ORIGINAL PAPER

Multiplexity analysis of networks using multigraph representations

Termeh Shafie[1] (ORCID) · David Schoch[2]

Abstract

Multivariate networks comprising several compositional and structural variables can be represented as multigraphs by various forms of aggregations based on vertex attributes. We propose a framework to perform exploratory and confirmatory multiplexity analysis of aggregated multigraphs in order to find relevant associations between vertex and edge attributes. The exploration is performed by comparing frequencies of the different edges within and between aggregated vertex categories, while the confirmatory analysis is performed using derived complexity or multiplexity statistics under different random multigraph models. These statistics are defined by the distribution of edge multiplicities and provide information on the covariation and dependencies of different edges given vertex attributes. The presented approach highlights the need to further analyse and model structural dependencies with respect to edge entrainment. We illustrate the approach by applying it on a well known multivariate network dataset which has previously been analysed in the context of multiplexity.

Keywords Random multigraphs · Edge multiplicity · Complexity · Multivariate networks · Data aggregation

1 Introduction

Multivariate networks consist of several different relations (also referred to as endogenous or structural variables) mapped on a common set of vertices with multiple covariates (also referred to as exogenous or compositional variables). Thus, the term can be seen as a generalisation of what is referred to as multiple, multi-relational, multilayered or multiplex networks (White et al., 1976; Fienberg et al.,

---

✉ Termeh Shafie
  termeh.shafie@manchester.ac.uk

[1] Department of Social Statistics, The University of Manchester, Manchester, England

[2] Department of Sociology, The University of Manchester, Manchester, England

1985; Lazega and Pattison, 1999; Pattison and Wasserman, 1999; Magnani and Wasserman, 2017; Bianconi, 2018). Analysing such networks provides unique opportunities in unravelling the mechanisms that govern the structure, but with the cost of increased methodological and computational complexity. A common approach to overcome these issues is to simplify the networks and apply single network analytic tools. A recent review on such methods for multilayer networks is given by Interdonato et al. (2020). Existing studies on multiplexity have used the concept in different contexts. These include visualization De Domenico et al. (2015), community detection Bothorel et al. (2015), Mucha et al. (2010), and model development for both cross sectional and longitudinal network data Krivitsky et al. (2020), Koehly and Pattison (2005), Snijders et al. (2013).

To date, multiplexity studies that incorporate vertex covariates in the analysis are still scarce. The approach proposed in this work allows for the analyses of multiplexity given single or combined vertex attributes. We simplify an observed multi-relational network while keeping available information on the vertices. This is done by aggregating with respect to single or combined vertex categories. Such networks can be represented using multigraphs with multiple inter-category edges of different kind, and intra-category edges represented as self-edges or loops (Shafie, 2015; 2016). Using this multigraph representation, we propose different tools to visually examine network properties and perform an exploratory analysis of the joint and marginal distribution of edges, i.e. the edge proportions among dyads in different aggregated vertex categories. The multigraph representation additionally enables the use of probabilistic multigraph models proposed in Shafie (2015, 2016) to analyse structural features given vertex covariates. Note that the models proposed in Shafie (2015, 2016) are different than those presented in e.g. Janson (2009, 2014), Ranola et al. (2010), Godehardt (1993) since the former allow self-edges while the latter models define them as structural zeros.

The considered multigraph models in this work are used with a special focus on entrainment of the multiple edges. Specifically, multiplexity statistics derived under the models are used to detect associations between the different types of edges, and to infer whether the occurrence of various edge types are statistically different and/ or independent of each other. To illustrate the proposed framework, we apply the methods on a multivariate network with five relations and two vertex attributes. This dataset has been analysed in several multiplexity studies, all of which only focus structural patterns and associations among the multiple type of edges. These applications include visualization techniques (Rossi and Magnani, 2015; Dickison et al., 2016), descriptive measures of node, tie and network level features (Dickison et al., 2016; Interdonato et al., 2020), computation of similarities between the different layers of the networks (Bródka et al., 2018), and exploration of the multiplex structure using generalised multidimensional scaling (Giordano et al., 2019).

The remainder of this paper is structured as follows. Multigraph representations, including aggregation technique and random multigraph models, are given in the following section. In Sect. 2, statistics used to analyse multiplexity are defined and their use is exemplified. Section 3 introduces the dataset used throughout the paper and the analysis is performed in Sect. 4. In Sect. 5, we present an extension to the

proposed framework which is not contingent on the presence of vertex covariates. We conclude with a discussion on the flexibility, limitations and possible extensions of the approach.

2 Multigraph representations

2.1 Network data aggregation

Multigraphs can be used to represent directly observed networks with multiple edges mapped on the same set of vertices. These graphs commonly exclude loops since observing self-edges is very unlikely in many applications. However, there are other ways in which a multigraph can be obtained as a representation of observed networks. For instance, using different aggregation techniques on the available vertex attributes in a multivariate network. As illustrated in Shafie (2015, 2016), vertices can be partitioned into categories based on single or combined attributes. Such aggregations result in multigraphs which are significantly smaller than the original graph and contain intra-category edges represented by loops. This approach naturally allows for a simultaneous study of compositional and structural variables including their covariation. These aggregation techniques are the main focus of our approach and are exemplified in Sect. 3.2 using the multivariate network data presented in Sect. 3.1.

2.2 Random multigraph models

An undirected multigraph with n vertices and m edges can be represented by its edge multiplicity sequence denoted $\mathbf{M} = (M_{ij} : (i,j) \in \mathcal{R})$, where $\mathcal{R} = \{(i,j) : 1 \leq i \leq j \leq n\}$ is the canonical site space for undirected edges, that is vertex pair sites where edges can be present. The edge multiplicity M_{ij} then denotes the number of (multiple) edges at these different sites $(i,j) \in \mathcal{R}$ for $1 \leq i \leq j \leq n$ and with $i = j$ representing edge loops. The number of available vertex pair sites for a multigraph on n vertices is given by $r = \binom{n+1}{2}$.

