NETWORK ANOVA RANDOM EFFECTS MODELS FOR NODE ATTRIBUTES

GABRIEL MONTE-ROJAS*
Instituto Interdisciplinario de Economía Política de Buenos Aires
(IIEP-BAIRES-UBA)-CONICET
Facultad de Ciencias Económicas, Universidad de Buenos Aires
Av. Córdoba 2122 2do piso, C1120AAQ, Ciudad Autónoma de Buenos Aires, Argentina

PEDRO ELOSEGUI
Banco Central de la República Argentina and Universidad Nacional de La Plata

Abstract. This paper develops a subgraph network random effects error components structure for network data to perform analysis of variance. In particular, it proposes a model for evaluating the network interdependence of nodes attributes allowing for edge and triangle specific components. The latter serve as a basal model for modeling more general network effects. Consistent estimators of the variance components and Lagrange Multiplier specification tests for evaluating the appropriate model of random components in networks structures is proposed. Monte Carlo simulations show that the tests have good performance in finite samples. The proposed tests is applied to the unsecured (Call) interbank market network in Argentina.

1. Introduction. Network effects, where the interaction among agents is considered relevant, is a topic of interest in the statistical literature. In network structured data, the information on the direct or indirect contact among the involved parties actually matters. This fact has attracted considerable attention, especially regarding possible spillover effects in diverse areas, such as education, production, financial and banking markets, trade and many others. See [12], [13] and [15] for a recent literature reviews. The kind of applications cover both the statistical and game theoretical analysis, see e.g. [14] and [18] for recent works.

In a given network structure, observations are not independent and the relationships among them depend on their relative network position. The network structure differs from the spatial statistical analysis, as the latter is naturally embedded into some metric space (i.e., natural geometry). It is also different from panel data models because there is no prior group structure to cluster observations.

2010 Mathematics Subject Classification. Primary: 05Cxx; Secondary: 68R10.
Key words and phrases. Networks, Nodes, Links, Triangles, Clusters.

We are grateful to the Guest Editors, Viktoriya Semeshenko, Gabriel Brida and Andrea Roverini, and to two anonymous referees for helpful comments and suggestions. Matías Pardini and Emilio Sáenz Guillén provided invaluable research assistance. We also thank the Gerencia Principal de Estadísticas Económicas of Banco Central de la República Argentina for the data used in the empirical application.

* Corresponding author: Gabriel Montes-Rojas.
The most obvious type of intra-network correlation arises when we consider observations given by vertices or nodes that have a common edge or link. If we consider a link-specific effect, this would result in a specific factor that arises only for those linked nodes and not for others. Nodes that share a link might be more correlated with each other than with others with no links. This simple idea can be extended to any subgraph structure, such as triangles, diamonds, cliques, stars, etc. That is, nodes that are within a common subgraph share more in common that nodes that are not. These common features may be subgraph specific, thus allowing for a given structure of nodes correlation depending whether they belong to the same subgraph or not.

This paper is concerned with statistically evaluating a real valued node attribute within a network structure. In particular, the variance and covariance structure, thus using analysis of variance (ANOVA). We define a random components structure where the components depend on local network features of the observations. In particular, for a given graph we construct the error components model by considering node-, link- and triangle-specific effects. The main purpose of this exercise is that the empirical researcher starts from a standard variance-covariance structure (i.e., independent error components), and then tests sequentially for potential components’ patterns (e.g., edges, triangles, diamonds, cliques, stars, etc.). The result is that the observed node attribute can be decomposed into subgraph specific components that adds to the total variability in the attribute. The developed error components model has interesting features. Contrary to the standard error components models, network effects will typically imply heteroskedasticity. Take for instance the vertex\$edge-only error components model where each vertex will have a vertex specific random component and an edge specific random component. Vertices that have one link are different from those that have two or more. The edge specific component will in fact generate a higher variance for vertices with more links. The ANOVA structure might have positive variance at some subgraph network structures but not others. Thus, identifying which random components are active and which are not is an important for correct specification.

This paper is organized as follows. Section 2 develops the subgraph network random effects error components model. Section 3 presents simple consistent variance-covariance components estimators. Section 4 constructs specification tests and Section 5 evaluates the finite sample performance using Monte Carlo experiments. Section 6 applies the proposed tests to the interbank market in Argentina. Section 7 concludes.

2. Network error components model.

2.1. Network definition and notation. Consider a graph $G = (V, L)$ as a mathematical structure consisting of a set $V$ of vertices (also commonly called nodes) and a set $L$ of edges (also commonly called links). Unless otherwise specified the graph is undirected where elements of $L$ are unordered pairs $(i, s)$ of distinct vertices $(i, s) \in V \times V$. In a directed graph the elements of $L$ are ordered pairs $(i, s) \in V \times V$. The number of vertices is $N = |V|$ and the number of edges is $M = |L|$. Without loss of generality, we will label the vertices simply with the integers $1, \ldots, N$, and the edges, $1, \ldots, M$. Note that $M \leq N(N - 1)/2$ for undirected graphs (and $M \leq N(N - 1)$ for directed ones). There are $2^{\binom{N}{2}}$ potential undirected networks.

