Mining Bursty Groups from Interaction Data

Alexander Gorovits
agorovits@albany.edu
University at Albany—SUNY

Lin Zhang
lzhang22@albany.edu
Tencent.com

Ekta Gujral
egujr001@ucr.edu
University of California, Riverside

Evangelos Papalexakis
epapalex@cs.ucr.edu
University of California, Riverside

Petko Bogdanov
pbogdanov@albany.edu
University at Albany—SUNY

ABSTRACT

Empirical studies and theoretical models both highlight burstiness as a common temporal pattern in online behavior. A key driver for burstiness is the self-exciting nature of online interactions. For example, posts in online groups often incite posts in response. Such temporal dependencies are easily lost when interaction data is aggregated in snapshots which are subsequently analyzed independently. An alternative is to model individual interactions as a multi-dimensional self-exciting process, thus, enforcing both temporal and network dependencies. Point processes, however, are challenging to employ for large real-world datasets as fitting them incurs super-linear cost in the number of events. *How can we efficiently detect online groups exhibiting bursty self-exciting temporal behavior in large real-world datasets?*

We propose a bursty group detection framework, called MYRON, which explicitly models self-exciting behavior within groups while also accounting for network-wide baseline activity. MYRON imposes bursty temporal structure within a scalable tensor factorization framework to decouple within-group interactions as interpretable factors. Our framework can incorporate different "shapes" of temporal burstiness via wavelet decomposition or kernels for self-exciting behavior. Our evaluation on both synthetic and real-world data demonstrates MYRON's utility in community detection. It is up to 40% more effective in detecting ground truth groups compared to state-of-the-art baselines. In addition, MYRON is able to uncover interpretable bursty patterns of behavior from user-photo interactions in Flickr.

CCS CONCEPTS

• **Mathematics of computing** → *Stochastic processes*; • **Computing methodologies** → **Factorization methods**.

KEYWORDS

Hawkes Process; Self-Exciting Processes; Tensor Factorization; Community Detection; Dynamic Networks; Temporal Graphs

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

ACM ISBN 978-1-4503-8446-9/21/11…$15.00
https://doi.org/10.1145/3459637.3482370

ACM Reference Format:
Alexander Gorovits, Lin Zhang, Ekta Gujral, Evangelos Papalexakis, and Petko Bogdanov. 2021. Mining Bursty Groups from Interaction Data. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3459637.3482370

1 INTRODUCTION

Data from many domains can be represented as an evolving network where edges correspond to interactions between actors and/or items. Examples abound: message exchanges in social networks, reviews of products on Amazon or businesses on Yelp, and packet exchange events among internet hosts. Frequently, network interactions are generated from an underlying group structure: related individuals, groups of related products or interests, or botnet IPs. Understanding such groups and their temporal behaviors can yield key insights about the data and aid downstream applications such as recommender systems, forecasting and anomaly detection.

The collective timing of interaction events contains critical information about the inherent groups of nodes [18, 30]. Intuitively, nodes within a group exhibit coherent temporal interactions and individual interactions beget subsequent within-group interactions in a self-exciting manner [52, 55]. For example, comments on a forum post typically elicit other comments in reply or product reviews might invite further discussion. Such "bursty" behavior has been described in settings from social to biological systems [14, 24, 36]. Importantly, activity of individual nodes in the group serves to drive overall group-level activity, as opposed to individual self-excitation. Hence, tracking the temporal activity of the group over time is crucial to identifying its members. Additionally, as pairs of nodes can participate in multiple groups (e.g. friend or work groups on a social network), the timing of interactions relative to others yields information about the group identity of edges that is difficult to derive from the interacting entities alone.

There are several key challenges in detecting groups with bursty behavior. First, explicitly modeling self-exciting behavior among all observed interactions through widely-adopted point process models such as Hawkes processes [21] is prohibitive for large datasets with millions of interactions [29]. Alternatively, arbitrary aggregation of the data as tensors [58] or evolving network snapshots [57] can offer scalability at the expense of losing inter-temporal influence and quality of group detection. A trade-off between the above two strategies is possible and is a key contributions of this work.

A second challenge is that groups may overlap and exhibit varying degrees of burstiness, with some spiking faster or higher [16]. Additionally, group-level bursty activity may be "buried" in background or other-group activity. Finally, groups present in a system

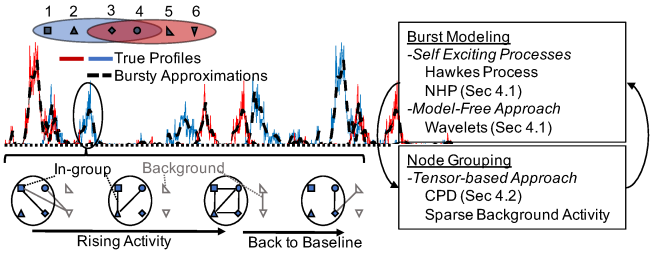


Figure 1: Six nodes are members of two overlapping groups (red and blue; top left), with temporal activities displayed in the central panel. The bottom panel demonstrates in-group interactions varying over time alongside random background activity potentially involving group nodes. The overall, dual-thrust approach of MYRON involves simultaneously fitting tensor factorization-based group structure and enforcing bursty temporal behavior on the temporal mode of said groups.

may exhibit variable temporal activity, even effectively "disappearing" and "reappearing" in the data. A robust detection method must be able to track such a discontinuous group as a consistent entity as opposed to clusters of nodes in individual bursts.

Figure 1 presents an example of our setting and an overview of our approach. We seek to identify group membership of nodes alongside overall group temporal behavior from unlabeled, attribute-free weighted dynamic graphs representing node interactions over time. We propose a framework for *Mining burstY gROups from iNteractions (MYRON)*, based on regularized tensor factorization that enforces group-level self-exciting structure on the temporal factors. Our discrete time approach allows for scalability to large datasets while enforcing long-range temporal influence among snapshots, thus striking a trade-off between point process models and aggregated snapshot approaches. MYRON allows flexibility in modeling the bursty group structure: we demonstrate two alternatives based on wavelets and point processes. Our evaluation on synthetic and real-world datasets demonstrates that MYRON is advantageous compared to state-of-the-art techniques for both group detection and temporal burst detection. We also demonstrate that a key parameter—the number of bursty groups—can be estimated within practical quality for a dataset of interactions.

Our contributions in this work are as follows:

- We propose a general framework for bursty group detection from interactions which offers a middle ground between expensive point processes and temporal-dependency-agnostic snapshot methods.
- We introduce two methods for imposing burstiness via: i) an interpretable Poisson model, and ii) a light-weight relaxation based on wavelet decomposition. We also propose a method for estimating the number of groups.
- Our exhaustive evaluation of MYRON on synthetic and real-world datasets demonstrates its advantage for group and temporal burst detection, with quality improvement over baselines of up to 30% in synthetic and 35% in real-world data.

2 RELATED WORK

Community detection: Community detection is a fundamental problem in network analysis. While community detection in static networks has enjoyed sustained research interest [13, 51], community detection in dynamic networks is a newer field; for a recent review, see [43]. A variety of approaches consider evolving communities of (smoothly) varying membership [30, 53]. This group of approaches focuses on tracking communities' memberships as opposed to their varying activity. Other methods, similar to our setting, focus on the evolving behavior of static-membership communities [11, 18, 44, 54]. Similar to us, many of them leverage tensors to represent dynamic interaction data [2, 15, 20, 39]. Others model the internal community structure as deviating from the global network behavior [22]. Global temporal network analysis for community detection has also received increased attention recently including methods employing local profiles [12] and uncovering anomalous temporal communities [2, 5, 34]. Recent works have demonstrated the importance of imposing constraints on community temporal behavior through Fused-LASSO regularization for temporal factors [18], wavelet-based treatment of the temporal mode in a Tucker decomposition [47], periodic behavior [54] and general periodicity, trends and graph smoothness via appropriate dictionary encoding [33]. In contrast to all the above, we explicitly model bursty self-exciting behavior and the influence of past in-group interactions on the future, resulting in superior group detection as we demonstrate empirically. In particular, bursty behavior results from long range dependencies that short-history models cannot capture.