In the following, we briefly present the details of two probabilistic multigraph models introduced by Shafie (2015, 2016). Applications of the models can be found in Frank and Shafie (2018) and Shafie (2015). The first model is a random multigraph model which can be seen as a generalisation of the Bernoulli model. The m edges of the multigraph are independently assigned to the sites $(i,j) \in \mathcal{R}$ and the edge multiplicity sequence $\mathbf{M}$ follows a multinomial distribution with parameters m and $\mathbf{Q} = (Q_{ij} : (i,j) \in \mathcal{R})$, where $\mathbf{Q}$ is the edge probability sequence with edge assignment probabilities Q_{ij} for each site $(i,j) \in \mathcal{R}$. The model is referred to as the *independent edge assignment model* (IEA) (Shafie, 2015) and the probability of a multigraph under this model is given by

$$P(\mathbf{M} = \mathbf{m}) = \binom{m}{\mathbf{m}} \mathbf{Q}^{\mathbf{m}} = \frac{m!}{\prod_{i \leq j} m_{ij}!} \prod_{i \leq j} Q_{ij}^{m_{ij}}. \tag{1}$$

The second model is a random multigraph model where the assumption of independence is dropped and a degree based dependence is incorporated. The observed edges are split, and the stubs (or half-edges) are randomly coupled to form edges. This model is referred to as the *random stub matching model* (RSM) given a fixed degree sequence. The edge probability sequence $\mathbf{Q}$ is defined as a function of these degrees $\mathbf{d} = (d_1, \ldots, d_n)$. The edge assignment probabilities are given by

$$Q_{ij} = \begin{cases} \binom{d_i}{2} \Big/ \binom{2m}{2} & \text{for} \quad i = j \\ d_i d_j \Big/ \binom{2m}{2} & \text{for} \quad i < j , \end{cases} \tag{2}$$

and the probability of a multigraph under this model is given by

$$P(\mathbf{M} = \mathbf{m}) = \frac{2^{m_2} \binom{m}{\mathbf{m}}}{\binom{2m}{\mathbf{d}}} = \frac{2^{m_2} m! \prod_{i=1}^{n} d_i!}{(2m)! \prod_{i \leq j} m_{ij}!} , \tag{3}$$

where $m_2 = \sum \sum_{i < j} m_{ij}$ (Shafie, 2016).

The independence assumption following the IEA model is known to be unrealistic in real world settings but can serve as a baseline for comparison and to suggest necessary model modifications which go beyond the independence assumption. Moreover, moments of certain statistics to analyse multigraph structures are easier to handle under this model (specific examples of such statistics are given in the following section). In contrast, the RSM model allows for degree based effects such as the Matthew effect (Merton, 1968) but comes with combinatorial difficulties when deriving closed formulas for the moments of statistics. These formulas are essential for analysing observed networks since computing full multigraph probability distributions under either of the two models increase quickly and becomes infeasible for larger networks. Specifically for the IEA model, the number of multigraphs represented by $\mathbf{M}$ is given by $\binom{m + r - 1}{m}$.

Under the RSM model, this number is far lower since it is constrained by the degree sequence. However, to the best of our knowledge, no explicit formula for this number under the RSM model exists for multigraphs with more than three vertices. A formula for multigraphs with three or fewer vertices can be found in Frank and Shafie (2018).

Shafie (2016) studies the possibility to approximate the distribution of $\mathbf{M}$ under the RSM model using the IEA model, thus allowing the closed expressions of moments of statistics to be used in the structural analysis, This approximation is based on ignoring the dependence of the edge assignments and letting the edge probability sequence be defined as a function of the degree or stub multiplicity

sequence $\mathbf{Q} = \mathbf{Q}(\mathbf{d})$, as shown in Eq. 2. While the RSM model can be seen as selecting pairs of stubs to form edges without replacement, the IEA approximation can is equivalent to selecting pairs of stubs to form edges with replacement. Thus, we have independent edge assignment of stubs and the number of multigraphs $\mathbf{M}$ is still restricted by outcomes consistent with the stub multiplicity sequence. In order to determine when this approximation is justified, Shafie (2016) studies the edge multiplicity distributions for loops and non-loops, and the two central moments of these multiplicities. Some examples of the general results found are that the edge multiplicity distributions under RSM is closely related to that of the approximate IEA distribution at low degree vertices. For a more detailed discussion on the justification of this approximation, see Shafie (2016).

2.3 Multiplexity statistics

The moments of certain statistics to analyse structural properties of multigraphs are, as mentioned above, easier handled under the IEA model. This holds when using the IEA model as a stand-alone model or as an approximation to the RSM model. Examples of such statistics include the number of loops and numbers of non-loops. These can be used to analyse the edges moving within and between aggregated vertex pair categories and effects such as homophily or heterophily (see Shafie, 2015) for more details and examples). We here focus on statistics which can be used to detect multiplexity. These statistics are defined using a complexity sequence, i.e. the distribution of edge multiplicities[1]. This sequence is given by $\mathbf{R} = (R_0, R_1, R_2, \ldots, R_m)$ where

$$R_k = \sum_{i<j} \sum I(M_{ij} = k) \quad \text{for} \quad k = 0, 1, \ldots, m , \tag{4}$$

and I is an indicator variable. Thus, R_0 denotes the number of vertex pair sites with no edge occupancy, R_1 single edge occupancy, R_2 double edge occupancy, and so forth.

Multiplexity can be analysed using combinations of the R_k statistics. Comparing $R_0 + R_1$ to $R_3 + \cdots + R_k$, for instance, may reveal tendencies toward strengthening ties given an edge type. The first two central moments of R_k under the IEA model are derived in Shafie (2015) and given by

$$\mathbb{E}(R_k) = \sum_{i \leq j} \sum \binom{m}{k} Q_{ij}^k (1 - Q_{ij})^{m-k} \quad \text{for} \quad k = 0, 1, \ldots, m , \tag{5}$$

and

---

[1] The term complexity stems from Wasserman and Faust (1994): *"if a graph contains loops and/or any pairs of nodes is adjacent via more than one line, a graph is complex"*.

$$\mathbb{V}(R_k) = \binom{m}{k} \sum_{s \in \mathcal{R}} \left[Q_s^k (1 - Q_s)^{m-k} \left(1 - \binom{m}{k} Q_s^k (1 - Q_s)^{m-k} \right) \right]$$
$$- \binom{m}{k} \sum_{\substack{s,t \in \mathcal{R} \\ s \neq t}} \sum \left[\binom{m-k}{k} Q_s^k Q_t^k (1 - Q_s - Q_t)^{m-2k} - \binom{m}{k} Q_s^k Q_t^k [(1-Q_s)(1-Q_t)]^{m-k} \right] \tag{6}$$

where s and t are site pairs in $\mathcal{R} = \{(i,j) \in V^2 : i \leq j\}$. For instance, the number of sites with no edge occupancy has expected value

$$\mathbb{E}(R_0) = \sum_{i \leq j} \sum (1 - Q_{ij})^m \tag{7}$$

and variance

$$\mathbb{V}(R_0) = \sum_{s \in \mathcal{R}} (1 - Q_s)^m Q_s^m - \sum_{\substack{s,t \in \mathcal{R} \\ s \neq t}} \sum [(1 - Q_s - Q_t)^m - (1 - Q_s)^m (1 - Q_t)^m], \tag{8}$$

and the number of sites with single edge occupancy is given by R_1 has expected value

$$\mathbb{E}(R_1) = m \sum_{i \leq j} \sum Q_{ij}(1 - Q_{ij})^{m-1} \tag{9}$$

and variance

$$\mathbb{V}(R_1) = m \sum_{s \in \mathcal{R}} Q_s (1 - Q_s)^{m-1} \left(1 - m Q_s (1 - Q_s)^{m-1} \right)$$
$$- m \sum_{\substack{s,t \in \mathcal{R} \\ s \neq t}} \sum \left[(m-1) Q_s Q_t (1 - Q_s - Q_t)^{m-2} - m Q_s Q_t [(1 - Q_s)(1 - Q_t)]^{m-1} \right]. \tag{10}$$

The expected values and variances of R_k for all $k = 1, \ldots, m$ under the IEA model are implemented in the R package `multigraphr`[2]. The resulting interval estimates can be used to find indicators for dependencies among edges, i.e. when intervals overlap for multiple edge types. We illustrate this in Sect. 4. All notation used in this paper is summarised in Table 1.