For our purposes consider a set of triangles in undirected graphs as $\text{Triangles} = \{(i, s, r) \in V^3, i < s < r, (i, s), (s, r), (i, r) \in L\}$, the number of triangles is $T \leq \binom{N}{3}$. The edges, $1, \ldots, M$. Note that $M \leq N(N - 1)/2$ for undirected graphs (and $M \leq N(N - 1)$ for directed ones). There are $2^{\binom{N}{2}}$ potential undirected networks.

For our purposes consider a set of triangles in undirected graphs as $\text{Triangles} = \{(i, s, r) \in V^3, i < s < r, (i, s), (s, r), (i, r) \in L\}$, the number of triangles is $T \leq \binom{N}{3}$. The edges, $1, \ldots, M$. Note that $M \leq N(N - 1)/2$ for undirected graphs (and $M \leq N(N - 1)$ for directed ones). There are $2^{\binom{N}{2}}$ potential undirected networks.
\(N(N - 1)(N - 2)/6\). The set of triangles could be defined differently for directed graphs.

The fundamental connectivity of a graph \(G\) may be captured in an \(N \times N\) binary adjacency matrix \(A\) with entries

\[a_{is} = \begin{cases} 1 & \text{if vertices } \{i,s\} \in L \\ 0 & \text{otherwise} \end{cases},\]

the edge-incidence matrix \(B\), an \(N \times M\) binary matrix with entries

\[b_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ is incident to edge } j \\ 0 & \text{otherwise} \end{cases},\]

and the triangle incidence matrix \(C\), an \(N \times T\) binary matrix with entries

\[c_{ik} = \begin{cases} 1 & \text{if vertex } i \text{ is incident to triad } k \\ 0 & \text{otherwise} \end{cases}.

### 2.2. Subgraph random effects in the undirected graph model.

Consider the following assumption on the probability space.

**Assumption 1.**

- Let \(G \in \mathcal{G}_N\) be a space of graphs of size \(N\), \(\sigma(\mathcal{G}_N)\) a \(\sigma\)-algebra in the sample space \(\mathcal{G}_N\), and \(\mathcal{P}_N\) a probability space on the measurable space on \(\mathcal{G}_N, \sigma(\mathcal{G}_N)\). Then \([\mathcal{G}_N, \sigma(\mathcal{G}_N), \mathcal{P}_N]\) form a probability space.

This paper is concerned with the additive random network structure of the following form.

**Assumption 2.**

- Assume that

\[\varepsilon_i := \nu_i + \sum_{j=1}^{M} b_{ij}\mu_j + \sum_{k=1}^{T} c_{ik}\delta_k, \quad (1)\]

where \(\{b_{ij}\} \) and \(c_{ik}\) are the row vectors of matrices \(B\) and \(C\) derived from graph \(G\), and \(\nu, \mu\) and \(\delta\) are mutually independent random variables that satisfy:

(i) \(\forall i,j,t\) \(E(\nu_i \mid G) = E(\mu_{ij} \mid G) = E(\delta_{ijt} \mid G) = 0\),

(ii) \(\forall i,j,t\) \(\text{Var}(\nu_i \mid G) = \sigma^2_\nu\), \(\text{Var}(\mu_{ij} \mid G) = \sigma^2_\mu\), \(\text{Var}(\delta_{ijt} \mid G) = \sigma^2_\delta\).

The random variable \(\varepsilon_i\) is the node attribute of interest. This paper evaluates whether this attribute is related to the network structure and quantifies its components, namely a node component, a link component and a triangle component.

The model can also be written as

\[\varepsilon_i = \nu_i + \sum_{i=1}^{N} \sum_{s>r} a_{is}\mu(s) + \sum_{i=1}^{N} \sum_{s>r} \sum_{r>s} a_{is}a_{sr}a_{ir}\delta(sr), \quad (2)\]

where \(\mu(s)\) and \(\delta(sr)\) correspond to the common edge and triangle effects, respectively.

In matrix notation the model above can be written as an \(N \times 1\) vector such as

\[\varepsilon = \nu + B\mu + C\delta, \quad (3)\]
and
\[ \Omega := E[\varepsilon \varepsilon^\top | G] = E[\nu \nu^\top + B \mu \mu^\top + C \delta \delta^\top | G] \]
\[ = \sigma_\nu^2 I_N + \sigma_\mu^2 BB^\top + \sigma_\delta^2 CC^\top, \]
where \( \nu \) is a \( N \times 1 \) random vector, \( \mu \) is a \( M \times 1 \) random vector, \( \delta \) is a \( T \times 1 \) random vector.

In the case with no network effects, defined as the vertex-only model, \( \Omega_v = \sigma_\nu^2 I_N \).

In this case, the network structure has no effect on the node attributes and each realization is the result of an iid random variable.

The random-effects vertex\&edge incidence model would be \( \Omega_{ve} = \sigma_\nu^2 I_N + \sigma_\mu^2 BB^\top \).

