AN ANALYSIS ON ENRON CORPORATION EMAIL NETWORK

Unveiling the Dynamics, Characteristics, and Indications Underlying the Enron Crisis

Contents

Keywords 3

Abstract 3

1. Introduction 3

2. Data Description 4

3. Analysis Platform 6

4. Enron Email Network Analysis 7

    4.1 Network Evolvement - comparison of the network level measure across time 7

    4.2 ERGM model - what characteristics make ties between employees more likely 12

    4.3 Analysis and Visualization of Particular Part of the Network 19

    4.4 A Closer Look at A Particular Key ‘Culprit’s Email Network – Kenneth Lay 32

5. Conclusion and Implication 37

6. Reference 38

7. Appendix 40

    Appendix Ⅰ - Ergm Result 40

    Appendix Ⅱ - Entire network visualization using igraph for each of the 47 months 89

    Appendix Ⅲ - Kenneth Lay’s network 114

An Analysis on Enron Corporation Email Network

- Unveiling the Dynamics, Characteristics, and Indications Underlying the Enron Crisis

Keywords

Social Network Analysis, Enron Corporation, Failure, Bankruptcy, Email, Behavior analysis, Interactive Visualization, Data Mining.

Abstract

The Enron corpus is a network containing over 500,000 emails sent from 150 employees of the Enron Corporation, which were made public due to legal action taken by the U.S. Federal Regulatory Commission (FERV) against the company. This work presents how we conduct the social network analysis (SNA) using the Enron corpus. Three different angles of looking at the network were applied. These different approaches are (a) comparing the network-level measures over time, (b) identifying group-level features using ERGM model, and (c) detect the activity and presence of important actors with respect to the Enron Scandal.

1. Introduction

Enron Corporation, an American energy, commodities, and services company, has been named by Fortune magazine 'America's most innovative company' for six consecutive years. It was America's seventh-largest company in the year 2001, with its annual revenue exceeding $100B. The company seemed to be operating in perfect condition, with its stock price kept rising and shareholders gained ever-increasing shareholder equity. However, the turning point was about to happen. On October 16, 2001, Enron reports a $618M loss, its first quarterly loss in the company's history, along with a $1.2B reduction in its shareholder equity. Not soon after, Enron filed Bankruptcy. As a result, nearly 100,000 people lose their jobs. The Enron scandal has become one of the largest corporate corruption and accounting frauds in American's history. Numerous of its senior management members were sentenced to jail.

So, what happened? In order to understand it, we would like to find out if there is any indication of the collapse of Enron before the crisis happened. Additionally, we would like to identify a 'crisis point' and find out some of the characteristics of its email network before and after this 'crisis point'. We would conduct both network-level analysis and node-level analysis on the email network.

2. Data Description

In the event of disclosure of Enron's having hidden millions of dollars of debt from failed projects by SPV and financial loopholes, the US SEC and Federal Energy Regulatory Commission(FERC) quickly instituted investigation in October 2001, where about 500000 internal emails were made public online for transparency originally, making it the largest public domain database of real-world company internal emails in the world. This grounds the foundational data for the engaging research topic of our study. Although the original enormous dataset was involved in a number of integrity problems and was truncated or transformed in some way to protect affected employee's privacy later, a lot of efforts have been made by a variety of scholars and contributors to provide a cleaner, more usable dataset so far. We use mainly the cleaned version posted in Arne Hendrik

Ruhe's website (http://www.ahschulz.de/enron-email-data/) for our project, primarily a RData file containing the email message data (such as senders and receivers) and a little further information for some specific people(referred to as the special "employeelist" of 149 people here). We did a little data wrangling including:

- removed invalid "dates"(in yyyy-mm-dd format) in the message data file, which could have resulted from recording error, such as dates with year like "0001-", "0002-" and any date after 2007 (e.g.: 2020) when Enron Corporation should have finished legal dissolution

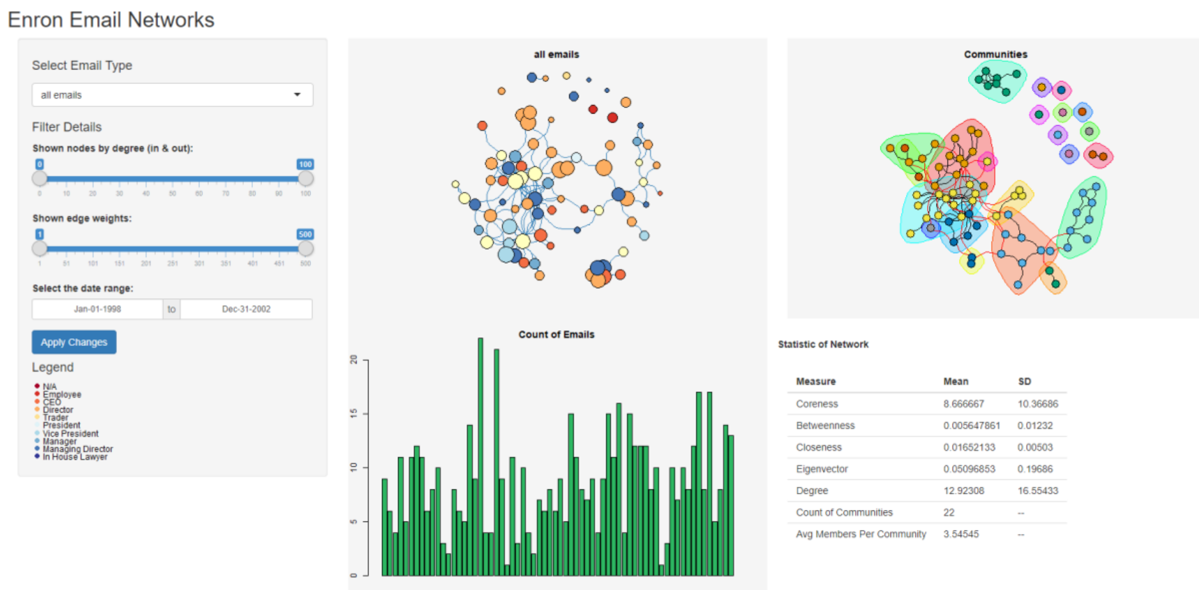
- removed any leading or trailing whitespaces in the "sender" and "receiver" variable in the email message related data

After finishing the data cleaning process, before integrating the data into edge list, we have 17500+ unique senders and 68084+ unique receivers in the email data with 140600+ unique email subjects within the time periods of 1215 unique dates. After integrating data into email network edge list (data extracted from the computers of the 149 people in the "employeelist" was applied here), we got an email network with 375000+ ties and 16169 nodes(with redundant observations removed), covering three email types – "TO"(280000+ ties), "CC"(47000+ ties) and "BCC"(47000+ ties), within the time window ranging from 1979-12-31 to 2002-09-22 in 47 months. We also noticed most of the people appearing in the huge email network do not have "status" (role) information (e.g.: "manager", "director", etc.). Furthermore, there could be significant emailing activities between a pair of actors even in a day sometimes: for example, the biggest number of emails between a unique pair of sender and receiver in the network within a day we counted is 333, which was mostly sent by an sender address called "stacey.w.white@enron.com".

3. Analysis Platform

Since our data varies along time mainly from October 1998 to June 2002, including the periods before, during and after the Enron crisis, investigating the transformation of the internal email network over time becomes essential.

Besides the general R approaches or packages (igraph, ggplot, network, etc..), we introduced Shiny package to construct an analysis platform, which makes it easy to build interactive web apps straight from R. We built the following dashboards based on Shiny, in which the control panel allows the user to choose the network and related statistic data based on time, in/out degree, and weights. This flexible and interactive platform provides an effective way to progress our further study.



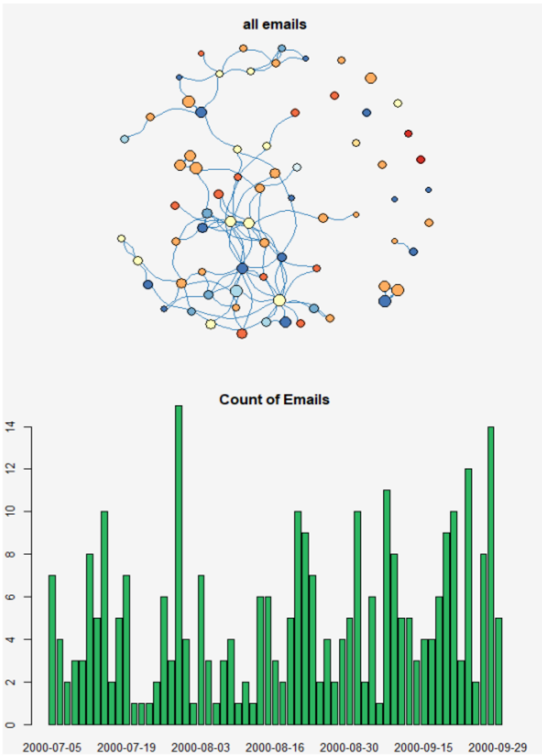
4. Enron Email Network Analysis

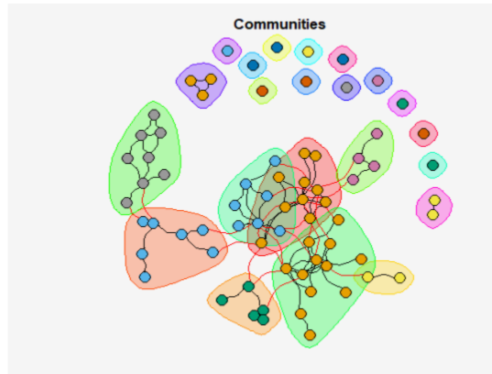
4.1 Network Evolvement - comparison of the network level measure across time

Enron's stock was performing well and hit its all-time high of $90.56 in August 2000. We consider time around this date as its 'prime days'. In October 2001, Enron reported its first quarterly loss and filed for bankruptcy soon after. Thus, we consider October 2001 to be its 'crisis point'. In order to better understand the change in characteristics before and after the crisis point, we have segmented out four snippets of time, which are 2000 quarter 3 and quarter 4, as well as 2001 quarter 3 and quarter 4. We would like to compare the network-level measures between these four time periods and see how the network evolved over time, in order to find out any indication of the collapse of Enron beforehand.

First, looking at the network level measures in 2000 Q3 (figure 1), we can see that people were constantly communicating via emails throughout the months. Additionally, most people are formed into clusters (also called communities), with only a few members being isolated in the peripheral. The table in the corner right shows different measures of network centrality, which we will later compare with centrality measures from other time periods.

Figure 1

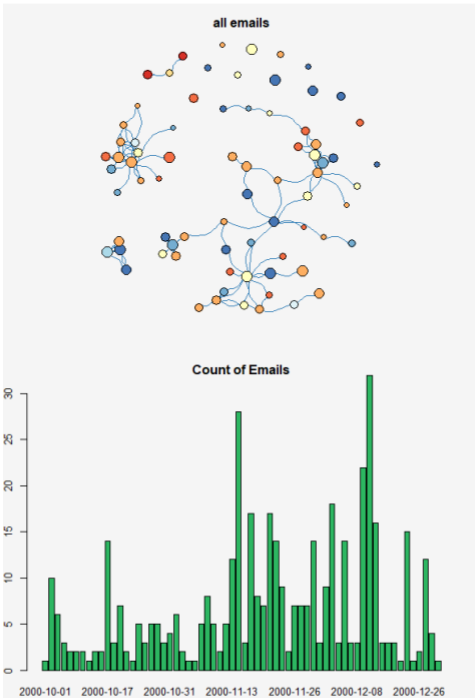


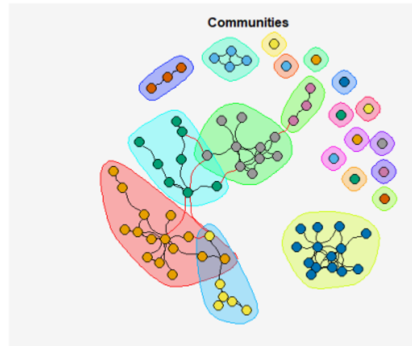


Measure	Mean	SD
Coreness	5.06	6.2526363766688
Betweenness	0.01	0.0141582809416378
Closeness	0.02	0.00539910951295343
Eigenvector	0.04	0.170150818534397
Degree	7.36	9.51989104499664
Count of Communities	22.00	--
Avg Members Per Community	3.27	--

Next, we investigate the network-level measures in 2000 Q4 (figure 2) and 2001 Q3. We can see that the features of the network in these two different time periods did not change that much compared with that of 2000 Q3. The average degree per node has increased from 7.36 to around 10. The average number of members per community are still around 3.5. Most of the people are still formed into groups, with a few people isolated in the peripheral. However, when we look at the count of emails in 2001 Q3, we can see that there is an obvious spike around the end of September and the beginning of October 2001. Since Enron has not released its quarterly loss to the public until October, we suspect that people within Enron have known about the upcoming ill-performed report a few days prior to its actual release date and started to panic, sending out emails asking what is really going on.

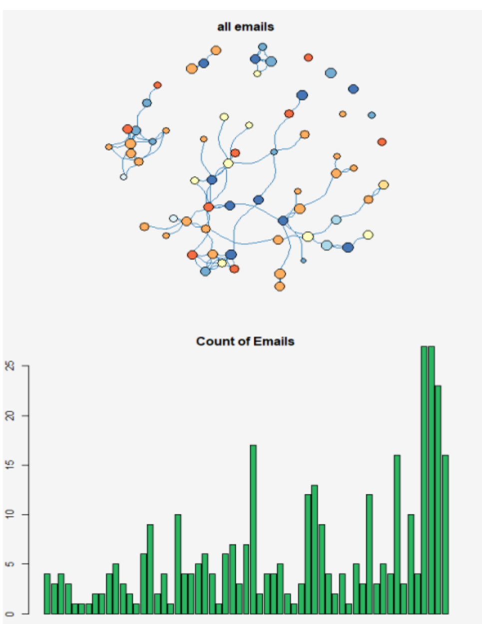
Figure 2

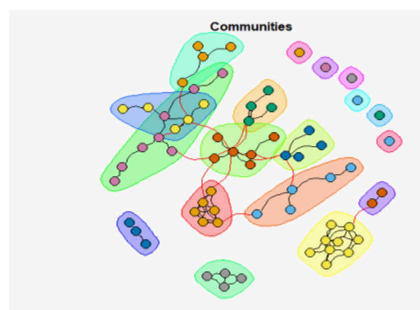




Measure	Mean	SD
Coreness	6.73	7.5037528148458
Betweenness	0.00	0.00498509210539595
Closeness	0.01	0.00227614209522121
Eigenvector	0.03	0.163334664574842
Degree	10.35	14.2359397997456
Count of Communities	20.00	--
Avg Members Per Community	3.75	--

Figure 3

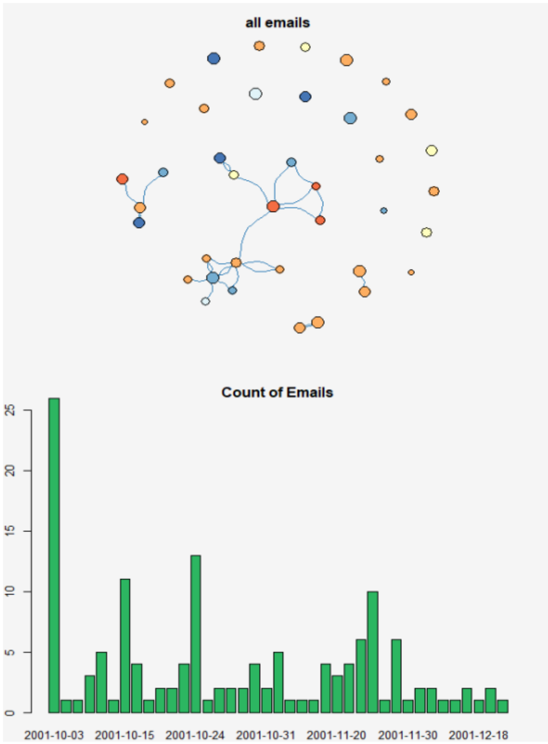


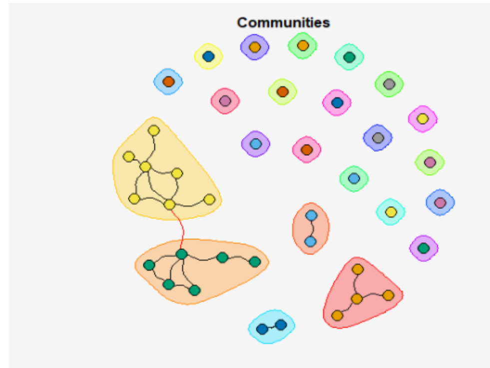


Measure	Mean	SD
Coreness	6.70	6.34876050160659
Betweenness	0.00	0.00688113859144126
Closeness	0.02	0.00323440891809276
Eigenvector	0.06	0.211912612061203
Degree	10.09	11.250609929076
Count of Communities	18.00	--
Avg Members Per Community	3.67	--

Finally, we look at the network level measures in 2001 Q4 (figure 4). We can find that there is an obvious spike at the beginning of October 2001. That was when Enron reported its first quarterly loss to the public. Soon after that, the count of emails sent by Enron employees dropped dramatically. We suspect that people have changed to other ways of communication, such as face to face, since they knew emails are traceable and thus did not want to talk about sensitive information via emails. Furthermore, when looking at the network graph, we can see that people stopped forming groups. Many people have left their groups and isolated themselves from others. The average degree per node has dropped to 4.41, and the average number of members per community have dropped to 1.7. We infer that the reason behind is because people wanted to stay away from troubles and crisis. Or perhaps they have formed into groups via other means of communication.

Figure 4

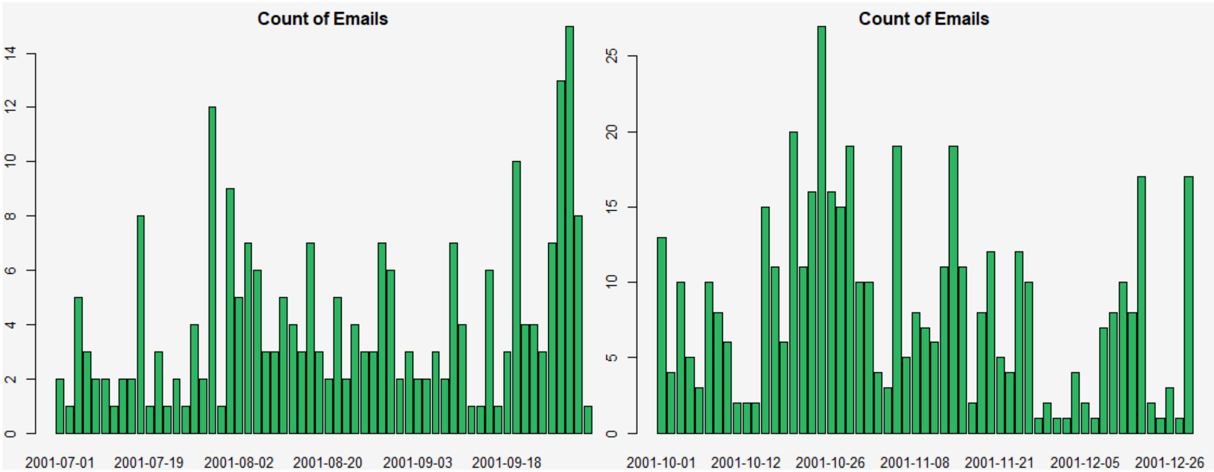




Measure	Mean	SD
Coreness	3.28	4.23611489052433
Betweenness	0.00	0.00259299478935049
Closeness	0.03	0.0023037086356927
Eigenvector	0.09	0.218879629665649
Degree	4.41	7.47176421327918
Count of Communities	23.00	--
Avg Members Per Community	1.70	--

On top of that, the amount of Bcc emails has skyrocketed after the crisis happened. When looking at the count of Bcc emails for 2001 Q3 and Q4 (figure 5), we can see that after the crisis point, people started to send out much more Bcc emails. Bcc stands for blind carbon copy. For emailing, people use Bcc when they want to copy others privately. Any recipients on the Bcc line of an email are not visible to others on the email (Bridges). We can consider Bcc emails as a kind of 'secret emails. After the crisis happened, Enron employees were more comfortable sending Bcc emails since those emails might contain sensitive information.

Figure 5



4.2 ERGM model - what characteristics make ties more likely

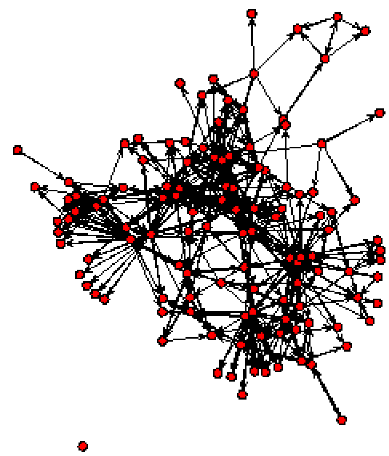
After we visually explored our data according to network charts before and after the crisis, we want to take a further step to figure out what kinds of characteristics make ties between employees more likely. We will utilize new technology: Statistical exponential family models.

Recent advances in the statistical modeling of random networks have had an impact on the empirical study of social networks. Statistical exponential family models (Strauss and Ikeda 1990) are a generalization of the Markov random network models introduced by Frank and Strauss (1986), which in turn derived from developments in spatial statistics (Besag, 1974). These models recognize the complex dependencies within relational data structures.

