Learning to embed lifetime social behavior from interaction dynamics

Benjamin Wild[1,*], David M Dormagen[1], Michael L Smith[2,3], Tim Landgraf[1,*],

1 Department of Computer Science, Freie Universität, Berlin, Germany
2 Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA
3 Department of Collective Behaviour, Max Planck Institute of Animal Behavior, Konstanz, Germany

* b.w@fu-berlin.de, tim.landgraf@fu-berlin.de

Abstract

Interactions of individuals in complex social systems give rise to emergent behaviors at the group level. Identifying the functional role that individuals take in the group at a specific time facilitates understanding the dynamics of these emergent processes. An individual's behavior at a given time can be partially inferred by common factors, such as age, but internal and external factors also substantially influence behavior, making it difficult to disentangle common development from individuality. Here we show that such dependencies on common factors can be used as an implicit bias to learn a temporally consistent representation of a functional role from social interaction networks. Using a unique dataset containing lifetime trajectories of multiple generations of individually-marked honey bees in two colonies, we propose a new temporal matrix factorization model that jointly learns the average developmental path and structured variations of individuals in the social network over their entire lives. Our method yields inherently interpretable embeddings that are biologically relevant and consistent over time, allowing one to compare individuals' functional roles regardless of when or in which colony they lived. Our method provides a quantitative framework for understanding behavioral heterogeneity in complex social systems, and is applicable to fields such as behavioral biology, social sciences, neuroscience, and information science.

Author summary

Group-level emergent behaviors are the result of interactions between individual group members. To understand these social dynamics, one must objectively measure the function of an individual in their group at any given time. Ideally, one would also like to compare individuals from different groups, for example, to measure how specific environmental conditions or other external factors influence group behavior. Unfortunately, such an objective measure is hard to obtain because the group and its dynamics constantly change, making it challenging to define an individual's role in the group as a function of its actions and interactions. We propose a principled approach to model individuals in complex social systems by considering that function often depends, at least partially, on common factors such as age. The model learns a meaningful and interpretable descriptor for all individuals, and can be used to understand how complex social systems function and the emergence of group behavior.

1 Introduction

Animals living in groups often coordinate their behavior, resulting in emergent properties at the group level. The dynamics of the inter-individual interactions produce, for example, the coherent motion patterns of flocking birds and shoaling fish, or the results of democratic elections in human societies. In many social systems, individuals differ consistently in how, when, and with whom they interact. The way an individual participates in social interactions and therefore contributes to the emergence of group-level properties can be understood as its functional role within the collective [1–4].

Technological advances have made it possible to track all individuals and their interactions, ranging from social insects to primate groups [5–10]. These methods produce datasets that have unprecedented scale and complexity, but identifying and understanding the functional roles of the individuals within their groups has emerged as a new and challenging problem in itself. Social network analysis of interaction networks has proven to be a promising approach because interaction networks are comparatively straightforward to obtain from tracking data, and the networks represent each individual in the global context of the group [2, 3, 11, 12].

In most social systems, the way individuals interact changes over time, due to new experiences, environmental changes, or physiological conditions. Furthermore, groups themselves also tend to change, both in size and composition [13–18]. Despite these changes over time, an objective measure of the functional role should identify individuals that serve a similar function (e.g. a guard versus a forager). Unfortunately, we are now facing a recursive definition of function: We are trying to derive the function of an individual from the network, but the network itself is also a function of the individuals' behavior (and other factors). Still, consider a group-living species in which only a subset of individuals engage in nursing duties. If we analyze the networks of different groups of the same species in different environmental conditions and group sizes, we still expect an objective measure of function to be shared among individuals engaged in nursing, regardless of these confounding factors. How can we extract such an objective measure from a constantly changing network of interactions without a fixed frame of reference?

In many social systems, individuals share common factors that partially determine the roles they take. For example, an individual's age can have a strong influence on behavior. In humans, factors such as socioeconomic status are comparatively easy to measure yet determine behavior and, therefore, interactions to a large extent. If individuals take on roles partially determined by a common factor, can we use this dependency to learn an objective measure of function? Here, we show that such common factors are a powerful inductive bias to learn semantically consistent functional descriptors of individuals over time, even in highly dynamic social systems.

In recent years, methods that automatically learn semantic embeddings from high-dimensional data have become popular. These methods map entities into a learned vector space. For example, in natural language models, a word can be represented as a vector, such that specific regions in the manifold of learned embeddings correspond to words with similar meaning. Similarly, recommender systems can learn meaningful embeddings of users and items, for example, movies, such that similar entities cluster in the manifold of learned embeddings [19–22].

Such embeddings are usually learned from the data without additional supervision. In recommender systems, a movie's genre is usually not given in a dataset of user ratings, yet the genre can be identified given the learned embeddings [23]. This capability of learning embeddings from raw data and using them in downstream tasks is desirable in datasets of social interactions, where raw data is often abundant but labels are hard to acquire. Furthermore, embeddings are interpretable. For example, vector arithmetic of word embeddings can be used to understand how semantic concepts the natural language model has learned from the data relate to each other [24]. For entities that change

over time, trajectories of embeddings can be analyzed, i.e., how one entity changes within the learned manifold of embeddings. Such analyses can, for example, reveal how environmental conditions such as resource availability affect behavioral changes within the group [25, 26].

Most real-world networks have a hierarchical organization with overlapping communities, and thus soft community detection algorithms are often used to group and describe entities [26–28]. Non-negative matrix factorization (NMF) is a principled and scalable method to learn embeddings from data that can be represented in matrix form, such as interaction networks. NMF has an inherent soft clustering property and is therefore well suited to derive embeddings from social interaction networks [29]. If the embeddings allow us to predict relevant behavioral properties, they serve our understanding as *semantic* representations.

In symmetric non-negative matrix factorization (SymNMF), the dot products of any two individuals' embeddings (*factor vectors*) reconstruct their interaction affinity [30, 31], see Figure 1 a and b). However, this algorithm has no straightforward extension in temporal settings where the interaction matrices change over time. The interaction matrices at different time points can be factorized individually, but there is no guarantee that the embeddings stay semantically consistent over time. The dot product is permutation invariant, therefore factorization can result in different embeddings depending on the optimization method being used, or noise in the data. Consider the hypothetical case of two groups of animals of the same species with two tasks, guards and nurses. Factorizing the interaction matrices of both groups will likely reveal two clusters, but there is no guarantee that the same cluster will be assigned to the same task for both groups. The same problem can occur in the case of only one group with new animals emerging and some dying over time without any changes in the distribution of tasks on the group level. In this case, the embeddings are not semantically consistent over time. The prediction of relevant behavioral properties will deteriorate, and individuals cannot be meaningfully compared against each other.

Several approaches to extend NMF to temporal settings have been proposed in a variety of problem settings. Previous work proposed factorization methods for time series analysis [32, 33], while others focus on the analysis of communities that are determined by their temporal activity patterns [34]. Jiao and coworkers consider the case of communities from graphs over time and enforce temporal consistency with an additional loss term [35]. Several previous works represent network embeddings as a function of time [36] and [37], but the meaning of these embeddings can still shift over time. Temporal matrix factorization is similar to the tensor decomposition problem, which has many proposed solutions, see review by [38]. In particular, time-shifted tensor decomposition methods have been used in multi-neuronal spike train analysis, when recordings of multiple trials from a population of neurons are available [39, 40].

