (Black & Wiliam, 1998; Gielen et al 2010).

### 2.2 Why is it meaningful to use peer assessment?

The use of peer assessment in education comes with a number of benefits for both teachers and students (Tsui & Ng, 2000). Peer feedback can increase pressure on students to perform well by introducing a social pressure. Cole (1991) finds this effect in fourth grade and Pope (2005) in university. An increase in time spent on assignment that are given peer feedback is documented in (Tsui & Ng, 2000; Pope, 2001; Gibbs & Simpson, 2004). When an entire class of students collaborate on evaluating homework and with a good tool in place

where the logistics are handled, the teacher can save a large amount of time and students can receive feedback on their work much faster (Sadler, 2006; Searby, 1997). Gibbs et al (2004) writes that getting imperfect feedback immediately might have much more impact than perfect feedback four weeks later. Gielen et al (2010) also talks about the possibility of introducing more feedback into the classroom through “several ‘intermediate’ peer assessment sessions on draft versions of for instance an essay or report” arguing that this “could answer to this need of regular feedback if teachers are not able or willing to increase its frequency themselves.”. When students assess and give feedback to work of other students, they get a deeper understanding of the material, both from seeing alternative solutions to the task they have solved themselves and from reflecting upon the work of others and formulating their thoughts into constructive feedback (Bloom et al, 1956; Boud, 1989; Sadler, 2006; Liu, 2006). Topping (2003) describes that students find peer feedback more understandable and useful due to other students being ‘on the same wavelength’, and teacher feedback is often misinterpreted by students due to a difference in discourse (Hounsell, 1987; Higgins, 2000; Gibbs et al., 2004; Yang et al., 2006). Giving feedback is also a good way to train the ability to receive and understand feedback (Rust et al., 2003; Bloxham & West, 2004). The use of peer evaluations in a class can even make the classroom friendlier, more cooperative and consequently build a greater sense of shared ownership for the learning process (Sadler & Good, 2006). The process of peer assessment is not limited to educational purposes, and the use of self-evaluation and peer review are an important part of future, adult, professional practice, and test grading is a good way to develop these skills (Boud, 1989).

### 2.3 The importance of reviewer allocation

There are many ways to practically carry out a peer feedback session. In the following, we assume a process where N students submit M submissions either alone or in groups before a fixed deadline to an online web service, after which each student then becomes an independent reviewer who has a fixed amount of time – usually in the order of days – to review L submissions that cannot include their own. The reviewers can review anytime they want within the review period but have to submit a review before they can start reviewing the next submission.

The assumed system is equivalent to the one employed in the online peer feedback system Peergrade (Wind & Jørgensen & Hansen 2017). Peergrade also allows students to hand in late in certain cases and allows teachers to extend deadlines for handing in and giving peer feedback. Furthermore, it also – necessarily – handles cases where students do not complete their submissions or reviews. Exceptions like these call for robust methods to allocate submissions to reviewers in a way that does not undermine the potential value of peer feedback.

During a peer feedback session it is important that reviewers are allocated to submissions such that the feedback distribution is even and fair. Feedback being even means that all submissions receive the same number of reviews. If there are N reviewers each giving feedback L times on a total of M handins, feedback can be distributed evenly if  $N \times L / M$  is an integer. This condition is for example met if students hand in individually ( $N = M$ ). In practise, however, students may work in unevenly sized groups, forget to hand in and neglect their duty to review. If reviewers are allocated to submissions at random, this can create situations where some submissions only receive a single or no reviews. When  $N \times L / M$  is not an integer, the goal is to never have a submission receiving 2 reviews more than another submission (since in this case the reviews could have been distributed more evenly).

To distribute feedback fairly means that each submission receives the same level of feedback quality. This is a more interesting metric to optimize for because there may otherwise arise cases where, at the end of a session, some submissions have only received poor feedback - likely coming from reviewers who are not qualified to appropriately judge the work at hand.

Solving the problem of even feedback distribution is trivial and – as we shall see later – follows naturally from the solution to the problem of fairness distribution.

### 2.4 Previous work

In Wind & Aslak (2017) we introduce and discuss the way we measure quality of feedback by letting students evaluate the quality of the feedback they receive. In this paper we will use this measure of feedback quality as a way to measure the distribution of feedback quality across submissions.

Piech et al (2013) are interested in inferring accurate grades using peer review. They develop mathematical models to infer accurate grades, and show that using these models it is possible to determine when the grade of a submission is accurate enough, and until then assign more reviewers to it.

Han et al (2017) proposes an algorithms for assigning reviews to submissions in a smart way. Their goal is to ensure accuracy in grading, and consequently their objective is to reduce the variance between grades on a paper. To make decisions about which paper a student should review, they use a model trained on how the students interacted with the course material (video watching and other similar metrics) to predict the competence of a student. This requires data gathered before the actual peer review process starts. Additionally they make two important assumptions about the review process: Firstly that all students will finish their reviews (their original dataset of 600 students is reduced to 210 due to this assumption) and secondly they ignore edge-cases like students handing in late and students starting a review but not finishing it in time.

## 2.5 Our contribution

In the following we demonstrate a novel way to intelligently allocate reviewers in peer feedback. We train a statistical model to infer the quality of feedback based on a dataset of feedback quality evaluations. This dataset contains more than 20,000 reviews where the receiver has indicated the usefulness. Using this model and historical data we calculate the feedback-giving skill of each student and uses that as input to an allocation algorithm that assigns submissions to reviewers, in order to optimize the feedback quality for all students. Note that in contrast to Piech et al (2013) and Han et al (2017) we are optimizing for the quality of written feedback, not accuracy of grades or distribution of academic performance for reviewers. We test the performance of our approach using real data from over 600 peer feedback sessions run in Peergrade.