In this case, the node attributes are affected by the presence of common links. This produces an heteroskedastic effect because common links would be associated with a separate random component. The relative magnitudes of \( \sigma_\nu^2 \) and \( \sigma_\mu^2 \) would help in evaluating the importance of links within the network.

The random-effects vertex\&triangle incidence model would have \( \Omega_{vt} = \sigma_\nu^2 I_N + \sigma_\delta^2 CC^\top \).

Joining both models gives a general model with vertex, edge and triangle random components in (4).

The results above can be easily extended to weighted networks where \( A \) is replaced by \( W \), and the \( B \) and \( C \) matrices are also constructed using the weighted components.

2.3. Comparison with spatial models. Consider now a simple spatially autocorrelated models, where the matrix \( A \) is related to distance, and for this case \( A \) is a symmetric contiguity matrix. See [3] for a discussion and a comparison of different models.

The simplest statistic for spatial dependence is the so-called Moran’s I statistic, defined as \( m = \varepsilon^\top A \varepsilon \). While this provides an intuitive way of quantifying spatial dependence it provides not guidance on how to model the random components. Another model of spatial autocorrelation is captured by a simple autocorrelation parameter \( \rho \) such that
\[ \varepsilon = \nu + \rho A \nu, \quad \nu \sim iid(0, \sigma_\nu^2). \]

In this model the node attribute \( \varepsilon \) is composed of an idiosyncratic component, \( \nu \), which may have an effect on the linked nodes through the adjacency matrix \( A \). Note that this produces the following structure in the variance-covariance matrix:
\[ \Omega_{exp1} = \sigma_\nu^2 (I_N + \rho A) (I_N + \rho A)^\top = \sigma_\nu^2 (I_N + 2 \rho A + \rho^2 A^2). \]

Although this would capture some of the features developed above, it imposes restrictions on the parameters, i.e., \( \rho \) and \( \rho^2 \) (see, e.g., [16, 17]).

The added value of the paper can be seen in comparison to this spatial model. These models require a pre-specification of the adjacency matrix, which may be subject to misspecification. Once this is given there is a fixed relation among the variance-covariance components that multiply the powers of \( A \), i.e., \( \rho \). The potential gains from using a subgraph network error component model rather than a spatial model would depend on the flexibility of adding additional components in \( \varepsilon \) instead of a correlation structure of a single component \( \nu \).

There is also a conceptual difference. Suppose a shock or perturbation in a given node. The spatial model would then evaluate how this shock propagates across the
network depending on A, which in turn relates to the pre-specified distance. On the contrary, the proposed subgraph network model would consider that a given shock may be the result of a perturbation in a particular subgraph structure (i.e., node, link, triangle). The propagation may occur only if there are common substructures among nodes. Furthermore, the shock effect may be of different magnitude depending on the type of subgraph effects that will appear.

3. ANOVA consistent variance components estimators. Here we consider simple consistent estimators of the variance components using ANOVA-type decompositions.

Consider the following statistics: \( S_1 = \frac{1}{N} \sum_{i=1}^{N} u_i^2 \), \( S_2 = \frac{1}{M} \sum_{i=1}^{N} \sum_{s \geq i} a_i s a_{ir} a_{is} u_i u_s \), \( S_3 = \frac{1}{N} \sum_{i=1}^{N-2} \sum_{r > i}^{N} \sum_{s > r}^{N} a_{ir} a_{is} u_i u_s \). \( S_1 \) contains the usual sum of squared errors.

Note that for each vertex there will be at most \( N - 1 \) edges to which it belongs and \( N - 2 \) triangles. Moreover, each edge will be repeated twice for undirected graphs, one for each vertex, and each triangle will be repeated three times, one for each vertex. Then, \( E[S_1 \mid G] = \sigma^2 \), \( E[S_2 \mid G] = \sigma^2 + \frac{2M}{N} \), and \( E[S_3 \mid G] = \sigma^2 + \frac{2M}{N} \). \( E[S_1 \mid G] \) is the (conditional) covariance of a vertex. \( S_2 \) contains the (conditional) covariance of two vertices that have a common edge.

Finally, \( S_3 \) computes the cross products for active triangles (i.e., \( a_{is} = a_{sr} = a_{ir} = 1, r > s > i \)). Note that for \( S_3 \) if we have a triangle, say \((i,s,r)\), then two nodes, say \(i\) and \(s\), must share both \(\mu_{(i,s)}\) and \(\delta_{(i,s,r)}\). Then, \( E[S_3 \mid G] = \sigma^2 + \sigma^2 \). \( E[S_3 \mid G] \) is the (conditional) covariance of two vertices that have common edge and triangle(s).

