Applied Network Science

RESEARCH Open Access

Network cards: concise, readable summaries of network data

James Bagrow[1,2*] and Yong-Yeol Ahn[3,4]

*Correspondence:
james.bagrow@uvm.edu

[1] Mathematics and Statistics, University of Vermont, Burlington, VT, USA
[2] Vermont Complex Systems Center, University of Vermont, Burlington, VT, USA
[3] Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA
[4] Network Science Institute, Indiana University, Bloomington, IN, USA

Abstract

The deluge of network datasets demands a standard way to effectively and succinctly summarize network datasets. Building on similar efforts to standardize the documentation of models and datasets in machine learning, here we propose *network cards*, short summaries of network datasets that can capture not only the basic statistics of the network but also information about the data construction process, provenance, ethical considerations, and other metadata. In this paper, we lay out (1) the rationales and objectives for network cards, (2) key elements that should be included in network cards, and (3) example network cards to underscore their benefits across a variety of research domains. We also provide a schema, templates, and a software package for generating network cards.

Keywords: Network data, Network summaries, Reporting guidelines, Tabular summary, Standardized reporting, Karate club, Plant-pollinator network, Temporal contact network, World airline network, Protein–protein interaction network

Introduction

Network structure can be found in numerous complex systems and it provides a unifying framework to study those systems collectively (Börner et al. 2007; Mitchell 2009; Newman 2018; Menczer et al. 2020). Beyond academic interests, we also live in a connected world and any actions we take online leave digital traces that echo various socioeconomic networks (Lazer et al. 2009). Due to its broad appeal and usefulness, the network perspective is widely used across domains and *network data* is ubiquitous in science and society.

Despite the universality and deluge of networks, there is currently no consensus nor standard procedures to report the characteristics of networks and their metadata. As argued in the case of documenting models and data in machine learning (Mitchell et al. 2019; Gebru et al. 2021), the lack of such standards—and the resulting lack of attention paid to important aspects of models and datasets—may lead to negative societal

outcomes. This is equally true for network datasets, particularly given the ubiquity and relevance of network datasets across academia and industry.

Here we introduce network cards,[1] standard tabular summaries of network data that capture both metadata about the network and statistics describing the data themselves. While we cannot expect a "one-size-fits-all" solution given the variety of networks that scientists consider, network card's ability to summarize the most basic network statistics will be broadly appealing and make network data more accessible. Network cards are intended to be flexible enough to describe a variety of rich network types such as multilayer and higher-order networks and can even extend to describe multiple networks simultaneously. Network cards are also complementary to other efforts such as more detailed "datasheets" (Gebru et al. 2021). Cards can provide a succinct summary of network-specific information which can be expanded upon when needed. And, when aspects of a datasheet—which is focused on datasets for machine learning—are not exactly applicable, the network cards can be used to document both network statistics and key metadata that pertain to data provenance, ethics, privacy, and other concerns.

A network card can provide researchers with a number of benefits. Glanceable information about a network dataset allows researchers to quickly digest the most salient features of the dataset. Network cards will answer the most basic questions about network data, such as: What are the nodes? What defines links? How big is the network? How dense? Awareness of the basic information provided by the network cards can prevent misinterpretation of network data and having a set of standardized statistics and information will encourage both data producers and users to pay attention to key details such as how and when the data were gathered or whether there are any important ethical considerations involving the dataset.

The need for concise network summaries

Most network researchers are familiar with the Zachary Karate Club (Zachary 1977) (Table 1), a very popular example network. But did you know that the original data contained eight different "interaction contexts" and can be considered a multiplex network? This rich context is now mostly lost and rarely discussed because the most widely disseminated dataset for the Karate Club was the version where all contexts are collapsed down to binary edges.

As another example, consider protein–protein interaction (PPI) networks. These data are collected through experimental assays that test whether proteins interact with one another. But not all assays are designed to detect dyadic (pairwise) interactions. For example, whereas Yeast Two-Hybrid (Y2H) does test pair interactions in isolation (Brückner et al. 2009), Affinity Purification Mass Spectrometry (AP-MS) uses tagged bait-prey protein pairs to identify interacting clusters (complexes) of proteins (Gingras et al. 2007). In other words, the results of AP-MS assays will over-represent *cliques*. These different data generating processes have profound consequences for the final network structure, with AP-MS-derived networks exhibiting far more clustering than Y2H. A researcher not recognizing these differences may draw inappropriate

[1] See also: github.com/network-cards.

Table 1 Example network card for the Zachary Karate Club

Name	Zachary Karate Club
Kind	Undirected, unweighted
Nodes are	Members of club at US university
Links are	Members consistently interacted outside club
Considerations	Heavily used as an example network
Number of nodes	34
Number of links	78
Degree*	4.588 [1, 17]
Clustering	0.571
Connected	Yes
Diameter	5
Assortativity (degree)	-0.476
Node metadata	None
Link metadata	None (original study included eight interaction contexts)
Date of creation	1977
Data generating process	Direct observation of club members during period 1970-72
Ethics	-
Funding	None
Citation	Zachary (1977) [8]
Access	https://networkrepository.com/karate.php (accessed 2022-02-12)

*Distributions summarized with average [min, max].

