Journal of Social Structure

JoSS Article: Volume 16

A Multigraph Approach to Social Network Analysis

Termeh Shafie

Department of Computer & Information Science

University of Konstanz

termeh.shafie@uni-konstanz.de

Abstract

Multigraphs are graphs where multiple edges and edge loops are permitted. The main purpose of this article is to show the versatility of a multigraph approach when analysing social networks. Multigraph data structures are described and it is exemplified how they naturally occur in many contexts but also how they can be constructed by different kinds of aggregation in graphs. Special attention is given to a random multigraph model based on independent edge assignments to sites of vertex pairs and some useful measures of the local and global structure under this model are presented. Further, it is shown how some general measures of simplicity and complexity of multigraphs are easily handled under the presented model.

Keywords: multigraph, edge multiplicity, edge loop, data aggregation, random multigraph, complexity.

1 Introduction

Network data involving relational structure representing interactions between actors are commonly represented by graphs where the actors are referred to as vertices or nodes, and the relations are referred to as edges or ties connecting pairs of actors. Research on social networks is a well established branch of study and many issues concerning social network analysis can be found in Wasserman and Faust (1994), Carrington et al. (2005), Butts (2008), Frank (2009), Kolaczyk (2009), Scott and Carrington (2011), Snijders (2011), and Robins (2013).

A common approach to social network analysis is to only consider binary relations, i.e. edges between pairs of vertices are either present or not. These simple graphs only consider one type of relation and exclude the possibility for self relations where a vertex is both the sender and receiver of an edge (also called edge loops or just shortly loops). In contrast, a complex graph is defined according to Wasserman and Faust (1994):

> If a graph contains loops and/or any pairs of nodes is adjacent via *more than one* line the graph is *complex*. [p. 146]

In practice, simple graphs can be derived from complex graphs by collapsing the multiple edges into single ones and removing the loops. However, this approach discards information inherent in the original network. In order to use all available network information, we must allow for multiple relations and the possibility for loops. This leads us to the study of multigraphs which has not been treated as extensively as simple graphs in the literature.

As an example, consider a network with vertices representing different branches of an organisation. The edges apparent in such a network may then comprise of information, money, and personnel flows including cooperation, support, friendship and antagonism. These different edges should be considered simultaneously in order to understand the inter-organisational behaviour. Robins and Pattison (2006) emphasise analysing these different edges jointly to understand social processes in the network and its implications for an organisation's performance. In an organisational network, it is also evident that different kinds of ties may appear within the same branch creating loops. For instance, friendships may be more common between individuals within a branch, which may also indicate a higher propensity to turn to these friends for advice or support. A multigraph has the capacity to gather and represent all of this information.

A common definition of multigraphs (also called multiple networks) is graphs having several kinds of ties on the same vertex set (e.g. Robins 2013; Ranola et al. 2010; Koehly and Pattison 2005). However, by convention, loops are in many cases excluded from consideration. We emphasise the fact that social network data often comprise of loops in their natural state and include this in our definition: multigraphs are graphs where multiple edges and loops are permitted.

In this paper we present a new multigraph approach that may be used to analyse networks with multiple edges and loops of different kinds. We describe multigraph data structures with examples of their natural appearance together with a description of the possibility to obtain multigraphs using blocking, aggregation and scaling. A novel way of representing multigraphs using edge multiplicities is introduced and we quantify graph complexity by the distribution of edge multiplicities. A random multigraph model based on independent edge assignments (IEA) to sites of vertex pair is given and we derive several complexity statistics under IEA. Further, it is described how these measures can be used to analyse local and global network properties and to convey structural dependencies in social networks. Finally, a brief discussion is given on the possibilities and limitations of the presented multigraph approach, together with suggestions for future reasearch topics.

Through out all sections we use the well-known data set of Padgett's Florentine families (Padgett 1987) as an empirical example to illustrate our presented approach. Note that our primary purpose for using this data set is not to derive new conclusions on processes that determine marital and financial relations given actor attributes between pairs of families. There is already a wide range of analyses available in the literature (e.g. Wasserman and Faust 1994; Padgett and Ansell 1989, 1993; Breiger and Pattison 1986). Our primary use of this data is to illustrate our multigraph approach, to show the applicability of the presented multigraph model, and to highlight the usefulness of the derived statistics.

2 Why Multigraphs?

A multigraph, also called multiple or multivariate network, consists of a set of actors, and a collection of relations that specify how pairs of actors are related to each other (Wasserman and Faust 1994, Pattison and Wasserman 1999). In a multiple network, there are different types of relations where each one can be seen as binary variable on the set of actors constituting a univariate network. It is well known that studying multiple networks is the key to fully understand the patterning and intertwining of relations that characterise structural features latent in a network (Robins and Pattison 2006; Koehly and Pattison 2005). The problem of finding the interdependencies of social relations has a long theoretical history (Pattison 1993; White 1963; White, Boorman and Breiger 1976). Some of these studies are summarised here and followed by a discussion on how they relate to the multigraph approach presented in this paper. Our approach is a simple way to analyse multigraphs and differs from previous approaches since it allows for the occurence of edge loops. We start by a description of multigraph data structures and how they may appear natural in many contexts but also how they may be obtained by different kinds of aggregation, blocking and scaling. It is emphasised how loops may appear in these different forms of structures and why they should not be ignored in the study of social networks. The several ways for obtaining multigraphs point to the importance of developing multigraph models to fully understand their structural properties.

2.1 Multigraph Data Structures and Possible Applications

A fundamental unit of analysis in social networks is that of a dyad which is defined as a vertex pair of distinct actors, with sites between them for possible edges. For multigraphs we refrain from using this notion since the two elements of a dyad may comprise of the same actor with possible loops. Instead, we use the concept of vertex pairs defined as ordered or unordered pairs of possibly undistinct vertices, where the former is for directed and the latter is for undirected graphs.

Multigraph data structures can be observed directly and are common in contexts where several edges can be mapped on the same vertex pair, for instance social interactions of different kinds between a group of individuals (e.g. friends, colleagues, neighbours) or contact types (phone call, email, instant message) between and within departments of an organisation. Note that within departmental edges correspond to loops. Other multigraphs that easily exhibit themselves are given by edge variables that count occurrences of events at different sites of vertex pairs. Less obvious multigraphs can also be identified or created from network data. This is now illustrated by a few examples that briefly indicate the various possibilities, together with suggestions for related research topics.

Social networks of contacts between people are seen as more than just binary relations if time is taken into consideration. Contacts might appear and disappear, and their durations might be recorded. Total contact time or total number of contacts taken during a fixed time period can be recorded at different sites and represented by multigraphs. To list a few examples; hyperlink connections between websites, money transactions between bank accounts, and connections between email users during a period of time. Note that in many cases, loops are directly observed as well. In the study of hyperlink connections between websites, loops appear as links from a given website to itself.

Other properties of (undistinct) pairs of vertices can also be interpreted as valued graphs, which can be represented as multigraphs by using appropriate integer scales for the values. Typical examples are strength or importance of contacts. Other examples are measures of capacity or potential for transmission of information or flow of goods.

Recent interest in very large graphs obtained from the internet can inspire to various kinds of aggregation that creates multigraphs. Two such aggregations are described in more detail below, where the first includes the possibility of loops while the second does not.

A large simple graph can be transformed to a multigraph by aggregating vertices to blocks of vertices, and counting edges within and between blocks. In other words, creating a vertex set (by any partitioning or blocking) generates data that usually and usefully may include loops. These loops need to be included in the analysis of the network to reveal social processes at play. Secondly, a collection of simple graphs representing different binary relations on the same or partly the same set of vertices can be aggregated over relations so that a single multigraph (without loops) is obtained. Note that big data is not necessary for the utility of multigraphs. Rather it is sufficient with big data to make good use of aggregation to multigraphs, and usefulness depends on how well the obtained multigraphs can be analysed. In other words, simplification by aggregation is important (like all simplifications in analysis) and therefore multigraphs are useful if good methods and models are at hand for understanding them.

A research topic related to the study of multigraphs is that of network sampling. When information about a population network is collected by sampling of paths, trees, or other subgraphs that may be partly overlapping, there might be useful information in observed multiplicities of vertices and edges. Generally, several sampled subgraphs can be combined to multigraphs reflecting both the underlying population graph and the properties of the sampling design. Snowball samples with several stages are usually treated with focus on distinct sites without multiplicities, which implies difficult problems for design based statistical inference. Such issues are surveyed by Frank (2009; 2011). The complications caused by overlapping snowball stages might suggest multigraph modelling as an alternative approach to handle snowball sample data.

