Failla *et al. Applied Network Science* (2023) 8:31
https://doi.org/10.1007/s41109-023-00555-6

Applied Network Science

RESEARCH

Open Access

Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions

Andrea Failla[1,2*], Salvatore Citraro[2] and Giulio Rossetti[2]

*Correspondence: andrea.failla@phd.unipi.it

[1] Department of Computer Science, University of Pisa, Pisa, Italy
[2] Institute of Information Science and Technologies "A. Faedo" (ISTI), National Research Council (CNR), Pisa, Italy

Abstract

Recent advances in network science have resulted in two distinct research directions aimed at augmenting and enhancing representations for complex networks. The first direction, that of high-order modeling, aims to focus on connectivity between sets of nodes rather than pairs, whereas the second one, that of feature-rich augmentation, incorporates into a network all those elements that are driven by information which is external to the structure, like node properties or the flow of time. This paper proposes a novel toolbox, that of Attributed Stream Hypergraphs (ASHs), unifying both high-order and feature-rich elements for representing, mining, and analyzing complex networks. Applied to social network analysis, ASHs can characterize complex social phenomena along topological, dynamic and attributive elements. Experiments on real-world face-to-face and online social media interactions highlight that ASHs can easily allow for the analyses, among others, of high-order groups' homophily, nodes' homophily with respect to the hyperedges in which nodes participate, and time-respecting paths between hyperedges.

Keywords: High-order networks, Feature-rich networks, Attributed networks, Stream graphs

Introduction

Complex networks provide a lens through which to illustrate plenty of behaviors that characterize humans as social animals. The elements of graph theory constituted the most helpful toolbox to represent and analyze social networks, with the intention to study complex behavior by mapping any possible kind of human contact, interaction, or relation as pairs of edges between unit elements called nodes. Network science, founded on such a basis, has been able to unravel many social patterns hidden at several scales of human relationships. Global network structures such as rich-clubs (Colizza et al. 2006) and core-periphery structures (Gallagher et al. 2021), together with meso-scale organizations in blocks or communities (Fortunato and Hric 2016), give an idea to the extent to which graphs are useful to grasp the knowledge of complex social architectures. However, the intrinsic nature of graphs to map dyadic patterns does not allow encoding explicitly group connectivity or high-order relations, which are fundamental in the social sphere. An increasing number of works recently started to address the mathematical tools of hypergraph theory (Aksoy et al. 2020) and simplicial complexes (Iacopini et al. 2019; Battiston et al. 2020) to implement multi-body representations of social systems (Torres et al. 2021). Such new lines of hyper-network science aim to point out the importance of high-order interactions when studying the social dynamics of groups (Veldt et al. 2023; Sarker et al. 2023) or nodes embedded in groups rather than within neighborhoods built upon pairwise connections (Failla et al. 2023).

Parallel to this new interest in augmented topologies, other lines of research in network science focus on representations combining the structure with the large amount of domain-specific elements often available from a social system, like people's qualities or preferences, or with any kind of information external to the system that can be related to the structure, e.g., the flow of time that could affect topological changes. Mining such semantically augmented networks helps to unhearth many interesting social properties, from assortative mixing patterns based on common preferences (Newman 2003) to the rules hidden in the formation and evolution of groups (Palla et al. 2007; Rossetti and Cazabet 2018). The term '*feature-rich*' networks (Interdonato et al. 2019) unifies all these augmented implementations that aim to add external, semantic information to a complex structure. Originally designed for pairwise networks, we believe that any complex topology could benefit from a feature-rich implementation, thus also networked representations built upon hypergraphs and simplicial complexes.

Hence, the objective of this work is to address the analysis of high-order patterns together with feature-rich elements. Generalizing the feature-rich framework, we aim to represent and analyze complex social phenomena along the following three dimensions: topology, dynamic features, and node attributes. To this purpose we introduce *ASH*, an *A*ttributed *S*tream-*H*ypernetwork implementation for representing high-order temporal networks with attributive information on nodes.

The rest of the work is organized as follows. Section 2 sums up the principal literature on the three main complex network contexts surrounding this work, namely the dynamic, the node-attributed, and the high-order representations for networks. Section 3 introduces a formalism for the Attributed Stream-Hypergraph, our framework for addressing node-attributed evolving high-order topologies. Section 4 discusses our main results on real-world scenarios, from face-to-face contacts to user interactions on online platforms. Section 5 concludes the work. Finally, in the "Appendix" we introduce a Python library to work with Attributed Stream Hypergraphs.

Related work

In the following, we provide an overview of the main enriched/augmented network implementations that are addressed in the work. First, we discuss dynamic and node-attributed network representations; then, we sum up the emerging contributions about high-order representations for complex systems.

Dynamics of networks. Many network data that represent human activity have an intrinsic dynamic nature, from e-mail exchanges (Klimt and Yang 2004) and financial transactions (Zhao et al. 2018), which are instantaneous forms of connections, to face-to-face interactions, that involve a certain duration, and friendships, that are generally stable and persistent over time. Hence, choosing a proper representation for modeling the dynamics of all these different social behaviors is not a straightforward task. Different temporal semantics impose different representations (Holme and Saramäki 2012; Rossetti and Cazabet 2018), being possible to categorize them according to the following properties: i) stability, e.g., when dynamic data are represented as a snapshot sequence from a time-window aggregation (Ribeiro et al. 2013; Chiappori and Cazabet 2021); ii) duration, e.g., when data are represented as interval graphs (Holme and Saramäki 2012); and iii) immediacy, e.g., when data are represented as a stream graph of temporal nodes and connections (Latapy et al. 2018). In this work, we will mainly focus on such stream graphs, that have been proven to extend and generalize classic centrality measures (Simard et al. 2021), and multi-layer structure as well (Parmentier et al. 2019). More generally, among the most interesting and cutting-edge analyses on dynamic networks, we can mention community detection (Rossetti and Cazabet 2018), link prediction (Divakaran and Mohan 2020), and mixing pattern estimation (Citraro et al. 2022), as well as works extending properties like reciprocity (Chowdhary et al. 2023) and structures like rich-clubs (Pedreschi et al. 2022) to dynamic environments.

Networks with attributes. Attributes or metadata often describe the properties of the nodes involved in networked data. Node attributes can be fruitfully used for improving results on classic network tasks, e.g., in community detection, where both tight connectivity and label homogeneity within communities need to be guaranteed (Chunaev 2020). Attribute-enriched implementations can support analyses on the combined structural and attributive dimensions, searching for possible relations between the properties of nodes and how they are likely to connect (McPherson et al. 2001; Newman 2003). Node attributes can be leveraged for estimating homophily and heterogeneous mixing patterns (Peel et al. 2018; Rossetti et al. 2021). Other tasks oriented to machine learning points of view can leverage on node metadata – e.g., the distribution of values within the adjacent neighborhood of a target node – for node classification and link prediction purposes (Bhagat et al. 2011). There is also an emerging effort toward the exploration of global patterns of connectivity in attributed data, which still is an unexplored topic in the literature on feature-rich networks. Attributed backboning, for instance, is the task of finding the subtree of a graph that spans over the nodes with a minimized connection cost, where such cost is determined by node affinitive attributes (Guan et al. 2019). Similarly, a (k,r)-core structure is a subgraph that is cohesive with respect to both node connectivity and similarity (Zhang et al. 2017).