3 Running example

3.1 Data description

To illustrate the proposed framework, we use the AUCS dataset, a multivariate network which consists of different type of ties and vertex attributes (Dickison et al., 2016; Rossi and Magnani, 2015), which has been analysed in several

---

[2] https://CRAN.R-project.org/package=multigraphr

multiplexity studies, e.g. Rossi and Magnani (2015), Dickison et al. (2016), Interdonato et al. (2020), Bródka et al. (2018), Giordano et al. (2019).[3] The network includes five different types of relations (co-authorship, Facebook friendship, shared leisure activities, regularly eating lunch together, and working with each others) among 61 employees of the Computer Science Department at Aarhus University. The data also includes two categorical vertex attributes, namely the affiliation to eight different research groups and five academic positions (professor, associate, postdoctoral researcher, PhD student, and administrative staff). In our analysis, we exclude the actors with missing values and those with multiple attribute categories assigned to them (e.g. an admin was affiliated with two research groups).[4] After pruning the data, we additionally excluded one research group with only one associated person. This leaves us with five networks on 52 vertices which is used in the subsequent analyses. The five relations are shown in Fig. 1 and some descriptive statistics are presented in Table 2.

3.2 Data transformation

A multigraph representation of the AUCS network is obtained by aggregating the edges based on single or combined vertex categories or vertex pair sites, as specified by the actor attributes, and with the different ties moving between and within the chosen categories. Hence, three different multigraphs can be constructed using the given vertex attributes. Two that are aggregated based on single vertex categories, academic position and research group, and one that is aggregated based on cross classifying both attributes. The resulting graphs are shown in Fig. 2.

Multigraphs as shown in Fig. 2 are the traditional way of displaying networks. However, multiple edge categories and the existence of loops can have a negative impact on the readability. We thus offer an alternative visualization using *waffle matrices* as shown in Fig. 3. The waffle matrices show occurrences of the five ties as cells of a matrix, with each row and column representing a vertex pair category. The diagonal thus represents edges within categories, i.e. the loops.

The multigraph based on the aggregation of academic position (cf. Figs. 2a and 3a) shows that eating lunch seems to be the most frequent interaction among all positions and that most interactions occur between the PhD students themselves, and between PhD students and Postdocs. This tendency is particularly clear for leisure activities and eating lunch together. We also note that the majority of publications occur between PhD students and Professors (likely their supervisors).

The multigraph based on the aggregation of research groups (cf. Figs. 2b and 3b) comprises seven vertices. The most frequent interactions occur within research groups, with the second group (G2) having the most internal interactions, especially with respect to eating lunch together and being friends on Facebook. Note that some research groups seem to not interact at all.

---

[3] The dataset can be found at http://multilayer.it.uu.se/datasets.html

[4] Note that this is not how missing and ambiguous attribute values should be treated in an empirical study. We opt for this solution as we use the application as a mere illustration of the proposed framework.

Table 1 Notations used in this paper

n	Number of vertices
m	Number of edges
$\mathcal{R}$	Canonical site space for undirected edges
r	Number of vertex pair sites $(i,j) \in \mathcal{R}$ given by $\binom{n+1}{2}$
M	Sequence with edge multiplicities M_{ij} for sites $(i,j) \in \mathcal{R}$
m	Sequence of realisations of edge multiplicities m_{ij} at sites $(i,j) \in \mathcal{R}$
Q	Sequence with edge assignment probabilities Q_{ij} at sites $(i,j) \in \mathcal{R}$
d	Sequence of vertex degrees $d_1, d_2, \ldots, d_n$
R	Sequence of edge multiplicity frequencies $R_0, R_1, R_2, \ldots, R_m$
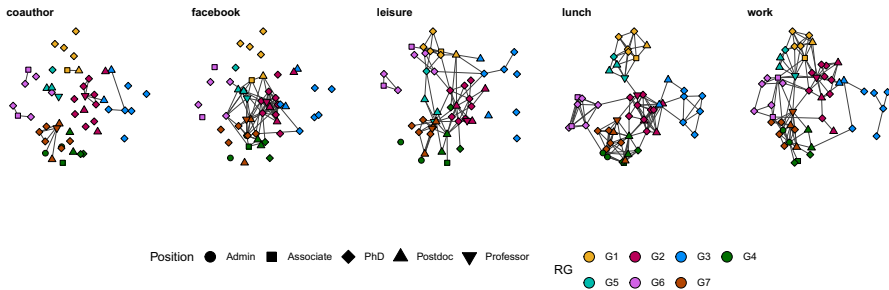
Fig. 1 The five types of relations of the considered network dataset. The available vertex attributes are research group (RG) and academic position

Table 2 Descriptive statistics for the five relation types in the used network dataset

Type	Edges	Non-isolated vertices	Transitivity	Max degree
coauthor	21	25	0.43	5
facebook	96	29	0.50	14
leisure	87	46	0.34	14
lunch	162	52	0.63	15
work	114	52	0.37	14

