A Frequency-Structure Decomposition for Link Streams

Esteban BAUTISTA, Matthieu LATAPY
Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
esteban.bautista-ruiz@lip6.fr, Matthieu.Latapy@lip6.fr

Résumé – Un flot de liens est un ensemble de triplets (t, u, v) représentant des interactions dans le temps. Leur analyse est essentielle pour de nombreuses applications. Ils sont généralement étudiés par des approches de traitement du signal et de théorie des graphes, qui permettent d'inspecter leurs propriétés dynamiques et structurelles. Cependant, les techniques actuelles ne permettent pas de lier de manière satisfaisante les informations de structure et fréquence qu'ils contiennent. Nous répondons à cette limitation en proposant une nouvelle décomposition fréquence-structure pour les flots de liens. Notre décomposition analyse la dimension temporelle avec une décomposition classique de traitement du signal, comme Fourier ou les ondelettes, et la dimension structurelle via une nouvelle décomposition pour les graphes que nous avons conçue pour analyser des séquences de graphes. Nous montrons que notre décomposition permet de concevoir naturellement des filtres pouvant récupérer des structures spécifiques avec des fréquences spécifiques.

Abstract – A link stream is a set of triplets (t, u, v) modeling interactions over time and their effective analysis is key for numerous applications. They are traditionally studied via signal processing and graph theory approaches, which allow to study their dynamical and structural properties. However, current techniques do not allow to accurately reveal the frequency-structure patterns contained in them. To overcome this limitation, this work introduces a novel decomposition for link streams. Our decomposition analyses the time dimension via traditional signal dictionaries, like Fourier or wavelets, and the structural dimension via a new decomposition for graphs that we tailored to analyze sequences of graphs. We show that our decomposition allows to naturally design filters that can recover specific structures with specific frequencies.

1 Introduction

Phone calls, financial transactions, or network traffic are examples of data possessing a temporal-relational structure naturally modeled by a link stream [1]: a set of triplets (t, u, v) indicating that u interacted with v at time t. Link streams carry a wealth of information and their proper study is key for numerous applications. Research in link streams has boosted in recent years, with many works leveraging them to achieve state-of-the-art results in various domains [2, 3, 4]. Yet, despite successes, techniques for link stream analysis are not fully satisfactory from the theoretical perspective. In particular, they do not allow to accurately reveal the fundamental frequency-structure patterns contained in them, a situation that we aim to address in this work.

Traditionally, link streams are studied as collections of time series or sequences of graphs, allowing to use signal processing and graph theory to extract frequency and structural patterns from them. These are then combined to find features that reveal temporal-relational patterns in the data (often a set of vertices interacting with some frequency signature). While such feature engineering approach works well in practice, it is not entirely satisfactory from the theoretical perspective: rather than listing all the patterns contained in the data, they amount to heuristically spot if a given one exists. In analogy to signal processing, this is similar to measuring the maximum difference between adjacent samples to spot if a signal contains a high frequency. It is certainly a meaningful feature, albeit in most cases it is more useful to extract the entire frequency information with the Fourier transform. Thus, in this work we aim to find the analogous of the Fourier transform for link streams: a decomposition that reveals all their fundamental frequency-structure patterns.

Existing techniques based on Fourier and spectral decompositions allow extraction of the frequency and structural information from link streams. However they make it hard to relate which structures are associated to which frequencies, limitation arising due to the bad adaptation of graph spectral decompositions for a time-varying context. This work addresses the issue by introducing a new decomposition for graphs that can be easily combined with signal decompositions, allowing to soundly analyze link streams in frequency and structure.

2 Definitions and state-of-the-art

Definitions. Let V denote a set of vertices and $\mathcal{E} = V \times V$ the edge space of size $M = |\mathcal{E}|$. We assume $\mathcal{E}$ to be indexed and $e_k \in \mathcal{E}$ its k-th element. Without loss of generality, we assume $|V|$ to be a power of 2. A discrete time link stream is defined as $L \subseteq \{(t, u, v) : t \in \mathbb{Z}, u \in V, v \in V\}$. We denote $e_i(t)$ the time series of e_i, where $e_i(t) = 1$ if $(t, e_i) \in L$ and 0 otherwise. We denote $G_t(V, E_t)$ the graph at time t, where $E_t = \{(u, v) : (t, u, v) \in L\}$.

