# Attributed Stream-Hypernetwork Analysis: a SocioPatterns Case Study

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

[1]Computer Science Dept., University of Pisa, IT
[2]KDD Lab, ISTI, National Research Council, IT

**Abstract**

Network science comes as a solid framework to describe a multitude of human behaviours. Face-to-face human interactions, for instance, are often represented by dynamic networks involving time-varying links. Such temporal models are shown to be effective as proxies for real communications between individuals. However, networks are intrinsically bounded to pairwise/dyadic connections, whereas complex human dynamics can naturally involve higher-order organization, namely relationships between groups of entities. In the last few years, hypergraph and simplicial complex models have been addressed as promising tools to better understand the dynamics of social groups. In the analysis of face-to-face interactions, the higher-order organization of temporal networks has been addressed by investigating collections of datasets initially designed for graph-based analysis. Yet even these higher-order representations continue to ignore the rich attributes or metadata often carried by the nodes. Such attributes can offer new interesting perspectives about the dynamics of the higher-order structure emerging from a stream of social interactions. In this work, we aim to address this gap by introducing attributed stream-hypernetwork models, i.e., higher-order temporal networks with attributive information on nodes. Considering the Primary and High School temporal networks from the well-known SocioPatterns project, we infer the higher-order temporal structure of interactions between children and high school students, and we characterize their non-trivial relationships with respect to their gender attribute.

**Keywords**

High-order Networks, Dynamic Networks, Feature-rich Networks, Data Analysis



## 1. Introduction

The evolving nature of human interactions is often approached through the lens of dynamic network analysis [1]. Dynamic networks indeed, built on top of graph theory tools, can represent a wide set of time-evolving human interactions. Dynamic network analysis, a fruitful subfield of network science, can shed lights on the complex laws governing such time-evolving connections as well as their duration. However, the intrinsic nature of dynamic network interactions can not explicitly go beyond dyadic patterns between pairs of nodes. Such network-based constraints must be taken into account if we aim to investigate human interactions from the point of view of groups instead of nodes. Hypernetwork science is the new, cutting-edge line

*SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy*
a.failla@studenti.unipi.it (A. Failla); salvatore.citraro@phd.unipi.it (S. Citraro); giulio.rossetti@isti.cnr.it (G. Rossetti)
http://giuliorossetti.net (G. Rossetti)

of research emerging in network science community that addresses such higher-order structures for representing complex systems. Thus, dynamic features and higher-order representations are the two main ingredients for investigating the complex nature of dynamic systems where time-varying group relations are involved, e.g., face-to-face and physical proximity interactions. Finally, a last ingredient we may want to consider is the rich set of attributes that nodes can carry, with the purpose of analyzing how node-attribute information distributes across groups and the higher-order contexts they define.
In this work we introduce *ASH*, an Attributed Stream-Hypernetwork model able to represent higher-order temporal networks with attributive information on nodes. To test the potentiality of this new model, we infer the attribute-rich higher-order temporal structure of contact patterns between children and high school students from the well-known SocioPatterns project collections.

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 dynamic networks, higher-order structures and attributed networks. Section 3 introduces our Attributed Stream-Hypernetwork model and discusses the main results obtained from the two SocioPatterns datasets in terms of topological, node-features and interactions' dynamics analyses. Finally, Section 4 concludes the work and discusses promising lines of research left open for the future.

## 2. Related work

In the following we will provide a brief overview of the main enriched network models and higher-order representations for complex systems used throughout the work. Dynamic and node-attributed graphs are the feature-rich [2] representations on which we will focus primarily; then, we will sum up the main emerging contributions in the analysis of higher-order interactions in complex networks.