We approach this problem in the honey bee, a popular model system for studying individual and collective behavior [41]. Honey bees allocate tasks across thousands of individuals without central control, using an age-based system: young bees care for brood, middle-aged bees perform within-nest labor, and old bees forage outside [42, 43]. While age is a good predictor for the task of an average bee, individuals often deviate drastically from this common developmental trajectory due to internal and external factors. Honey bee colonies are also organized spatially: brood is reared in the center, honey and pollen are stored at the periphery, and foragers offload nectar near the exit. Therefore, an individual's role is partially reflected in its location, which provides the unique opportunity to evaluate whether learned embeddings based on the interaction data alone are meaningful.

A recent work proposes a method based on spectral decomposition to extract a semantic embedding (*Network age*) from honey bee interaction matrices and shows that
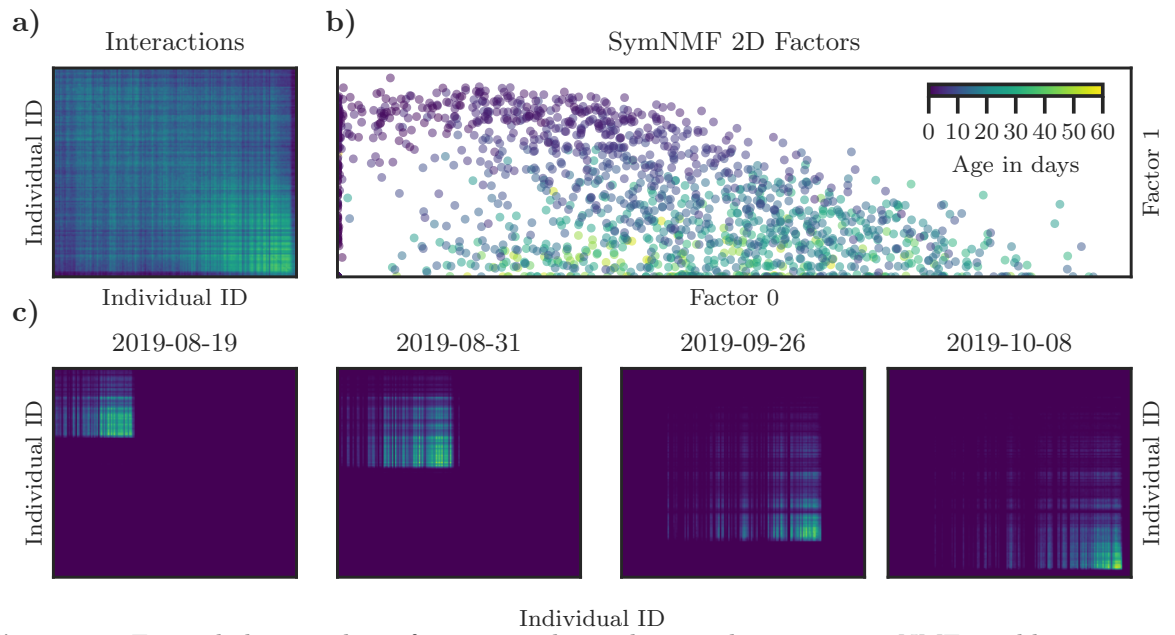
Figure 1. For a daily snapshot of a temporal social network, symmetric NMF is able to extract meaningful factor representations of the individuals. Colors represent the interaction frequencies of all individuals (**a**). The age-based division of labor in a honey bee colony is clearly reflected in the two factors - same-aged individuals are likely to interact with each other (**b**). For long observation windows spanning several weeks, the social network changes drastically as individuals are born, die, and switch tasks (**c**). Here, we investigate how a representation of temporal networks can be extracted, such that the factors representing individuals can be meaningfully compared over time, and even across datasets.

these embeddings can be used to predict task allocation, survival, activity patterns, and future behavior [12]. The method proposed here is conceptually similar but solves several remaining challenges. Here, we introduce Temporal NMF (TNMF), which yields consistent semantic embeddings even for individuals from disjoint datasets, for example, data from different colonies, or for long-duration recordings that contain multiple lifetime generations.

TNMF jointly learns a) a functional form of the average trajectory of embeddings along the common factor, b) a set of possible functional deviations from the average trajectory, and c) for each individual, a soft-clustering assignment (*individuality embedding*) to these deviations. We show that these representations can be learned in an unsupervised fashion, using only interaction matrices of the individuals over time. We analyze how well the model is able to disentangle common development from individuality using a synthetic dataset. Furthermore, we introduce a unique dataset containing lifetime trajectories of multiple generations of individually-marked honey bees in two colonies. We evaluate how well the embeddings learned by TNMF capture the semantic differences of individual honey bee development by evaluating their predictiveness for different tasks and behaviorally relevant metrics compared to several baseline models proposed in previous works.
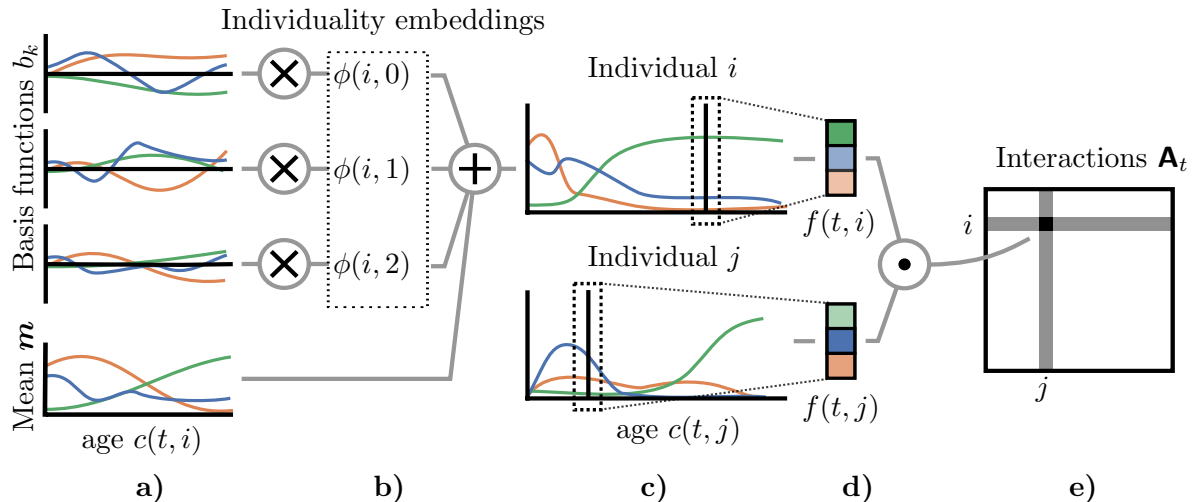
Figure 2. Overview of the method: We learn a parametric function describing the *mean life trajectory* $m(c(t,i))$ and a set of basis functions of individual variation $b(c(t,i))$, where $c(t,i)$ is the age of individual i at time t (**a**). For each individual, an embedding is learned consisting of one scalar per basis function that scales the contribution of the respective basis function - this vector of weights makes up the *individuality embedding* of an individual (**b**). The mean trajectory $m(c(t,i))$ plus a weighted sum of the basis functions $b(c(t,i))$ constitute the *lifetime trajectory* of each individual (**c**). At each time point, factors can be extracted from the individual lifetime trajectories (**d**) to reconstruct the interaction affinity between individuals (**e**). Note that the lifetime trajectories are functions of the individuals' ages, while interactions can occur at any time t.

2 Materials and Methods

2.1 Temporal NMF algorithm

SymNMF factorizes a matrix $A \in \mathbb{R}_+^{N \times N}$ such that it can be approximated by the product FF^T, where $F \in \mathbb{R}_+^{N \times M}$ and $M \ll N$:

$$\hat{F} = \underset{F \geq 0}{\operatorname{argmin}} \left\| A - FF^T \right\|^2 \qquad A_{i,j} \approx f(i) \cdot f(j)^T \qquad f(i) = F_{i,:} \quad f(i) \in \mathbb{R}_+^M \qquad (1)$$

When applied to social networks, $f(i)$ can represent the role of an entity within the social network A [30, 31] - however, in temporal settings, factorizing the matrices for different times separately will result in semantically inconsistent factors.