State-of-the-art. Link streams are objects that can be studied as signals and graphs. In the signal perspective, the goal is to employ signal processing on the signals $e_k(t)$ to leverage their frequency signature and extract features for events of interest. This is commonly done with tools like the Fourier or Wavelet transforms and with a wide range of filters [2, 5, 6, 7]. In the graph perspective, the goal is to use graph theory on the graphs G_t to leverage their structural signature and extract features for events of interest. This is commonly done by extracting graph statistics or tracking structures, like cliques or communities [8, 9, 10, 4].

Towards the goal of decomposing link stream L into its frequencies and structures, the best current solution consists in (i) extracting all the frequency information of L through a Frequency analysis of the signals $e_i(t)$; and (ii) extracting all the structural information of L through spectral decompositions of the graphs G_t. However, these two approaches are hard to combine in order to assess which structures are associated to which frequencies. We stress that it is important to address this step as phenomena of interest in link streams, like network attacks or financial frauds, are hypothesized as interactions with structure and frequency signatures [1]. The main difficulty in combining the two approaches above lies in the poor adaptation of spectral decompositions to a time-varying setting. Spectral decompositions consider a matrix $\mathbf{S}$ encoding a graph, like the adjacency, and through matrix diagonalization techniques express it as $\mathbf{S} = \sum_k b_k \mathbf{Q}_k$, where $\mathbf{Q}_k$ are rank-1 matrices that capture the whole structural information of $\mathbf{S}$. Even though spectral decompositions have been used in link streams with success [10, 4], they have the theoretical drawback that the matrices $\mathbf{Q}_k$ change for different graphs (as well as the b_k's) and are hard to compare. This makes it very difficult to relate the structures across snapshots and link them to a frequency. Moreover, this limitation also makes it difficult to understand the effect of time-domain operations, like aggregating snapshots or filtering the signals $e_k(t)$, in the structural domain (and vice-versa).

3 Proposed decomposition

We now introduce our decomposition for link streams. We first amend the limitation of spectral decompositions by developing a new graph decomposition that has a fixed basis for all the graphs in the sequence. We then show that this graph decomposition can be easily combined with signal decompositions, allowing us to introduce a decomposition for link streams. Lastly, we show that filters can be easily studied with our framework.

3.1 A new decomposition for graphs

Our goal is to decompose different graphs under one same dictionary. Contrary to common decompositions that rely on matrix factorizations, we interpret all the possible graphs between a set of vertices V as functions supported on the edge-space $\mathcal{E} = V \times V$. This way, we can propose a decomposition for all the functions supported on such edge-space. In more precise terms, we see an unweighted graph $G(V, E)$ as a function $f_G : \mathcal{E} \to \mathbb{R}$, where $f_G(e) = 1$ if $e \in E$ and 0 otherwise. Then, our goal is to develop a meaningful decomposition for f_G. To do this, we notice that interesting structures in graphs can be of different scales. For example, it can be a large group of vertices forming a community, or a small clique confined to the boundaries of the graph. Interestingly, in any of these cases, if such a group is represented by $V_s \subset V$, then the pattern of interest reflects as a dense function supported on $V_s \times V_s$. This means that such function can be very well approximated by a constant function supported on $V_s \times V_s$. Hence, building from this observation, we propose to develop a multi-resolution analysis by piece-wise constant functions of f_G as our decomposition.

In signal processing, such analysis corresponds to the Haar wavelet transform and is performed by dilating and shifting a scaling function. In our case, there is no natural notion of shifting or dilation, yet we can still achieve it by recursively partitioning $\mathcal{E}$. To show this, let us set $\mathcal{E}_0^{(-\log_2(M))} = \mathcal{E}$ and recursively partition this set according to the following rule: $\mathcal{E}_k^{(-(j+1))} = \mathcal{E}_{2k}^{(-j)} \cup \mathcal{E}_{2k+1}^{(-j)}$, with $\mathcal{E}_{2k}^{(-j)} \cap \mathcal{E}_{2k+1}^{(-j)} = \emptyset$ and $|\mathcal{E}_{2k}^{(-j)}| = |\mathcal{E}_{2k+1}^{(-j)}|$, until we obtain singletons $\mathcal{E}_k^{(0)} = e_k$. Then, based on this partitioning, we can define a multi-resolution analysis through the scaling functions defined as:

$$\phi_k^{(-j)}(e) = \begin{cases} \sqrt{2^{-j}} & e \in \mathcal{E}_k^{(-j)} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