Back to the R language, ergm is a collection of functions to plot, fit, diagnose, and simulate from exponential-family random graph models (ERGMs). It introduces software tools for the representation, visualization, and analysis of network data. Especially, the ergm package implements maximum likelihood estimates of ERGMs to be calculated using Markov Chain Monte Carlo, as well as provides tools for simulating networks and assessing model goodness-of-fit.

We will apply “statnet” package to do this analysis. To illustrate our process, we will take 2000Q3 as an example, to show how we investigate what kinds of characteristics make ties between employees more likely.

Step 1: Build up the network: instead of igraph package we used before, we need to apply another package “network” to recreate the same network for 2000Q3. the following plot shows the email network of 2000Q3:



Step 2: Assign the attributes to all the vertices(employees). We create 3 attributes: ‘gender’, 'department' and 'seniority'. The raw data does not provide such information, however, we created them manually: gender could be inferred from names; department could be searched from google or LinkedIn, while seniority could be assigned to senior if its title is a director or above.

For example, we have Mary Hain in the employee list without any detailed information, however, we could deduce ‘Mary’ is obviously a female name. After that we googled Mary Hain Enron, we got the following information:

SFGATE https://www.sfgate.com/business/article/Attorney-may-hold-Enron-key-2819946.php

Attorney may hold Enron key

David Lazarus Published 4:00 am PDT, Wednesday, May 22, 2002

All the attention in political circles may be on former **Enron** bigwigs **Ken Lay** and **Jeff Skilling** and on price-gouging schemes with nicknames like "Death Star" and "Get Shorty."

But if Enron is to get its comeuppance for shamelessly exploiting California's power woes, the central figure may turn out to be an obscure lawyer named **Mary Hain**.

Now, we could assign her department as 'legal' and seniority to 'junior'.

Step 3: Process the ergm regression. We apply edges, mutual and nodematch/nodefactor to study the email network, then get the estimate parameter of our network.

Below are the details of the parameters of the model:

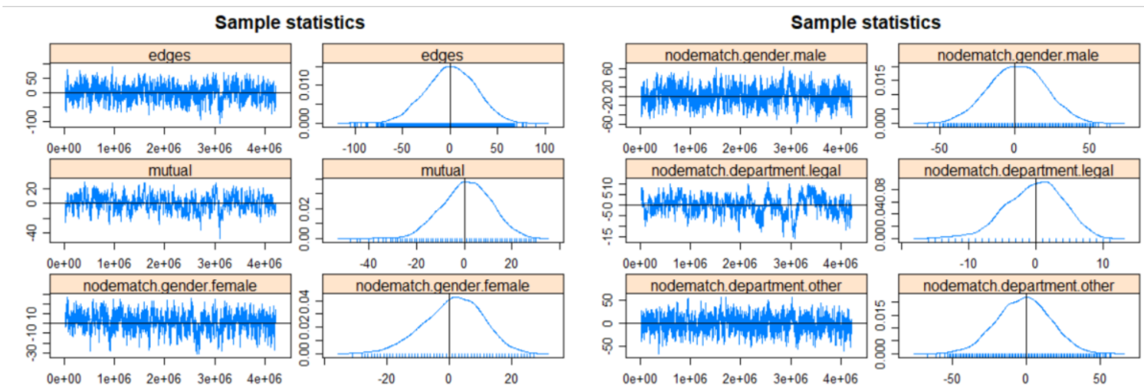
(1) Edges: the number of network edges, which is related to the network density, and reflects the effect of the network edges. It is equivalent to the constant term of the regression model.

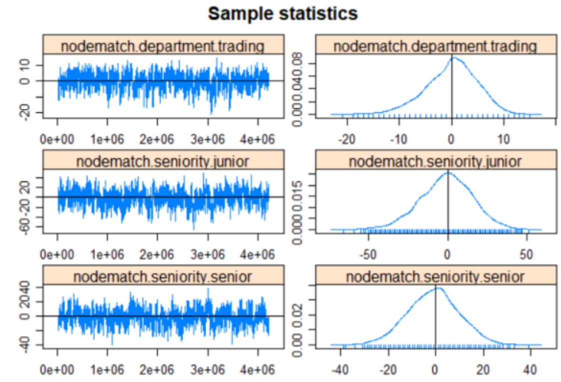
(2) Mutual: If an outgoing edge receives a direct feedback edge, it is recorded as a reciprocity relationship (a mutual). A positive regression coefficient of this item is indicating that compared with the case of random edge connections, the actual network is mutually beneficial orientated.

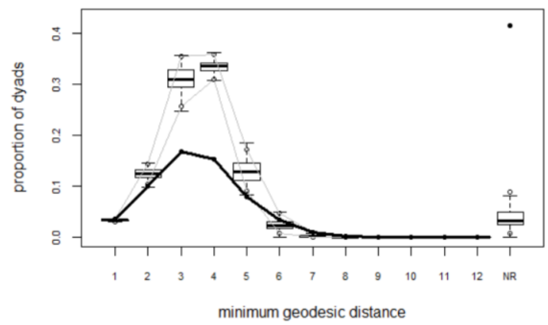
(3) Nodefactor: The node attribute effect reflects that if a node has certain attributes (classification attributes), then the probability of the node forming an edge is higher.

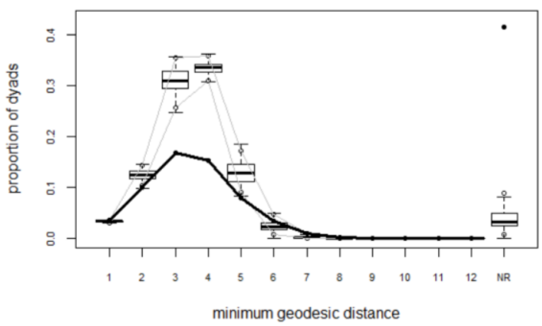
(4) Nodematch. This study defines it as the number of network edges where the nodes at both ends have the same attributes. This statistic can capture the effect of node homogeneity on the network. The attribute studied here is also the area where the network node is located.

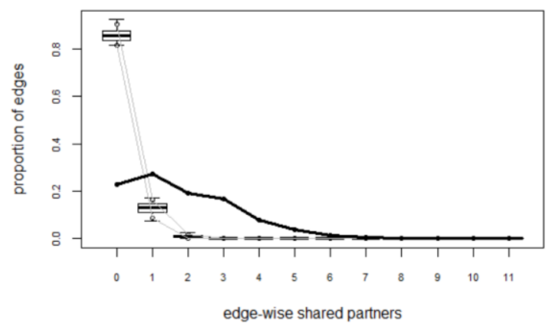
Step 4: Conduct MCMC diagnostics on a model fit and gof calculates p-values for geodesic distance, degree, and reachability summaries to diagnose the goodness-of-fit of exponential family random graph models:

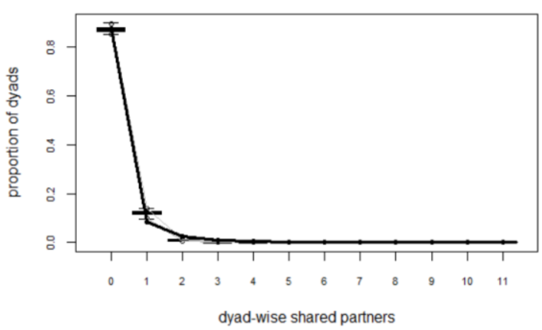


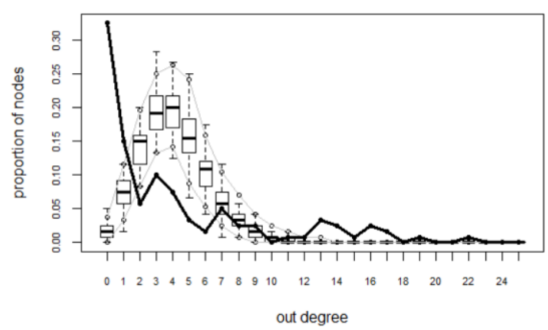

Sample statistics
nodematch.department.trading
nodematch.department.trading
nodematch.seniority.junior
nodematch.seniority.junior
nodematch.seniority.senior
nodematch.seniority.senior

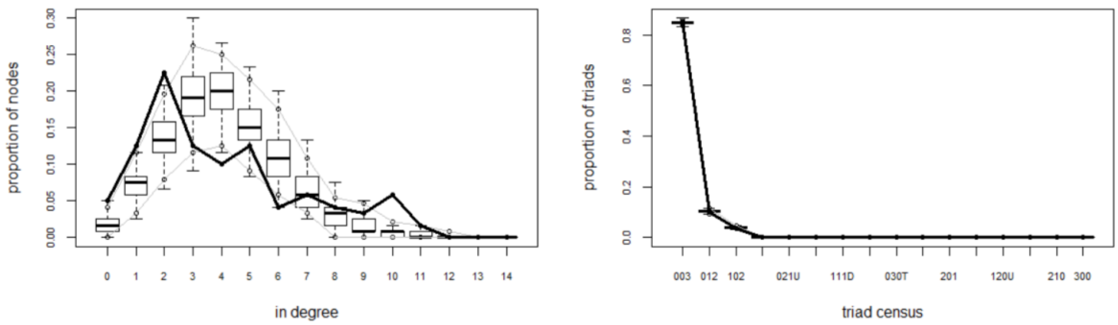

proportion of dyads
minimum geodesic distance


proportion of dyads
minimum geodesic distance

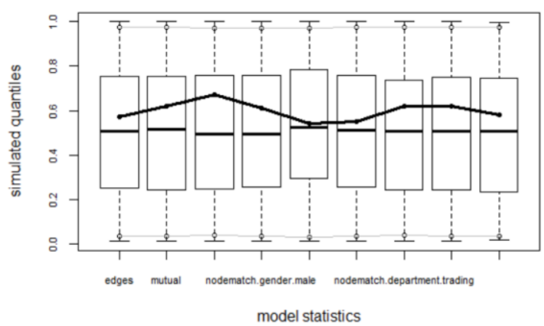

proportion of edges
edge-wise shared partners


proportion of dyads
dyad-wise shared partners


proportion of nodes
out degree



Goodness-of-fit diagnostics



Goodness-of-fit analysis is a comparison of the actual network and the fitted model. The thick black solid line in the analysis results represents the observation network, while the light black line represents the measurement result of the simulated network at the 95% confidence interval.

When the black line falls between the light black lines, it indicates that the simulated network can well represent the structural characteristics of the observation network. Otherwise, the goodness of fit is low.

As the graphs illustrate, most indicators can reflect the characteristics of the real network. Overall, the model fits satisfactorily and can capture statistical characteristics' key mechanism. After all the procedures, we list all the output from ergm as below:

nodefactor	**2000Q3**		**2000Q4**		**2001Q3**		**2001Q4**	
edges	-3.119	***	-3.289	***	-3.326	***	-3.128	***
mutual	3.568	***	4.045	***	3.534	***	3.805	***
gender.male	-0.039		-0.089		-0.142	**	-0.228	***
department.other	-0.418	***	-0.371	***	-0.188	*	-0.162	*
department.trading	-0.502	***	-0.350	***	-0.394	***	-0.272	**
seniority.senior	0.100	.	0.165	***	0.345	***	0.286	***
nodematch								
edges	-3.962	***	-4.096	***	-3.853	***	-3.603	***
mutual	3.579	***	4.015	***	3.479	***	3.788	***
gender.female	0.443	***	0.434	***	0.395	***	0.322	***
gender.male	0.083		0.051		-0.034		-0.188	***
department.legal	1.106	***	1.045	***	1.945	***	1.703	***
department.other	-0.065		-0.041		0.205	**	0.146	**
department.trading	0.127		0.405	**	-0.073		-0.097	
seniority.junior	0.051		0.014		-0.188	*	-0.191	***
seniority.senior	0.184		0.416	***	0.566	***	0.445	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1								

We can conclude from pre-crisis to post-crisis:

From the gender aspect, 'male' sent fewer emails, and tend to send more email to 'female' rather than to 'male'. From the department aspect, the legal team or other team become more likely to send emails inside the legal team or other team. However, the trading team shows the opposite behavioral. From the seniority aspect, seniors sent more emails post-crisis, and tend to send emails to other seniors. instead, juniors sent emails to juniors pre-crisis, but sent to outside of juniors post-crisis.

In summary, with ERGM approach, we can tell that during the transition time of crisis, these email network indicates abnormal behaviors appeared across the whole company. However, it is still incapable to answer the question of who the troublemaker is ultimately. We need to dive deeper into the data set to figure it out.

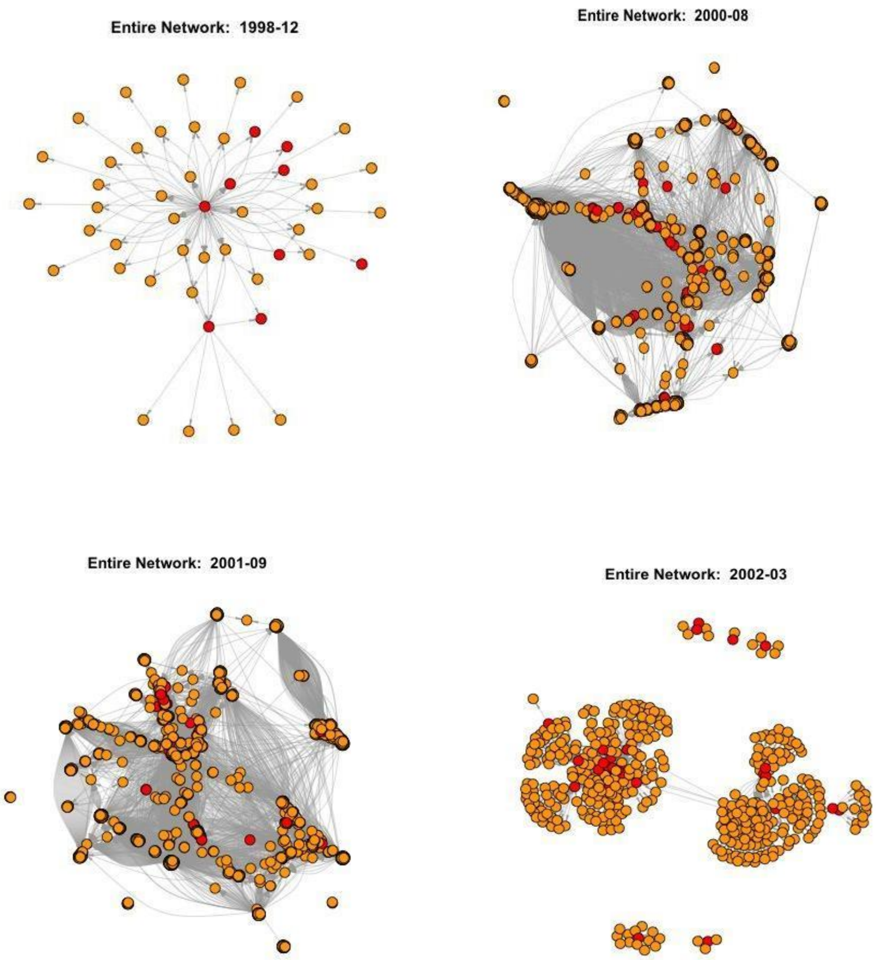
4.3 Analysis and Visualization of Particular Part of the Network

After obtaining perception from the entire network containing about 52000+ observations for all email addresses in the Enron email network data for each month, it is time to further drill down to some particular parts of the entire network and their dynamics for more thorough examinations. Remember that we have a special list of 149 people, including many of the senior management and sensitive staff in the Enron Corporation, among whom a good number were involved in the scandal or even prosecuted for a multitude of misdeeds later for the Enron Collapse. The analysis here focuses on the network comprised of these 149 people[1] and probes into different aspects of the network characteristics.

First of all, in order to get an impression of where the actors on this special list position in the broader entire email network (all three types of emails "To", "CC" & "BCC" included here), as well as how it evolves, graphs of the entire network are plotted at each time period of the 47 months ***(for all graphs, see Appendix)***. Note that the nodes denoting actors on the special list have been

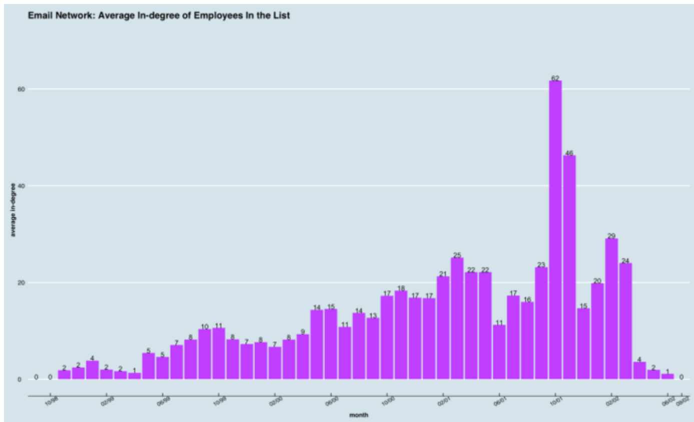
1 By combining information from this special employee list and the original large email network data, two people on the list, Andrew Lewis(with role of "Director") and Darron Giron(with role of "Employee"), have been found to have missing data in the email network, probably due to inconsistent email address or lack of records. These two people are not currently included from the analysis.

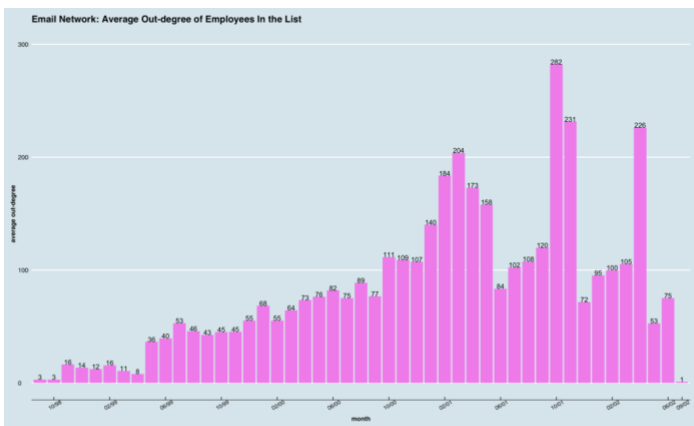
marked red for easier distinction. Similar to the analysis approach before, we pick snapshots of the network structures at particular points – “1998-12”(earlier periods in our data), “2000-08”(golden times of Enron’s stock price), “2001-09”(right before the crisis broke out), and “2002-03”(after Enron’s filing for bankruptcy):



From these snapshots of entire network visualization, we may notice that at the very early periods in our data (e.g.: “1998-12”), the number of nodes is relatively small and all nodes seem to distribute in space more evenly; regarding those nodes in red(people in the special list), it seems there was almost always one in the center of the emailing network (but the person in the center changed across different months), a handful of others in the list acted like brokers, yet still some others rather stayed relatively periphery in the network. Later during the more prosperous times in Enron's stock history (e.g.: "2000-08"), it appears that the number of emailing ties among nodes increased significantly, and those nodes in red spread more randomly among other "common" nodes not in the special list. While people tended to form a lot of emailing ties, some nodes started staying closer to each other in the graph visualization while some others did not but rather stood by themselves. When it came to the periods right before the bad news exploded (e.g.: "2001-09"), we may see that some smaller-group-like pattern made up by the nodes still existed, yet general nodes appeared to be a little further away from each other with the attachment slightly loosening. A few more isolates may come into view. Within the special node list, some actors may even become nearer among themselves. At last, during some amount of time after Enron went bankrupt (e.g.: "2002-03"), some kinds of more apparent, seemingly close-packed cluster structures emerged, and there are also coexisted small groups of nodes isolated from the larger cluster structures at this point, including certain nodes in the special list as well. The brokerage role, bridging some subdivided groups together and filling in the structural hole, of the special nodes in red seemed to recur here interestingly, about which would be discussed more later.

Now that we have got a rough perception of what the entire network looks like visually with those in the special list signed, let’s look at the average level centrality measure statistics specifically for the people in the special list.[2]





2 An especially noteworthy caveat of the number calculation here is that the absolute magnitude of certain measure could be biased by the data source itself. For example, it could be noticed that the betweenness centrality here of those in the “special” list is tremendously high sometimes, which turns to be remarkably higher than the rest of the nodes (and the whole network in general); but the cause might be rooted in the fact that this dataset used comes exactly from those 149 “special” people’s computers during investigation.