Here we present a novel temporal NMF algorithm (*TNMF*) which extends SymNMF to temporal settings in which $\mathbf{A} \in \mathbb{R}_+^{T \times N \times N}$ changes over time t. We assume that the entities $i \in \{0, 1, \ldots, N\}$ follow to some extent a common trajectory depending on an observable property (for example the age of an individual). We represent an entity at a specific point in time t using a factor vector $f^+(t,i)$ such that

$$\hat{\mathbf{A}}_{t,i,j} = f^+(t,i) \cdot f^+(t,j)^T \qquad \qquad \hat{\mathbf{A}} \in \mathbb{R}_+^{T \times N \times N} \quad f^+(t,i) \in \mathbb{R}_+^M \qquad (2)$$

In contrast to SymNMF, we do not directly factorize $\mathbf{A}_t$ to find the optimal factors that reconstruct the matrices. Instead, we decompose the problem into learning an average trajectory of factors $\boldsymbol{m}(c(t,i))$ and structured variations from this trajectory $\boldsymbol{o}(t,i)$ that depend on the observable property $c(t,i)$:

$$\boldsymbol{f}(t,i) = \boldsymbol{m}(c(t,i)) + \boldsymbol{o}(t,i) \qquad \boldsymbol{f}^+(t,i) = \max(0, \boldsymbol{f}(t,i)) \tag{3}$$

$$c : \mathbb{N}^{T \times N} \to \mathbb{N} \quad \boldsymbol{m} : \mathbb{N} \to \mathbb{R}_+^M \quad \boldsymbol{o} : \mathbb{N}^{T \times N} \to \mathbb{R}^M$$

128

This decomposition is an inductive bias that allows the model to learn semantically consistent factors for entities, even if they do not share any data points (e.g., there is no overlap in their interaction partners), as long as the relationship between functional role and $c(t,i)$ is stable. Note that in the simplest case $c(t,i) = t$, TNMF can be seen as a tensor decomposition model, i.e. the trajectory of all entities is aligned with the temporal dimension t of $\mathbf{A}$. In our case, $c(t,i)$ maps to the age of individual i at time t. 129 130 131 132 133 134

While many parameterizations for the function $\boldsymbol{o}(t,i)$ are possible, we only consider one particular case in this work: We learn a set of *individuality basis functions* $\boldsymbol{b}(c(t,i))$ (shared among all entities) that define a coordinate system of possible individual variations and the *individuality embeddings* ϕ, which capture to what extent each basis function applies to an entity:

$$\boldsymbol{o}(t,i) = \sum_{k=0}^{K} \phi_{i,k} \cdot b_k(c(t,i)) \qquad \phi : \mathbb{R}^{N \times K} \quad b_k : \mathbb{N}^T \to \mathbb{R} \tag{4}$$

where K is the number of learned basis functions. This parameterization allows us to disentangle the forms of individual variability (*individuality basis functions*) and the distribution of this variability (*individuality embeddings*) in the data. 135 136 137

We implement the functions $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$ with small fully connected neural networks with non-linearities and several hidden layers. The parameters θ of these functions and the entities' embeddings ϕ are learned jointly using minibatch stochastic gradient descent:

$$\hat{\theta}, \hat{\phi} = \operatorname*{argmin}_{\theta,\phi} \left\| \boldsymbol{A} - \hat{\boldsymbol{A}} \right\|^2 \tag{5}$$

Note that non-negativity is not strictly necessary, but we only consider the non-negative case in this work for consistency with prior work [30, 31]. Furthermore, instead of one common property with discrete time steps, the factors could depend on multiple continuous properties, i.e. $c : \mathbb{R}^{T \times N} \to \mathbb{R}^P$, e.g. the day and time in a intraday analysis of social networks. 138 139 140 141 142

We find that the model's interpretability can be improved using additional regularization terms without significantly affecting its performance. We encourage sparsity in both the number of used factors and individuality basis functions by adding L_1 penalties of the mean absolute magnitude of the factors $\boldsymbol{f}(t,i)$ and basis functions $\boldsymbol{b}(c(t,i))$ to the objective. We encourage individuals' lifetimes to be represented with a sparse embedding using an L_1 penalty of the learned *individuality embeddings* ϕ. 143 144 145 146 147 148

We also introduce an optional adversarial loss term to encourage the model to learn embeddings that are semantically consistent over time, i.e. to only represent two entities that were present in the dataset at different times with different embeddings if this is strictly necessary to factorize the matrices $\mathbf{A}$. We jointly train a discriminative network $d(\phi_i)$ that tries to classify the time of the first occurrence of all entities based on their *individuality embeddings* ϕ. The negative cross-entropy loss of this model is added as a regularization term to equation 5 in a training regime similar to generative adversarial networks [44]. Note that a high cross-entropy loss of the discriminative network $d(\phi_i)$ 149 150 151 152 153 154 155 156

implies that the distribution of *individuality embeddings* ϕ is consistent over time. See appendix S1.1 for more details and S2 for an ablation study of the effect the individual regularization terms have on the results of the model.

We implemented the model using PyTorch [45] and trained it in minibatches of 256 individuals for 200 000 iterations with the Adam optimizer [46]. We calculate the reconstruction loss $\left\|\mathbf{A}_t - \hat{\mathbf{A}}_t\right\|^2$ only for valid entries, i.e., we mask out all matrix elements where one of the individuals is not alive at the given time t. See appendix S1.3 for the architecture of the learned functions, a precise description of the regularization losses, and further hyperparameters. The code of our reference implementation is publicly available: `https://github.com/nebw/temporal_nmf`.

2.2 Data
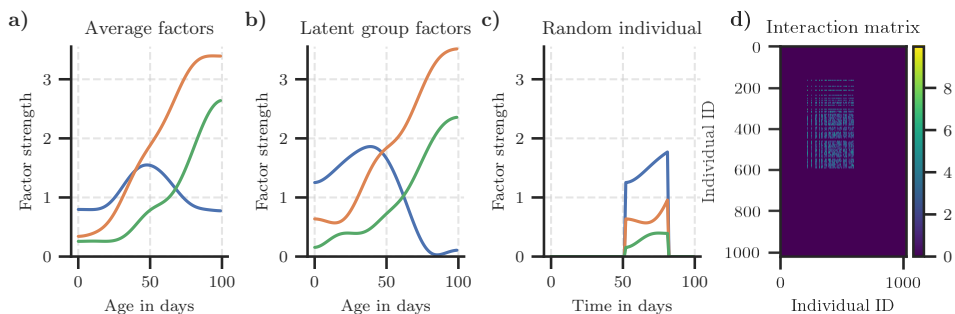
2.2.1 Synthetic data

Figure 3. Example of one synthetic dataset. **a)** Common lifetime trajectory of all entities. **b)** The lifetime trajectory of one latent group. **c)** The factors of one individual in the dataset of the latent group visualized in **b**. **d)** Generated interaction matrix for one day.

We created synthetic datasets using a generative model of interactions based on a common latent trajectory of factors and groups with structured variations from this trajectory. We compute the number of interactions between two individuals as the dot product of their latent factors and additive Gaussian noise. Using these datasets we can evaluate whether the model successfully converges and is able to correctly identify which individual belongs to which latent group, even in the presence on high amounts of observational noise. While we believe that such a latent structure exists in most complex social systems, it is not directly observable, and thus, for data from a real system, we can only evaluate the model on proxy measures (see section 2.3) that are observable.