A network card is a concise, three-panel, tabular summary of a network and associated information. The three panels summarize, from top, overall information about the network, the structure of the network such as its size and density, and meta-information such as where the data originated and any ethical considerations associated with the data

*Distributions summarized with average [min, max]

and biased conclusions, which may lead to potential harms down the line. This leads us to ask, how to best retain critical information such as these experimental details, external to a network's structure, when disseminating the network data?

A contributing factor towards losing critical details about experiments and data over time may be information overload. The scientific literature is estimated to double in size every 15–25 years (Bornmann and Mutz 2015; Fortunato et al. 2018; Bornmann et al. 2021). Furthermore, particularly in the case of network data, we are challenged not only by the growth of the absolute volume of papers, but also the breadth of the works that deal with network data. Network science is a highly interdisciplinary field and researchers interested in network data come from all domains of research (Börner et al. 2007) and any efforts to retain critical experimental and data details should be both succinct—to mitigate information overload—and broadly accessible.

Alongside the literature's exponential growth, many fields of research are in the midst of a replication crisis, where past work has been called into question (Ioannidis 2005; Collaboration 2015; Nissen et al. 2016; Cockburn et al. 2020). Causes of this crisis include poor statistical practice (Loken and Gelman 2017; Benjamin et al. 2018; Gosselin 2020) and poor data documentation (Kanwal et al. 2017; Taylor et al. 2018; Rupprecht et al. 2020). Documenting the provenance of data is crucial for data-driven studies of networks, and there is a real need for a systematic, standard way to describe the various

details of a network dataset, details that are not strictly part of the network topology itself and so are often lost as researchers share data files describing that topology but nothing else.

To retain critical details accompanying datasets in the face of information overload while accommodating broad interdisciplinary interest, we argue that standardization, portability, and succinctness of presentation are critical. Standardization is crucial not only because it acts as a shared *checklist* that keep researchers from omitting important details, but also because it allows the development of shared understanding and tools. The more portable it is, the easier to prevent the loss of critical metadata. Succinctness of presentation is critical: a researcher should (correctly) understand the results of a scientific study as quickly (and accurately) as possible. As the writer of a scientific study, this can require hard choices when describing the results, using enough jargon for the intended audience but not more, enough technical detail for someone to replicate the study but not so much they cannot follow the results, and enough interpretation so the results are communicated and contextualized clearly but correctly.

Taken together, these factors—information overload, interdisciplinary network interest, and the need to document data—point towards the need for a succinct, standardized, broadly readable, and portable means to summarize network datasets. Our goal here is to propose a solution to meet these needs, the network card.

Network cards

Network cards were designed to achieve three properties:

1. Concise. A card should be compact and efficient, occupying one page at most. Concise presentation also helps the portability of the cards.
2. Readable. Any researcher familiar with networks should instantly understand all card contents. Cards should be as approachable to non-specialists as possible.
3. General and flexible. Works for all types of networks. Can be adapted to special circumstances.

Inspired by summaries of regression models, we propose a three-panel tabular layout for network cards, with the first panel providing overall information about the network, the second panel focusing on structural information related to the network's topology, and the third panel describing further meta-information such as availability of metadata, how the data were generated or gathered, and any ethical concerns to consider.

Table 1 shows a network card for the famous Zachary Karate Club, one of the prototypical example networks used in the literature. From a glance at the card, a reader can deduce a number of salient details including where the network data came from, what constitutes nodes and links, the network's size, and where to go for more information. Although the Karate Club is such a heavily used example, as we discussed above, less familiar may be some of the associated features of the club, such as when the data were gathered, and the fact that multiple interaction contexts were captured for the social ties in the network. If this network card can "follow" the data and be readily available, it would be less likely that such information gets lost.

Table 2 Network card for a plant–pollinator network

Name	M_PL_058
Kind	Undirected, unweighted
Nodes are	Plants and pollinators
Links are	Pollination interactions
Considerations	Bipartite [32 plants, 81 pollinators]
Number of nodes	113
Number of links	319
Degree*	5.646 [1, 28]
Clustering	0
Connected	2 components [98.23% in largest]
Component size	[111, 2]
Diameter	n/a
Largest component's diameter	6
Assortativity (degree)	-0.379
Node metadata	Species name
Link metadata	None
Date of creation	Spring 2005
Data generating process	Field observation at study sites in Natural Park of Cap de Creus in Catalonia, Spain; data retrieved from web-of-life.es
Ethics	Work complied with the current laws of Spain
Funding	Integrated European Project Assessing Large Scale Risks to Biodiversity with Tested Methods, Ministerio de Ciencia y Tecnología projects Efecto de las Especies Invasoras en las Redes de Polinización, Determinantes Biológicos del Riesgo de Invasiones Vegetales
Citation	Bartomeus *et al.* (2008) [24]
Access	https://www.web-of-life.es/networkjson.php?id=M_PL_058 (accessed 2022-03-10)

*Distributions summarized with average [min, max].