And the wavelet functions defined as:

$$\theta_k^{(-j)}(e) = \begin{cases} \sqrt{2^{-j}} & e \in \mathcal{E}_{2k}^{(-j)} \\ -\sqrt{2^{-j}} & e \in \mathcal{E}_{2k+1}^{(-j)} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Scaling and wavelet functions are pair-wise orthonormal by construction. They allow to represent f_G at different levels of resolution indicated by the superscript $(-j)$, where the coarsest one is $(-\log_2(M))$ and the finest one is (0). This is done as follows:

$$f_G(e) = \sum_k s_k^{(-j)} \phi_k^{(-j)}(e) + \sum_{\ell \le j} \sum_k w_k^{(-\ell)} \theta_k^{(-\ell)}(e), \tag{3}$$

where $s_k^{(-j)} = \langle f_G, \phi_k^{(-j)} \rangle$ denotes a scaling coefficient and $w_k^{(-j)} = \langle f_G, \theta_k^{(-j)} \rangle$ a wavelet coefficient. The first term in the right hand side represents the approximation of f_G by piece-wise constant functions supported on the sets $\mathcal{E}_k^{(-j)}$, for all k. The second term contains all the details necessary to recover f_G from its coarse-grain approximation. In particular, the wavelet coefficients and functions of level $(-\ \ell)$ contain all the information necessary to recover the approximation of f_G at resolution level $(-(\ell - 1))$ from the one at level $(-\ell)$. This naturally makes the notion of filtering arise: suppressing the wavelet coefficients and reconstructing back results in the coarse-grain approximation f_G, which, in the graph domain is equivalent to replacing $f_G(e)$ with the average value of f_G in the set $\mathcal{E}_k^{(-\ell)}$ that contains e.

The above derivations show that a multi-resolution analysis of graphs is possible, but the crucial step of how to partition the edge-space is not covered. To address it, we observe that the multi-resolution analysis is more meaningful when the function supported on $\mathcal{E}_k^{(-j)}$ is a constant function. So, the key idea is to find sets $\mathcal{E}_k^{(-j)}$ that contain mostly active or inactive edges. While there can be several ways to search such sets, a natural and adaptive approach consists in looking for rank-1 motifs in the adjacency matrix, which is what matrix diagonalization methods do. We therefore develop a methodology to perform such partitioning by leveraging the second largest singular vector of the adjacency matrix of G. We stress that using a matrix factorization in this situation will not be a problem for link streams, as it will be performed just once and before the link stream is analyzed. Our partitioning operates as follows. We first set $\alpha_0^{(0)} = V$. Then, for each set $\alpha_k^{(j)}$, we take the submatrix (of the adjacency matrix) of rows that are indexed by $\alpha_k^{(j)}$ (we take all columns). We compute the second largest left singular vector of the submatrix and sort it. The elements of $\alpha_k^{(j)}$ associated to the top half entries form the set $\alpha_{2k}^{(j+1)}$ and the remainder $\alpha_{2k+1}^{(j+1)}$. We perform this until the sets $\alpha_k^{(j_{max})}$ are singletons. For space reasons, we omit the entire derivation of how the sets $\mathcal{E}_k^{(-j)}$ are constructed. Our important result is that the above process maps each vertex $u \in \alpha_k^{(j_{max})}$ to a new unique index k. Then, based on this relabelling of vertices, we can define a one-to-one mapping between the partitioned edge-space and the integers on the interval $[1, M]$. This allows us to map edge (u, v), relabelled as (x, y), to a position on the integer line according to the following function:

$$f(x,y) = \begin{cases} k^2 + f(x, y-k) & x \leq k \text{ and } y > k \\ 2k^2 + f(x-k, y) & x > k \text{ and } y \leq k \\ 3k^2 + f(x-k, y-k) & x > k \text{ and } y > k, \end{cases} \tag{4}$$

where k is the previous integer to $max(x, y)$ that is a power of 2. The interesting property of this function is that it maps $\phi_k^{(-j)}(e)$ to the k-th scaling function at resolution $(-j)$ of a discrete Haar time series analysis on the interval $[1, M]$. The implication is that we can compute our graph decomposition using a classical filter-bank for Haar multi-resolution analysis of time-series, which has complexity $O(M)$.