2.2 Related Work

Many methods for describing structural properties in multi-relational networks correspond to studying structural similarities of actors and finding patterns among the relations. These methods are referred to as social position and role analysis. The common aim of these methods is to simplify the representation of the network data to uncover subsets of actors into positions based on their relational similarity (White et. al 1976) and describe the associations among these relations based on how they combine to connect actors or positions (Boorman and White 1976; Breiger and Pattison 1986).

A positional analysis maps actors into equivalence classes, where an equivalence class consists of all actors who are identical on a specified property. Among the first equivalence definitions is Lorraine and White's (1971) notion of structural equivalence which lead to the introduction of blockmodel by White et. al (1976). A blockmodel is a hypothesis about a network and presents simplified and general features of a network's structure through the relations between pairs of social positions or blocks, where each position is defined by a collection of structurally equivalent actors. For two actors to be structurally equivalent in a multi-relational network, they must have identical relations to and from all other actors on all edges. When reflexive ties or loops are substantively important in the study, graph equivalence (Guttman 1977) has been defined which is less general than structural equivalence.

The interdependence analysis of network relations are mostly algebraic in character, where labelled paths for multiple network relations are constructed (Boorman and White 1976; Pattison 1993) or of more general connectivity structure (Doreian 1980;1986). These methods are deterministic which means that algebraic model hypothesis on the set of network relations can not be evaluated. For this purpose, relationships between the algebraic expression of path dependencies and classes of possible network statistics are used in exponential random graph models (ERGMs). Exponential random graph models are statistical modelling approaches that can be used to find common patterns of small network configurations. These configurations identify common patterns and regularities in actors' immediate network environment (Lusher et. al 2012). For multiple network studies, they are particularly useful because they give evidence of how different networks relate to one another. Pattison and Wasserman (1999) present the exponential multigraph model (ERMM) which is based on hypotheses about the dependencies among multiple network tie variables. Applications of ERMMs can be found in Koehly and Pattison (2005) and Lazega and Pattison (1999). We will return to the ERMM and its representation for our multigraph model in Section 5.

In positional analysis, there are different ways to represent the relations within and between blocks: an image matrix, a reduced graph and a density table (Faust and Wasserman 1992). In an image matrix, each edge is coded as either present or absent between each pair of positions. A reduced graph has the blocks as vertices and uses the relational information in an image matrix to define the edges between the vertices. Further, if the block structure is represented by colors, we have a colored multigraph. A density table is a matrix that has blocks as its rows and columns, and the values in the matrix are the proportion of edges that are present from the actors in the row positions to the actors in the column positions. All of these representation forms are options for presenting aggregated multigraph data. However, it should be noted that for very large multigraphs, an image matrix and a reduced graphs are not easily interpreted. A good representation option similar to a density table is a table with densities replaced by the edge multiplicities within and between the position categories.

There are clear similarities between the aggregation techniques presented for multigraphs and positional analysis used for blockmodelling. The purpose of positional analysis is to simplify the data information by representing it in terms of the positions identified by the equivalence definition and to provide an interpretation of the results. As mentioned previously, any aggregation of an original graph to a multigraph is useful as long as there are good methods of interpreting them. This holds whether aggregation is based on single or combined categories of actor attributes, different equivalence definitions, edge attributes or relational algebras, or over different time periods.

Another common method in network analysis is to compare observed networks against null distributions of graphs with certain properties. For the univariate network, one can compare features of an observed graph to the distribution of the corresponding feature in simulated uniform graph distribution conditional on one or more network properties. If the observed graph's feature is extreme compared to graphs in this distribution, it may be assumed that the observed graph is unlikely to have been produced from that distribution, and thus may convey some social structural phenomena. For the case of multiple networks, a similar procedure can be used with conditional uniform multigraph distributions. Pattison et. al. (2000) present a collection of properties of multiple networks to condition on and illustrate how a sequence of conditional uniform multiple graph distributions can be used to evaluate hypothesised sets of relations of ordering and equality among labelled paths in a multiple network (so called algebraic constraints).

Under the assumption of dyadic independence, a null distribution commonly used for univariate networks is the Bernoulli random graph distribution in which edges have fixed probability of occurence. Wasserman (1987) generalised this distribution to multivariate networks without loops that can been used to test hypothesis about multi-relational structure. The multigraph model we present in Section 4 is the multi-relational equivalence of the Bernoulli graph with independent edges assigned to sites of vertex pairs, including the possible (and multiple) occurence of loops. The presented model can in similar fashion be used as a null distribution for which observed multi-relational network properties can be compared to. For instance, we can compare actual multigraphs to the random multigraph model to check whether it could have been generated by a random process.

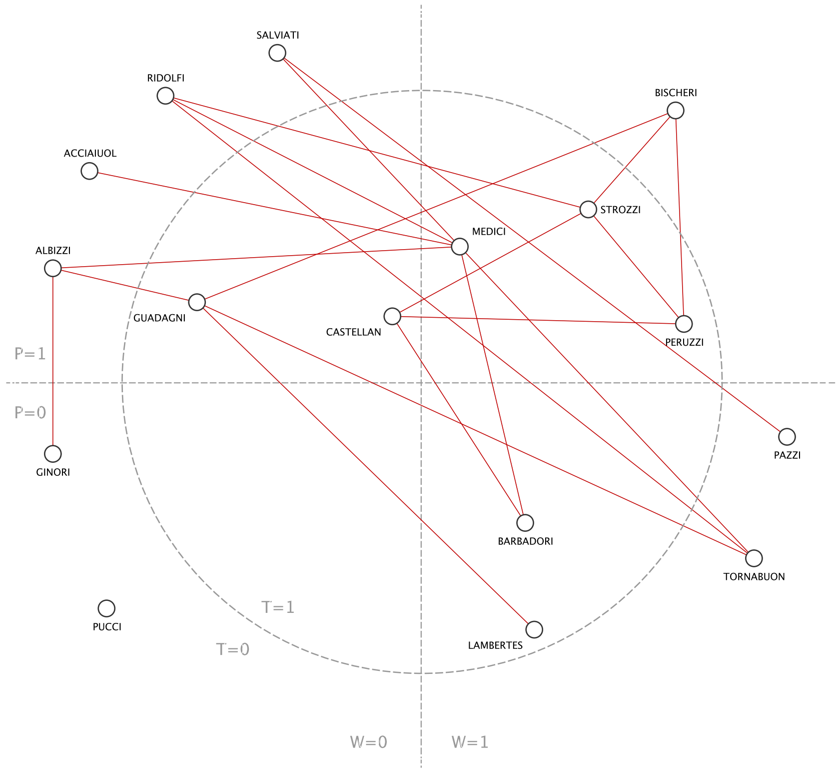
In the context of dyad independent models and our descriptive multigraph analysis presented in Section 3, we mention stochastic blockmodels with latent vertex categories or classes. Extending the approach of Snijders and Nowicki (1997), Nowicki and Snijders (2001) introduce the stochastic version of latent blockmodels for multiple network. This model assumes that the probability distribution of the relations between two vertices depends only on the latent classes to which the vertices belong and that the relations are independent conditional on these classes.

2.3 Example: Florentine Network

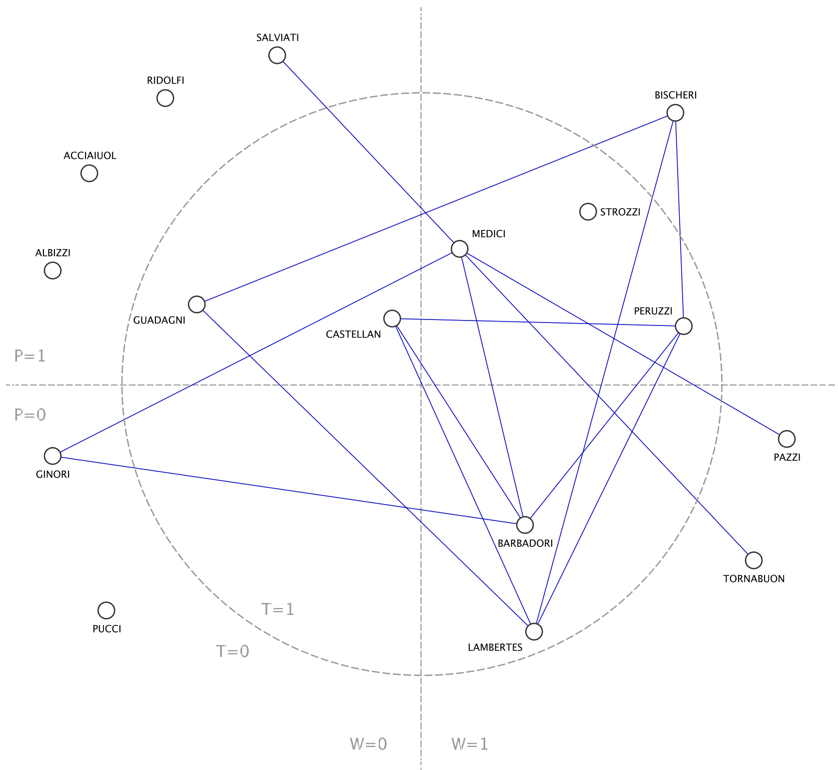
To illustrate some of the theoretical discussions on multigraphs, we use a running example based on social relations among Renaissance Florentine families. This network data was collected by Padgett (1987) with Kent's (1978) compilation from the history of this period as the primary source (see also Padgett and Ansell 1989, 1993). More specifically, we use the same subset of this data as Breiger and Pattison (1986) which is a sample of 16 out of 116 leading Florentine families because of their historical prominence. The actors in this network are the 16 families with two measured relations that are both binary and symmetrical: marital (M) and financial (F), such as credits, loans and business partnership. Further, there are three actor attributes: net wealth in thousands of lira in 1427 (denoted W), number of priorates i.e. number of seats on the Civic Council between 1282 and 1344 (denoted P), and the number of business and marriage ties in the total network data set consisting of the 116 families (denoted T). We will refer to these attributes as economic, politic and social influence, respectively.

