Current Zoology, 2021, 67(1), 81–99
doi: 10.1093/cz/zoaa079
Article

OXFORD

Article

Multilayer network analyses as a toolkit for measuring social structure

Kelly R. Finn*

Neukom Institute, Dartmouth College, Hanover, NH 03755, USA

*Address correspondence to Kelly R. Finn. E-mail: Kelly.R.Finn@dartmouth.edu.
Handling editor: Matthew Hasenjager

Received on 30 May 2020; accepted on 22 December 2020

Abstract

The formalization of multilayer networks allows for new ways to measure sociality in complex social systems, including groups of animals. The same mathematical representation and methods are widely applicable across fields and study systems, and a network can represent drastically different types of data. As such, in order to apply analyses and interpret the results in a meaningful way the researcher must have a deep understanding of what their network is representing and what parts of it are being measured by a given analysis. Multilayer social networks can represent social structure with more detail than is often present in single layer networks, including multiple "types" of individuals, interactions, or relationships, and the extent to which these types are interdependent. Multilayer networks can also encompass a wider range of social scales, which can help overcome complications that are inherent to measuring sociality. In this paper, I dissect multilayer networks into the parts that correspond to different components of social structures. I then discuss common pitfalls to avoid across different stages of multilayer network analyses—some novel and some that always exist in social network analysis but are magnified in multi-layer representations. This paper serves as a primer for building a customized toolkit of multilayer network analyses, to probe components of social structure in animal social systems.

Key words: animal behavior, multilayer networks, relationships, sociality, social networks, social structure, subgroups

Multilayer network analysis is an increasingly popular approach for studying social systems. The aim of this paper is to help behavioral scientists who are new to multilayer network analysis navigate the structure of a multilayer social network, generate ideas about how it could be used with their data, and point out things to consider when constructing and analyzing their multilayer networks. Social network analysis is now widely used in animal behavior research, usually representing the social structure of an animal group as a single network (i.e., a monolayer or single layer network). Multilayer networks are a recent advance in network science (Kivelä et al. 2014), and can retain more detail about an animal group by using multiple connected "layers" of networks stacked together, allowing for a more complete representation of the social situations of animals in a group (Finn et al. 2019). Multilayer networks therefore facilitate defining and measuring specific components of social structure, making them useful to anyone who studies animal social behavior.

The first part of this paper explains why multilayer social network analyses form an excellent toolkit for measuring components of social structure. Next, I describe where the parts of a multilayer network are represented in a matrix, the mathematical representation that analyses often act on. Then, I break down the different parts of a multilayer network, from the most micro to the most macro social scale as they are functionally related to social dynamics. Finally, I describe many of the caveats to this approach to orient researchers to the many decisions and considerations faced when conducting multilayer social network analyses. I give suggestions to overcome these challenges, and/or suggest future work that is needed to address them.

Measuring Social Structure

It has long been realized that sociality exists across many scales (Hinde 1976, 1978; Hinde and Datta 1981), making it intrinsically difficult to define and measure. Further complicating matters, interactions and data widely vary across different species, making it difficult to generalize methods (Bergman and Beehner 2015). Hinde (1976) conceptualized the patterning of interactions over time as a relationship, and the patterning of relationships in a group as social structure. While numerous ways have been proposed to quantitatively measure a dyadic relationship (Silk et al. 2013), there are much fewer that measure an individual's overall social role, subgroups, or the entire structure of a group, while considering multiple possible types of interactions (Fischer et al. 2017) and other nuances such as time, space, and context. Quantifying these components of social structure is valuable, as they are believed to correspond with many important outcomes such as stress physiology (Sapolsky 1982; Brent et al. 2011; Balasubramaniam et al. 2016; Vandeleest et al. 2016; Wooddell et al. 2017; Schrock et al. 2019), disease transmission (Drewe 2010; Balasubramaniam et al. 2019), group stability (Beisner et al. 2015), and collective colony responses (Pinter-Wollman 2015). Quantifying characteristics of how groups are structured, sometimes conceptualized as notions of "social complexity," is also valuable for testing hypotheses about the evolution of sociality, such as predation pressures (Groenewoud et al. 2016) or other ecological factors (Wittemyer et al. 2005) driving its evolution, or social complexity itself driving the evolution of cognition (Whiten and Byrne 1988; Kummer et al. 1997; Dunbar 1998) and the evolution of communication (Freeberg et al. 2012; Sewall 2015). More generally, having a framework to think about social structure is important for any researcher studying a social animal (Lehmann and Ross 2011). If we lack a sense of an animal's overall social situation, do we really have a sense of the meaning or function of a given social interaction in our data? What does a given set of interaction data actually represent in the scope of the animals' social lives?

Fortunately, the computational tools available to social scientists are beginning to catch up with the complexity and detail that we have always known exists. In the current age, social scientists are increasingly using computational methods, researchers from mathematics, physics, and computer science are seeking social data to assess their toy models of social processes, and the division between "hard" and "soft" sciences is blurred or even trivial. Social network analysis was instrumental to bridging this gap. Networks themselves are interesting mathematical objects with numerous properties to be explored and measured, yet they also provide a framework to directly represent and analyze social data. Social networks offer a way to measure how relationships are organized in a social group, and can encode sociality across multiple social scales, from the social spheres of individuals (e.g. ego networks) to the overall group structure (Croft et al. 2008; Wey et al. 2008). However, a single network may miss nuances across relationships types, and some researchers have identified that multiple social networks may be necessary to characterize parts of a group's social structure and dynamics (Lehmann and Ross 2011; Barrett et al. 2012; Lehmann et al. 2012). Multilayer networks, however, can contain an even wider range of social units—they can simultaneously represent micro-level interactions, higher-level patterning of interactions into relationships and individual roles, macro scale structure across the whole group such as dominance hierarchies, and finally the entire group structure across all social domains, all of which are important components of sociality (Hobson et al. 2019).

The added structure of multilayer networks can allow researchers to overcome some common criticisms of single layer network representations of their study systems. For instance, a criticism of social networks is that they do not always factor in all of content that exists in real social ties (Borgatti et al. 2014)—interactions that are functionally different or occur in different contexts might not be distinguishable from each other if in the same network. However, multiple network layers can include several different types of interactions, or interactions that occur in multiple different contexts. Similarly, layers can represent interactions that occur in different time windows, addressing another criticism that networks are static representations and ignore dynamics (Borgatti et al. 2014). Since multilayer networks can link interactions across layers, they also allow for a better representation of interdependencies that exist across different types of interactions, which cannot be encoded in single layer networks. Multilayer networks are promising tools because they can encode such additional information about interactions within animal groups into a single framework (Barrett et al. 2012; Silk et al. 2018; Smith-Aguilar et al. 2018; Finn et al. 2019; Beisner et al. 2020; Pereira et al. 2020).

I suggest that measures of various parts of multilayer networks are ideal to create a "toolkit" to describe the numerous properties of social structure in socially sophisticated animal groups. Similar toolkit approaches exist for measuring other complex structures such as hierarchies (Zafeiris and Vicsek 2018), dyadic relationships (Silk et al. 2013), and time series (Goldberger et al. 2002), all of which themselves can exist within social structure. Despite the promise of multilayer social network analysis, it can still be a challenge to identify which characteristics of social structure one should quantify to assess a particular outcome, and which tools to use to best measure and summarize it. Conducting multilayer network analyses requires the researcher have a good conceptualization of what their network represents, what parts make sense to measure, and what exactly the tools they use are measuring. The remainder of the paper provides a detailed breakdown of multilayer social networks to make them more accessible, and serves as a guide for modeling a social system as a multilayer network to help researchers quantify specific components of social structure in their study systems to answer their own specific research questions.

The Supra-adjacency Matrix

The mathematical formulation of multilayer networks (De Domenico et al. 2014; Kivelä et al. 2014; Aleta and Moreno 2018), and the utility of various multilayer analyses for studying numerous topics in animal behavior (Pilosof et al. 2017; Silk et al. 2018; Smith-Aguilar et al. 2018; Finn et al. 2019; Atkisson et al. 2020; Pereira et al. 2020) have been reviewed elsewhere. Briefly, compared with a single layer network (Figure 1A), a multilayer network is comprised of multiple "layers" of networks (Figure 1B). Like single layer networks, *nodes* (Figure 2A) can represent some entity, often individuals (though they need not be), and are connected by *edges* (Figure 2B), which often represent an interaction or relationship (though they need not be). In multilayer networks, nodes in the same layer can be connected with *intralayer edges*, or nodes in different layers can be connected with *interlayer edges*. Multiple "stacks" of layers can be grouped into separate *aspects*, and nodes can even be connected across aspects (see Figure 3C). Different layers can represent different node types, edge types, time points, or any sort of distinction a researcher wants to make about social interactions (Finn et al. 2019). If layers represent different types of interactions or different time periods, there may be multiple "copies" of nodes across the layers (e.g. the same individuals are represented on both an aggression and an affiliative layer). Such "copies" of nodes

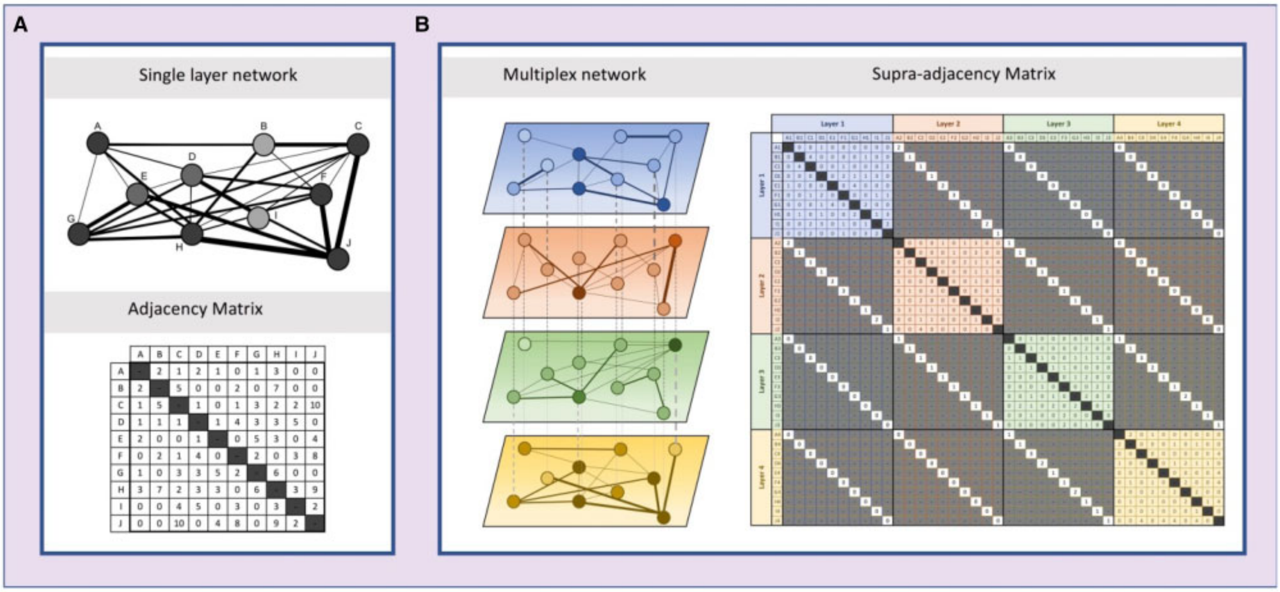

Figure 1. Single and multilayer networks and their matrices. A single layer network (A) is comprised of nodes (labeled A–J) connected by undirected edges whose weights are indicated by the thickness of the lines. If there are multiple types of connections that form edges, they could instead be separated into multiple distinct network layers as a multiplex network (B). Here, there is a copy or "node-tuple" of each A–J node on each layer, which are connected to each other across layers.

are sometimes referred to as "node-tuples" (see Figure 2A; Kivelä et al. 2014). Multiplex networks are special cases of a multilayer network, where interlayer edges only exist between node-tuples of the same node (see Figure 2A; Kivelä et al. 2014) (e.g. individuals are only connected to themselves across layers). Note that for most of this paper I discuss examples of multilayer networks that are multiplex social networks, where different layers are different types of interactions that occur among the same set of individuals, but other representations are also possible.

Single layer networks are represented by an adjacency matrix, where the IDs of individuals are represented across both columns and rows of a matrix (see Figure 2A, upper half). If individuals interact, the weight of the network edge (e.g., number or amount of interactions) is indicated where they align on the adjacency matrix (see Figure 2B, upper half). The diagonal cells of this matrix are cells that match individuals with themselves. If self-loops (i.e., an edge connecting a node to itself; e.g., an individual directing a behavior toward itself) are not allowed, the diagonal of this matrix is zero or holds no values (see Figures 1 and 2). If the edges do not have a direction, the adjacency matrix is symmetrical across the diagonal—the values in the upper triangle of cells exist as a mirror reflection across the diagonal in the lower triangle of cells. If the edges do have a direction, the numbers on the upper and lower the triangles represent interactions directed from one individual to another and can be different values. If the edges have weight (e.g., indicate the frequency or duration of interactions between 2 individuals), the values in the cells can be any number. If the edges are unweighted (e.g., indicate that an interaction happened, but no detail about the amount), the values in cells are binary 1 or 0. Figure 1A shows a single layer network that is weighted and undirected, and its associated adjacency matrix.