## 2.6 Data

Our data set consists of peer evaluation data from the platform Peergrade ([www.peergrade.io](http://www.peergrade.io)) by Wind et al (2017) from 623 assignments run during 2017 with a total of 40,931 reviews completed. For each assignment a number of students (between 12 and 315) participate. An assignment usually corresponds to a specific task (eg. write an essay or complete a set of math problems) that the students were asked to solve and submit. For each assignment the students in the related course upload their submission either alone or with other students in a group. After students have handed in their work, the peer assessment session starts. In this session, each student will individually assess and give feedback to a number of submission's (usually 3) from other students. The feedback process is double blind (students don't know who they review or who review their own work).

When the peer assessment session is over the students are able to view the feedback given to them by their peers. They are then able to rate the quality of the feedback (the so-called feedback score) on a scale:

- Not at all useful (corresponds to a score of 1)
- Not very useful (corresponds to a score of 2)
- Somewhat useful, although it could have been more elaborate (corresponds to a score of 3)
- Very useful, although minor things could have been better (corresponds to a score of 4)
- Extremely useful, constructive and justified (corresponds to a score of 5)

The total number of reviews with a feedback score used for this paper is 13,337. The feedback used is available for other researchers that want to work with it. To get access to the data, contact the authors.

## 3. Method

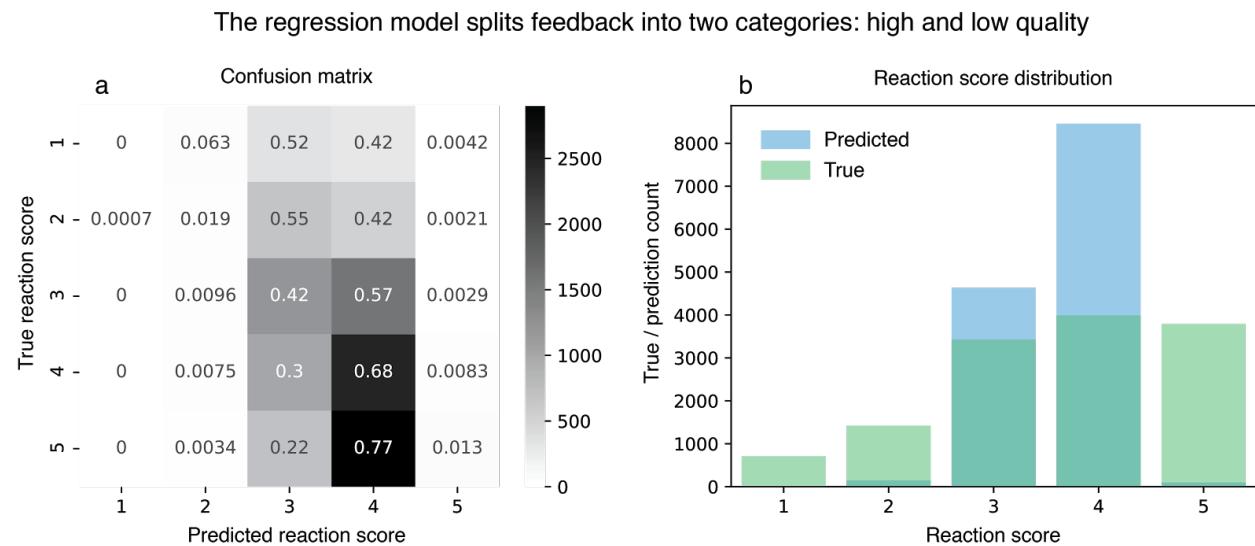
Our approach to the reviewer allocation problem comprises two subproblems. The first is that we need to know how good a piece of feedback is, and how good each student is at giving feedback in general. For this we build a scoring model which uses historical feedback performance data to evaluate the feedback-giving skill level of the students. The second challenge is to construct an algorithm which uses this information together with the current distribution of feedback quality at any time, to allocate the next piece feedback in a way that maximizes fairness in the end.

### 3.1 Feedback skill scoring

To ensure that all students receive feedback of similar quality during a peer feedback session, it is imperative that we can score the feedback-giving skill of each student. After a feedback session in Peergrade, students can react to their received feedback with a quality score between 1 and 5 (the feedback score). When a student has received many feedback scores it gives a strong indication of their skill level. However, this feature is only used around 10% of the time so for most students relying on it alone is either impossible, or incurs a high risk of misclassification. To get an informed estimate we train a regression model to predict the feedback score given the feedback text. Using this model we can predict what score a piece of feedback *would have* most likely received. Taking the average of obtained/historical feedback scores and predictions of feedback scores for all pieces of feedback that a student has given provides a fair and accurate measure of the student's overall skill level.

Our prediction model uses a Random Forest regressor (Breiman 2001) with feedback encoded in five features, which are: (1) text length, (2) number of evaluation rubrics filled out, (3) time spent giving feedback, (4) frequency of typos in the text and (5) text sentiment (Nielsen, 2011). These features (except for the text sentiment) are normalised for each assignment to make it possible to compare feedback from different peer feedback sessions. Using this model we infer feedback scores for the feedback that has not yet received a feedback score. Finally we take the average feedback score a feedback giver has received as their feedback-giving skill.

Validation tests with 10-fold cross validation yields a generalization correlation between predicted and true scores of 0.33. The mean squared error (MSE) measured on unseen data is 1.19, which compares well to a majority class baseline of 1.44 and a randomized class baseline of 2.64. Fig. 1 shows further validation of the regressor performance. The main observation is that our regressor is able to successfully distinguish between "good" and "bad" feedback, but has trouble telling 1, 2 and 3 apart and 4 and 5 apart. Most feedback is predicted as 3 or 4 because the fitting algorithm optimizes the MSE during training such that predicting extreme low or high scores incurs a possibly higher MSE. For the purpose of the forthcoming analysis, however, it is good enough that we are able to distinguish good and bad feedback.