**Dynamics of networks**. Nowadays temporal information is more and more available from networks. However, choosing a proper temporal representation is not straightforward. Friendships are persistent over time, whereas face-to-face interactions involve a certain duration, and e-mails, messages or financial transactions are even instantaneous. Different temporal semantics imposes different modelling strategies [1]. Among the most suitable models encoding dynamic features in networks, we can identify those focusing on i) stability, ii) duration, and (iii) immediacy, where:

i a dynamic network is represented as a sequence of autonomous and independent graphs;
ii a set of intervals preserve the time-varying dynamics of connections whose a dynamic network is composed;
iii a dynamic network comes as a stream of interactions over time – e.g., the *stream graph* [3] is emerging as one of the most fruitful models able to model both instantaneous and lasting links, and to capture the presence/absence of nodes as well.

Among the most interesting and cutting-edge dynamic network mining tasks we recall community detection [4], link prediction [5], and mixing pattern estimation [6].

**Networks with metadata**. Similarly to temporal information, metadata or attributes describing the properties and the characteristics of nodes are often available from network data. Such metadata-enriched models can support new mining and analyses about the relationships between the structural and the attributive information inferred from complex systems. Generally speaking, works focusing on the relations between attributes and structure aim to study their correlation or influence, searching for bridges between tabular and networked data [7].

Node attributes can be fruitfully used for improving the community detection task, where both tight internal connectivity and label-homogeneity within communities need to be guaranteed [8], and for estimating heterogeneous mixing patterns complex networks [9, 10]. The distribution of metadata surrounding a single node, e.g., the distribution of features within the node's ego-network, can be used in the node classification and in the link prediction tasks [11].

**Higher-order networks**. Although traditional analyses addressed mostly pairwise interactions, many network dynamics can be better modelled by higher-order system representations involving complex interactions between groups of nodes. Being an emerging line of research [12, 13], the expressive power of such models describing higher-order interactions in complex systems is yet largely unexplored. The interest in the physics of higher-order interactions is growing [14], and it has been mainly explored in the context of diffusion analysis, e.g., studying social contagion with simplicial complexes, [15], as well as in time-varying settings [16]. Higher-order structures varying in time are an important and emerging trend of research [17].

In most of such analyses, the higher-order organization of static/dynamic networks is addressed by investing datasets originally designed for graph-based analysis. One of the most intriguing and fundamental works in the future will have to focus on the inference of statistically significant higher-order interactions [18]. Moreover, higher-order-based techniques are emerging to generalize well-known graph-based techniques or to conservatively shift to them, as in the case of s-line graph analysis for hypergraph models [19].

## 3. Hypernetwork Science

To study dynamic high-order social interactions, simply borrowing results from the existing literature is not enough. Hypergraphs and simplicial complexes, to name the nowadays most used high-order representation frameworks, have both strengths and weaknesses. None of them has been adequately defined in the presence of evolving topologies. Indeed, their applicability to online social environments needs to be carefully analyzed to understand if the constraint they come with aligns with the semantics expressed by social interaction networks. 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 for social interactions (e.g., nationality, gender, age…). In order to start filling the existing gap in high-order dynamic and feature-rich modeling of social systems, here we propose the framework of Attributed Stream Hypergraphs (henceforth, ASH).

**Definition 3.1** (Attributed Stream Hypergraph (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* $\mathrm{A}$ *and* $\Omega$ *the initial and final instants;*
- $V$ *is the set of the nodes of the temporally flattened hypergraph;*
- $W \subseteq T \times V$ *is the set of temporal nodes;*
- $E \subseteq T \times V^n$ *is the set of temporal hyperedges such that* $(t, N) \in E$ *implies that* $N \subseteq V$ *and* $\forall u_i \in N, (t, u_i) \in W$*;*
- $L$ *is the set of temporal node attributes such that* $L(t, u)$ *with* $(t, u) \in W$ *and* $t \in T$*, identifies the set of categorical values associated to* $u$ *at time* $t$*.*

ASHs are a conservative extension of well known modeling frameworks (namely, Hypergraphs [12] and Stream Graphs [3]). From them it inherits several analytical peculiarities and, from their union, it is able to provide novel insights that the original models are not able to unveil independently. Moreover, integrating time evolving node attributes, it allows to study not only how individuals' characteristics changes (e.g., opinions, political leaning) but also how such changes relate/affect the topological structure surrounding them.