Like single layer networks, multilayer networks have matrix representations called "supra-adjacency matrices." These can be conceptualized as a matrix of matrices. Interactions between individuals that interact on the same layer are represented the same way as they would be in an adjacency matrix for a single layer network, and there are one of these matrices for each network layer. In a supra-adjacency matrix, adjacency matrices for each network layer are "glued" together diagonally, and interlayer edges that connect nodes across layers are indicated in the remaining squares on the larger matrix of matrices. In an adjacency matrix, all individuals are paired with all individuals, and in the supra-adjacency matrix, all of the layers are paired with all layers. On these non-diagonal matrices in the supra-adjacency matrix, the diagonal cells indicate where nodes link to copies of themselves on other layers. In a multiplex network, these will be the only interlayer edges. Nodes can be attached to all copies of themselves on all layers, or only to the adjacent layer (as seen in Figure 3B). If the entire multiplex network is undirected, including the interlayer edges, the entire supra-adjacency matrix is symmetrical across the diagonal—the matrices in the upper triangle of matrices exist as a mirror reflection across the diagonal in the lower triangle of matrices. If interlayer edges are directed, the lower triangle that corresponds to the upper triangle of the supra-adjacency matrix will be different. A weighted and undirected multiplex network and its associated supra-adjacency matrix are shown in Figure 1B.

A number of single layer matrix operations have been or could be generalized to supra-adjacency matrices [e.g., Page Rank versatility (De Domenico et al. 2015d), infomap community detection (De Domenico et al. 2015a)]. In addition, this network structure allows for the creation of new measures and analyses that do not exist for single layer networks.

The Functional Parts of a Multilayer Social Network

Individuals—node-tuples and interlayer edges

At the most micro level, social structures are comprised of individuals and interactions. While both are present in single layer networks, already there are differences between single and multilayer

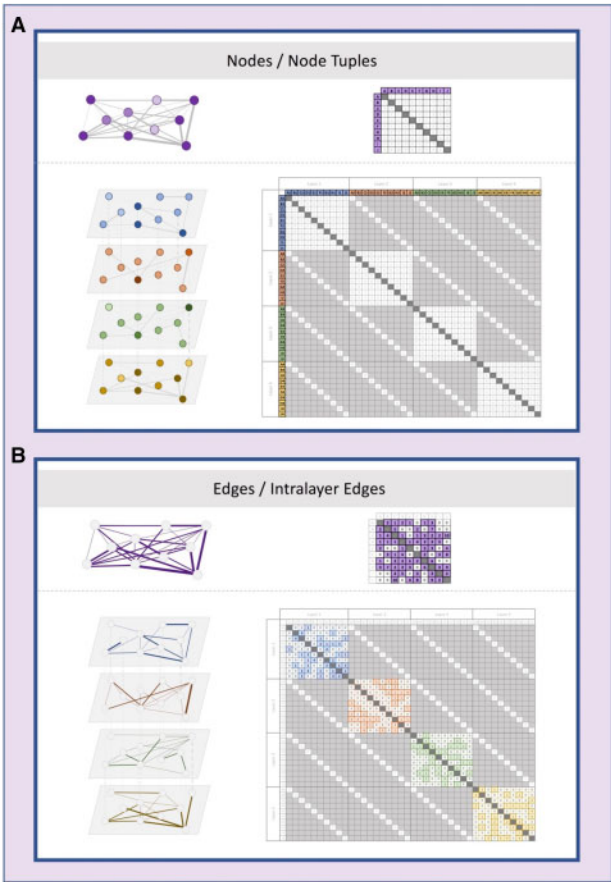

Figure 2. Nodes and intralayer edges in single and multilayer networks. The nodes in a single layer network (purple circles in the top of panel A) are represented as columns and rows of the corresponding adjacency matrix (purple cells in the top of panel A). The nodes of a multiplex network are also represented as the columns and rows of the corresponding supra-adjacency matrix, grouped by which layer they are in (each layer is a different color in panels A and B). Repeated copies of the nodes from each layer (i.e., node-tuples, colored circles in the bottom of panel A) are represented as repeated sets of rows and columns in the corresponding supra-adjacency matrix, pasted next to each other (colored cells in the bottom of panel A). The edges in a single layer network (purple lines in the top of panel B) are represented as values in the cells of the corresponding adjacency matrix (purple cells in the top of panel B). If there are no self-loops allowed, the diagonals are blank or zero (dark gray cells in panels A and B). In the adjacency matrix, values are zero where edges do not exist between different nodes, and non-zero with values that represent the edge weight where edges do exist (purple cells in the top of panel B). The edges of each layer of a multiplex (colored lines in the bottom of panel B) are represented as adjacency matrices for each layer, pasted together diagonally across the supra-adjacency matrix (blue, orange, green, and yellow cells in the bottom of panel B).

representations. A multiplex network may contain a set of node-tuples for each individual, which themselves are connected by interlayer edges and can "interact" (see Figure 4A). An intuitive example is a multiplex transportation network (Strano et al. 2015; Chodrow et al. 2016)—perhaps each node is a city, each layer is a mode of transportation between cities (e.g., flights, train, and car), and interlayer edges are travel within cities between for instance, a train station and an airport. In this case, both intralayer and interlayer edges could be weighted by travel time. In a multiplex social network where there are copies of an individual on each layer, interlayer edges are perhaps less intuitive when it comes to assigning values to

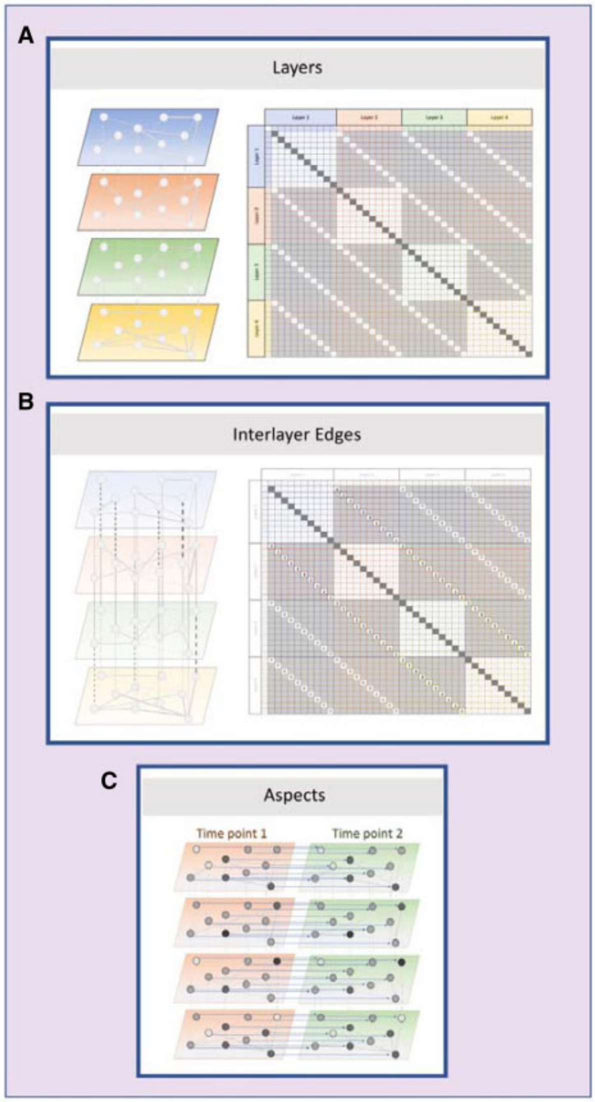

Figure 3. Unique multilayer network characteristics. There are parts of a multilayer network and their corresponding locations on a supra-adjacency matrix that do not exist on single layer networks or their adjacency matrices. In the same way that individuals are organized across columns and rows in an adjacency matrix, layers (colored rectangles in panel A) are organized as sets of columns and rows in the corresponding supra-adjacency matrix (colored sections in panel A). This structure creates cells that correspond to all combinations of all node-tuples. In a multiplex network where interlayer edges only connect to node-tuples of the same node, interlayer edges (dotted lines connecting nodes across layers in panel B) only exist across the diagonals of the matrices that represent connections across different layers (bold values in panel B). In this network, node-tuples are only connected on networks that are next to each other, so some of these values are 0 for layers that are not adjacent. If node-tuples were connected to their counterparts on all layers, all of these diagonals would have non-zero values. If the network is not a multiplex network and allowed node-tuples to be connected to node-tuples of different individuals across layers, cells other than the diagonals in these interlayer matrices could have values. If interlayer edges are directed, these values would not be symmetrical across the diagonal of the whole supra-adjacency matrix, as they are in panel B. Finally, multilayer networks can be represented as stacks of different layers (i.e., aspects, orange and green stack of layers in panel C), and nodes can be connected across aspects. While the supra-adjacency matrix is not shown here, it would increase in scale in a similar way from the jump from single to multiple layers, but to include copies of all of these parts for multiple aspects that themselves could be connected.

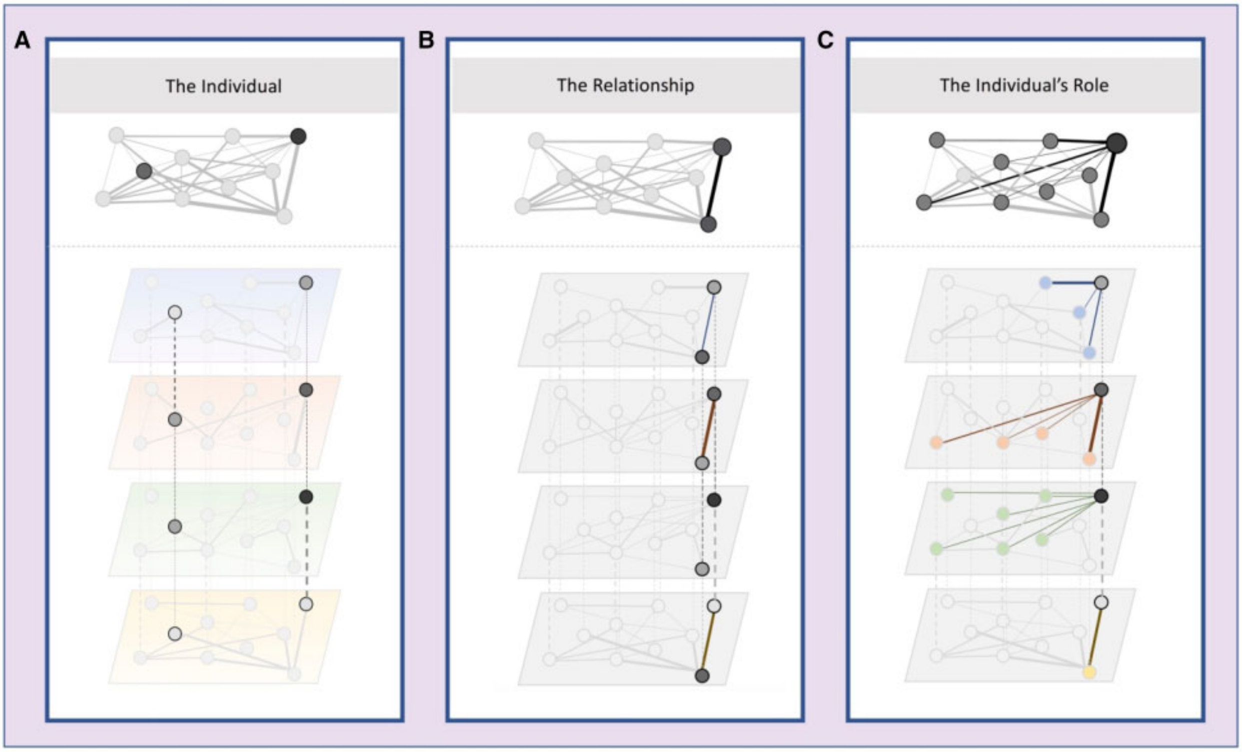

Figure 4. Smaller scale parts of a single and multiplex network. In both single and multiplex networks, individuals, relationships, and an individual's role exist as smaller scale parts of the network. For individuals (A), while 2 different individuals are represented only as nodes in the single layer network (A, top), there are multiple copies or node-tuples of each of them in each layer in the multiplex network, and they are connected by interlayer edges (A, bottom). For a relationship (B), there is only one edge connecting 2 individuals in a single layer network (B, top), while there can be edges on each layer between 2 individuals in a multiplex network (B, bottom). For an individual's role (C), an ego network includes the individual and it's alters connected by edges (C, top), while in a multiplex network, an ego network includes the node-tuples for an individual and it's alters on each layer, and the edges that connect them on each layer.

them or using them to represent data. As is the case for edges in single layer networks, what exactly they represent is up to the researcher, and there are endless possibilities. Some analyses do not require or use interlayer edge weights (e.g., interlayer correlations or some measures of layer compressibility; De Domenico et al. 2015b), and others allow the researcher to explore outcomes across a range of interlayer couplings (De Domenico et al. 2015a).

It is a valid option to explore outcomes across a range of interlayer edge weights, and it could even be the case that if certain outcomes are already known or expected, the interlayer edge weights can be what is calculated. For example, if a researcher already knows which individuals are in true sub-groups or communities, one could compute the interlayer edge weights that best reproduce these communities from interaction data, and learn how coupled various interaction types are within the social dynamics that produced those communities. Similarly, if interlayer edge weights that connect individuals are different across individuals in the same layers (e.g., individual 1 is strongly connected across layers A and B, while individual 2 is weakly connected across layers A and B), one could potentially infer how coupled various interaction types are within the individuals that produced those communities.

If uniquely assigned to individuals, interlayer edges can help preserve more detail about how individuals are likely to behave and represent individual differences in a more detailed way. Researchers can represent "types" of individuals in ways that match the differentiation a species is able to make among group members given their cognitive abilities (Hobson et al. 2019), or in ways that reflect functionally different contributions to group dynamics. Interlayer edges could, for example, be weighted to reflect attributes of an individual's personality, and the unique weighting and direction of interlayer edges could reflect different personalities. For instance, layers in a multiplex network could represent different group-wide contexts (e.g., feeding, traveling, and resting), and all intralayer edges could be aggression. An individual that behaves very flexibly might have weak interlayer coupling if how they behave across contexts varies drastically, whereas an individual that behaves very consistently and is always aggressive might have strong interlayer coupling. Another possibility is that interlayer edges could represent temporal coupling of interactions. If one layer represents grooming and another represents aggression, a group or species that has high rates of post-conflict reconciliation, or often receives post-conflict social support, might have heavily weighted directed edges from aggression to grooming in a multiplex, while other groups or species might have lightly weighted or absent interlayer edges in that direction if grooming rarely follows aggression.