**Figure 1: Regression splits the data into high and low quality feedback.** (a) Regressor confusion matrix. The dark squares in columns 3 and 4 show that the regressor predicts these values for most input. The annotations correspond the frequency at which feedback with true score corresponding to the row value is predicted as that corresponding to the column value. In large, positively rated feedback is predicted as 4, while negatively rated feedback is predicted as 3. (b) Distribution of true and predicted feedback scores. Due to the cost function used in the optimization protocol for the model we use, mispredicting 1, 2 and 5 are sufficiently expensive that the model avoids predicting these scores almost entirely.

### **3.2 Feedback allocation**

Generally feedback allocation can be done in two ways; either before the review session starts (pre-allocation) or live during the review session (online-allocation). The challenge with pre-allocation is that if just a fraction of the students do not complete their reviews, then there is a high likelihood of a submission getting no feedback at all.

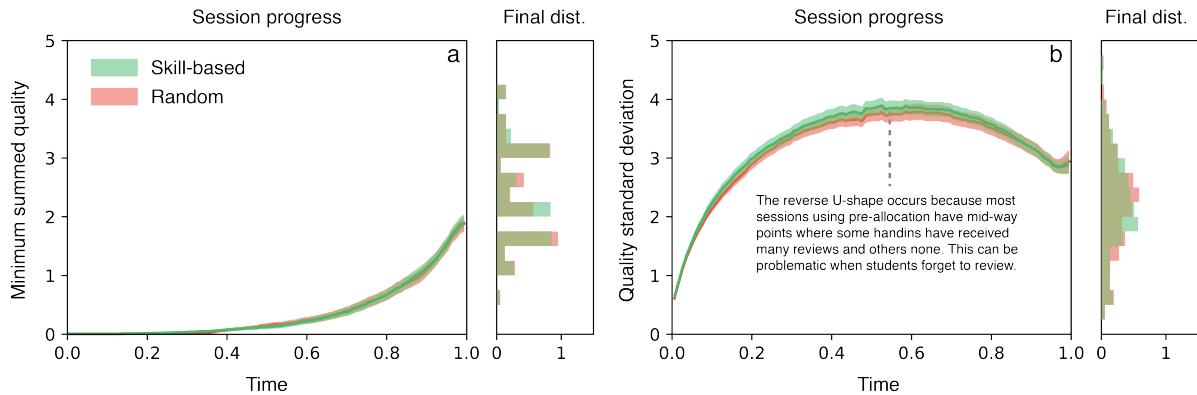
Assuming that we can predict the quality of a review before it has been given, we want to use this information to make a fair allocation. In this paper we will be comparing three different allocation-algorithms; a random allocation, a skill-based allocation (our proposed method) and an oracle-allocation. In a random allocation, reviews are allocated without considering feedback quality - but still in a way that ensures evenness (every submission should receive approximately the same number of reviews). This random allocation can happen either before the review session (pre-allocation) or during the review session (online allocation). The skill-based allocation considers the quality of already completed reviews in the ongoing peer feedback session, and estimates quality of future reviews in order to make the review-allocation more fair (every submission should receive high quality feedback). The oracle-allocation is used as a benchmark and assumes knowledge of the actual future feedback (information which is not available at the allocation-time) in order to make the optimal allocation in an online-manner. The objective is to make an allocation algorithm that gets as close to the oracle-allocation as possible.

In order to test our allocation algorithms, we use actual peer feedback data from the system Peergrade. We use existing peer feedback sessions that were run in the past, and use them to simulate different allocation scenarios. For a completed peer feedback session we have the reviews, their qualities (for some reviews), who did them and when they were completed. To evaluate an allocation-strategy we can “run” the peer feedback session again, but this time allocating the reviews differently. By simulating different ways reviews are allocated we can measure the impact of different allocation strategies. This approach is based on the assumption that the quality of a review is not dependent on the submission being reviewed. This assumption will be addressed later. The strength of using this approach is that our simulations closely resemble actual peer feedback sessions in terms of feedback quality, temporal dynamics and real-life quirks (such as students not completing their assigned reviews).

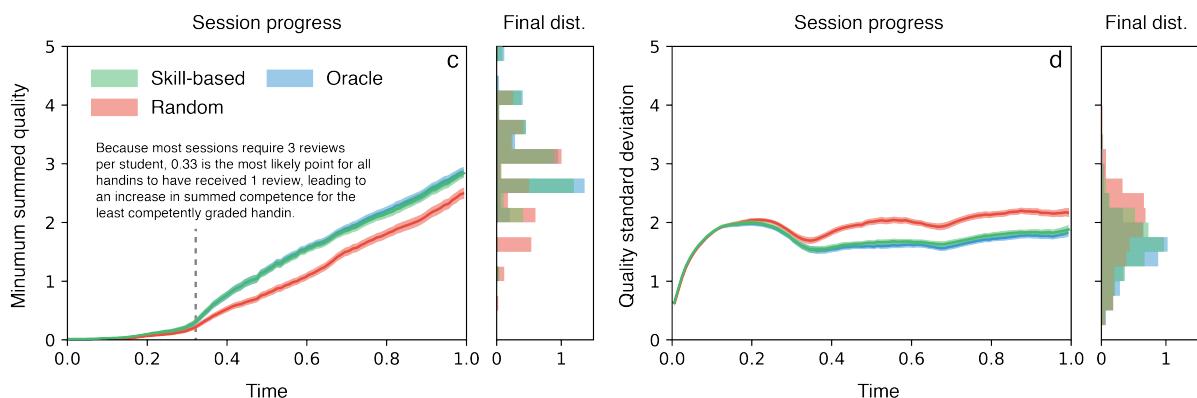
## **4. Results**

We will use two different measures to evaluate the quality of a reviewer allocation. The first measure is the “minimum summed quality” which is the sum of feedback quality of feedback given to the submission which received the lowest quality feedback. To optimize (maximize) this first measure, the submission receiving the worst feedback should receive as much good feedback as possible. The second measure is the “quality standard deviation” which is the standard deviation of feedback quality given to different submissions. To optimize (minimize) this measure all submissions should receive feedback of similar quality.