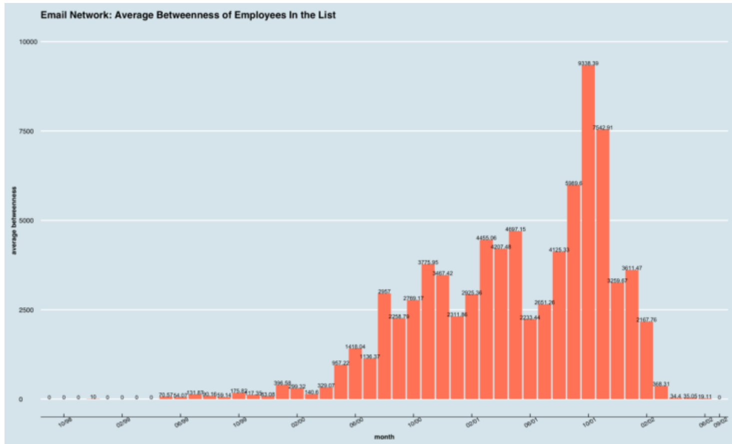
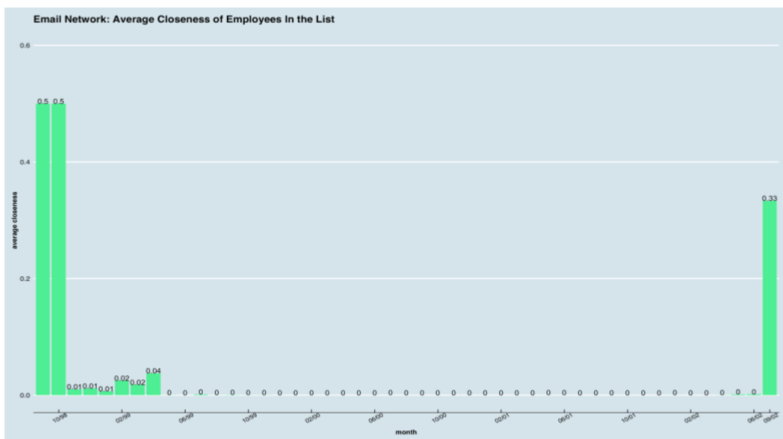
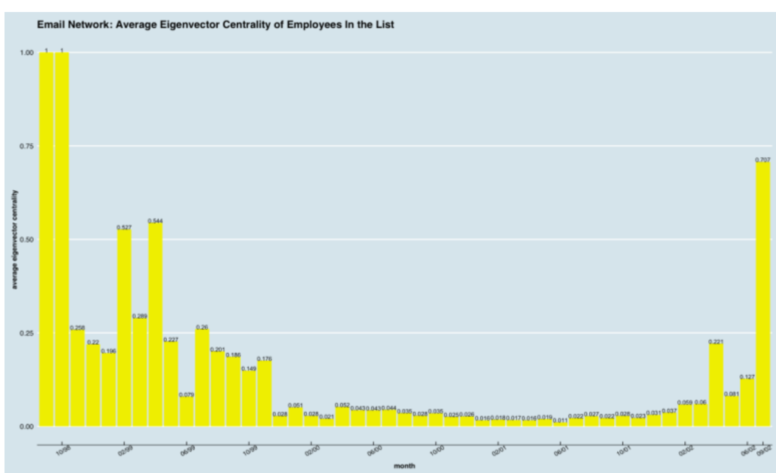


Figure 6: Average In-degree, Out-degree & Betweenness Centrality Measures for People in the Special List





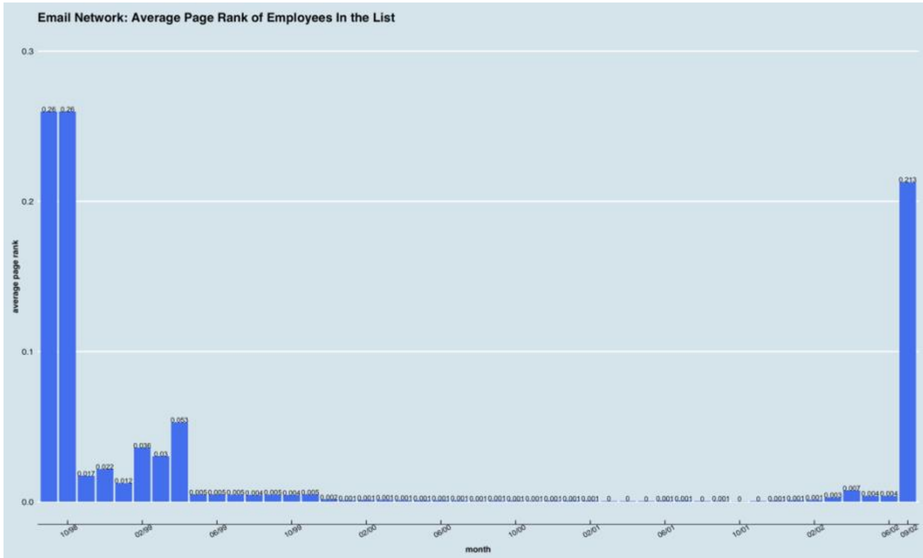


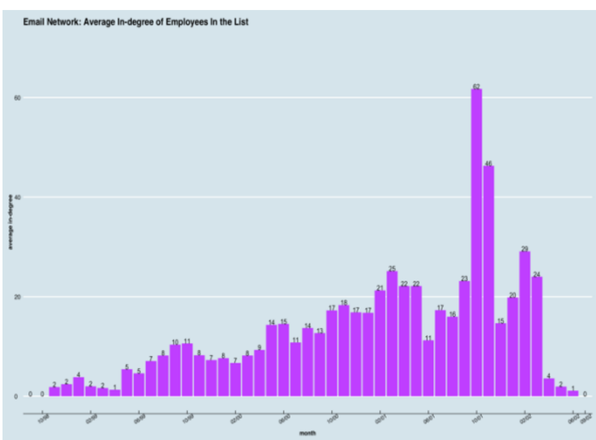
Figure 7: Average Closeness, Eigenvector & Page Rank Centrality Measures for People in the Special List

We use R ggplot2 package to plot the changes of the average of 6 main centrality measures over each month above – in-degree centrality, out-degree centrality, closeness centrality, betweenness centrality, eigenvector centrality and page rank centrality – for the network consisting of emails involving people in the special list only. The major observations we get there turn out to be interesting. For three of the 6 centrality measures discussed here – in-degree, out-degree and betweenness centrality, they tend to reflect a similar trend of this particular network, whose magnitude basically grew steadily for the first 31 months(till around March 2001), decreased a little bit during the next few months, increased again to reach the historical high, and then winded down towards the end(except one month around February 2002) as the Enron quickly collapsed. Notably enough, these three measures all arrived at their peak of magnitude at exactly the point of time when Enron's astounding hidden losses got revealed for the first time – in October 2001. Based on the dynamics of these three measures, we might interpret that the people in the special list presented a higher degree of activeness and stronger role of brokerage in the email network around the troubling times than the normal periods; an explanation for this phenomenon could be that these special actors were deeply caught in the seething concerns and had to deal with the exposure through frequent emailing then; while they also tried to impose intense control of information and conversations within the corporation at the same time.

However, when we turn to look at the other three centrality measures – closeness, eigenvector and page rank centrality, there presents another narrative. Opposite to the trend shown by the three centrality measures analyzed above, these three measures here displayed the same tendency that the magnitudes achieved their biggest levels at the very beginning and end of the entire time window observed, when there are only a small number of actors in the broader email network; but during the longer periods in the midst of our observations when the whole network got more populated, those people in the special list had much smaller magnitudes regarding these three measures, including the periods around the exposure of the bad news, as if they remained fairly "quiet" in between. This might imply that the people in the special list stayed relatively far away from most of the remaining "normal" nodes in the entire network; neither did they have a significant influence exerted in the interim.

Taking both groups of the three centrality measures into consideration, it may be observed that the actors in the special list exhibited a combination of high degree and betweenness centrality with low closeness. Thus, it might be reasoned to state that those actors in the special list developed some interesting patterns of potential embedded smaller clusters that were somehow far from the center of the entire network, and more intriguingly, might have monopolized the ties from a small number of actors to many others, during the unsettled periods of Enron Crisis. This could help us better understand how those “sensitive” people were behaving when sinister events that fundamentally undermined a large corporate stroke.

Given the overall understanding of the dynamics within this particular part of the broader network containing 149 actors, now we inspect in further detail the potential effects of people’s “status” (or role) on their behaviors in terms of the 6 main centrality measures discussed previously. We have 10 different categories in the “status” variable in our dataset, including “Employee”, “Trader”, “Manager”, “Director”, “President”, and more senior “CEO” and others. Note that the plots of the 6 main centrality measures are provided at each of the 4 particular points of time chosen here – “1998-12”(earlier periods in data), “2000-07”(prime time of Enron’s stock price), “2001-09”(a month right before the uncovering of the huge losses), and “2002-01”(one month after Enron filed for bankruptcy), in which the measure of each actor has been demonstrated through the grouping of coloring block representing different “statuses”(roles).



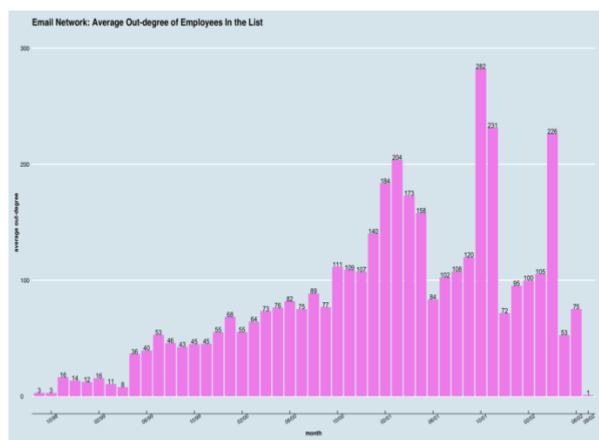


Figure 8: Six Main Centrality Measures of Each Actor in the Special List during December 1998

Figure 9: Six Main Centrality Measures of Each Actor in the Special List during July 2000

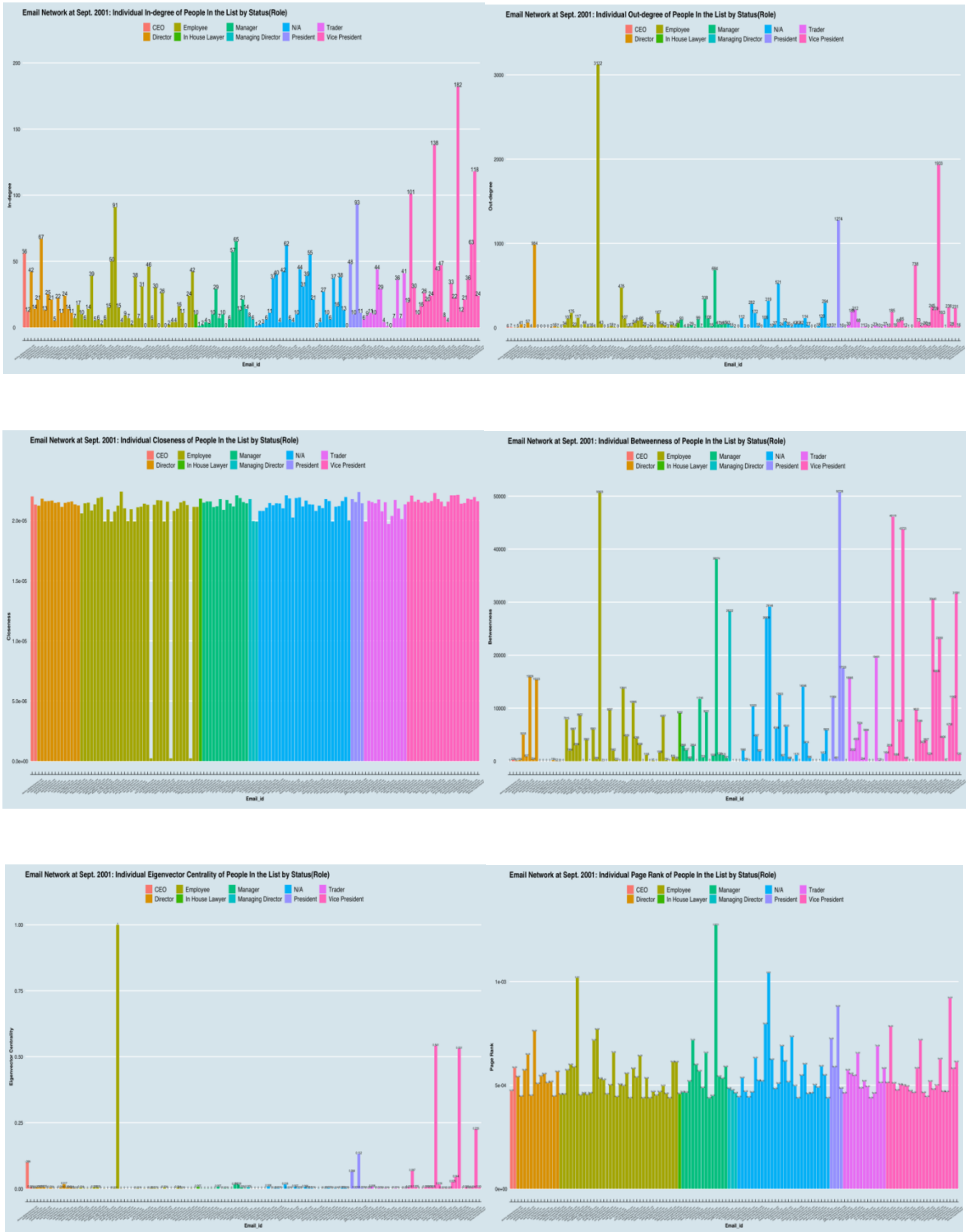


Figure 10: Six Main Centrality Measures of Each Actor in the Special List during September 2001

Figure 11: Six Main Centrality Measures of Each Actor in the Special List during January 2002

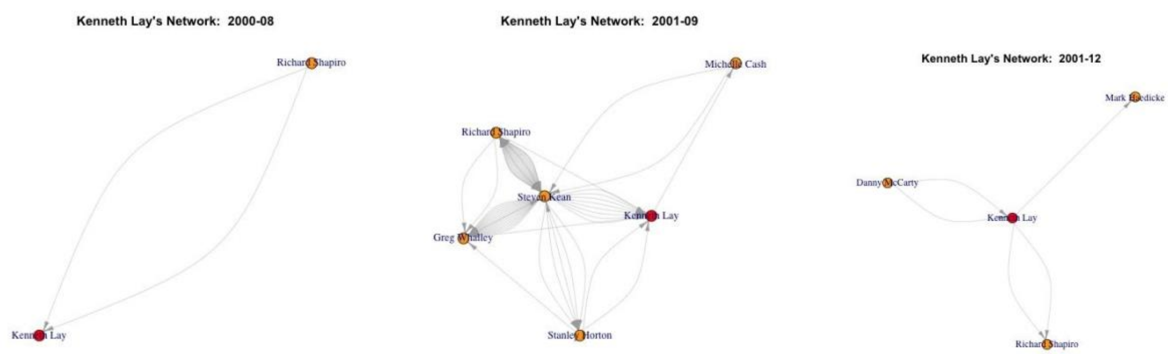
Reviewing all the plots depicting the part of the network consisting of the 149 special actors across different times above, a quick first impression would be that overall, much less variation or partition was observed (with a few exceptions) in two of the centrality measures – closeness centrality and page rank centrality – than any other centrality measure in terms of this network of 149 special actors, probably indicating that few people in this special list really came to be much closer to certain specific other than the rest, but they rather tended to scatter among the nodes. Another noteworthy finding here turns out to be that the top management of Enron Corporation, namely the "CEO" type of people(CEO, CFO, etc. might be included), seemed to somehow "retreat to the backstage" with relatively low values in many of the centrality measures here, especially with regard to the eigenvector centrality which measures influence. In comparison, people of some other "statuses", such as the "employees", behaved more active, close to others, and more influential sometimes. There might be quite a lot potential reasons for this: the top management might have been less involved, or "dominant" in the specific daily routines of work which required lots of emails; or in Enron's internal organization, a more "employee-initiated" culture or atmosphere might have been more prevalent, while most common employees might not be very clear about what the top corporate managements were manipulating "on the quiet" then. In addition, the scales of some of the measure statistics here verify what we have discussed in previous sections again – the closeness centrality of people in the special list around the occurrence of the crisis was much smaller than that during the normal times, indicating those people were getting more distant with each other, while the betweenness centrality became even more than 4 times as big as that during the usual times in general, serving as a possibly clear signal of stricter restrictions of information flow.

There might be some intriguing indication about the effects of “status” (role) those special people engaged here. The “Employee” and “N/A”(we are not quite sure about their status) type of people seemed to remain high activeness, closeness, betweenness and influence in the first two points of time chosen in particular, which represented those earlier, more “flourishing” and “successful” times for Enron Corporation. Nevertheless, in the later “failing” periods, people of two types of “statuses” -- “Vice President” and “President”, who approached the highest rank within the corporation only second to the “CEO” s, started standing out, especially regarding the betweenness measurement. The people of “Vice President” status emerged during the month right before the bad exposure, and those of “President” status showed up more as soon as the bankruptcy occurred. One potential inference we make here could be that instead of coming straight forward to handle the troubles themselves during such inflamed days, the apprehensive top management might choose to pass their will towards one lower level of senior management while still attempting to keep themselves concealed in some way.

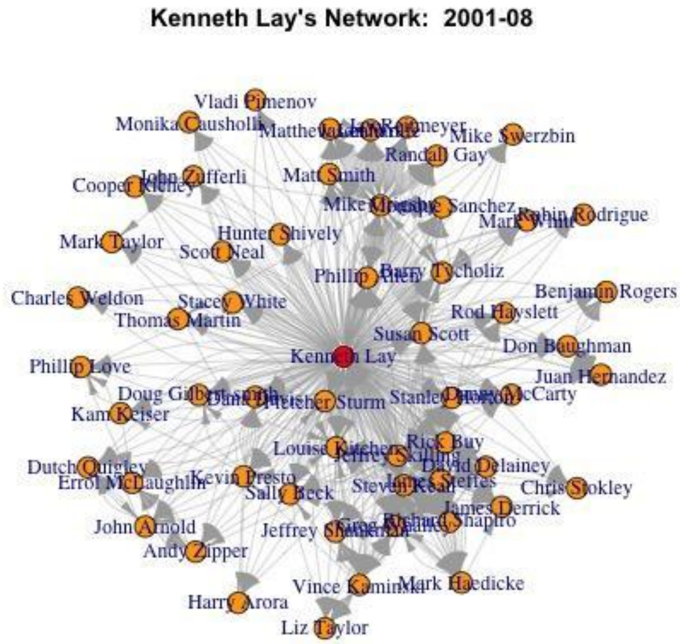
4.4 A Closer Look at A Particular Key ‘Culprit’s Email Network – Kenneth Lay

Having gained insights about the part of the broader network composed of the 149 people in the special list, we eventually zoom in on specific nodes to take a detailed investigation on particular person of importance in our case. Kenneth Lay (or “Kenneth Lee Lay” in full name), the founder of Enron Corporation in 1985 and former CEO and Chairman of the company, was one of the “key villains” heavily involved in the stunning Enron Scandal, and later in the trial of him and Jeffrey Skilling(CEO & COO) in 2006, he was convicted of 10 counts of securities fraud and false statements and could have been sentenced up to 45 years in prison before he died on July 5, 2006, several months earlier than the final sentencing and legal execution. Lay had a unique positioning in approaches of exploring the "origin of evil" in our study, so we made a closer look at his network as an example of key-person examination here.

Like our approach in previous sections, we first roughly plot the email network directly involving him across all months with data about him. This periods in our data starts from October 1999 and ends at January 2002, with 21 months in total, but the time point is not always continuous. We presented here snapshots of his network at three particular points of time – "2000-08"(prime time of Enron's stock price), "2001-09"(one month before the exposure of the scandal), and "2001-12"(when Enron filed for bankruptcy):

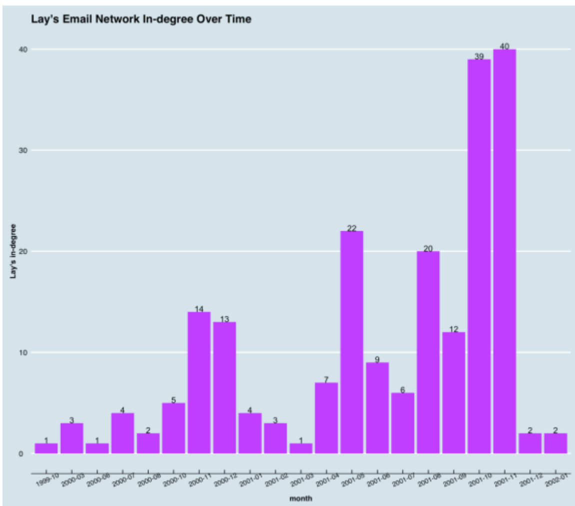


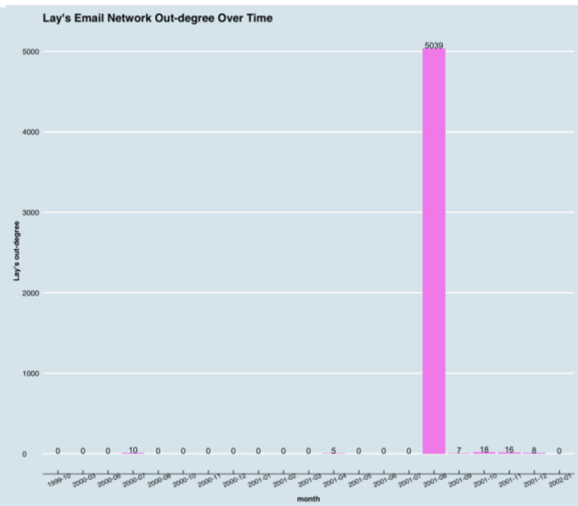
It could be interesting that in our data, Kenneth Lay had only a few email contacts most of the time. However, in light of all visualizations across all 21 months, we found certain month when there could be exceptionally more emailing activities happening than other months in Lay's network, e.g., August 2001 ("2001-08"):

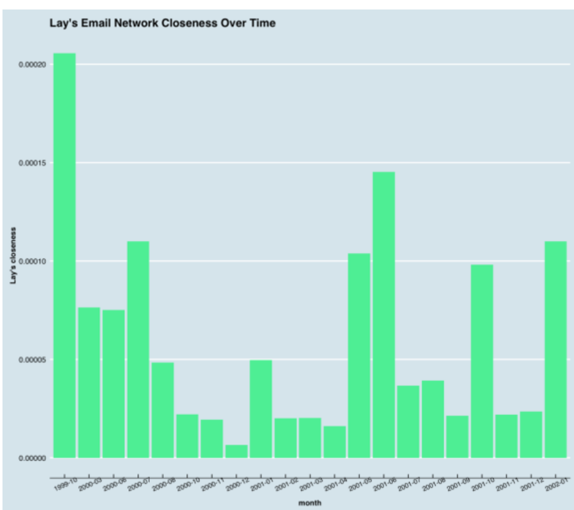


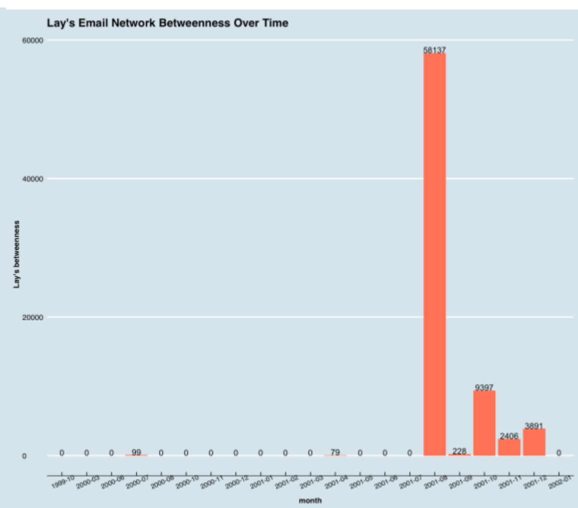
Observing Lay's network at particular points of time, a person we may notice that have "emerged" in some way is Richard Shapiro – it turned out that he was the senior vice president of the Enron Corporation and more intriguingly, "chief Enron lobbyist" during those most scandalous times of the Enron collapse. Indicated by an online article "Bernanke lobbyist authored Enron/Cheney energy plan", Richard Shapiro was highly likely an insider as well in Enron top management's collusion, and very possibly, may even have involved in some "secretive negotiations" between Enron and the Vice President of the Energy Task Force(created by then-U.S. President George W. Bush), Dick Cheney, with Lay and other Enron senior management during that sensitive time (Kevin Connor, *Eyes on the Ties*). Notice that from the visualizations above of Lay's network, it seems that Richard Shapiro established contact with Lay since a very early point of time, and they kept email communication later from time to time. This intriguing small discovery might have been a preceding indication of some "secret plot" being hatched among part of the Enron's senior management circle involving Lay. Further, would that even implicate certain possible "hidden coalition" of interest between the energy giant, Enron, and the political forces of the Bush government then? We cannot say for sure, but this might be a clue that worth more careful probe.