This multi-relational network is a multigraph in its original form, with two types of ties on the same set of actors, but with no loops. However, we use this multigraph to show how aggregation can be done based on single and combined categories of actor attributes, with edges moving within and between these categories. We will start by using a network visualisation common for multiple networks where the two different edge types are depicted in separate diagrams with fixed vertex positions. Figure 1 shows (a) the marriage alliances and (b) financial ties based on the dichotomisation of the three attributes W, P and T. These binary values are chosen to reflect weak (= 0) and strong (= 1) economic, political and social influence.[1] Figure 1 can be used to get an impression of differences and similarities between the two edge types across different actor attributes. For instance we note that marital relations are more dense in the top part of graph (a) where $P = 1$, and financial relations are dense for $W = 1$ given in the right part of graph (b). We aggregate these graphs into multigraphs over all possible combinations of the actor attributes as shown in Figure 2. We return to a more thorough discussion on conditional dependencies of edge types given actor attributes in the next section.

[1] The attribute values are set to 1 if $W > 40$, $P > 0$ and $T > 10$, respectively. For a full list of this dichotomisation, see Table 3 in Appendix.



(a)



(b)

Figure 1: Graphs of (a) marriage and (b) financial relations between the 16 Florentine families arranged to illustrate their binary actor attributes wealth (W), number of priorates (P), and total number of ties (T).

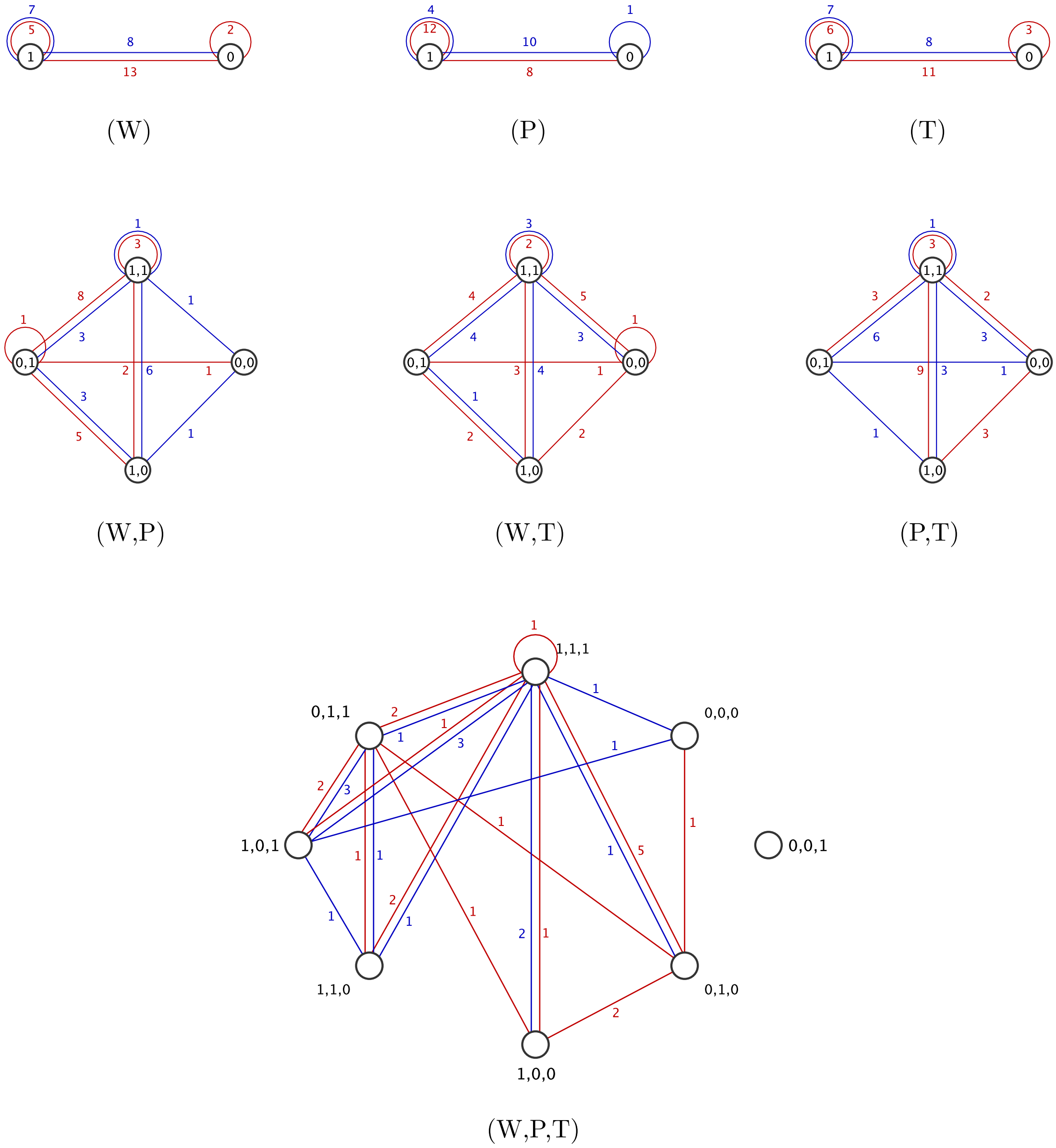


Figure 2: Aggregated multigraphs of the 16 Florentine families with marriage (red) and financial (blue) relations moving between and within categories based on actor attributes wealth (W), number of priorates (P), and total number of ties (T). The vertex labels indicate binary attribute combinations and the edge labels indicate the edge multiplicities.

3 Exploratory Analysis of Multigraphs

Probabilistic graph models allow us to understand the uncertainty connected to the observed network and in particular, we gain information of the distribution of different network outcomes for a given specified model. Furthermore, probabilistic models allow for inferences about different network structures and properties.

Network modelling is often simplified if independence between the vertices and independence between the edges can be assumed. However, in order to specify realistic probability distributions for social networks, formalising hypothetical structural dependencies based on a given data is required since social networks often exhibit different forms of dependencies. Dependency among dyads (or higher-order structures) is what makes network modelling difficult and is often the essence of social network theories.

Although, the multigraph model presented in this paper is based on independent edge assignments to sites of vertex pairs, we briefly review some different forms of dependencies that may arise in multigraph. For a more thorough discussion on structural dependencies we refer the reader to Snijders (2011), Robins and Pattison (2006), Koehly and Pattison (2005), and Pattison and Wasserman (1999). Further, we discuss descriptive methods for finding dependence between vertex pairs in multigraph data which is essential for finding a suitable network model. In particular, we refer to the methods of cluster analysis and cross classification presented in Frank and Nowicki (1993), Frank et al. (1985a), and Frank et al. (1985b).

3.1 Dependence Structures in Multigraphs

Multigraphs are dynamic and different types of edges influence and reshape each other. These multiple relations may interlock and determine one another. A common form of dependence is on the basic edge-level and is called multiplexity or entrainment, where the presence of a tie affects the presence or absence of another. We might also note dependence in form of dyadic exchange, also called reciprocity, where one type of edge from actor i to actor j may be conditionally dependent on other edges from j to i. Another possible dependence structure is connected to role algebra and the patterns of interlocking between different kinds of edges, so called role interlocking in social networks (White 1977; Boorman and White 1976; Lorraine and White 1971). Pattison (1993) presents these interrelationships by partial ordering labelled paths, where labelled paths systematically trace connections among sequences of individuals. A last form of dependence mentioned here is that of positional effects in blockmodels where edge patterns between pairs of actors in different social positions appear. This implies that actors in similar positions may exhibit similar conditional dependencies among their relations (Nowicki and Snijders 2001; Snijders and Nowicki 1997). Further, and as Pattison and Wasserman (1999) remark, several of these different dependence arguments may be combined resulting in the interdependence of interlocking roles.