Bursty behavior modeling and burst detection: Bursty processes in data streams or point processes have been modeled as Markovian [25] and through long-memory self-exciting point processes such as the Hawkes process [21]. Hawkes processes have been employed to uncover the underlying network structure among entities [28], to group event sequences [50], and for large-scale inference [27]. This line of work, however, typically suffers from limited scalability due to long-range inter-event relationships and the computational cost of maximum-likelihood methods typically employed. Work by Linderman and Adams [29] seeks to alleviate the limited scalability by temporal aggregation within a Bayesian framework, though without grouping in the node domain. There is also recent research on continuous-time-aware tensor decomposition[49, 57] which combines ideas from Hawkes point processes and tensor factorization methods, however the approaches in this group suffer from limited scalability to large-scale networks frequently encountered in the online domain.

Dynamic graph mining and evolutionary clustering. A parallel line of work to ours is the paradigm of evolutionary community detection [9, 35], where varying community memberships may be observed over time. In our approach, even though the communities discovered by the tensor factorization are static, we are able to incorporate such considerations by assuming a threshold upon which a community is considered to have evolved to a different one [3, 37] (which may be application-dependent), and increase the number of tensor factors in order to accommodate for a larger number of communities to be discovered, since our method allows for potentially significant overlap between communities.

3 PRELIMINARIES AND NOTATION

We introduce key notation used throughout the paper in Table 1. The input data in our problem setting is a set of temporal interactions of the form (i, j, t), where $i \in V1, |V1| = N1$ and $j \in$

$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N1 \times N2 \times T}$	3-way temporal interaction tensor
$\mathbf{X}_{(i)}$	The mode-i matrix of $\mathbf{X}$.
$\mathbb{1}$	A tensor consisting of all ones.
$Ci \in \mathbb{R}_{\geq 0}^{Ni \times K}$	Group membership matrix for mode i, K is the number of groups/factors
$A \in \mathbb{R}_{\geq 0}^{T \times K}$	Temporal profile matrix of groups/factors
$S \in \mathbb{R}_{\geq 0}^{T \times K}$	Shape-based approx. for temporal profile
$s \in \mathbb{R}_{\geq 0}^{K}$	Factor scaling vector
b_X	scalar global background activity level
$\|B\|_F$	The Frobenius norm $\|B\|_F^2 = \sum_i \sum_j B_{ij}^2$
λ	Importance param. for shape constraint
$\alpha_0, \bar{\alpha}$	Initial and baseline NHP intensity
γ, β	per-event intensity gain and persistence

Table 1: Key notation used throughout the paper.

$V2, |V2| = N2$ are the interacting entities (or nodes), and $t \in [1, T]$ is the discrete time of the interaction occurrence. Note that when analysing user-user interactions (i.e., $V1 \equiv V2$) the problem becomes one of community detection, while in the general case $V1$ and $V2$ could represent different sets of entities, e.g., people and photos. We aggregate all interactions in a three-way tensor $\mathbf{X} \in \mathbb{R}_{\geq 0}^{N1 \times N2 \times T}$ with entities corresponding to the number of interactions, i.e., we allow for multiplicity of pairwise interactions within the same time point. This allows for handling node addition/removal over time by working with the union of all nodes.

The canonical polyadic decomposition (CPD) [42] is a tensor extension of SVD and factorizes a tensor $\mathbf{X}$ as a sum of K rank-one tensors $[[C1, C2, A]]$, where $Ci \in \mathbb{R}_{\geq 0}^{Ni \times K}$ are factor matrices for the first two modes which can be interpreted as groupings of the corresponding entities, while $A \in \mathbb{R}_{\geq 0}^{T \times K}$ represents the activity profiles (temporal factors) corresponding to each of the K groups. $C1$ and $C2$ are in general different (e.g. in bipartite graphs), however, in the case of undirected graphs on a single type of node, they convey the same information.

4 PROBLEM FORMULATION

Our model is based on the following (informal) generative process:

(1) *Interactions arise within overlapping groups of entities whose strength of group association may vary;*

(2) *The intensity of within-group interactions exhibits a bursty self-exciting structure over time.*

(3) *Background "non-group" interactions of a fixed intensity arise between random pairs.*

Our goal is to recover the bursty group structure corresponding to the above behavior from temporal interaction data. We formalize the problem as a regularized tensor factorization as follows:

Bursty Group Detection: Given a tensor $\mathbf{X} \in \mathbb{R}_{\geq 0}^{N1 \times N2 \times T}$ of weighted temporal interactions, group number K, and temporal fit parameter λ, *find* factor matrices $Ci \in \mathbb{R}_{\geq 0}^{Ni \times K}, i \in 1, 2$ representing node group membership and $A \in \mathbb{R}_{\geq 0}^{T \times K}$ representing corresponding group temporal behaviors based on:

$$\min_{C1, C2, A, b_X, S} ||\mathbf{X} - [[C1, C2, A]] - b_X \mathbb{1}||_F^2 + \lambda ||A - S||_F^2 \quad \text{s.t. } C1, C2, A, b_X \geq 0, \tag{1}$$

where S is the bursty shape approximation matrix, and b_X models the global average interaction level.

The first term in the objective above is a tensor reconstruction loss augmented by a background interaction level controlled by b_X, while the second is a penalty enforcing alignment of the temporal factor with a bursty activity simultaneously learned in the shape matrix S. Scalar λ balances the fit and burstiness.

Our solution framework, called MYRON, is illustrated in Fig.1. MYRON addresses all challenges in bursty community detection: limited scalability of continuous time methods, overlapping group structure with varied behaviors, and background noise. In the rest of this section we discuss two alternative approaches of modeling burstiness in S and justify the specific form of the background model b_X.

4.1 Enforcing bursty structure by S.

One of our key design principles is that within-group activity over time is self-exciting and forms bursts of interactions, i.e. in-group interactions boost successive in-group interactions. In addition, the intensity of this self-exciting behavior is specific to individual groups. Within our objective we enforce this through a reference shape S which we model as group-specific bursty time series. A regularization-based approach for this objective allows for a controlled level of variation in valid profiles and penalizes non-bursty temporal fits for groups. We consider two models for this reference bursty shape: (i) a non-homogeneous Poisson process as an "aggregate" extension of a Hawkes process, and (ii) a time series with sparse wavelet reconstruction via the *Daubechies basis* which resembles bursts in time. The former is rooted in a long history of self-exciting process models, while the latter offers an efficient yet high-quality alternative.

Non-Homogeneous Poisson (NHP) burstiness: Self-exciting point processes are a common, if computationally expensive, way to model burstiness of event sequences. We adapt the exponential kernel Hawkes process formulation [21] for event intensity at time t, $\alpha_t = \alpha_0 + \sum_{\tau < t} \Phi(t - \tau)$, and introduce a *Non-homogeneous Poisson process (NHP)* for our aggregate group-level activity. We define the same intensity, or number of expected in-group interactions in interval $[t, t+1)$, α_t, as: $\alpha_t = \bar{\alpha} + (\alpha_{t-1} - \bar{\alpha}) * \beta(\Delta_{t-1}) + \gamma \hat{\alpha}_{t-1}$. Here $\bar{\alpha}$ is a persistent baseline intensity describing the arrival of within-group non-burst events, β is a decay kernel, γ is additional intensity from events in the previous period, and $\hat{\alpha}_t = \frac{N_t}{\Delta_t}$ is the empirical rate of events in that period. This per-interval intensity is described by a baseline, decayed intensity from the prior period, and additional intensity from recent events.

Assuming a constant time interval length (similar to [29]), i.e. $\exists c : \forall t, \Delta_t = c$, yields a constant decay kernel β. We can rewrite the intensity recursively as:

$$\alpha_t = \bar{\alpha}(1 - \beta^t) + \beta^t \alpha_0 + \gamma \sum_{i=1}^{t} \beta^{i-1} \hat{\alpha}_{t-i}, \tag{2}$$

where α_0 is the initial intensity. Note that since Δ_t is absent from the above equation, we can rescale time such that $\Delta_t = 1$ without loss of generality and think in units of "time windows". Our NHP model is an aggregate version of the exponential kernel Hawkes process and the two are asymptotically equivalent. Specifically, one can show the following relationship:

Theorem 4.1. *Let $\beta(\Delta_t) = \beta^{\Delta_t}$, then the model from Eq. 2 converges to an exponential-kernel Hawkes process as $\Delta_t \to 0$ and $t \to \infty$.*

Proof. As the window size decreases, we reach a point where $\alpha_t \in \{0, 1\}$, i.e. there is at most one event per period. Now, as $t \to \infty$, we approach $\bar{\alpha} + \gamma \sum_{i=1}^{t} \beta^{i-1} \mathbf{1}_{\alpha_i=1}$, which we can rewrite as $\bar{\alpha} + \sum_{\tau<t} \gamma \beta^{t-\tau}$, thus, arriving at the standard Hawkes formulation. □

For interpretability, we require several natural restrictions on the parameters in Eq. 2. All parameters should be non-negative: negative intensity has no meaningful interpretation, while $\gamma < 0$ implies self-attenuation as opposed to self-excitation. In addition, $0 \leq \beta < 1$, as we expect a decaying influence with time, thus avoiding intensity buildup in the absence of events.