In the absence of triangle effects, i.e., \( \sigma^2 = 0 \), the model simplifies to \( \sigma^2 = E[S_1 \mid G] - E[S_2 \mid G] \frac{2M}{N} \). \( \sigma^2 = E[S_2 \mid G] \), such that the non-negativity restrictions are \( E[S_2 \mid G] \geq 0 \) and \( \frac{E[S_2 \mid G]}{E[S_1 \mid G]} \geq \frac{2M}{N} \), such that the ratio of the variance of a vertex to the covariance of two random vertices needs to be bigger than the average number of edges per vertex. First, take for instance a cycle graph, a 2-regular graph with all vertices of degree 2 such that \( M = N \). For this case the variance of the vertices need to be at least twice the covariance. Second, consider a complete graph with \( M = N(N - 1)/2 \). In this case, the ratio of variance to covariance needs to grow faster than the number of vertices.

In the absence of edge effects, \( \sigma^2 = 0 \), the model simplifies to \( \sigma^2 = E[S_1 \mid G] - E[S_3 \mid G] \frac{2M}{N} \). \( \sigma^2 = E[S_3 \mid G] \), such that the non-negativity restrictions are \( E[S_3 \mid G] \geq 0 \) and \( \frac{E[S_3 \mid G]}{E[S_1 \mid G]} \geq \frac{3T}{N} \), such that the ratio of the variance of a vertex to the covariance of two random vertices needs to be bigger than the average number of triangles per vertex.

For the edge and triangle effects model, solving for \( (\sigma^2, \sigma^2, \sigma^2) \) gets \( \sigma^2 = E[S_1 - E[S_2 \mid G] \frac{2M}{N} \) \( \sigma^2 = E[S_3 \mid G] \), \( \sigma^2 = \frac{2M}{N} \), and \( \sigma^2 = \frac{3T}{N} \).
For this case the non-negativity restrictions imply: (i) $E[S_1|G] \geq 1$, (ii) $E[S_2|G] \geq \frac{3\theta}{M}$, (iii) $E[S_3|G] \geq \frac{2\theta}{M}$, and (iv) $E[S_4|G] \geq \frac{3\theta}{M}$. Restriction (i) implies that the covariance among vertices that belong to a triangle must be larger than the covariance of vertices that share a link. Restriction (ii) states that the ratio $E[S_2|G]$ cannot exceed the average number of triangles per edge. Restriction (iii) correspond to the number average number of links per vertex. Restriction (iv) is a combination of the above with no clear interpretation.

4. Specification tests for undirected unweighted graphs. We consider Lagrange Multiplier (LM) tests for testing the presence of different network random components structures assuming that the error components are Gaussian. Consistent estimators of $\theta$ under the null can be obtained using an ANOVA-type analysis as in Section 3. Hence our tests will be based on Neyman’s $C(\alpha)$ principle, which produces tests that are asymptotically equivalent to likelihood based LM tests under $\sqrt{N}$-consistent non-maximum likelihood estimation of the nuisance parameters. See [7] for a discussion.

Consider a partition of $\theta = (\theta_1^T, \theta_2^T)^T$, where $\theta_2$ contains the parameters under the corresponding null hypothesis $H_0^2$: $\theta_2 = 0$, and $\theta_1$ the nuisance parameters that need to be estimated. In our particular case, $\theta$ will be partitioned into either $\theta_1 = \sigma^2_\nu, \theta_2 = \sigma^2_\mu$ when we want to test for the presence of edge network effects assuming $\sigma^2_\nu = 0, \theta_1 = \sigma^2_\nu, \theta_2 = \sigma^2_\mu$ when we want to test for the presence of edge network effects assuming $\sigma^2_\nu = 0, \theta_1 = \sigma^2_\nu, \theta_2 = (\sigma^2_\mu, \sigma^2_\delta)$ when we want to test for the presence of edge and triangle network effects, $\theta_1 = (\sigma^2_\mu, \sigma^2_\delta), \theta_2 = \sigma^2_\delta$ when we want to test for the presence of triangle effects assuming edge effects or $\theta_1 = (\sigma^2_\mu, \sigma^2_\delta), \theta_2 = \sigma^2_\mu$ when we want to test for the presence of triangle effects assuming edge effects.

Consider now [10] locally size-robust type statistics (BY test hereafter). For this, consider a new partition of $\theta = (\theta_1, \theta_2, \theta_3)^T$ where we want to test for the null hypothesis $H_0^1$, we consider $\theta_1$ as nuisance parameters to be estimated, but the validity of the test is affected by the validity of $H_0^3$: $\theta_3 = 0$. Global valid tests for $H_0^1$ would require consistent estimators of $\theta_3$ as in the construction of the conditional LM statistics above. In practice, however, estimators of $\theta_3$ may be cumbersome. Moreover, they may suffer identification conditions under the null. Thus, [10] has been successfully implemented to test one particular null without estimating the other nuisance parameter $\theta_3$. This procedure is valid under $\sqrt{N}$-local deviations of $H_0^3$ and different empirical studies confirmed its validity for non-local deviations. In our particular case, the parameter will be partitioned as $\theta_1 = \sigma^2_\nu, \theta_2 = \sigma^2_\mu, \theta_3 = \sigma^2_\delta$. This procedure allows us to test for triangle effects without estimating the edge effects variance component, even when we are estimating under the joint null hypothesis $H_0^2 \& H_0^3$: $\sigma^2_\mu = \sigma^2_\delta = 0$, which is just least-squares estimation. The statistic is constructed as in [8, 9] for non-maximum likelihood estimation. Then, $LM_{2(3)-1}(\hat{\theta}) \overset{d}{\rightarrow} \chi^2_{\text{dim}(\theta_2)}$, for $\hat{\theta}$ being a consistent estimator under the joint null hypothesis $H_0^2 \& H_0^3$: $\sigma^2_\mu = \sigma^2_\delta = 0$ and for $\theta_3 = o(1/\sqrt{N})$.