The largest multigraph is given by the aggregation of both research group and academic position, which yields a graph with 20 vertices, excluding 15 isolates (Fig. 2c). For a multigraph of this size, the waffle matrix is not easily readable. We therefore exclude it here. However, we do note the following from the cross classification presented in Fig. 2c; a small cluster of heavy interaction is observed from and to the second research group, and in particular among the PhD students.
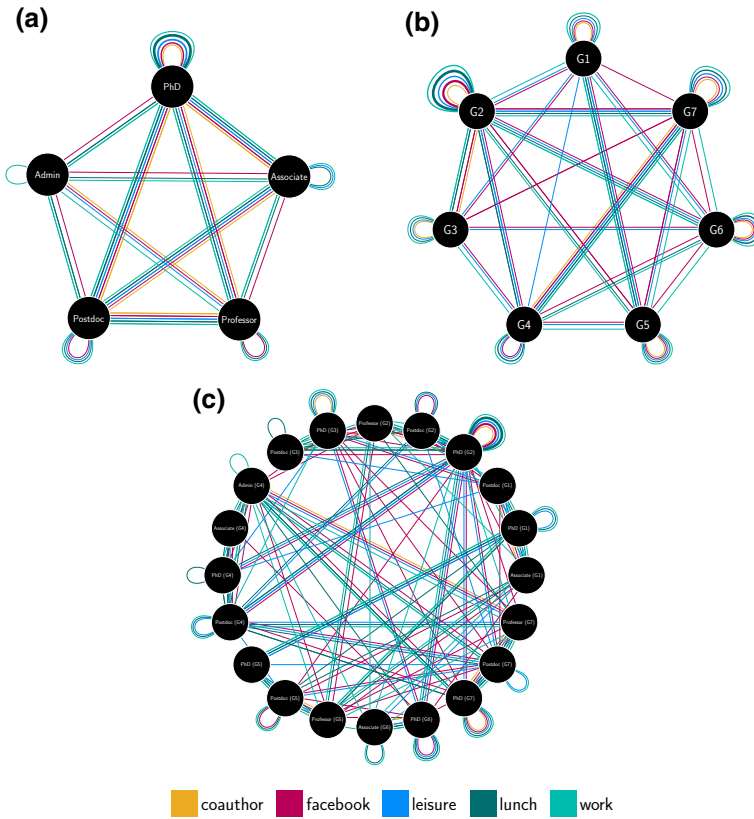
Fig. 2 Aggregated multigraphs based on actor attributes (**a**) academic position, (**b**) research group, and (**c**) combination of academic position and research group (with isolated vertices omitted). Thickness of edges represents multiplicity counts

Additionally, we observe a cluster of ties representing interactions between certain academic positions within research group seven.

4 Multiplexity analysis

4.1 Joint and marginal distribution of edge types

In order to explore the covariation of the different edges given vertex attributes, the joint and marginal distribution of the $\binom{52}{2} = 1326$ dyads within and between categories (or vertex pairs sites) can be systematically analysed for each aggregated multigraph. Doing so, we may be able to detect conditional dependencies of edges given vertex attributes. Practically, this can be done using contingency tables in which the proportions of edges within and between the different categories are represented and analysed (Shafie, 2015). If we, for example, are interested in
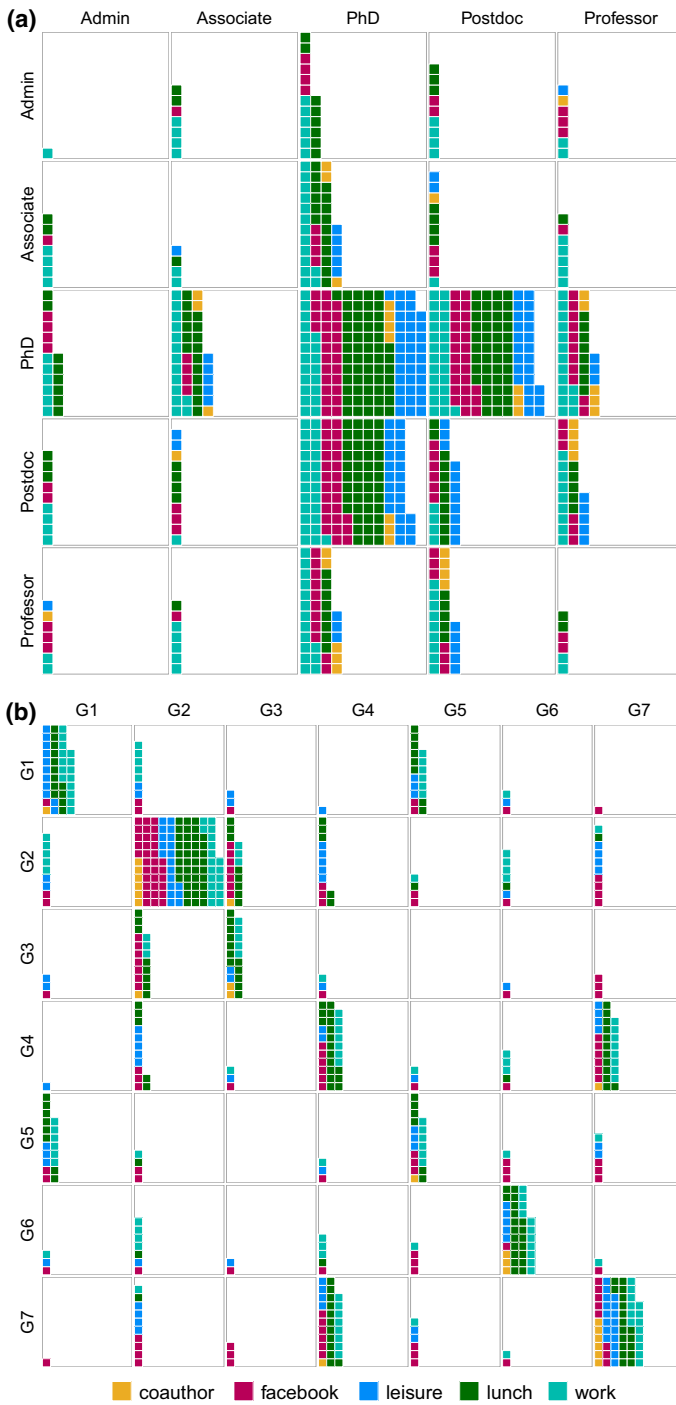
◄**Fig. 3** Aggregated multigraph based on (**a**) academic position with 5 vertices and (**b**) research group with 7 vertices. Each cell represents a vertex pair category and the diagonal represent within category edges (loops)

statistics that measure the local edge density, then varying edge proportions among the dyads in different categories would indicate the need for a model with dependent edges.

For cases with more than two types of ties, a clustering of the dyad distributions may be needed. Alternatively, a systematic review of pairwise ties can be used for cross tabulation and visualization. As an illustration, we consider the multigraph aggregated by research group and the joint and marginal distribution of work and leisure ties. The proportion of edges (work, leisure, or both) within and between the different vertex pairs sites are shown in Fig. 4. The frequency of edges relative to the total number of possible edges in each vertex pair category (y axis) is given as dots, where the size of dots represents observed frequencies. We note a tendency for the two ties to appear simultaneously within the same research group (as represented by the black dots). To obtain consistent results, we emphasize the need to systematically check these dyad distributions for all the different edges in the multigraph under study.