To impose an NHP-like structure on temporal factors in Eq. 1, the shape S in the regularization term $||A - S||_F^2$ is instantiated with a matrix of time series resulting from bursty NHP fits of within-group events. In other words, $S_{:k} = \{\alpha_t^{(k)}\}$, where $\{\alpha_t^{(k)}\}$ is the NHP intensity fit for the k-th group's temporal component. Note that in addition to enforcing a bursty shape, an NHP-based solution also yields burstiness parameters characterizing group activity.

Daubechies wavelets: While the NHP-based burstiness definition is interpretable and rooted in a long-standing body of work on point process modeling, its recursive nature comes with a considerable computational cost. Thus, we also propose an efficient alternative based on "fitting" factor activity with an appropriate wavelet basis decomposition. The Discrete Wavelet transform [31] models a time series as a combination of scaled and shifted wavelet filters which form a multi-resolution representational basis. Informally, the wavelet transform will represent an arbitrary temporal group profile in our framework as consisting of a small number of scaled and shifted kernel "shapes". In particular, if the kernel shape is "burst-like", then the temporal profile will be represented by a number of appropriately placed and sized bursts. By retaining only high-energy wavelet coefficients, one can reduce the noise [56] in group activity and focus only on the bursts themselves. For the purposes of our objective function, S in the wavelet case is a matrix of group-specific time series reconstructed by sparse wavelet transform. The shape of the Daubechies wavelet [10] captures a "rising front" behavior similar to the NHP formulation, hence we adopt it for regularization. Alternative wavelet bases can also be employed.

4.2 Tensor model with background activity.

In the data fit term of Eq. 1 we augment CPD to include a constant rank-one factor representing a background activity, namely:

$$\mathbf{X} \approx \sum_{K} [c1_k \otimes c2_k \otimes a_k] + b_X \mathbb{1}, \tag{3}$$

where K is the number of bursty groups (or factorization rank) and $c1_k, c2_k, a_k$ are vectors representing single-factor loadings in the respective tensor mode. This modeling decision explicitly allows for a fraction of node interactions to arise outside the context of recurring bursty groups. In the case of user-user interactions, this background behavior can be due to weak ties [19]. Alternatives to a constant background level can also be considered (e.g. global periodic behavior or smooth trend), but the simpler approach was advantageous across datasets, and thus, we focus on it for our presentation. The introduction of the baseline activity tensor $b_X \mathbb{1}$ is akin to removing the global mean of the input $\mathcal{X}$. Another means to this end could be to pre-process the data by subtracting the mean from the tensor element-wise. Such an approach will "densify" typically sparse input data resulting in significant computational and memory overhead [4].

5 OPTIMIZATION SOLUTION

We next present optimizations for of our objective in Eq. 1 which requires solving for three factor matrices $C1, C2, A$, the baseline level b_X, and the detailed parameters θ of the temporal shape, which in the NHP case consist of per-group $(\alpha_0, \bar{\alpha}, \gamma, \beta)$, and in the wavelet of a set of per-group vectors of wavelet coefficients. We outline the steps for solving the compound problem below.

5.1 An alternating optimization solution

Alternating Optimization is a common technique in a variety of settings, particularly in the case of tensor factorization [23, 26]. We update individual components of the problem in a cyclic manner while holding other components fixed, cycling through updates of $C1, C2, A, S, b_X$. Like coordinate descent, as long as each individual update is optimal, this is a non-increasing operation to the objective, as each step minimizes an objective already minimized for another component by the previous step. Since the overall objective is bounded, this monotone behavior guarantees convergence. The structure of the problem leads to related updates for all variables:

- $C1, C2$ are nonnegative factors updated via

$\arg\min_{Ci} ||X - [[C1, C2, A]] - b_X \mathbb{1}||_F^2$ s.t. $Ci_{nk} \geq 0, \forall (n, k)$,

- A is updated via a "shape-based" constraint

$\arg\min_A ||X - [[C1, C2, A]] - b_X \mathbb{1}||_F^2 + \lambda ||A - S||_F^2$,

- S is updated by fitting parameters to the estimated A from above to obtain an approximation for S,
- b_X is updated based on the current estimates of $C1, C2$ and A.

Updates for A, C and $C2$ via ADMM. The Alternating Direction Method of Moments (ADMM) is commonly used [6] for dividing a complex optimization problem into simpler components. It has been recently applied to other regularized tensor decomposition objectives [18, 23]. Intuitively, ADMM works by solving each subproblem separately while enforcing similarity between the solutions. For the temporal factor A, the ADMM objective is:

$$\begin{aligned} \min_{C1, C2, \tilde{A}, A, b_X} \quad & ||\mathcal{X} - [[C1, C2, \tilde{A}]] - b_X \mathbb{1}||_F^2 + \lambda ||A - S||_F^2 \\ \text{s.t. } & C1, C2, A, \tilde{A}, b_X \geq 0, A = \tilde{A} \end{aligned} \tag{4}$$

The ADMM update involves iterating over the three updates:

$$\begin{cases} \tilde{A} \leftarrow (\check{C}^T \check{C} + \rho I)^{-1} (\check{C}^T \mathbf{X}_{(3)} + \rho (A + R_A)^T) & (a) \\ A \leftarrow \arg\min_A \lambda ||A - S||_F^2 + \rho/2 ||R_A + A - \tilde{A}^T||_F^2 & (b) \\ R_A \leftarrow R_A + A - \tilde{A}^T, & (c) \end{cases} \tag{5}$$

where $\check{C}$ denotes the Khatri-Rao product of $C1, C2$. The first update can be efficiently computed via the lower Cholesky decomposition of $\check{C}^T \check{C} + \rho I$ and the matricized tensor times Khatri-Rao product (MTTKRP) representation of $\check{C}^T \mathbf{X}_{(3)}$, similar to prior approaches [18, 23]. With S known, the solution to 5(b) becomes $A = (\lambda S + \frac{\rho}{2} \tilde{A}) / (\lambda + \frac{\rho}{2})$.

ADMM updates for $C1$ and $C2$ are similar:

$$\begin{cases} \tilde{C} \leftarrow (\check{C}^T \check{C} + \rho I)^{-1} (\check{C}^T \mathbf{X}_{(3)} + \rho (C + R_C)^T) & (a) \\ C \leftarrow \arg\min_{C>0} \rho/2 ||R_C + C - \tilde{C}^T||_F^2 & (b) \\ R_C \leftarrow R_C + C - \tilde{C}^T, & (c) \end{cases} \tag{6}$$

where C is the updated factor ($C1$ or $C2$) and $\check{C}$ is replaced with the product of the fixed factors (e.g. $C2$ and A when updating $C1$). Non-negativity is enforced by retaining positive values in Eq. 6(b).

Algorithm 1 MYRON

Require: Tensor X, group count K
Ensure: Factorization $\{C1, C2, A\}$, scaling s
1: Initialize $C1, C2, A$
2: **while** Factors not converged **do**
3: $C1, s \leftarrow \text{FACTOR}(X_{(1)}, b_X, C2 \odot A, s, K, \text{nonnegative})$
4: $C2, s \leftarrow \text{FACTOR}(X_{(2)}, b_X, C1 \odot A, s, K, \text{nonnegative})$
5: $A, s \leftarrow \text{FACTOR}(X_{(3)}, b_X, C2 \odot C1, s, K, \text{shape})$
6: $S, params \leftarrow \text{ShapeApprox}(A)$
7: $b_X \leftarrow \frac{\sum_{ijt} X_{ijt} - \sum_k (\sum_i C1_{ik})(\sum_j C1_{jk})(\sum_t A_{tk})}{N1 \times N2 \times T}$

Solving for the shape S. The solution to the shape subproblem depends on the specific constraint considered. In each case, however, the idea is similar - we fit the temporal profiles A to a specific model of temporal activity, and use the resultant idealized approximation as our shape matrix S. When A is known, solving for S in either the NHP or wavelet formulation requires first fitting a set of parameters describing the burstiness of proposed traces.