High-order networks. Although traditional network science mostly addressed pairwise network representations, many dynamics can be better thought of as high-order representations involving relations between groups of nodes. As an emerging line of research (Battiston et al. 2020; Joslyn et al. 2020; Torres et al. 2021), the expressive power of such high-order relations is yet largely unexplored. The interest in the physics of high-order interactions is growing (Battiston et al. 2021), being extensively explored in the area of diffusive processes on networks, e.g., for studying social contagion with simplicial complexes (Iacopini et al. 2019), in time-varying settings as well (Chowdhary et al. 2021). High-order structures varying in time are an important and emerging trend of research (Cencetti et al. 2021; Comrie and Kleinberg 2021). They have been applied to study the network structure of scientific revolutions (Ju et al. 2020), or the evolution of high-order linguistic networks in scientific

Table 1 List of symbols and notation used in the article

Symbol	Description
$\mathcal{S}$	Stream Hypergraph
T	Set of time instants
t	A time instant belonging to T
V	Set of nodes
u	A node u, belonging to V
W	Set of temporal nodes
(t, u)	A temporal node observed at time t, belonging to W
E	Set of temporal hyperedges
N	A subset of nodes
(t, N)	A temporal hyperedge observed at time t, belonging to E
L	Set of node attributes
l	Node attribute value
$l_{(t,u)}$	Attribute value of node u at time t
P	Sequence of hyperedges
T_u	Set of time instants where node u is present
V_t	Set of nodes active at time t
$D(t, u)$	Temporal star of u at time t

texts (Christianson et al. 2020). There is also an increasing interest in the analysis of high-order interactions with attributes, e.g., measures for estimating homophily in hypergraphs and simplicial complexes (Veldt et al. 2023; Sarker et al. 2023), or integrating node attributes through annotated high-order models (Chodrow and Mellor 2020). The high-order structure of static/dynamic networks is often addressed by investigating datasets originally designed for graph-based analysis, thus one of the most intriguing future challenges is the inference of statistically significant high-order interactions from complex systems (Musciotto et al. 2021). Finally, some lines of works tend to be more conservative, as in the case of the s-line graph analysis for hypergraphs (Aksoy et al. 2020), where the hyperedge-projection of the hypergraph is used to apply, for instance, graph-based centrality measures to characterize hyperedges rather than nodes.

Attributed Stream Hypergraphs

To study dynamic high-order social interactions, simply borrowing results from the existing literature is not enough. Hypergraphs and/or simplicial complex has been not adequately defined in the presence of evolving topologies. Moreover, individuals embedded in a social system can often be characterized by multiple features — *profiles* that contextualize some of the key properties playing a role in social interactions (e.g., nationality, gender, age...). In this section we introduce ASH, our Attributed Stream Hypergraph model, adequately defined for evolving high-order interactions with semantically enriched nodes.

Table 1 summarizes the list of symbols and notation used throughout the work. We formally define ASHs as follows:

Definition 1 (ASH) Let $\mathcal{S} = (T, V, W, E, L)$ be a stream hypergraph, where:

- $T = [\mathrm{A}, \Omega]$ is the set of discrete time instants, with A and Ω the initial and final instants, and $t \in T$ identifies a time instant belonging to T;
- V is the set of the nodes of the temporally flattened hypergraph, namely the set of all nodes appearing during the ASH's lifespan;
- $W \subseteq T \times V$ is the set of temporal nodes such that $(t, u) \in W$ identifies a node u observed at time t;
- $E \subseteq T \times V^n$ is the set of temporal hyperedges such that $(t, N) \in E$ implies that $N \subseteq V$ and $\forall u_i \in N, (t, u_i) \in W$;
- $L = \{l_1, ...l_m\}$ is the set of m node attributes such that $l_{(t,u)}$ with $(t, u) \in W$ and $t \in T$, identifies the categorical value of the attribute l associated to u at time t.

ASHs bring together high-order interactions, temporal dynamics, and node attributes. It should be noted that other modeling frameworks can be thought of as particular instances of an ASH, where one of the three dimensions is *switched off*. For instance, given an ASH $\mathcal{S} = (T, V, W, E, L)$, it is possible to switch off a dimension that results in one of the following representations:

- an attributed stream graph (Citraro et al. 2022) for $|N| = 2, \forall (t, N) \in E$, where $|N|$ identifies the number of nodes included in hyperedge $(t, N) \in E$;
- a static node-attributed hypergraph (Veldt et al. 2023) for $|T| = 1$ (i.e., there is no temporal dynamics), which implies $W = V$ and $E \subseteq V^n$;
- a stream hypergraph (without node attributes) for $L = \emptyset$.

Inheriting from stream graphs and hypergraphs

ASHs are a conservative extension of stream graphs (Latapy et al. 2018) and hypergraphs (Battiston et al. 2020; Aksoy et al. 2020), thus inheriting from such frameworks their peculiar concepts. For instance, ASHs inherit from stream graphs the peculiarities of temporal nodes and temporal edges, since the nature of nodes and edges is analyzed with respect to the times they appear in the temporal stream. Nodes/edges can be thought of as temporal entities that can be present or absent at a certain time in the stream, so that the *contribution* of a node/edge is said to be *equal to 1* – i.e., represented as a whole quantity – only if it is present all the time in the stream. With a rapid example, the contribution of an edge uv is computed as follows: $m_{uv} = \frac{|T_{uv}|}{|T|}$, where $|T_{uv}|$ represents the number of time instants where uv is present, and $|T|$ is the overall number of time instants. Naturally, the main difference with stream graphs is that, in an ASH, the temporal presence of an interaction is accounted for hyperedges. This aspect captures the fact that nodes/edges might not be present all the time, thus $|W|$, the sum of active nodes across all temporal instants, and $|T \times V|$, the sum of all possible active nodes across all temporal instants, might differ significantly. The contribution of temporal hyperedges is computed under the same rationale, i.e., the sum of active hyperedges across all temporal instants over the sum of all possible active hyperedges across all temporal instants. Finally, in the case when all nodes are present all the time in the stream, the representation is called *link stream*, and it is a possibility allowed for ASHs as well.