**Pre-allocation:** Skill and random based allocation perform the same



**Online allocation:** Skill based allocation leads to more fair quality distribution



**Figure 2: Online allocation with respect to student reviewing skill, ensures more evenly distributed feedback quality across submissions.** Summing the quality of feedback received for each submission and taking the minimum value gives a fair estimate of how fair an allocation algorithm distributes feedback quality across a collection of submissions. During a session this value can be computed for each feedback submission to produce a time-series of the performance of an allocation algorithm. Here, we produce such a time-series for each peer feedback session in our dataset (623), and estimate a sliding mean and confidence interval to obtain a curve that illustrates the evolution of performance for different algorithms. We compare five algorithms here, (a-b) two that preallocate submissions to reviewers, and (c-d) three that allocate a submission to a reviewer in the moment she requests it (online method). The skill-based algorithms compute the student's feedback skill at the moment she requests a submission, by taking the average of past true and predicted feedback scores (Fig 1.b). We also compare to an *oracle* algorithm, which assigns submissions to reviewers with knowledge of the outcome feedback score of the review. Here, we strictly use predicted feedback scores, since the short time-scale of a real-world peer feedback session renders student feedback scores inaccessible for use in allocation.

Pre-allocation algorithms (Fig. 2a-b) perform at the same level, whether the allocation process is informed by student skill levels or not. In Fig. 2a we observe that the minimum summed quality remains very low throughout the session and only increases towards the end to reach 1.95 and 1.96 for random and skill respectively (95% confidence intervals overlapping). In Fig. 2b we see that that the quality standard deviation (the “unevenness” in quality distribution) reaches high values midway through the session and decreases towards the end, settling at 2.96 and 2.85 for random and skill, respectively (95% confidence intervals overlapping). Both of these tendencies are artifacts of a central flaw in the pre-allocation approach, which is that allowing unevenness during the span of the peer feedback session makes the process highly vulnerable to disturbances such as students not fulfilling their duty to review or server failures that cut the session short.

Online algorithms continuously allocate for fairness, which is why the minimum summed quality increases from zero at an early point (Fig. 2c-d). This occurs characteristically around  $t = 0.33$ , which is because the bulk of sessions require students to hand in alone and review three times. The skill-based online allocation

algorithm causes the minimum summed quality to increase from this point towards a final value of 2.84. This is significantly better than the random online allocation algorithm which leads to a final minimum summed quality of 2.50. The online oracle algorithm does not perform significantly better than the skill-based algorithm, reaching a final value of 2.86. This is likely because students are consistent enough in the quality of their review that knowing the outcome of a review is practically as good as knowing the performance history of a reviewer. Compared to the global mean feedback score of 3.65, which the average minimum summed quality that we measure here cannot exceed for tautological reasons, 2.85 is fair.

## 5. Discussion

In order to measure the impact of different strategies for allocating reviewers in peer feedback we used a simulated approach based on real data. This approach has the benefit that our simulations resemble real-life peer feedback sessions when it comes to timing of reviews, quality of reviews and how some students do not complete giving their feedback. In order to make this simulation, we need to assume that the quality of a piece of feedback is not dependent on the submission receiving the feedback. As shown in Wind & Aslak (2017), this assumption does not hold completely, but the alternative is to use different allocation strategies in real peer feedback sessions and try to compare them, which would introduce stronger assumptions.

Classification of low prediction scores proves difficult when only considering text features. In Wind & Aslak (2017) we show that the peer feedback score is a weak but significant function of the numerical evaluation given by the reviewer, pointing to a feedback scoring bias due to “revenge”. We also show that feedback scores are higher on highly rated handins. Both of these non-text-based factors may explain that low feedback scores are difficult to predict only using text. At the same time, it is worth considering whether it is wise to include features in the model produced by student subjectivity, since these may fit to a bias that is generally undesired.

When comparing different strategies for allocation we use the “minimum summed quality” as a measure of allocation quality. This measure is chosen to ensure that no student receives very poor feedback. Other measures be chosen which will lead to other review allocations. All of our allocation strategies work under the constraint that each submission should receive the same number of reviews (the evenness constraint). Relaxing this constraint would possibly allow for a better allocation in terms of “minimum summed quality”, but students that receive fewer reviews might perceive this as unfair (even though their reviews are of higher quality).

### 5.1 Future work

In this paper we developed a new allocation strategy that tries to optimize the total quality of feedback given to each submission. It might also be interesting to allocate reviews in order to have students review submissions of different qualities or in a way where each submission receives reviews from students with different academic competences (how good their submissions are). Furthermore it would be interesting to develop an allocation strategy combining these different allocation goals.

All allocation strategies we presented in this paper worked under the evenness constraint. Relaxing this constraint might allow for more fair allocations, but could have implications on the perceived fairness from feedback receivers. It would be relevant to investigate this perceived fairness further.

Our approach is based on the possibility to infer the quality of feedback from the feedback text itself, and on predicting the quality of feedback in an upcoming review. Improving on these problems will lead to improvements in allocation of reviews as well - and might be the subject of a future study on their own.