We model a common lifetime trajectory of factors using a smoothed Gaussian random walk in $\mathbb{R}^+$ with $\sigma_{\text{walk}} = 1$ for the steps of the random walk and $\sigma_{\text{smoothing}} = 10$ for the Gaussian smoothing kernel. See Figure 3 a) for one example of a generated lifetime trajectory with three factors. We then randomly create latent groups by creating smoothed Gaussian random walks that define how these groups differ from the common lifetime trajectory. See Figure 3 b) for the lifetime trajectory of one latent group. For each group, we also define different expected mean lifetimes. We set the average lifetime of an entity to 30 days with a standard deviation of 10 days. We then randomly assign 1024 individuals to those latent groups and also assign random dates of emergence and disappearance of these individuals in the dataset. We then compute the individual factor trajectories for each individual, as can be seen in Figure 3 c). Finally, for 100 days of

simulated data, we generate interaction matrices by computing the dot products of the factors of all individuals (Figure 3 d).

We then measure how well the *individuality embeddings* ϕ of a fitted model match the true latent groups from the generative model using the adjusted mutual information score [47]. Furthermore, we measure the mean squared error between the ground truth factors and the best permutation of the factors $\boldsymbol{f}^+$. We evaluate the model on 128 different random synthetic datasets with increasing Gaussian noise levels in the interaction tensor.

2.2.2 Honey bee data

Honey bees are an ideal model system with a complex and highly dynamic social structure. The entire colony is observable most of the time. In recent years, technological advances have made it possible to automatically track individuals in entire colonies of honey bees over long periods of time [6,10,48]. We analyze a dataset obtained by tracking thousands of individually marked honey bees at high temporal and spatial resolution, covering entire lifespans and multiple generations.

Two colonies of honey bees were continuously recorded over a total of 155 days. Each individual was manually tagged at emergence, so the date of birth is known for each bee. Timestamps, positions, and unique identifiers of all (N=9286) individuals from these colonies were obtained using the BeesBook tracking system [10,12,48]. See Table 1 for dates and number of individuals. Temporal affinity matrices were derived from this data as follows: For each day, counts of proximity contact events were extracted. Two individuals were defined to be in proximity if their markers' positions had an euclidean distance of less than 2 cm for at least 0.9 seconds. The daily affinity between two individuals i and j based on their counts of proximity events $p_{t,i,j}$ at day t was then computed as: $\mathbf{A}_{t,i,j} = \log(1 + p_{t,i,j})$, $\mathbf{A} \in \mathbb{R}^{N_t \times N_i \times N_i}$, where N_t is the number of days and N_i the number of individuals in the dataset.

The datasets also contains labels that can be used in proxy tasks (see section 2.3) to quantify if the learned embeddings and factors are semantically meaningful and temporally consistent.

The datasets are open access and available under the *Creative Commons Attribution 4.0 International* license: `https://zenodo.org/record/3862966` [49].

Dataset	Dates	Days	Individuals	Interaction pairs
BN16	2016-07-23 to 2016-09-17	56	2443	43 174 748
BN19	2019-07-25 to 2019-11-01	99	6843	167 366 381

Table 1. The honey bee datasets contain the number of proximity-inferred interactions extracted from tracking data of all individuals in two long-term recordings spanning a total of 155 days and 9286 individuals.

In both datasets, we define $c(t, i)$ as the age in days of an individual i at time t.

2.3 Evaluation

Reconstruction: We measure how well the original interaction matrices $\mathbf{A}$ can be reconstructed from the factors. We do not require the model to reconstruct the interaction matrices as well as possible because we only use the reconstruction as a proxy objective to learn a meaningful representation. Still, a high reconstruction loss could indicate problems with the model, such as excessive regularization.
Consistency: We measure to what extent the *individuality embeddings* ϕ change over time. For each model, we train a multinomial logistic regression model to predict the

source cohort (date of birth) and calculate the area under the ROC curve ($\text{AUC}_{\text{cohort}}$) using a stratified 100-fold cross-validation with scikit-learn [50]. The baseline models do not learn an individuality embedding; therefore we compute how well the model can predict the cohort using the mean factor representation of the individuals over their lives. We define consistency as $1 - \text{AUC}_{\text{cohort}}$ of this linear model. Note that a very low temporal consistency would indicate that the development of individual bees changes strongly between cohorts and colonies, which we know not to be true.

Mortality and Rhythmicity: We evaluate how well a linear regression model can predict the mortality (number of days until death) and circadian rhythmicity of the movement [12] (R^2 score of a sine with a period of 24 h fitted to the velocity over a three-day window). These metrics are strongly correlated with an individual's behavior (e.g. foragers exhibit strong circadian rhythms because they can only forage during the daytime; foragers also have a high mortality). We follow the procedure given in [12] and report the 100-fold cross-validated R^2 scores for these regression tasks.

Time spent on different nest substrates: For a subset of the data, from 2016-08-01 to 2016-08-25, nest substrate usage information is also available. This data contains the proportion of time each individual spends in the brood area, honey storage, and on the dance floor. This data was previously published and analyzed [12,51]. The task of an honey bee worker is strongly associated with her spatial distribution in the hive. We therefore expect a good representation of the individuals' functional role to correlate with this distribution.

For this data, we expect the factors $\boldsymbol{f}^+$ and *individuality embeddings* ϕ to be semantically meaningful and temporally consistent if they reflect an individual's behavioral metrics (mortality and rhythmicity) and if they do not change strongly over time (measured in the consistency metric).

2.4   Baseline models

Biological Age: Task allocation in honey bee is partially determined by temporal polyethism. Certain tasks are usually carried out by individuals of about the same age, e.g. young bees are usually occupied with nursing tasks. We therefore use the age of an individual as a baseline descriptor.

Symmetric NMF: We compute the factors that optimally reconstruct the original interaction matrices using the standard symmetric NMF algorithm [31,52], for each day separately, using the same number of factors as in the TNMF model.

Optimal permutation SymNMF: We consider a simple extension of the standard SymNMF algorithm that aligns the factors to be more consistent over time. For each pair of subsequent days, we consider all combinatorial reorderings of the factors computed for the second day. For each reordering, we compute the mean L_2 distance of all individuals that were alive on both days. We then select the reordering that minimizes those pairwise L_2 distances and greedily continue with the next pair of days until all factors are aligned. Furthermore, we align the factors across colonies (where individuals cannot overlap) as follows: we run this algorithm for both datasets separately and align the resulting factors by first computing the mean embedding for all individuals grouped by their ages. As before, we now select from all combinatorial possibilities the reordering that minimizes the L_2 distance between the embeddings obtained from both datasets. See section S3.1 for pseudo code.