Another key concept that can be generalized to ASHs is that of path. Paths on graphs have already been extended to hypergraphs within the s-analysis framework (Aksoy et al. 2020). This framework builds on the idea that hyperedge paths (or any walk, equivalently) not only have a *length*, i.e., the number of hyperedges crossed during the walk, but also a *width*, i.e., the cardinality of the minimum intersection between subsequent hyperedges. For instance, an s-walk of width 3 (*3-walk*, equivalently) is a sequence of hyperedges where each edge intersects on at least 3 nodes with its predecessor (except for the hyperedge at the beginning) as well as with its successor (except for the hyperedge at the end). However, the dynamic nature of ASHs comes with the added constraint of temporal contiguity. In other words, in a temporal setting, each subsequent hyperedge along an s-walk must come with non-decreasing, adjacent time instants. This also implies that, aside from length and width, a temporal s-walk also has a *duration*, namely the number of time instants occurring between the beginning and the end of the walk. Hence, we define a *time-respecting s-walk* as follows:

Definition 2 (Time-respecting s-walk) A time-respecting s-walk of length k, width s, and duration d is a sequence $P = \{(t_0, N_0), (t_1, N_1), \ldots, (t_{k-1}, N_{k-1})\}$ such that:

- $P \subseteq E$;
- $(t_i, N_i) \in E$;
- $t_i \leq t_{i+1}$ for all is, where $i \in \mathbf{Z}^+ \wedge i < k$ identifies the position of a hyperedge along the walk, with $\mathbf{Z}^+$ identifying the set of positive integers;
- $s \in \mathbf{Z}^+ \wedge s \leq |(N_i) \cap (N_{i+1})|$;
- $d = t_{k-1} - t_0$;

By leveraging the above formulation, the notions of shortest, fastest, fastest-shortest, shortest-fastest, and foremost time-respecting s-walks can be deduced as already done for stream graphs (Latapy et al. 2018), e.g., shortest paths are the ones with minimal length k, fastest paths are the ones with minimal duration $t_k - t_0$, fastest-shortest are the fastest paths among the shortest ones, and shortest-fastest, viceversa; foremost paths, independently from length and duration, are the ones that reach first the destination.

Another concept that can be extended dynamically is that of node's *star*, namely the set of hyperedges where the node is present. This can be limited to include only hyperedges that are active at a specific point in time.

Definition 3 (Temporal Star) Let $u \in V$ be a node in the ASH. The temporal star of u at time t is the set of temporal hyperedges that include u in t, and is denoted $D(t, u) = \{(t, N) : (t, N) \in E \wedge u \in N\}$, with $N \subseteq V$ and $\forall\, u_i \in N, (t, u_i) \in W$.

Temporal star analysis allows quantifying node-level properties of temporal high-order structures (Comrie and Kleinberg 2021), as well as eventually extending to the temporal dimension concepts like hyperego-network density and overlap (Lee et al. 2021).

Towards temporal mixing patterns estimation

Apart from combining the stream graph's evolutionary nature with the hypergraph's high-order structure, ASHs can integrate time-evolving node attributes, i.e., labels that (may) change in time. This peculiarity allows studying not only how individuals' characteristics change (e.g., opinions, political leaning) but also how such changes relate/affect the topological structure surrounding them. As a node's attribute values might vary through time, one can quantify the extent of their *consistency*, namely to what extent a node's attribute value remains constant over time.

Definition 4 (Consistency) Let $u \in V$ be a node, and $l \in L$ be an attribute such that $l_{(t,u)}$ denotes the attribute value of u at time t; let T_u identify the set of time instants where u is present. The *Consistency* of u with regards to l ranges in [0, 1] and is computed as:

$$Consistency(u, l) = 1 - \left(- \sum_{t \in T_u} p(l_{(t,u)}) \log p(l_{(t,u)}) \right). \tag{1}$$

Said differently, consistency corresponds to the complementary of the entropy (cf. next Definition 6) of u's attribute values across time.

Henceforth, we may be interested in quantifying (a) hyperedges' homogeneity, which is a rising hot topic in attributed high-order analyses (Veldt et al. 2023; Sarker et al. 2023), and also (b) the level of homogeneity of a target node with respect to the set of hyperedges it belongs. In case (a), several options are possible, e.g., as cleanly proposed in Veldt et al. (2023), Sarker et al. (2023) with statistically validated measures.

More straightforward ways to measure hyperedges' homogeneity can consist in finding an aggregate value that sums up the *characteristics* of a hyperedge with respect to the labels carried by the nodes within it. For instance, a characteristic value can be the frequency of the most frequent class within a hyperedge. Hence, we can use hyperedges' purity (Citraro and Rossetti 2020) as follows:

Definition 5 (Temporal Purity) Let $(t, N) \in E$ be a temporal hyperedge and $l \in L$ be a node attribute. Let $max_{l \in L}(\sum_{n \in N} l_{(t,n)})$ be the most frequent categorical value within (*t*, *N*). The temporal purity of (*t*, *N*) is the relative frequency of the most frequent value and ranges in $[\frac{1}{|l|}, 1]$, where $|l|$ is the cardinality of the attribute:

$$Purity(t, N, l) = \frac{\max_{l \in L}(\sum_{n \in N} l_{(t,n)})}{|(t, N)|}, \tag{2}$$

Similarly, another characteristic value that can be used to describe a hyperedge is entropy, which quantifies the degree of disorder related to the nodes' attribute values within the hyperedge.

Definition 6 (Entropy) Let $(t, N) \in E$ be a temporal hyperedge and $l \in L$ be a node attribute. Let $A_{(t,N),l}$ be the set of the attribute values of l in (*t*, *N*). The entropy of (*t*, *N*) with respect to l ranges in [0, 1] and is computed as follows:

$$H(t,N,l) = -\sum_{i}^{|A_{(t,N),l}|} p(i)\log p(i) \tag{3}$$

In case (b), our focus is on a target node $u \in V$ aiming to analyze u's homogeneity with respect to its attribute value $l_{(t,u)}$. We can still associate each hyperedge in $D(t, u)$ with a characteristic value. The ones described so far, i.e., purity and entropy, result in continuous values. However, we can also characterize a hyperedge by means of a categorical value. Here, for instance, we describe each hyperedge in the temporal star of a target node u by means of the most frequent attribute value within the hyperedge, namely $max_{l\in L}(\sum_{n\in N} l(t,n))$, with $(t,N) \in D(t,u)$. Having such categorical value can allow us to compute the relative frequency of such characteristic values with respect to the label of the target node.

Definition 7 (Star Homogeneity) Let $u \in V$ be a node with attribute value $l_{(t,u)}$, $l \in L$. Let be $D(t, u)$ the temporal star of u, and $max_{l\in L}(\sum_{n\in N} l_{(t,n)})$, with $(t,N) \in D(t,u)$ the most frequent categorical value of a hyperedge belonging to the star of u. The star homogeneity of u with respect to $l_{t,u}$ is the relative frequency of the hyperedges in $D(t, u)$ that share with u its same attribute value $l_{t,u}$. It ranges in [0, 1] and is defined as follows:

$$Star\ Homogeneity(t,u,l) = \frac{|\{(t,N) \in D(t,u) : max_{l\in L}(\sum_{n\in N} l_{(t,n)}) = l(t,u)\}|}{|D(t,u)|}, \tag{4}$$

Star homogeneity quantifies a node's degree of embeddedness across all of the contexts/interactions it finds itself in.

Finally, consistency, purity, entropy, and star homogeneity, can be averaged to capture the global behavior of the ASH:

$$Consistency(\mathcal{S},l) = \frac{1}{|V|}\sum_{u\in V} Consistency(u,l), \tag{5}$$

$$Purity(\mathcal{S},l) = \frac{1}{|E|}\sum_{(t,N)\in E} Purity(t,N,l). \tag{6}$$

$$H(\mathcal{S},l) = \frac{1}{|E|}\sum_{(t,N)\in E} H(t,N,l) \tag{7}$$

$$Star\ Homogeneity(\mathcal{S},l) = \frac{1}{|V_t|}\sum_{u\in V} Homogeneity(t,u,l). \tag{8}$$

Note that Eq. 5 averages over the number of nodes in the *ASH*, as it quantifies nodes' overall behavior; Eq. 6 and 7 average over the number of hyperedges as they quantify a property of the high-order relation; Eq. 8 averages over $|V_t|$, i.e., the number of nodes that are active at time t, as star homogeneity is relative to a specific point in time.