Finally, interlayer edges in a multiplex could even reflect physiological processes. For example, to understand disease transmission, a multiplex network could be comprised of layers of different types of social interactions that could spread a disease [e.g., biting and grooming (Drewe 2010)] (De Domenico et al. 2016; Finn et al. 2019), and interlayer edges could reflect an individual's likelihood of infection, or the likelihood that exposure via one behavioral type would become transmissible through another behavioral type of exposure. In such a scenario, if interlayer edge weights were high, it would be more likely that an individual could transmit the disease through a different behavioral means than it was acquired. There are of course many more possible uses of interlayer edges not listed here, and even more in a multilayer network that is not multiplex, where interlayer edges can connect different individuals across layers. What is appropriate for a given network construction is entirely dependent on the study system, data, research question, and chosen analysis.

Interactions and relationships—intralayer edges and N-grams

Depending on the data and study system, edges in a single or multilayer network may or may not represent a relationship. In social species that use many different behaviors to interact, it is unlikely that one edge type can encompass the entire relationship, but rather may represent a "type" of relationship such as a dominance relationship (Hinde 1976; Hinde and Datta 1981). Multilayer networks can contain multiple of these "relationship types" (see Figure 4B) and potentially represent an entire relationship, or at least a larger amount of a dyad's "relationship space." A set of edge weights between a dyad across multiple layers (i.e., vector of edge weight values) is sometimes referred to as an "N-gram" or "l-gram" (Atkisson and Finn 2020; Wu et al. 2020). For example, in a multiplex with 3 layers (threat, chase, bite), if individual A and individual B threaten and chase each other but do not bite each other, their relationship can be represented as the unweighted N-gram [1,1,0]. An N-gram could also be weighted to reflect the edge weights of each layer; if individual A and B threaten each other 2 times, chase each other 4 times, and bite each other zero times, their weighted N-gram would be [2,4,0]. When layers contain directed interactions, N-gram representations would need to contain multiple values for each layer to represent both directions. Say A and B both threaten each other once, A chases B 3 times while B chases A one time, and they never bite each other. The 2 directed N-grams would be: A->B [1,3,0] and B->A [1,1,0]. Alternatively, each direction could be treated as though it is a separate layer, containing an "initiate" and "receive" layer: A-> B [1,1,3,1,0,0]; B-> A [1,1,1,3,0,0] (Beisner et al. 2015; Atkisson and Finn 2020).

N-gram representations of relationships can encode the number of layers in which dyads interact in, the weights and directions of those interactions, and how these edges are distributed across layers. With this more nuanced categorization of social bonds, researchers can better assess differentiation and the degree of differentiation of relationships, an important component of social structure (Whitehead 2008; Bergman and Beehner 2015). For example, one might consider a dyad in a single layer network connected with an edge weight of 4 to have a stronger bond than a dyad connected with an edge weight of one. In a multilayer network, a dyad that has an aggression layer edge weight of 5 and a grooming layer edge weight of 0 might be considered strongly antagonistic, another dyad that has an aggression layer edge weight of 2 and a grooming layer edge weight of 1 might be considered mildly antagonistic, and a dyad with an aggression layer edge weight of 1 and grooming layer edge weight of 8 might be considered strongly affiliative.

It may be useful to consider a relationship within the context of the other relationships an individual has. A single edge can contain a direction and a weight and can be assessed relative to an individual's other edges. The proportion of an individual's interactions that are with one partner can index how important a relationship might be to an individual (of course, this may vary based on edge type). Similarly, the amount by which individuals are similarly important to each other could reflect balance or imbalance in the relationship. For instance, Figure 5A,B shows the connections of individual "Blue" (blue node) and individual "Red" (red node) in grooming networks. In Figure 5A, the connection may be more "important" to Blue, since half of blue's grooming is with Red, whereas only one-sixth of Red's grooming is with Blue, creating an imbalanced grooming relationship. In contrast, in Figure 5B, one-third of the grooming for both Red and Blue is with each other, creating a more balanced grooming relationship.

Similarly, one can assess relative importance and balance of relationships by comparing a dyad's N-grams to their other N-grams for a more nuanced characterization of importance. A relationship might be relatively important to one individual in only one or in many behavioral domains. Similarly, a dyad might have an imbalanced relationship in only one domain, or in all domains. Perhaps a relationship could even be balanced if the individuals have equivalent imbalances in different social domains. There are numerous ways these attributes could be conceptualized. For instance, in Figure 5C, Red and Blue groom each other a lot, but do not interact on other layers (huddle and aggression). Because neither have other grooming partners, their relationship still may be important, despite only interacting in one behavioral domain. In contrast, in Figure 5D, they only interact in the grooming layer and they both have many other grooming partners, so this relationship may not be very important. Alternatively, instead of being important because they are each other's only interaction partner in a behavioral domain (e.g., Figure 5C), in Figure 5D, Red's relationship with a different individual "Orange" (orange node) may be important because they interact in all layers—grooming, huddling, and fighting may indicate they are relevant to each other in many social contexts. The redundancy or uniqueness of a relationship may also contribute to how important a relationship is. For instance, in Figure 5E, the N-gram for Red's relationship for both Blue and Orange is [1,1,0]. Red has a total of 3 of these N-grams, indicating multiple relationships with the same patterning where they groom and huddle but do not fight. Thus, Red's relationships with Blue and Orange may be redundant social support. While it may be important to have redundant social support, the specific relationship with any of them may not be uniquely important. In contrast, the N-grams of all of blue's relationships are unique, and Blue's relationship with Red is their strongest exclusively affiliative relationship.

The notion of balance in multilayer relationships can also have diverse manifestations. For instance, in Figure 5F, Blue and Red's relationship may be imbalanced—Blue's connection with Red comprises a smaller proportion of their connections than Red's connection to Blue on all layers, making Blue a more important connection for Red on all layers. Alternatively, Figure 5G shows interactions in grooming, huddling, and mating layers, where their overall relationship may be more balanced—Red comprises a larger proportion of Blue's grooming, whereas Blue comprises a larger proportion of Red's huddling, and they both exclusively mate with each other. Multiple behavioral layers can allow researchers to better assess differential investment across behavioral domains, possible tradeoffs that exist in relationships, reciprocity that occurs "across currencies."

Finally, as is the nature of social networks, the connections and attributes of an individual's other connections could influence the significance of other relationships. For instance, in Figure 5H, Blue has more grooming and huddling partners than Red, and it may appear as though Red and Blue's grooming and huddling relationships may be more important for Red than for Blue. However, all of Blue's grooming and huddling partners also have many other grooming and huddling partners, whereas Red's few other grooming

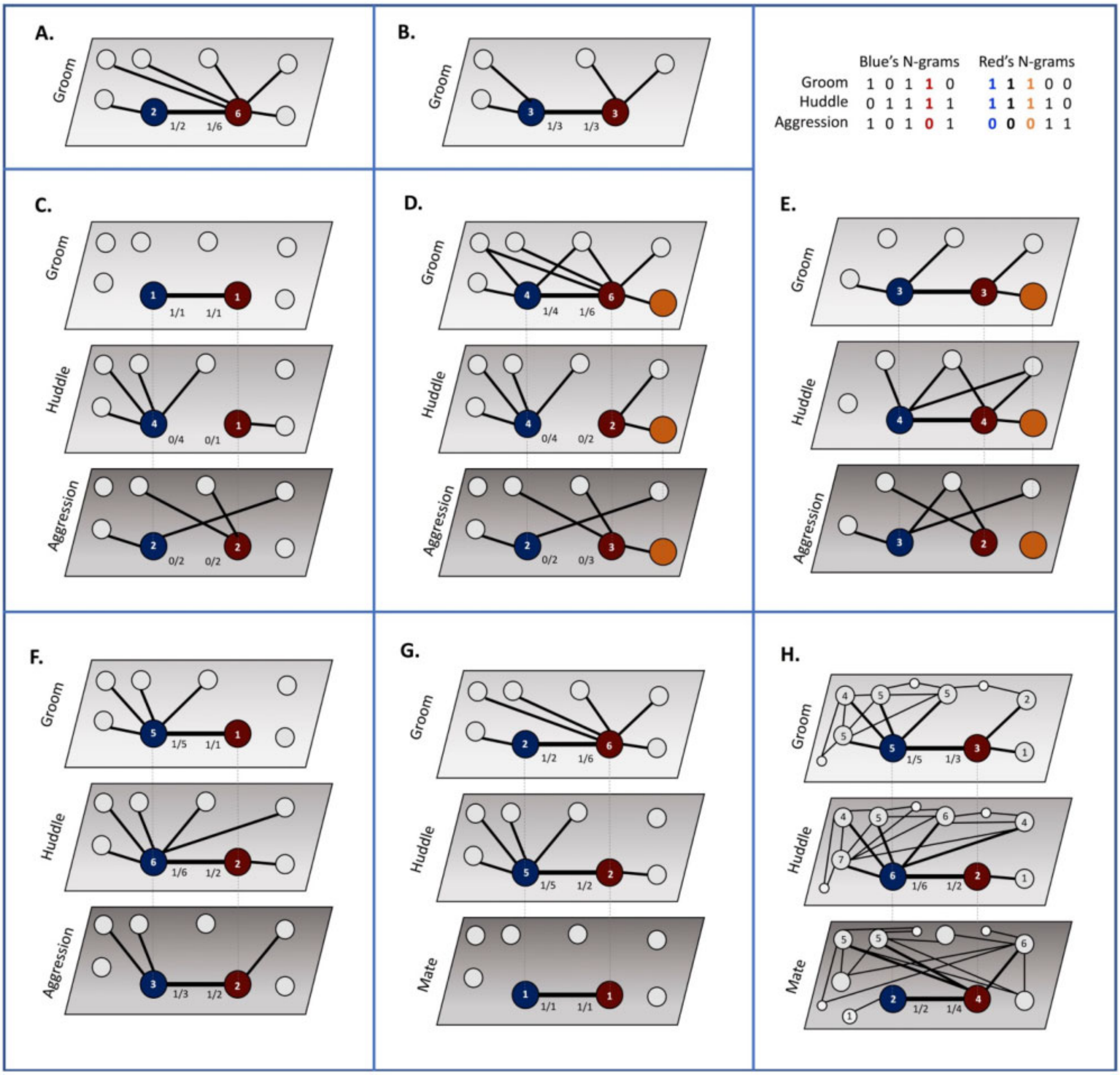

Figure 5. Examples of relationships and their relative importances and balance. Panels A and B show the connections of individual Blue (blue node) and individual Red (red node) in a grooming network, and panels C–H show their connections in a multiplex social network. Their alters are represented as either gray or orange nodes. All connections are unweighted, and the degrees of the relevant individuals are shown inside the node-tuples, and the proportion of connections the blue and red individual have that are with each other is noted below the edge that connects them on each layer when relevant. The behavioral dimension that the layer represents is labeled to the left of each layer. In panel E, the N-grams of all of Blue and Red's relationships are shown as vertical vectors, with binary values corresponding to whether or not they interacted on each layer.

and huddling partners also only have one or no other partners. This could indicate the difference between having a small number of close ties versus many weak ties, or personality characteristics such as introversion and extraversion. If Red's ties with fewer others are indeed stronger social bonds, Red's connection with more gregarious Blue may actually be a less important grooming and huddling tie. In the mating layer, it may appear as though Red is more important to Blue, since Blue only has one other mate while Red has 3 other mates. However, Red's other mating partners also have many other mating partners (5, 5, and 6), making the mating investment less certain in those ties. Blue's other mating partner exclusively mates with Blue, making it a more certain mating investment. Thus, Red's tie to blue is actually Red's most certain mating investment, whereas Blue's tie to Red is their least certain mating investment. Considering these relationship details across the entire relationship of Red and Blue in these particular behaviors, it is much less clear how important or balanced their overall relationship is, but there are also many more ways to probe that question.

Individual roles—multilayer ego networks, centralities, and versatilities

An individual's role in a society emerges from its unique patterning of interactions and relationships with other group members. This can correspond to a node's position in the overall network when the network is comprised of relevant interactions with others as edges (see Figure 4C). In a single layer network, this can be summarized many different ways including the total number of interaction partners (i.e., degree), the total strength of all of these connections (i.e., weighted degree), how diversely the edge weights are distributed across interaction partners [e.g., partner diversity in Silk et al. (2013) and Cheney (1992); disparity (Barthélemy et al. 2005), and various other centrality measures (e.g., eigenvector centrality, betweenness centrality, etc.)]. In a multiplex network, an individual's role emerges from the intersecting behavioral dimensions by which the individual is involved (Barrett et al. 2012). Again, there are multiple possible conceptualizations: an individual exists in a slice through social domains [i.e., the layers or "behavioral dimensions" (Barrett et al. 2012)], and also exists in a slice through their connections (i.e., their set of N-grams). The additional richness of a multilayer representation allows for individuals to be characterized as summaries of how they behave across different social domains, summaries of how they behave across different relationships, or combinations of the 2.

Considering just an individual's node tuples and its connections to others (i.e., alters) on each layer (e.g., Figure 4C), we can consider a vertical slice (through layers) and compression (of relationships) of an individual's social space into a vector of length L (where L is the number of layers), filled with values of any of the above individual measures calculated on each network layer. For instance, a vector of an individual's degrees from all layers is considered the multidegree (Kivelä et al. 2014). As the diversity of social domains may be a relevant component of how a group is structured (Whitehead 2008; Ramos-Fernandez et al. 2018), measures of variance across these "vertical" values can be informative about how consistently or diversely an individual invests interactions to different social domains. Alternatively, we can consider a horizontal slice (through relationships) and compression (of layers) of an individual's social space into a vector of length A (where A is the number of alters), filled with values of any of the above measures calculated on N-grams. For instance, an individual could be described as a vector of the number of layers an individual interacts in for each of its alters. As social differentiation may be a relevant characteristic of how groups are structured (Whitehead 2008; Bergman and Beehner 2015), measures of variance across these "horizontal" values can be informative about how consistently or diversely an individual invests interactions toward different individuals.