In the following we provide a first example of how ASH can be used to study real world complex social systems. Our analysis will focus on some aspects of the three dimensions modeled by ASH: high-order topology, node semantics and time.

### 3.1. SocioPatterns data

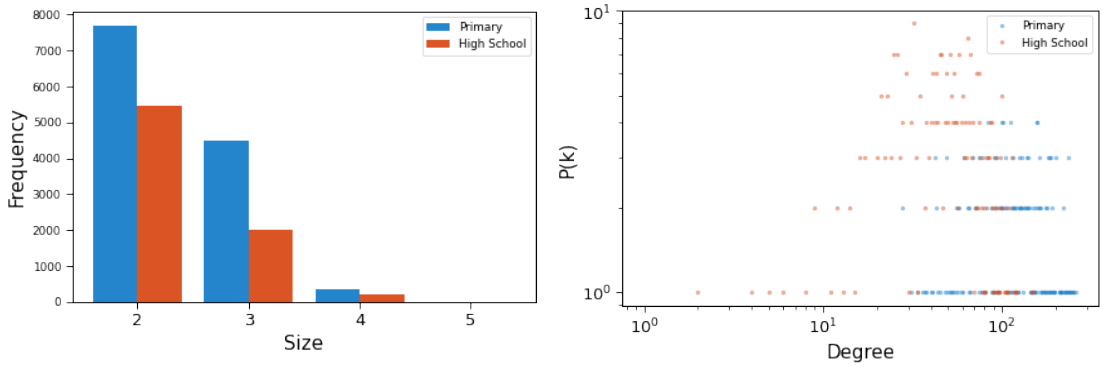
SocioPatterns[1] is a project collecting a variety of physical proximity and face-to-face interactions across several environments, e.g., hospitals, workplaces and schools. These data were classically collected to study human behaviours in terms of temporal interactions, confirming, for instance, the existence of few long-lasting contacts and a multitude of brief contacts while analyzing their duration in spontaneous human interactions [20], or observing that the contacts are shaped by the organization of the offices in workplaces [21]. The face-to-face data are also extremely useful to estimate the transmission of infectious diseases in contexts like schools [22] and workplaces [21]. Recently, such data have been studied by applying an higher-order temporal perspective [17], or by enhancing the analysis of human behaviours in terms of temporal assortative mixing estimation [6].

In this work we will focus on the two collections expressing face-to-face contacts in primary [23] and high school [24] contexts. We focus on both days of the primary school and all four days of the high school. Nodes are enriched with attributive information about the gender of students. For representing the temporal higher-order structure, we leverage a similar method as that introduced in [17], namely: if at 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.

### 3.2. Topological analysis

In the case of hypergraphs modeling social interactions, hyperedges represent groups of individuals co-acting simultaneously. In this respect, we consider individuals connected by an

[1] http://www.sociopatterns.org/




(a) (b)

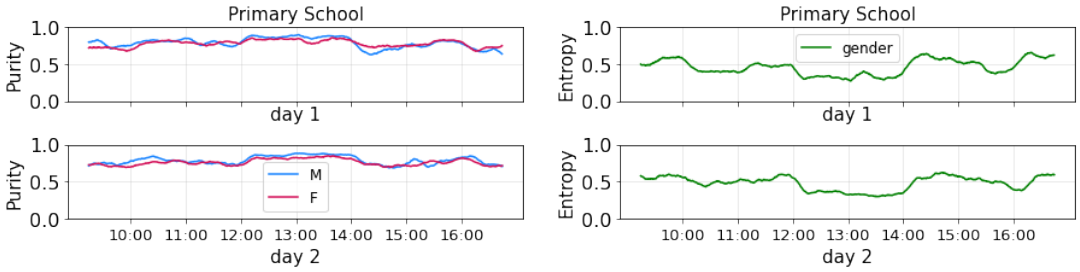
**Figure 1:** Hyperedge distribution (left) and node degree distribution in log-log scale (right) of the two hypernetworks