The test considered are:

- $LM_{\mu}$: LM test for $H_0 : \sigma^2_\mu = 0$ when $\sigma^2_\nu$ is estimated as mean squared error (MSE) after OLS estimation and $\sigma^2_\delta = 0$ is assumed.
- $LM_\delta$: LM test for $H_0 : \sigma^2_\delta = 0$ when $\sigma^2_\mu$ is estimated as MSE after OLS estimation and $\sigma^2_\nu = 0$ is assumed.
• \( LM_{\mu, \delta} \): LM test for \( H_0 : \sigma^2_{\mu} = \sigma^2_{\delta} = 0 \) when \( \sigma^2_{\mu} \) is estimated as MSE after OLS estimation.
• \( LM_{\mu(\delta)} \): BY test for \( H_0 : \sigma^2_{\mu} = 0 \) when \( \sigma^2_{\mu} \) is estimated as MSE after OLS estimation and \( \sigma^2_{\delta} = 0 \) is allowed to have local deviations.
• \( LM_{\delta(\mu)} \): BY test for \( H_0 : \sigma^2_{\delta} = 0 \) when \( \sigma^2_{\delta} \) is estimated as MSE after OLS estimation and \( \sigma^2_{\mu} = 0 \) is allowed to have local deviations.
• \( LM_{\delta-\mu} \): LM test for \( H_0 : \sigma^2_{\delta} = 0 \) when \((\sigma^2_{\mu}, \sigma^2_{\delta})\) is estimated as in Section 3 after OLS estimation.

5. Monte Carlo experiments. Here we explore the small sample performance of the proposed tests through a Monte Carlo experiment. We will consider the following model:

\[
\varepsilon_i = \nu_i + \sum_{i=1}^{N} \sum_{s>i}^{N} a_{is} \mu_{(is)} + \sum_{i=1}^{N} \sum_{r>s}^{N} a_{ir} a_{ir} \delta_{(isr)},
\]

\[i = 1, 2, \ldots, N,\]

where \( A = \{a_{ir}\} \) is an adjacent contiguity matrix. We assume \( \nu_i \sim iid \ N(0,10), \mu_{(is)} \sim iid \ N(0, \sigma^2_{\mu}) \) and \( \delta_{(isr)} \sim iid \ N(0, \sigma^2_{\delta}) \).

We consider \( N \in \{100, 225, 400\} \) and simulate two types of networks. First, we consider an Erdős-Rényi random graph where links are randomly generated with a given probability \( p_N \), i.e., \( Prob(a_{ir} = 1) = p_N, i, r = 1, \ldots, N, i \neq r \). For the Erdős-Rényi graphs we have on average a constant proportion of vertices and edges, \( N/M \), using \( p_{100} = 0.05, p_{225} = 0.05 \times 100/225, p_{400} = 0.05 \times 100/400 \). In this case, the number of triangles per node is also constant on average. Second, a queen-type spatial structure where edges are generated according to queen contiguity, i.e., for a squared board with number of rows and columns \( n = \sqrt{N} \), for \( i = 1, \ldots, N, a_{ir} = 1 \) if \( r \in \{i-1, i+1, i-n-1, i-n, i-n+1, i+n-1, i+n, i+n+1 \} \) with \( 1 \leq r \leq N \), and \( a_{ir} = 0 \) otherwise. Note that the considered spatial-type model has a similar number of triangles and edges for each node, i.e., 8 edges and triangles for a node that is not on the border of the board.

First, we consider the empirical size results where we impose the absence of both edge and triangle random effects, \( \sigma^2_{\mu} = \sigma^2_{\delta} = 0 \) in Table 7. In all cases, marginal, joint and robust tests have the appropriate size, for all levels of significance.

Second, we consider the empirical power and size-robustness for \((\sigma^2_{\mu}, \sigma^2_{\delta}) \in \{0, 1, \ldots, 10\}^2 \) in Figures 1 and 4. In each case the (a) figure report the rejection rates as we increase the value of \( \sigma^2_{\mu} \), and the (b) the rejection rates for different values of \( \sigma^2_{\delta} \).