4.2 The sequence of edge multiplicities

We analyse the multiplexity statistics R_k given in Eq. 4 for the multigraphs in Fig. 2 in two different ways. First, we explore the distribution of these statistics given all five relations and all three aggregated multigraphs, as shown in Fig. 5. This gives an indication for patterns regarding single, double, triple, etc. occupancy of the different edges. Note that the cut off point $k = 10$ in Fig. 5 is chosen since there are only a few occurrences of more than ten edges of each type in the used data set. In principle, the value can be set to the maximally observed value. Note that the count
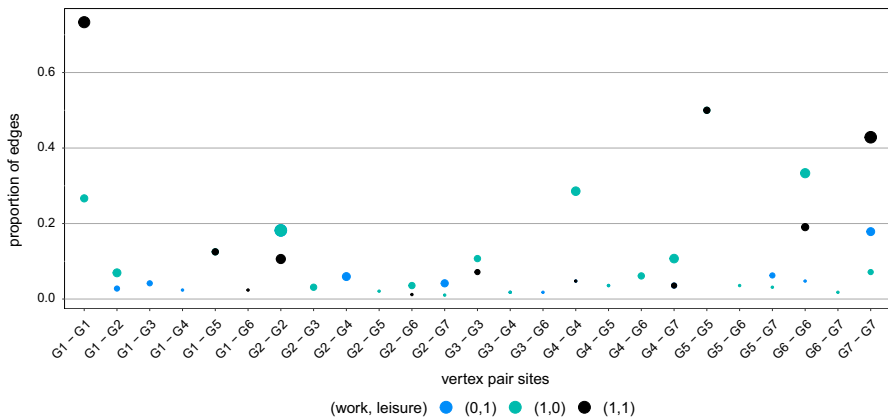

Fig. 4 Relative frequencies of leisure and work ties within and between vertex pair categories based on research group. Size of dots represent observed frequencies
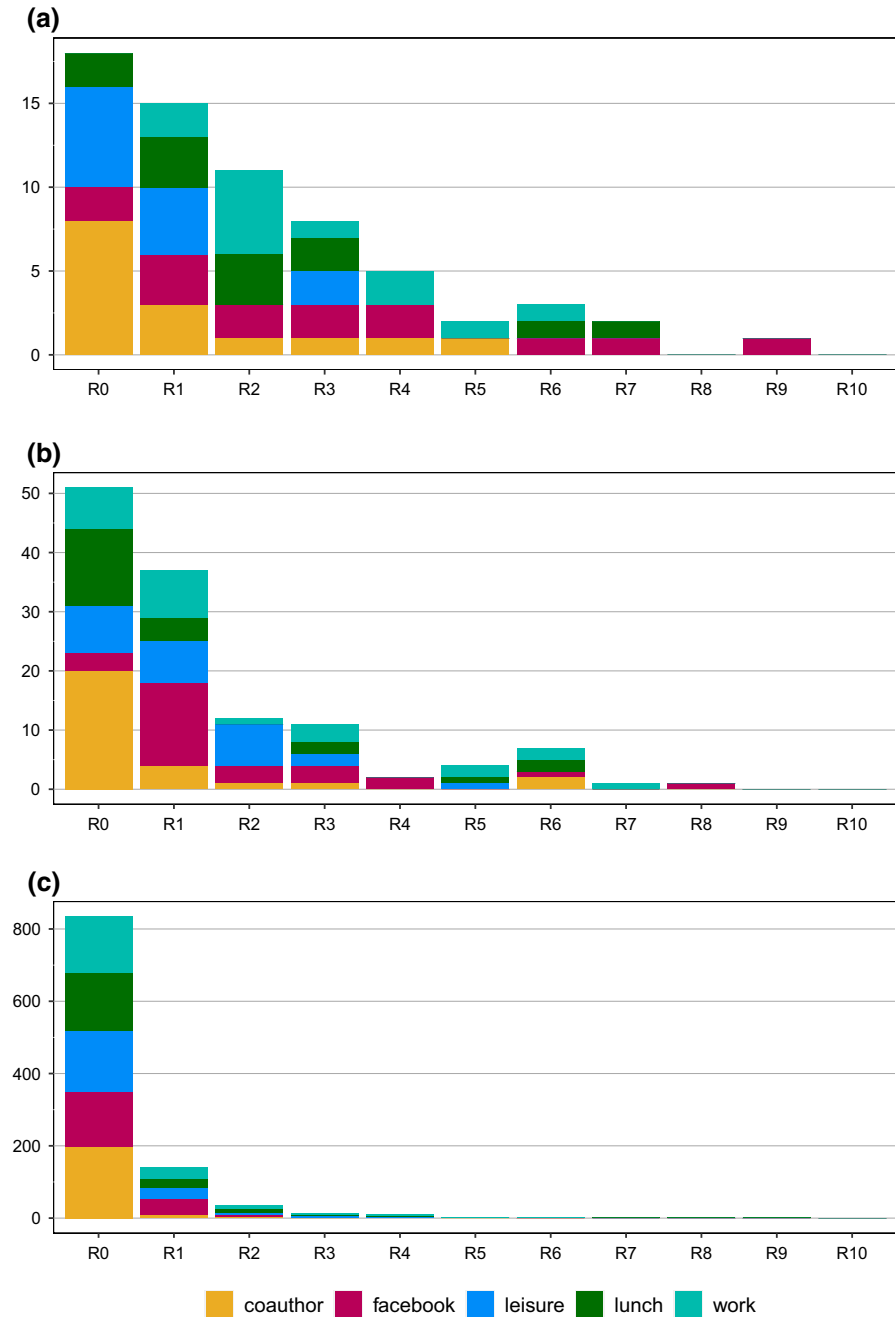
Fig. 5 Frequencies of edge multiplicities R_k for $k = 1, \ldots, 10$ in the three aggregated multigraphs in Fig. 2 based on academic position (**a**), research group (**b**) and the combination of both (**c**)

values illustrated as bars of different colours for different edge types in Fig. 5 are always relative to the same number: the number of possible vertex pair sites given by r. Specifically, for the three aggregated multigraphs considered, these numbers equal 15 (cf. Fig. 5a), 28 (cf. Fig. 5b) and 210 (cf. Fig. 5c). Note that when dividing the observed R_k by these numbers, the fractions should be interpreted differently for null occupancy of edges (R_0) and single, double or multiple occupancies of edges ($R_1, \ldots, R_k$). A high proportion of null occupancies may indicate a sparse network and a high proportion of multiple occupancies may point to a dense network.