We now turn to methods for detecting different structural dependencies. In cluster analysis, the actors are categorised into clusters so that actors within a subset are more similar than individuals in different subsets. These clusters can be found through vertex aggregation described in previous section so that we have homogeneity within each cluster, and heterogeneity between the clusters. After clusters have been identified, the approach is to try separate models within and between different clusters.

Another way to decide whether there is need for a model with edge dependencies is to cross classify the dyads (or vertex pairs) according to different statistics of interest. The dyad counts in different categories can then be counted and analysed. To find interesting interaction effects between the statistics, log-linear and logit analysis can be applied. This approach will then also allow us to analyse the edge proportions among the dyads in different categories. For instance, if we are interested in statistics that measure the local edge density, then the discovery of different edge proportions among the dyads in different categories indicate the need for a model with dependent edges. Note however that an initial clustering of the dyad distributions may be necessary to obtain a reasonable number of dyad distributions for which log-linear model fitting can be used.

3.2 Example: Florentine Network

We now turn to Padgett's network of Florentine families where cross classification of vertex pairs can be used to determine the covariation of actor and edge attributes. By systematically going through the relative frequency or proportions of edges within and between vertex pair categories, we can detect conditional dependencies of edges given actor attributes. However, since this data material is small, we simply arrange the data and use these edge proportions to detect tendencies in terms of how economic (W), political (P) and social influence (T) covary with marital (M) and financial (F) relations. This arrangement is given in Table 1 for the first row of multigraphs in Figure 2, where each of the three multigraphs represent aggregation based on attributes (a) wealth, (b) number of priorates, and (c) total number of ties. Since we are considering binary actor attributes, we get three vertex pair categories for each of the three aggregated multigraphs. Table 1 represents the frequency distribution of the presence and absence of marriage (M) and business (F) ties within and between these three categories, for each aggregated multigraph. Note that the total number of vertex pairs for each multigraph is equal to $\binom{16}{2} = 120$.

(a) Multigraph aggregated by wealth

W	W		F=0	F=1	Total	Total
0	0	M=0	26 (0.93)	0 (0)	26 (0.93)	
		M=1	2 (0.07)	0 (0)	2 (0.07)	
		Total	28 (1.00)	0 (0)	28 (1.00)	28 (0.23)
1	1	M=0	19 (0.68)	4 (0.14)	23 (0.82)	
		M=1	2 (0.07)	3 (0.11)	5 (0.18)	
		Total	21 (0.75)	7 (0.25)	28 (1.00)	28 (0.23)
0	1	M=0	48 (0.75)	3 (0.05)	51 (0.80)	
		M=1	8 (0.12)	5 (0.08)	13 (0.20)	
		Total	56 (0.87)	8 (0.13)	64 (1.00)	64 (0.53)
Total		M=0	93 (0.78)	7 (0.06)	100 (0.84)	
		M=1	12 (0.10)	8 (0.07)	20 (0.17)	
		Total	105 (0.88)	15 (0.13)		120 (1.00)

(b) Multigraph aggregated by priorates

P	P		F=0	F=1	Total	Total
0	0	M=0	14 (0.93)	1 (0.07)	15 (1.00)	
		M=1	0 (0)	0 (0)	0 (0)	
		Total	14 (0.93)	1 (0.07)	15 (1.00)	15 (0.13)
1	1	M=0	33 (0.73)	0 (0)	33 (0.73)	
		M=1	8 (0.18)	4 (0.09)	12 (0.27)	
		Total	41 (0.91)	4 (0.09)	45 (1.00)	45 (0.38)
0	1	M=0	46 (0.76)	6 (0.10)	52 (0.86)	
		M=1	4 (0.07)	4 (0.07)	8 (0.14)	
		Total	50 (0.83)	10 (0.17)	60 (1.00)	60 (0.50)
Total		M=0	93 (0.78)	7 (0.06)	100 (0.84)	
		M=1	12 (0.10)	8 (0.07)	20 (0.17)	
		Total	105 (0.88)	15 (0.13)		120 (1.00)

(c) Multigraph aggregated by ties

T	T		F=0	F=1	Total	Total
0	0	M=0	33 (0.92)	0 (0)	33 (0.92)	
		M=1	3 (0.08)	0 (0)	3 (0.08)	
		Total	36 (1.00)	0 (0)	36 (1.00)	36 (0.30)
1	1	M=0	12 (0.57)	3 (0.14)	15 (0.71)	
		M=1	2 (0.10)	4 (0.19)	6 (0.29)	
		Total	14 (0.67)	7 (0.33)	21 (1.00)	21 (0.18)
0	1	M=0	48 (0.76)	4 (0.06)	52 (0.82)	
		M=1	7 (0.11)	4 (0.06)	11 (0.17)	
		Total	55 (0.87)	8 (0.12)	63 (1.00)	63 (0.53)
Total		M=0	93 (0.78)	7 (0.06)	100 (0.84)	
		M=1	12 (0.10)	8 (0.07)	20 (0.17)	
		Total	105 (0.88)	15 (0.13)		120 (1.00)

Table 1: Frequencies of marriage (M) and financial (F) edges within and between vertex pairs of multigraphs aggregated by binary actor attributes (a) wealth (W), (b) number of priorates (P), and (c) total number of ties (T).

We start by looking at the proportions or percentages (in parenthesis) within the cells of each vertex pair category and note the following. Financial relations ($F = 1$) have no occurence (0%) for WW=00 and T=00, indicating that there are no financial edges within the category of families (i.e. loops) with low wealth and low total number of ties, respectively. Marital relations (M=1) have no occurence (0%) for PP=00, that is there are no marital loops for families with low number of priorates. Next, we look at cases where F=1 and M=1 have a high number of occurences, but first we need to define what we mean with high. The frequency of vertex pairs with edge combinations FM=00, 01, 10, 11 are equal to 93, 12, 7 and 8, or as percentages; 78, 10, 6, and 7 % (this can be read in the bottom part of each sub-table of Table 1). Thus, $F = 1$ has 13 % of all vertex pairs and M=1 has 17 % of all vertex pairs. One can reason that edge occurences of each type exceeding these values are considered high. This occurs for $F = 1$ which has 25 % of the vertex pairs with WW=11, and 33% with TT=11. Further, M=1 has 27 % of the vertex pairs with PP=11, and 29 % with TT=11. From these results we can conclude that there is a tendency for many financial relations within the most wealthy families and within the most social families, respectively. There are also many marital relations within the most politically influential families but also within the most social. We also detect tendencies for financial relations between families with high and low political influence with 17 % of the vertex pairs with PP=01.

Although not presented here, a similar investigation for all aggregated multigraphs in Figure 2 was performed. The main results are briefly summarised. Marital relations have high frequencies within families with high wealth and high priorates, while financial relations are most frequent between wealthy families with low and high political influence (i.e. between vertices WP=10 and WP=11). When looking at the 8 vertex multigraph aggregated on all three attributes, we note that financial relations are most frequent between vertices WPT=101 and WPT=011, while marital relations occurs most frequently between vertices WPT=110 and WPT=111. This can be interpreted as social influence being an important factor for financial alliances, but not for entering marriages.

4 Random Multigraph Model

After formulating hypotheses about the interdependencies of the (multiple) edges, appropriate multigraph models should be developed. Statistical graph models serve several purposes, for instance explaining and predicting social relations and behaviour. A desirable goal of a model is to best represent our observed data, i.e. to reproduce the structures we witness in our observed network. Models may also represent theories we may have about the observed data, and fitting a model permits us to see if our theoretical conception about the data can be validated.

Uncertainty is certainly present in many multigraphs and this is sufficient reason for stochastic approaches. Statistics can formulate precise statements about uncertainty, i.e. what would happen if we measured the data at another point in time, on a different set of actors or under different environmental factors? By incorporating randomness, statistical models deal with expected values so we can draw inference about whether observed data are consistent with expectations.

A random multigraph is a family of multigraphs with a probability distribution, and appropriately chosen it can be a model for a considered application. We present a random multigraph model where undirected edges are independently assigned to sites of pairs of vertices according to a common probability model. It is well known that the independence assumption is unrealistic for social networks. However, it can serve as a baseline for comparison, as done for univariate networks by the Bernoulli dependence assumption (see Section 2.2). Further, we can estimate expected outcomes of local and global structural properties under this model and use the variances to determine the reliability of the estimates. Moreover, it is possible to perform statistical goodness-of-fit tests between an observed multigraph and the model presented here. Thus, by performing these tests, deviations from independence or other models in the observed multigraph data can be detected.