3.2 Link stream decomposition

We now introduce our decomposition for link streams. We follow the standard methodology of extracting frequency using the signal perspective and structures from the graph perspective. The important difference is that we now use our decomposition for graphs. As a first step, we fix a graph dictionary that will be used to analyze the entire sequence. In the absence of extra information, the natural way to do this consists in aggregating the link stream into a single graph. This reveals all the regions of the edge space where the activity occurs and are relevant to track. Based on the aggregated link stream, we can then use our procedure to partition the edge space to fix the basis. As a second step, we represent the link stream in a suitable matrix format where the rows are indexed by time and the columns by edges. Namely, we represent link stream L through the matrix $\mathbf{L}$ so that $\mathbf{L}_{t,j} = 1$ if $(t, e_j) \in L$ and 0 otherwise. Notice that in this format the j-th column of $\mathbf{L}$ corresponds to $e_j(t)$ and the t-th row to $f_{G_t}(e)$.

Based on this rewriting, we can then define the matrix $\mathbf{\Psi} = [\psi_1, \psi_2, \dots]$ where the columns are the atoms of a signal dictionary, like Fourier or wavelets. This allows us to represent the frequency analysis $\mathbf{L}$ as the simple matrix product $\mathbf{L} = \mathbf{\Psi A}$, where $\mathbf{A}$ is a matrix that contains all the frequency information of $\mathbf{L}$ and its entry $\mathbf{A}_{i,j}$ encodes the importance of frequency i in $e_j(t)$. Similarly, we define the matrix

$$\mathbf{\Phi}^\top = [\phi_0^{(-\log_2(M))}, \theta_0^{(-\log_2(M))}, \theta_0^{(-\log_2(M)+1)}, \dots],$$

where the columns are the scaling and wavelet functions of our graph decomposition. This allows us to represent the structural analysis of $\mathbf{L}$ as $\mathbf{L} = \mathbf{B\Phi}$, where $\mathbf{B}$ is a matrix that contains all the structural information of $\mathbf{L}$ and its entry $\mathbf{B}_{i,j}$ encodes the j-th wavelet coefficient of graph G_i.

Our main result is that if we decompose $\mathbf{A}$ into the graph dictionary (to recover the structural information from it), and $\mathbf{B}$ into the signal dictionary (to recover the frequencies from it), we obtain in both cases the same matrix of coefficients $\mathbf{C}$ that encode for the frequency-structure information of $\mathbf{L}$. This is, we can express the link stream as:

$$\mathbf{L} = \mathbf{\Psi C \Phi}. \tag{5}$$

Notice that $\mathbf{C}$ contains all the frequency and structural information about $\mathbf{L}$, and its entry $\mathbf{C}_{i,j}$ quantifies the importance of structure j oscillating at frequency i in the link stream. Eq. (5) constitutes our proposed decomposition for link streams.

3.3 Filters in link streams

Our decomposition in Eq. (5) allow us to track the combined importance of frequencies and structures. A natural application is to then recover interesting frequencies and structures by filtering out the rest. We notice that this can be trivially done by simply suppressing the undesired entries of the matrix $\mathbf{C}$. However, suppressing entries from the matrix arbitrarily is an operation that may be hard to model as linear systems processing $\mathbf{C}$, or represented easily in the $\mathbf{L}$ domain. We now show that we can combine frequency and graph filters to target the recovery of specific frequencies and structures from $\mathbf{C}$, while keeping the benefits of having linear systems processing the stream with interpretations in the $\mathbf{C}$ and $\mathbf{L}$ domains.

Starting with frequency filters, we recall that, in the time domain, the filtering operation can be modeled as the product of a circulant matrix $\mathbf{H}$ (whose columns are shifted versions of the impulse response) with the signal. If such a filter is applied to the signals $e_k(t)$, this operation can be simply expressed as $\widetilde{\mathbf{L}} = \mathbf{HL}$. By decomposing $\mathbf{L}$ and using the fact that circulant matrices are diagonalized by the Fourier basis, we can express the impact of the filter in the $\mathbf{C}$ domain as $\widetilde{\mathbf{L}} = \mathbf{\Psi \Lambda_H C \Phi}$, where $\mathbf{\Lambda_H}$ is a diagonal matrix. Similarly, we can model the impact of structural filters in the $\mathbf{L}$ and $\mathbf{C}$ domains. Recall that