hyperedge if they interacted within a time frame of 20 seconds. Figure 1a compares the hyperedge size distribution of the two hypernetworks. Primary and High School show similar trends, with smaller-sized (unique) interactions being the most frequent. It should be noted that Primary records almost twice as many hyperedges as High School. As the number of individuals increases, the amount of hyperedges rapidly decreases, ranging down to extremely rare 5-way interactions, which occur less than ten times in each network. A possible explanation is that smaller interactions are generally the easiest to engage into, with larger ones requiring time and energy of more and more individuals [25]. This can also be confirmed by analyzing the frequency of interactions, that is how often the nodes of the same hyperedge interact through time. Both hypernetworks count nearly 70 000 interactions, which is peculiar considering that students in High School were monitored for twice as much time as the schoolchildren. Pairwise relations are once again the most frequent in both networks ($\sim 90\%$), possibly due to the slightness of the time frame used. Still, multiadic ones take up a non-negligible part of the whole.

A remarkable difference concerns the nodes' hyperdegrees. We refer to a node's hyperdegree as the number of incident hyperedges, that is the number of hyperedges that include that node. As shown in Figure 1b, both distributions follow a bell-like trend, with most values concentrating in the vicinity of the average and a few located near the minimum/maximum. However, degrees in Primary are generally higher, averaging at 125 connections ($\sigma = 55.16$); in contrast, High School holds lower — the average being 54 — but more clustered values ($\sigma = 27.34$).

### 3.3. Node-Features analysis

Of the 242 nodes in Primary, 115 are males, 112 are females, and the remaining part is of unknown gender (i.e., teachers). High School, instead, counts 176 males, 146 females and 7 teachers over 329 individuals. Note that despite having less nodes, Primary has more than double the amount of teachers. Indeed, Primary counts a teacher every 16 students, whereas High School counts one every 47.




(a) (b)

**Figure 2:** (a) Purity and (b) Entropy trends in the Primary School contact dataset.

We use $Entropy$ to describe the behaviour of nodes w.r.t. the gender attribute, as it quantifies the degree of disorder of the attribute. Let $A$ be the set of nodes' attributes, the entropy of an attribute $a \in A$ is computed as follows:

$$Entropy(a) = -\sum_{i}^{|A|} p(i|a) log p(i|a) \tag{1}$$

We highlight low entropy values in both networks, namely 0.54 in Primary and 0.46 in High School. This suggests that interpersonal relations in both scenarios are mainly uniform w.r.t. gender, with younger students having slightly less homogeneous contacts. It should be noted that these values take into account the existence of teachers, whose gender is unknown, thus increasing entropy; nonetheless, excluding edges that involve teachers lowers both entropy values by just $\sim 0.01$.

Another relevant metric is $Purity$, i.e., the relative frequency of the most frequent attribute value within a hyperedge. Formally, the purity of hyperedge $e$ is computed as follows:

$$Purity_e = \frac{\max_{a \in A}(\sum_{v \in e} a(v))}{|e|} \tag{2}$$

Then, $Purity$ is normalized by the number of hyperedges:

$$Purity = \frac{1}{|E|} \sum_{e \in E} Purity_e \tag{3}$$

Purity values confirm the previous observations with Entropy. Purity scores are high in both schools, namely 0.76 for Primary and 0.79 for High School, meaning that the majority of nodes in almost every hyperedge share the same gender. By studying mostly-male and mostly-female hyperedges separately, we also relieved that the former ones are $\sim 3\%$ purer than the latter ones, in both scenarios.