Experiments

In this section we provide basic experiments to test ASH's potentialities. We define the two following case studies:

- We provide a characterization of face-to-face interactions within the SocioPatterns project,[1] focusing particularly on children in a primary school (Stehlé et al. 2011), and medical staff and patients in a hospital ward (Vanhems et al. 2013);
- We build a case study on online social network discussions about political topics (Morini et al. 2021) to describe aspects related to pairwise-based vs. high-order-based representations.

We promote two different analyses coherently with the different nature/persistence of connectivity in face-to-face contacts and online discussions. In face-to-face interactions we are more interested in analyzing temporal mixing patterns and time-respecting paths with different aggregation windows. In online discussions, where more stability can be reached in users' intreraction, we promote a comparison between pairwise and high-order representations in characterizing users' mixing patterns. The experiments are conducted by leveraging our Python library handling ASH-structured data,[2] which is introduced and discussed in the "Appendix".

High-order temporal dynamics in face-to-face contacts

As mentioned, we analyze the dynamics of children in a primary school and individuals in a hospital ward. Some detailed information is provided as follows:

- *Primary School* (Stehlé et al. 2011): this dataset contains face-to-face interactions between children during the whole school day: node metadata include children's gender and class;
- *Hospital* (Vanhems et al. 2013): this dataset contains the temporal contact data between medical doctors (MED), nurses and paramedics (NUR), administrative staff (ADM), and patients (PAT) in a short-stay geriatric unit of a University hospital. Data were collected for a week.

For representing the temporal higher-order structure, we leverage a similar method as that introduced in Cencetti et al. (2021), namely: if at time t there are $n * (n+1)/2$ dyads between the members of a set of n nodes such that they are involved in a fully connected clique, such links are promoted to form a n-hyperedge. We aim to characterize the hypernetworks with regard to their structure, node features, and dynamics, at different time aggregations (1 min, 5 min, 10 min, 30 min, and 1 h).

[1] www.sociopatterns.org

[2] https://github.com/GiulioRossetti/ASH

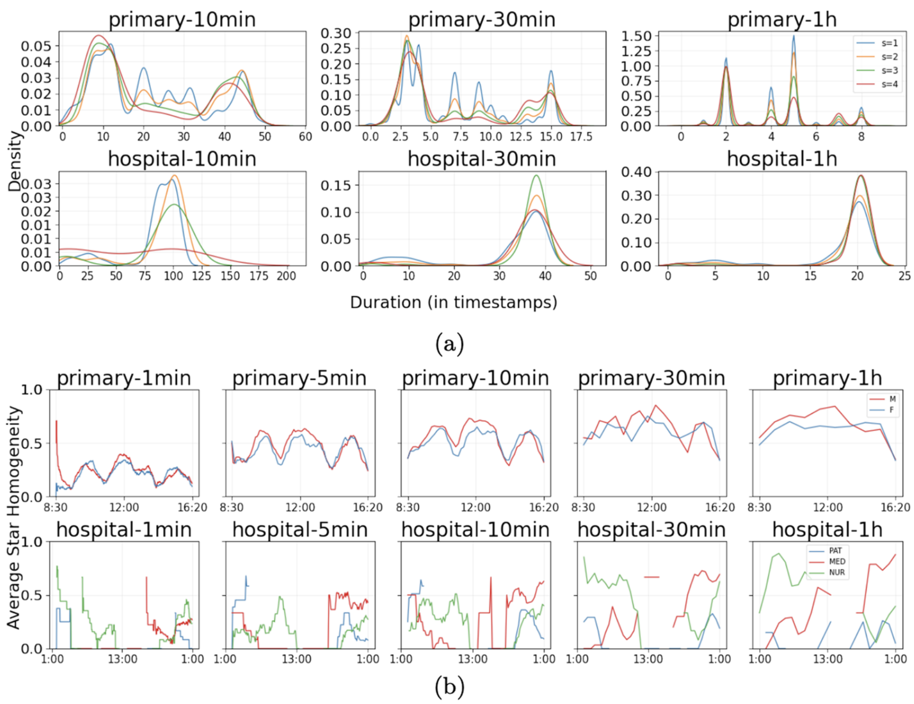

Fig. 1 Duration of shortest s-paths (**a**) and temporal trends of average star homogeneity (**b**) for Primary School and Hospital Ward at different aggregation windows

Time-respecting paths in face-to-face contacts

ASHs allow to define time-respecting paths between incident hyperedges. As mentioned, hypergraph paths add the notion of width, broadening the interest in observing how much this parameter affects the length/duration of the walks. Figure 1a shows the distribution of the duration of shortest s-paths (for s equal to 1, 2, 3, and 4) for both primary school and hospital ward at different aggregation windows. In primary school, we observe heterogeneous distributions with several peaks. Conversely, larger aggregation windows in the hospital ward show that the majority of hyperedges tend to be reached distant in time, intuitively related to the fact that patients can be reached only through the interaction between nurse and medical staff. As expected, for s=1, we observe a larger number of walks, and this difference is more visible when the aggregation window is smaller (10 min). In primary school, larger widths (s>1) highlight a large presence of shortest-lived and longest-lived paths with respect to the entire distribution, whereas in the hospital ward, larger widths let us observe the presence of longer lasting paths than the ones observed with s=1.

Temporal mixing patterns in face-to-face contacts

ASHs allow to study high-order dynamic mixing patterns. In the Sociopatterns datasets we can estimate them at different temporal scales. Figure 1b describes the temporal trends of average star homogeneity (cf. Equation 7) for the two hypernetworks at different aggregation windows. Clear temporal patterns emerge in the primary school dataset, where the attribute observed is children's gender. Children's interactions are more

Table 2 Reddit Data Network statistics (averaged across semesters)

Topic	# nodes	# edges	# Pro-Trump	# Anti-Trump	# Neutral
Gun Control	4991	15298	3346	1645	–
Minorities Discrimination	5540	12605	3318	2222	–
Political Sphere	4509	7079	1280	2395	834

randomly mixed during lesson breaks, and no differences emerge between male and female behavior. In the hospital ward, nurses' interactions are homophilic at late-night/early morning, while MDs' are more homophilic in the evening. The temporal aggregation windows have an impact on both datasets. In primary school, larger windows let more homogeneous interactions emerge. However, larger windows do not let us distinguish between class hours and breaks. Conversely, larger windows in the hospital ward let us observe that some categories are more homophilic than others, e.g., nurses, while patients tend to be disassortative all the time, coherently with the fact that they stay in different rooms (Vanhems et al. 2013) and they are visited only by nurses and medical staff.

Homophilic behaviors in pairwise and group political discussions on Reddit

We focus on data collected from the debate between Trump supporters and anti-Trump citizens during the first two and half years of Donald Trump's presidency, covering a period between January 2017 and July 2019. The debates cover both controversial/polarizing sociopolitical issues and broader discussions within the US political ideologies, as follows:

- *Gun Control*: this topic is identified by collecting lists of subreddits that either support gun legalization or are against it;
- *Minorities Discrimination*: identified by considering groups that promote gender/racial/sexual equality and groups with more conservative attitudes;
- *Political Sphere*: identified by covering different US political ideologies such as Republicans, Democrats, Liberals, and Populists.