## 6. Conclusion

In this paper we have presented a new way of allocating reviews in a peer feedback session. Our approach learns the feedback-giving skill of each student and uses that as input to an algorithm that assigns submissions to students, such that feedback quality is optimized for all students. We tested the performance of our allocation strategy using real data from over 600 peer feedback sessions from the system Peergrade. By comparing with an “uninformed random” allocation algorithm as well a “super-informed oracle” algorithm we demonstrate that we can allocate reviewers to handins such that all submissions receive similar quality

feedback and where we significantly outperform the random allocation. Our method is also robust to reviews not being completed and other real-life quirks.

## 7. References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Bloom, B.S., 1956. Taxonomy of educational objectives: The classification of educational goals .
- Bloxham, S. and West, A. 2004. Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment and Evaluation in Higher Education*, 29: 721–733.
- Boud, D. and Falchikov, N., 1989. Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher education*, 18(5), pp.529-549.
- Breiman, L. (2001). Machine Learning, 45(1), 5-32. Random Forests.
- Cole, D.A., 1991. Change in self-perceived competence as a function of peer and teacher evaluation. *Developmental Psychology*, 27(4), p.682.
- Gibbs, G. and Simpson, C. 2004. Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1: 3–31.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P. and Struyven, K., 2010. Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), pp.304-315.
- Gielen, S., Tops, L., Dochy, F., Onghena, P. and Smeets, S., 2010. A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British educational research journal*, 36(1), pp.143-162.
- Higgins, R. 'Be more critical!': rethinking assessment feedback. Paper presented at the space British Educational Research Association Conference, Cardiff University.space, September, 7–10, Wales.
- Hounsell, D. 1987. "Essay writing and the quality of feedback". In *Student learning: research in education and cognitive psychology*, Edited by: Richardson, J., Eysenck, M. W. and Piper, D. W. Milton Keynes: Open University Press.
- Liu, N.F. and Carless, D., 2006. Peer feedback: the learning element of peer assessment. *Teaching in Higher education*, 11(3), pp.279-290.
- Nielsen, Finn Årup. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." arXiv preprint arXiv:1103.2903 (2011).
- Pope, N. 2001. An examination of the use of peer rating for formative assessment in the context of the theory of consumption values. *Assessment and Evaluation in Higher Education*, 26: 235–246.
- Pope, N. K. L. 2005. The impact of stress in self- and peer assessment. *Assessment and Evaluation in Higher Education*, 30: 51–63.
- Rust, C., Price, M. and O'donovan, B. 2003. Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28: 147–164.
- Sadler, D. R. 1998. Formative assessment: revisiting the territory. *Assessment in Education*, 5: 77–84.
- Sadler, P.M. and Good, E., 2006. The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1), pp.1-31.
- Searby, M. and Ewers, T., 1997. An evaluation of the use of peer assessment in higher education: A case study in the School of Music, Kingston University. *Assessment & Evaluation in Higher Education*, 22(4), pp.371-383.
- Mikolov, Tomas & Chen, Kai & Corrado, Greg and Dean, Jeffrey. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013.
- Topping, K. J. 2003. "Self- and peer assessment in school and university: reliability, validity and utility". In *Optimizing new modes of assessment: in search of qualities and standards*, Edited by: Segers, M., Dochy, F. and Cascallar, E. Dordrecht: Kluwer Academic.
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9, 147e170.
- Wind, David Kofoed, Rasmus Malthe Jørgensen, and Simon Lind Hansen. "Peergrade - Better Feedback For Your Students". Peergrade. N.p., 2017. Web. 6 Jan. 2017.
- Yang, M., Badger, R. and Yu, Z. 2006. A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15: 179–200.

## 5 Software projects

Here I present two Python packages that I have developed to facilitate my research. The software is open source and written in a general way so it can be used by other researchers. Both projects are documented as short summary articles. The first is published in the Journal of Open Source Software, while the other has been submitted to the arXiv.

1. The first paper presents *Netwulf*, a Python package that enables reproducible interactive visualization of networks directly from Python. Netwulf accepts a network in either `networkx.Graph` format [29] or as a node-link dictionary and launches a new browser window with a live visualization of the network powered by JavaScript and the `d3-force` network layouting API [30]. The user can modify both style and physics of the network by dragging sliders and moving nodes around. The final image can then be exported back to Python where it is rendered as a `matplotlib` figure [31], enabling further style manipulation and saving to multiple formats.
2. The second paper presents *Infostop*, a Python package that offers a simple API for clustering mobility trajectories into points of interest, or so-called *stop locations*. It uses Infomap for clustering points as nodes in a network [32], which has the implication that stop locations which are so close that samples obtained from them slightly overlap are *not* necessarily given the same label, which is otherwise the case with most spatial clustering algorithms. As such Infomap can, for example, detect fine-grained stop locations such as rooms within buildings.

# Netwulf: Interactive visualization of networks in Python

Ulf Aslak<sup>1, 2</sup> and Benjamin F. Maier<sup>3, 4</sup>

<sup>1</sup> Center for Social Data Science, University of Copenhagen, DK-1353 København K 2 DTU Compute, Technical University of Denmark, DK-2800 Kgs. Lyngby <sup>3</sup> Robert Koch-Institute, Nordufer 20, D-13353 Berlin <sup>4</sup> Department of Physics, Humboldt-University of Berlin, Newtonstr. 15, D-12489 Berlin

DOI: [10.21105/joss.01425](https://doi.org/10.21105/joss.01425)

## Software

- [Review ↗](#)
- [Repository ↗](#)
- [Archive ↗](#)