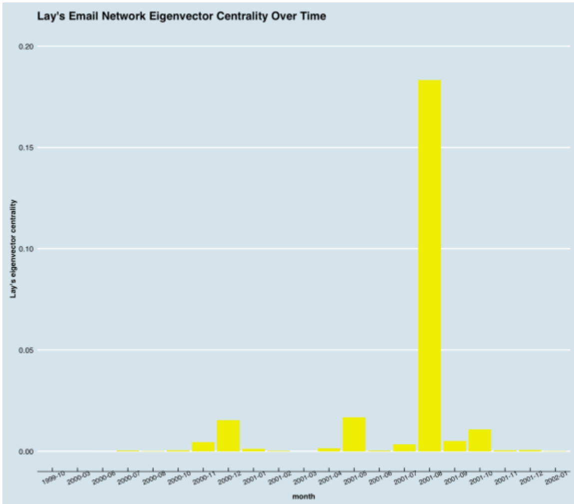
Then we plot the 6 main descriptive centrality measures of Lay's network over time:

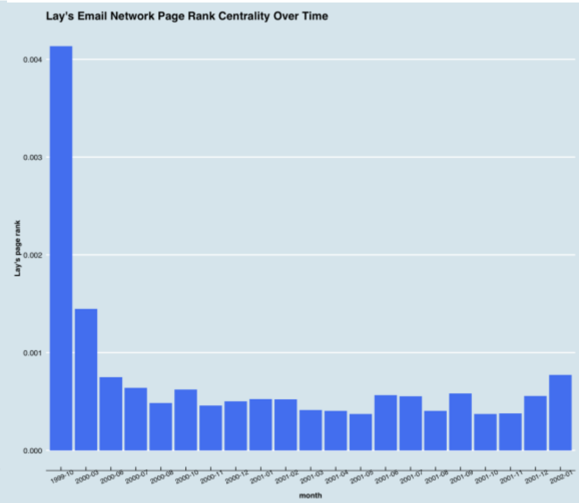












Interestingly, while the in-degree of Lay's network exceeded those of the other months exactly during the exposure of bad news (like "2001-10" and 2001-11"), possibly due to the in-coming queries about what was going on, we observe that several of the centrality measures actually presented some unusual patterns a few months before the actual exposure: for example, the out-degree, betweenness, and eigenvector centrality all achieved their peak during the month of August 2001("2001-08"), which gave astonishingly higher values than any of the other months. It would lead us to some thoughts that maybe these "uncommon" behavioral patterns before the actual occurrence of the event could serve as certain "warning signs" of the organization's internal complication, possibly calling for more attention of external supervision.

Finally, we also pulled out all the edge lists of Lay's email network for each month, to see which contact(s) might have had the most frequent conversation with Kenneth Lay. We look at the snapshots at "2000-07", "2001-09"(one month before exposure), "2001-11" & "2001-12"(during the Crisis and bankruptcy), with number of emails between pairs of sender and receiver sorted in descending order (***see Appendix for Tables***). Interestingly again, several people that might have played certain "special" role in this case of scandal might be spotted here, some of which echoed what we have just discussed once more and others pointing out more directions of the "story" – Richard Shapiro, Steven J. Kean(Lay's Chief of Staff), Greg Whalley(President, head of Trading Operation), Stanley Horton(President, CEO of Enron Transportation Services subsidiary), Louise Kitchen(President, head of Energy Trading division), Liz Taylor, etc.. These people might worth further elaborate investigation in future studies.

5. Conclusion and Implication

The comparison of network-level measures across time revealed that after the crisis point, people within Enron have evidently changed their behavior. People have changed from email to other forms of communication (e.g. face-to-face); they stopped forming groups and wanted to keep a lower profile instead; they did not want their email recipients to see each other and started sending out massive amount of bcc emails.

On the other hand, ERGM shows us similar phenomenon that during the transition time of crisis, trading/legal apartment, male/female, senior/junior presented different behavior pre- and post-crisis. People tend to email to high-level person and tend to share emails within legal team rather than sending it outside legal team.

Further examination of the particular part of the network involving the 149 people in the special list evolving over time also revealed that on average, three of the centrality measures -- in-degree, out-degree, and betweenness centrality reflect a similar trend of higher activeness and stronger brokerage role, imposing a possibly intense control of information flow, of these people in the special list around the occurrence of crisis; while from another perspective, the other three of the centrality measures -- closeness, eigenvector centrality and page rank centrality present an opposite pattern of change of statistics, indicating less intimacy (or further distance) and weaker influence of the people in the list during the interim period leading up to the crisis. While those special actors of “Employee” type of status seemed to remain active, closer to others and more influential especially in “normal” times of work, people of particular statuses such as “Vice President” and “President” tend to stand out more in terms of brokerage role and influence when the Enron Crisis broke out.

The dynamics of the network of an individual actor of importance is also examined. By visualizing and examining the network patterns and ties of one of the “key culprits”--Kenneth Lay, we might spot some particular actors in the upper-level management suspicious as part of “origin of misdeeds”, which may be worth further investigation, such as Enron’s “chief lobbyist” Richard Shapiro. Some unusual patterns of high out-degree, betweenness and eigenvector centrality of Lay’s email network several months preceding the actual exposure of the scandal might also be noteworthy indications of the company’s internal complication.

We conclude by identifying the need for further investigation on (a) analyzing the content of emails, (b) including more background information of employees in the ERGM model, and (c) examining some of the other key factors such as Enron’s “chief lobbyist” Richard Shapiro.

6. Reference

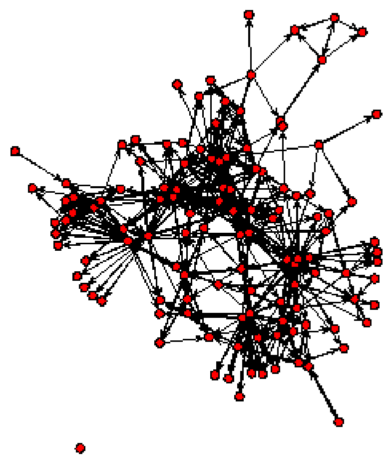
[1] Ahschulz.de. (2019). Enron Data. Available at: http://www.ahschulz.de/enron-email-data/.
[2] Cs.cmu.edu. (2019). Enron Email Dataset. Available at: https://www.cs.cmu.edu/~enron/.
[3] Chicagotribune.com. (2019). Chicago Tribune - We are currently unavailable in your region. [online] Available at: https://www.chicagotribune.com/news/ct-xpm-2002-01-24-0201240281-story.html.

[4] Nytimes.com. (2019). Timeline: A chronology of Enron Corp. Available at: https://www.nytimes.com/2006/01/18/business/worldbusiness/timeline-a-chronology-of-enron-corp.html.
[5] Journal of Accountancy. (2019). The Rise and Fall of Enron. Available at: https://www.journalofaccountancy.com/issues/2002/apr/theriseandfallofenron.html.
[6] Microsoft 365 Blog. (2019). 5 Tips on using Bcc in Outlook Email - Microsoft 365 Blog. [online] Available at: https://www.microsoft.com/en-us/microsoft-365/blog/2012/03/23/5-tips-on-using-bcc-in-outlook-email/.

7. Appendix

Appendix I - Ergm Result

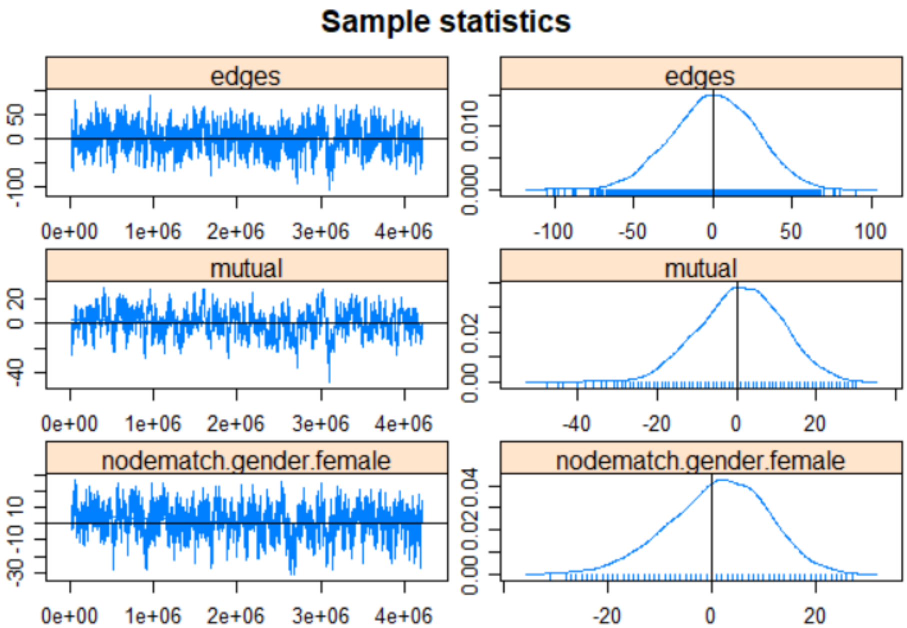
2000Q3

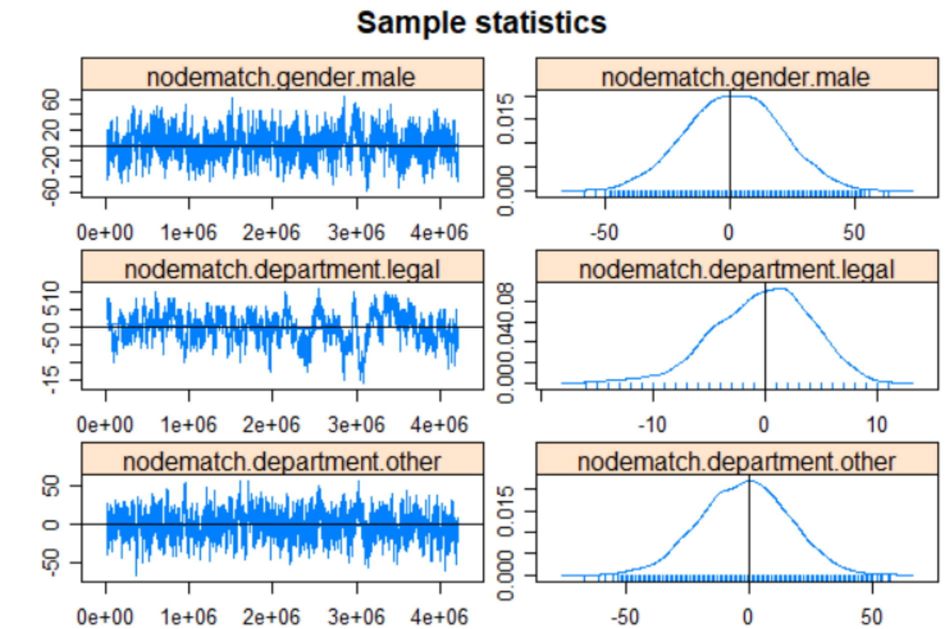


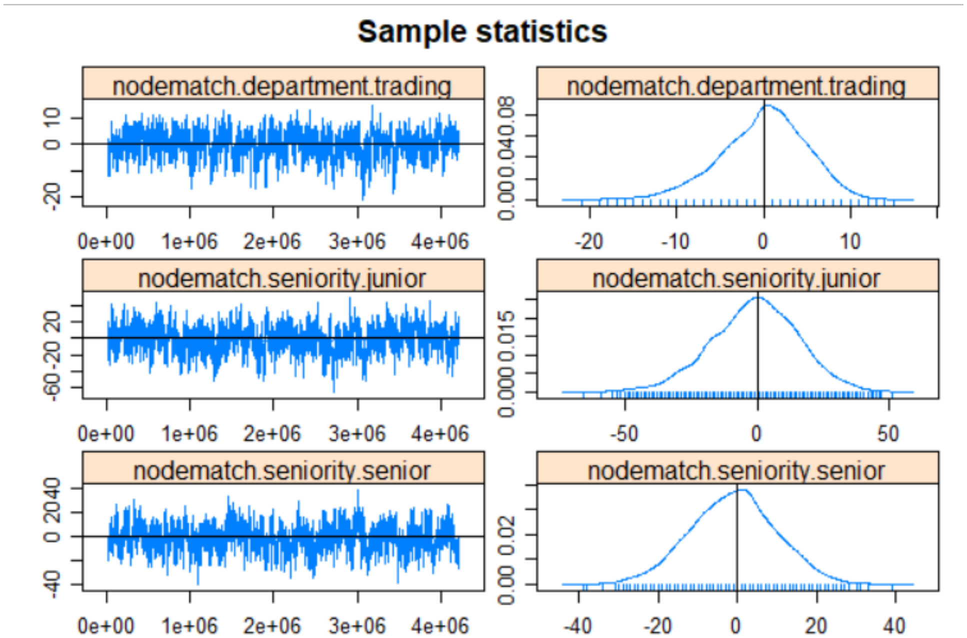
2000Q3 Nodematch

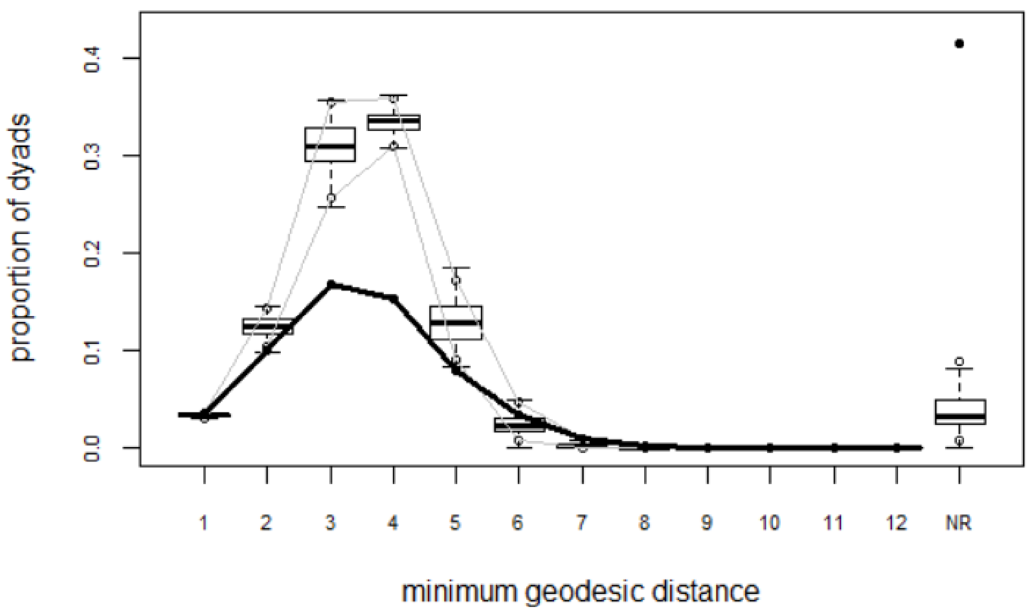
```
==========================
Summary of model fit
==========================

Formula:   net ~ edges + mutual + nodematch("gender", diff = T) +
    nodematch("department", diff = T) + nodematch("seniority",
    diff = T)

Iterations:  2 out of 20

Monte Carlo MLE Results:
                           Estimate Std. Error MCMC % z value Pr(>|z|)
edges                      -3.96178    0.10403      0 -38.082  < 1e-04 ***
mutual                      3.57957    0.15729      0  22.757  < 1e-04 ***
nodematch.gender.female     0.44332    0.12492      0   3.549 0.000387 ***
nodematch.gender.male       0.08290    0.08591      0   0.965 0.334576
nodematch.department.legal  1.10574    0.25467      0   4.342  < 1e-04 ***
nodematch.department.other -0.06540    0.08348      0  -0.783 0.433402
```

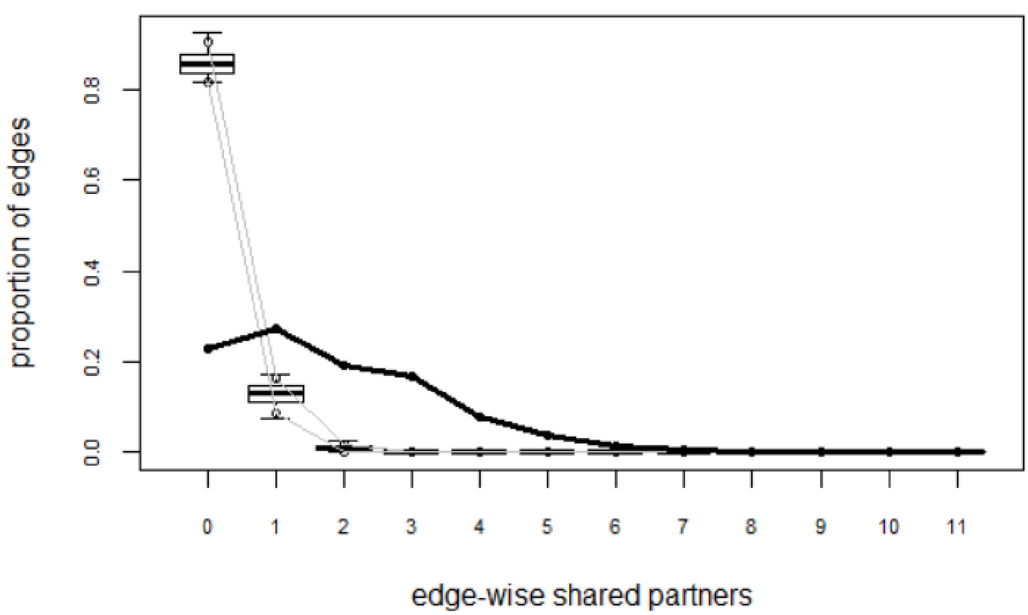
```
nodematch.department.trading  0.12651    0.21648      0   0.584 0.558960
nodematch.seniority.junior   -0.05114    0.08561      0  -0.597 0.550236
nodematch.seniority.senior    0.18405    0.11320      0   1.626 0.103981
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 19796  on 14280  degrees of freedom
 Residual Deviance:  3770  on 14271  degrees of freedom

AIC: 3788    BIC: 3856    (Smaller is better.)
```