• *The NHP case:* The problem consists of fitting the four parameters $\alpha_0, \bar{\alpha}, \beta, \gamma$. This can be done independently for each group $k = 1...K$. We iterate over the following updates:

$$\begin{cases} \bar{\alpha}^* = \frac{\mathcal{N} - \sum_t \beta^t A_t - \alpha_0 \sum_t (1-\beta^t)\beta^t - \gamma \sum_t \xi_t (1-\beta^t)}{\sum_t (1-\beta^t)^2} & (a) \\ \alpha_0^* = \frac{(1-\beta^2)(\sum_t A_t \beta^t - \bar{\alpha} \sum_t \beta^t (1-\beta^t) - \gamma \sum_t \xi_t \beta^t)}{1-\beta^{2T}} & (b) \\ \gamma^* = \frac{\sum_t A_t \xi_t - \bar{\alpha} \sum_t \xi_t (1-\beta^t) - \alpha_0 \sum_t \beta^t \xi_t}{\sum_t \xi_t^2} & (c) \\ \beta^* = \arg\min_\beta \sum_t [A_t - (\bar{\alpha}(1-\beta^t) + \beta^t \alpha_0 + \gamma \sum_{i=1}^t \beta^{i-1} A_{t-i})]^2, & (a) \end{cases} \quad (7)$$

where $\xi_t = \sum_{i=1}^t \beta^{i-1} A_{t-i}$ and $\mathcal{N} = \sum_{t=0}^T A_t$. All but the β^* update can be computed analytically from first derivatives. The complexity is driven by Eq 7(d) and specifically the computation of $(A_t - \alpha_t)$ which can be done in $O(T)$. Precomputing ξ each iteration is done via a single pass through the time series. Overall, the complexity per update is $O(KT)$, which can be reduced by approximating the objective. For instance, truncating the influence memory (e.g. consider A_{t-i} for $i \in [t - O(1), t]$) relaxes the dependence on T in computing β^* and minimally affects the solution when β is small.

• *The Wavelet case:* Given a fixed A, we can compute the complete wavelet decomposition via Mallat's pyramid algorithm [32] which involves iteratively applying approximation and detail filters specific to the wavelet form. A fraction of the highest (by absolute value) coefficients are retained. This thresholding is theoretically optimal in terms of minimizing the mean squared error (MSE) for sparse wavelet reconstruction [46]. A fraction of 2% yielded maximal performance across datasets in our evaluation. To obtain the shape matrix S, we reconstruct each time series from the sparse coefficient vector by a reverse transform.

5.2 The overall MYRON algorithm

We combine the above updates in the steps of MYRON outlined in Alg. 1. The outer loop iterates over each of the three factor matrices with the remainder of the parameters fixed (lines 3-5) following the ADMM process derived in Sec. 5.1 and detailed in Alg. 2. In line 6, ShapeApprox() refers to the shape-specific fitting (NHP or Wavelet) described in Section 5.1 (Solving for the shape S). The mean of the residual tensor is computed in Step 7, as the difference of the sum of all entries of $\mathbf{X}$ and the sum of the factor-based reconstruction.

In addition to the steps described above, we perform two optimizations that aid computation and convergence:

Algorithm 2 FACTOR

Require: Tensor $\mathbf{X}$, mean b_X, product of known factors W, scale s, group count K, constraint type *cons*
Ensure: Fitted factor H, scale s
1: Initialize H, R
2: $\hat{W} \leftarrow s \odot W$
3: $G \leftarrow \hat{W}^T \hat{W}$
4: $\rho \leftarrow \text{trace}(G)/k$
5: $L \leftarrow \text{LowerCholesky}(G + \rho I)$
6: $F = \hat{W}^T(\mathbf{X} - b_X \mathbb{1})$
7: **while** Not Converged **do**
8: $\tilde{H} \leftarrow (L^T)^{-1} L^{-1} [F + \rho(H + U)^T]$
9: **if** *cons* == nonnegative **then**
10: $H \leftarrow max(0, \tilde{H}^T - U)$
11: **else if** *cons* == shape **then**
12: $H \leftarrow \left[\lambda S + \frac{\rho}{2}(\tilde{H}^T - U)\right] / (\lambda + \frac{\rho}{2})$
13: $R \leftarrow R + H - \tilde{H}^T$
14: $H \leftarrow H./\text{colmax}(H)$
15: $s \leftarrow s * \text{colmax}(H)$

(1) *Fitting Sparse Global Behavior:* Although using the mean of the overall adjacency tensor as global behavior is a simple preprocessing step on paper, simply subtracting the mean of the overall $\mathbf{X}$ would yield a large dense tensor which will be inefficient to decompose. Instead, we represent the aforementioned mean-scaled constant Kruskal tensor in its decomposed form (i.e. three vectors of ones and the single scalar b_X) and utilize the associativity of the MTTKRP to compute $\hat{W}^T(X_{(*)} - b_X \mathbb{1}_{(*)})$ as follows:

$$MTTKRP(\hat{W}^T, X_{(*)}) - MTTKRP(\hat{W}^T, b_X \mathbb{1}_{(*)}).$$

(2)*Enforcing Scale:* The approach in Alg. 1 may suffer from scaling problems between the individual factors. Namely, if the total energy in the activity of some factor is greater than that of others, the Frobenious norm shape regularizer will favor smooth fitting of the k-th factor while essentially ignoring all others, thus yielding poor fits for less active communities or potentially splitting a larger one and missing others entirely. This problem can be exacerbated by the scaling freedom of CPD. In particular, different initialization of $C1, C2, A$ can lead to varying amounts of energy assigned to the temporal mode of any given group. We solve this problem by normalizing each factor as we fit it, holding each column of $C1, C2, A$ at most 1 and maintaining the group "strength" in an overall scaling vector s. Prior work has shown that conducting similar iteration re-normalization of factors for CPD improves convergence [41].

We measure overall convergence via changes in the normalized reconstruction error. We use a convergence threshold of 10^{-7} unless stated otherwise. Relaxing this threshold or bounding the number of iterations yields significant speed-up at minimal quality cost.

The complexity is dominated by calls to FACTOR (Alg. 2), which is in turn dominated by a MTTKRP [18] (line 6) and the matrix inversion in (line 8). Employing Lower Cholesky decomposition and sparse operations one can reduce the complexity to $O(K^3 + mK + t_l m S_L)$, where m, t_l, S_L represent the number of non-zeros in the matrix-tensor product within fixed factors, the number of while loop iterations (typically small), and non-zeros in the Cholesky decomposition of the fixed factor product, respectively.

5.3 Estimating the group count K

The group count K is a key parameter in MYRON and hence we propose and evaluate a method to estimate it. One heuristic approach

Algorithm 3 MYRON-CCD: Detect k^* via burst-aware CCD

Require: Tensor $\mathbf{X}$
Ensure: Activity-aware optimal k^*
1: **for** k=minK:maxK **do**
2: $[C1, C2, A, s] \leftarrow$MYRON$(X, k)$
3: $[U, \Sigma, V] \leftarrow SVD(A)$
4: **for** r=1: k **do**
5: $Ut = U(:, 1:r)$
6: Set $Ut(|Ut| < 0.1 * max(Ut)) = 0$
7: $\Pi_r = Ut * Ut^T$
8: $\mathbf{X}_r = \mathbf{X} \times_3 \Pi_r$
9: $A_r = \Pi_r * A$
10: c(r) = efficient_corcondia($\mathbf{X}_r$,$C1, C2, A_r$,s)
11: Store $(k, \max(c))$
12: k^* = multi_obj_opt({(k, c)})
13: **return** k^*

for k estimation in tensor decomposition is the core consistency diagnostic (CCD) [7]. To account for the contribution of strong bursts to temporal behavior, we modify CCD to account for periods of bursts in the temporal dimension by "aggregating" the input tensor in time according to the temporal factor A. Intuitively, a good estimate for A will allow us to aggregate the timeline and along with the community factors $C1$ and $C2$ will also yield a "good" factorization for a temporally aggregated input tensor, where goodness is measured in terms of high CCD value.