Data collection, users' ideology inference, and network construction are properly described in the reference paper (Morini et al. 2021), being able to identify three users' families, pro-trump, antitrump, and neutral classes, that we use as our categorical attribute values. Leveraging the original temporal network,[3] here we infer the hypergraph structure by means of all the maximal cliques. As in the reference analysis, Morini et al. (2021), we consider a time window of six months when analyzing system interactions' dynamics. Average statistics for the pairwise graphs are shown in Table 2.

Analytical setting

We set a four-fold framework to analyze ideological homogeneity from different network-based perspectives as in the following:

[3] Original network data available at https://github.com/virgiiim/EC_Reddit_CaseStudy.

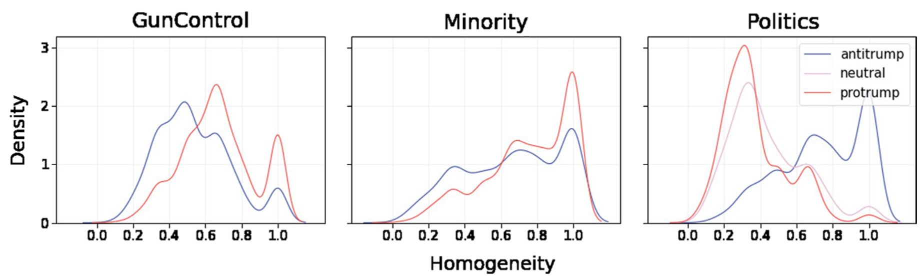

Fig. 2 KDE distributions of pairwise ego-networks' homogeneity among the three different Reddit communities

i. we promote an analysis on dyadic interactions, measuring how much users are homogeneously embedded in their pairwise ego-networks;
ii. we shift the focus from individual users to groups, and we measure the homophily of such groups represented as hyperedges;
iii. we come back to individual users, adopting user's point of view by measuring how much a user is embedded in the hyperedges where he/she participates;
iv. we introduce a time-aware analysis to track stability or variations in ideological homogeneity.

As a preliminary question, we aim to explore whether different behaviors emerge among individual users (i) and groups (ii): can high-order interactions capture patterns that are invisible to dyadic interactions? Then, we aim to understand the role of single users in the several hyperedges where they participate (iii), as a meeting point between the two previous issues: can high-order neighborhoods capture patterns that graph ego-networks cannot? Finally, the focus on interactions' dynamics (iv) would allow us to track stable or mutable patterns as time goes by.

It should be noted that computations in (i) and (iii) are different from (ii). In (i) and (iii) we aim to measure the homogeneity of users' contexts with respect to the political leaning of a specific target node. In (i), a *context* is represented by the set of adjacent nodes in the ego-network of a target node, while in (iii) the context is the set of hyperedges where the target node participates in. We use a measure of homogeneity to estimate target nodes' similarity within nodes' own contexts. We can use Eq. 7 for both pairwise and high-order nodes' ego-network, since in the former case we compute the relative frequency of the attribute values among the node's first-order neighborhood, and in the latter case we use the most frequent value as the characteristic values of a hyperedge. Conversely, in (ii) the focus is on hyperedges' homogeneities. Thus, we use Eq. 2, which computes the relative frequency of the most frequent attribute value within the hyperedge.

Pairwise ego-networks reveal both homophilic and heterophilic users' preferences

Figure 2 outlines graph ego-networks' homogeneities in the three topics considered. For computing the pairwise network homogeneity, we use the following measure:

$$Homogeneity(t,u) = \frac{|v \in \Gamma(t,u) : l(t,u) = l(t,v)|}{|\Gamma(t,u)|}, \quad (9)$$

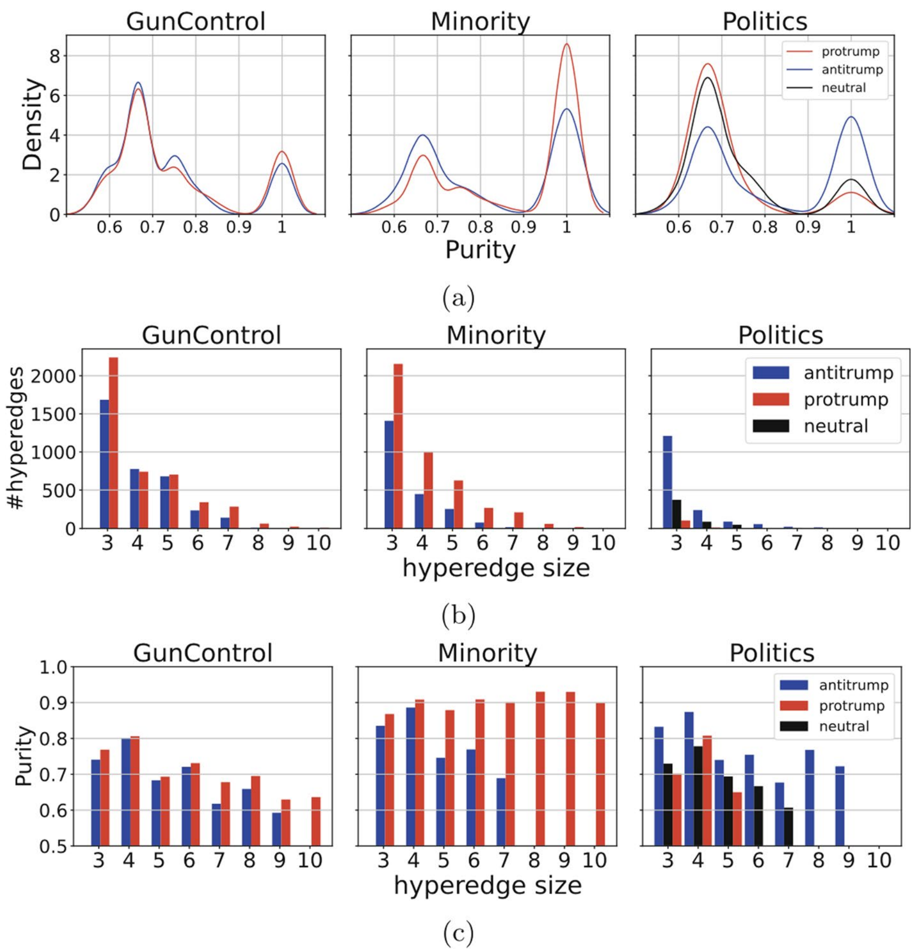

Fig. 3 KDE distributions of hyperedges' purity (**a**), number of pure hyperedges (**b**), and average purity (**c**) in function of hyperedge size among the three different Reddit communities

where $\Gamma(t, u)$ is the set of u's adjacent nodes at time u.

Results are aggregated over the semesters. The analysis of pairwise interactions captures both homophilic and heterophilic patterns, telling us that such political discussions manifest heterogeneity. For instance, in *Politics*, protrump and neutral users show heterophilic behavior, while antritrump are more homogeneous. *Minority* is overall more homophilic than *GunControl*, where interactions seem to be also more randomly mixed.