4.1 Independent Edge Assignment (IEA)

Random multigraph models on a vertex set $V = \{1, \dots, n\}$ assign edges randomly to sites of ordered or unordered pairs of vertices. If (X_{2k-1}, X_{2k}) is the site of edge k for $k = 1, \dots, m$, the sequence $\mathbf{X} = (X_1, \dots, X_{2m})$ represents a general directed multigraph with m edges. Here $\mathbf{X}$ is a sequence in R^m where R is the site space for edges which is equal to V^2. An undirected multigraph can be obtained by identifying the ordered vertex pairs (i, j) and (j, i) with the unordered vertex pair and define its canonical representation by the ordered pair (i, j) where $i \leq j$. The canonical site space for undirected edges is then given by $R = \{(i, j) : 1 \leq i \leq j \leq n\}$. With this convention the sequence $\mathbf{X}$ can be interpreted as an undirected multigraph represented by a sequence $\mathbf{Y} = (Y_1, \dots, Y_{2m})$ where (Y_{2k-1}, Y_{2k}) is defined as

$$(Y_{2k-1}, Y_{2k}) = \begin{cases} (X_{2k-1}, X_{2k}) & \text{if } X_{2k-1} \leq X_{2k} \\ (X_{2k}, X_{2k-1}) & \text{if } X_{2k} \leq X_{2k-1} \end{cases} \tag{1}$$

for $k = 1, \dots, m$.

A basic random multigraph is obtained by independent assignment of edges to vertex pairs in the site space R. Let $\mathbf{Q} = (Q_{ij} : (i, j) \in R)$ denote the assignment probabilities given to different possible sites. They satisfy $Q_{ij} \geq 0$ and $\sum\sum_{(i,j)\in R} Q_{ij} = 1$. Assume that m edges labelled $k = 1, \dots, m$ are independently assigned to vertex pairs $(Y_{2k-1}, Y_{2k}) \in R$, and that the multigraph is given by the sequence $\mathbf{Y} = (Y_1, \dots, Y_{2m})$ of sites for its edges. If we do not want to distinguish the edges, it is sufficient to represent the multigraph by its edge multiplicities at different sites. Let the sequence of random edge multiplicities be denoted by $\mathbf{M} = (M_{ij} : (i, j) \in R)$ where the counts

$$M_{ij} = \sum_{k=1}^{m} I(Y_{2k-1} = i, Y_{2k} = j) \tag{2}$$

satisfy $M_{ij} \geq 0$ and $\sum\sum_{(i,j)\in R} M_{ij} = m$.

The assumptions imply that $\mathbf{M}$ is multinomially distributed with parameters m and $\mathbf{Q}$. This is written $\mathbf{M} \sim \text{mult}(m, \mathbf{Q})$ for $\mathbf{Y} \sim \text{IEA}(m, \mathbf{Q})$. The number of possible outcomes $\mathbf{M}$ can be combinatorially obtained as the number of ways to distribute m balls (edges) in r urns (sites) and is given by

$$\binom{m + r - 1}{r - 1} , \tag{3}$$

where r is the number of possible sites in R and equal to $r = \binom{n+1}{2}$ for undirected multigraphs. The probability of $\mathbf{M} = \mathbf{m}$ is given by

$$P(\mathbf{M} = \mathbf{m}) = \binom{m}{\mathbf{m}} \mathbf{Q}^{\mathbf{m}} = \frac{m!}{\prod_{i \leq j} m_{ij}!} \prod_{i \leq j} Q_{ij}^{m_{ij}} , \tag{4}$$

for all possible outcomes $\mathbf{m} = (m_{ij} : (i, j) \in R)$. From this it follows that the random edge multiplicity M_{ij} is binomially distributed with parameters m and Q_{ij} for $i \leq j$. Thus, the expected values and variances of the edge multiplicities are given by

$$E(M_{ij}) = mQ_{ij} , \tag{5}$$

and

$$Var(M_{ij}) = mQ_{ij}(1 - Q_{ij}) , \quad \text{for } i \leq j . \tag{6}$$

Note that the loop multiplicities at vertex i correspond to when $i = j$. In the limit of large m, the edge multiplicity M_{ij} is approximately Poisson distributed with parameter $\lambda_{ij} = mQ_{ij}$ for $i \leq j$.

4.2 Complexity and Simplicity Measures

Random graph models can be characterised by the expected values of certain statistics, e.g. expected number of edges, triangles, or vertices of certain degrees. Using the results obtained for edge multiplicities under IEA, we derive expected values and variances of some statistics that can be used to study simplicity and complexity of multigraphs. These two concepts are flexible and may have many interpretations. For example, we may consider simplicity as "below a critical level" when we aggregate multiple edges so that simplicity would be of interest if we want to avoid higher levels for some particular reason.

The numbers of loops denoted M_1 and the number of non-loops denoted M_2 are two statistics that are useful in the study of multigraphs. In particular, for aggregated multigraphs, they can be used to analyse edges within and between vertex categories. The expected values of the M_1 and M_2 under IEA are directly obtained as expected values of local multiplicities according to

$$E(M_1) = m \sum_{i=1}^{n} Q_{ii} \ , \tag{7}$$

and

$$E(M_2) = m \sum\sum_{i<j} Q_{ij} \ . \tag{8}$$

We also obtain $E(M_2) = m - E(M_1)$ by using the linear relationship $M_2 = m - M_1$. This linear relationship also implies that

$$\mathrm{Var}(M_2) = \mathrm{Var}(M_1) \ . \tag{9}$$

This common variance is given by

$$\mathrm{Var}(M_1) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(M_{ii}, M_{jj}) = m\left[\sum_{i=1}^{n} Q_{ii}(1-Q_{ii}) - \sum\sum_{i\neq j} Q_{ii}Q_{jj}\right] \ , \tag{10}$$

which in a simpler form can be written as

$$\mathrm{Var}(M_1) = mQ_1Q_2 \ , \tag{11}$$

where $Q_1 = \sum_{i=1}^{n} Q_{ii}$ and $Q_2 = \sum\sum_{i<j} Q_{ii}Q_{jj}$.

A variable that has been used by several authors to study simplicity is $M_1 + M_3$ where M_3 is the number of pairs of equal non-loops. This number is formally given by

$$M_3 = \sum\sum_{i<j} \binom{M_{ij}}{2} \ . \tag{12}$$

The sum $M_1 + M_3$ is a variable that is 0 if and only if the multigraph is simple. The expected value of M_3 is given by

$$E(M_3) = \frac{m(m-1)}{2} \sum\sum_{i<j} Q_{ijij} \ , \tag{13}$$

where Q_{ijij} is the probability of two distinct edges k and ℓ, where $k \neq \ell$, to appear between vertex i and j, where $i < j$.

Consider now the sequence of frequencies of sites with multiplicities $0, 1, \ldots, m$ given by $\mathbf{R} = (R_0, \ldots, R_m)$ where

$$R_k = \sum\sum_{i\leq j} I(M_{ij} = k) \qquad \text{for} \quad k = 0, 1, \ldots, m \ . \tag{14}$$

The distribution of multiplicities that is given by **R** is called the complexity of the graph (Frank and Shafie, 2012). The complexity sequence **R** contains the distribution of multiplicities among the sites. Summary measures of this distribution are of interest as measures of complexity focusing on special properties of the graph. For instance, the proportion of multiple sites or the average multiplicity among multiple sites are simple measures of complexity focusing on any kind of deviation from graphs without multiple edges. These measures are presented in Frank and Shafie (2012) together with a measure that linearly combines the frequencies of different multiplicities given by

$$\sum_{k=2}^{m} \binom{k}{2} R_k \ . \tag{15}$$

This measure counts the number of pairs of edges associated with the same site and if loops are forbidden, it is positive if and only if the graph is not simple. Further, under the IEA model it follows that

$$E(R_k) = \sum\sum_{i \leq j} \binom{m}{k} Q_{ij}^k (1 - Q_{ij})^{m-k} \quad \text{for} \quad k = 0, 1, \ldots, m \ . \tag{16}$$