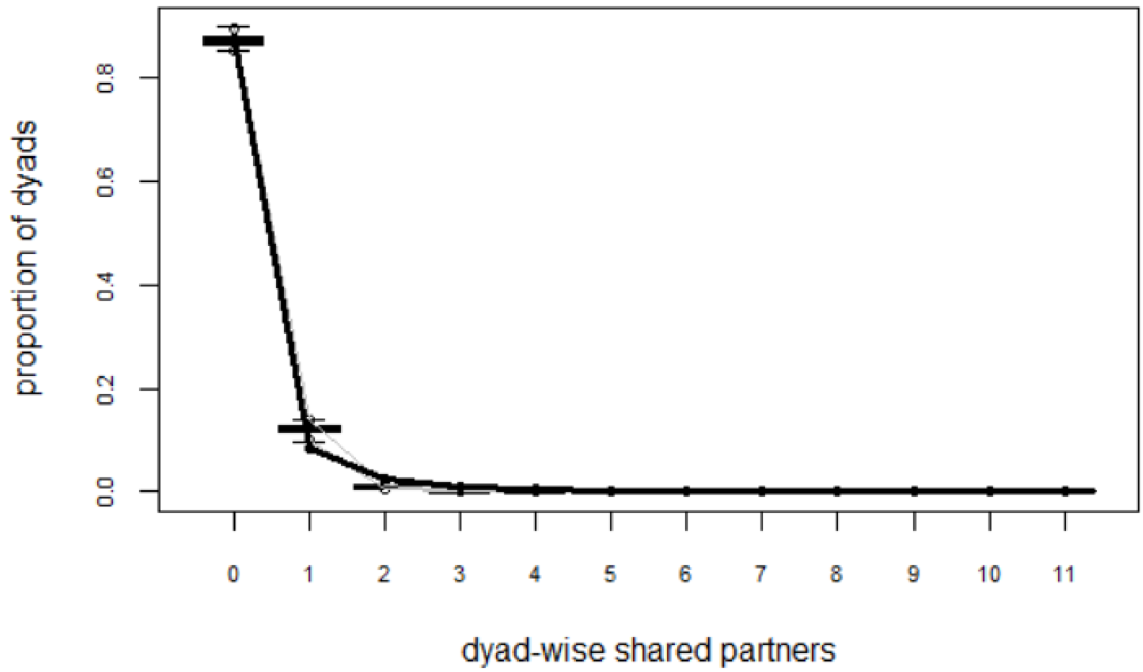



Sample statistics
nodematch.gender.male
nodematch.gender.male
nodematch.department.legal
nodematch.department.legal
nodematch.department.other
nodematch.department.other
0e+00
1e+06
2e+06
3e+06
4e+06

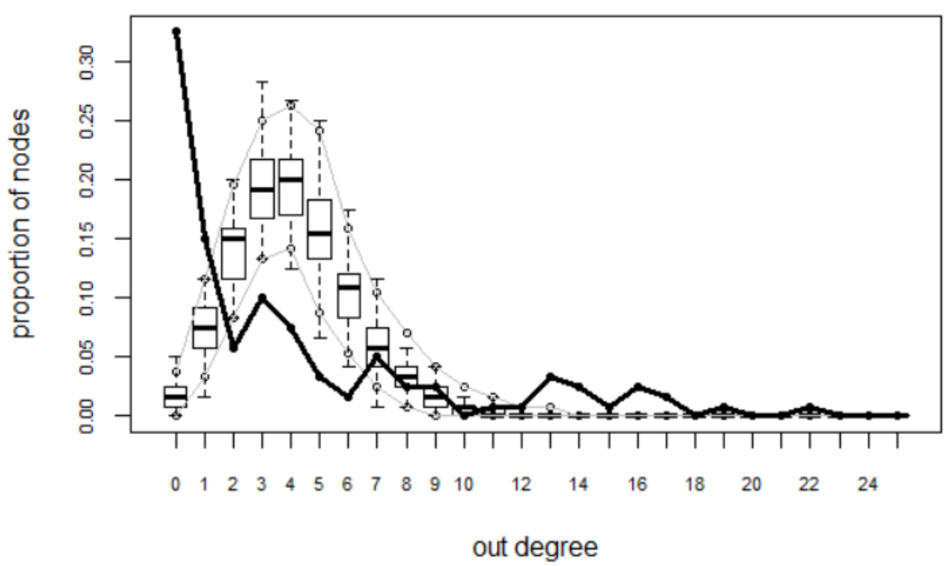

Sample statistics
nodematch.department.trading
nodematch.department.trading
nodematch.seniority.junior
nodematch.seniority.junior
nodematch.seniority.senior
nodematch.seniority.senior
0e+00
1e+06
2e+06
3e+06
4e+06


proportion of dyads
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
8
9
10
11
12
NR
minimum geodesic distance

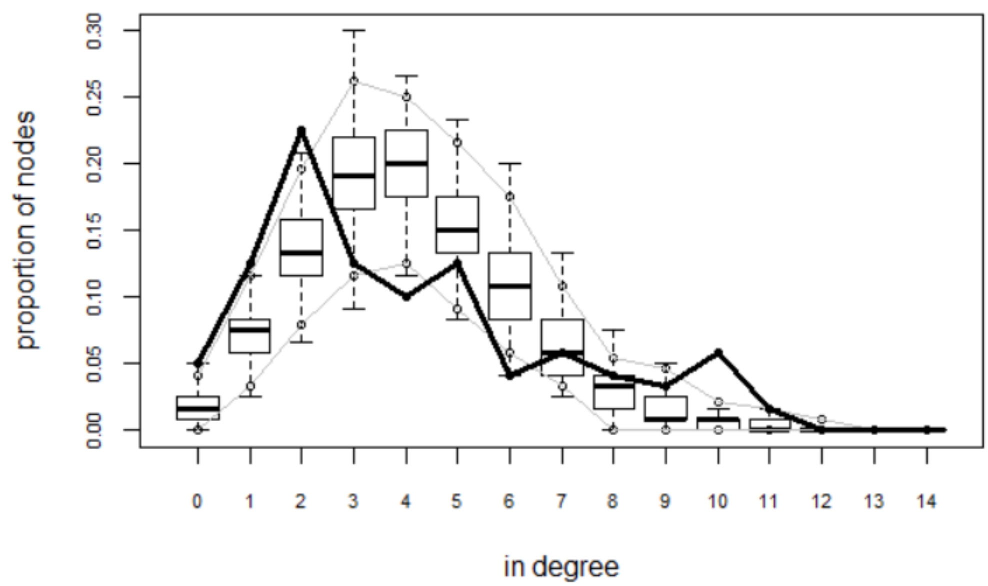

proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
edge-wise shared partners



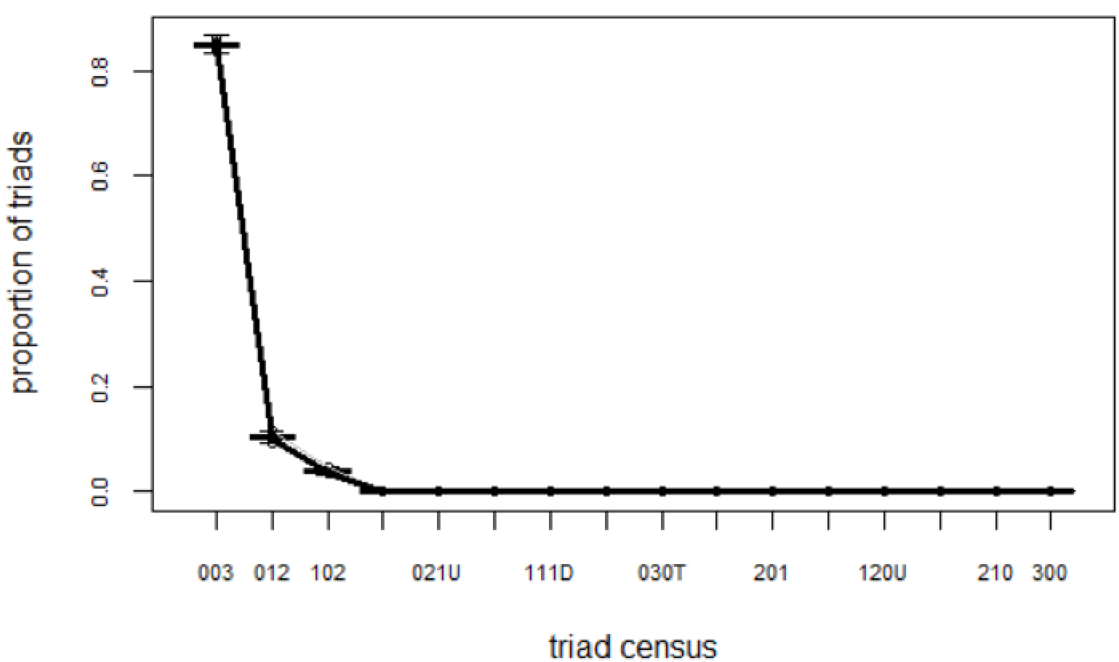
proportion of dyads
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
dyad-wise shared partners



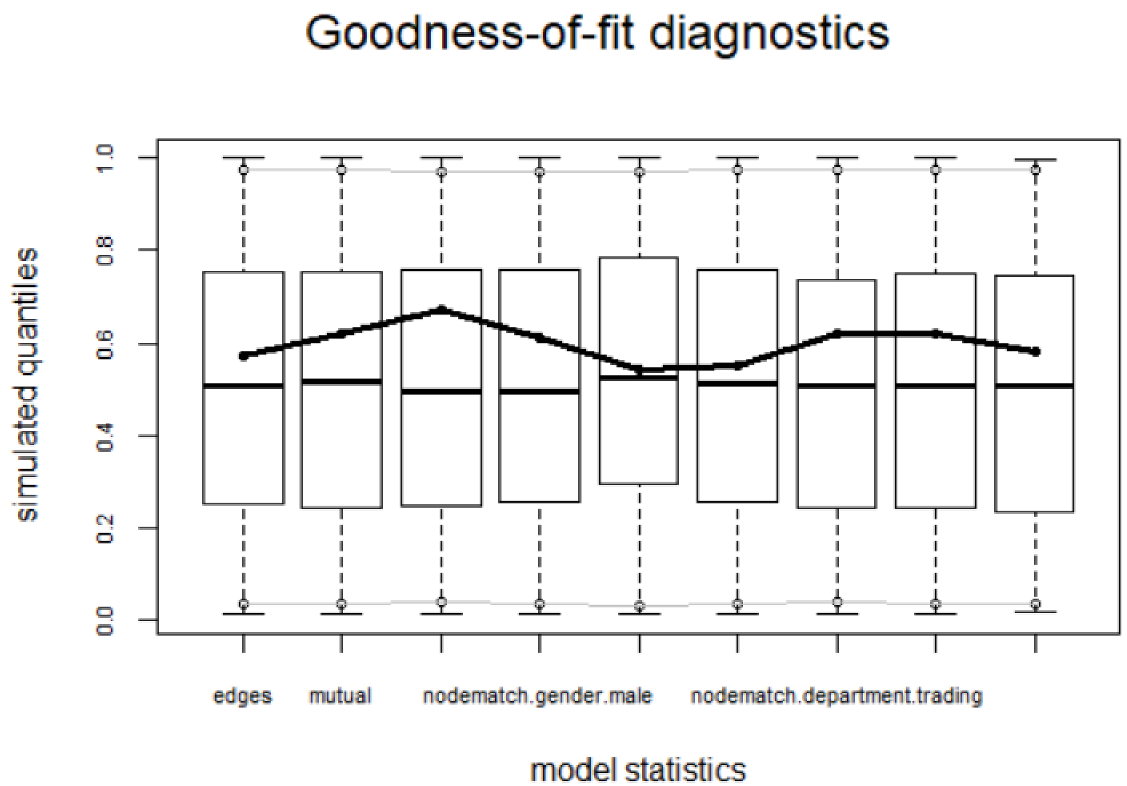
proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
out degree



proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
in degree



proportion of triads
0.0
0.2
0.4
0.6
0.8
003
012
102
021U
111D
030T
201
120U
210
300
triad census