Tensor decomposition: We also compare against a constrained non-negative tensor decomposition model with symmetric factors $\boldsymbol{F} \in \mathbb{R}_+^{N \times M}$ and temporal dynamics
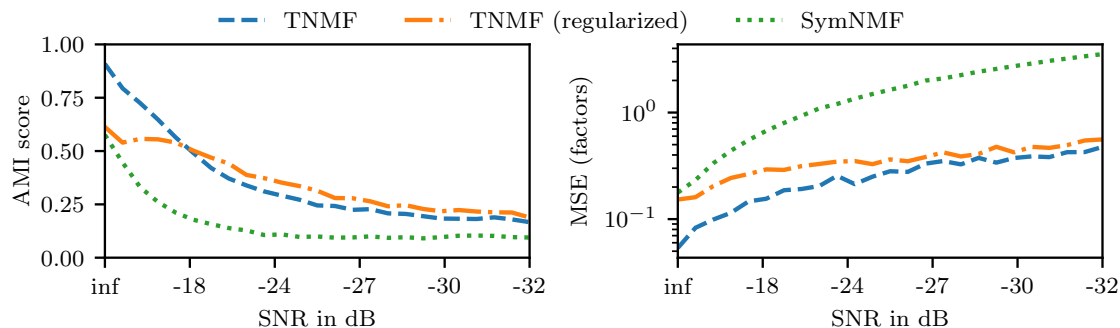
Figure 4. AMI score and mean squared error between true factors and the best permutation of learned factors for increasing noise levels. The median values over 128 trial runs are shown.

constrained to the diagonals, i.e. $\mathbf{D} \in \mathbb{R}_+^{T \times M \times M}$ and $\mathbf{D}_t = \mathrm{diag}(\boldsymbol{d}_t)$, $\boldsymbol{d}_t \in \mathbb{R}_+^M$.

$$\hat{\mathbf{A}}_t = \boldsymbol{F}\mathbf{D}_t\boldsymbol{F}^T \tag{6}$$

$$\hat{\boldsymbol{F}}, \hat{\mathbf{D}} = \operatorname*{argmin}_{\boldsymbol{F},\mathbf{D}} T^{-1} \sum_{t=0}^{T} \left\| \boldsymbol{A}_t - \hat{\boldsymbol{A}}_t \right\|^2 \tag{7}$$

Temporal NMF models: We evaluate variants of the temporal symmetric matrix factorization algorithms proposed by [35] and [36].

For the tensor decomposition and temporal NMF baselines, we follow the procedure given above for the *Optimal permutation SymNMF* to find the optimal reordering to align the factors obtained by applying models to the two datasets separately.

3 Results

3.1 Synthetic data

We factorize the interaction matrices of the 128 synthetic datasets with varying levels of Gaussian noise. We confirmed that our model converges in all datasets and evaluate whether we can distinguish the individuals' ground truth group assignments. To that end, we extract the *individuality embeddings* ϕ from the models and measure how well they correspond to ground truth data using the adjusted mutual information (AMI) score. Furthermore, we measure the mean squared error between the best permutation of learned factors $\boldsymbol{f}^+$ and the ground truth factors.

We find that for low levels of noise, our model can identify the truth group assignments with high accuracy, and are still significantly better than random assignments even at very high levels of noise (see figure 4). Note that for this experiment, we evaluated a model with the same hyperparameters as used in all plots in the results section (see Table 2) and a variant without explicit regularization except the L_1 penalty of the learned *individuality embeddings* ϕ ($\lambda_{\mathrm{embeddings}}$, because this regularization is required to meaningfully extract clusters), which was set to 0.1. See appendix 2.2.1 for more details on the synthetic datasets.

3.2 Honey bees

Mean lifetime model: The model learns a sparse representation of the developmental trajectory of a honey bee in the space of social interactions. Only two factors are
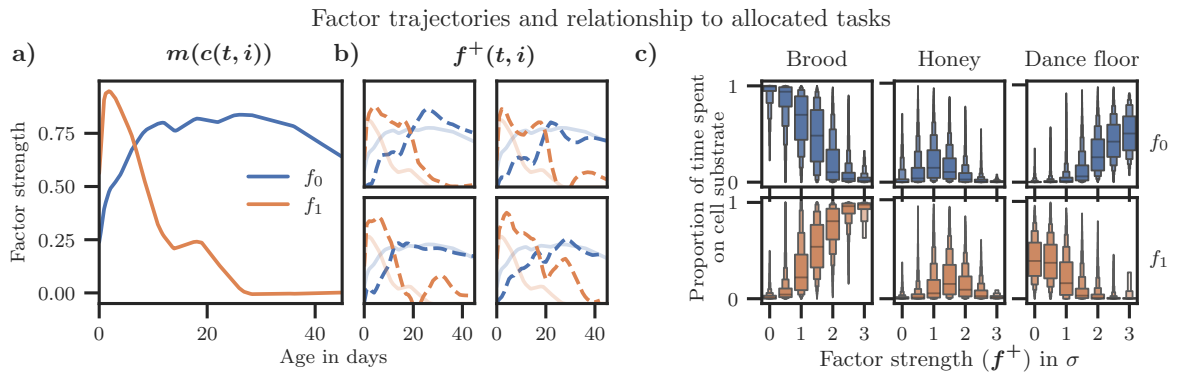
Factor trajectories and relationship to allocated tasks

Figure 5. a) Mean lifetime trajectories according to $\boldsymbol{m}(c(t,i))$. The model learns a sparse representation of the functional position of the individuals in the social network. f_0 (blue) mostly corresponds to middle-aged and older bees, and f_1 (orange) predominantly describes young bees. Only factors with a mean magnitude of at least 0.01 are shown. **b)** Even though the model uses only these two factors, it is still expressive enough to capture individual variability, as can be seen in randomly sampled individuals' lifetime trajectories. **c)** The individual factors $\boldsymbol{f}^+$ and the proportion of time the individuals spent on different nest substrates. The strong correlation indicates that the learned factors are a good representation of the individuals' roles in the colonies. Note that the factors have been divided by their standard deviation here for ease of comparability.

effectively used (they exceed the threshold value of 0.01). These factors show a clear trend over the life of a bee, indicating that the model captures the temporal aspects of the honey bee division of labor (See Figure 5).

Interpretability of factors: To understand the relationship between the factors and division of labor, we calculate how the factors map to the fraction of time an individual spent on the brood area, honey storage, or dance floor (where foragers aggregate). Time spent on these different substrates is a strong indicator of an individual's task. The factor f_1, which peaks at young age (Figure 5), correlates with the proportion of time spent in the brood area, while a high f_0 indicates increased time spent on the dance floor. Therefore, the model learned to map biologically relevant processes.

Individuality basis functions and individuality embeddings: Due to the regularization of the embeddings, the model learns a sparse set of *individuality basis functions*. As encouraged by the model, most individuals can predominantly be described by a single basis function. That means that while each honey bee can collect a unique set of experiences, most can be described with a few common *individuality embeddings* which are consistent across cohorts and colonies. In the context of honey bee division of labor, the basis functions are interpretable because the factors correspond to different task groups. For example, $b_{12}(c(t,i))$ (accounting for $\approx 10.7\%$ of the individuals) describes workers that occupy nursing tasks much longer than most bees. As the *individuality embeddings* ϕ only scale the magnitude of the basis functions, they can be interpreted in the same way. *Individual lifetime trajectories* in the factor space can be computed based on the *mean lifetime trajectories* ($\boldsymbol{m}$), *individuality basis functions* ($\boldsymbol{b}(c(t,i))$) and *individuality embeddings* (ϕ). See figure 6 for examples of *individual lifetime trajectories* from workers that most strongly corresponded to the common *individuality basis functions*.