Figures 1 and 2 report the tests for detecting edge heterogeneity, that is, \( \sigma^2_{\mu} > 0 \). Note that the marginal tests \( LM_{\mu} \) has the largest power performance for changes in \( \sigma^2_{\mu} \) (see top figures (a)), followed by the joint tests \( LM_{\mu, \delta} \). However, the marginal test also rejects in the direction of \( \sigma^2_{\delta} > 0 \), as the bottom (b) figures show. Indeed, it is not robust to the presence of triangle effects. In order to detect departures from \( \sigma^2_{\mu} = 0 \) the BY tests are constructed, robust to the presence of \( \sigma^2_{\delta} > 0 \), without estimating \( \sigma^2_{\delta} \). The BY robust test have good power performance as can be seen in Figures 1-(a), in fact, close to the joint test, but it has low power in the Queen spatial more complex network model, as shown in Figure 2-(a). Nevertheless, the BY test is robust to deviations in \( \sigma^2_{\delta} > 0 \), as seen in Figures 1-(b) and 2-(b).
Tests for triangle effects have a similar performance to those of edge effects. As in the previous paragraph, the tests have the expected rejection rates in the direction of $\sigma_2^2 \sigma_0 > 0$, and the BY robust test have correct size for $\sigma_2 > 0$. Note that the conditional test $LM_{\delta-\mu}$ estimates $\sigma_2^2$, and as such it should be robust to misspecification in edge effects. For this case the BY robust tests outperforms it in terms of size and power in the Erdős-Rényi random graph model, and it is very close to the conditional tests in the Queen spatial structure.

6. Empirical application. Network analysis of the interbank market is a useful tool for policymakers as it can help to prevent financial contagion and problems related to ‘too-big-to-fail’ financial institutions. In fact, empirical networks have been used for stress test exercises, see [24] for a review. Despite been developed to assess centrality in other contexts, when adapted to the context of financial networks, network centrality measures can guide bank regulation authorities in their assessment of the systemic importance of financial and non-financial institutions. In fact, in the financial economic literature network analysis has mostly been applied to payment systems, interbank markets, and more recently extended to capture the mutual exposure of financial institutions to other asset classes, including derivatives contracts, in a multilayer networks framework [19, 4, 20, 21].

Also, network positioning may affect banks’ liquidity management and interest rates by different mechanisms. First, in line with [1], dense interconnections serve as a mechanism for the propagation of shocks, leading to a more fragile financial system. In that sense, banks that are more connected may be perceived by the market as more fragile. However, the same banks can be perceived as ‘too-interconnected-to-fail’ such that rather than fragile, those banks are perceived as more likely to be bailout, see [5]. This is similar to the too-big-to-fail effect observed in other interbank markets. Second, as argued by [11], financial institutions with more extensive and strategic financial networks, can more efficiently acquire and process information due to their better access to order flows, see [22]. Third, banks with higher centrality within the network have better access to liquidity and are able to charge larger intermediation spreads. Previous empirical evidence [2, 6, 23] suggests that being systemically more important, in term of size or connectedness, can explain part of the cross-sectional variation in banks’ borrowing costs before and during the global financial crisis.

We apply this analysis to the interbank overnight unsecured Call market in Argentina. In this market, banks directly connect each other (“call”) to lend or borrow excess reserves and to satisfy Central Bank minimum liquidity requirements. While the participating number of banks in the market is small (approx. 50 banks), Argentina is known for recurrent financial and liquidity shocks, partial and/or full, private or public, defaults on different contracts, all of which affects liquidity management of the financial sector.

We use daily data for the period 1st January 2015 to 31st December 2018. While the Call market has contracts ranging from 1 up to more than 30 days long, we concentrate in overnight contracts keeping the sample of transactions on 1 to 4 days range only. Networks are constructed on a monthly basis to capture the rich strategic relationships that can be formed among bank pairs. Thus we analyze 48 cross-section network structures (one for each month).
Consider a network $G_t$ of $N_t$ banks, where $t$ indexes time (i.e. months). The dependent variable of interest is $\pi_{it}$, that is, the net profit obtained in a given month from all lending and borrowing transactions, and defined as

$$\pi_{it} = \sum_{h=1}^{H_t} \sum_{j=1, j\neq i}^{N_t} (L_{ij,h} r_{ij,h} - B_{ji,h} r_{ji,h}),$$

where $h = 1, 2, ..., H_t$ are the days within month $t$, $N_t$ corresponds to the banks that participate in the Call market during month $t$, $L_{ij,h}$ is the total amount lent to $j$ on day $h$ at rate $r_{ij,h}$, $B_{ij,h}$ is the total amount $i$ borrow from $j$ on day $h$ at rate $r_{ji,h}$.

We consider a separate network for each month from a total of 48 months (4 years). In each case we evaluate the proposed tests, $LM_{\mu,\delta}$, $LM_{\mu}$, $LM_{\delta}$, $LM_{\mu(\delta)}$ and $LM_{\delta(\mu)}$, and we compare them with [3] canonical tests for spatial dependence. In particular, we use Moran’s I LM test for spatial dependence. The figures 5-7 computes p-values in log-scale from all the tests considered here.

Several features of the exercise can be highlighted.