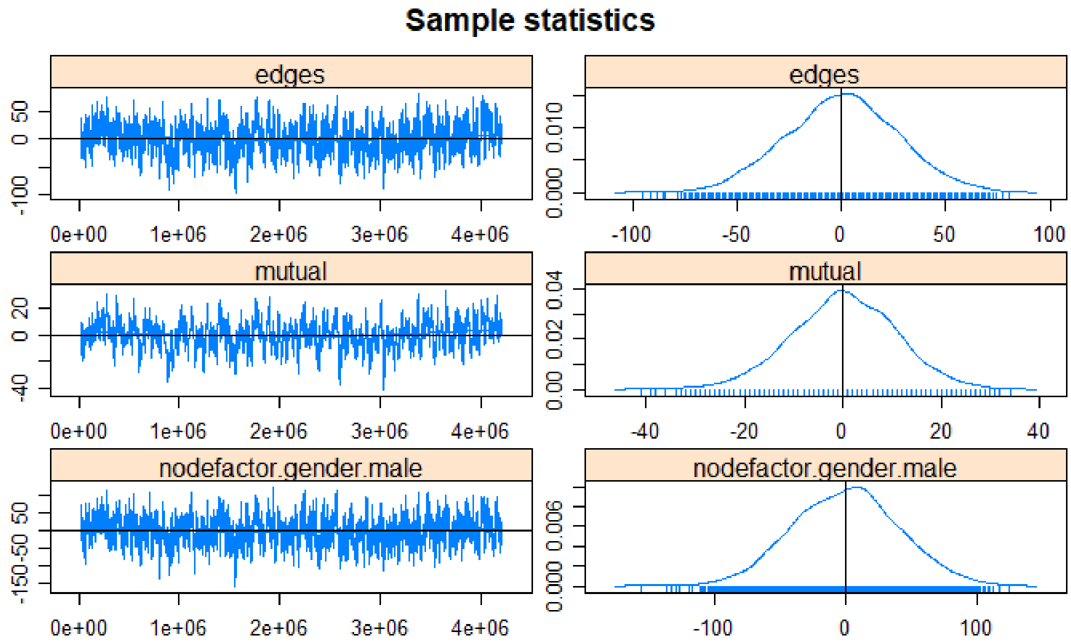


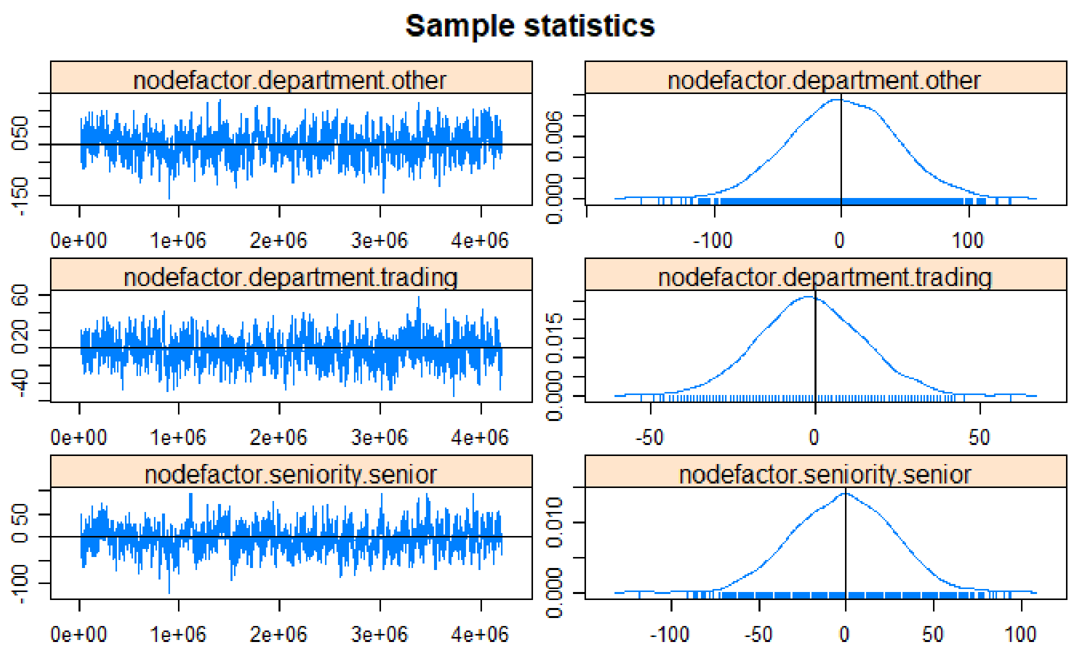
2000Q3 Nodefactor

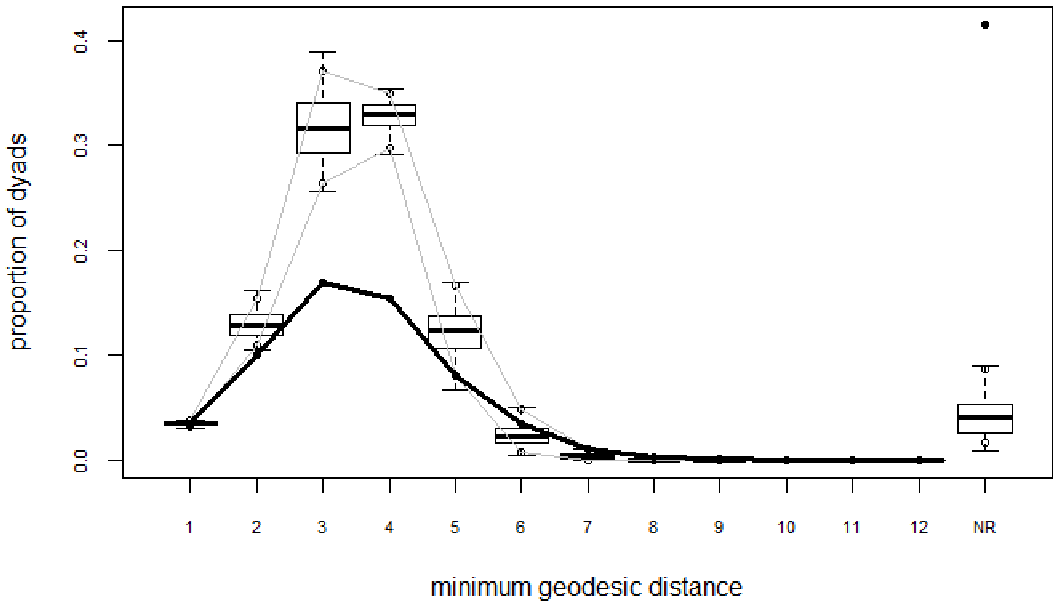
```
==========================
Summary of model fit
==========================

Formula:   net ~ edges + mutual + nodefactor("gender") + nodefactor("department") +
    nodefactor("seniority")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                            Estimate Std. Error MCMC % z value Pr(>|z|)
edges                       -3.11907    0.16606      0 -18.783   <1e-04 ***
mutual                       3.56815    0.15633      0  22.824   <1e-04 ***
nodefactor.gender.male      -0.03894    0.06483      0  -0.601   0.5481
nodefactor.department.other -0.41829    0.09123      0  -4.585   <1e-04 ***
nodefactor.department.trading -0.50205  0.11361      0  -4.419   <1e-04 ***
nodefactor.seniority.senior  0.09960    0.05841      0   1.705   0.0881 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Null Deviance: 19796 on 14280 degrees of freedom
Residual Deviance: 3775 on 14274 degrees of freedom

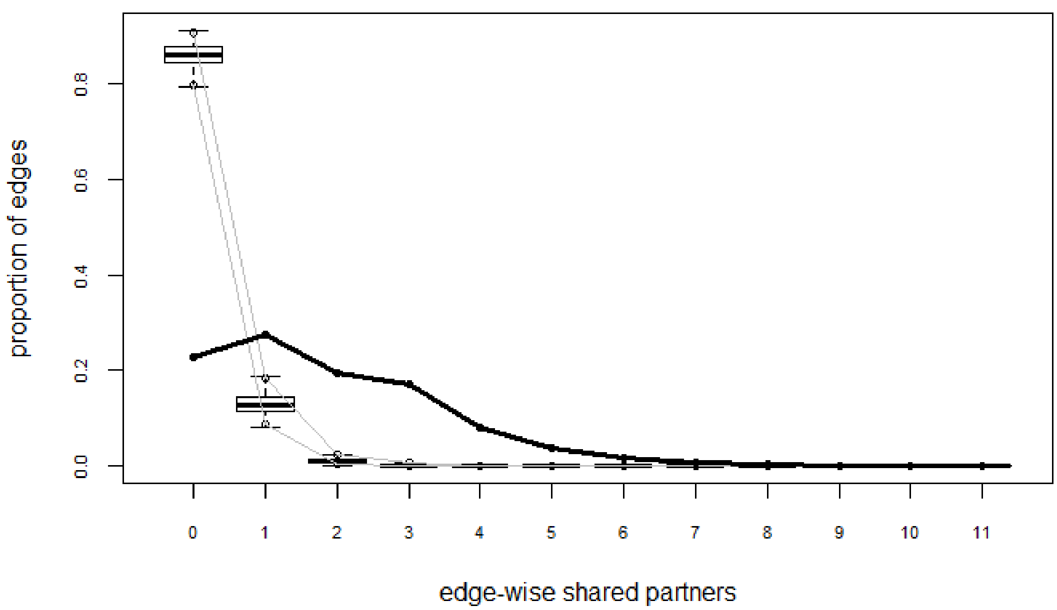
AIC: 3787 BIC: 3832 (Smaller is better.)



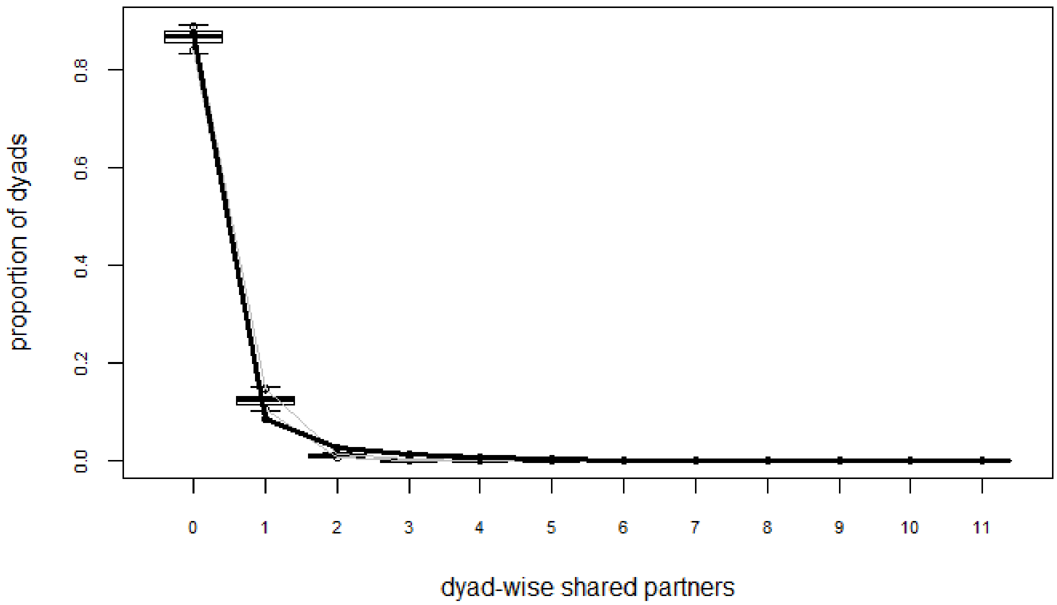


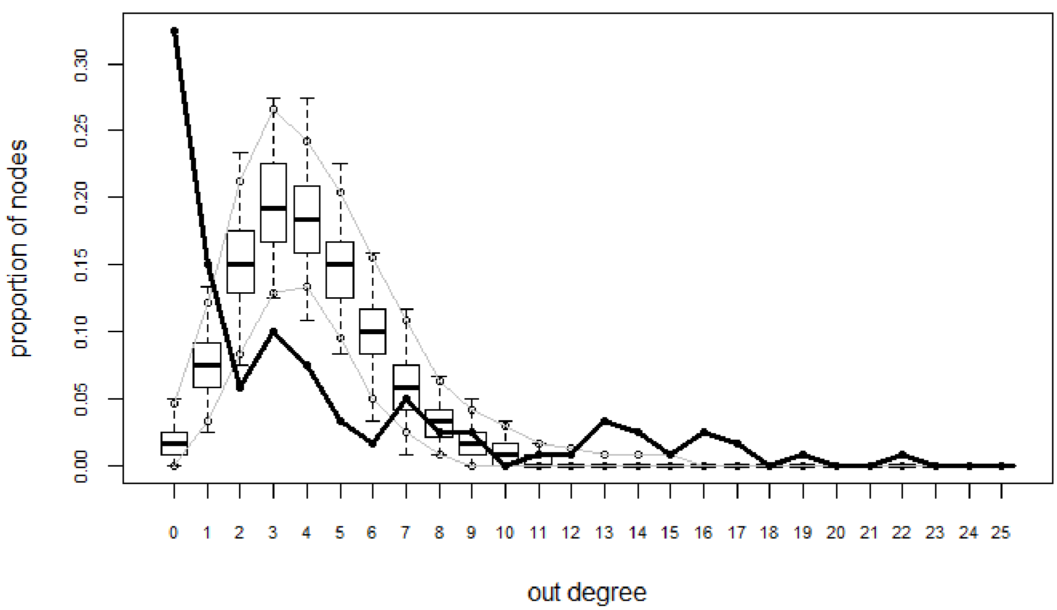


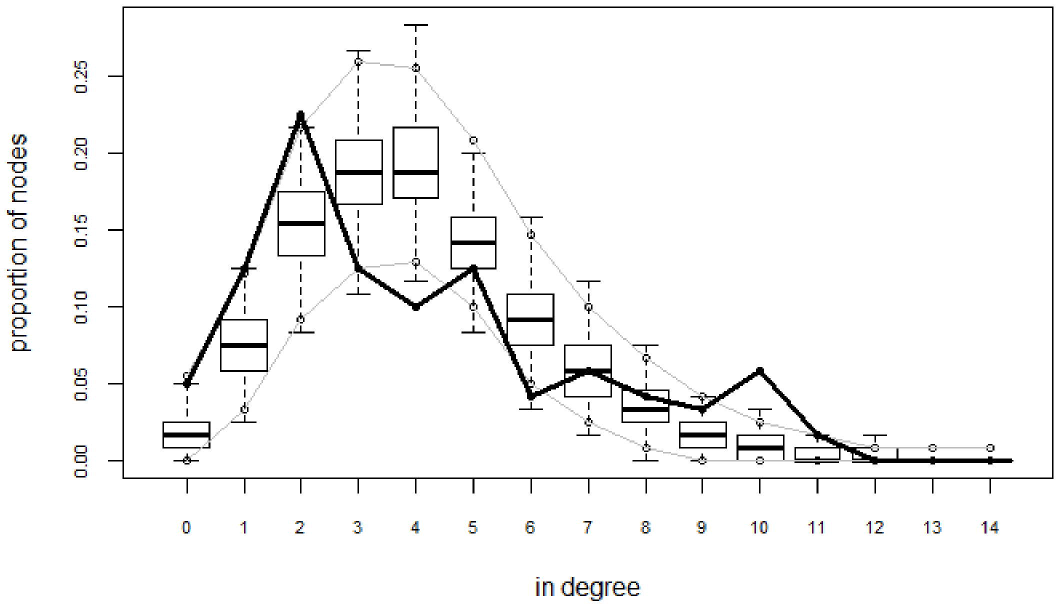
proportion of dyads
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
8
9
10
11
12
NR
minimum geodesic distance

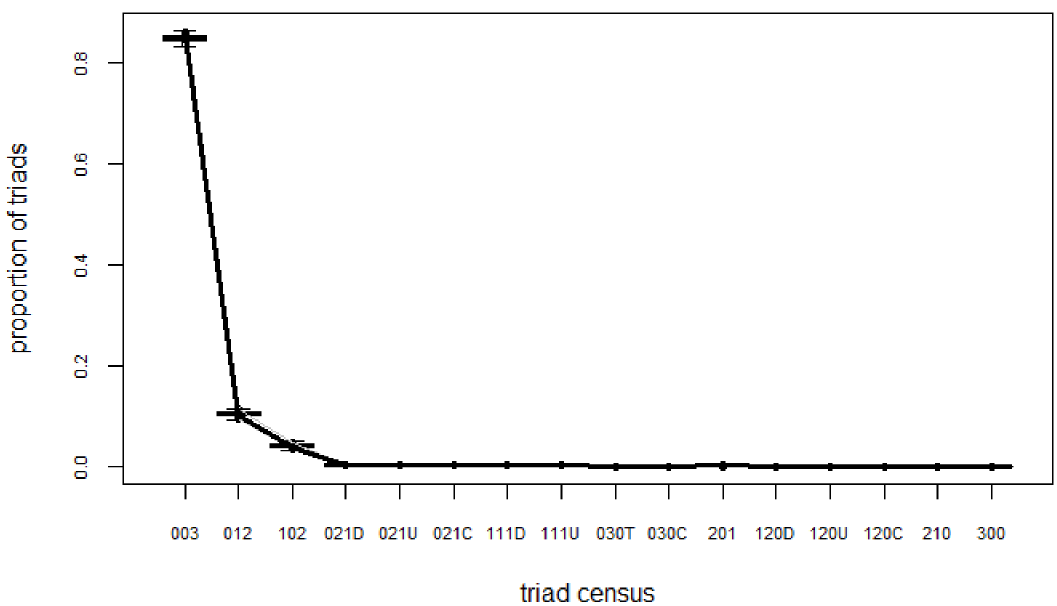


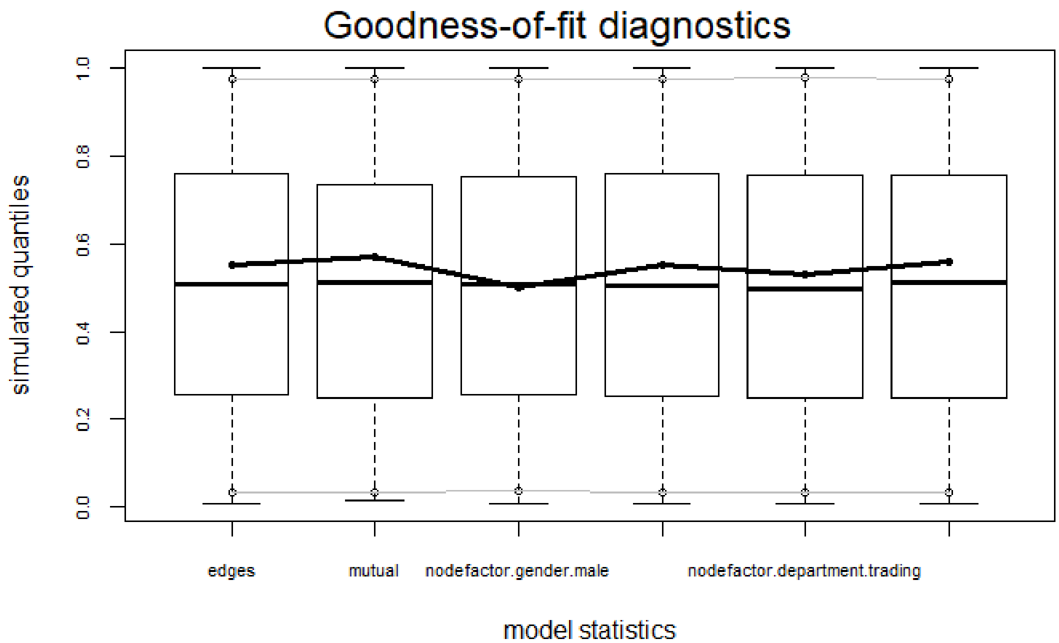
proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
edge-wise shared partners


proportion of dyads
0.0
0.2
0.4
0.6
0.8
0 1 2 3 4 5 6 7 8 9 10 11
dyad-wise shared partners

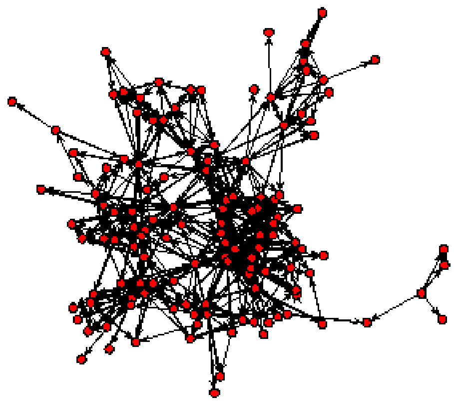

proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
out degree


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
in degree


proportion of triads
0.0
0.2
0.4
0.6
0.8
003 012 102 021D 021U 021C 111D 111U 030T 030C 201 120D 120U 120C 210 300
triad census


Goodness-of-fit diagnostics
simulated quantiles
0.0
0.2
0.4
0.6
0.8
1.0
edges
mutual
nodefactor.gender.male
nodefactor.department.trading
model statistics

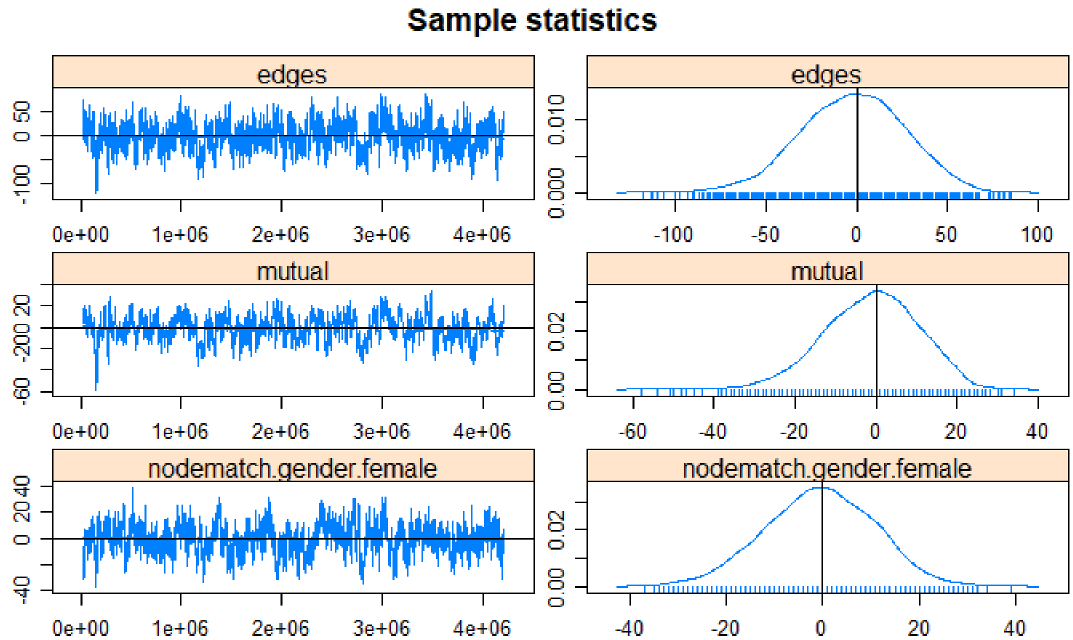
2000Q4



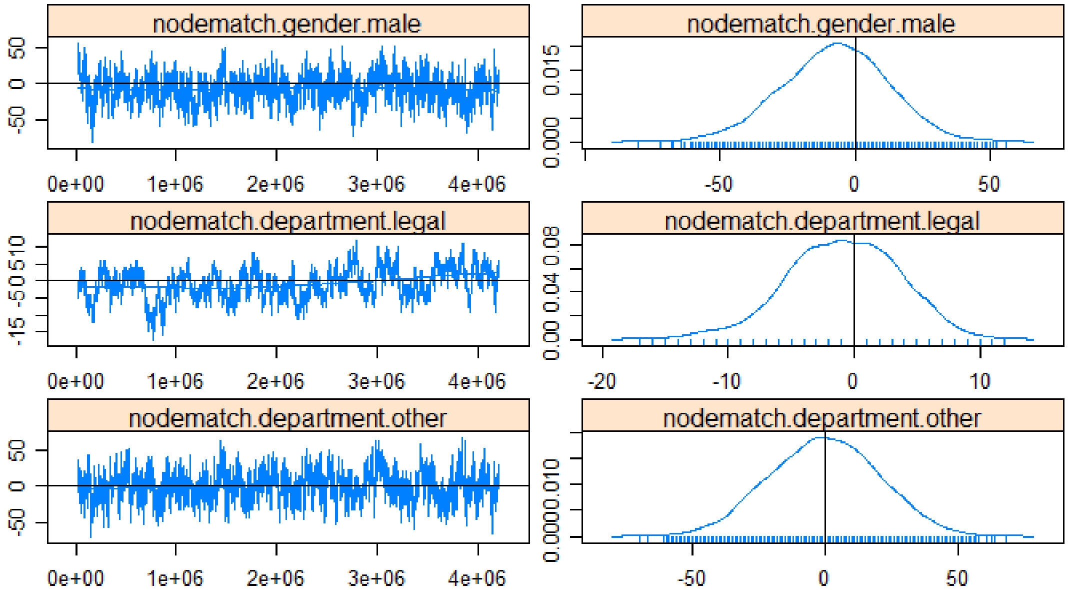
2000Q4 nodematch

```
==========================
Summary of model fit
==========================

Formula:   net ~ edges + mutual + nodematch("gender", diff = T) + 
    nodematch("department", diff = T) + nodematch("seniority", 
    diff = T)

Iterations:  2 out of 20 

Monte Carlo MLE Results:
                          Estimate Std. Error MCMC % z value Pr(>|z|)    
edges                     -4.09634    0.09054      0 -45.242  < 1e-04 ***
mutual                     4.01482    0.14537      0  27.617  < 1e-04 ***
nodematch.gender.female    0.43374    0.10676      0   4.063  < 1e-04 ***
nodematch.gender.male      0.05118    0.08124      0   0.630  0.52872    
nodematch.department.legal    1.04517    0.23558      1   4.437  < 1e-04 ***
nodematch.department.other   -0.04060    0.07451      0  -0.545  0.58584    
nodematch.department.trading  0.40465    0.15041      0   2.690  0.00714 ** 
nodematch.seniority.junior    0.01355    0.08091      0   0.167  0.86703    
nodematch.seniority.senior    0.41552    0.09965      0   4.170  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

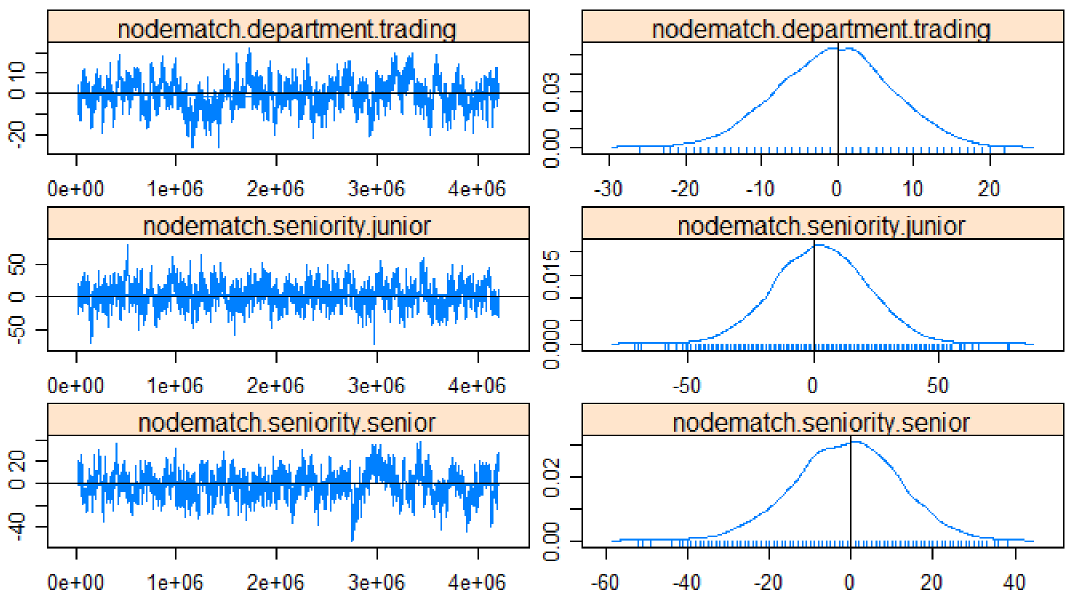
```
   Null Deviance: 22536  on 16256  degrees of freedom
 Residual Deviance:  4458  on 16247  degrees of freedom

AIC: 4476    BIC: 4545    (Smaller is better.)
```