We operationalize the above intuition and show the steps of MYRON-CCD in Alg. 3. We iterate over a range of candidate group counts K and apply MYRON to yield factors $C1, C2, A$ (Step 2). We then apply efficient CCD [40] to associate the optimal CCD c with each potential rank k (Steps 3-11) by first projecting A and the input tensor $\mathbf{X}$ onto the top SVD vectors for all potential ranks up to k. Subsequently, we use the multi-objective optimization procedure described in [38] to determine the best pair (k^*, c^*) and thereby the estimated group count K^*. In our case, we may expect periods of non-bursty activity to be less informative and therefore aggregated into a single time stamp. MYRON's accounting for global baseline behavior also serves to smooth out noise. We evaluate MYRON-CCD and compare it to burst-agnostic estimators also employing CCD. We show that MYRON-CCD's estimates improve with increasing ground truth K.

6 EXPERIMENTS

6.1 Datasets.

Synthetic Data: To evaluate the performance of our method on data with known behavior, we generate a set of small synthetic datasets representing communities within an unweighted dynamic graph. We create two overlapping internally interacting groups, which allows for controlled variation in experimental parameters, along with overlaying a background activity level. Edges are generated from two groups of thirty nodes, with 10 nodes participating in both groups, according to group-level activity traces. These traces are generated from an NHP with known parameters ($\bar{\alpha} = 0.1, \alpha_0 = 0, \gamma = 0.5, \beta = 0.45$), with one group being offset by 10 timepoints from the other to retain a difficult separability, and contain three small bursts (see the top panel in Fig. 3(e)). Edges are produced by randomly drawing the number of edges matching the activity of the group from the pairs of nodes within it. We overlay global behavior by producing a third group containing all 100 nodes with non-bursty behavior, where the number of edges in a given time period is drawn from a Poisson distribution with a constant or varying intensity depending on the experiment.

Synthetic data for K-finding experiments in section 6.6 is generated in a similar fashion, though the temporal profiles are created randomly per group with $\alpha \in [0, 0.5], \gamma \in [0.5, 1] \times \beta, \beta \in [0, 1]$ for 500 discrete time points. Communities are drawn as a random set of 50 nodes from within 200 with arbitrary overlap.

Real-world Datasets: We also evaluate our methods on several real-world interaction datasets which represent either user-user (Reddit, Github Repos) or user-item (Delicious, Flickr, Github Topics) dynamic interaction graphs.

• **Reddit:** This dataset is obtained from a full dump of Reddit comments collected in 2015 [1]. We combine multiple subreddits (each a ground truth group), with responses to comments or a top-level post recorded as bi-directional edges between the posting users. The final dataset contains "programming", "gaming", "politics", "geek", and "lost" for the first four months of 2009. Data is aggregated hourly.

•**Flickr:** This is a benchmark image captioning dataset for sentence-based image description. The dataset used in this paper is obtained by selecting the six most frequent tags from [45]. Tensor modes represent *user-image-date*, where date is at daily granularity.

• **Delicious[45]:** This dataset was obtained by crawling *Del.icio.us* portals during 2006 and 2007. The most recent postings were monitored over a period of several months to collect an initial list of user names. User pages were crawled for corresponding postings and stored as tags. The tensor modes represent *user-page-date* and a non-zero entry indicates that a user has tagged a web page in a given day. We use a set of frequently-used tags as groups, with the most-tagged pages and most active users within those tags as nodes (weighted by participation frequency).

• **Github:** We present two Github-derived datasets using hourly data from the first three months of 2016. Github Repos consists of users interacting with a set of six repos (those ranked 100-105 by activity within those three months, to avoid bot-heavy and single-user repos), with a ground truth of activity-weighted user-repo interaction. Github Topics consists of users who interact with repos tagged with five language-based keywords java, javascript, python, go, and C. We use repo events within these topics as our ground truth.

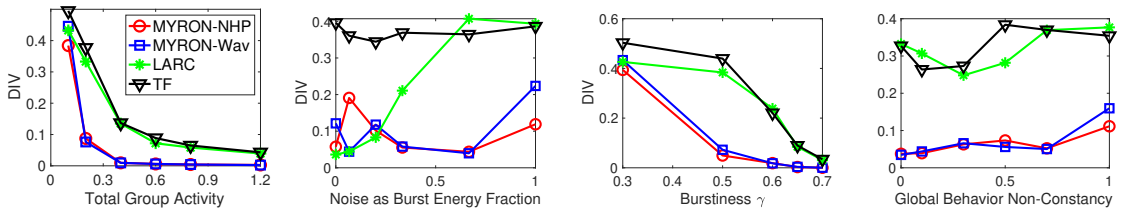
6.2 Experimental setup.

To demonstrate the benefits of incorporating proper temporal dependence and removal of the global trend, we compare two versions of our method, MYRON-NHP and MYRON-Wav, against two alternative methods. The first is direct non-negative tensor factorization via CPD (TF), which imposes no additional structure on the factorization. The second baseline LARC [18], assumes a piecewise-constant temporal behavior on groups. We downloaded LARC's implementation from the authors' website[1]. To tune the values of the two regularization parameters we performed a grid search between 0.01 and 10 (exponential step). The default parameters in the authors' implementation resulted in the best overall performance and we used them across experiments. We obtain non-negative tensor factorization (TF) solutions also from the implementation

[1] http://www.cs.albany.edu/~petko/lab/software/LARC_CODE.zip

	Statistics			Signed error ± std for estimation of K				MYRON-NHP		MYRON-Wav		LARC [18]		TF [15]	
Dataset	$\|\mathcal{V}\|$	T	K	MYRON-NHP	MYRON-Wav	TF-CCD	CCD	DIV	Time (s)	DIV	Time (s)	DIV	Time (s)	DIV	Time (s)
Reddit[1]	35196×35196	2880	5	10†	10†	10†	−3†	**0.521**	148.9	0.592	171.3	0.891	149.7	0.943	8.0
Delicious[45]	10k×10k	1430	10	**−1 ± 4.1**	−3 ± 6.3	5 ± 0.9	3 ± 1.4	**0.48**	307	0.664	72.7	0.763	288.3	0.819	19.5
Flickr[45]	3478*×100k	705	6	2.6 ± 6	2.3 ± 6.3	**0.5 ± 4.8**	−1.6 ± 4.6	**0.585**	98.6	0.755	43.7	0.945	68.7	0.941	1.4
Github Repos	8294×8294	2183	6	2.6 ± 3.2	5 ± 3.4	**−1.9 ± 1.1**	−2 ± 0.9	**0.584**	188.5	0.608	57.9	0.598	99.5	0.643	22.6
Github Topics	9020*×137819	2184	5	15†	15†	15†	−2†	**0.864**	45.5	0.894	31.5	0.901	146.0	0.911	7.4

Table 2: Real-world dataset statistics (columns 1-4), quality of k estimation (columns 5-8) and quality and running time of ground-truth group discovery (columns 9-17). (* Denotes the mode with ground truth groups.; † The estimates in these experiments are either the maximum tests K of 15 or or the minimum of 2 (extremes in the search space), indicating that no good matches were detected.



(a) Activity Level (b) Noise (c) Baseline Burstiness (d) Global Behavior

Figure 2: Performance comparison on ground-truth group detection for varying (a) activity volume, (b) noisy interactions, (c) burstiness, and (d) global behavior.

of LARC with only non-negative regularization on the temporal factor.

When running MYRON, we set the λ parameter at 0.2 for all experiments. Values in $[0.01, 10]$ were searched, with 0.2 yielding the best performance overall (though the effect of λ was not strong). Runtimes are given with a convergence bound of 10^{-5}, which yielded minimal changes in performance compared to 10^{-7} (see Figure 3(d)). MYRON-Wav retains the top 2% of the wavelet coefficients.

In all experiments we use group number K equal to the ground truth number of groups, which is obtained through the construction of the datasets. In general, K is an unknown parameter, but as our metrics are based on alignment with ground truth it is more difficult to measure the quality of a fit with a different number of communities. We also evaluate the quality of estimating K.