Submitted: 25 April 2019

Published: 31 August 2019

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

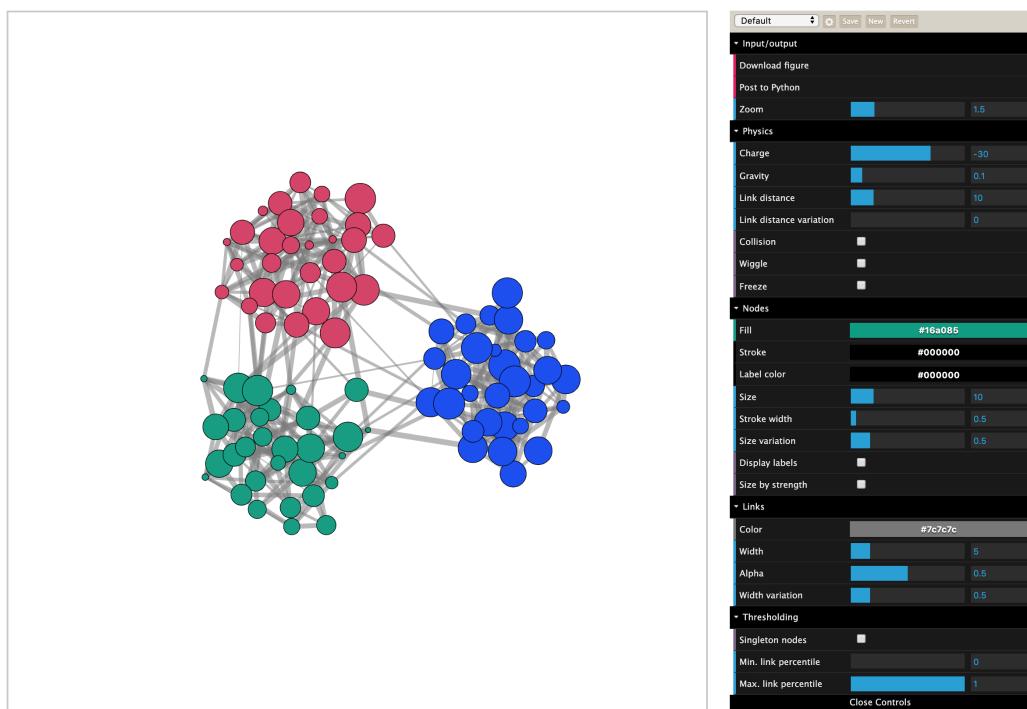
## Summary

Network visualization is an effective way to illustrate properties of a complex system. It is an important tool for exploring and finding patterns, and is used by researchers and practitioners across many fields and industries. Currently, there exist a number of tools for visualizing networks. *Networkx* (Aric A. Hagberg & Swart, 2008) is a popular Python package for network analysis which provides limited functionality for computing layouts and plotting networks statically. Layout computations are done in Python or using the php-based software *Graphviz* (Ellson et al., 2001), which is slow. Another Python package for network analysis and visualization is *graph-tool* (Peixoto, 2014), which relies on a high number of external C++-libraries for installation/compilation which can be overwhelming for beginners. Furthermore, its visualization functions are non-interactive, as well. *Gephi* (Bastian, Heymann, & Jacomy, 2009) and *Cytoscape* (Shannon et al., 2003) are dedicated interactive visualization and analysis software programs. They are both Java-based and run desktop clients with a GUI where users save and load networks as separate files. *Webweb* (Wapman & Larremore, 2019) enables interactive visualization for Python and Matlab networks using the d3.js (Bostock, Ogievetsky, & Heer, 2011) force layout. Its main purpose, however, is exploration of network features and exporting one-time visualizations as SVG or HTML.

For many users, these tools offer the necessary functionality to visualize networks in most desired ways. However, since a growing population of network researchers and practitioners are relying on Python for doing network science (Overflow, 2019), it is increasingly pressing that a fast and intuitive Python tool for reproducible network visualization exists.

*Netwulf* is a light-weight Python library that provides a simple API for interactively visualizing a network and returning the computed layout and style. It is build around the philosophy that network manipulation and preprocessing should be done programmatically, but that the efficient generation of a visually appealing network is best done interactively, without code. Therefore, it offers no analysis functionality and only few exploration features, but instead focuses almost entirely on fast and intuitive layout manipulation and node/link styling. Interaction with Netwulf typically works as follows:

1. Users have a network object, G, in either dictionary or *networkx.Graph* format. They then launch a Netwulf visualization by calling `netwulf.visualize(G)`.
2. The command opens a new browser window containing G as an interactive, manipulable, stylable network. Here, the user can, for instance, explore how different configurations of physics parameters like *node charge* and *gravity* influence the layout, they can change properties like node color and link opacity, and even threshold the network data for weak or strong links. When the user has finalized the layouting process, they may either:
  1. Save the image directly from the interactive visualization as a PNG file.



**Figure 1:** Interactive visualization of a modular network in Netwulf.

- Post the style and computed node positions back to Python in a dictionary format, which allows for further manipulation in the Python backend. Moreover, using the function `netwulf.draw_netwulf`, the network can be redrawn using the common Python drawing library `matplotlib` (Hunter, 2007), which further enables saving the visualization in any format.

The interactive visualization is implemented in JavaScript, relies on d3.js (Bostock et al., 2011) for computing layouts, and uses the HTML5-object canvas for rendering. This makes it, to our knowledge, the most performant tool for interactive network visualization in Python to date.

## Acknowledgements

Both authors contributed equally to the software, documentation, and manuscript. B. F. M. is financially supported as an *Add-On Fellow for Interdisciplinary Life Science* by the Joachim Herz Stiftung.