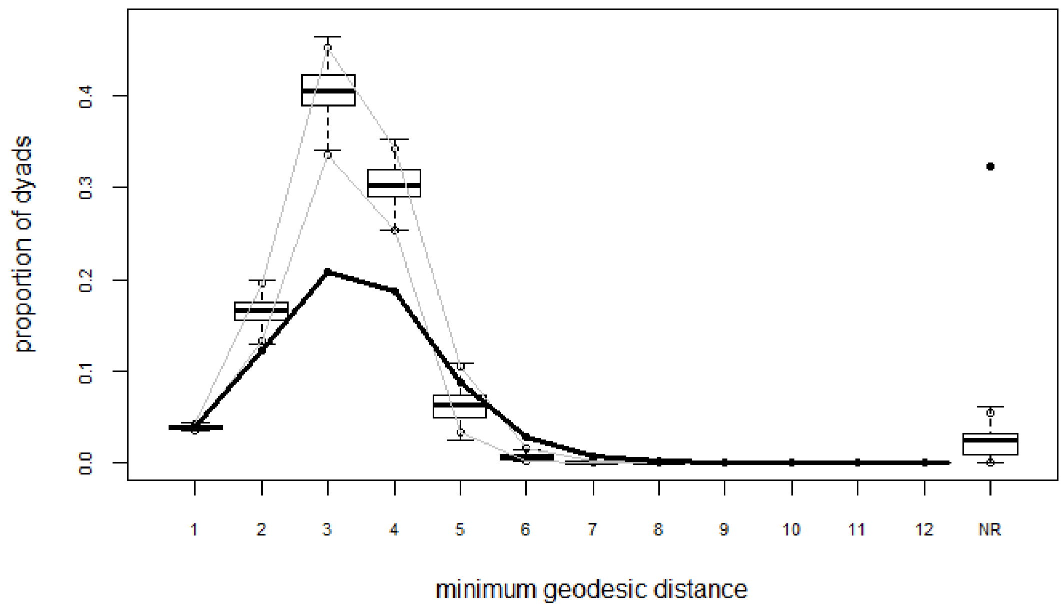


Sample statistics

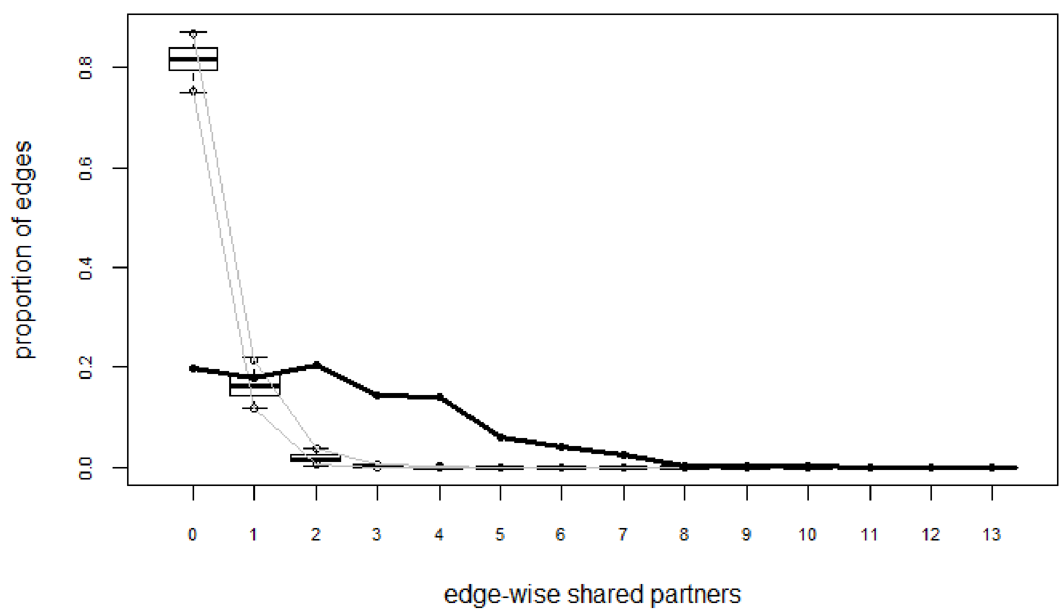


Sample statistics

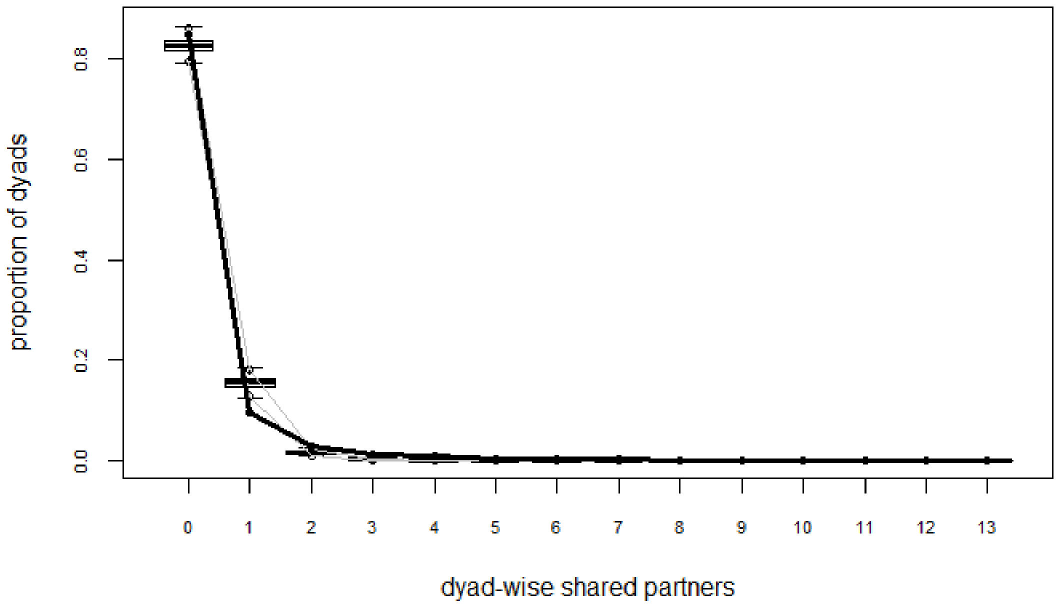


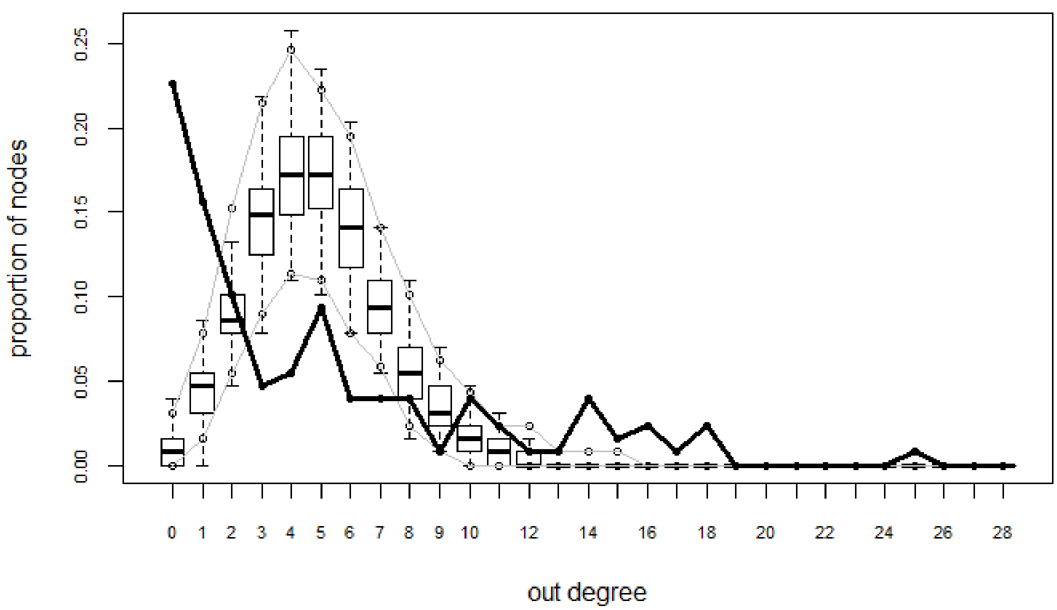


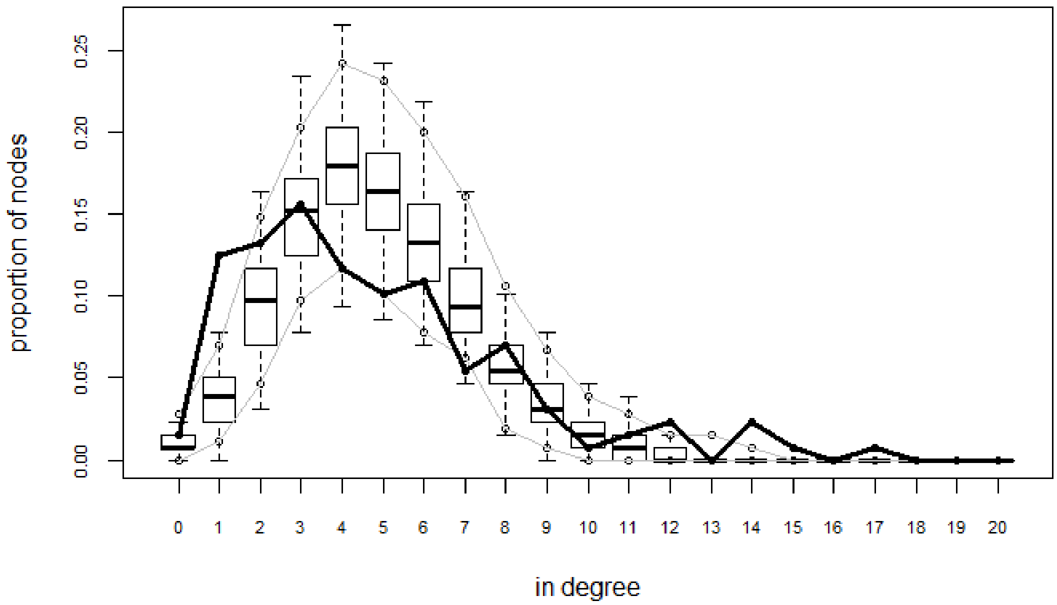
proportion of dyads
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
8
9
10
11
12
NR
minimum geodesic distance

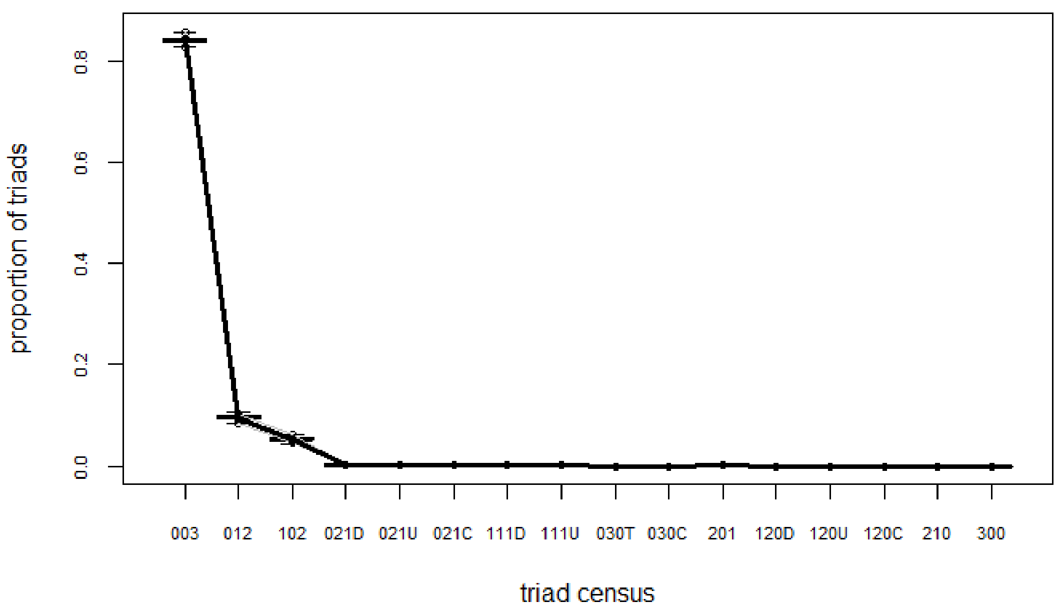


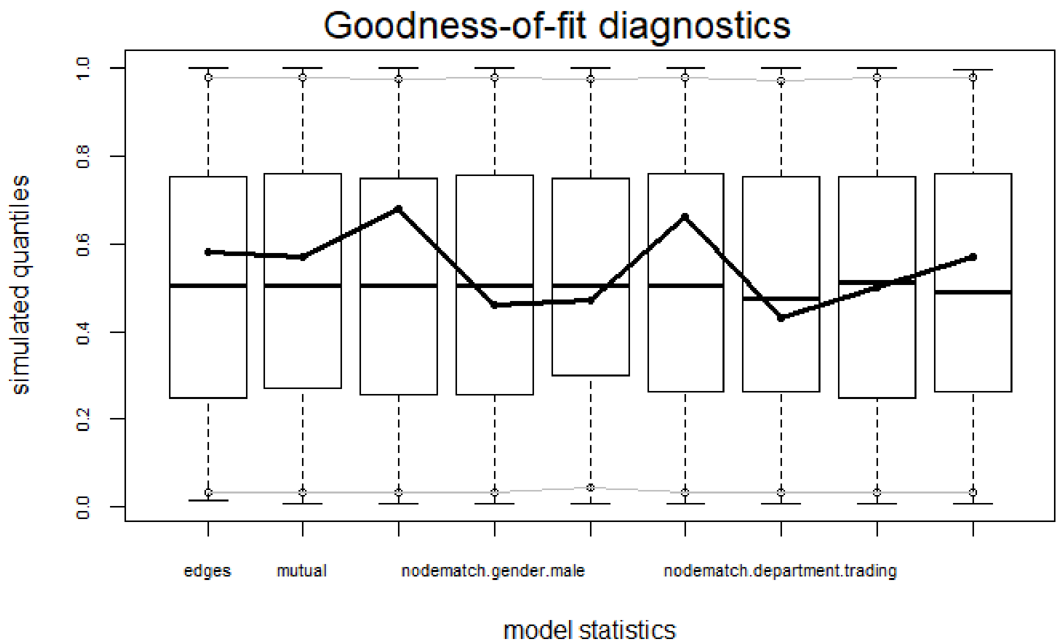
proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
edge-wise shared partners


proportion of dyads
0.0
0.2
0.4
0.6
0.8
0 1 2 3 4 5 6 7 8 9 10 11 12 13
dyad-wise shared partners


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0 1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28
out degree


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
in degree


proportion of triads
0.0
0.2
0.4
0.6
0.8
003 012 102 021D 021U 021C 111D 111U 030T 030C 201 120D 120U 120C 210 300
triad census

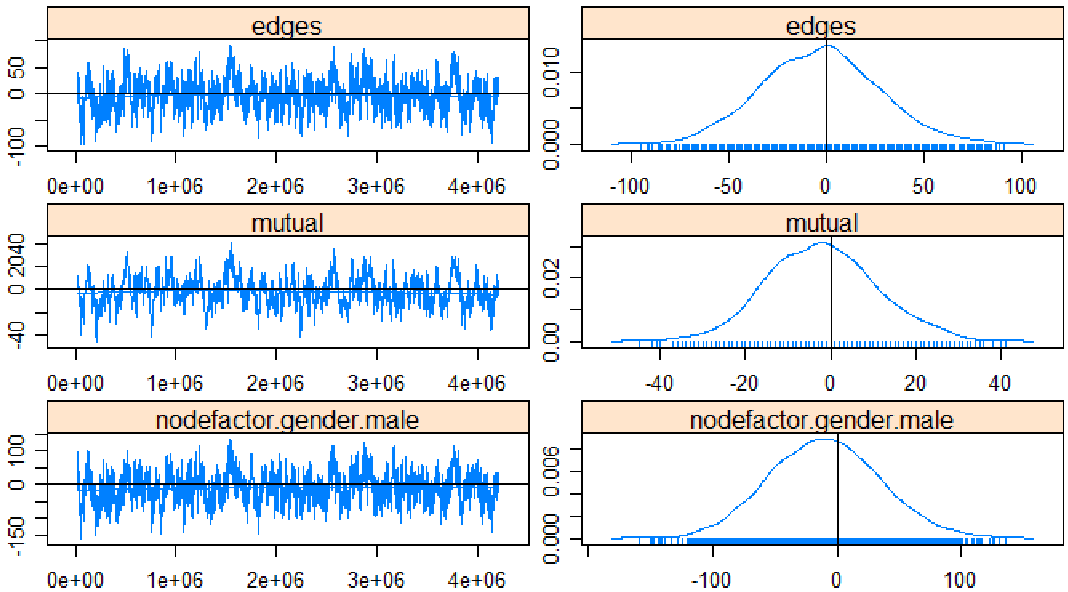


2000Q4 Nodefactor

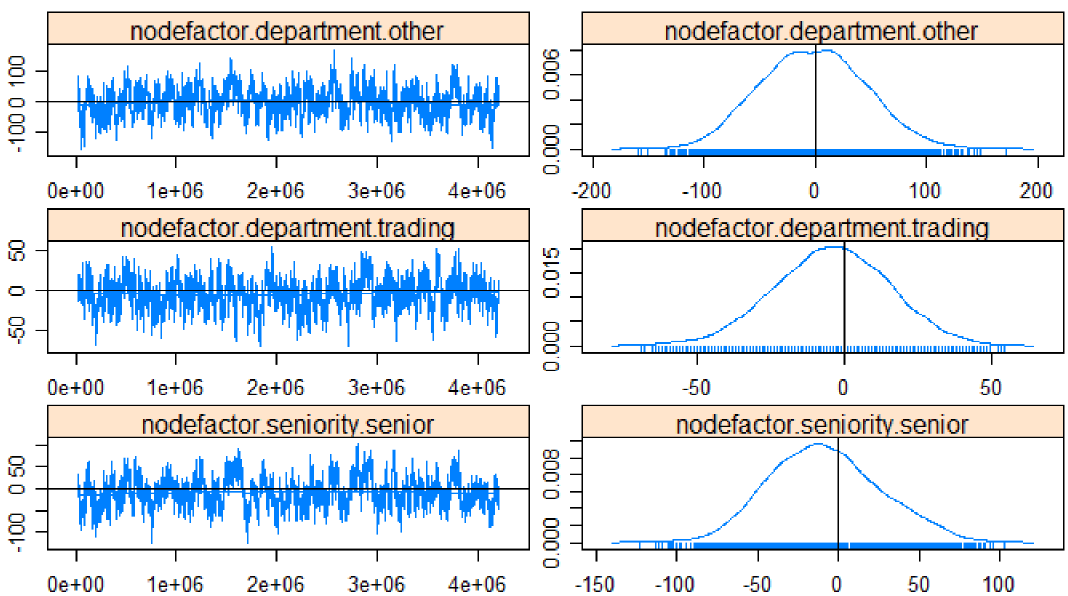
```
==========================
Summary of model fit
==========================

Formula:   net ~ edges + mutual + nodefactor("gender") + nodefactor("department") +
    nodefactor("seniority")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                         -3.28857    0.16053      0 -20.485  < 1e-04 ***
mutual                         4.04471    0.14213      0  28.458  < 1e-04 ***
nodefactor.gender.male        -0.08887    0.05720      0  -1.554 0.120286
nodefactor.department.other   -0.37066    0.08558      0  -4.331  < 1e-04 ***
nodefactor.department.trading -0.35012    0.10149      0  -3.450 0.000561 ***
nodefactor.seniority.senior    0.16487    0.04725      0   3.490 0.000484 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 22536  on 16256  degrees of freedom
 Residual Deviance:  4482  on 16250  degrees of freedom
```

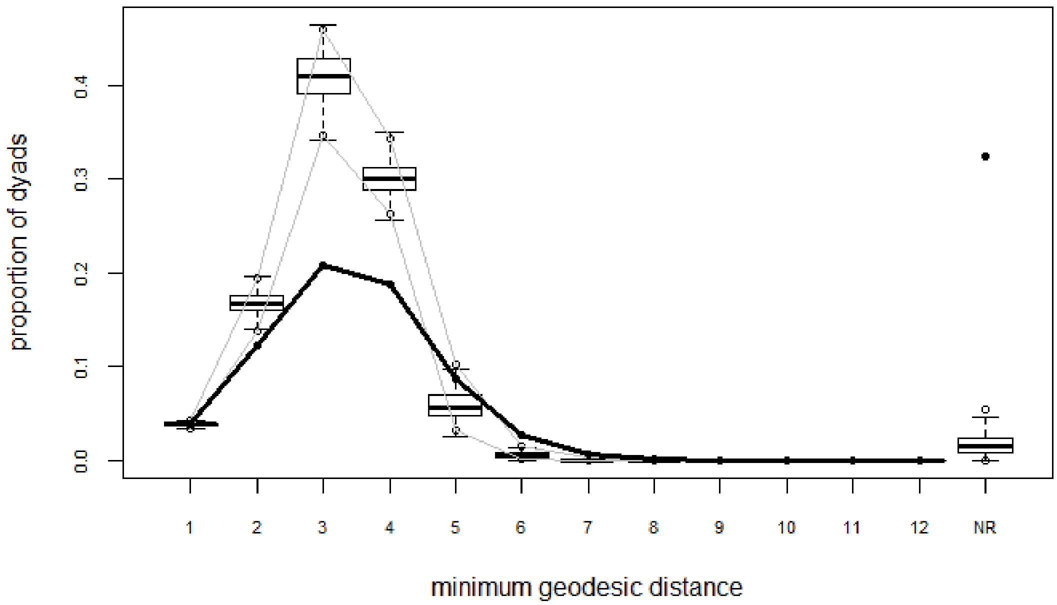
AIC: 4494   BIC: 4540   (Smaller is better.)