This example shows how to highlight the details of a bipartite network

*Distributions summarized with average [min, max]

We now discuss the contents of each panel of a network card in greater detail. "Appendix A" contains a complete description, derived from our schema, of all entries broken down by panel.

Overall information

The top of each network card provides an overall description of the network, a name, whether the network is undirected or directed, whether there are link weights, and any further considerations worth bringing attention to. Sometimes left unstated, but absolutely crucial, are explicit definitions for the nodes and links. Almost always the first question a network scientist asks when a collaborator brings them unfamiliar network data is what constitutes the nodes and what relationship defines the links. We believe that providing a prominent and explicit venue for displaying these definitions is one of the most valuable aspects of network cards.

Structure

The second panel of a network card provides basic summary statistics for describing the structure of the network. We intentionally rely on the most common and broadly

understood network statistics, ensuring cards are readable to as many researchers as possible. Our goal is to summarize the size, density, and connectivity of the network, so we report the number of nodes and links, the average degree, average clustering coefficient, whether the network is connected, and the degree assortativity of the network. If the network is connected, we report the network's diameter. If it is not connected, we report the number of connected components, the proportion of nodes in the largest component, a summary of the distribution of nodes per component (component size), and the diameter of the largest connected component. Table 2, which we discuss in greater detail in Sect. 4, demonstrates these statistics.

Examining some network cards, one realizes there is some redundancy in these quantities. For example, the average degree $\langle k \rangle$ comes directly from the numbers of nodes N and links M (via $\langle k \rangle = 2M/N$), which we already report. While it is thus not strictly necessary, we believe it is worth including an entry describing the degree because (1) we can summarize more of the distribution than its first moment (see our notes on statistical summaries, below), (2) it saves readers some (albeit simple) mental arithmetic while reading, and (3) it provides a nice way to compare networks: often, differences in average degree are more important than different numbers of nodes or links. The clustering coefficient is another case where its inclusion is useful even if it appears unnecessary: for a bipartite network, clustering is zero by definition. However, we feel it is worth always including to maintain consistency and in this case to reinforce as a "sanity check" this natural feature of a bipartite network. (Imagine the dismay of someone writing code to process their data and discovering triangles in what should be a bipartite network!)

Meta-information

Beyond general information about the network and summaries of the network's topology, we sought to capture additional salient details that describe the network without directly relating to the network data itself. The presence of metadata describing the elements of the network is top of mind, and we include entries for noting metadata describing nodes and describing links. Examples of node metadata include demographic variables for individuals in a social network or gene ontology terms for proteins in a PPI network. Examples of link metadata include the mode of communication in a contact network (e.g., text or phone call in a mobile phone network), and the link type (or interaction context) in a multilayer or multiplex network. Often these metadata draw on unique facets of the network under study and can enable novel research questions by drawing comparisons between features of the network and values in the metadata.

Next, we include information on the date of creation, allowing researchers to specify the time period of the dataset. In particular, we believe it is important to specify whether the network came from ongoing data collected over a long period. One situation where this is useful is for comparing networks (see below): we can use the date entry to distinguish, for instance, a network measured in 2020–2021 from that network measured during 2022–2023.

An entry describing the data generating process is also included in the meta-information box. Broadly construed, this entry is intended to briefly describe how the network data were collected, measured, captured, or otherwise quantified. A social network derived from mobile phone billing records can thus be distinguished from a social

network derived from chest-worn proximity sensors and from a social network derived from manual observation by researchers. All are "social networks", but these very different generating processes will have profound impacts on the form of the network and how the data should be interpreted. Of course, networks are captured by all different manners of data generating processes, making it necessary to be flexible in what details to include, and some processes can be quite complex and difficult to describe succinctly. Nevertheless, brief descriptions of the process, even relying on citations for more information, are invaluable and essential.

Lastly, and in many ways most importantly, we provide entries for reporting ethical concerns, funding disclosures, a citation, and an access field describing where the data were retrieved. The ethics entry can be used highlight particular concerns such as whether the network data required informed consent or whether the network data can be freely shared or not. Researchers may also wish to indicate whether the data are inappropriate for certain applications.

Important considerations

Some important further considerations are worth noting.

Importance of meta-information

While summary statistics of the topology can be extracted programmatically, data formats are often poor documents of ethical concerns, the meanings of nodes and links, the data generating process, and so forth, all of which are important for reproducibility and better understanding the data's nature. Network cards are therefore a valuable additional document, providing a succinct record of these critical but often ephemeral information that are not readily captured in the network's topology.