### 3.4. Interactions' Dynamics analysis

Both Figure 2 and Figure 3 focus on the average temporal trends of hyperedges' purities and entropies along time. While doing so, purity scores are disaggregated by gender labels, with the aim

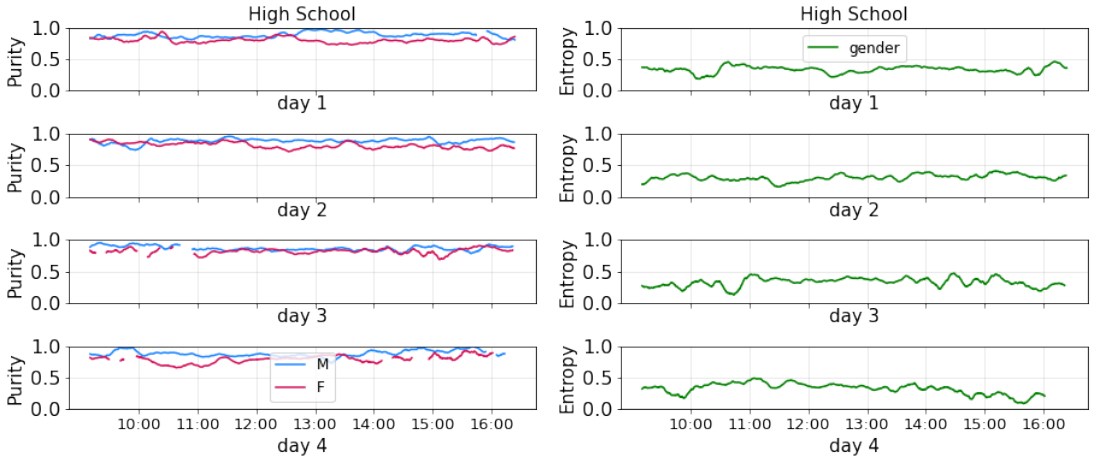



**Figure 3:** (a) Purity and (b) Entropy trends in the High School contact dataset.

to observe whether different patterns emerge from the labels. On average, both purity/entropy values are high/low, indicating a strong homogeneity by gender in both hypernetworks. Some little differences seem to emerge between the two datasets, since (i) primary school gender entropy is higher than the high school entropy, indicating that a more mixed interaction global pattern occurs in children contacts; (ii) some periodical patterns emerge in children contacts, whereas no periodicity is observed in the high school dataset; in detail, the entropy score trend in children contacts decreases at lunchtime [23], indicating how more homogeneous groups tend to be shaped by an "external" context. To conclude, no significant differences seem to emerge between male and female students by looking at the purity trends, even if slightly higher male scores than female ones could be observed sometimes in high school student contacts; however, the absence of periodical patterns does not allow us to better interpret/discuss such behaviours.

## 4. Discussions and Conclusions

In this work we tested the possibilities of ASH, our new model to represent and analyze complex streams of higher-order interactions enriched with information about node profiles. We extracted the temporal higher-order structure of two well-known datasets from the SocioPatterns project, namely Primary and High School face-to-face contacts. We enriched nodes with attributive information about the students' gender. Our main focus was on a qualitative study on the dynamics of group interactions that a graph-based study on the set of pairwise links could not highlight. We divided the analysis along three perspectives. On the one hand, we studied the global properties of the higher-order systems, e.g., highlighting how $n$-hyperedges distribute across group dimensions. On the other hand, we highlight the strong homogeneity of such groups w.r.t. the gender attribute. Finally, along the temporal dimension we discovered differences between the two schools: children seem to be characterized by periodical patterns,

e.g., lower entropy values during lunchtime, and their entropy is on average higher than high school students.

In future works, we plan to provide quantitative justifications of such preliminary exploratory observations. We also plan to define domain-specific measures based on a solid characterization of node profiles, leveraging them to better analyze the dynamics of higher-order contact structures. More insights about the dynamics of groups in face-to-face interaction data will be provided by extending this preliminary case study to the wide set of SocioPatterns datasets, as well as other contact patterns databases [26], and by testing different aggregation time frames.

## Acknowledgments

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 n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu).