Figure 5 reveals several tendencies in the data. The coauthor relation has the most null occupancies of edges (R_0) in the multigraphs based on academic position and research group. This, however, is due to the sparsity of the edge type. It is generally important to pay attention to the edge densities when interpreting the presented bar charts. In Fig. 5a, for instance, leisure and coauthor have a similar number of null occupancies but differing densities in the respective network. While the coauthor network is indeed sparse, leisure edges are concentrated in only a few vertex pair categories (cf. Fig. 3a). We also note that working relationships most frequently come in pairs (R_2) for the multigraph based on research group (b). For the latter, note that $R_0 + R_1$ for work ties is smaller than $R_2 + R_3 + \cdots + R_{10}$, indicating a tendency to strengthen the ties. In contrast, we also see that the multigraph aggregated based on both attributes has a majority of null and single occupancies (R_0 and R_1) for the different edges meaning there is no indication of certain relations being strengthened.

4.3 Application of the probability models

As a last step, we apply the IEA model, and the IEA approximation of the RSM model. Note that while the analyses thus far have primarily been descriptive, showing only tendencies and indications of multiplexity, we now move toward the confirmatory part of the analyses where we can formally test whether ties are associated to each other using the fitted models. Specifically, interval estimates of the R_k statistics are used to inform on the covariation of the five ties given the aggregated vertex categories in the three multigraphs shown in Fig. 2. Although unrealistic, the IEA model can still provide insight into the multiplex structure of the multigraphs and may inspire further model specifications. Thus, when testing whether the occurrence (or non-occurrences) of the various edge types are significantly different, overlapping intervals would imply failure to reject the null and indicate dependency among those edges.

Under the IEA model, the maximum likelihood estimate of the edge assignment probabilities are given by the empirical fraction of each edge type, i.e. $\hat{Q}_{ij} = m_{ij}/m$ for $1 \leq i \leq j \leq n$. Under the RSM approximation, these probabilities are given by the observed degree sequence of each edge type, as given in Eq. 2. Expected values and variance of R_k are estimated and intervals $\hat{E} \pm 2\sqrt{\hat{V}}$ are depicted in Fig. 6 where the IEA model is shown in the left column (a,c,d) and the RSM approximation is given in the right column (b,d,f). The rows represent the three aggregated multigraphs in Fig. 2. The figure allows us to check the fit of the model, that is
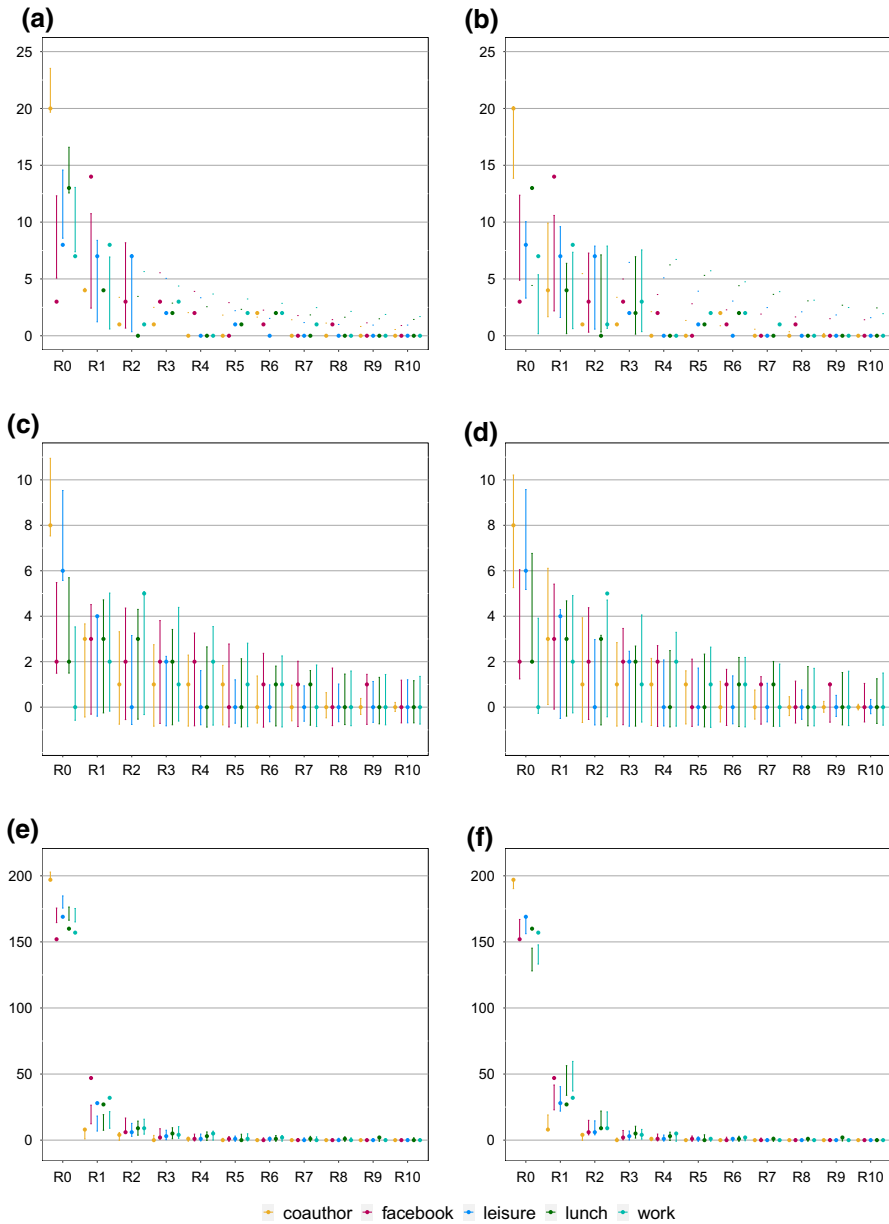
Fig. 6 Interval estimates for multiplicity statistics R_k under the IEA model (**a**, **c**, **e**) and the IEA approximation of the RSM model (**b**, **d**, **f**) for the three aggregated multigraphs in Fig. 2 based on a–b) academic position, c–d) research group, and e–f) a combination of both

whether observed frequencies of the statistics are captured by the model specified intervals. As already mentioned, overlapping intervals given the different edges hints to edge entrainment and dependence, thus implying model revision is needed.

Interpreting Fig. 6, we note the following. Both the IEA model and its approximation of the RSM model provide good fits for the multigraph aggregated based on research group as the observed values are almost all captured by their respective intervals (Fig. 6c and d). In fact, the simple IEA model in Fig. 6c provides the best fit as the intervals capture all observations (note that R_3 under the RSM approximation falls outside of the interval). This allows us to use statistics under the IEA model, such as number of loops or number of non-loops, to study homophily or heterophily effects within and between vertex categories. Moreover, multiplexity statistics defined in Sect. 4 can be used to study possible edge dependencies and associations between the different edges. We note that all intervals R_k for $k \geq 1$ in Fig. 6c and d are overlapping, thus indicating that their occurrences are not significantly different, nor are they independent implying that some form of edge dependency is needed in the model specification.