Statistical summaries

Most networks are large and distributions of numeric network properties are typically employed, perhaps the most popular being the network's degree distribution, the number of connections per node. To include such information in a network card, we recommend using basic summary statistics, the simplicity of these statistics being key to our design goals. For most situations, a distribution can be summarized using the mean value along with a measure of range such as its minimum and maximum. However, many network statistics are heavy-tailed or broadly distributed and a more robust choice than the mean is the median. Therefore, we propose basic summaries of the form, "average [min, max]" or "median [5th percentile, 95th percentile]". One exception is if the quantity being summarized has only a few values. This can happen, for instance, when summarizing the component sizes of a network that is not connected but has, say, only two connected components—it does not make much sense to include summary statistics for only two values. We recommend simply including a list of the observed values when there are five or fewer values; our implementation (see below) will do this automatically. We encourage users of network cards to always include a footnote describing exactly how distributions are summarized such as we use in Table 1.

Absence of graphics

Graph layout visualizations are often used to present networks and we considered including an entry for such a figure. We decided against it for two reasons. One, layouts are not always helpful or appropriate. They work best for sparser and smaller networks. Many networks are simply too large or dense to be readable in a two-dimensional projection, as anyone who has found themselves staring at a "hairball" graphic can attest. Two, including graphics makes it more challenging to support formats such as plain text or spreadsheets. (This also rules out other graphical summaries such as histograms.) A purely written format makes such representations easy to produce.

Comparing multiple networks

One particular strength of concise tabular summaries is that they are readily extended to multiple networks simply by adding additional columns. In other words, if a one-network card can be thought of as a two-column table, with the first column labeling each entry and the second for the entry's contents, then a "multicard" will have one column for labels and one column for each network. A multicard can be useful in several scenarios, such as comparing different network extraction techniques for the same data, capturing snapshots of a dynamic network, or even comparing replications of one study across different experiments. We show an example multicard in Sect. 4.

Special networks

Networks are complex, and there is a whole zoology devoted to different forms of networks capturing all manner of different structural and dynamic properties. Our standard network card accounts for all combinations of directed, undirected, weighted, and unweighted networks, but many other types of networks exist, some of which can also be handled well by our standard card but others may require further consideration. We discuss ways to accommodate such networks.

Bipartite networks, where nodes form two disjoint sets and links exist only between nodes in different sets, can be accommodated with a standard card, but it is worth explicitly denoting the network's bipartiteness, as we do in Table 2. Some statistics in the structure panel, such as the degree distribution, may be replicated to report the distribution for both node sets separately.

Temporal networks can be most naturally accommodated by either transforming them into a static network and noting its temporal nature as metadata (see Table 3), comparing multiple snapshots of the network with a multicard, or adapting some entries of the card (particularly those in the structure panel) to describe the temporal network directly. While this last option seems most appealing, it can in fact be the most challenging, as considerable research continues on how best to quantify temporal networks, and the results of this work have not yet experienced wide adoption compared to the more basic measures we used. And in terms of describing the results of transforming the dynamic network into a static network, a network multicard with one column for each choice of transform can actually work very well to demonstrate and compare those transformations against one another.

Signed networks, where edges have positive and negative values associated with them, can be handled in a standard network card simply by denoting signedness in the

Table 3 Network card for a temporal contact (close proximity) network

Name	TNet Malawi Pilot
Kind	Undirected, weighted
Nodes are	Study participants
Links are	Close proximity interactions
Link weights are	Number of interactions
Considerations	
Number of nodes	86
Number of links	347
Degree*	8.070 [1, 31]
Clustering	0.527
Connected	2 components [97.67% in largest]
Component size	[84, 2]
Diameter	n/a
Largest component's diameter	5
Assortativity (degree)	0.0363
Node metadata	None
Link metadata	Time and duration of interaction
Date of creation	2019-12-16 to 2020-01-10
Data generating process	Study participants in a rural village in Malawi wore a low-power sensor on the chest to measure their proximity to other participants. Time of contact was recorded and transformed to link weights
Ethics	Written consent was obtained from all participants or their guardians (both, in the case of adolescents). Study approved by Ethical Committee at the University of Zurich (OEC IRB #2018-046) and Ethical Committee at College of Medicine in Malawi (P.10/19/2825)
Funding	UNICEF Malawi and support from the Lagrange Project funded by the CRT Foundation
Citation	Ozella *et al.* (2021) [25]
Access	http://www.sociopatterns.org/datasets/contact-patterns-in-a-village-in-rural-malawi/ (accessed 2022-03-24)

*Distributions summarized with average [min, max].

This example highlights using a card to document transformation of a dynamic network to a (weighted) static network as well as describing ethical concerns (informed consent)

*Distributions summarized with average [min, max]

considerations entry. This entry can also include a basic statistic for the overall proportion of negative links, for example: "links are signed [23.5% links are negative]". For a signed network, it is crucial to denote the meaning of signedness in the "Links are" entry, for example: "links are ally (positive weight) or adversary (negative weight) ties".

Multilayer networks, where nodes are associated with one or more contexts or layers and links exist between or across layers, can also be handled by standard cards by denoting their layers as considerations and as node/link metadata. The structure of the different layers can be illustrated, at least when there are not too many layers, by either expanding each statistic, for example reporting layerwise the sizes, densities, and so forth, or by using a multicard where each column describes a layer of the network. It may be worth including multilayer-specific statistics in the structure panel of the card, but again relying on uncommon measures may not be worth the loss in readability to some audiences. Important considerations to note are if the network is multiplex, where nodes "replicate" across each layer and, closely related to multiplex, if the network is multirelation, where multiple links can exist between nodes (cf. Table 1). Both situations

can be well described using the node and link metadata entries as well as the considerations entry as needed.