There are numerous ways one could calculate the distribution of an individual's interactions across their whole ego multiplex network (i.e., the connection of an individual and their alters), that compresses and summarizes across both layers and relationships in various ways, all offering different possible interpretations of the diversity and redundancy in their social interactions and relationships. For example, in some data, unique N-grams can be conceptualized as distinct "types" of relationship. The entropy of an individual's interactions as they occur across different types of relationships [i.e., how diversely an individual's interactions are distributed across relationship types or "weighted N-gram entropy" (Atkisson and Finn 2020)] could index how evenly or consistently an individual's interactions are spread over different relationships types, or if they tend to invest more in or preferentially associate within one type of relationship.

Such a measure is conceptually similar to a multilayer disparity measure, and the Individual Level Relationship Diversity (ILRD) proposed by Fischer et al. (2017), with subtle differences. They all describe patterns of interactions as relationships and assess the distribution of an individual's interactions across relationships. N-gram entropy coarse grains relationships into relationship types (i.e., N-grams), while disparity counts each interaction partner as a different "type," and ILRD identifies unique relationships types with a multi-step process, performing a clustering analysis on values computed for multiple dyadic level summary statistics from multiple behaviors. N-gram entropy then measures how diversely an individual's interactions are distributed across relationship types using Shannon entropy, whereas disparity measures the spread of interactions across unique interaction partners using variance (or perhaps in a multilayer version, a vector of spreads, or the variance of vectors), and ILRD measures the extent to which one relationship type dominates and individual's relationships using a version of the Simpson diversity index. Such subtle differences across measures may be important when tailoring an analysis to a specific question and study system.

In another example, the amount of overlap in edges within relationships across layers could be calculated, and an individual could be represented by a vector of the edge overlap values across all layer combinations, or a single overlap measure. Constructing the multiplex ego network of individuals is a great way to begin conceptualizing which of the numerous components of an individual's social situation one might want to capture for individuals in their unique dataset, and how these components could be compressed into a summary statistic to represent a particular dimension of sociality that address their research question.

For centrality measures that extend beyond an individual's immediate ties (e.g., eigenvector centrality where the connections of connections are also considered), there are numerous aggregation methods that have been used to describe an individual's role across multiple layers, that summarize the structure at various stages of compression [see Supplementary Materials 2 in Finn et al. (2019) for a discussion of some of these methods]. Some methods calculate a multilayer centrality by calculating a centrality measure on each layer, then aggregating this measure across layers into one summarizing value (Kivelä et al. 2014). When layers represent different types of interactions, it may be the case that different centrality measures might represent which individuals are "important" to the group. For example, a Borda count (de Borda 1781) or Kemeny aggregation (Kemeny and Snell 1962) has been used to rank individuals combining centrality measures across layers where layers use either eigenvector, degree, or betweenness centrality (Pósfai et al. 2019; Beisner et al. 2020). A modified Borda count has also been used to calculate "meta-centrality" to identify "super-spreaders" by combining multiple centrality measures calculated from the same layer, aiming to create a measure that is more generalizable across different networks (Madotto and Liu 2016).

In contrast to aggregated centralities where values are calculated separately for the position of node tuples on each separate layer before aggregation, multilayer versatilities aggregate values calculated for the position of node tuples in the full multiplex structure (Kivelä et al. 2014; De Domenico et al. 2015d). For instance, multilayer Page Rank centrality calculates Page Rank for a node-tuple with the steady state of a random walker through one layer, then aggregates these values for all an individual's node-tuples. In contrast, Page Rank versatility calculates Page Rank for a node-tuple with the steady state of a random walker across the whole multilayer network, moving from node to node over both intra and interlayer edges, then aggregate these values for all an individual's node-tuples (De Domenico et al. 2015d). The centrality version does not use

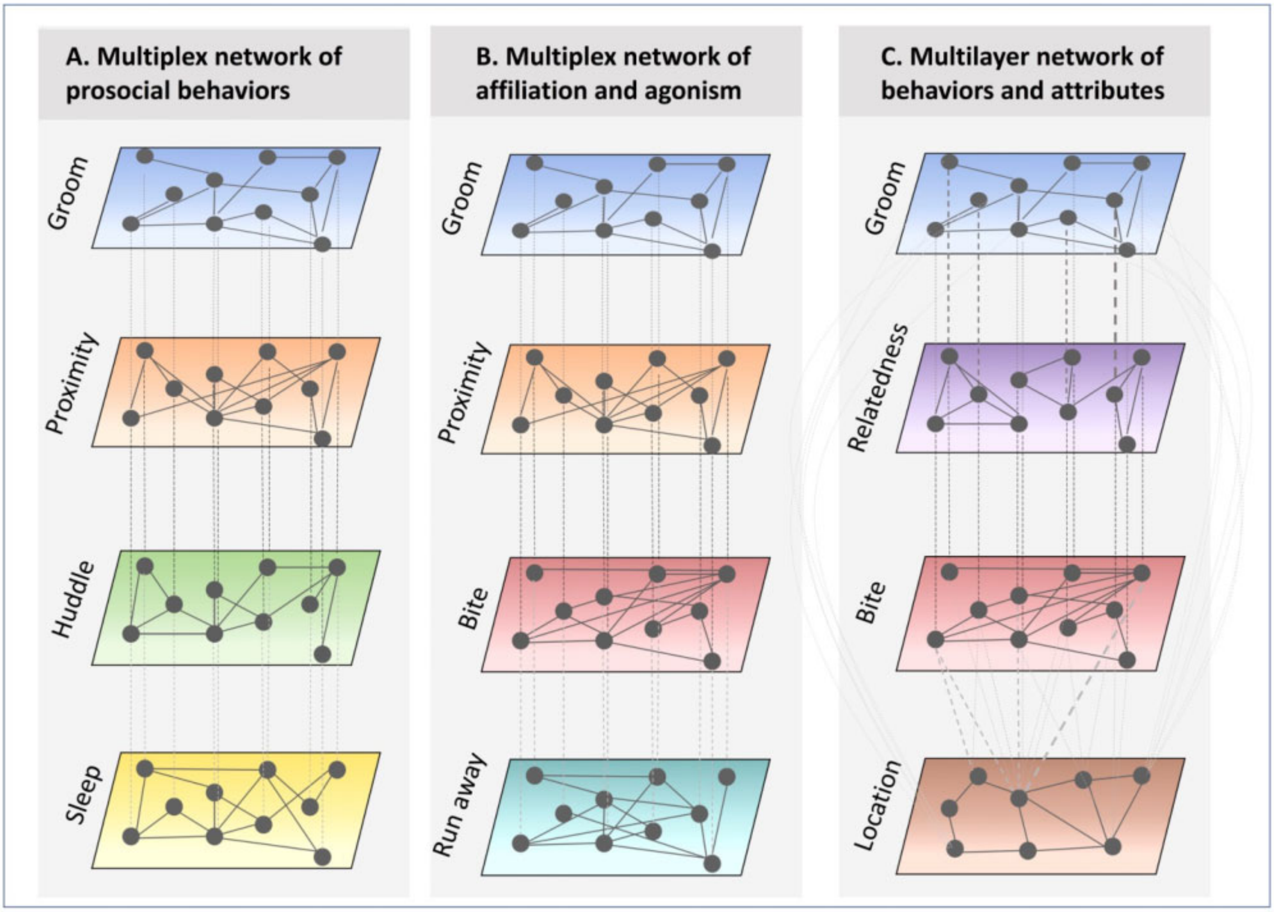

Figure 6. Different types of layers. Three hypothetical multilayer networks are represented in panels A–C. The network in panel A is a multiplex network that has layers which reflect various prosocial behaviors. It could represent the affiliative social structure of a group and its structure could feasibly be measured as such. The network in panel B is also a multiplex network but includes layers of both affiliative and agonistic behaviors. Because the behaviors are functionally different and sometimes of opposing functions (e.g., bite versus run away), it is less clear what pathways through or combinations of the layers would be representing. Other measures such as layer coupling could still be informative to understand how layers relate to or influence each other. The network in panel C is a multilayer network that includes affiliative and agonistic behaviors, relatedness, and a location layer. This network has some interlayer edges connecting individuals on interaction layers to location nodes where that behavior occurred. It is less clear what it's overall structure represents, but it could still potentially be summarized in ways that are informative about who does which behaviors where, and how much influence there is from family structure.

interlayer edges, so it may be easier to account for network differences such as density so that one layer is not disproportionately represented. In addition, the centrality version avoids potential confusion regarding what "flow" across network layers represents, when layers represent interactions that are functionally very different. While the versatility version does require more careful attention to weighting of edges for different layers and may require more careful interpretation, it allows for the opportunity to incorporate interlayer edges, and is actually measuring the structure across the entire multiplex network.

For example, Figure 6 shows 3 different multilayer networks that have 4 different sets of layers. The first is a multiplex network that has different prosocial or affiliative behaviors on each layer: groom, proximity, huddle, and sleep. A multilayer versatility measure could be useful on this network for questions regarding which individuals are of greatest risk for disease spread. Because the edges on each layer represent a behavior functionally related to physical proximity related to disease spread, flow across layers could make sense. A versatility measure could be useful for questions regarding which individuals are most important for social cohesion, though if edges were weighted, it would be less clear what interlayer edge weights would represent. In the second network, the function of edges in each layer is more distinct from each other, as both affiliative and agonistic behaviors are present. Flow across layers makes less sense, but a multilayer centrality measure could still be useful to indicate the influence or prominence an individual has in a group. Other individual descriptors using N-grams could also be informative here. The last network is not a multiplex, as one layer represents a spatial network of locations where certain behaviors occurred, in addition to behavioral layers and an attribute layer of relatedness. Here, neither centrality nor versatility measures would be interpretable.

Transmission and flow—multilayer pathways

In a single layer network, social dynamics arise from the interactions between individuals and as a result, meso and global network structures emerge. Thus, we can identify patterns between small sets of individuals, and/or patterns across the whole network, that may be informative about the underlying dynamics that led to that structure. When a network is not completely connected nor randomly connected, there exist potentially meaningful differences in how certain individuals are connected to each other, either by direct links (A -> B) or longer indirect pathways that span across the connections of multiple individuals (A -> C -> B) (see Figure 7A; Wey et al. 2008; Pinter-Wollman et al. 2014).

Some analyses use network pathways to detect clusters or communities where there is a set of individuals where most are connected to each other with direct or short pathways (Krause et al. 2007; Wey et al. 2008). Other analyses use pathways to assess global properties such as acyclicity, which reflects the general direction of pathways across a whole network and relates to social structures like dominance hierarchies (Finn 2019). Indirect pathways in a network could index indirect relationships between individuals formed via transitive inference (Beisner et al. 2016), if for instance, an individual observes a dominance interaction between its interaction partner and a third party, and is able to infer its dominance or subordinance to the third party. Cognitively sophisticated forms of reasoning such as this may determine which interactions in a group are likely to influence an individual's behavior, and which interactions one may care to quantify. In a multilayer network, the number of pathways that exist in a network by which individuals could be connected increases with multiple network layers, and increases even more when interlayer edges exist, though care should be taken that all edges that are measured make sense as a plausible pathway in flow-based measures.

Some network analyses use a random walker to move across edges in the network to infer characteristics of its structure based on how often the random walker reaches certain nodes and how often certain pathways are traveled. Often, which direction a random walker goes is determined probabilistically by the weights and directions of both the intralayer and interlayer edges (if weighted and directed) (De Domenico et al. 2015a). Should the random walker be more, less, or equally likely to travel to the node-tuple of the same individual on a different layer, compared with other nodes in the same layer? For such measures, and even for measures that don't explicitly employ a random walker technique but assess characteristics about the "flow" (e.g., of information or disease transmission) through edges (e.g., measures of acyclicity, network diameter, betweenness centrality, and closeness centrality), thinking about a random walker traveling across pathways is a fairly intuitive way to consider how interlayer edges might influence certain multilayer network measures, and help guide thinking about if and how interlayer edges should be incorporated into pathways that are used in an analysis. As described for centrality and versatility measures in Figure 6, pathways that span multiple network layers may or may not be appropriate, as they may or may not have sensical interpretations. Edges from different layers and edges between layers need to match in a coherent way for a pathway to be useful.

Subgroups—multilayer communities

Subgrouping is a common focus of how groups are structured (Whitehead 2008; Grueter et al. 2012; Ramos-Fernandez et al. 2018). For instance, subgroups in a social group could be characterized by individuals that interact in different social contexts, and could do so consistently or with high variability (Ramos-Fernandez et al. 2018). Others have described subgrouping or communities within a nested or modular social stratification (Grueter et al. 2012). Communities can vary in size, number, or in how differentiated they are from each other (Whitehead 2008). Multilayer community inference methods could facilitate quantifying these components of social structure, by assessing community structure over layers that represent interaction types or time windows (e.g., number of communities, number of layers by which node-tuples are from within communities, and number of communities an individual's node-tuples are assigned to).