Results are presented as averages of five runs, with initializations identical across methods for each run. For timing purposes, experiments were run on a 20 core 2GHz Intel(R) Xeon(R) Gold 6138 server with 240GB of RAM; note that none of the methods are run in parallel, i.e. implementations utilize only a single core.

The primary comparison metric presented is Jensen-Shannon divergence (DIV), which measures the difference between two distributions. We treat the weights on the node factors as a distribution over nodes. Lower DIV values correspond to a better agreement with ground truth distributions of node weights within a community. Clusters are matched to ground truth in a greedy best-match fashion. We follow the same quality evaluation procedure as in the LARC baseline [18].

Implementation of MYRON is available at http://www.cs.albany.edu/~petko/lab/code.html.

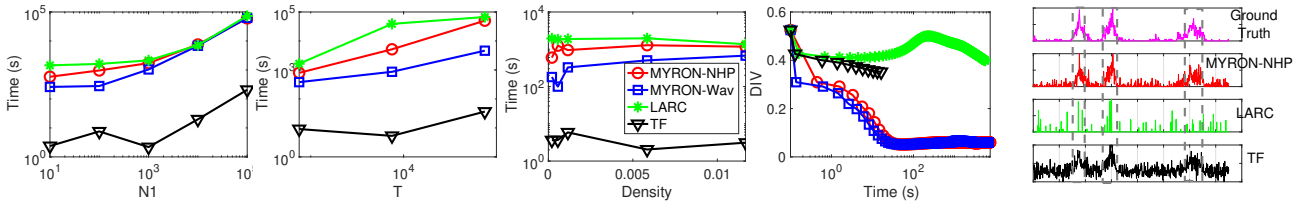
6.3 Detection of ground truth groups.

MYRON-NHP and MYRON-Wav effectively detect ground truth groups with a variety of behaviors. To determine sensitivity to variation in group and global activity levels , in Figs. 2(a) and 2(b) we vary the total energy (i.e. number of edges) within the group or within the global behavior respectively, while keeping the other fixed. In Fig. 2(a), low-activity groups lead to poor performance for all methods, however MYRON gains accuracy faster with increasing burst energy, even as early as 20% of the global energy within each group. Conversely, increasing global activity effectively serves as noise, masking in-group relationship between nodes. Low noise poses no challenge for all method but TF, however LARC quickly begins to add group membership due to background (non-group) interactions. At the same time MYRON is mostly insensitive to significant levels of non-group activity.

Fig. 2(c) shows the importance of modeling bursty behavior in particular by varying the γ parameter in the generated NHP groups. Low values of γ yield mostly static traces that exhibit neither high nor frequent bursts, whereas larger values yield strong and frequent bursts. The total group activity is normalized to be equal across all settings. All methods perform poorly at low burstiness, however MYRON quickly improves. High burstiness groups are more pronounced and easier to detect for all methods.

Figure 2(d) demonstrates that despite relying on mean centering, MYRON retains performance even with non-constant global behavior. The global group behavior is defined as $pB1 + (1-p)B2$, where $B1$ is a linearly increasing trace representing a common setting of overall network growth over time whereas $B2$ is a constant intensity global behavior. None of the methods are noticeably affected by variations in the global behavior, but the relatively high volume of this behavior (equivalent to 0.75 in Figure 2(b)) obscures the groups to LARC and TF.

Results from group detection quality in real-world datasets are presented in Tbl. 2 (last 8 columns). Overall, MYRON-NHP outperforms competitors despite occasionally being marginally slower. MYRON-Wav also does well to varying degrees likely dependent



(a) 1st Mode Node Count (b) Time Points (c) NNZ density (d) Convergence in Time (e) Burst Detection

Figure 3: Scalability on different dimensions (a-c) and convergence behavior (d) of all methods on Synthetic data. Panel (e) compares detection of three pronounced bursts (outlined), which MYRON-NHP fits in period and "shape".

on how well the wavelet basis describes the underlying data structure. However, MYRON-Wav is notably faster on most datasets. We present performance on bipartite graphs as well, in particular through the Delicious, Flickr, and Github Topics datasets, demonstrating that MYRON is not confined to community detection alone.

6.4 Scalability.

Figure 3 investigates the scaling of competing techniques with respect to the number of entities (Fig. 3(a)), T (Fig. 3(b)) and the fraction of non-zero entries (Fig. 3(c)). Fig. 3(a) shows qualitatively similar polynomial growth across all methods with larger $N1$, corresponding to similarity in the fitting process for node-mode factors - slowdown is largely due to increased convergence complexity. MYRON performs noticeably faster than LARC with only a slight relaxation in tolerance. Increasing T (Fig. 3(b)) also exhibits similar behavior with method-specific overhead, though the benefit of MYRON-Wav is clearer: NHP is more complex with time (which also accounts for the initial gap at small $N1$ and T. Running times are minimally affected by tensor density, at least in the regimes tested (Fig. 3(c)). Real interaction data is likely to be very sparse, e.g. constant communication within large cliques is a different setting altogether. Overall, MYRON's running time compares favorably with other discrete-time tensor-based methods, which themselves are significantly more scalable than continuous-time methods such as multi-dimensional Hawkes processes.

Fig. 3(d) provides more detail on the convergence properties of each method, for a $10k \times 100 \times 750$ tensor. Both versions of MYRON converge to near-optimal DIV only marginally slower than non-negative TF's convergence. Hence stopping early (or using a looser convergence criterion) can yield fast and good-quality results.

6.5 Burst detection.

In Fig. 3(e) we present the estimated bursts position for competing methods in synthetic data with a ground truth bursty profile (top). MYRON-NHP closely reconstructs the bursty profile of the ground truth behavior. Since LARC aims to obtain a piece-wise on/off behavior without too many switches it "breaks" bursts into short spikes, and thus, fails to detect the presence of longitudinal bursts. TF does not impose any shape regularization on the temporal information and simply minimizes the reconstruction error. While it manages to partially recover bursts, it detects too much non-burst activity leading to noisy detected groups and higher DIV in group detection.

6.6 Estimating K.

Next we evaluate the quality of MYRON-CCD (Sec. 5.3) for estimating the number of groups in synthetic and real datasets. We set $minK = 2$ and $maxK = 15$. We compare the estimates to those of two burst-agnostic alternatives: TF-CCD, which follows the same procedure as MYRON-CCD but uses TF instead of MYRON in step 2; and CCD [40] which is the original method for TF K estimation.

Fig. 4 presents the error of the estimates of all competing methods on synthetic datasets of 200 nodes over 500 time steps containing varying numbers of independently bursty, Hawkes-process generated groups. Results are averages of 5 different datasets per value of K. While accurate detection of group count is difficult (as evidenced by error throughout and significant variance), MYRON outperforms non-burst-aware alternatives especially for larger K. In particular the wavelet approach smooths out noise in the data and predicts K with increasing accuracy. In contrast using the method Alg. 3 with burst-agnostic TF as the factorization approach consistently overestimates K, likely due to an increased impact of noise. CCD without adjustment yields consistent underestimates of increasing error.

Real data results (Tbl. 2, columns 5-8) feature significant variance, but also show that a burst-aware approach estimates a reasonable number of communities when the ground truth K is big (e.g. for Delicious where K=10). Of note is that while the specific K is hard to estimate exactly, estimates by MYRON-CCD are reasonably close to ground truth values. While MYRON-CCD Error overestimates K when the ground-truth is slow (both Tbl. 2 and Fig. 4), running MYRON light slightly larger K is likely to capture the true communities regardless, alongside either background contributions or multiple version of the same community.

In general, very accurate detection of group count remains a difficult problem. For application purposes other approaches may also be considered, including for example cross-validation for value imputation [8, 48]. We note that our approach is not computationally trivial, requiring significant memory for the purposes of reconstructing a potentially dense "compression" of the input dynamic graph. However, our analysis demonstrates that i) finding a reasonable estimate for group count is not impossible, and ii) methods that properly account for temporal behavior in the groups have advantages.