Lastly, higher-order networks or hypergraphs have recently seen increased interest. A network card can immediately indicate whether the network is higher-order using the 'Links are' entry, for example: "Links are: social groups (hyperlinks)". Among the possible choices for additional statistics to describe the hypergraph's structure, we recommend at minimum adding a statistic for the distribution of link size (nodes per link). This basic quantity should be broadly understood by readers and captures much useful information. Other statistics specific to higher-order networks may be worth including as well, again under the caveat that broad readability of the card should be maintained as much as possible.

Implementation

To make network cards easy to use, we have created an open source implementation available online (github.com/network-cards). Our package facilitates generating network cards as tables and spreadsheets, and to read and write network cards in a standard format (we provide a schema). Currently, our package works with Python; over time, we hope to support other languages, such as R, MATLAB, and Julia. We also provide templates for the common network types that researchers can complete manually.

Examples of network cards

We have assembled some example networks meant to highlight the usefulness of network cards as summaries across research problems involving network data. These examples span social, biological and ecological research.

Our first example, already introduced in Table 1, is the famous Zachary Karate Club. As noted earlier, it displays both commonly known properties of the network as well as hardly-discussed features (multiple interaction contexts).

Table 2, also previously seen, shows a network card for a bipartite plant–pollinator network (Bartomeus et al. 2008). This network, collected from field observations in Spain, highlights the card's flexibility at concisely capturing metadata. In particular, it describes the bipartiteness including the numbers of plants and pollinators, the available metadata (species names are associated with each node), and the study details (when and where the data were collected). We also see the network is relatively dense, is not globally connected due to a single disjoint link, and the nodes are strongly degree dissortative.

Next, Table 3 shows the network card for a temporal contact network (Ozella et al. 2021). This network was captured from proximity sensors worn by participants. We transform the temporal network data to a weighted network and the network card succinctly documents our processing. As expected for most social networks due to triadic closure, this network is triangle-heavy, with a clustering coefficient greater than 0.5. Like the previous example, we again see that the network consists of two disconnected components, where one contains most nodes. This network card also illustrates how one can describe ethical concerns for the study, including acquiring informed consent from study participants.

Table 4 Network card for a directed network

Name	OpenFlights routes
Kind	Directed, weighted
Nodes are	Airports
Links are	Direct routes flown between airports (source node: departing airport, target node: arriving airport)
Link weights are	Number of routes
Considerations	Historical records, updated 2014
Number of nodes	3425
Number of links	37595 (1 self-loop)
— Bidirectional links	48.8%
Degree (in/out)*	10.9766 [0, 238]
Degree+	21.9533 [1, 477]
Clustering	0.4692
Connected	Disconnected
Assortativity (degree)	-0.0104
Node metadata	IATA airport codes
Link metadata	None (airline IDs, codeshare status, equipment IDs available in original data)
Date of creation	2014
Data generating process	Open flights data retrieved, codeshare routes removed, routes grouped by airport codes to get directed links and weights
Ethics	-
Funding	None
Citation	None
Access	https://openflights.org/data.html (accessed 2022-09-26)

*Distributions summarized with average [min, max].

+Undirected.

This example highlights structural fields specific for directed networks. For directed networks we recommend explicitly describing the directionality of links by referring to source and target nodes

*Distributions summarized with average [min, max]

+Undirected

As another useful example, in Table 4 we present a network card for a directed network, in this case the network of direct flight routes flown between airports. Here, nodes are airports, represented with IATA airport codes, and links are directed, weighted links counting the number of direct flights between pairs of airports. Structurally, we report the number of bidirectional links as a proportion of all links, as well as summaries of the in-degrees, out-degrees and the degree treating the network as undirected. As we show in this example, for directed networks, we recommend always explicitly defining link directionality by referring to the source and target nodes in the "Links are" entry.

Table 5 meanwhile, shows the network card for the recently released HuRI, the human reference interactome (Luck et al. 2020). This example shows how biological information can be put into a card, describing the gene metadata associated with nodes in the network and a brief description of the high-throughput assays used to infer protein–protein interactions (PPIs). A researcher interested in these data will immediately know where they can turn to enrich their study with node metadata, in

Table 5 Network card for a protein–protein interaction network

Name	HuRI
Kind	Undirected, unweighted
Nodes are	Human proteins
Links are	Binary protein interactions
Considerations	
Number of nodes	8 272
Number of links	52 548 [480 self-loops]
Degree*	12.705 [1, 500]
Clustering	0.0592
Connected	72 components [98.51% in largest]
Component size*	114.889 [1, 8 149]
Diameter	n/a
Largest component's diameter	12
Assortativity (degree)	-0.115
Node metadata	GENCODE v27 gene annotations
Link metadata	None
Date of creation	2019
Data generating process	Links inferred using a high-throughput three-panel yeast two-hybrid assay applied to pairs of protein-encoding genes taken from human ORFeome v9.1
Funding	National Institutes of Health and others
Citation	Luck *et al.* (2020) [26]
Access	http://www.interactome-atlas.org/download (accessed 2022-04-01)

*Distributions summarized with average [min, max].