In order to find the variance of R_k, let s and t be site pairs in $R = \{(i,j) \in V^2 : i \leq j\}$ so that

$$\mathrm{Var}(R_k) = \sum\sum_{s,t \in R} \mathrm{Cov}(I_s, I_t) \ .$$

For $s = t$

$$\mathrm{Cov}(I_s, I_t) = \binom{m}{k} Q_s^k (1 - Q_s)^{m-k} \left(1 - \binom{m}{k} Q_s^k (1 - Q_s)^{m-k}\right) \ ,$$

and for $s \neq t$, we use the fact that (M_s, M_t) is trinomially distributed with parameters $(Q_s,\ Q_t)$ so that

$$\mathrm{Cov}(I_s, I_t) = \binom{m}{k}\binom{m-k}{k} Q_s^k Q_t^k (1 - Q_s - Q_t)^{m-2k} - \left[\binom{m}{k} Q_s^k (1 - Q_s)^{m-k} \binom{m}{k} Q_t^k (1 - Q_t)^{m-k}\right] \ ,$$

where $k \geq 0$ and $2k \leq m$. Thus,

$$\begin{aligned} \mathrm{Var}(R_k) &= \sum\sum_{s,t \in R} \mathrm{Cov}(I_s, I_t) \\ &= \binom{m}{k} \sum_{s \in R} \left[Q_s^k (1 - Q_s)^{m-k} \left(1 - \binom{m}{k} Q_s^k (1 - Q_s)^{m-k}\right)\right] \\ &\quad - \binom{m}{k} \sum\sum_{\substack{s,t \in R \\ s \neq t}} \left[\binom{m-k}{k} Q_s^k Q_t^k (1 - Q_s - Q_t)^{m-2k} - \binom{m}{k} Q_s^k Q_t^k [(1 - Q_s)(1 - Q_t)]^{m-k}\right] \ . \end{aligned} \tag{17}$$

From this we can handle other statistics related to complexity. The number of sites with no occupancy given by R_0 has expected value

$$E(R_0) = \sum\sum_{i \leq j} (1 - Q_{ij})^m \tag{18}$$

and variance

$$\mathrm{Var}(R_0) = \sum_{s \in R} (1 - Q_s)^m Q_s^m - \sum\sum_{\substack{s,t \in R \\ s \neq t}} [(1 - Q_s - Q_t)^m - (1 - Q_s)^m (1 - Q_t)^m] \ , \tag{19}$$

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

$$E(R_1) = m \sum\sum_{i \leq j} Q_{ij}(1 - Q_{ij})^{m-1} \tag{20}$$

and variance

$$\begin{aligned} \mathrm{Var}(R_1) = m \sum_{s \in R} Q_s(1 - Q_s)^{m-1} \left(1 - mQ_s(1 - Q_s)^{m-1}\right) \\ - m \sum\sum_{\substack{s,t \in R \\ s \neq t}} \left[(m-1)Q_sQ_t(1 - Q_s - Q_t)^{m-2} - mQ_sQ_t[(1 - Q_s)(1 - Q_t)]^{m-1}\right] . \end{aligned} \tag{21}$$

Further, the expected value of the number of multiple edges is

$$E(m - R_1) = m \left(1 - \sum\sum_{i \leq j} Q_{ij}(1 - Q_{ij})^{m-1}\right) , \tag{22}$$

and the number of multiple occupancy sites is thus given by

$$E(r - R_0 - R_1) = r - \sum\sum_{i \leq j}(1 - Q_{ij})^m - m \sum\sum_{i \leq j} Q_{ij}(1 - Q_{ij})^{m-1} . \tag{23}$$

Although the presented statistics are derived under the assumption of independence, we apply them to our example network to highlight how they may be used to explore structural dependencies in multigraphs. This is done in th next section where we illustrate how the presented statistics with their expectations and variances can be used to suggest model modifications.

4.3 Example: Florentine Network

We apply the IEA model on the Florentine network and use several derived statistics under IEA to see tendencies in terms of economic, political and social influence on marital and financial relations among the 16 families. In particular, estimated expected outcome ± variability for each of the selected statistics are used to analyse the occurence of marital and financial edges in the aggregated multigraphs from Figure 2.

In order to apply the IEA model, we first estimate the assignment probabilities $\mathbf{Q}$ for the edge multiplicities $\mathbf{M}$. This multiplicity sequence is multinomially distributed and the maximum likelihood estimates of the unknown cell probabilities of a multinomial distribution are simply the empirical fraction of edges that are present in the data set. Thus, the estimated assignments probabilities for the original graphs and aggregated multigraphs are given by $\hat{\mathbf{Q}} = \mathbf{m}/m$ or $\hat{Q}_{ij} = m_{ij}/m$ for $i \leq j \in R$. The observed edge multiplicities $\mathbf{m}$ for each Florentine multigraph is given in Table 4 in Appendix.

Next, we use $\hat{\mathbf{Q}}$ to estimate expected outcomes ($\hat{E}$) and variances ($\hat{V}$) of the statistics. We use these estimates to create approximate 95 % confidence intervals given by $\hat{E} \pm 2\sqrt{\hat{\mathrm{V}}}$. Table 2 gives confidence intervals for the statistics M_1 and M_2, (number of loops and non-loops), and R_k where $k = 0, 1$ (number of null and single occupancy sites), for all multigraphs given in Figure 2. By investigating these confidence intervals across different statistics and multigraphs, we can explore the covariation of ties given different actor attributes.

We start by looking at the top part of Table 2 and the interval estimates for M_1 and M_2. Focusing on multigraphs with two vertices that are aggregated by single actor attributes wealth (W), number of priorates (P) and total number of ties (T), we can interpret tendencies in the edge occurences of each type within and between vertex categories. The expected number of within category marriages (M) is highest (=12) for the multigraph G(M;P) which is aggregated by number of priorates. This indicates a tendency for families to enter marital alliances with other families in the same political category. Reversely, the expectation of financial relations (F) within the same political category is lowest (=5) for this multigraph. The confidence intervals for M_1 for each edge type overlap. With one exception, this is also the case for the M_2 intervals representing the number of edges between vertex categories. This exception is the multigraph aggregated over all three binary actor attributes and is mainly due to no present loops in the (aggregated) data. The fact that the intervals for both M_1 and M_2 are overlapping could be an indicator that there is covariation between the two relation types meaning that some form of dependence process is at play. Thus, the IEA model may not be a good fit to this network.

The number of no edges occurring at vertex pair sites is given by R_0. This number can be interpreted as the tendency for isolated vertices to appear in a multigraph which is of importance in the context of network diffusion. The higher the expectation of this statistic, the higher the tendency for no ties within or between vertex categories. Looking at the interval estimates for R_0 in Table 2, we note highly overlapping intervals for marital and financial relations over all aggregated multigraphs. This does not indicate any clear tendencies in terms of vertex isolation. The number of single occupancy of sites R_1 is however more informative for this purpose. Using M_1 and R_1 allow us to detect multigraphs with isolated vertices. A low expectation of R_1 but high expectation of M_1 indicate single tie occurences within vertex categories (single loops) so that the vertex pair in question may be isolated from other pairs in the multigraph. From Table 2, we note that this occurs for multigraph G(M;P). Thus, the political power of one family does not play a role in entering marriage alliances to other families with different levels of power. This is consistent with our interpretations using M_1.

A similar investigation can be done for R_1 together with M_2 which shows the tendency for single occupancy of edges between vertex categories of the multigraph (i.e. simplicity). If multiple ties are then defined as the propensity of strengthening existing relations (as in valued graphs), a high expectation of these two statistics indicates no such tendency. Among the presented multigraphs in Table 2 (with the same number of vertices), this occurs for G(M;W,T) which means that economic and social influence may not be determining factors for strengthening alliances through marriages.

We conclude this section by a short discussion and extension of our IEA analysis. By calculating R_k for $k = 3, \dots, m$ we can extract more information in terms of how relations covary and further explore structural dependencies. In particular, if interval estimates of R_k for marital and financial ties are overlapping over all k, there may be multiplexity present (see Section 3.1). This would suggest that the presented IEA model based on independency is not an appropriate fit to the data. Similarily, the sum $R_0 + R_1$ can be used in comparison to $R_3 + \cdots + R_k$ to detect the propensity of multiple edges to occur. Another way to test the fit of the model is by running simulations of the IEA model with estimated edge assignment probabilities, and calculate the expectation of statistics representing network properties of interest. Deviation in these values from those observed in the network data would then indicate a poor fit.