First, the proposed subgraph tests are different from test for spatial dependence. In order to show this we compute the joint test results for edge&triangle effects, $LM_{\mu,\delta}$, and the Moran’s I test for spatial error and spatial lag specifications. Figure 5 shows the p-values (in log-scale) of the tests for each month. The analysis reveals that in most cases, when $LM_{\mu,\delta}$ rejects (i.e. when the tests detect the presence of subgraph network random effects, of any type), spatial tests do not; and when spatial tests reject, only a few tests reject. This indicates that our proposed subgraph network structure is different from what would be captured by an model of spatial autocorrelation.

Second, there is heterogeneity across months in terms of the presence of edge, triangle, and edge&triangle effects. Figure 6 plots the test results for $LM_{\mu}$ (marginal LM test for edge effects) and $LM_{\delta}$ (marginal LM test for triangle effects). The figure reveals that there are more cases of only edge effects, while less so with triangle effects detected. When we explore this further using the robust LM tests in Figure 7, $LM_{\mu(\delta)}$ for edge effects controlling for triangle effects, and $LM_{\delta(\mu)}$ for triangle effects controlling for edge effects, a similar pattern applies.

7. **Conclusion.** This paper develops a simple model of subgraph network random effects that can be used to estimate the random structure with network data. It focuses on evaluating the appropriate level of effects, using the example of links' and triangles' effects as random components, and constructing a battery of specification tests.

Monte Carlo evidence shows that the tests correctly identify the type of network structure. Applying these tests to the Argentinean’s unsecured Call interbank market reveals heterogeneity across time in terms of the type of subgraph network random effects.

**REFERENCES**

[1] D. Acemoglu, A. Ozdaglar and A. Tahbaz-Salehi, Systemic risk and stability in financial networks, *American Economic Review*, 105 (2015), 564-608.
[2] P. Angelini, A. Nobili and C. Picillo, The interbank market after August 2007: What has changed and why?, *Journal of Money, Credit and Banking*, 43 (2011), 923-958.
[3] L. Anselin, A. Bera, R. Florax and M. Yoon, Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, 26 (1996), 77–104.
[4] L. Bargigli, G. di Iasio, L. Infante, F. Lillo and F. Pierobon, The multiplex structure of interbank networks, *Quantitative Finance*, 15 (2015), 673–691.

[5] S. Battiston, M. Puliga, R. Kaushik, P. Tasca and G. Caldarelli, Debrrank: Too central to fail? Financial networks, the fed and systemic risk, *Scientific Reports*, 2.

[6] M. L. Bech, J. T. E. Chapman and R. J. Garratt, Which bank is the central bank?, *Journal of Monetary Economics*, 57.

[7] A. Bera and Y. Bilias, Rao’s score, Neyman’s c(α) and Silvey’s ln tests: An essay on historical developments and some new results, *Journal of Statistical Planning and Inference*, 97 (2001), 9–44.

[8] A. Bera, G. Montes-Rojas and W. Sosa-Escudero, General specification testing with locally misspecified models, *Econometric Theory*, 26 (2010), 1838–1845.

[9] A. Bera, G. Montes-Rojas and W. Sosa-Escudero, A new robust and most powerful test in the presence of local misspecification, *Communications in Statistics - Theory and Methods*, 46 (2017), 8187–8198.

[10] A. Bera and M. Yoon, Specification testing with locally misspecified alternatives, *Econometric Theory*, 9 (1993), 649–658.

[11] G. G. Booth, U. G. Gurun and H. Zhang, Financial networks and trading in bond markets, *Journal of Financial Markets*, 18 (2014), 126–157.

[12] A. G. Chandrasekhar, Econometrics of network formation, in *The Oxford Handbook of the Economics of Networks*, Oxford University Press, Oxford, 2016.

[13] A. de Paula, Econometrics of network models, *CEMMAP Working Paper CWP52/15*, (2017).

[14] S. de Raco and V. Semeshenko, Labor mobility and industrial space in Argentina, *Journal of Dynamics & Games*, 6 (2019), 107–118.

[15] B. S. Graham, Network data, *NBER Working Paper 26577*, (2019).

[16] H. H. Kelejian and I. R. Prucha, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, 40 (1999), 509–533.

[17] H. H. Kelejian and I. R. Prucha, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics*, 157 (2010), 53–67.

[18] Y. Kumagai, Social networks and global transactions, *Journal of Dynamics & Games*, 6 (2019), 211–219.

[19] S. Langfield, Z. Liu and T. Ota, Mapping in the UK interbank system, *Journal of Banking & Finance*, 45 (2014), 288–303.

[20] J.-L. Molina-Borboa, S. Martínez-Jaramillo, F. Lopez-Gallo and M. van der Leij, A multiplex network analysis of the mexican banking system: Link persistence, overlap and waiting times, *Journal of Network Theory in Finance*, 1 (2015), 99–138.

[21] S. Poledna, J. L. Molina-Borboa, S. Martínez-Jaramillo, M. van der Leij and S. Thurner, The multi-layer network nature of systemic risk and its implications for the costs of financial crises, *Journal of Financial Stability*, 20 (2015), 70–81.

[22] A. Temizsoy, G. Iori and G. Montes-Rojas, The role of bank relationship in the interbank market, *Journal of Economic Dynamics & Control*, 59 (2015), 118–141.