These observations are coherent with the analyses performed on the original data paper (Morini et al. 2021), where in *Minority* and *Politics* it is more likely to observe echo-chambers – oriented towards a protrump political leaning in *Minority*, and antitrump in *Politics*, while *GunControl* discussions are less polarized.

Hyperedges' purity emphasizes heterogeneity

Fig. 3a shows ideological homogeneity within the hyperedges in the three topics considered, captured by purity. Results are aggregated over the semesters. The discussions in *Minority* are the *purest* ones, a result which is coherent to what already observed at the meso-scale graph-based community level in the original paper (Morini et al. 2021): *GunControl* does not present strongly polarized communities (i.e., echo chambers) among

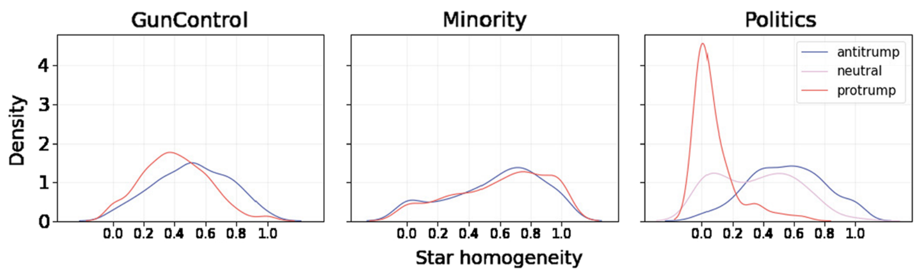

Fig. 4 KDE distributions of hypergraph star ego-networks' homogeneity among the three different Reddit communities

different semesters (Morini et al. 2021) as well as it seems that only a bunch of contexts present quite perfect purity (Fig. 3a, leftmost); in *Minority* on average, more than half of total users are trapped in echo chambers (Morini et al. 2021), and hyperedge purities show a quite similar pattern as well, with a tendency of protrump users to form more homogeneous groups (Fig. 3a, center); also *Politics* presents high homogeneity contexts, where antitrump users are more likely to form homogeneous groups (Fig. 3a, rightmost). Moreover, we analyze these patterns with respect to the hyperedge size. Figure 3b, c highlight, respectively, the number (b) and the average purity (c) of pure groups in function of the group size. For instance, in *Minority* we observe that only protrump pure discussions involve groups with more than 7 participants, and that they are quite pure, 0.9. The same does not happen in *GunControl*, while in *Politics* the biggest contexts involve antitrump users only but with a lower purity than the one of protrump users in *Minority*.

Users are involved in heterogeneous debates

As can be observed in Fig. 4, the topics show diversified behaviors when the analysis shifts to star egos. Indeed, there is no more trace of the heterogeneous patterns observed in Fig. 3a. The key insight, however, relates to another type of heterogeneity in user debates. While engaging in relatively homogeneous contexts (Fig. 3), it seems that users find themselves in rather mixed collections of debates. That is to say, although homophilic behavior is highlighted in most debates (i.e., hyperedges), the set of contexts a node is involved in (i.e., its star) is generally diversified with respect to ideology/political leaning. This is especially true in *GunControl*, where protrump users appear to engage in a more heterogeneous set of debates than their counterparts, as opposed to what was noted in Fig. 3a. The same holds for *Politics*, which displays a peak in heterogeneous protrump stars while the antitrump ones show more homophilic behavior. *Minority*, instead, still shows strong homogeneity traits for both antitrump and protrump users, thus confirming previous observations.

Interactions' dynamics: users' preferences tend to be consistent in time

As far as the temporal dimension is concerned, a certain degree of consistency w.r.t. debates homogeneity/heterogeneity can be observed. As a matter of fact, the average star homogeneity outlines almost-flat trends (Fig. 5), indicating minor variations. Here, *GunControl* and *Minority* reveal near-constant heterogeneity/homogeneity for both political alignments; lastly, *Politics* displays only a small bump during the third semester

Table 3 Reddit Data Network statistics (averaged across semesters)

Topic	#stay	#stay_pct	mean_consistency (stay)	std_ consistency (stay)
GunControl	580	0.116	0.591	0.478
Minority	735	0.132	0.715	0.441
Politics	574	0.127	0.748	0.310

Nodes that stay in the network for more than two semesters/timestamps; mean and std of the consistency values for such nodes

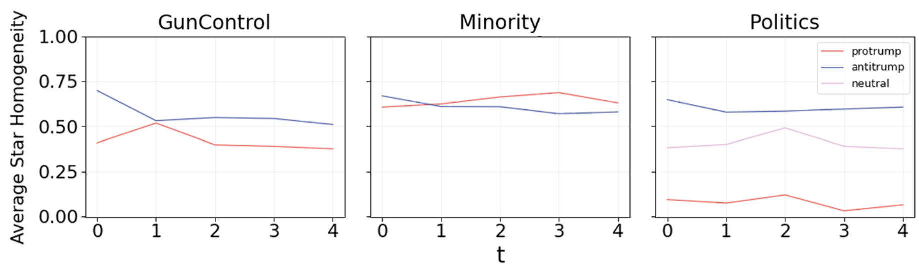

Fig. 5 Average hypergraph star ego-networks' homogeneity over time among the three different Reddit communities

concerning protrump and neutral users. What makes this result so interesting is the fact that only $\sim 11\%$ of the nodes stay in the network for more than a semester (see Table 3), but still coherence is observed regardless of the continuous turnover of nodes. Such coherence is also confirmed by the *Consistency* values of the remaining nodes (Table 3, fourth column), which hint at a resilience to opinion change.

Discussion and conclusion

In this work, we proposed an Attributed Stream Hypergraph (ASH) representation for taking into account both high-order relations (Battiston et al. 2020) and feature-rich information (Interdonato et al. 2019) through which to describe complex social systems. With ASH, social phenomena represented by means of high-order interactions can also be studied together with additional information that goes beyond the network structure, namely nodes' semantics and time. We have shown how this paradigm can be used to analyze social interactions along the (i) structural, (ii) attributive, and (iii) temporal dimensions. The high-order architecture inherited by hypergraphs (i) allows to more realistically model social interactions which naturally occur in groups of varying sizes. Node metadata can be used to construct node profiles (ii), which can be used to assess differences and similarities in the behaviors of different classes. The temporal dimension (iii) can shed light on recurring patterns over time.

The novelty of ASH stands in the possibility to combine all these aspects together. Being interested in analyzing a wide variety of different temporal network data, we focused here on face-to-face contacts and discussions on online platforms. In the experiments, we recognized the significance of capturing the temporal dynamics of both datasets, by building different case studies according to the different nature of the data analyzed. Face-to-face contacts, like the well-known ones from SocioPatterns, were suitable for a broad characterization involving the enriched expressivity provided by our augmented representation. In particular, we observed whether different instantiations of the parametrized dimensions inner in our model – like the width of hyperedges for computing time-respecting s-walks – could lead to different results, thus interpretations. Time-respecting s-walks can allow to identify stable, densely-connected subhypergraphs, and to generalize dynamic centrality scores to hypergraphs (Simard et al. 2021). In face-to-face contacts we noticed that time-respecting s-walks and their duration can vary according to the width of hyperedges. This observation hints to carefully consider the importance of this parameter for further studies involving time-respecting paths in high-order networks, e.g., for assortative mixing studies (Citraro et al. 2022) or information diffusion models (Antelmi et al. 2021).