The results can guide further model revisions, for instance, when specifying multiplexity statistics in ERGMs for multivariate and multilayered networks (Koehly and Pattison, 2005; Krivitsky et al., 2020). While these model extensions go beyond the scope of this paper, we briefly outline how they can be implemented. The core principle of using ERGMs is to specify the models using conceptually suitable and relevant structural properties which are founded in social theory. Moreover, these properties need to be motivated by a hypothesis driven analysis. The presented multigraph framework not only helps to hypothesise about structural features with respect to multiplexity, but most importantly allows the filtering of relevant vertex covariates and edge types with respect to multiplexity. We exemplify this using the results in Fig. 6c and d where almost all observed values of R_k are captured by their estimated intervals, thus indicating an adequate fit. This interdependencies, shown as overlapping intervals, can be used to formulate hypotheses about multiplexity effects given the research group affiliations. For example, we might be interested in examining if staff members, given the research group, tend to eat lunch or spend leisure time with those they also work together with. It is also evident that co-authorship is entrained with work tie, but not vice versa. Such hypotheses can then serve as guidelines for statistics to include in multivariate or multilayered ERGMs. Since multiplexity is included as interaction terms between pairs of tie variables, caution must be taken to not include too many parameters in the model. This in particular since we are using covariates (here research group) implying homogeneity constraints yielding equal parameters for isomorphic configurations can not be considered. In order to avoid this, we can restrict hypotheses to models including pairwise ties and selecting those that are almost perfectly overlapping in Fig. 6c and d. For example, this holds for Facebook and lunch ties in Fig. 6c. Thus, an initial fitted ERGM would result in a positive multiplexity parameter for these ties (while controlling for density and research group). The second model could then include more multiplexity configurations that are of higher order than two vertices. Consider for example degree based effects which can be read from Fig. 6d (i.e. the approximated RSM model given degree sequences). These degree based effects can then be included in an ERGM using alternating k-star or the geometrically weighted degree configurations (see Krivitsky et al. (2020) for more details on such statistics).

5 Extension to the presented framework

The presented framework is not contingent on the presence of vertex attributes. Multigraphs can be obtained by only using the edge information of networks. Observed edge variables can be transformed into vertex variables by considering statistics such as centrality. Aggregations will be based on these transformed vertex variables, either alone or in combination with each other.

We illustrate this by using the same dataset yet discarding the two existing vertex attributes. Consider the relations Facebook friendship, leisure activities and eating lunch together. Once transformed into vertex variables, these three relations can be defined as a measurement of online (Facebook) and offline (leisure and lunch) social influence. This is done by calculating the degree of each vertex based on the three relations, which we then dichotomize to reflect low and high social influence. This dichotomisation is based on a median split of the degree distributions, as shown in Fig. 7.

Table 3 shows how these three binary vertex variables can be used to aggregate the work and co-authorship networks into a multigraph on six vertices based on online and offline social influence. The resulting multigraph is shown in Fig. 8. This multigraph can be used to examine whether there is an association between (online and offline) social influence and the two work ties by a similar approach as given in Sect. 4.

With only two type of ties present (work and coauthor), it is easy to explore the marginal and joint edge distributions of the dyads (cf. Sect. 4). By visualizing them as shown in Fig. 9, we note the following. When ordering the vertex pair sites in increasing order of social influence, we note a slight upward trend, i.e. the propensity of work ties increases with higher values of social influence. This indicates that the presence of the two edges is contingent on the attribute 'social influence', as defined in Table 3. Note that the large discrepancy between the number of co-authorship and work ties in each of the networks leads to almost all of the coauthor ties being present simultaneously with work ties. As shown in Sect. 4, the multiplexity statistics R_k can then be used to detect potential associations and dependencies among the different ties.
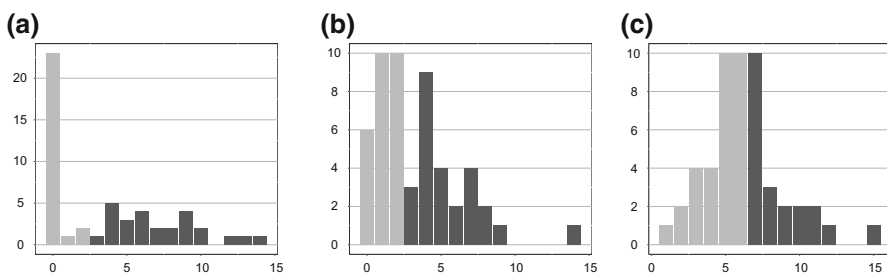

Fig. 7 Degree distributions of the three edges used to define online and offline social influence; a) facebook, b) leisure and c) lunch. The values are dichotomised to reflect low and high based on a median split of the distributions

Table 3 Dichotomised online and offline social influence, cross-classified and aggregated into a multigraph with vertex categories represented in the third column

Outcome space for social influence		
Online = (facebook)	Offline = (lunch, leisure)	(Online, Offline)
0	0 = (0,0)	(0,0)
1	1 = (0,1) or (1,0)	(0,1)
	2 = (1,1)	(0,2)
		(1,0)
		(1,1)
		(1,2)
		Vertices in aggregated multigraph

Fig. 8 Aggregated multigraph with coauthor and work relations moving within and between categories based on online and offline social influence Table 3
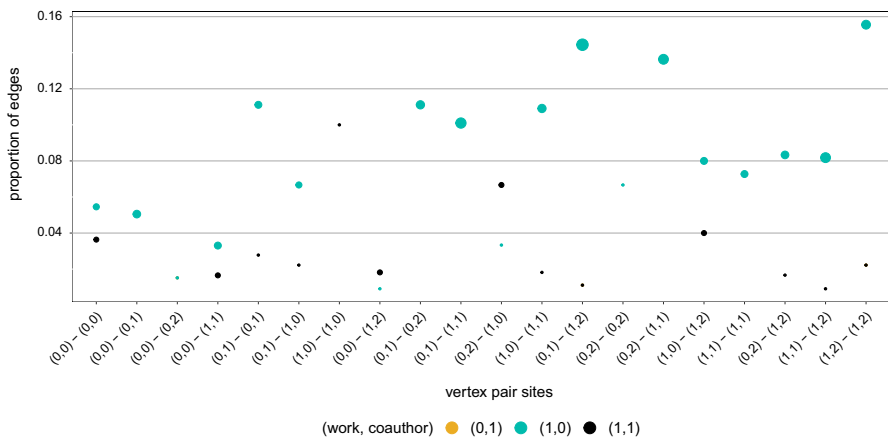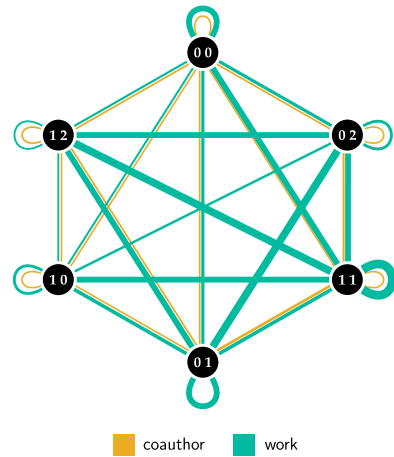
Fig. 9 Relative frequencies of work and co-authorship ties within and between vertex pair categories. Size of dots represent observed frequencies

6 Discussion

We presented a framework for analysing multiplexity in multivariate networks, based on simplifications of an observed network. The main benefit of this framework is that it allows to analyse multiplexity in the presence of vertex covariates.