[23] A. Temizsoy, G. Iori and G. Montes-Rojas, Network centrality and funding rates in the e-mid interbank market, *Journal of Financial Stability*, 33 (2017), 340–365.

[24] C. Upper, Simulation methods to assess the danger of contagion in interbank markets, *Journal of Financial Stability*, 7 (2011), 111–125.
Table 1. Empirical size

| N  | $LM_\mu$ | $LM_\delta$ | $LM_{\mu,\delta}$ | $LM_{\mu(\delta)}$ | $LM_{\delta(\mu)}$ | $LM_{\delta-\mu}$ |
|----|----------|-------------|-------------------|-------------------|-------------------|-------------------|
| 100| 0.009    | 0.016       | 0.0145            | 0.009             | 0.0165            | 0.0115            |
| 225| 0.012    | 0.0115      | 0.015             | 0.013             | 0.012             | 0.009             |
| 400| 0.013    | 0.012       | 0.0085            | 0.0095            | 0.0075            | 0.007             |

| Size 1% |
|---------|
| 100     |
| 225     |
| 400     |
| 100     |
| 225     |
| 400     |

| Size 5% |
|---------|
| 100     |
| 225     |
| 400     |
| 100     |
| 225     |
| 400     |

| Size 10% |
|----------|
| 100      |
| 225      |
| 400      |
| 100      |
| 225      |
| 400      |

Spatial queen structure

| N  | $LM_\mu$ | $LM_\delta$ | $LM_{\mu,\delta}$ | $LM_{\mu(\delta)}$ | $LM_{\delta(\mu)}$ | $LM_{\delta-\mu}$ |
|----|----------|-------------|-------------------|-------------------|-------------------|-------------------|
| 100| 0.0115   | 0.0105      | 0.0105            | 0.01              | 0.011             | 0.0115            |
| 225| 0.0175   | 0.0065      | 0.012             | 0.0145            | 0.0135            | 0.014             |
| 400| 0.0085   | 0.0085      | 0.0095            | 0.012             | 0.011             | 0.011             |

| Size 1% |
|---------|
| 100     |
| 225     |
| 400     |
| 100     |
| 225     |
| 400     |

| Size 5% |
|---------|
| 100     |
| 225     |
| 400     |
| 100     |
| 225     |
| 400     |

| Size 10% |
|----------|
| 100      |
| 225      |
| 400      |
| 100      |
| 225      |
| 400      |

Notes: Monte carlo experiments based on 2000 replications.

Figure 1. LM tests for edge effects, $\sigma^2_{\mu} = 0$, Erdős-Rényi random graph

(a) (b)

Notes: Monte carlo experiments based on 2000 replications. Solid line: $LM_\mu$. Dashed line: $LM_{\mu,\delta}$. Dotted line: $LM_{\mu(\delta)}$.

Received February 2020; revised April 2020.
Figure 2. LM tests for edge effects, $\sigma^2_\mu = 0$, Queen spatial structure

Notes: Monte Carlo experiments based on 2000 replications. Solid line: LM$_\mu$. Dashed line: LM$_{\mu\delta}$. Dotted line: LM$_{\mu(\delta)}$.

Figure 3. LM tests for triangle effects, $\sigma^2_\delta = 0$, Erdős-Rényi random graph

Notes: Monte Carlo experiments based on 2000 replications. Solid line: LM$_\delta$. Dashed line: LM$_{\mu\delta}$. Dotted line: LM$_{\delta(\mu)}$. Dash-dot line: LM$_{\delta-\mu}$.
Figure 4. LM tests for triangle effects, $\sigma_\delta^2 = 0$, Queen spatial structure

(a) \hspace{1cm} (b)

Notes: Monte carlo experiments based on 2000 replications. Solid line: $LM_\delta$. Dashed line: $LM_{\mu,\delta}$. Dotted line: $LM_{\delta(\mu)}$. Dash-dot line: $LM_{\delta-\mu}$.

Figure 5. Subgraph joint tests for edge and triangle effects and spatial Moran’s I LM test

Note: P-values in log-scale. Dashed line is the 10% critical value and dotted line to the 5% critical values. Horizontal axis corresponds to joint test for edge and triangle effects ($LM_{\mu,\delta}$). Vertical axis corresponds to Moran’s I LM tests for spatial error.
Figure 6. Subgraph tests for edge and triangle effects

Note: P-values in log-scale. Dashed line is the 10% critical value and dotted line to the 5% critical values. Horizontal axis corresponds to tests for edge effects ($LM_{\mu}$). Vertical axis corresponds to tests for triangle effects ($LM_{\delta}$).

Figure 7. Robust subgraph tests for edge and triangle effects

Note: P-values in log-scale. Dashed line is the 10% critical value and dotted line to the 5% critical values. Horizontal axis corresponds to tests for edge effects robust to triangle effects ($LM_{\mu(\delta)}$). Vertical axis corresponds to tests for triangle effects robust to edge effects ($LM_{\delta(\mu)}$).