While applying the ASH framework on US-politics-bound communities on Reddit we observed strong homophilic behaviors among groups/hyperedges with respect to users' political leaning. However, while focusing on the preferences of single nodes, namely on how much a target node is homogeneously embedded with respect to the representative political leaning of the groups/hyperedges it belongs to, we mostly observe a relevant decrease in nodes' homophilic behaviors. As a consequence, we observe that users prefer participating in contexts whose representative leaning is different than the target node's own label, although hyperedges are strongly homophilic per se. When studying enriched high-order representations such as ASHs, an interesting point to investigate is the difference between ASHs and their particular instances when one of the dimensions is *switched off*, for instance the high-order dimension. Interestingly, the previously described patterns are not observed when looking at the pairwise ego-networks only, indicating the different expressive power provided by the enriched model.

In future works, we plan to focus on the constraints that stream hypergraphs could eventually raise, such as the issues of under/overfitting social data or the robustness of the measures to missing data. Our findings highlight how different temporal aggregations and graph vs. hypergraph representations deeply affect the output of analytical pipelines. Thus, some of the most interesting challenges in the future will be understanding the impact of different representations (e.g., graphs vs. hypergraphs), of high-order structure inference methods (e.g., via cliques (Cencetti et al. 2021), overlapping communities, or other statistical methods Contisciani et al. 2022), and of different measures to study mixing behaviors. We also plan to introduce synthetic ASH generators to be used in the validation of analysis results.

Lastly, we plan to update and maintain the `ASH` Python library, hoping it will simplify and make more accessible to researchers and practitioners feature-rich hypernetwork analysis.

Appendix: `ASH`: a library for Attributed Stream Hypergraphs

Despite the ever-growing interest in the analysis of high-order topologies, very few software packages are available to work with hypernetworked data. Nonetheless, most of them entirely discard temporal information – restricting analyses to the static scenario – and lack analytical tools to study attributive dynamics. To address these issues, we developed `ASH`, a Python library allowing to easily handle multiadic data while retaining temporal and attributive information. In this section, we introduce the library rationale and describe some of the main features.

Classes The library's core lies in a homonymous class that offers basic functionalities related to hypergraph building, statistics, temporal information, and hypergraph transformations. Nodes and hyperedges are assigned a unique identifier at creation (integers for nodes, strings for hyperedges), as well as initial and final temporal ids (both integers) which identify the presence of a node/hyperedge between those points in time. These allow to successively retrieve information about nodes and edges through time in an efficient way.

Node metadata is appropriately enclosed in a separate class that represents *node profiles*. Indeed, since nodes can have multiple attributes (i.e., inherent features) as well as other statistics (e.g., centrality scores), it is useful to enclose them in a different structure, also to handle updates and comparisons.

(Dynamic) Hypergraph measures and trasformations The library offers a variety of methods to compute basic node and hyperedge statistics. Node neighborhoods, hyperedge distributions and such can be computed both on the flattened (i.e., static, aggregated) hypergraph, or by only including hyperedges active during a specific time period. Dynamic network measures were generalized to hypergraphs, such as node and hyperedge contribution and uniformity (Latapy et al. 2018). The package provides several hypergraph transformations such as bipartite projection, dual hypergraph, and s-line graph, as well as hypergraph decomposition to graphs via clique expansion. ASHs can also be sliced both structurally (i.e., induced sub-hypergraph) and/or temporally (i.e., hypergraph temporal slice).

Paths, distances, and centralities. ASH provides full support for the s-analysis framework (Aksoy et al. 2020), which is used to generalize classic graph measures to hypergraphs. This allows to compute paths, distances, connected components, clustering, as well as several centrality measures, and extend them along the temporal dimension. Time-respecting s-walks can be measured both in terms of length, weight, and duration.

Attribute Analysis We provide several measures to quantify mixing behaviors on dynamic hypergraphs, including but not limited to those introduced in this work. These quantities can characterize both nodes (e.g., measures that take into account a specific node and its surroundings) and hyperedges (e.g., measures quantifying homophilic behaviors in high interactions).

Visual Analytics The ASH library includes a dedicated module to facilitate the visualization of measures, node degree distribution, hyperedge size distribution, as well as time series representing the structural and attributive characteristics of the ASH through time.

I/O Finally, input/output facilities are handled by a dedicated module. In detail, it is possible to read/write node profiles from/to.csv and.json files, read/write interactions from/to csv, and also read/write the whole `ASH` from/to.json files.

All in all, the library is efficient and scales well to large hypergraphs, especially considering the complexity of the underlying model and the layers of information it operates on (nodes, relations, temporal information, and node profiles). Such performance is primarily due to its mapping systems, namely a node-to-edge and node-to-star mappings, providing for $\mathcal{O}(1)$ access to hyperedge and node stars respectively, as well as an `ash`-native time-to-edge mapping that permits hyperedge lookups by temporal id. `ASH` is built on top of two core libraries, `halp` and `DynetX`, to achieve these functionalities. Most of the hypergraph-related functionalities build on top of the `UndirectedHypergraph` class from the `halp` library, providing an efficient implementation and a wide range of utilities. The temporal dimension is handled by `DynetX`, a library for dynamic network analysis that provides support for temporal networks, allowing modeling the evolution of networks over time. To date, `ASH` is the only comprehensive software package to efficiently handle hypergraph-structured data enriched with node attributes. In fact, its competitors either do not scale well to large hypergraphs or primarily focus on other structures (e.g., directed hypergraphs). Moreover, other libraries can only model static systems (i.e., there is no support for temporally-aware data), and lack measures and algorithms to study attribute-dependent wiring patterns. Our library takes the best of both worlds by integrating the s-analysis framework with node attribute analytics, and adding temporal support on top.

Abbreviations

ASH(s) Attributed Stream Hypergraph(s)
KDE Kernel Density Estimation
w.r.t. With respect to

Acknowledgements

This work is supported by the European Union - Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 - Integrating Activities for Advanced Communities", Grant Agreement no. 871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu).

Author contributions

AF, SC, and GR designed research, performed research, and wrote the paper. AF and SC contributed to the acquisition, analysis, and interpretation of the experiments. AF and GR contributed to the design of the software package rationale and its implementation. GR coordinated and supervised all of the research. All authors read and approved the final manuscript.