Some community detection algorithms have been generalized to multilayer networks (Kivelä et al. 2014; De Domenico et al. 2015a). The community detection algorithms (or perhaps more accurately called community inference or community assignment algorithms because they do not necessarily *detect* a ground truth) that exist for multilayer networks allow a researcher more flexibility in deciding how communities should be assigned, as the coupling between layers can be adjusted (for an example of this, see Supplementary Material 2 in Finn et al. 2019). Briefly, in multilayer community assignment, node-tuples of the same individual can be assigned to the same or different communities (see Figure 7B), and the coupling between layers can determine if it should be more strongly biased to group node-tuples of the same individual into the same community, or more strongly biased by the intralayer edges a node-tuple has with other individuals on each layer. As such, the researcher can decide if they want to assign node-tuples to communities while emphasizing differences in relationship types or emphasizing consistent subgroups of individuals (Supplementary Material 2 in Finn et al. 2019).

For instance, the first multiplex in Figure 6 has functionally similar behavioral layers (affiliation), so one could assign strong coupling between layers (de-emphasizing differences across layers), and use community assignment to identify different social cliques. Alternatively, in the second multiplex in Figure 6 where the layers are functionally different behaviors, one could assign weaker coupling between layers (emphasizing differences across layers) and use community assignment to identify subgroups more representative of different relationship types. One could even more creatively use community assignment in the third multilayer network, where in addition to 2 behavioral layers, there is a layer with nodes of locations, and a layer of edges that reflect relatedness. While inferred communities may not represent subgroups in the same capacity, the composition of communities could be informative about who is doing what, where, and with whom. For example, if one location is grouped with a larger number of node-tuples from one of the behavioral networks, it could reflect that certain interactions or contexts are linked to a location. If individual's node tuples from the relatedness layer are more likely to be in the same communities as their node tuples from the grooming layer, it could represent that grooming is more related to nepotism.

Group connectedness and fragmentation—multilayer clustering

Group connectivity is another common description of social structure (Lehmann and Dunbar 2009; Lehmann et al. 2012). For instance, some social structures are characterized by having fragmented subgroups while still maintaining overall social cohesion (Lehmann and Dunbar 2009). Such attributes could be indexed by both measures of clustering and community assignment. However, clustering and presence of distinct communities do not necessarily co-vary—clustering indexes "connectiveness" of subgroups, instead of who is in each subgroup or how many exist. Clustering can be measured locally (i.e., to what extent is an individual in a strongly clustered region) or globally (i.e., to what extent are individuals in the group clustered), often assessing the number and density of triangles (where 3 individuals are all connected to each other) in regions of the network. Some clustering measures have been

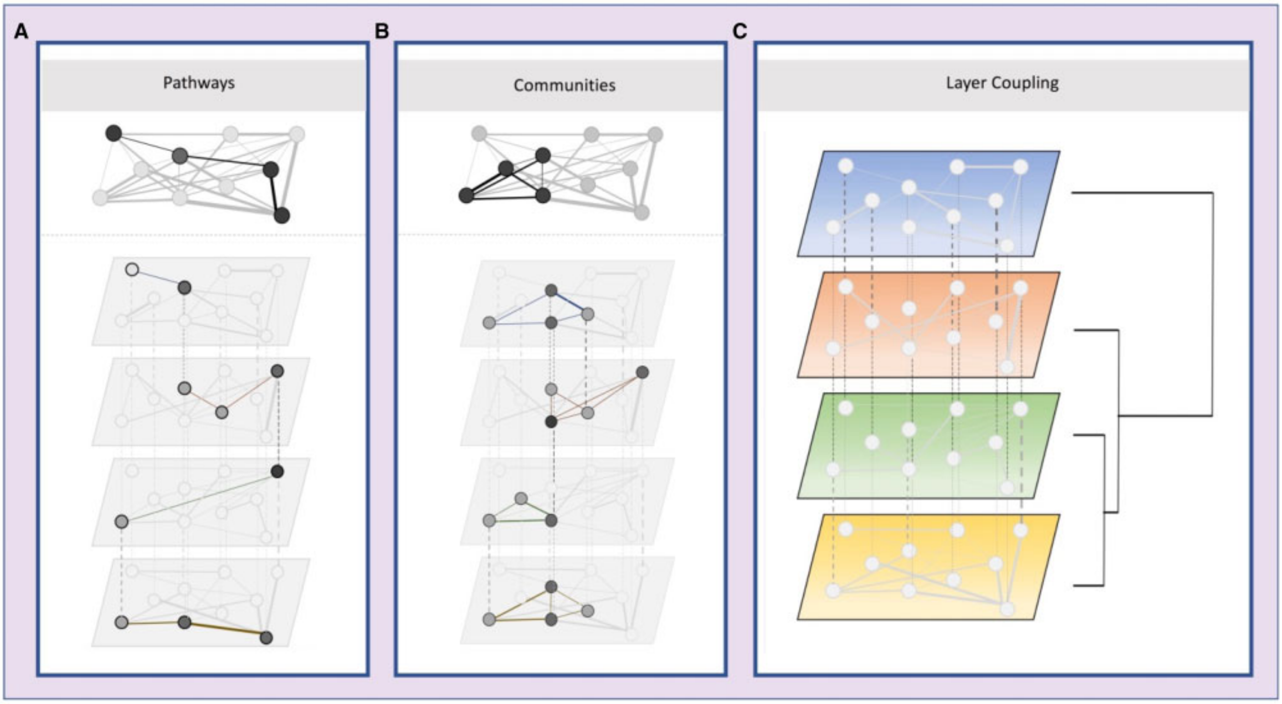

Figure 7. Larger scale structure arising from social processes. Structure arises in subgroups and across the entire network from the dynamics of interacting individuals. In a single layer network, nodes can be connected directly or through pathways (panel A, top), and in multiplex networks, there are many additional pathways that connect individuals that span across both intra and interlayer edges (panel A, bottom). For communities of more connected sets of individuals, individuals are usually assigned to only one community in single layer networks (panel B, top), whereas node-tuples of an individual can either be all in the same communities or in different communities in a multiplex network (panel B, bottom). Layer coupling can only be measured with multilayer network layers (panel C). Layers can be grouped based on how similarly structured they are, or how strongly connected they are to each other (e.g., the dendrogram in panel C indicates which layers are most similar to each other).

generalized to multilayer networks (Kivelä et al. 2014). Still, care should be taken in how they are used to ensure they are capturing structure in the network that is meaningful. For instance, does it make sense for clustering to also reflects how strongly node tuples of the same individual are connected to each other, as well as how closely they are connected to others? This may or may not make sense for different datasets.

In a multiplex social network, clustering measures could be useful in a number of cases by measuring how individuals are clustered based on sets of behaviors instead of just one. For instance, if there are 2 similar behaviors (e.g., huddling and grooming) that both reflect the same general social domain (e.g., affiliation), and interacting with both indicates stronger bonds than only interacting with one, clustering measures of a multiplex network of both behaviors might be a better representation of how "cliquey" a group is. This would be especially true if one of the behaviors sometimes serves other purposes in the absence of the other [e.g., grooming for political reasons (Wooddell et al. 2019)]. See Section 4.2.3 in Kivelä et al. (2014) for a review and discussion of multilayer clustering methods, and numerous other recent clustering analyses for multilayer networks (Chen and Hero 2017; Chen et al. 2019; El Gheche et al. 2020).

Behavioral dimensions and domains—layers

Layers are of course what allow for the more detailed descriptions of individuals and relationships in multilayer networks. Different interaction types, time periods, seasons, or contexts (e.g., interactions during provisioning versus not) can be separated into different layers such that they are not conflated, yet still exist in one structure. Different time windows can be represented as different layers (or different aspects), reflecting different group memberships or changes in relationships. This can allow researchers to differentiate between stable versus unstable relationships, or long-term social bonds versus brief interactions. The number of layers also indexes the number of dimensions by which relationships can differ—more socially sophisticated groups may have more types of interactions (Barrett et al. 2012), more contexts in which they interact (Freeberg et al. 2012), and/or greater use of cues or signals (Anderson and McShea 2001), all of which can be represented as layers.

What constitutes a distinct layer? Especially when there are many different presentations or levels of a specific behavioral type (e.g., intensities of aggression or submission), it may not be clear which behaviors should be grouped together or separated into discrete layers. This can also be an issue in single layer analyses, though it is no longer an invisible problem when you must actively decide what should be multiple distinct network layers. For example, even within the same species, which behaviors are used to create an aggression layer can vary quite a bit [see Finn et al. (submitted for publication) for a discussion about this in Barbary macaques], making results across studies less comparable. How then, can one decide which behaviors are functionally similar enough to be in the same layer, or different enough such that they should be kept separate?

While some methods exist to aid such decisions, blind application of them is unlikely to yield results that are optimal for your unique dataset and study. Methods such as layer reducibility (De Domenico et al. 2015b) are designed for a mathematical optimization of distinctly different layers that may or may not match the aims of behavioral researchers, which are probably to use layers of behaviors that serve different biological or social functions. It is likely important to combine multiple behaviors that are functionally equivalent or redundant to the study species (Hobson et al. 2013), not just in their network structure. Two layers might show very similar structure that might be optimally combined based on some algorithm, but it may be the case that based on knowledge of the study system, the behaviors are functionally different, in which case one may want to keep them separate. That 2 layers of functionally different behaviors are structured similarly may even be informative about that social group.

It is also important to understand what parts of the network are being used in methods that aim for some optimization, so that one can consider whether or not those network structures are functionally important or what one may want to optimize. Some methods may not factor in the weight or direction of edges (De Domenico et al. 2015b), which could be very important for your networks. For instance, a method might use percent of edge overlap to decide if 2 layers should be combined, but perhaps you are making decisions about agonistic behaviors, and care more about how similar the direction in the flow of edges across the networks are more than you care that the same animals used the behaviors.

Decisions about combining or separating behaviors into separate layers should be made considering both your knowledge about the species/study system, properties of the networks, and what those properties mean given your species/study system. One may need to use a combination of species knowledge and descriptive network statistics to make decisions, and those decisions might vary for different research questions. See van der Marel et al. (2020) for a framework and suggestions to guide these decisions.

Behavioral covariation—interlayer coupling

Within the same social group, networks of different types of interactions may be structurally very similar or distinct from each other (Lehmann et al. 2012; Beisner et al. 2020), and therefore measures of layer coupling could be useful for characterizing a group's social structure or social dynamics. For instance, some species may use different behaviors for different goals or have context-specific social investments (e.g., association choices may decrease competition, while grooming choices may reestablish relationships) (Lehmann and Boesch 2009). Multiple network structuring mechanisms such as this would then likely generate layers with very different structure.

Interlayer coupling could be used to describe a layer, pairs of layers, or the overall network, and can be based on either connectivity with interlayer edges, or similarity of intralayer edges. If one layer is more coupled to all other layers, it could be a keystone network that disproportionately influences other layers and overall group dynamics (Fushing et al. 2014). Measures of layer coupling can also be used to aid decision-making about which layers to use during network construction, as descriptive statistics to describe the similarity of layers.

An index that uses interlayer edges reflects coupling based more on attributes such as flow between layers, strength of connection between layers, or "dynamical spillover" (Vijayaraghavan et al. 2015). In the same way that single or multilayer networks can be described with a distribution of node-level measures (e.g., degree distribution or distribution of a versatility measure), sets of network layers or an entire multilayer network can be described as the diversity or distribution of interlayer edge weights between layers. For instance, if 2 layers are more strongly coupled than others, their interlayer edge weights might be skewed toward higher values more than other sets of layers. There could also exist differences in the variation of interlayer edges between layers—two layers may have relatively consistent interlayer edge weights or directions connecting nodes across layers, whereas the interlayer edges weights connecting nodes across another 2 layers may widely vary.

Interlayer coupling can also be described not by how strongly layers are tied by interlayer edges, but by how similar their intralayer edges are to each other. Such measures are based more on attributes such as similarity, overlap, or correlation between layers. A network could have layers that are equally different from each other, or some that are more similar to each other (see dendrogram in Figure 7C). There are a number of approaches that could be used to assess similarity between layers, many of which are compiled into lists elsewhere [4.2.5 in Kivelä et al. (2014); Evolutionary models section and Supplementary Materials 1 in Finn et al. (2019); De Domenico et al. 2015c]. Some methods and implementations can even group layers based on how similar they are to each other, similar to a community assignment of layers [see Kao and Porter (2018), Stanley et al. (2016), and the implementation of both layer correlations and layer reducibility (De Domenico et al. 2015b) in Muxviz De Domenico et al. (2015c)].

Social situation—the global multilayer network

An entire social group can be described with summaries of the distributions of any of the smaller social units described above, combinations of them, or measures of the entire structure. For instance, the distribution of social roles in the group could be homogenous, diverse, or skewed (Whitehead 2008), and the distribution of a multilayer centrality or versatility measures could capture this, to the extent that the measures reflect functionally different roles in a social group. More globally, the social structure could be indexed by how individuals and interactions are organized across the entire multilayer network (Barrett et al. 2012).

The full structure of a multilayer social network describes a social group with the patterning at which all of the smaller social units are organized, capturing characteristics of interest such as redundancies in a social system (Anderson and McShea 2001), how diversely individuals interact, and the degree to which individuals preferentially associate with others (Whitehead 2008). Considering the number of layers and edge attributes (weights, direction), one can think about all the different possible combinations of interactions individuals could have—when all possibilities exist across all individuals, the network is maximally entropic, such that probabilistically it is completely uncertain who is going to interact and in what capacity, since each interaction is equally likely. This would not necessarily be the most "complex" social structure, as one could argue that everyone would then have the same relationship (Ramos-Fernandez et al. 2018). In contrast, if no individuals interact, the network would be minimally entropic (Barrett et al. 2012), which one could argue is also not very complex. As there are countless properties of a network that could be measured, it is not likely that one measure will wholly reflect everything a researcher wishes to index about a social group, or even perfectly map on to one notion of sociality one may have (e.g., complexity or uncertainty in the example above). Instead, one may need to carefully build a toolkit of measures that are appropriate for their properties of interest.

Multilayer Social Network Analysis in Practice

The above sections aim to help researchers studying social behavior navigate the multilayer network structure, conceptualize their study system as a multilayer network, and begin thinking about which components of sociality could be measured from that representation. The remainder of the paper serves to help researchers think deeply about the measures they use and offer tips to reason through the decisions that must be made when using multilayer social network analysis.