6.7 Case study: analysis of a Flickr dataset.

Next we visualize and discuss the temporal behavior of detected groups by MYRON in Flickr. The Flickr dataset was crawled by Gorlitz et al. [17] and spans the period from January 2004 to December 2005. From the raw data, which consists of (User, Image, Tag, Time) tuples, we extract (User, Image, Time) tuples corresponding to six

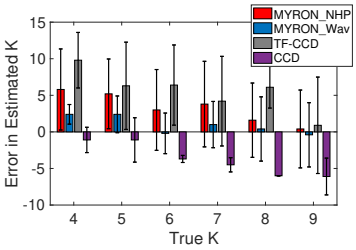


Figure 4: Signed error ± std (5 trials) comparison of MYRON-CCD and alternatives for estimating a increasing ground truth K in a synthetic datasets.

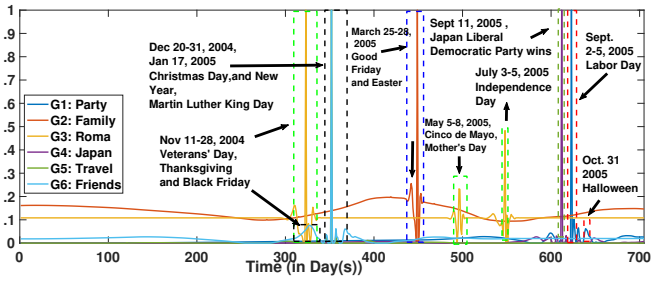


Figure 5: A case study of MYRON-Wav's detection of bursts in the Flickr dataset involving six frequent tags

frequently used tags: Party, Family, Roma, Japan, and Friends; the tags are then used as our ground truth groups.

The results of employing MYRON in Flickr are visualized in Fig. 5, where the x-axis is the time (day), and the y-axis is the factor value corresponding to each reconstructed group. The bursts of the curves can be explained by different tagging intensity within a particular group. For example, group 4 (Japan) has limited activity throughout the data collection period, however, a burst focused on a political event (a Liberal Democratic Party victory in multiple branches of government) is detected by MYRON-Wav. Another interesting tag/group, Roma (group 3), is active throughout the year. The images tagged by this group are mostly historic places like La Fontana dei Quattro Fiumi (Fountain of Four Rivers), Piazza Navona, Tiber and St. Peter church. As we do not have any demographic information of users, we mapped other groups i.e. Party, Family and Friends, to USA events. Results show bursts that are consistent with holiday seasons. We do not observe any significant bursts in group 5 (Travel), but this tag is active during holidays like Christmas and New year's eve, and may simply have a higher overall level of activity that is less distinctly captured.

7 DISCUSSION

Detection from raw interactions: Our experiments demonstrate that MYRON is able to detect ground truth communities containing bursts, solely using interaction data, more accurately than close baselines which do not explicitly account for bursty activity. Hence our methodology is especially applicable and necessary for social media and online forum discussion datasets whose activity is inherently self-exciting. The community burstiness that we leverage can be in the form of particular events to discuss, release-centered activity in a github repository within a topic "group", or topic-level events or salient moments as in the case study above.

Burst modeling: The distinction between the NHP and Wav variants of MYRON deserves special mention. Modeling bursty behavior via any means is beneficial, as both variants outperform baselines. However, in most of our datasets, the NHP version performs better quality-wise, bolstered by a more general modeling of bursty processes. This quality comes at a higher computational cost, particularly for longer time scales. In addition to group detection, NHP also provides a more general, faithful, and intepretable description of the "type" of burstiness of a given group in terms of the NHP process parameters. These parameters have physical equivalents in standard Hawkes parameters and can be employed to simulate more group interaction with "faithful" temporal properties.

On the other hand, wavelets essentially pinpoint burst locations without describing the entire temporal behavior. The Wav formulation also offers trivial migration to other wavelet forms for temporal group behavior which may fit particular data better. Many standard wavelets are less bursty in shape, but the MYRON framework is able to accommodate these regardless.

Limitations and extensions: Self-excitation is only one potential form of community behavior, though a common one. Trends over time and periodicity are other possible temporal patterns. If these occur at the global level, their contribution can be modeled by augmenting the baseline behavior term b_X in MYRON. At the community level, other temporal "shapes" can be considered, including discrete Fourier transforms, other wavelet forms, or a variety of other temporal behaviors that can be modeled as an approximate representation of noisy time series. We leave such extensions to future work.

8 CONCLUSION

In this paper we introduced a general and robust framework for bursty group detection from interaction data based on tensor factorization, called MYRON. We incorporated burstiness via two alternative models: i) an interpretable non-homogeneous Poisson model which generalizes classic Hawkes process models for individual events and ii) a light-weight alternative employing Daubechies wavelet decomposition. We performed extensive evaluation of our framework on synthetic and real-world datasets spanning different types of online interaction data. Our evaluation demonstrated the advantage of MYRON for group and temporal burst detection in comparison to recent state-of-the-art baselines. Our methodology enabled improvement of quality for group detection of up to 30% in synthetic and 40% in real data. In addition, MYRON was able to detect interpretable bursty behavior, which we linked to real-world events, when employed to mine the user-photo interactions in the Flickr dataset.

ACKNOWLEDGMENTS

The work was supported by the NSF Smart and Connected Communities (SC&C) grant CMMI-1831547. Research was also supported in part by the National Science Foundation under CAREER grant no. IIS 2046086 and grant no. 1901379. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding parties.

REFERENCES

[1] [n.d.]. Reddit Comments Crawl https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/.

[2] Miguel Araujo, Spiros Papadimitriou, Stephan Guennemann, Christos Faloutsos, Prithwish Basu, Ananthram Swami, Evangelos E. Papalexakis, and Danai Koutra. 2014. Com2: Fast Automatic Discovery of Temporal ('Comet') Communities.

[3] Thomas Aynaud and Jean-Loup Guillaume. 2010. Static community detection algorithms for evolving networks. In *8th International symposium on modeling and optimization in mobile, Ad Hoc, and wireless networks*. IEEE, 513–519.

[4] Brett W Bader and Tamara G Kolda. 2007. Efficient MATLAB computations with sparse and factored tensors. *J. Sci. Comp.* 30, 1 (2007), 205–231.

[5] Petko Bogdanov, Misael Mongiovi, and Ambuj K. Singh. 2011. Mining Heavy Subgraphs in Time-Evolving Networks. In *ICDM*.

[6] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* 3, 1 (Jan. 2011), 1–122. https://doi.org/10.1561/2200000016

[7] Rasmus Bro and Henk AL Kiers. 2003. A new efficient method for determining the number of components in PARAFAC models. *J. Chemom.* 17, 5 (2003), 274–286.

[8] Rasmus Bro, Karin Kjeldahl, Age K Smilde, and HAL Kiers. 2008. Cross-validation of component models: a critical look at current methods. *Analytical and bioanalytical chemistry* 390, 5 (2008), 1241–1251.

[9] Zhengzhang Chen, Kevin A Wilson, Ye Jin, William Hendrix, and Nagiza F Samatova. 2010. Detecting and tracking community dynamics in evolutionary networks. In *2010 IEEE International Conference on Data Mining Workshops*. IEEE, 318–327.

[10] Ingrid Daubechies. 1992. *Ten lectures on wavelets*. Vol. 61. Siam.

[11] Daniel J DiTursi, Gaurav Ghosh, and Petko Bogdanov. 2017. Local Community Detection in Dynamic Networks. In *Proc. ICDM*.

[12] Sofia Fernandes, Hadi Fanaee-T, and João Gama. 2018. Dynamic Graph Summarization: A Tensor Decomposition Approach. *Data Min. Knowl. Discov.* 32, 5 (Sept. 2018), 1397–1420. https://doi.org/10.1007/s10618-018-0583-9

[13] Santo Fortunato and Andrea Lancichinetti. 2009. Community Detection Algorithms: A Comparative Analysis: Invited Presentation, Extended Abstract. In *Proc. ICST VALUETOOLS* (Pisa, Italy). ICST, Article 27, 2 pages. http://dl.acm.org/citation.cfm?id=1698822.1698858

[14] Sabrina Gaito, Matteo Zignani, Gian Paolo Rossi, Alessandra Sala, Xiaohan Zhao, Haitao Zheng, and Ben Y Zhao. 2012. On the bursty evolution of online social networks. In *Proc. of ACM SIGKDD HotSocial*. ACM, 1–8.

[15] Laetitia Gauvin, André Panisson, and Ciro Cattuto. 2014. Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach. *PloS one* 9, 1 (2014), e86028.