## References

- Aric A. Hagberg, D. A. S., & Swart, P. J. (2008). *Exploring network structure, dynamics, and function using NetworkX*. (G à el Varoquaux Travis Vaught & J. Millman, Eds.) (Vol. 7). Pasadena, CA USA: SciPi2008.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In. doi:[10.13140/2.1.1341.1520](https://doi.org/10.13140/2.1.1341.1520)

- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data - Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12). doi:[10.1109/TVCG.2011.185](https://doi.org/10.1109/TVCG.2011.185)
- Ellson, J., Gansner, E., Koutsofios, L., North, S., Woodhull, G., Description, S., & Technologies, L. (2001). Graphviz open source graph drawing tools. In *Lecture Notes in Computer Science* (pp. 483–484). Springer-Verlag.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3), 90–95. doi:[10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Overflow, S. (2019). Stack overflow developer survey. Retrieved April 23, 2019, from
- Peixoto, T. P. (2014). The graph-tool python library. *figshare*. doi:[10.6084/m9.figshare.1164194](https://doi.org/10.6084/m9.figshare.1164194)
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), 2498–2504. doi:[10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
- Wapman, H. K., & Larremore, D. B. (2019). Webweb: A tool for creating, displaying, and sharing interactive network visualizations on the web. *Journal of Open Source Software*. doi:[10.21105/joss.01458](https://doi.org/10.21105/joss.01458)

# Infostop: Python package for detecting stop locations in mobility data

Ulf Aslak\* and Laura Alessandretti

*Centre for Social Data Science, University of Copenhagen, DK-1353 København K and  
DTU Compute, Technical University of Denmark, DK-2800 Kgs. Lyngby*

(Dated: September 2, 2019)

Detecting places of interest, or so-called *stop locations* in movement trajectories of people is of great interest to researchers in multiple fields. The task is seemingly trivial and involves first filtering out non-stationary measurements, and second clustering the remaining stationary points. Difficulties, however, emerge when frequently visited places are close to each other. In such cases the samples may overlap which makes spatial clustering algorithms likely to detect separate but overlapping locations as being one and the same. In this paper, we describe the *Infostop* algorithm for stop location detection, which is highly robust to overlap because it leverages the flow-based network community detection algorithm Infomap for labeling places. Infostop has a simple API and can both be used for labeling a time-ordered sequence of coordinates (GPS or otherwise), and a collection of points. Infostop is written in Python, uses vectorized distance computations and C++-based libraries for speed, and can be installed with pip.

## I. SUMMARY

Understanding human mobility is of great interest to scientists in multiple fields. While most mobility data is given as a time-series of GPS coordinates, e.g. collected from a smartphone, often the unit of greatest interest to researchers is the series of places that an individual visited—so called *places of interest* or *stop locations*. To detect stop locations a number of challenges must be overcome. First, the raw data is typically huge, requiring often some sort of downsampling pass over the collection to reduce the volume at minimal signal expense. Second, and most difficult, the remaining coordinates must be clustered. There exists a number of methods for doing this [1–6], but none deal with the fact that the coordinate clusters can be overlapping. For example, if an individual frequently stops at two separate locations that are near each other (e.g. adjacent buildings on a campus) location measurements from all visits should fall into two separate clusters; however, due to sampling noise they may overlap. Clustering algorithms that operate in euclidean space, such as DBSCAN [7] or other agglomerative clustering methods are not well suited for this problem as anything that overlaps tend to get merged.

To allow for detection of overlapping clusters, Infostop treats the coordinates as nodes in a network and constructs links based on a distance threshold. It then clusters, or finds *communities*, in this network using the Infomap algorithm [8]. Infomap is a flow based network clustering algorithm, which assigns separate labels to every pocket of nodes that constrain mobility of a random walker traversing the network. In this network representation, if two locations are close to each other, and their samples slightly overlap, the number links connecting within-location samples will vastly outnumber the number of links connecting between-location samples. Individual stop locations, although close, will therefore capture the random walker and earn a separate cluster label.

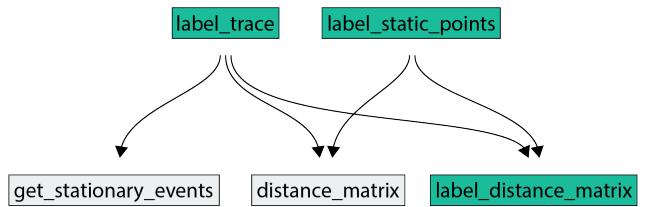


FIG. 1. Function dependency network.

## II. USAGE

Infostop has three main functions (see Fig. 1).

1. **label\_trace:** : **Given a time-ordered sequence of location measurements from one individual, assign stop location labels to each measurements.** (a) Using `get_stationary_events` detect which points are stationary and store only the median (lat, lon) of each stationarity event. A point belongs to a stationarity event if it is less than `r1` meters away from the median of the time-previous collection of stationary points. (b) Using `distance_matrix` compute the pairwise distances between all stationarity event medians. (c) Using `label_distance_matrix`, construct a network that links nodes (i.e. event medians) which are closer than `r2`. Assign a cluster label to each node using the Infomap algorithm, and map those labels back onto the input sequence. If input measurements are non-stationary, they are assigned the label `-1`. Optionally, also assign label `-1` to locations only visited once.
2. **label\_static\_points:** **Given a collection of locations points, assign cluster labels to each point.** It assumes that points correspond to stationary location measurements; therefore Infostop only needs to execute steps 1.b-1.c.
3. **label\_distance\_matrix:** **Given a distance matrix, assign cluster labels to each point.** Here, the user already knows the distance matrix of the points they want

\* [ulfaslak@gmail.com](mailto:ulfaslak@gmail.com)



FIG. 2. Example of locations detected with Infostop.

to cluster, and Infostop therefore only executes step 1.c.

The default distance metric is the haversine distance, thus expecting the input location measurements to be (lat, lon) coordinates. However, the user can easily specify another dis-

tance function if they like. As such, Infostop—and in particular its `label_static_points` function—can be used as a general clustering algorithm in any space that allows distance measurement between points.