	Observed				Under IEA					
					M_1			M_2		
Multigraph (G)	n	m	m_1	m_2	$\hat{E}$	$\hat{V}$	$\hat{E} \pm 2\sqrt{\hat{V}}$	$\hat{E}$	$\hat{V}$	$\hat{E} \pm 2\sqrt{\hat{V}}$
G(M;W)	2	20	7	13	7	4.55	7 ± 4.27	13	4.55	13 ± 4.27
G(F;W)	2	15	7	8	7	3.73	7 ± 3.86	8	3.73	8 ± 3.86
G(M;P)	2	20	12	8	12	4.8	12 ± 4.38	8	4.8	8 ± 4.38
G(F;P)	2	15	5	10	5	3.33	5 ± 3.65	10	3.33	10 ± 3.65
G(M;T)	2	20	9	11	9	4.95	9 ± 4.45	11	4.95	11 ± 4.45
G(F;T)	2	15	7	8	7	3.73	7 ± 3.86	8	3.73	8 ± 3.86
G(M;W,P)	4	20	4	16	4	3.2	4 ± 3.58	16	3.2	16 ± 3.58
G(F;W,P)	4	15	1	14	1	0.93	1 ± 1.93	14	0.93	14 ± 1.93
G(M;W,T)	4	20	3	17	3	2.55	3 ± 3.19	17	2.55	17 ± 3.19
G(F;W,T)	4	15	3	12	3	2.4	3 ± 3.10	12	2.4	12 ± 3.10
G(M;P,T)	4	20	3	17	3	2.55	3 ± 3.19	17	2.55	17 ± 3.19
G(F;P,T)	4	15	1	14	1	0.93	1 ± 1.93	14	0.93	14 ± 1.93
G(M;W,P,T)	8	20	1	14	1	0.95	1 ± 1.95	19	0.95	19 ± 1.95
G(F;W,P,T)	8	15	0	15	0	0	0 ± 0	15	0	15 ± 0

	Observed				Under IEA					
					R_0			R_1		
Multigraph (G)	n	r	r_0	r_1	$\hat{E}$	$\hat{V}$	$\hat{E} \pm 2\sqrt{\hat{V}}$	$\hat{E}$	$\hat{V}$	$\hat{E} \pm 2\sqrt{\hat{V}}$
G(M;W)	2	3	0	0	0.12	0.11	0.12 ± 0.66	0.29	0.22	0.29 ± 0.94
G(F;W)	2	3	1	0	1.00	0.00	1.00 ± 0.02	0.00	0.00	0.00 ± 0.07
G(M;P)	2	3	1	0	1.00	0.00	1.00 ± 0.01	0.00	0.00	0.00 ± 0.04
G(F;P)	2	3	0	1	0.36	0.24	0.36 ± 0.98	0.43	0.29	0.43 ± 1.07
G(M;T)	2	3	0	0	0.04	0.04	0.04 ± 0.39	0.14	0.13	0.14 ± 0.71
G(F;T)	2	3	1	0	1.00	0.00	1.00 ± 0.02	0.00	0.00	0.00 ± 0.07
G(M;W,P)	4	10	4	16	4.88	0.65	4.88 ± 1.62	1.18	0.82	1.18 ± 1.82
G(F;W,P)	4	10	1	14	5.14	0.85	5.14 ± 1.84	1.41	0.96	1.41 ± 1.96
G(M;W,T)	4	10	2	2	3.14	0.95	3.14 ± 1.95	1.78	1.35	1.78 ± 2.33
G(F;W,T)	4	10	5	1	5.44	0.34	5.44 ± 1.16	0.75	0.61	0.75 ± 1.57
G(M;P,T)	4	10	5	0	5.24	0.23	5.24 ± 0.96	0.68	0.61	0.68 ± 1.56
G(F;P,T)	4	10	4	3	5.14	0.85	5.14 ± 1.84	1.41	0.96	1.41 ± 1.96
G(M;W,P,T)	8	36	24	7	27.00	2.65	27.00 ± 3.25	3.74	2.52	3.74 ± 3.17
G(F;W,P,T)	8	36	26	7	28.67	2.35	28.67 ± 3.06	3.20	2.12	3.20 ± 2.91

Table 2: Observed and estimated complexity and simplicity statistics under independent edge assignments of marital (M) and financial (F) edges in the Florentine network for multigraphs aggregated by combinations of binary actor attributes wealth (W), number of priorates (P), and total number of ties (T).

5 Discussion and Future Topics

When graphs are used for modelling of data in social networks and other applications, it is sometimes important to allow multiple edges and/or edge loops and consider models for multigraphs that capture more than just a binary relationship. Multigraphs, however, have not been treated as extensively as simple graphs in the literature. A few situations are presented in this article to illustrate how multigraphs appear naturally in various applications. Further, we discuss how sampling and edge generating mechanisms can provide multiple ties. Other illustrations involve various aggregation procedures similar to those of positional analysis and blocking. Therefore, multigraphs can appear directly in applications but can also be constructed by the analyst if relevant models are available.

A random multigraph model based on the assumption of independence edge assignments (IEA) is presented and measures for analysing the local and global structure of multigraphs under this model are introduced. This simplified model has its limitations since we are ignoring the stochastic dependence between edge assignment which is apparent in many real world networks. The discussion of independence between edges (and also between vertices) must be viewed from both a local and a global perspective.

On one hand, there are local social processes that generate dyadic relations and these processes may depend on the surrounding social structure. These local dependencies must be taken into consideration when modelling and analysing multigraphs, as done with exponential random graph models (ERGM) for univariate networks (Lusher et. al. 2012). In an ERGM, a graph G is represented through summary statistics $z(G)$ and assigns probabilities to graphs according to these statistics given by

$$P_\theta(G) = \kappa^{-1} \exp\left(\sum_i \theta_i z_i(G)\right) \ , \tag{24}$$

where θ_i is the weighting parameter for statistic i and κ is the normalising term ensuring that the sum of the probability mass function over all multigraphs is equal to one. The simplest ERGM is represented by including one statistic representing the number of edges and assuming independence among the edges (i.e. the Bernoulli graph, see Section 2.2). For multivariate networks, the IEA model can be expressed as an exponential random multigraph model (ERMM) so that the probability of a multigraph (represented by its multiplicity sequence) is given by

$$P_\theta(\mathbf{M} = \mathbf{m}) = \kappa^{-1} \exp\left(\sum_{i \leq j}\sum m_{ij}\theta_{ij}\right) \ , \tag{25}$$

where m_{ij} is the statistic with parameter $\theta_{ij} = \log(Q_{ij})$. Two network configurations are relevant to this model: single or multiple edges and loops. However, the expected statistics presented in Section 4.2 could be the core terms of an ERMM allowing us to consider more complex configurations in multigraphs. In particular, these configurations should incorporate dependence effects in multigraphs, e.g. the interlocking of multiple relations (see Section 3.1), and the effect of loops on local tie formations (e.g. dyads, triads or of higher order). We leave this as suggestion for following research.

On the other hand, global dependence that is not induced by local processes seems unrealistic for large networks where the actors may not even be aware or come in contact with each other. This indicates a need to focus on analysing induced subgraphs of small size to support the modelling. For instance, by using cluster analysis and cross classification of dyad distributions (Frank and Nowicki 1993, Frank et al. 1985a, and Frank et al. 1985b), it is possible to find factors that need to be controlled for in the model. Further, induced subgraphs may be analysed as described above; through local social processes and by using ERMMs.

An alternative random multigraph model, where consideration is taken to the stochastic dependence between assignment of edges to sites of vertex pairs, is a special kind of preferential attachment model (Barabàsi & Albert, 1999) which in the literature sometimes is called the configuration model (Janson, 2009; Newman, 2003; Bollobàs, 2001, Bender & Canfield, 1978). Under this model, ties are formed by randomly coupling pairs of stubs (i.e. semi-edges) according to a fixed degree sequence (or stub multiplicity sequence). The advantage of this model is that we are taking into account degree based processes, but that comes with a price: this dependence makes it much more tedious to derive measures of simplicity and complexity, as those presented for the IEA model in Section 4.2 of this paper. There are however methods for obtaining an approximate IEA model from a configuration model. One such method is to ignore the dependence between edges in the configuration model and assume independent edge assignments of stubs. This is done using assignment probabilities in the IEA model that corresponds to edge assignments under the configuration model defined by the fixed degrees. This can also be viewed as repeated assignments with replacements of stubs, whereas the configuration model is repeated assignments without replacement of stubs. When the dependence between edges can be ignored, the IEA model is much easier to analyse. It is therefore of interest to determine when this ignorance can be justified. This can be done by a thorough comparison of the two models using their edge multiplicity and multigraph distributions, respectively. This is also suggested for future research.

In summary, this paper presents the presumed applicability of a multigraph approach in network analysis. However, as with any new approach, applications to real world data are the only way to truly get an idea of its limitations and possibilities. To that end, we encourage researchers across different branches of social science to apply the presented multigraph method and to test the introduced model's goodness-of-fit to real data.

Acknowledgements *The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement n. 319209.*

References

Barabàsi, A. L. & Albert, R., 1999. Emergence of scaling in random networks. Science, 286, 509–512.

Bender, E. A., Canfield, E. R., 1978. The asymptotic number of labeled graphs with given degree sequences. Journal of Combinatorial Theory, Series A 24(3), 296–307.

Boorman, S. A., White, H. C., 1976. Social structure from multiple networks. II. Role structures. American Journal of Sociology, 1384–1446.

Bollobàs, F., 2001. Random Graphs. Cambridge Studies in Advanced Mathematics (No. 73), 2^{nd} ed., Cambridge: Cambridge University Press, New York, NY. press.