[16] K-I Goh and A-L Barabási. 2008. Burstiness and memory in complex systems. *EPL (Europhysics Letters)* 81, 4 (2008), 48002.

[17] Olaf Görlitz, Sergej Sizov, and Steffen Staab. 2008. PINTS: peer-to-peer infrastructure for tagging systems.. In *IPTPS*. 19.

[18] Alexander Gorovits, Ekta Gurjal, Evangelos Papalexakis, and Petko Bogdanov. 2018. LARC: Learning Activity-Regularized Overlapping Communities Across Time. In *Proc. of ACM SIGKDD*.

[19] Mark S Granovetter. 1977. The strength of weak ties. In *Social networks*. Elsevier, 347–367.

[20] Ekta Gujral and Evangelos E Papalexakis. 2018. SMACD: Semi-supervised Multi-Aspect Community Detection. In *"In Proc. of SDM*. SIAM.

[21] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[22] Joyce C Ho, Joydeep Ghosh, and Jimeng Sun. 2014. Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In *Proc. of ACM SIGKDD*. ACM, 115–124.

[23] Kejun Huang, Nicholas D Sidiropoulos, and Athanasios P Liavas. 2016. A flexible and efficient algorithmic framework for constrained matrix and tensor factorization. *IEEE Trans. Signal Process.* 64, 19 (2016), 5052–5065.

[24] Márton Karsai, Kimmo Kaski, Albert-László Barabási, and János Kertész. 2012. Universal features of correlated bursty behaviour. *Scientific reports* 2 (2012), 397.

[25] Jon Kleinberg. 2002. Bursty and Hierarchical Structure in Streams. In *KDD*. https://doi.org/10.1145/775047.775061

[26] Tamara G Kolda and Brett W Bader. 2009. Tensor decompositions and applications. *SIAM review* 51, 3 (2009), 455–500.

[27] Rémi Lemonnier, Kevin Scaman, and Argyris Kalogeratos. 2017. Multivariate Hawkes processes for large-scale inference. In *Proc. AAAI*.

[28] Scott Linderman and Ryan Adams. 2014. Discovering latent network structure in point process data. In *Proc. ICML*. 1413–1421.

[29] Scott W Linderman and Ryan P Adams. 2015. Scalable bayesian inference for excitatory point process networks. *arXiv preprint arXiv:1507.03228* (2015).

[30] X. Ma and D. Dong. 2017. Evolutionary Nonnegative Matrix Factorization Algorithms for Community Detection in Dynamic Networks. *IEEE Trans. on Knowl. and Data Eng.* 29, 5 (May 2017), 1045–1058. https://doi.org/10.1109/TKDE.2017.2657752

[31] Stéphane Mallat. 1999. *A wavelet tour of signal processing*. Academic press.

[32] Stephane G Mallat. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 7 (1989), 674–693.

[33] Maxwell McNeil, Lin Zhang, and Petko Bogdanov. 2021. Temporal Graph Signal Decomposition. In *ACM International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD 2021)*.

[34] Misael Mongiovi, Petko Bogdanov, Razvan Ranca, Ambuj K. Singh, Evangelos Papalexakis, and Christos Faloutsos. 2013. NetSpot: Spotting Significant Anomalous Regions on Dynamic Networks. In *SDM*.

[35] Misael Mongiovi, Petko Bogdanov, and Ambuj K. Singh. 2013. Mining evolving network processes. In *ICDM*.

[36] Seth A Myers and Jure Leskovec. 2014. The bursty dynamics of the twitter information network. In *Proc. ICWWW*. ACM, 913–924.

[37] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. 2007. Quantifying social group evolution. *Nature* 446, 7136 (2007), 664–667.

[38] Evangelos E Papalexakis. 2016. Automatic unsupervised tensor mining with quality assessment. In *Proc. of SDM*. SIAM, 711–719.

[39] Evangelos E Papalexakis, Leman Akoglu, and Dino Ienco. 2013. Do more views of a graph help? community detection and clustering in multi-graphs. In *Proc. of Int. Conf. on FUSION*. IEEE, 899–905.

[40] Evangelos E Papalexakis and Christos Faloutsos. 2015. Fast efficient and scalable core consistency diagnostic for the parafac decomposition for big sparse tensors. In *ICASSP, 2015 IEEE Int. Conf. on*. IEEE, 5441–5445.

[41] Evangelos E Papalexakis, Nicholas D Sidiropoulos, and Rasmus Bro. 2012. From k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors. *IEEE Trans. Signal Process.* 61, 2 (2012), 493–506.

[42] Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. 2017. Introduction to Tensor Decompositions and their Applications in Machine Learning. *CoRR* abs/1711.10781 (2017). http://dblp.uni-trier.de/db/journals/corr/corr1711.html#abs-1711-10781

[43] Giulio Rossetti and Rémy Cazabet. 2018. Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)* 51, 2 (2018), 1–37.

[44] Polina Rozenshtein, Nikolaj Tatti, and Aristides Gionis. 2014. Discovering dynamic communities in interaction networks. In *Proc. ECML PKDD*. Springer, 678–693.

[45] Shaden Smith, Jee W. Choi, Jiajia Li, Richard Vuduc, Jongsoo Park, Xing Liu, and George Karypis. 2017. *FROSTT: The Formidable Repository of Open Sparse Tensors and Tools*. http://frostt.io/

[46] Eric J Stollnitz, Tony D DeRose, Anthony D DeRose, and David H Salesin. 1996. *Wavelets for computer graphics: theory and applications*. Morgan Kaufmann.

[47] Jimeng Sun, Charalampos E Tsourakakis, Evan Hoke, Christos Faloutsos, and Tina Eliassi-Rad. 2008. Two heads better than one: pattern discovery in time-evolving multi-aspect data. *DMKD* 17, 1 (2008), 111–128.

[48] Yorgos Tsitsikas and Evangelos E Papalexakis. 2020. NSVD: Normalized Singular Value Deviation Reveals Number of Latent Factors in Tensor Decomposition. In *Proceedings of the 2020 SIAM International Conference on Data Mining*. SIAM, 667–675.

[49] Hongteng Xu, Dixin Luo, and Lawrence Carin. 2018. Online Continuous-Time Tensor Factorization Based on Pairwise Interactive Point Processes.. In *IJCAI*. 2905–2911.

[50] Hongteng Xu and Hongyuan Zha. 2017. A Dirichlet mixture model of Hawkes processes for event sequence clustering. In *Proc. NIPS*. 1354–1363.

[51] Jaewon Yang and Jure Leskovec. 2012. Community-Affiliation Graph Model for Overlapping Network Community Detection. In *Proc. ICDM (ICDM '12)*. IEEE, 1170–1175. https://doi.org/10.1109/ICDM.2012.139

[52] Junjie Yao, Bin Cui, Yuxin Huang, and Yanhong Zhou. 2010. Detecting bursty events in collaborative tagging systems. In *ICDE*. IEEE, 780–783.

[53] Xiangxiang Zeng, Wen Wang, Cong Chen, and Gary G Yen. 2019. A consensus community-based particle swarm optimization for dynamic community detection. *IEEE transactions on cybernetics* (2019).

[54] Lin Zhang, Alexander Gorovits, and Petko Bogdanov. 2019. PERCeIDs: PERiodic CommunIty Detection. In *Proc. ICDM*. IEEE, 816–825.

[55] Qingyuan Zhao, Murat A. Erdogdu, Hera Y. He, Anand Rajaraman, and Jure Leskovec. 2015. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In *Proc. of ACM SIGKDD* (Sydney, NSW, Australia). ACM, 1513–1522. https://doi.org/10.1145/2783258.2783401

[56] RuiZhen Zhao, XiaoYu Liu, Ching-Chung Li, Robert J Sclabassi, and MinGui Sun. 2009. Wavelet denoising via sparse representation. *Science in China Series F: Information Sciences* 52, 8 (2009), 1371–1377.

[57] Shandian Zhe and Yishuai Du. 2018. Stochastic nonparametric event-tensor decomposition. In *Proc. NIPS*. 6856–6866.

[58] Yunyue Zhu and Dennis Shasha. 2003. Efficient elastic burst detection in data streams. In *Proc. of ACM SIGKDD*. ACM, 336–345.