*Distributions summarized with average [min, max]

this case using standard GENCODE gene annotations. HuRI also exhibits self-loops, capturing a small set of self-interacting proteins, and the card naturally draws attention to this information.

Luck et al. contrast HuRI with pre-existing PPI networks, one ("Lit-BM") extracted binary interactions from literature curated datasets, another ("HI-union") combined all previous screening experiments conducted by the research group with HuRI. We compare these networks with HuRI in a multicard shown in Table 6. With these summaries, we can succinctly define the similarities and differences in the networks, both in their data generating processes and in their structure.

Discussion

In this paper we propose network cards, simple and accessible tabular summaries of network datasets. Network cards are intended to be concise, readable, and flexible. Using a corpus of example networks, we highlight the information contained within network cards and how researchers can employ them in their own work. To help researchers use network cards, we have created a schema, fill-in templates, and an open source software package for generating cards, all available at github.com/network-cards.

We envision network cards being useful in the following situations: (1) as tables in manuscripts and supporting material; (2) as summaries display on pages of online repositories of network data; (3) included with data downloads as metadata alongside "READMEs" and other information; (4) as reporting guidelines or checklists adopted specifically for studies using network data; (5) shown as part of internal presentations with collaborators working on shared data; (6) lastly, and with the caveat that dense

Table 6 A network "multicard". Here the HuRI network (Table) is compared against two other networks from the same study (Luck et al. 2020)

Name	Lit-BM	HuRI	HI-union
Kind	Undirected, unweighted	Undirected, unweighted	Undirected, unweighted
Nodes are	Human proteins	Human proteins	Human proteins
Links are	Binary protein interactions	Binary protein interactions	Binary protein interactions
Considerations			HI-union includes HuRI
Number of nodes	6 047	8 272	9 094
Number of links	13 441 [683 self-loops]	52 548 [480 self-loops]	64 006 [764 self-loops]
Degree*	4.446 [1, 415]	12.705 [1, 500]	14.077 [1, 641]
Clustering	0.0618	0.0592	0.0621
Connected	248 components [92.31% in largest]	72 components [98.51% in largest]	70 components [98.81% in largest]
Component size*	24.383 [1, 5 582]	114.889 [1, 8 149]	129.914 [1, 8 986]
Diameter	n/a	n/a	n/a
Largest component's diameter	13	12	11
Assortativity (degree)	-0.0876	-0.115	-0.131
Node metadata	GENCODE v27 gene annotations	GENCODE v27 gene annotations	GENCODE v27 gene annotations
Link metadata	None	None	None
Date of creation	2019	2019	2005–2019
Data generating process	Links taken from literature-curated data set	Links inferred using a high-throughput three-panel yeast two-hybrid assay applied to pairs of protein-encoding genes taken from human ORFeome v9.1	Links taken from union of previous PPI screens
Funding	National Institutes of Health and others	National Institutes of Health and others	National Institutes of Health and others
Citation	Luck *et al.* (2020) [26]	Luck *et al.* (2020) [26]	Luck *et al.* (2020) [26]
Access	`http://www.interactome-atlas.org/download` (accessed 2022-04-01)		

*Distributions summarized with average [min, max].

A reader can quickly ascertain similarities and differences between the networks

*Distributions summarized with average [min, max]

technical information may not be appropriate for some venues, with broad adoption, network cards may also be useful in presentations during conferences and meetings.

Broad adoption of network cards may lead to three potential benefits. One, it becomes easier to understand papers using network data. Readers can more quickly grasp the most salient details of the network or networks employed in the study—what are the nodes, what are the links, when was the network data collected—when those details are presented in a standard manner the reader is accustomed to. Quick and accessible information summaries are increasingly important as the volume of scientific research grows (Kostoff and Hartley 2001).

A second potential benefit of network cards stems from their highlighting of non-structural information such as ethical concerns or the presence of metadata, serving as a useful checklist. By drawing attention to these facets of the data, readers are better equipped to understand the appropriateness and broader consequences of working with the network data, especially important for data that come with ethical or privacy concerns. Highlighting these details is important if the data are available and the reader wishes to use it themselves (Gebru et al. 2021). Those details, which may be lost when considering only the network structure, may reveal that the data may not be suitable for certain purposes. The succinctness of network cards can increase the chances for researchers to correctly identify which data to use for themselves.