The sheer number of necessary decisions and lack of standardized approaches may seem daunting, and maybe the approach appears underdeveloped. However, in some sense, this is true for all statistical approaches. The choices we make when collecting, processing, and analyzing data (e.g., the ethograms, the sampling, sample size, units/metrics, behavioral proxies, assignment of variables, control of confounds, choice of statistical analysis, statistical models, and choice of nulls) or "researcher's degrees of freedom," for any analysis matter, and it's not the case that time tested analytical approaches and conventions are necessarily best practices [e.g., p-hacking (Smaldino and McElreath 2016; Munafò et al. 2017)]. Because more detail from one's data is being used, there may be more decisions to be made. However, these decisions are not inherently any different than the types of choices researchers make during any other analysis—it is just the case that they are not always thought about, and for these analyses they must be.

I do not believe there will be—and do not think there should be—plug-and-chug network analyses. What social networks (both single and multilayer) represent is unique to a given study system and dataset (Wey et al. 2008; Farine and Whitehead 2015). As such, the analysis one uses on any network must be applied uniquely and interpreted uniquely to a given dataset and study question (James et al. 2009). This is even more so the case with multilayer networks, as the representation gets more complicated with each additional node or edge type. We can begin creating guidelines to follow, and collaborations with network scientists can of course be helpful. However, nothing can replace the researcher (1) knowing what their data represents about a real world system, (2) understanding what the multilayer network of their data represents in that system, and (3) knowing what a measure or analysis is assessing about that structure. Without these things, one cannot be confident in any interpretation of a measure. It is vital that whoever constructs the network and runs the analysis has a very clear idea of what the network represents, and what about that structure the analysis is capturing (Farine and Whitehead 2015). Here, I discuss some of the common pitfalls one faces using single layer network analyses that are exacerbated with multiple layers, and then outline some of the pitfalls one may face that are unique to multilayer networks.

Exacerbated pitfalls of general social network analyses

Forming an analysis plan

One's approach to multilayer social network analysis might be different if they are starting from the very beginning where they are designing a study including the data collection, versus trying to use multilayer network analyses to learn something from an existing dataset. However, regardless of the starting point, the following 3 things need to both be translated from the study system to a multilayer network (or vice versa) and must all conceptually match: the research question, the data, and the analysis (see Table 1 and Figure 8). A network must be constructed from data that can be measured in a way that is informative about the study system. While this is true for single layer network analysis, this is more difficult with multiple network layers, because they often contain multiple types of data in different layers which need to be measured together in an interpretable way.

Mining versus designing a network. If you are designing a study where you expect you will use multilayer network analysis, you can formulate your study question, decide on the analyses to answer the question, and then design the network you would need to answer the question (starting point at The Research Question in Figure 8). Then, you can collect your data to exactly reflect the characteristics of your study system needed for the analysis to answer your question. While this is ideal, it is often the case data have already been collected to answer a question without considering the network construction or what specific analysis will be used. In this case, more care is needed to make sure the network constructed from the data are a good representation of the component of sociality they are interested in, and that the measure captures this component.

Sometimes, a researcher may even begin with an interesting multilayer network dataset but lacks a specific study question (starting point at the data in Figure 8). It is likely the case that one can still learn many things from their data with multilayer network analysis from this starting point, but there may be limitations to what one can do. Even more care should be taken to make sure they understand what the data represents about the study system and the analysis captures something about it that is interpretable. We even see researchers, usually from quantitative fields, begin with a novel method, then search for a dataset they can use it on (starting point at the analysis in Figure 8). The analysis trajectory that is most prone to a mismatch is when data are explored as a multilayer network, without a particular question, as the interpretation of a result for any pre-existing analytical method on a pre-existing dataset may be unclear, or even inappropriate for the dataset. Despite the starting point, it is important to consider the research question, the data, and the analysis in relation to each other and ensure they match in a coherent way.

When beginning with questions and a dataset (or even a question about a study system), it may be the case that one needs to develop new measures to capture the part of the system they wish to assess. Many existing methods are published in quantitative journals by quantitative researchers, and the measures are not necessarily designed for a certain study system. As a result, they may conclude a general interpretation that may or may not be true for your data and question [e.g., optimal layer reducibility methods (De Domenico et al. 2015b) discussed above].

Measures versus classes of measures

Sometimes classes of measures are touted to have a general meaning (e.g. centrality measures, communities in community detection, measures of modularity, and measures of assortativity). In reality, there are multiple different analyses that fall into these classes that sometimes measure different parts of a network (Farine and Whitehead 2015). For example, there are many different measures of "centrality" (e.g., betweenness, page rank, and eigenvector) that can all index the "importance" of an individual or "how central" it is. However, each of these captures a different characteristic of how edges are patterned in a network. Similarly, community detection algorithms assign individuals to communities, but countless methods exist that determine communities in different ways, some of which may assign communities that are similar or different to the types of subgroups you may be interested in Fortunato and Hric (2016).

Table 1. Forming a coherent social network analysis

	The research question	The data	The analysis
The research question	What is the thing you are trying to learn or identify about your study system? Over which behavioral domains, time scales, or social scales does it exist? If your system was represented as a network, how would you phrase this as a network question?	The data must be constructed into a network to contain the structure and information necessary to answer your question. Is the thing about your study system that you are interested in measuring represented and contained in a network of this data?	Could this measure act on the parts and social scale relevant to your research question? Could an interpretation of analysis outcome answer your question? What would you learn from different values of this measure about your system?
The data		You must be able to construct a network from your data that makes sense to your study system. What can the interlayer and intralayer edges reflect? What are the nodes? What are distinct layers? What parts of your system does this network represent?	The measurement must make sense to the data. What parts of a network is the analysis combining and summarizing? What data do those parts represent? Does it make sense to combine them in this way?
The analysis			Which measurement or analysis are you going to use, and what part of a multilayer network does it measure? What does it capture about how nodes, edges, or layers are related to each other? What is the range of possible results and what would they mean?

Three important components—the study question, the data, and the analysis—must all match each other within the context of a study system and within the context of a network structure. There can exist mismatches between any 2 of these components, or between how all 3 fit together. Each square in this table gives examples of questions, to assess that each of these components is clear and combinations of them are coherent.

In addition, what a measure functionally represents could be quite different with different edge types (e.g., the same centrality measure in an agnostic versus an affiliative network) (Wey et al. 2008; Farine and Whitehead 2015; Beisner et al. 2020). While many datasets can all be represented by a network structure, the analyses done on that structure may not have the same interpretations across datasets. It may appear that some network measures have a "meaning" (e.g., importance of an individual to a group), but there is no functional interpretation of any network measure that exists absent the study system and data, nor do they generalize across all datasets.

The temptation and risk of overgeneralizing across a class of measures may be greater in a multilayer network, as the outcome spans multiple behavioral domains. For instance, it may seem that a multilayer versatility may reflect even better how important or central an individual is, as it considers multiple behaviors; or, it may seem that the communities that node-tuples are assigned to may be closer to a ground truth, since it includes multiple behaviors and it doesn't constrain an individual to only one community. However, the reality is that there is still a lot of variability across values generated by different measures within a class of measures. For instance, Muxviz (De Domenico et al. 2015c) contains implementations of 13 different multiplex centrality or versatility measures. Figure 9 shows the values from all of these measures calculated on the same 2-layer multiplex network of grooming and associations in a baboon group, with data from Franz et al. (2015a, 2015b) (also analyzed for Page Rank versatility in Finn et al. 2019). While Finn et al. (2019) showed that individuals were ranked differently using Page Rank on the separate layers, aggregated network, or multiplex, individuals are also ranked differently using these different centralities measures on the multiplex. Even though the layers are both related to affiliation, and all measures are centralities that more or less represent how "central" an individual is, which individuals have the highest values is far from consistent. The reality is also that it may be more difficult to match these multilayer analyses to both a dataset and a question in a sensical way, and one must have a good understanding of how each measure is calculated and what their data are in order to decide which one is most appropriate. All interpretations of network measures are relative to what the data are, so the interpretation of multilayer measures is relative to the unique combination of multiple types of data that make the network.

Designing a measure. As mentioned before, there is a decent chance that there does not exist a measure or analysis that well captures the component of social structure one wants to quantify. This is not necessarily a problem and may in fact be a great opportunity for a new collaboration and/or an additional methods paper. It is often the case that network scientists are willing (even eager) to collaborate with social scientists that have data and intimately know a real-world system. If a researcher conceptualizes their system and data as a multilayer network, identifies the scale of sociality they are interested in, and describes what parts or characteristics of the network they want to quantify, it may become clear how they should compress and summarize the data. If not, chances are that by communicating these things to a network scientist, together they will be able to construct an appropriate measure. As network science and multilayer networks have become increasingly popular in quantitative fields as well, it is likely there exists such a scientist in the physics, mathematics, or computer science department at most universities. Social network analysis is intrinsically an interdisciplinary endeavor, and cross disciplinary collaborations should be the norm.

Statistical significance and randomization

Network analysis often requires additional statistical techniques such as network randomizations to account for structure that may exist merely because the data are being represented as a network. It can be a challenge to determine what is an appropriate

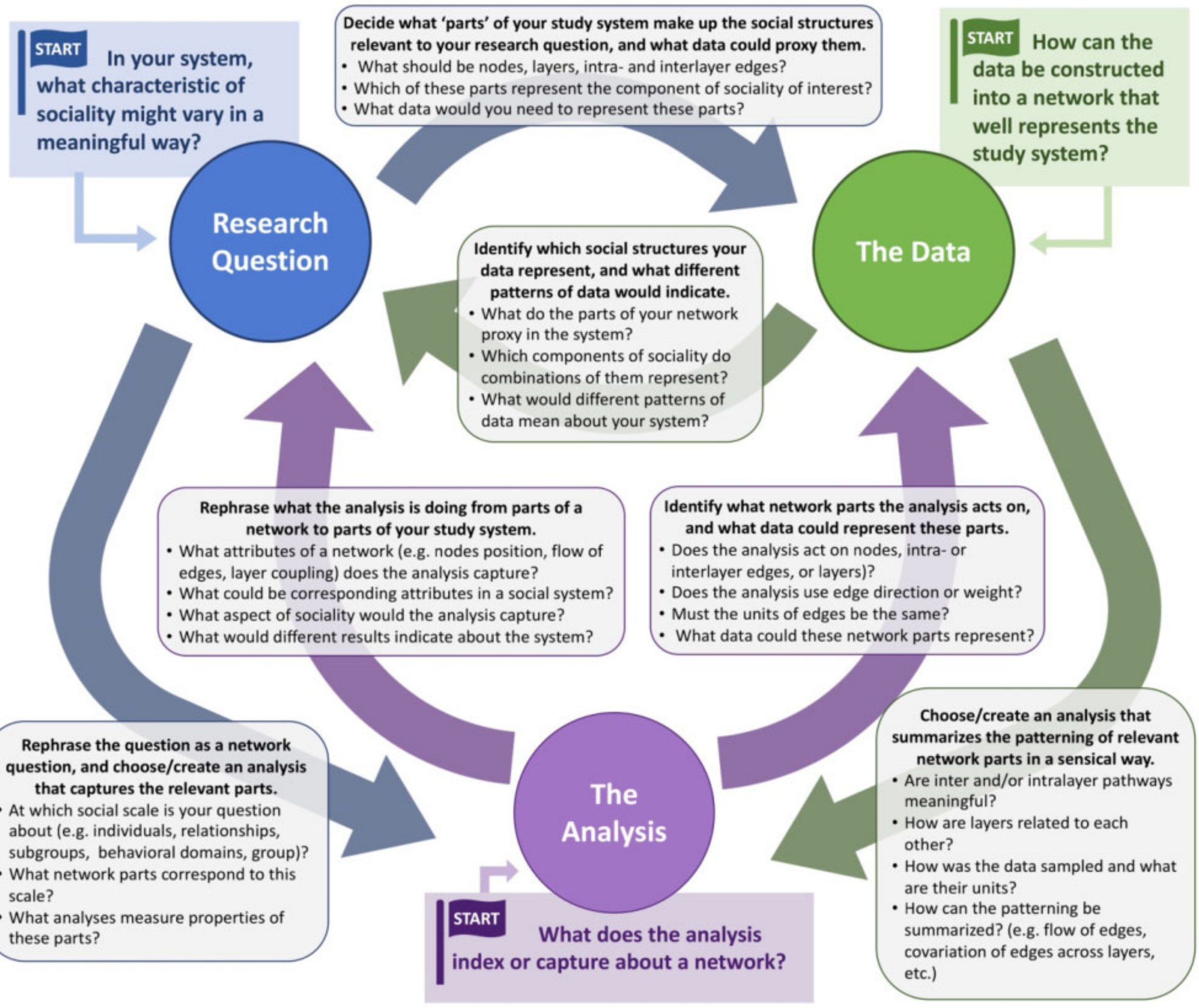

Figure 8. Flowchart of an analysis plan. Ideally, multilayer social network analysis begins with a research question, though in practice researchers sometimes start with a dataset or a novel analytical method. This flowchart outlines some of the important questions and considerations researchers should be able to explain, from various starting points. Regardless of the starting point, it is important that the research question the data, and the analysis conceptually match.

randomization even in a single layer network, though it is still important to make sure measures are not merely an artifact of the data being structured as a network, or an artifact of how the data were sampled [see Farine (2017) for an overview of randomizations as null models for animal social networks, Fosdick et al. (2018) and Newman (2018) for more about network randomizations, and Supplementary Materials in Finn et al. (submitted for publication) for an example of designing randomizations on a dataset, and Hobson et al. (2020) for an extensive discussion of reference models]. Which parts of the networks one should randomize or preserve varies depending on what question is being asked with a given measure (i.e., what other parts should be controlled for in the randomization), and also varies depending on how the data were sampled (i.e., to preserve any bias in the randomizations that may have induced from the sampling methods, to ensure the structure isn't just an artifact of that bias) (Farine 2017). In a multilayer network, there are many more parts to potentially randomize or consider. The same considerations that apply for a single layer network apply to each layer of a multilayer network. If data for layers was collected using different sampling, there may need to be different types of randomizations on each layer. In addition, depending on whether researchers are using interlayer edges and whether those also are assigned values from data, randomizations may need to be used on these edges as well.