Evaluation: We verify that the learned representations of the individuals are meaningful (i.e., they relate to other properties of the individuals, not just their interaction matrices) and semantically consistent over time and across datasets using the metrics described in the section *Evaluation*. We compare variants of our model with different adversarial loss scaling factors and factor L_1 regularizations, the baseline models, and
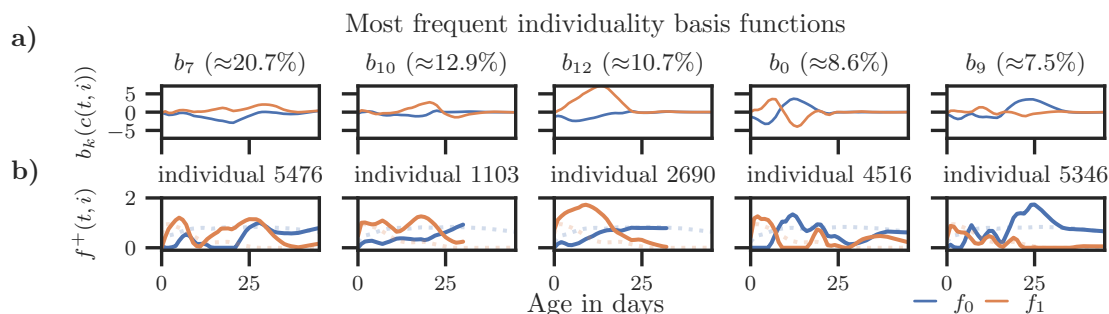
a)

Most frequent individuality basis functions

Figure 6. a) Magnitude of factor offsets for the five most common individuality basis functions over age $b_k(c(t,i))$. The percentage of individuals that most strongly correspond to the individual basis functions is shown in the column titles. More than 60% of the individuals strongly correspond to one of the five basis functions shown here. **b)** Because the basis functions describe *individuality offsets* from the mean lifetime trajectory, it may be easier to interpret them by visualizing individual examples. For each of the basis functions (top row), we show a lifetime trajectory of an individual that corresponds to that basis function (bottom row). Note that individuals can die or disappear at any time (solid lines). The mean lifetime trajectories are shown as dotted lines in the background.

the individuals' ages. We expect a good model to be temporally consistent and semantically meaningful. All variants of our model outperform the baselines in terms of the semantic metrics *Mortality* and *Rhythmicity*, except for the [36] model, which performs comparably well in the *Mortality* metric. The adversarial loss term further increases the *Consistency* metric without negatively affecting the other metrics. A very strong adversarial regularization (see row with $\lambda_{\mathrm{adv}} = 1$ in Table 2) prevents the model from learning a good representation of the data. See Table 2 for an overview of the results. We also evaluate the tradeoff between the different metrics using a grid search over the hyperparameters (see appendix 3.2).

Scalability: The functions $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$ are learned neural networks with non-linearities. The objective is non-convex and we learn the model parameters using stochastic gradient descent. Optimization is therefore slower than the standard NMF algorithms that can be fitted using algorithms such as Alternating Least Squares [53]. We found that the model converges faster if the reconstruction loss of the age based model $\boldsymbol{m}(c(t,i))$ is additionally minimized with the main objective in equation 5. Due to the minibatch training regime, our method should scale well in larger datasets. Small neural networks were sufficient to learn the functions $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$ in our experiments. Most of the runtime during training is spent on the matrix multiplication $f^+(t,i) \cdot f^+(t,j)^T$ and the corresponding backwards pass.

Tradeoff between temporal consistency and semantic meaningfulness: We performed a grid search over the hyperparameters λ_{f}, λ_{adv}, λ_{basis}, and $\lambda_{\mathrm{embeddings}}$ (see Table 1) to evaluate whether models can only be either semantically meaningful or temporally consistent. For this analysis, we define *Semantic meaningfulness* as the sum of the *Rhythmicity* and *Mortality* metrics introduced in section 2.3. We find that models that are very temporally consistent fail to learn semantically meaningful information. Interestingly, the models with the best tradeoff between the two metrics are almost as semantically meaningful as those models with low temporal consistency and the highest semantic meaningfulness. This analysis suggests that regularization encourages the model to only represent different individuals differently if this is strictly necessary to factorize the data. See Figure 5.

	Model				
Method	Variant	$\left\lVert \boldsymbol{A} - \hat{\boldsymbol{A}} \right\rVert^2$ ↓	Consistency ↑	Mortality ↑	Rhythmicity ↑
Age	-	-	-	0.02	0.20
SymNMF	Vanilla	0.9	0.18	0.01	0.02
SymNMF	Optimal permutation	0.9	0.12	0.09	0.35
Tensor decomposition	-	1.36	0.03	0.06	0.09
DNMF [35]	$\gamma = 0.1$	0.9	0.19	0.02	0.05
DNMF [35]	$\gamma = 1$	1.15	0.15	0.01	0.04
s-TMF [36]	$\beta = 0.01, d = 5$	1.59	0.03	0.17	0.06
TNMF	No regularization	1.21	0.17	0.30	0.48
TNMF	$\lambda_{\mathrm{adv}} = 0, \lambda_{\mathrm{f}} = 0.01$	1.26	0.18	0.10	0.40
TNMF	$\lambda_{\mathrm{adv}} = 0.1, \lambda_{\mathrm{f}} = 0.01$	1.28	0.35	0.20	0.42
TNMF	$\lambda_{\mathrm{adv}} = 1, \lambda_{\mathrm{f}} = 0.01$	1.88	0.5	0.03	0.25
TNMF	$\lambda_{\mathrm{adv}} = 0, \lambda_{\mathrm{f}} = 0.1$	1.31	0.19	0.09	0.38
<u>**TNMF**</u>	$\lambda_{\mathrm{adv}} = 0.1, \lambda_{\mathrm{f}} = 0.1$	1.33	0.37	0.10	0.42

Table 2. The evaluation metrics for TNMF and the baseline models described in section 2.3. See appendix S1.3 and S3 for descriptions of the hyperparameters used. Note that the SymNMF model reconstruction loss can be seen as a lower bound for the matrix factorization models considered here, and imposing a temporal structure or regularization causes all models to explain less variance in the data. However, for all models except TNMF this does not result in a significant increase of the other metrics. The underlined model is used in all plots in the results section.

4 Discussion

Temporal NMF factorizes temporal matrices with overlapping and even disjoint communities by learning an embedding of individuals as a function of a common factor, such as age, and a learned representation of the individuals' individuality. This explicit dependency on a common factor that partially determines the function of an individual constitutes an inductive bias. We show that the model learns semantically consistent representations of individuals, even in challenging cases, such as the datasets analyzed in this work.

The individual components of the model are straightforward to visualize and interpret. The learned individuality embeddings ϕ can be understood as soft-cluster assignments relating to the whole lifetime of an individual, while the factor vectors $\boldsymbol{f}^+(t, i)$ can be interpreted as cluster assignments of the individuals at a specific point in time, i.e. two individuals with similar factor vectors are likely to interact if they exist in the same group at the same time. Furthermore, the model encourages sparsity, making the results easier to interpret because the model only uses as many factors and clusters as necessary.

We identified a crucial trade-off that comes with temporal consistency: For a specific point in time, the ability to predict behaviorally relevant attributes will likely be worse for a model that learns temporally consistent representations compared to a non-consistent model with the same capacity. Conversely, in more challenging cases, e.g. when taking long periods of time or data from disjoint communities into consideration, temporally consistency is indispensable for a good representation. Furthermore, we found that models can be temporally consistent, semantically meaningful, or both; selecting the correct model requires an inductive bias, but regularization of the model also influences the results.