Lastly, a third benefit of broad use of network cards comes through automation. It is more common for papers to examine a corpus of hundreds or even thousands of different networks (Kunegis et al. 2013; Rossi and Ahmed 2015; Kujala et al. 2018; Broido and Clauset 2019; Voitalov et al. 2019; Lynn et al. 2020). Analyzing networks at this scale is invaluable for revealing broad patterns and trends across research domains (Ikehara and Clauset 2017). But such scale requires automation: code must be written to analyze each network programmatically. Network cards admit a machine-readable JSON format, for which we provide a schema. If network cards are created when networks are added to large corpora, then subsequent analysis programs can read those cards at the same time they read the network data itself. In other words, standardizing the representation of network meta-information using cards has the potential to make that meta-information computationally accessible which can then drive a deeper understanding of network corpora.

Appendix A: Card contents

Here we list all the entries that constitute the standard three-panel network card. These entries are organized by panel and the descriptions were derived from a schema we have drafted to help with the standardization process.

Overall	Describes the network type (directed, weighted, etc.), what do nodes and links represent, what other key considerations do the data entail.
Name	A written identifier for the network
Kind	Is the network undirected, unweighted, etc..
Nodes are	The definition of nodes.
Links are	The definition of links (edges).
Link weights are	The definition of link (edge) weights.
Considerations	What considerations should be taken into account regarding the network's overall properties.
Structure	Summary statistics for the size, density, and other properties of the network structure.
Number of nodes	The number of nodes in the network.
Number of links	The number of links (edges) in the network.
Degree	A summary of the degree distribution.
Clustering	The average clustering of the network.
Connected	A description and summary of the network's connectivity.
Component size	An optional description about the sizes of the network's components.
Diameter	The diameter of the network, if connected.
Largest component's diameter	The diameter of the largest component's induced subgraph, if network is not connected.

Assortativity (degree)	The degree assortativity of the network.
Meta-information	Further details such as the presence of any node or link metadata, data collection documentation, ethical considerations, citations, and any funding acknowledgments.
Node metadata	A description of any metadata associated with nodes.
Link metadata	A description of any metadata associated with links.
Date of creation	A description of when the network data were gathered or created.
Data generating process	A description of how the network data were generated.
Ethics	A description of ethical considerations for the network data.
Funding	A description of funding related to the network data.
Citation	A citation and/or DOI associated with the network data.
Access	A URL or other description of where data can be obtained.

For more details, please see github.com/network-cards.

Author contributions
JB and Y-YA designed and conducted the research and wrote the manuscript.

Funding
J.B. acknowledges support by Google Open Source under the Open Source Complex Ecosystems And Networks (OCEAN) project. Y.-Y.A. acknowledges support by the Air Force Office of Scientific Research under award number FA9550-19-1-0391.

Availability of data and materials
Network data for ***Zachary Karate Club*** (Table 1 Zachary 1977) was retrieved from https://nrvis.com/download/data/soc/soc-karate.zip on 12 February 2022. Network data for ***M_PL_058*** (Table 2 Bartomeus et al. 2008) was retrieved from https://www.web-of-life.es/networkjson.php?id=M_PL_058 on 10 March 2022. Network data for ***TNet Malawi Pilot*** (Table 3 Ozella et al. 2021) was retrieved from http://www.sociopatterns.org/datasets/contact-patterns-in-a-village-in-rural-malawi/ on 24 March 2022. Network data for ***OpenFlights routes*** (Table 4) was retrieved from https://openflights.org/data.html on 26 September 2022. Network data for ***HuRI***, ***Lit-BM***, and ***HI-union*** (Tables 5 and 6 Luck et al. 2020) were retrieved from http://www.interactome-atlas.org/download on 1 April 2022. Data files as used in the study have been deposited at https://doi.org/10.6084/m9.figshare.20286648.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 11 July 2022 Accepted: 31 October 2022
Published online: 16 December 2022

References

Bartomeus I, Vilà M, Santamaría L (2008) Contrasting effects of invasive plants in plant–pollinator networks. Oecologia 155(4):761–770. https://doi.org/10.1007/s00442-007-0946-1

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald AG, Hadfield JD, Hedges LV, Held L, Hua Ho T, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchler M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore DA, Morgan SL, Munafó M, Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Rouder J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J, Johnson VE (2018) Redefine statistical significance. Nat Hum Behav 2(1):6–10. https://doi.org/10.1038/s41562-017-0189-z

Börner K, Sanyal S, Vespignani A (2007) Network science. Annu Rev Inf Sci Technol 41(1):537–607. https://doi.org/10.1002/aris.2007.1440410119

Bornmann L, Mutz R (2015) Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J Assoc Inf Sci Technol 66(11):2215–2222. https://doi.org/10.1002/asi.23329