Breiger, R. L., Pattison, P. E., 1986. Cumulated social roles: The duality of persons and their algebras. Social Networks, 8(3), 215–256.

Butts, C. T., 2008. Social network analysis: a methodological introduction. Asian Journal of Social Psychology 11(1), 13–41.

Carrington, P., Scott, J., Wasserman, S. (Eds.), 2005. Models and Methods in Social Network Analysis, Cambridge University Press, New York, NY. Press.

Doreian, P., 1980. On the evolution of group and network structure. Social Networks, 2(3), 235–252.

Doreian, P., 1986. Measuring relative standing in small groups and bounded social networks. Social Psychology Quarterly, 247–259.

Faust, K., Wasserman, S., 1992. Blockmodels: interpretation and evaluation. Social Networks 14(1), 5–61.

Frank, O., 2009. Estimation and sampling in social network analysis. In: R. Meyers (Ed.), Encyclopedia of Complexity and Systems Science, Springer Verlag, New York, pp. 8213–8231.

Frank, O., 2011. Survey Sampling in Networks. In: P. J. Carrington, & , J. Scott (Eds.), The SAGE Handbook of Social Network Analysis, pp. 389–403, Sage.

Frank, O., Nowicki, K., 1993. Exploratory statistical analysis of networks. Annals of Discrete Mathematics 55, 349– 366.

Frank, O., Shafie, T., 2012. Complexity of families of multigraphs, JSM Proceedings, Section on Statistical Graphics, Alexandria, VA: American Statistical Association, 2908–2921.

Frank, O., Hallinan, M., Nowicki, K., 1985a. Clustering of dyad distributions as a tool in network modeling. Journal of Mathematical Sociology 11(1), 47–64.

Frank, O., Komanska, H., Widaman, K. F., 1985b. Cluster analysis of dyad distributions in networks. Journal of Classification 2(1), 219–238.

Guttman, L., 1977. A definition of dimensionality and distance for graphs. In: Lingoes, J.C (Ed.), Geometric Representation of Relational Data. Ann Arbor, MI: Mathesis.

Janson, S., 2009. The probability that a random multigraph is simple. Combinatorics, Probability and Computing 18(1–2), 205–225.

Kent, D., 1978. The Rise of The Medici: Faction in Florence, 1426-1434. Oxford: Oxford University Press.

Koehly, L. M., Pattison, P., 2005. Random graph models for social networks: Multiple relations or multiple raters. In: P. Carrington, J. Scott, S. Wasserman (Eds.), Models and Methods in Social Network Analysis, Cambridge University Press, New York, NY, pp. 162–191.

Kolaczyk, E., 2009. Statistical Analysis of Network Data. Springer Verlag, New York.

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

Lorrain, F., White, H. C., 1971. Structural equivalence of individuals in social networks. The Journal of mathematical sociology, 1(1), 49–80.

Lusher, D., Koskinen, J., Robins, G. (Eds.), 2012. Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications. Cambridge University Press.

Nowicki, K., Snijders, T. A. B., 2001. Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association, 96(455), 1077–1087.

Padgett, J. F., 1987. Social mobility in hieratic control systems. In: R.L. Breiger (Ed.), Social Mobility and Social Structure. New York: Cambridge University Press.

Padgett, J. F., Ansell, C. K., 1989. From faction to party in Renaissance Florence. Department of Political Science, University of Chicago, 33.

Padgett, J. F., Ansell, C. K., 1993. Robust Action and the Rise of the Medici, 1400-1434. American Journal of Sociology, 1259–1319.

Pattison, P., 1993. Algebraic Models for Social Networks. Cambridge University Press.

Pattison, P., Wasserman, S., 1999. Logit models and logistic regressions for social networks: II. Multivariate relations. British Journal of Mathematical and Statistical Psychology, 52(2), 169–193.

Pattison, P., Wasserman, S., Robins, G., Kanfer, A. M., 2000. Statistical evaluation of algebraic constraints for social networks. Journal of Mathematical Psychology, 44(4), 536–568.

Ranola, J. M., Ahn, S., Sehl, M., Smith, D. J., Lange, K., 2010. A Poisson model for random multigraphs. Fioinformatics, 26(16), 2004–2011.

Robins, G., Pattison, P., 2006. Multiple networks in organisations. *Draft report.*

Robins, G., 2013. A tutorial on methods for the modeling and analysis of social network data. Journal of Mathematical Psychology, 57, 261–274.

Scott, J., Carrington, P. (Eds.), 2011. Handbook of Social Network Analysis. Sage Publications, London.

Snijders, T. A. B., 2011. Statistical models for social networks. Annual Review of Sociology 37, 131–153.

Snijders, T. A., Nowicki, K., 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of Classification, 14(1), 75–100.

Wasserman, S., 1987. Conformity of two sociometric relations. Psychometrika, 52(1), 3–18.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

White, H., 1963. An Anatomy of Kinship: Mathematical Models for Structures of Cumulated Roles. Englewood Cliffs, NJ: Prentice-Hall.

White, H. C., Boorman, S. A., Breiger, R. L., 1976. Social structure from multiple networks. I. Blockmodels of roles and positions. American Journal of Sociology, 730–780.

Appendix

Table 3: The original and binary attribute variables of the 16 Florentine families. The attributes are W=wealth, P=number of priorates and T=total number of ties.

	W		P		T	
ACCIAIUOL	10	0	53	1	2	0
ALBIZZI	36	0	65	1	3	0
BARBADORI	55	1	0	0	14	1
BISCHERI	44	1	12	1	9	0
CASTELLAN	20	0	22	1	18	1
GINORI	32	0	0	0	9	0
GUADAGNI	8	0	21	1	14	1
LAMBERTES	42	1	0	0	14	1
MEDICI	103	1	53	1	54	1
PAZZI	48	1	0	0	7	0
PERUZZI	49	1	42	1	32	1
PUCCI	3	0	0	0	1	0
RIDOLFI	27	0	38	1	4	0
SALVIATI	10	0	35	1	5	0
STROZZI	146	1	74	1	29	1
TORNABUON	48	1	0	0	7	0

Table 4: The edge multiplicities **m** presented as upper triangular matrices for all different graphs and multigraphs (G), with marital (M) and financial (F) relations, and aggregated by attributes W=wealth, P=number of priorates and T=total number of ties.

G	**m**
G(M)	$\begin{bmatrix} 0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&1&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$
G(F)	$\begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$
G(M;W,P,T)	$\begin{bmatrix} 0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&2&0&0&1&5 \\ 0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&0&1&2 \\ 0&0&0&0&0&0&2&1 \\ 0&0&0&0&0&0&0&2 \\ 0&0&0&0&0&0&0&1 \end{bmatrix}$
G(F;W,P,T)	$\begin{bmatrix} 0&0&0&0&0&1&0&1 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0&2 \\ 0&0&0&0&0&1&1&1 \\ 0&0&0&0&0&0&3&3 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0&0 \end{bmatrix}$
G(M;W,P)	$\begin{bmatrix} 0&1&0&0 \\ 0&1&5&8 \\ 0&0&0&2 \\ 0&0&0&3 \end{bmatrix}$
G(F;W,P)	$\begin{bmatrix} 0&0&1&1 \\ 0&0&3&3 \\ 0&0&0&6 \\ 0&0&0&1 \end{bmatrix}$
G(M;W,T)	$\begin{bmatrix} 1&1&2&5 \\ 0&0&2&4 \\ 0&0&0&3 \\ 0&0&0&2 \end{bmatrix}$
G(B;W,T)	$\begin{bmatrix} 0&0&0&3 \\ 0&0&1&4 \\ 0&0&0&4 \\ 0&0&0&3 \end{bmatrix}$
G(M;P,T)	$\begin{bmatrix} 0&0&3&2 \\ 0&0&0&3 \\ 0&0&0&9 \\ 0&0&0&3 \end{bmatrix}$
G(F;P,T)	$\begin{bmatrix} 0&1&0&3 \\ 0&0&1&6 \\ 0&0&0&3 \\ 0&0&0&1 \end{bmatrix}$
G(M;W)	$\begin{bmatrix} 5&13 \\ 0&2 \end{bmatrix}$
G(F;W)	$\begin{bmatrix} 7&8 \\ 0&0 \end{bmatrix}$
G(M;P)	$\begin{bmatrix} 12&8 \\ 0&0 \end{bmatrix}$
G(F;P)	$\begin{bmatrix} 4&10 \\ 0&1 \end{bmatrix}$
G(M;T)	$\begin{bmatrix} 6&11 \\ 0&3 \end{bmatrix}$
G(F;T)	$\begin{bmatrix} 7&8 \\ 0&0 \end{bmatrix}$