Availability of data and materials

The datasets of face-to-face interactions are available on the Sociopatterns website. www.sociopatterns.org. The Reddit data is available on GitHub https://github.com/virgiiim/EC_Reddit_CaseStudy.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 31 March 2023 Accepted: 28 May 2023
Published online: 09 June 2023

References

Aksoy SG, Joslyn C, Marrero CO, Praggastis B, Purvine E (2020) Hypernetwork science via high-order hypergraph walks. EPJ Data Sci 9(1):16

Antelmi A, Cordasco G, Spagnuolo C, Szufel P (2021) Social influence maximization in hypergraphs. Entropy 23(7):796

Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G, Petri G (2020) Networks beyond pairwise interactions: structure and dynamics. Phys Rep 874:1–92

Battiston F, Amico E, Barrat A, Bianconi G, Ferraz de Arruda G, Franceschiello B, Iacopini I, Kéfi S, Latora V, Moreno Y, et al. (2021) The physics of higher-order interactions in complex systems. Nat Phys 17(10):1093–1098

Bhagat S, Cormode G, Muthukrishnan S (2011) Node classification in social networks. In: Social network data analytics, pp 115–148. Springer

Cencetti G, Battiston F, Lepri B, Karsai M (2021) Temporal properties of higher-order interactions in social networks. Sci Rep 11(1):1–10

Chiappori A, Cazabet R (2021) Quantitative evaluation of snapshot graphs for the analysis of temporal networks. In: International conference on complex networks and their applications, pp 566–577. Springer

Chodrow P, Mellor A (2020) Annotated hypergraphs: models and applications. Appl Netw Sci 5(1):1–25

Chowdhary S, Andres E, Manna A, Blagojević L, Di Gaetano L, Iñiguez G (2023) Temporal patterns of reciprocity in communication networks. EPJ Data Sci 12(1):7

Chowdhary S, Kumar A, Cencetti G, Iacopini I, Battiston F (2021) Simplicial contagion in temporal higher-order networks. J Phys Complex 2(3):035019
Christianson NH, Sizemore Blevins A, Bassett DS (2020) Architecture and evolution of semantic networks in mathematics texts. Proc R Soc A 476(2239):20190741
Chunaev P (2020) Community detection in node-attributed social networks: a survey. Comput Sci Rev 37:100286
Citraro S, Rossetti G (2020) Identifying and exploiting homogeneous communities in labeled networks. Appl Netw Sci 5(1):1–20
Citraro S, Milli L, Cazabet R, Rossetti G (2022) {\Delta}-conformity: multi-scale node assortativity in feature-rich stream graphs. Int J Data Sci Anal 1–12
Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. Nat Phys 2(2):110–115
Comrie C, Kleinberg J (2021) Hypergraph ego-networks and their temporal evolution. In: 2021 IEEE international conference on data mining (ICDM), pp 91–100. IEEE
Contisciani M, Battiston F, De Bacco C (2022) Inference of hyperedges and overlapping communities in hypergraphs. Nat Commun 13(1):7229
Divakaran A, Mohan A (2020) Temporal link prediction: a survey. N Gener Comput 38(1):213–258
Failla A, Citraro S, Rossetti G (2023) Attributed stream-hypernetwork analysis: homophilic behaviors in pairwise and group political discussions on reddit. In: Complex networks and their applications XI: proceedings of the eleventh international conference on complex networks and their applications: complex networks 2022—Volume 1, pp 150–161. Springer
Fortunato S, Hric D (2016) Community detection in networks: a user guide. Phys Rep 659:1–44
Gallagher RJ, Young J-G, Welles BF (2021) A clarified typology of core-periphery structure in networks. Sci Adv 7(12):9800
Guan S, Ma H, Wu Y (2019) Attribute-driven backbone discovery. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 187–195
Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519(3):97–125
Iacopini I, Petri G, Barrat A, Latora V (2019) Simplicial models of social contagion. Nat Commun 10(1):1–9
Interdonato R, Atzmueller M, Gaito S, Kanawati R, Largeron C, Sala A (2019) Feature-rich networks: going beyond complex network topologies. Appl Netw Sci 4(1):1–13
Joslyn CA, Aksoy SG, Callahan TJ, Hunter LE, Jefferson B, Praggastis B, Purvine E, Tripodi IJ (2020) Hypernetwork science: from multidimensional networks to computational topology. In: International conference on complex systems, pp 377–392. Springer
Ju H, Zhou D, Blevins AS, Lydon-Staley DM, Kaplan J, Tuma JR, Bassett DS (2020) The network structure of scientific revolutions. arXiv preprint arXiv:2010.08381
Klimt B, Yang Y (2004) Introducing the enron corpus. In: CEAS
Latapy M, Viard T, Magnien C (2018) Stream graphs and link streams for the modeling of interactions over time. Soc Netw Anal Min 8(1):1–29
Lee G, Choe M, Shin K (2021) How do hyperedges overlap in real-world hypergraphs?-patterns, measures, and generators. In: Proceedings of the web conference 2021, pp 3396–3407
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Sociol 27(1):415–444
Morini V, Pollacci L, Rossetti G (2021) Toward a standard approach for echo chamber detection: Reddit case study. Appl Sci 11(12):5390
Musciotto F, Battiston F, Mantegna RN (2021) Detecting informative higher-order interactions in statistically validated hypergraphs. Commun Phys 4(1):1–9
Newman ME (2003) Mixing patterns in networks. Phys Rev E 67(2):026126
Palla G, Barabási A-L, Vicsek T (2007) Quantifying social group evolution. Nature 446(7136):664–667
Parmentier P, Viard T, Renoust B, Baffier J-F (2019) Introducing multilayer stream graphs and layer centralities. In: International conference on complex networks and their applications, pp 684–696. Springer
Pedreschi N, Battaglia D, Barrat A (2022) The temporal rich club phenomenon. Nat Phys 18(8):931–938
Peel L, Delvenne J-C, Lambiotte R (2018) Multiscale mixing patterns in networks. Proc Natl Acad Sci 115(16):4057–4062
Ribeiro B, Perra N, Baronchelli A (2013) Quantifying the effect of temporal resolution on time-varying networks. Sci Rep 3(1):1–5
Rossetti G, Cazabet R (2018) Community discovery in dynamic networks: a survey. ACM Comput Surv (CSUR) 51(2):1–37
Rossetti G, Citraro S, Milli L (2021) Conformity: a path-aware homophily measure for node-attributed networks. IEEE Intell Syst 36(1):25–34
Sarker A, Northrup N, Jadbabaie A (2023) Generalizing homophily to simplicial complexes. In: Complex networks and their applications XI: proceedings of the eleventh international conference on complex networks and their applications: complex networks 2022—Volume 2, pp 311–323. Springer
Simard F, Magnien C, Latapy M (2021) Computing betweenness centrality in link streams. arXiv preprint arXiv:2102.06543
Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J-F, Quaggiotto M, Van den Broeck W, Régis C, Lina B et al (2011) High-resolution measurements of face-to-face contact patterns in a primary school. PLoS ONE 6(8):23176
Torres L, Blevins AS, Bassett D, Eliassi-Rad T (2021) The why, how, and when of representations for complex systems. SIAM Rev 63(3):435–485
Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, Kim B-A, Comte B, Voirin N (2013) Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PLoS ONE 8(9):73970
Veldt N, Benson AR, Kleinberg J (2023) Combinatorial characterizations and impossibilities for higher-order homophily. Sci Adv 9(1):3200
Zhang F, Zhang Y, Qin L, Zhang W, Lin X (2017) When engagement meets similarity: efficient (k, r)-core computation on social networks. Proc VLDB Endow 10(10):998–1009. https://doi.org/10.14778/3115404.3115406
Zhao L, Wang G-J, Wang M, Bao W, Li W, Stanley HE (2018) Stock market as temporal network. Phys A 506:1104–1112