Previous works have demonstrated that biologically relevant findings can be obtained using network analysis of social interaction networks [4, 5, 26, 54–62]. A recent method, *Network Age* [12], proposes using spectral decomposition of honey bee interaction networks into succinct descriptors of the individual's social network that can be used
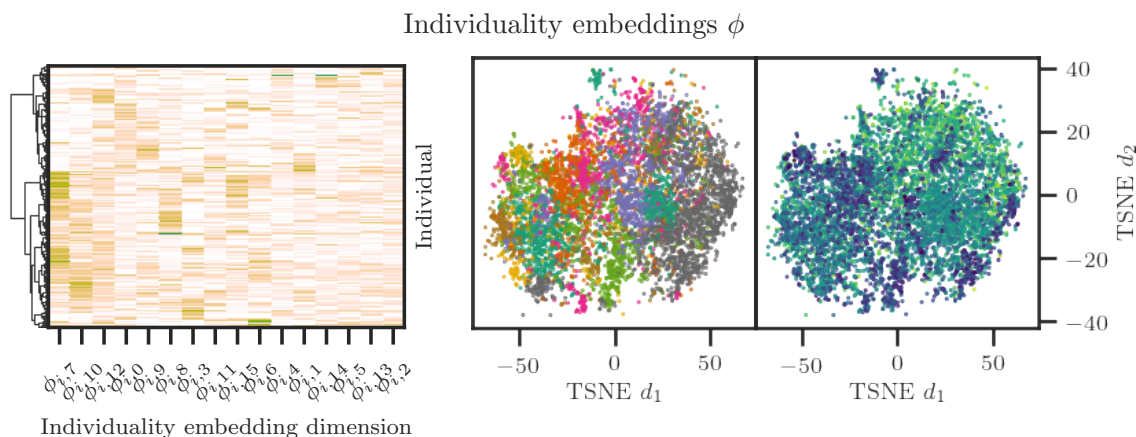
Individuality embeddings ϕ

Individuality embedding dimension

Figure 7. Left: Hierarchical clustering of individuality embeddings: Most individuals strongly correspond to a single individuality basis function, making it easy to cluster their lifetime social behavior (i.e. each individual has a high value in a single dimension for their individuality embedding). Because each cluster is strongly associated with a specific individuality basis function, and because each basis function is interpretable (Figure 5), these blueprints of lifetime development can also be intuitively understood and compared. **Right:** TSNE plots of the individuality embeddings colored by cluster (left) and the maximum circadian rhythmicity of an individual during her lifetime (right), indicating that the embeddings are semantically meaningful.

to predict task allocation, survival, activity patterns, and future behavior. Symmetric nonnegative matrix factorization and Laplacian-based spectral clustering have been shown to be equivalent [29]. Thus, TNMF can be understood as a further development of *Network Age*. TNMF learns representations of individuals based on their social interaction network that can facilitate the analysis of developmental trajectories, division of labor, and individual variance in behavior. Furthermore, TNMF provides temporally consistent embeddings and with that rectifies a remaining limitation of *Network Age*. We confirmed that, on the honey bee dataset, TNMF obtains biologically meaningful lifetime trajectories with promising prospects for experimental application. TNMF may help advance our understanding of the colony function and the interplay between environmental factors and individual and collective responses. The method presented here offers a way to investigate the impact of stress factors, such as pesticides, parasitic mites, and agricultural monoculture, on the social structure of colonies and therefore may present an avenue to improve honeybee health.

We applied our method to the honey bee model, which has numerous individuals, with an entangled and highly dynamic social structure. Our method, however, can be applied to any setting in which matrix factorization is commonly used, such as recommender systems, network analysis, audio processing, bioinformatics, etc. Interaction matrices in networked systems have a broad class of use-cases. In any system with dynamically interacting units, our model reduces high-dimensional interaction patterns to low-dimensional embeddings. The only requirement is that the interactions follow our generic model of an average path from which individual units can deviate. The method could serve as a means to generate hypotheses: Clustering individuals in the embedding space may reveal functional groups, and the basis functions can indicate relevant time points in individual developments that can be investigated in follow-up studies. Note that time in our model may be replaced by any variable along which one wants to study the matrix dynamics. While we evaluate TNMF on honey bees, the method may be used to study human social networks and their underlying dynamics.

A deeper understanding of human interaction dynamics may benefit aspects of human life, such as health, technology, and work. The method could, for example, be used to identify individuals with a higher risk of contracting or transmitting a disease, or help assess the effect of pandemics and potential interventions. 415 416 417 418

By publishing the honey bee dataset and our reference implementation of the TNMF algorithm, we hope to encourage the scientific community to build upon our efforts. 419 420

Acknowledgments 421

References

1. Gordon DM. Ant encounters: interaction networks and colony behavior. Princeton University Press; 2010.

2. Pinter-Wollman N, Hobson EA, Smith JE, Edelman AJ, Shizuka D, de Silva S, et al. The dynamics of animal social networks: analytical, conceptual, and theoretical advances. Behavioral Ecology. 2014;25(2):242–255. doi:10.1093/beheco/art047.

3. Krause J, James R, Franks DW, Croft DP. Animal Social Networks. Oxford University Press; 2015.

4. Farine DR, Whitehead H. Constructing, conducting and interpreting animal social network analysis. Journal of Animal Ecology. 2015;84(5):1144–1163.

5. Mersch DP, Crespi A, Keller L. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. Science. 2013;340(6136):1090–1093.

6. Gernat T, Rao VD, Middendorf M, Dankowicz H, Goldenfeld N, Robinson GE. Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks. Proceedings of the National Academy of Sciences. 2018;115(7):1433–1438. doi:10.1073/pnas.1713568115.

7. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience. 2018;21(9):1281–1289. doi:10.1038/s41593-018-0209-y.

8. Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, et al. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. eLife. 2019;8:e47994. doi:10.7554/eLife.47994.

9. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SSH, Murthy M, et al. Fast animal pose estimation using deep neural networks. Nature Methods. 2019;16(1):117–125. doi:10.1038/s41592-018-0234-5.

10. Boenisch F, Rosemann B, Wild B, Dormagen D, Wario F, Landgraf T. Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence. Frontiers in Robotics and AI. 2018;5. doi:10.3389/frobt.2018.00035.

11. Brask JB, Ellis S, Croft DP. Animal social networks – an introduction for complex systems scientists. arXiv:200509598 [physics, q-bio]. 2020;.

12. Wild B, Dormagen DM, Zachariae A, Smith ML, Traynor KS, Brockmann D, et al. Social networks predict the life and death of honey bees. Nature Communications. 2021;12(1):1110. doi:10.1038/s41467-021-21212-5.

13. Gordon DM, Mehdiabadi NJ. Encounter rate and task allocation in harvester ants. Behavioral Ecology and Sociobiology. 1999;45(5):370–377. doi:10.1007/s002650050573.

14. Naug D. Structure of the social network and its influence on transmission dynamics in a honeybee colony. Behavioral Ecology and Sociobiology. 2008;62(11):1719–1725. doi:10.1007/s00265-008-0600-x.

15. Sendova-Franks AB, Hayward RK, Wulf B, Klimek T, James R, Planqué R, et al. Emergency networking: famine relief in ant colonies. Animal Behaviour. 2010;79(2):473–485. doi:10.1016/j.anbehav.2009.11.035.

16. Ament SA, Wang Y, Robinson GE. Nutritional regulation of division of labor in honey bees: toward a systems biology perspective. WIREs Systems Biology and Medicine. 2010;2(5):566–576. doi:10.1002/wsbm.73.

17. Aplin LM, Farine DR, Morand-Ferron J, Cockburn A, Thornton A, Sheldon BC. Experimentally induced innovations lead to persistent culture via conformity in wild birds. Nature. 2015;518(7540):538–541. doi:10.1038/nature13998.

18. Davidson JD, Gordon DM. Spatial organization and interactions of harvester ants during foraging activity. Journal of The Royal Society Interface. 2017;14(135):20170413. doi:10.1098/rsif.2017.0413.

19. Frome A, Corrado GS, Shlens J, Bengio S, Dean J, Ranzato MA, et al. DeViSE: A Deep Visual-Semantic Embedding Model. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in Neural Information Processing Systems 26. Curran Associates, Inc.; 2013. p. 2121–2129.

20. Asgari E, Mofrad MRK. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. PLOS ONE. 2015;10(11):e0141287. doi:10.1371/journal.pone.0141287.

21. Camacho-Collados J, Pilehvar MT. From word to sense embeddings: a survey on vector representations of meaning. Journal of Artificial Intelligence Research. 2018;63(1):743–788. doi:10.1613/jair.1.11259.