The user may also provide timestamps with the location measurements as a third column in the input array to the `label_trace` function and specify the parameters `min_staying_time` and `max_time_between`, to control first how long the individual should remain stationary for the event to count as stationary, and second what the maximum time between two consecutive location samples may be for them to fall into the same stationary event. This is highly useful if the time-resolution of the location data is variable.

Figure 2 visualizes a sample of labeled locations from an anonymous user on a university campus. Only the medians of each stationary location events are shown. It is apparent that many of the clusters are either very near or even overlap to some degree.

- 
- [1] L. Gong, H. Sato, T. Yamamoto, T. Miwa, and T. Morikawa, Journal of Modern Transportation **23**, 202 (2015).
  - [2] X. Cao, G. Cong, and C. S. Jensen, Proc. VLDB Endow. **3**, 1009 (2010).
  - [3] R. Montoliu and D. Gatica-Perez, in *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia* (ACM, 2010) p. 12.
  - [4] T. M. T. Do and D. Gatica-Perez, IEEE Transactions on Mobile Computing **13**, 638 (2013).
  - [5] N. Wan and G. Lin, Computers, Environment and Urban Systems **39**, 63 (2013).
  - [6] F. Zhao, A. Ghorpade, F. C. Pereira, C. Zegras, and M. Ben-Akiva, Transportation research procedia **11**, 218 (2015).
  - [7] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, in *Kdd*, Vol. 96 (1996) pp. 226–231.
  - [8] M. Rosvall, D. Axelsson, and C. T. Bergstrom, The European Physical Journal Special Topics **178**, 13 (2009).

## References

- [1] Melanie Mitchell. *Complexity: A guided tour*. Oxford University Press, 2009.
- [2] Hiroki Sayama. *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks, 2015.
- [3] Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology*, 218(1):1–11, 2002.
- [4] Vishwesha Guttal, Paweł Romanczuk, Stephen J Simpson, Gregory A Sword, and Iain D Couzin. Cannibalism can drive the evolution of behavioural phase polyphenism in locusts. *Ecology letters*, 15(10):1158–1166, 2012.
- [5] Sepideh Bazazi, Paweł Romanczuk, Sian Thomas, Lutz Schimansky-Geier, Joseph J Hale, Gabriel A Miller, Gregory A Sword, Stephen J Simpson, and Iain D Couzin. Nutritional state and collective motion: from individuals to mass migration. *Proceedings of the Royal Society B: Biological Sciences*, 278(1704):356–363, 2010.
- [6] Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences*, 104(17):7301–7306, 2007.
- [7] Nancy B Grimm, Stanley H Faeth, Nancy E Golubiewski, Charles L Redman, Jianguo Wu, Xuemei Bai, and John M Briggs. Global change and the ecology of cities. *science*, 319(5864):756–760, 2008.
- [8] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96, 2012.
- [9] Christian Thiemann, Fabian Theis, Daniel Grady, Rafael Brune, and Dirk Brockmann. The structure of borders in a small world. *PloS one*, 5(11):e15422, 2010.
- [10] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [11] Enys Mones, Arkadiusz Stopczynski, Alex Sandy Pentland, Nathaniel Hupert, and Sune Lehmann. Optimizing targeted vaccination across cyber-physical networks: an empirically based mathematical simulation study. *Journal of The Royal Society Interface*, 15(138):20170783, 2018.

- [12] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.
- [13] Bjarke Mønsted, Piotr Sapiezyński, Emilio Ferrara, and Sune Lehmann. Evidence of complex contagion of information in social media: An experiment using twitter bots. *PloS one*, 12(9):e0184148, 2017.
- [14] Vedran Sekara, Arkadiusz Stopczynski, and Sune Lehmann. Fundamental structures of dynamic social networks. *Proceedings of the national academy of sciences*, 113(36):9977–9982, 2016.
- [15] Ulf Aslak, Martin Rosvall, and Sune Lehmann. Constrained information flows in temporal networks reveal intermittent communities. *Physical Review E*, 97(6):062312, 2018.
- [16] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through optimization: How facebook’s ad delivery can lead to skewed outcomes. *arXiv preprint arXiv:1904.02095*, 2019.
- [17] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. Measuring large-scale social networks with high resolution. *PloS one*, 9(4):e95978, 2014.
- [18] Nathan Eagle and Alex Sandy Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.
- [19] Juuso Karikoski and Matti Nelimarkka. Measuring social relations: case ota-sizzle. In *2010 IEEE Second International Conference on Social Computing*, pages 257–263. IEEE, 2010.
- [20] Christopher T Monk and Robert Arlinghaus. Eurasian perch, *perca fluviatilis*, spatial behaviour determines vulnerability independent of angler skill in a whole-lake reality mining experiment. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(3):417–428, 2017.
- [21] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462, 2006.
- [22] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779, 2008.
- [23] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818, 2010.

- [24] Pierre Deville, Chaoming Song, Nathan Eagle, Vincent D Blondel, Albert-László Barabási, and Dashun Wang. Scaling identity connects human mobility and social interactions. *Proceedings of the National Academy of Sciences*, 113(26):7047–7052, 2016.
- [25] Eric Sheppard and Robert B McMaster. *Scale and geographic inquiry: Nature, society, and method*. John Wiley & Sons, 2008.
- [26] Adam Moore. Rethinking scale as a geographical category: from analysis to practice. *Progress in human geography*, 32(2):203–225, 2008.
- [27] Nathan F Sayre. Ecological and geographical scale: parallels and potential for integration. *Progress in human geography*, 29(3):276–290, 2005.
- [28] Anssi Paasi. Place and region: looking through the prism of scale. *Progress in human geography*, 28(4):536–546, 2004.
- [29] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
- [30] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D<sup>3</sup> data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.
- [31] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90, 2007.
- [32] Martin Rosvall, Daniel Axelsson, and Carl T Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, 2009.