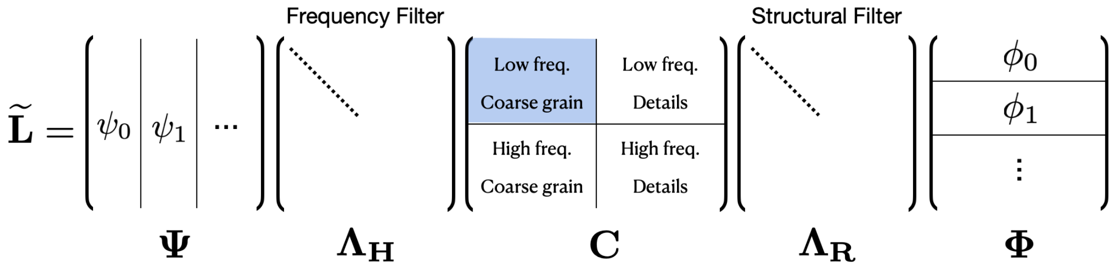

Figure 1: Example of an application of frequency and structural filters to recover information from the link stream. In the example, filters are designed to recover the backbone activity: coarse-grain structures with low frequencies.

a structural filter essentially amounts to replace the graph with its coarse grain approximation (or details). This process can be represented in the edge domain by a matrix $\mathbf{R}$ that multiplies the row vector $f_G(e)$ from the right. If this filter is applied to the graph sequence, we can model it in the $\mathbf{L}$ domain as $\widetilde{\mathbf{L}} = \mathbf{LR}$. Now, by decomposing $\mathbf{L}$ and using the fact that, by construction, $\mathbf{R}$ is diagonalized by $\mathbf{\Phi}$, we obtain its representation in the $\mathbf{C}$ domain as $\widetilde{\mathbf{L}} = \mathbf{\Psi C \Lambda_R \Phi}$, where $\mathbf{\Lambda_R}$ is a diagonal matrix. Thus, the frequency filters suppress the rows of $\mathbf{C}$ while the structural filters suppress its columns. Then, we can combine both approaches as:

$$\widetilde{\mathbf{L}} = \mathbf{\Psi \Lambda_H C \Lambda_R \Phi}$$

in order to recover specific ranges of frequencies and structures, like coarse grain structures slowly oscillating as illustrated in Figure 1.

4 Conclusion

We introduced a new frequency-structure decomposition for link streams. Our decomposition allows to reveal all the fundamental frequency-structure patterns contained in a link stream. Moreover, it offers the possibility to design filters to extract information from it. This offers great potential for better characterizing the signature of real-world events or to design more powerful techniques for feature extraction. The next step consists in leveraging our theoretical insights to address real-world applications.

5 Acknowledgements

This work is funded in part by the ANR (French National Agency of Research) under the Limass (ANR-19-CE23-0010) and FiT LabCom grants.

References

[1] M. Latapy, T. Viard, and C. Magnien, "Stream graphs and link streams for the modeling of interactions over time," *Social Network Analysis and Mining*, vol. 8, no. 1, pp. 1–29, 2018.

[2] S. Bhatia, B. Hooi, M. Yoon, K. Shin, and C. Faloutsos, "Midas: Microcluster-based detector of anomalies in edge streams," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 3242–3249, 2020.

[3] Y.-Y. Chang, P. Li, R. Sosivc, M. H. Afifi, M. Schweighauser, and J. Leskovec, "F-fade: Frequency factorization for anomaly detection in edge streams," *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021.

[4] S. Huang, Y. Hitti, G. Rabusseau, and R. Rabbany, "Laplacian change point detection for dynamic graphs," *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

[5] Y. Abera and D. Hailemariam, "Spatio-temporal mobile data traffic modeling using fourier transform techniques," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 20–24, IEEE, 2018.

[6] R. Fontugne, P. Abry, K. Fukuda, D. Veitch, K. Cho, P. Borgnat, and H. Wendt, "Scaling in internet traffic: A 14 year and 3 day longitudinal study, with multiscale analyses and random projections," *IEEE/ACM Transactions on Networking*, vol. 25, pp. 2152–2165, 2017.

[7] A. Mahmood, M. L. M. Kiah, M. R. Zaba, A. N. Qureshi, M. S. S. Kassim, Z. H. A. Hasan, J. Kakarla, I. S. Amiri, and S. R. Azzuhri, "Capacity and frequency optimization of wireless backhaul network using traffic forecasting," *IEEE Access*, vol. 8, pp. 23264–23276, 2020.

[8] Z.-M. Ren, M. S. Mariani, Y. Zhang, and M. Medo, "A time-respecting null model to explore the structure of growing networks," *ArXiv*, vol. abs/1703.07656, 2017.

[9] P. Moriano, J. Finke, and Y.-Y. Ahn, "Community-based event detection in temporal networks," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[10] T. Wu, C.-S. Chang, and W. Liao, "Tracking network evolution and their applications in structural network analysis," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 562–575, 2018.