Depending on the measure used and study question, edges can either be randomized within layer only, or across all layers. For example, if each layer is a different behavior and a researcher is interested in measuring some characteristic about an individual's role, it may be appropriate to only randomize connections within layers. Similarly, randomizations in multilayer motif analyses may only randomize within layers, as they aim to detect significant patterns that span multiple layers (Smoly et al. 2017). In contrast, if one was interested in how layers were related to each other, they may want to randomize across layers to preserve the frequencies at which behaviors were used, or even randomize the values within multi-degree vectors for each individual to preserve the total amount that individuals interacted. Intralayer edges could be randomized only among intralayer edges, or across interlayer edges as well. Another option, depending on the data and question, may be to randomize isolated parts of the network at one time (e.g., only one

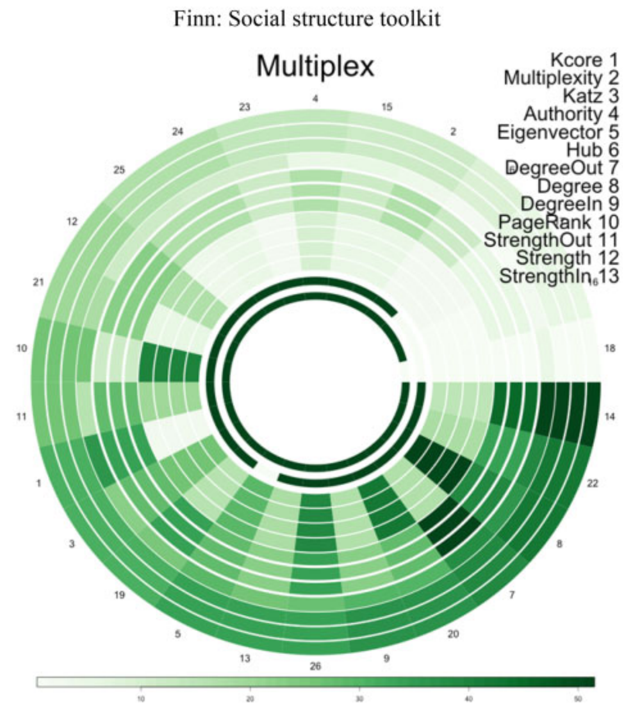

Figure 9. Multilayer centrality and versatility measures. Thirteen different centrality or versatility measures can be calculated on Muxviz (De Domenico et al. 2015c). All of them were calculated here on a 2-layer multiplex network of grooming and associations in a baboon group with data from Franz et al. (2015a, 2015b). The results are displayed in the annular visualization created in Muxviz—the darker the green, the higher the relative value for each centrality measure. Values for each centrality measure are shown as a ring in the visualization, with Kcore being the innermost, to Strength of Indegree being the outermost, following the order of the legend. Individual IDs are represented as numbers along slices of the circular figure.

layer), but do several randomizations that each randomize different parts, to assess the effects of various potential confounds. For each randomization, one should know what parts of the network structure are being randomized or preserved, and functionally what that means. What about the network structure does the randomization remove? What about the network structure is being preserved for each part that is not randomized? In general, each randomization should serve a unique purpose to statistically validate specific structural properties, and control for specific possible confounds. Like any other network analyses, randomizations will never be a one size fits all solution and depend on the unique study question, dataset, and analysis used.

Pitfalls specific to multilayer networks

Differences across layers

Besides the magnification of existing potential pitfalls in social network analyses, there are new challenges unique to multilayer networks to keep in mind. There are numerous attributes that can differ across network layers that can influence the values calculated in certain analyses. The following characteristics are important to consider when constructing and measuring a multilayer network.

Network properties. Not all behaviors occur at the same rates, which could cause networks of different behaviors to be different sizes, even if they were sampled at comparable intervals. If this is the case, some layers may be disproportionately represented or have inflated influence by certain analyses (Finn et al. 2019). For example, in measures that use a random walker, a random walker may end up spending more time on network layers that have greater participation (i.e., more individuals use a behavior), higher densities, or higher average degree, and therefore have a greater influence on the outcome. This may or may not be desirable. If the intention is to treat each layer as different but equally influential social domains, you may need to employ normalization procedures to lessen this inflation, or use methods that explicitly treat layers as separate, but equal contributors to the estimate (Beisner et al. 2020). Generally, comparing networks of different sizes can be challenging (James et al. 2009), and measures that compare layers will face similar problems.

Sampling. If the data from some behaviors were sampled differently than others (e.g., one used scan sampling while another used event sampling), some behaviors could have much higher frequencies in the data, even if they happened similar amounts. This creates similar problems as having networks of behaviors that occur at different rates. One may need to use different normalizations for edges on different layers, or account for this during randomizations.

Units. In addition, if different sampling occurred for different behavioral data, layers may represent data that were measured in fundamentally different units. For example, some edges might be weighted by the total number of days an animal was observed engaging in behaviors, while other edges might be weighted by a count of the number of times behaviors occurred. Beyond sampling differences, edges or layers could also differ in units due to the nature of representing categorically different attributes (Finn et al. 2019). If some behaviors occur for long periods of time (i.e., behavioral states), while others occur very quickly (i.e., behavioral events), some edge types might represent a duration engaged in a behavior, while others might represent a count of instances of a behavior. If some layers are not a behavioral type (e.g., relatedness), the unit might be something completely different (e.g., percent of genetic similarity). For some multilayer network measures, the results generated from layers that contain edges of different units could be difficult to interpret or uninterpretable. Different units might just rule out the ability to use certain analyses, or each layer may need to be further course-grained or normalized to make sense. For instance, if some edges represented total days observed, others represented total scans observed, and others represented total instances observed, using unweighted n-grams (i.e., the presence or absence of the behavior overall) may be a better option to analyze relationship types. Otherwise, with continuous edge weights it is unlikely any 2 relationships will appear similar even if they are (which is a problem if the intention is to identify relationships types), and/or it may be hard to reason about what the values of each edge mean relative to the whole relationship. Another option could be to discretize variables into categories of low, medium, and high, so each layer has a similar range of values that represent the relative amount they were used.

Which layers to use?

One of the most immediately obvious caveats to a multilayer network approach to quantifying social structure is that in order to get a reasonable representation of the social situations of a group, you would need to include all important and relevant layers (e.g., interaction types and contexts). Which types of social interactions are unique and important for a species might not actually be known (or even knowable). While it is the case that many datasets contain multiple interaction types or contexts, it is unlikely that many datasets contain all interaction types or all contexts that are relevant to an individual's social life. If for instance, you wanted to compare the social structure of different species (Pasquaretta et al. 2014), but left out an important layer or layers for some of the species, the networks would not be good representations. For any given study, however, it may not be the case that all social domains are necessary or appropriate to include, unless the research aim is to make a complete and thorough assessment of a group's social structure. Many or most study questions likely focus on specific components of social structure, and possibly even a narrow range of behavioral domains (e.g., agonism). Adapting the common phrase about modeling: all multilayer network representations are wrong, but some are useful.

When deciding which behavioral layers to include when running a particular analysis, many of the decisions may overlap with deciding what should be distinct layers. A researcher may have separately decided which layers should be distinct layers to represent their study system, in which case they then need to decide which are relevant to include in the analysis. Alternatively, a researcher could make decisions at the same time about which layers to separate, combine, or include in the analysis. For example, if a study question is related to agonism or a group's hierarchy, should layers of affiliative behaviors also be included? Considering what the analysis measures about a network's structure, which behaviors should that structure be made of? Unless the aim of the analysis is purely exploratory, it is likely not the best approach to include all possible layers, and instead they should be selected carefully with reference to the study question and method. There may be additional practical considerations as well, such as a maximum number of layers that could be used due to computational constraints, or constraints of statistical power for a given method and group size.

Code and software

Unfortunately, even of the existing analyses, not all have easy to implement software and code. For those that do have accessible implementations (see a list of some in Table 1 of Finn et al. 2019), it is still important that the researcher understands exactly what the analysis is doing. Especially with regards to interlayer edges, some software platforms may be using edges in unanticipated ways, or ways that are inappropriate for some data. For example, while Muxviz (De Domenico et al. 2015c) is a very useful platform for multilayer network visualization and analysis, some commands use data in potentially unexpected ways. Some analyses in Muxviz have the option to also calculate measures on the aggregate single layer network. Sometimes this preserves interlayer edges from the multilayer network that connected node-tuples of the same individual, as self-loops connecting a node to itself in the aggregate network. This could inflate a very basic measure such as degree, if one did not intend to count interlayer edges or self-loops.

Interlayer edges that connect node-tuples of the same individual may also create undesired inflation of multilayer measures that are not aggregated if run blindly. For instance, consider a scenario with a 2-layer multiplex network where betweenness versatility is calculated, which measures the number of shortest pathways between all dyads that pass through an individual. If an individual is connected to itself on both layers A and B with an interlayer edge between node-tuple A and node-tuple B, but node-tuple A does not interact with any individuals in its layer, node-tuple B is going to be involved in every shortest path between node-tuple A and all node-tuples for all other individuals on both network layers, substantially inflating the betweenness centrality of node-tuple B. Such effects may or may not be desirable depending on the research aim, so it is important to think them through.

In addition, some measures might not work if the data have directed or weighted edges, or such attributes might not be used in an analysis. Therefore, it is important to locate either the source code, or the original articles that the implementations are based on, to understand what the analyses are doing. When in doubt, it is always a good idea to create toy networks and test them out on the software, to make sure it is doing what one thinks it is doing. A researcher can build an intuition for when the values should be different and create model networks to check the output of the software against their intuition. For instance, one can start with a small network with only a few nodes and layers, manipulate the edges, and see if the measure changes in the way they expect. They can create a network that they think should produce a high value of a measure, and a network they think should produce a low value of a measure. By simulating a set of networks that vary by the characteristic they are trying to measure, they can then double check that the measure and implementation they are using is capturing this characteristic across the simulated set.

Final remarks

While multilayer networks will not solve all our problems, they can be extremely useful tools for quantifying social behavior if used thoughtfully and carefully. The promise of multilayer network analysis is not in what it can do for us, but what we can do with it. It will never be a plug-and-chug analysis, nor will it ever sort through all the complexities of our system by itself and tell us what we want to know. The sorting and decision-making still falls into the hands of the researcher, so it is important, as it is with any statistical model, to have a good sense of what the data are and what the analysis is doing with it. This is important on a conceptual level with regards to the study system, at an analytical level on the network, and that these 2 things map onto each other coherently through the data. There is a good chance a measure does not already exist for the exact thing a researcher wants to measure in a network constructed from their unique data, but there is also a good chance such a measure could be created. Do not be afraid to create your own measures or collaborate with a network scientist to create an analysis that does what you want! Because of the vast diversity of systems multilayer networks can represent, the future of multilayer network analysis depends as much, if not more, on empirical researchers identifying meaningful structure in their systems and data they want to quantify, as it does on quantitative researchers developing new tools. The relationship space of our study systems is the final frontier.

Author contributions

KRF conceptualized, wrote, and created visualizations for this manuscript.

Acknowledgments

I would like to thank the organizers of both this special column and the 2019 ABS multilayer network symposium—David Fisher, Matthew Silk, and Matthew Hasenjager—as well as the attendees of the ABS symposium, for inspiring this work. I would also like to thank all of my past and current collaborators with whom I have worked with on multilayer networks—Marton Posfai, Curtis Atkisson, Chris Welker, Brianne Beisner, Brenda McCowan and laboratory members, Raissa D'Souza and laboratory members, Noa Pinter-Wollman, Matt Silk, and Mason Porter. All these collaborations were formative to my work with multilayer networks and highly valued. I would also like to thank the Neukom Institute for funding over the course of this manuscript. Finally, I thank the 4 anonymous reviewers for their thoughtful feedback and edits on this paper, and Curtis Atkisson for helpful pre-submission feedback and review.

References

Aleta A, Moreno Y, 2018. Multilayer networks in a nutshell. ArXiv.

Anderson C, McShea DW, 2001. Individual versus social complexity, with particular reference to ant colonies. *Biol Rev Camb Philos Soc* **76**: 211–237.

Atkisson C, Finn KR, 2020. Redundant relationships in multiplex food sharing networks increase food security in a nutritionally precarious environment. arXiv:2011.12817v1.

Atkisson C, Górski PJ, Jackson MO, Hołyst JA, D'Souza RM, 2020. Why understanding multiplex social network structuring processes will help us better understand the evolution of human behavior. *Evol Anthropol* **29**: 102–107.

Balasubramaniam KN, Beisner BA, Hubbard JA, Vandeleest JJ, Atwill ER et al. 2019. Affiliation and disease risk: social networks mediate gut microbial transmission among rhesus macaques. *Anim Behav* **151**: 131–143.

Balasubramaniam KN, Beisner BA, Vandeleest J, Atwill E, McCowan B, 2016. Social buffering and contact transmission: network connections have beneficial and detrimental effects on *Shigella* infection risk among captive rhesus macaques. *PeerJ* **4**: e2630.

Barrett L, Henzi SP, Lusseau D, 2012. Taking sociality seriously: the structure of multi-dimensional social networks as a source of information for individuals. *Phil Trans R Soc Lond B Biol Sci* **367**: 2108–2118.

Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A, 2005. Characterization and modeling of weighted networks. *Phys A Statist Mech Appl* **346**: 34–43.

Beisner BA, Braun N, Pósfai M, Vandeleest J, D'Souza R, McCowan B, 2020. A multiplex centrality metric for complex social networks: sex, social status, and family structure predict multiplex centrality in rhesus macaques. *PeerJ* **8**: e8712.

Beisner BA, Hannibal DL, Finn KR, Fushing H, McCowan B, 2016. Social power, conflict policing, and the role of subordination signals in rhesus macaque society. *Am J Phys Anthropol* **160**: 102–112.

Beisner BA, Jin J, Fushing H, McCowan B, 2015. Detection of social group instability among captive Rhesus macaques using joint network modeling. *Curr Zool* **61**: 70–84.

Bergman TJ, Beehner JC, 2015. Measuring social complexity. *Anim Behav* **103**: 203–209.

Borgatti SP, Brass DJ, Halgin DS, 2014. Social network research: confusions, criticisms, and controversies. *Res Sociol Organ* **40**: 1–29.

Brent LJN, Semple S, Dubuc C, Heistermann M, MacLarnon A, 2011. Social capital and physiological stress levels in free-ranging adult female rhesus macaques. *Physiol Behav* **102**: 76–83.

Chen PY, Hero AO, 2017. Multilayer spectral graph clustering via convex layer aggregation: theory and algorithms. *IEEE Trans Signal Inf Process Netw* **3**: 553–567.

Chen Z, Chen C, Zheng Z, Zhu Y, 2019. Tensor decomposition for multilayer networks clustering. *Proc AAAI Conf Artif Intell* **33**: 3371–3378.

Cheney DL, 1992. Intragroup cohesion and intergroup hostility: the relation between grooming distributions and intergroup competition among female primates. *Behav Ecol* **3**: 334–345.

Chodrow PS, Al-Awwad Z, Jiang S, González MC, Herranz R et al., 2016. Demand and congestion in multiplex transportation networks. *PLoS ONE* **11**: e0161738.

Croft DP, James R, Krause J, 2008. *Exploring Animal Social Networks*. Princeton, NJ: Princeton University Press.

de Borda JC, 1781. Mémoire sur les élections au scrutin. In: ScrutinMémoires Del'Académie Royale Des Sciences de Paris Pour l'Année. 657–665.

De Domenico M, Granell C, Porter MA, Arenas A, 2016. The physics of spreading processes in multilayer networks. *Nat Phys* **12**: 901–906.

De Domenico M, Lancichinetti A, Arenas A, Rosvall M, 2015a. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys Rev X* **5**: 011027.

De Domenico M, Nicosia V, Arenas A, Latora V, 2015b. Structural reducibility of multilayer networks. *Nat Commun* **6**: 6864.

De Domenico M, Porter MA, Arenas A, 2015c. MuxViz: a tool for multilayer analysis and visualization of networks. *J Complex Netw* **3**: 159–176.

De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y et al., 2014. Mathematical formulation of multilayer networks. *Phys Rev X* **3**: 041022.

De Domenico M, Solé-Ribalta A, Omodei E, Gómez S, Arenas A, 2015d. Ranking in interconnected multilayer networks reveals versatile nodes. *Nat Commun* **6**: 6868.

Drewe JA, 2010. Who infects whom? Social networks and tuberculosis transmission in wild meerkats. *Proceedings Biol Sci* **277**: 633–642.

Dunbar RIM, 1998. The social brain hypothesis. *Evol Anthropol* **6**: 178–190.

El Gheche M, Chierchia G, Frossard P, 2020. OrthoNet: multilayer network data clustering. *IEEE Trans Signal Inf Process Netw* **6**: 13–23.

Farine DR, 2017. A guide to null models for animal social network analysis. *Methods Ecol Evol* **8**: 1309–1320.

Farine DR, Whitehead H, 2015. Constructing, conducting, and interpreting animal social network analysis. *J Anim Ecol* **84**: 1144–1163.

Finn KR, 2019. *Studying Animals as Complex Systems: From Perception to Social Structure*. Davis: University of California.

Finn KR, Silk MJ, Porter MA, Pinter-Wollman N, 2019. The use of multilayer network analysis in animal behaviour. *Anim Behav* **149**: 7–22.

Fischer J, Farnworth MS, Sennhenn-Reulen H, Hammerschmidt K, 2017. Quantifying social complexity. *Anim Behav* **130**: 57–66.

Fortunato S, Hric D, 2016. Community detection in networks: a user guide. *Phys Rep* **659**: 1–44.

Fosdick BK, Larremore DB, Nishimura J, Ugander J, 2018. Configuring random graph models with fixed degree sequences. *SIAM Rev* **60**: 315–355.

Franz M, Altmann J, Alberts SC, 2015a. Data from: knockouts of high-ranking males have limited impact on baboon social networks. *Curr Zool* **61**: 107–113.

Franz M, Altmann J, Alberts SC, 2015b. Knockouts of high-ranking males have limited impact on baboon social networks. *Curr Zool* **61**: 107–113.

Freeberg TM, Dunbar RIM, Ord TJ, 2012. Social complexity as a proximate and ultimate factor in communicative complexity. *Phil Trans R Soc B Biol Sci* **367**: 1785–1801.

Fushing H, Jordà Ò, Beisner B, McCowan B, 2014. Computing systemic risk using multiple behavioral and keystone networks: the emergence of a crisis in primate societies and banks. *Int J Forecast* **30**: 797–806.

Goldberger AL, Peng C-K, Lipsitz LA, 2002. What is physiologic complexity and how does it change with aging and disease? *Neurobiol Aging* **23**: 23–26.

Groenewoud F, Frommen JG, Josi D, Tanaka H, JungwirthA et al. 2016. Predation risk drives social complexity in cooperative breeders. *Proc Natl Acad Sci U S A* **113**: 4104–4109.

Grueter CC, Matsuda I, Zhang P, Zinner D, 2012. Multilevel societies in primates and other mammals: introduction to the special issue. *Int J Primatol* **33**: 993–1001.

Hinde RA, 1976. Interactions, relationships, and social structure. *Man* **11**: 1–17.

Hinde RA, 1978. Dominance and role-two concepts with dual meanings. *J Soc Biol Syst* **1**: 27–38.

Hinde RA, Datta S, 1981. Dominance: an intervening variable. *Behav Brain Sci* **4**: 443–443.

Hobson EA, Avery ML, Wright TF, 2013. An analytical framework for quantifying and testing patterns of temporal dynamics in social networks. *Anim Behav* **85**: 83–96.

Hobson EA, Ferdinand V, Kolchinsky A, Garland J, 2019. Rethinking animal social complexity measures with the help of complex systems concepts. *Anim Behav* **155**: 287–296.

Hobson EA, Silk MJ, Fefferman NH, Larremore DB, Rombach P et al., 2020. A guide to choosing and implementing reference models for social network analysis. arXiv.

James R, Croft DP, Krause J, 2009. Potential banana skins in animal social network analysis. *Behav Ecol Sociobiol* **63**: 989–997.

Kao TC, Porter MA, 2018. Layer communities in multiplex networks. *J Stat Phys* **173**: 1286–1302.

Kemeny JL, Snell JG, 1962. *Mathematical Models in the Social Sciences*. New York (NY): Blaisdell.

Kivelä M, Arenas A, Barthélemy M, Gleeson JP, Moreno Y et al., 2014. Multilayer networks. *J Complex Netw* **2**: 203–271.

Krause J, Croft DP, James R, 2007. Social network theory in the behavioural sciences: potential applications. *Behav Ecol Sociobiol* **62**: 15–27.

Kummer H, Daston L, Gigerenzer G, Silk JB, 1997. The social intelligence hypothesis. In: Weingart P, Mitchell SD, Richerson PJ, Maasen S, editors. *Human by Nature: Between Biology and the Social Sciences*. Mahwah, NJ: Erlbaum. 157–179.

Lehmann J, Andrews K, Dunbar RIM, 2012. Social networks and social complexity in female-bonded primates. *Proc Br Acad* **158**: 403–423.

Lehmann J, Boesch C, 2009. Sociality of the dispersing sex: the nature of social bonds in West African female chimpanzees, Pan troglodytes. *Anim Behav* **77**: 377–387.

Lehmann J, Dunbar RIM, 2009. Network cohesion, group size and neocortex size in female-bonded Old World primates. *Proc R Soc B Biol Sci* **276**: 4417–4422.

Lehmann J, Ross C, 2011. Baboon *Papio anubis* social complexity—a network approach. *Am J Primatol* **73**: 775–789.

Madotto A, Liu J, 2016. Super-spreader identification using meta-centrality. *Sci Rep* **6**: 38994.

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD et al., 2017. A manifesto for reproducible science. *Nat Hum Behav* **1**: 0021.

Newman MEJ, 2018. *Networks*. 2nd edn. Oxford: Oxford University Press.

Pasquaretta C, Leve M, Claidiere N, van de Waal E, Whiten A et al., 2014. Social networks in primates: smart and tolerant species have more efficient networks. *Sci Rep* **4**: 7600.

Pereira AS, Rebelo ID, Casanova C, Lee PC, Louca V, 2020. The multidimensionality of female mandrill sociality: a dynamic multiplex network approach. *PLoS ONE* **15**: e0230942.

Pilosof S, Porter MA, Pascual M, Kéfi S, 2017. The multilayer nature of ecological networks. *Nat Ecol Evol* **1**: 0101.

Pinter-Wollman N, 2015. Nest architecture shapes the collective behaviour of harvester ants. *Biol Lett* **11**:20150695.

Pinter-Wollman N, Hobson EA, Smith JE, Edelman AJ, Shizuka D et al., 2014. The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behav Ecol* **25**: 242–255.

Pósfai M, Braun N, Beisner BA, McCowan B, D'Souza RM, 2019. Consensus ranking for multi-objective interventions in multiplex networks. *New J Phys* **21**: 055001.

Ramos-Fernandez G, King AJ, Beehner JC, Bergman TJ, Crofoot MC et al., 2018. Quantifying uncertainty due to fission–fusion dynamics as a component of social complexity. *Proc R Soc B Biol Sci* **285**: 20180532.

Sapolsky RM, 1982. The endocrine stress–response and social status in the wild baboon. *Horm Behav* **16**: 279–292.

Schrock AE, Leard C, Lutz MC, Meyer JS, Gazes RP, 2019. Aggression and social support predict long-term cortisol levels in captive tufted capuchin monkeys *Cebus [Sapajus] apella*. *Am J Primatol* **81**: e23001.

Sewall KB, 2015. Social complexity as a driver of communication and cognition. *Integr Comp Biol* **55**: 384–395.

Silk J, Cheney D, Seyfarth R, 2013. A practical guide to the study of social relationships. *Evol Anthropol* **22**: 213–225.

Silk MJ, Finn KR, Porter MA, Pinter-Wollman N, 2018. Forum: can multilayer networks advance animal behavior research? *Trends Ecol Evol* **33**: 376–378.

Smaldino PE, McElreath R, 2016. The natural selection of bad science. *R Soc Open Sci* **3**: 160384.

Smith-Aguilar SE, Aureli F, Busia L, Schaffner C, Ramos-Fernández G, 2018. Using multiplex networks to capture the multidimensional nature of social structure. *Primates* **60**: 277–295.

Smoly IY, Lerman E, Ziv-Ukelson M, Yeger-Lotem E, 2017. MotifNet: a web-server for network motif analysis. *Bioinformatics* **33**: 1907–1909.

Stanley N, Shai S, Taylor D, Mucha PJ, 2016. Clustering network layers with the strata multilayer stochastic block model. *IEEE Trans Netw Sci Eng* **3**: 95–105.

Strano E, Shai S, Dobson S, Barthélemy M, 2015. Multiplex networks in metropolitan areas: generic features and local effects. *J R Soc Interface* **12**: 20150651.

van der Marel A, Prasher S, Carminito C, O'Connell C, Phillips A et al., 2020. A framework to evaluate whether to pool or separate behaviors in a multilayer network. ArXiv 2007.09743.

Vandeleest JJ, Beisner BA, Hannibal DL, Nathman AC, Capitanio JP et al., 2016. Decoupling social status and status certainty effects on health in macaques: a network approach. *PeerJ* **4**: e2394.

Vijayaraghavan VS, Noël P-A, Maoz Z, D'Souza RM, 2015. Quantifying dynamical spillover in co-evolving multiplex networks. *Sci Rep* **5**: 15142.

Wey TW, Blumstein DT, Shen W, Jordán F, 2008. Social network analysis of animal behaviour: a promising tool for the study of sociality. *Anim Behav* **75**: 333–344.

Whitehead H, 2008. *Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis*. Chicago (IL): University of Chicago Press.

Whiten A, Byrne RW, 1988. The Machiavellian intelligence hypotheses. In: Byrne RW, Whiten A editors. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Oxford University Press. 1–9.

Wittemyer G, Douglas-Hamilton I, Getz WM, 2005. The socioecology of elephants: analysis of the processes creating multitiered social structures. *Anim Behav* **69**: 1357–1371.

Wooddell L, Vandeleest J, Nathman A, Beisner B, McCowan B, 2019. Not all grooming is equal: differential effects of political vs affiliative grooming on cytokines and glucocorticoids in rhesus macaques. *PeerJ* **7**: e27961v1.

Wooddell LJ, Hamel AF, Murphy AM, Byers KL, Kaburu SSK et al., 2017. Relationships between affiliative social behavior and hair cortisol concentrations in semi-free ranging rhesus monkeys. *Psychoneuroendocrinology* **84**: 109–115.

Wu H, James RG, D'Souza RM, 2020. Correlated structural evolution within multiplex networks. *J Complex Netw* **8**: cnaa014.

Zafeiris A, Vicsek T, 2018. *Why We Live in Hierarchies? A Quantitative Treatise*. Springer. 12–39.