Bornmann L, Haunschild R, Mutz R (2021) Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Hum Soc Sci Commun 8(1):224. https://doi.org/10.1057/s41599-021-00903-w
Broido AD, Clauset A (2019) Scale-free networks are rare. Nature. Communications 10(1):1017. https://doi.org/10.1038/s41467-019-08746-5
Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci 10(6):2763–2788. https://doi.org/10.3390/ijms10062763
Cockburn A, Dragicevic P, Besancon L, Gutwin C (2020) Threats of a replication crisis in empirical computer science. Commun ACM 63(8):70–79. https://doi.org/10.1145/3360311
Collaboration OS (2015) Estimating the reproducibility of psychological science. Science 349(6251):4716. https://doi.org/10.1126/science.aac4716
Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B, Vespignani A, Waltman L, Wang D, Barabási A-L (2018) Science of science. Science 359(6379):0185. https://doi.org/10.1126/science.aao0185
Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H III, Crawford K (2021) Datasheets for datasets. Commun ACM 64(12):86–92. https://doi.org/10.1145/3458723
Gingras A-C, Gstaiger M, Raught B, Aebersold R (2007) Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 8(8):645–654. https://doi.org/10.1038/nrm2208
Gosselin R (2020) Statistical analysis must improve to address the reproducibility crisis: the access to transparent statistics (acts) call to action. BioEssays 42(1):1900189. https://doi.org/10.1002/bies.201900189
Ikehara K, Clauset A (2017) Characterizing the structural diversity of complex networks across domains. arXiv preprint arXiv:1710.11304
Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8):124. https://doi.org/10.1371/journal.pmed.0020124
Kanwal S, Khan FZ, Lonie A, Sinnott RO (2017) Investigating reproducibility and tracking provenance—a genomic workflow case study. BMC Bioinform 18(1):337. https://doi.org/10.1186/s12859-017-1747-0
Kostoff RN, Hartley J (2001) Structured abstracts for technical journals. Science 292(5519):1067. https://doi.org/10.1126/science.292.5519.1067a
Kujala R, Weckström C, Darst RK, Mladenović MN, Saramäki J (2018) A collection of public transport network data sets for 25 cities. Sci Data 5(1):180089. https://doi.org/10.1038/sdata.2018.89
Kunegis J (2013) KONECT: the Koblenz network collection. In: Proceedings of the 22nd international conference on world wide web. ACM, Rio de Janeiro, pp 1343–1350. https://doi.org/10.1145/2487788.2488173
Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) Computational social science. Science 323(5915):721–723. https://doi.org/10.1126/science.1167742
Loken E, Gelman A (2017) Measurement error and the replication crisis. Science 355(6325):584–585. https://doi.org/10.1126/science.aal3618
Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, Choi D, Coté AG, Daley M, Deimling S, Desbuleux A, Dricot A, Gebbia M, Hardy MF, Kishore N, Knapp JJ, Kovács IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD, Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin S-F, Choi SG, Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajdaoui H, Goebels F, Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N, MacWilliams A, Markey D, Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y, Sheykhkarimli D, Sheynkman GM, Simonovsky E, Taşan M, Tejeda A, Tropepe V, Twizere J-C, Wang Y, Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las Rivas J, Gaudet S, Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP, Calderwood MA (2020) A reference map of the human binary protein interactome. Nature 580(7803):402–408. https://doi.org/10.1038/s41586-020-2188-x
Lynn CW, Papadopoulos L, Kahn AE, Bassett DS (2020) Human information processing in complex networks. Nat Phys 16(9):965–973. https://doi.org/10.1038/s41567-020-0924-7
Menczer F, Fortunato S, Davis CA (2020) A first course in network science. Cambridge University Press, Cambridge
Mitchell M (2009) Complexity: a guided tour. Oxford University Press, Oxford
Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T (2019) Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. ACM, Atlanta, pp 220–229. https://doi.org/10.1145/3287560.3287596
Newman MEJ (2018) Networks: an introduction, 2nd edn. Oxford University Press, Oxford
Nissen SB, Magidson T, Gross K, Bergstrom CT (2016) Publication bias and the canonization of false facts. eLife 5:21451. https://doi.org/10.7554/eLife.21451
Ozella L, Paolotti D, Lichand G, Rodríguez JP, Haenni S, Phuka J, Leal-Neto OB, Cattuto C (2021) Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi. EPJ Data Sci 10(1):46. https://doi.org/10.1140/epjds/s13688-021-00302-w
Rossi RA, Ahmed NK (2015) The network data repository with interactive graph analytics and visualization. In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence. AAAI'15, pp 4292–4293 (2015). https://doi.org/10.5555/2888116.2888372
Rupprecht L, Davis JC, Arnold C, Gur Y, Bhagwat D (2020) Improving reproducibility of data science pipelines through transparent provenance capture. Proc VLDB Endow 13(12):3354–3368. https://doi.org/10.14778/3415478.3415556
Taylor SJE, Eldabi T, Monks T, Rabe M, Uhrmacher AM (2018) Crisis, what crisis—Does reproducibility in modeling and simulation really matter? In: 2018 Winter simulation conference (WSC). IEEE, Gothenburg, pp 749–762. https://doi.org/10.1109/WSC.2018.8632232

Voitalov I, van der Hoorn P, van der Hofstad R, Krioukov D (2019) Scale-free networks well done. Phys Rev Res 1(3):033034. https://doi.org/10.1103/PhysRevResearch.1.033034
Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33(4):452–473. https://doi.org/10.1086/jar.33.4.3629752

Publisher's Note