22. Nelson W, Zitnik M, Wang B, Leskovec J, Goldenberg A, Sharan R. To Embed or Not: Network Embedding as a Paradigm in Computational Biology. Frontiers in Genetics. 2019;10. doi:10.3389/fgene.2019.00381.

23. Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems. Computer. 2009;42(8):30–37. doi:10.1109/MC.2009.263.

24. Mikolov T, Yih Wt, Zweig G. Linguistic Regularities in Continuous Space Word Representations. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Atlanta, Georgia: Association for Computational Linguistics; 2013. p. 746–751. Available from: https://aclanthology.org/N13-1090.

25. Smith ML, Davidson JD, Wild B, Dormagen DM, Landgraf T, Couzin ID. The dominant axes of lifetime behavioral variation in honey bees; 2021.

26. Richardson TO, Kay T, Braunschweig R, Journeau OA, Rüegg M, McGregor S, et al. Ant behavioral maturation is mediated by a stochastic transition between two fundamental states. Current Biology. 2021;doi:10.1016/j.cub.2020.05.038.

27. Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. Nature. 2005;435(7043):814–818. doi:10.1038/nature03607.

28. Ahn YY, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. Nature. 2010;466(7307):761–764. doi:10.1038/nature09182.

29. Ding C, He X, Simon HD. On the equivalence of nonnegative matrix factorization and spectral clustering. In: in SIAM International Conference on Data Mining; 2005.

30. Wang F, Li T, Wang X, Zhu S, Ding C. Community discovery using nonnegative matrix factorization. Data Mining and Knowledge Discovery. 2011;22(3):493–521. doi:10.1007/s10618-010-0181-y.

31. Shi X, Lu H, He Y, He S. Community detection in social network with pairwisely constrained symmetric non-negative matrix factorization. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); 2015. p. 541–546.

32. Yu HF, Rao N, Dhillon IS. Temporal Regularized Matrix Factorization for High-dimensional Time Series Prediction. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems 29. Curran Associates, Inc.; 2016. p. 847–855.

33. Mackevicius EL, Bahle AH, Williams AH, Gu S, Denisenko NI, Goldman MS, et al. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. eLife. 2019;8:e38471. doi:10.7554/eLife.38471.

34. Gauvin L, Panisson A, Cattuto C. Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach. PLoS ONE. 2014;9(1). doi:10.1371/journal.pone.0086028.

35. Jiao P, Lyu H, Li X, Yu W, Wang W. Temporal community detection based on symmetric nonnegative matrix factorization. International Journal of Modern Physics B. 2017;doi:10.1142/S0217979217501028.

36. Yu W, Aggarwal CC, Wang W. Temporally Factorized Network Modeling for Evolutionary Network Analysis. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. WSDM '17. New York, NY, USA: Association for Computing Machinery; 2017. p. 455–464.

37. Wang W, Yu W, Cheng W, Aggarwal CC, Chen H. Link Prediction with Spatial and Temporal Consistency in Dynamic Networks. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017; p. 3343–3349.

38. Kolda T, Bader B. Tensor Decompositions and Applications. SIAM Rev. 2009;doi:10.1137/07070111X.

39. Mørup M, Hansen LK, Arnfred SM, Lim LH, Madsen KH. Shift-invariant multi-linear decomposition of neuroimaging data. NeuroImage. 2008;42(4):1439–1450. doi:10.1016/j.neuroimage.2008.05.062.

40. Williams AH. Combining tensor decomposition and time warping models for multi-neuronal spike train analysis. bioRxiv. 2020; p. 2020.03.02.974014. doi:10.1101/2020.03.02.974014.

41. Elekonich MM, Roberts SP. Honey bees as a model for understanding mechanisms of life history transitions. Comparative Biochemistry and Physiology Part A, Molecular & Integrative Physiology. 2005;141(4):362–371. doi:10.1016/j.cbpb.2005.04.014.

42. Seeley TD. Adaptive significance of the age polyethism schedule in honeybee colonies. Behavioral Ecology and Sociobiology. 1982;11(4):287–293. doi:10.1007/BF00299306.

43. Johnson BR. Division of labor in honeybees: form, function, and proximate mechanisms. Behavioral Ecology and Sociobiology. 2010;64(3):305–316. doi:10.1007/s00265-009-0874-7.

44. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. arXiv:14062661 [cs, stat]. 2014;.

45. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc Fd, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019. p. 8026–8037.

46. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: Bengio Y, LeCun Y, editors. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings; 2015.

47. Vinh NX, Epps J, Bailey J. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In: Proceedings of the 26th Annual International Conference on Machine Learning. ICML '09. New York, NY, USA: Association for Computing Machinery; 2009. p. 1073–1080.

48. Wario F, Wild B, Couvillon MJ, Rojas R, Landgraf T. Automatic methods for long-term tracking and the detection and decoding of communication dances in honeybees. Frontiers in Ecology and Evolution. 2015;3. doi:10.3389/fevo.2015.00103.

49. Wild B, Dormagen DM, Smith ML, Landgraf T. Learning to embed lifetime social behavior from interaction dynamics - Data; 2020. Available from: https://zenodo.org/record/3862966.

50. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12(Oct):2825–2830.

51. Wild B, Dormagen D, Landgraf T. Social networks predict the life and death of honey bees - Data; 2021. Available from: https://zenodo.org/record/4438013.

52. Kuang D, Yun S, Park H. SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering. Journal of Global Optimization. 2015;62(3):545–574. doi:10.1007/s10898-014-0247-2.

53. Kim J, He Y, Park H. Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. Journal of Global Optimization. 2014;58(2):285–319. doi:10.1007/s10898-013-0035-4.

54. Naug D, Camazine S. The role of colony organization on pathogen transmission in social insects. Journal of Theoretical Biology. 2002;215(4):427–439. doi:10.1006/jtbi.2001.2524.

55. Lusseau D, Newman MEJ. Identifying the Role That Animals Play in Their Social Networks. Proceedings: Biological Sciences. 2004;271:S477–S481.

56. Sueur C, Petit O. Organization of Group Members at Departure Is Driven by Social Structure in Macaca. International Journal of Primatology. 2008;29(4):1085–1098. doi:10.1007/s10764-008-9262-9.

57. Claidière N, Messer EJE, Hoppitt W, Whiten A. Diffusion Dynamics of Socially Learned Foraging Techniques in Squirrel Monkeys. Current Biology. 2013;23(13):1251–1255. doi:10.1016/j.cub.2013.05.036.

58. Rosenthal SB, Twomey CR, Hartnett AT, Wu HS, Couzin ID. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. Proceedings of the National Academy of Sciences. 2015;112(15):4690–4695. doi:10.1073/pnas.1420068112.

59. Cheney DL, Silk JB, Seyfarth RM. Network connections, dyadic bonds and fitness in wild female baboons. Royal Society Open Science. 2016;3(7):160255. doi:10.1098/rsos.160255.

60. Strandburg-Peshkin A, Papageorgiou D, Crofoot MC, Farine DR. Inferring influence and leadership in moving animal groups. Philosophical Transactions of the Royal Society B: Biological Sciences. 2018;373(1746):20170006. doi:10.1098/rstb.2017.0006.

61. Planckaert J, Nicolis SC, Deneubourg JL, Sueur C, Bles O. A spatiotemporal analysis of the food dissemination process and the trophallactic network in the ant Lasius niger. Scientific Reports. 2019;9(1):1–11. doi:10.1038/s41598-019-52019-6.

62. Hasenjager MJ, Hoppitt W, Leadbeater E. Network-based diffusion analysis reveals context-specific dominance of dance communication in foraging honeybees. Nature Communications. 2020;11(1):1–9. doi:10.1038/s41467-020-14410-0.