## References

[1] P. Holme, J. Saramäki, Temporal networks, Physics reports 519 (2012) 97–125.

[2] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron, A. Sala, Feature-rich networks: going beyond complex network topologies, Applied Network Science 4 (2019) 1–13.

[3] M. Latapy, T. Viard, C. Magnien, Stream graphs and link streams for the modeling of interactions over time, Social Network Analysis and Mining 8 (2018) 1–29.

[4] G. Rossetti, R. Cazabet, Community discovery in dynamic networks: a survey, ACM Computing Surveys (CSUR) 51 (2018) 1–37.

[5] A. Divakaran, A. Mohan, Temporal link prediction: A survey, New Generation Computing 38 (2020) 213–258.

[6] S. Citraro, L. Milli, R. Cazabet, G. Rossetti, $\delta$-conformity: Multi-scale node assortativity in feature-rich stream graphs, arXiv preprint arXiv:2111.15534 (2021).

[7] M. Zanin, D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik, S. Boccaletti, Combining complex networks and data mining: why and how, Physics Reports 635 (2016) 1–44.

[8] P. Chunaev, Community detection in node-attributed social networks: a survey, Computer Science Review 37 (2020) 100286.

[9] L. Peel, J.-C. Delvenne, R. Lambiotte, Multiscale mixing patterns in networks, Proceedings of the National Academy of Sciences 115 (2018) 4057–4062.

[10] G. Rossetti, S. Citraro, L. Milli, Conformity: a path-aware homophily measure for node-attributed networks, IEEE Intelligent Systems 36 (2021) 25–34.

[11] S. Bhagat, G. Cormode, S. Muthukrishnan, Node classification in social networks, in: Social network data analytics, Springer, 2011, pp. 115–148.

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

[13] L. Torres, A. S. Blevins, D. Bassett, T. Eliassi-Rad, The why, how, and when of representations for complex systems, SIAM Review 63 (2021) 435–485.
[14] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, et al., The physics of higher-order interactions in complex systems, Nature Physics 17 (2021) 1093–1098.
[15] I. Iacopini, G. Petri, A. Barrat, V. Latora, Simplicial models of social contagion, Nature communications 10 (2019) 1–9.
[16] S. Chowdhary, A. Kumar, G. Cencetti, I. Iacopini, F. Battiston, Simplicial contagion in temporal higher-order networks, Journal of Physics: Complexity 2 (2021) 035019.
[17] G. Cencetti, F. Battiston, B. Lepri, M. Karsai, Temporal properties of higher-order interactions in social networks, Scientific reports 11 (2021) 1–10.
[18] F. Musciotto, F. Battiston, R. N. Mantegna, Detecting informative higher-order interactions in statistically validated hypergraphs, Communications Physics 4 (2021) 1–9.
[19] S. G. Aksoy, C. Joslyn, C. O. Marrero, B. Praggastis, E. Purvine, Hypernetwork science via high-order hypergraph walks, EPJ Data Science 9 (2020) 16.
[20] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, A. Vespignani, Dynamics of person-to-person interactions from distributed rfid sensor networks, PloS one 5 (2010) e11596.
[21] M. Génois, C. L. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, A. Barrat, Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers, Network Science 3 (2015) 326–347.
[22] V. Gemmetto, A. Barrat, C. Cattuto, Mitigation of infectious disease at school: targeted class closure vs school closure, BMC infectious diseases 14 (2014) 1–10.
[23] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, et al., High-resolution measurements of face-to-face contact patterns in a primary school, PloS one 6 (2011) e23176.
[24] R. Mastrandrea, J. Fournet, A. Barrat, Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys, PloS one 10 (2015) e0136497.
[25] G. K. Zipf, Human Behaviour and the Principle of Least Effort, Addison-Wesley, 1949.
[26] P. Sapiezynski, A. Stopczynski, D. D. Lassen, S. Lehmann, Interaction data from the copenhagen networks study, Scientific Data 6 (2019) 1–10.