The simplification of the multivariate network is based on aggregation techniques. Groups of vertices are aggregated into single meta vertices based on single or combined attributes. The aggregated network can then be represented as a multigraph with the different kinds of edges moving between and within the meta vertices. Several visual and exploratory tools were presented to detect tendencies towards tie entrainment and (conditional) dependencies among the different ties. The confirmatory analysis is performed using specific statistics derived under suitable probability models for multigraphs. While these simple models may not fit the observed data, the derived statistics can still be used to detect dependencies or other important features that need to be controlled for in a model revision.

The presented approach is particularly useful in networks with several vertex covariates. However, we also illustrated how the approach can be used when no attributes are present by transforming edge variables into vertex variables. Note that these transformations should be guided by the research question and underlying social theories. The aggregation techniques are also applicable in the presence of other exogenous network variables. Consider, for example, networks with multiple types of ties observed over several time periods. Instead of vertex aggregation, we can aggregate edges over time. Thus, we obtain a multigraph for each edge category over the total time period. These multigraphs can then again be analysed using methods and models presented in this paper.

Attention must be paid to the density of various edge types and vertex variable distributions when applying our methodology. In the extended approach presented in Sect. 5, it is easy to control for the vertex variable distribution and ensure the different categories are equally sized by using percentiles of the empirical distribution. In the specific example, edge variables are transformed into vertex variables using the median to ensure that the number of observations in each of the two outcome categories are approximately equal (see Fig. 7). In cases where actor attributes are skewed, outcome categories can be merged, split or recoded in an empirically meaningful way to balance the distribution. In case of a large discrepancy between the different observed edge densities, caution should be taken when looking at the sequence R_k. The values should not be interpreted in isolation but in conjunction with other information from the aggregated multigraphs, for instance, statistics such as density, or observations from visual inspection of the waffle matrices. This is done to rule out any tendencies due to observed edge densities (or sparsity) and tendencies due to concentrated edge occupancies in only a few vertex pair categories.

Some limitations of the proposed multiplexity framework must be noted. First, it is only applicable to undirected networks. Extending the models to the directed case is technically involved and requires a lot of additional considerations due to the

increased combinatorial complexity of the model specifications. Second, visual inspections of the waffle matrices are only feasible for small networks. Finally, the presented framework does not reveal the direction (positive or negative) of the associations between the different edge types. In order to delineate this information, complementary approaches such as those presented for ERGMs are required.

References

Bianconi G (2018) Multilayer networks: structure and function. Oxford University Press, Oxford

Bothorel C, Cruz J, Magnani M, Micenková B (2015) Clustering attributed graphs: models, measures and methods. Netw Sci 3(3):408–444

Bródka P, Chmiel A, Magnani M, Ragozini G (2018) Quantifying layer similarity in multiplex networks: a systematic study. Royal Soc Open Sci 5(8):171747

De Domenico M, Porter MA, Arenas A (2015) MuxViz: a tool for multilayer analysis and visualization of networks. J Comp Netw 3(2):159–176

Dickison ME, Magnani M, Rossi L (2016) Multilayer social networks. Cambridge University Press, Cambridge

Fienberg SE, Meyer MM, Wasserman SS (1985) Statistical analysis of multiple sociometric relations. J Am Stat Assoc 80(389):51–67

Frank O, Shafie T (2018) Random multigraphs and aggregated triads with fixed degrees. Netw Sci 6(2):232–250

Giordano G, Ragozini G, Vitale MP (2019) Analyzing multiplex networks using factorial methods. Soc Netw 59:154–170

Godehardt EA (1993) Probability models for random multigraphs with applications in cluster analysis. In: Annals of Discrete Mathematics (Vol. 55, pp. 93-108). Elsevier

Interdonato R, Magnani M, Perna D, Tagarelli A, Vega D (2020) Multilayer network simplification: approaches, models and methods. Comput Sci Rev 36:100246

Janson S (2009) The probability that a random multigraph is simple. Comb Prob Comput 18(1–2):205–225

Janson S (2014) The probability that a random multigraph is simple. II. J Appl Prob 51(A):123–137

Koehly LM, Pattison P (2005) Random graph models for social networks: multiple relations or multiple raters. In: Scott J, Wasserman S (eds) Carrington PJ. Models and Methods in Social Network Analysis, Cambridge University Press, pp 162–191

Krivitsky PN, Koehly LM, Marcum CS (2020) Exponential-family random graph models for multi-layer networks. Psychometrika 85(3):630–659

Lazega E, Pattison PE (1999) Multiplexity, generalized exchange and cooperation in organizations: a case study. Soc Netw 21(1):67–90

Magnani M, Wasserman S (2017) Introduction to the special issue on multilayer networks. Netw Sci 5(2):141–143

Merton RK (1968) The Matthew effect in science: the reward and communication systems of science are considered. Science 159(3810):56–63

Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community structure in time-dependent, multiscale, and multiplex networks. Science 328(5980):876–878

Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks: ii. multivariate relations. Br J Math Stat Psychol 52(2):169–193

Ranola JM, Ahn S, Sehl M, Smith DJ, Lange K (2010) A Poisson model for random multigraphs. Bioinformatics 26(16):2004–2011

Rossi L, Magnani M (2015) Towards effective visual analytics on multiplex and multilayer networks. Chaos Solit Fract 72:68–76

Shafie T (2015) A multigraph approach to social network analysis. J Soc Struct 16

Shafie T (2016) Analyzing local and global properties of multigraphs. The J Math Sociol 40(4):239–264

Snijders TA, Lomi A, Torló VJ (2013) A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. Soc Netw 35(2):265–276

Wasserman S, Faust K (1994) Social network analysis: methods and applications, vol 8. Cambridge University Press

White HC, Boorman SA, Breiger RL (1976) Social structure from multiple networks. i. blockmodels of roles and positions. Am J Sociol 81(4):730–780