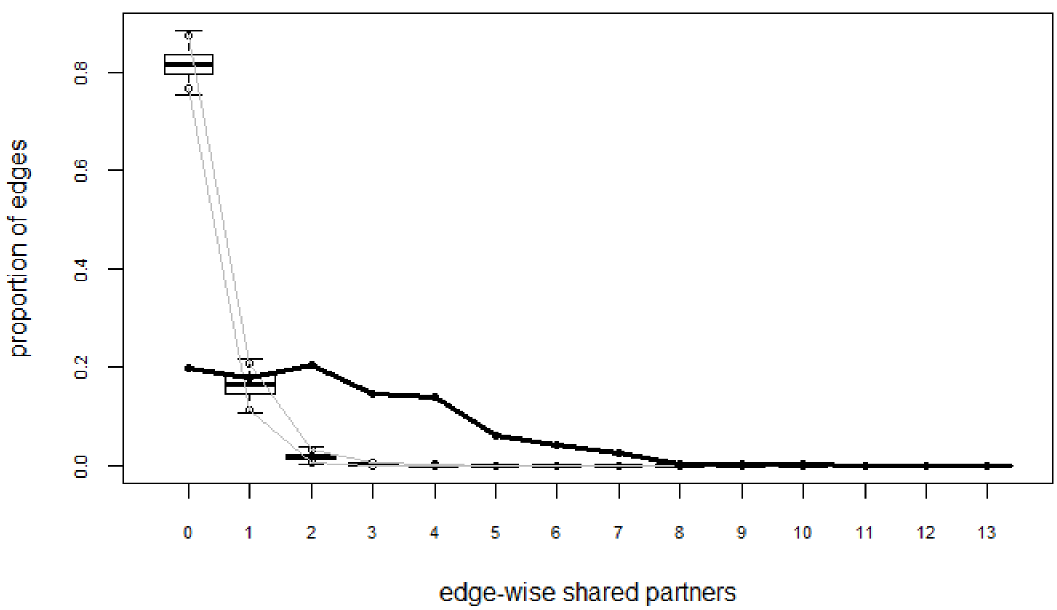
Sample statistics

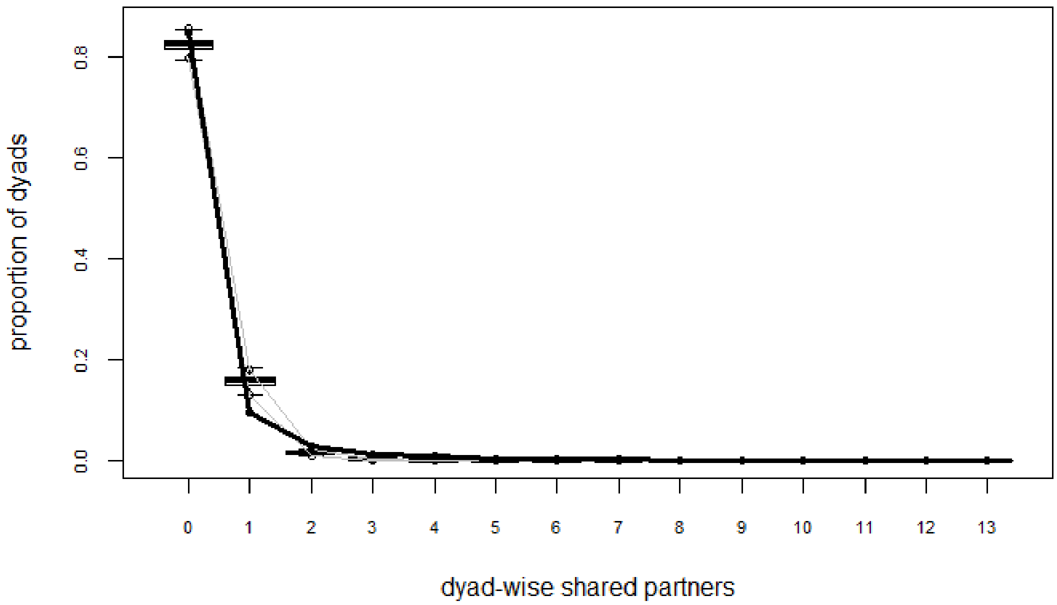


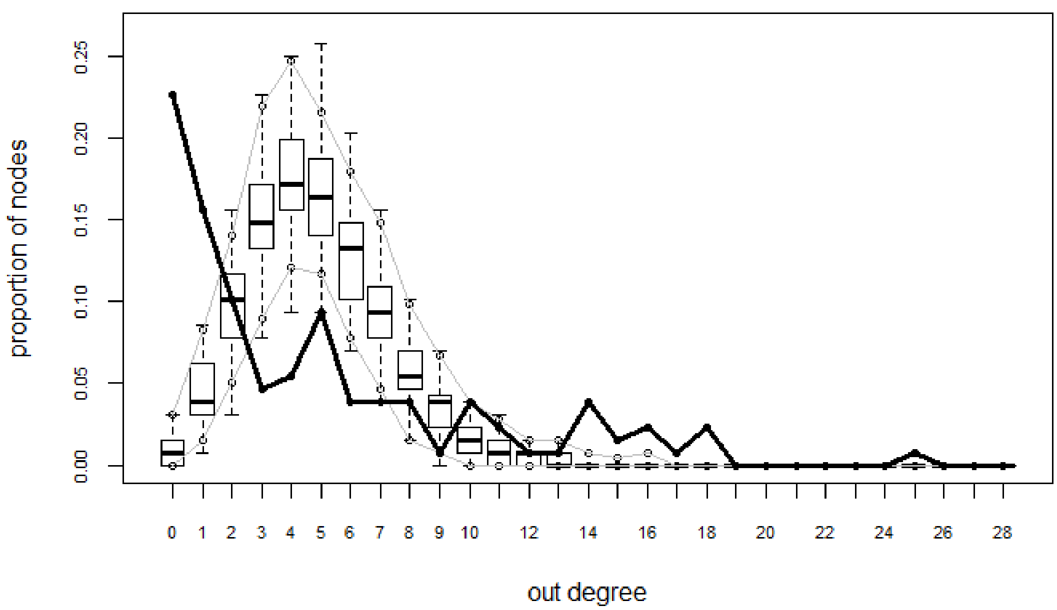
Sample statistics

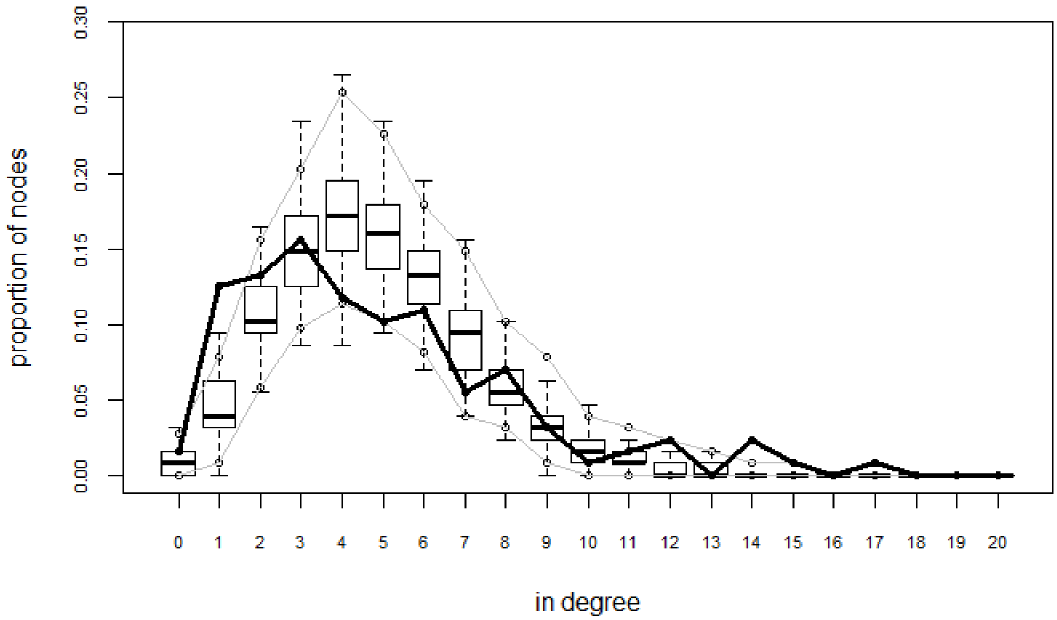



proportion of dyads
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
8
9
10
11
12
NR
minimum geodesic distance

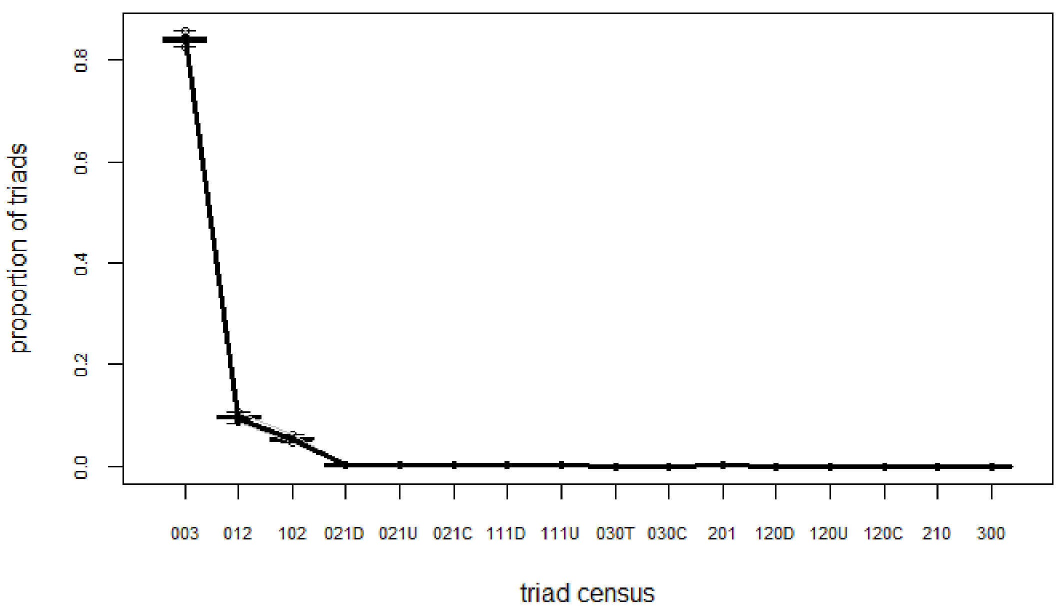

proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
edge-wise shared partners


proportion of dyads
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
dyad-wise shared partners

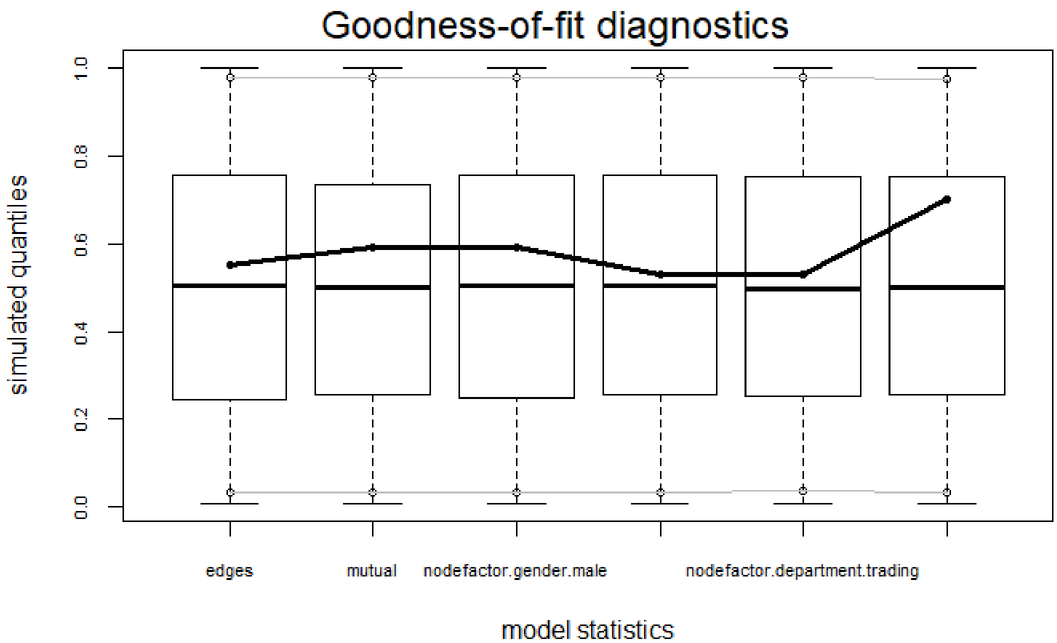

proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
26
28
out degree



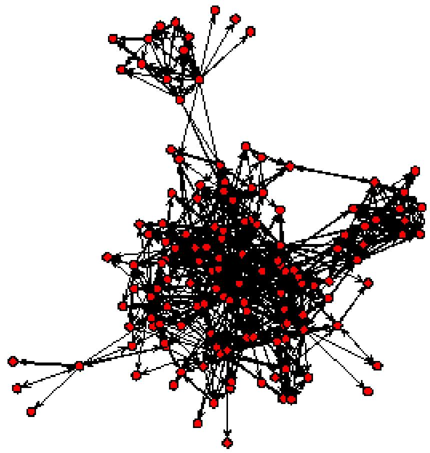
proportion of nodes
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
in degree



proportion of triads
0.0
0.2
0.4
0.6
0.8
003 012 102 021D 021U 021C 111D 111U 030T 030C 201 120D 120U 120C 210 300
triad census


Goodness-of-fit diagnostics
simulated quantiles
1.0
0.8
0.6
0.4
0.2
0.0
edges
mutual
nodefactor.gender.male
nodefactor.department.trading
model statistics

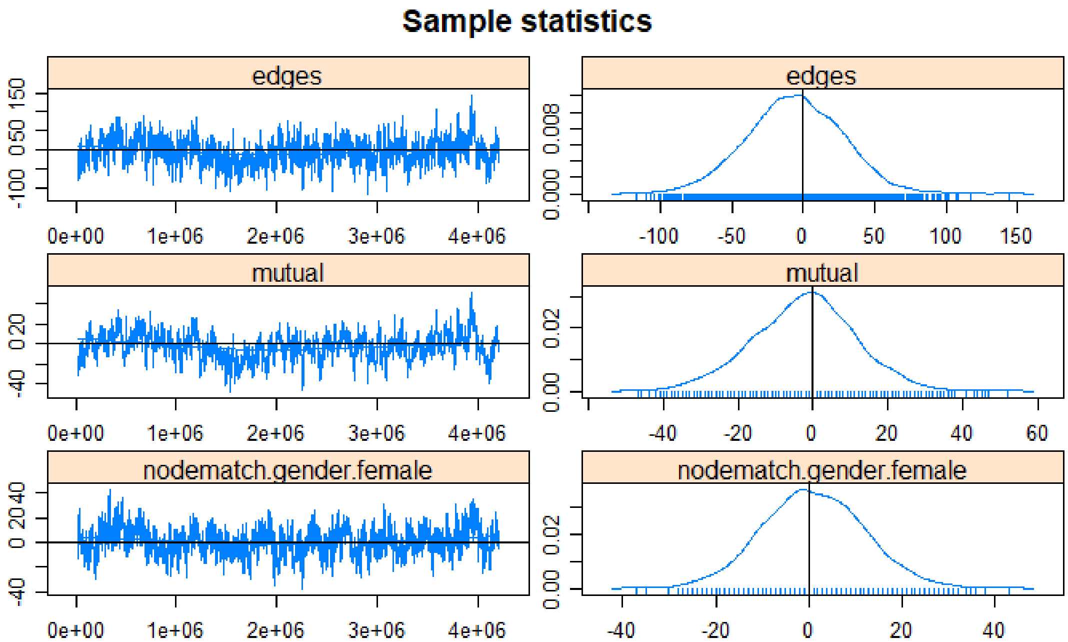
2001Q3

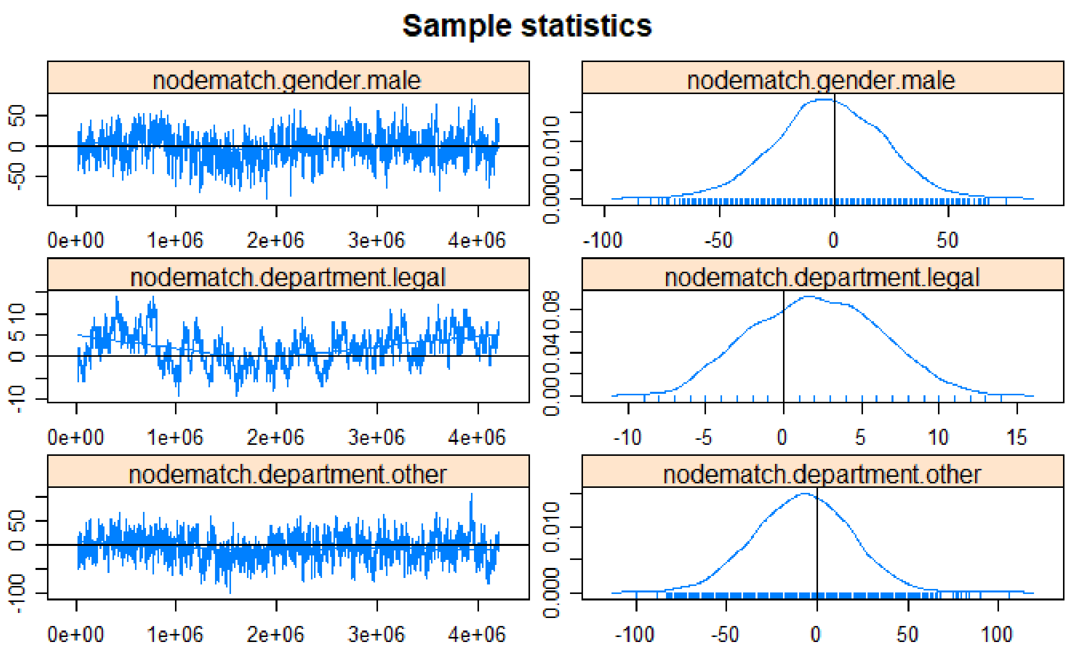


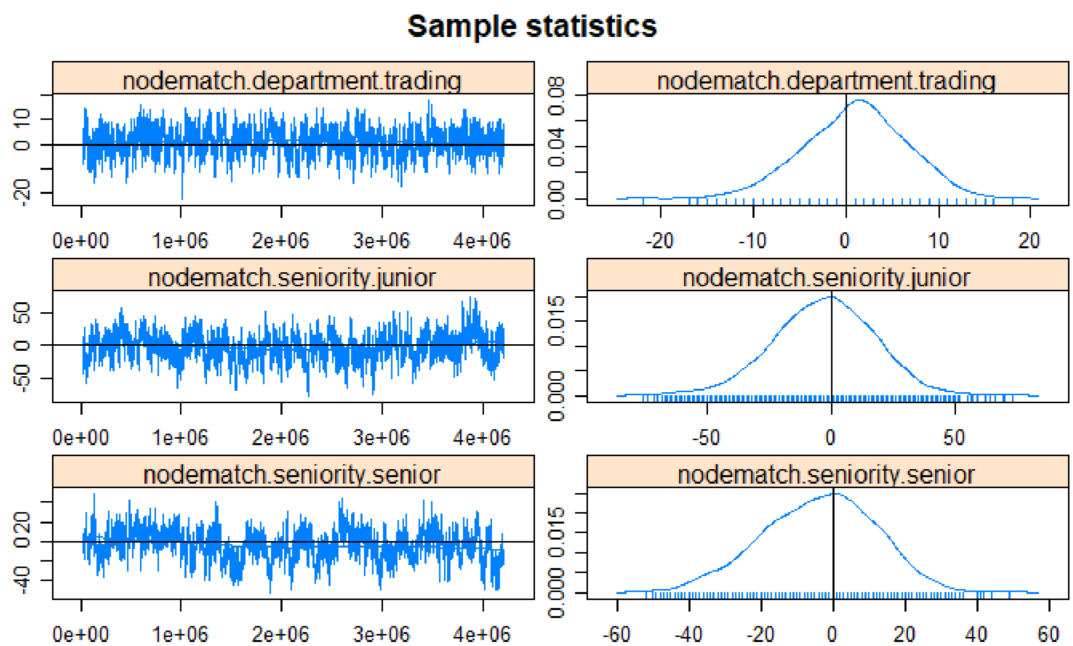
2001Q3 nodematch

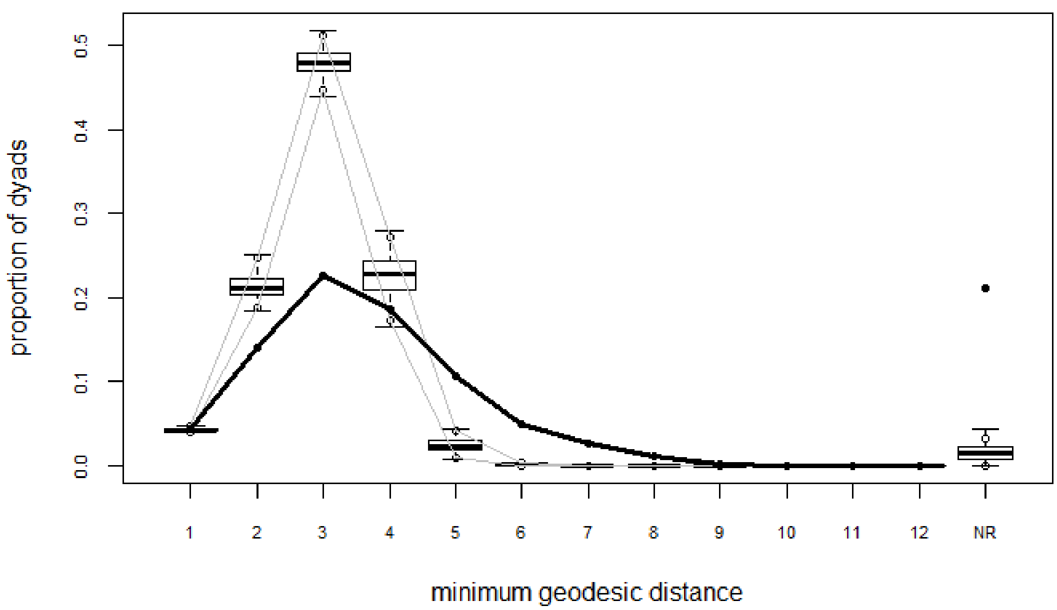
```
==========================
Summary of model fit
==========================

Formula:   net ~ edges + mutual + nodematch("gender", diff = T) +
    nodematch("department", diff = T) + nodematch("seniority",
    diff = T)

Iterations:  2 out of 20

Monte Carlo MLE Results:
                            Estimate Std. Error MCMC % z value Pr(>|z|)
edges                       -3.85251    0.08305      0 -46.386  < 1e-04 ***
mutual                       3.47942    0.12509      0  27.815  < 1e-04 ***
nodematch.gender.female      0.39477    0.10645      0   3.709 0.000208 ***
nodematch.gender.male       -0.03360    0.06797      0  -0.494 0.621127
nodematch.department.legal   1.94536    0.27519      1   7.069  < 1e-04 ***
nodematch.department.other   0.20525    0.06674      0   3.075 0.002103 **
nodematch.department.trading -0.07289   0.18850      0  -0.387 0.698995
nodematch.seniority.junior  -0.18826    0.07383      0  -2.550 0.010771 *
nodematch.seniority.senior   0.56575    0.08097      0   6.987  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

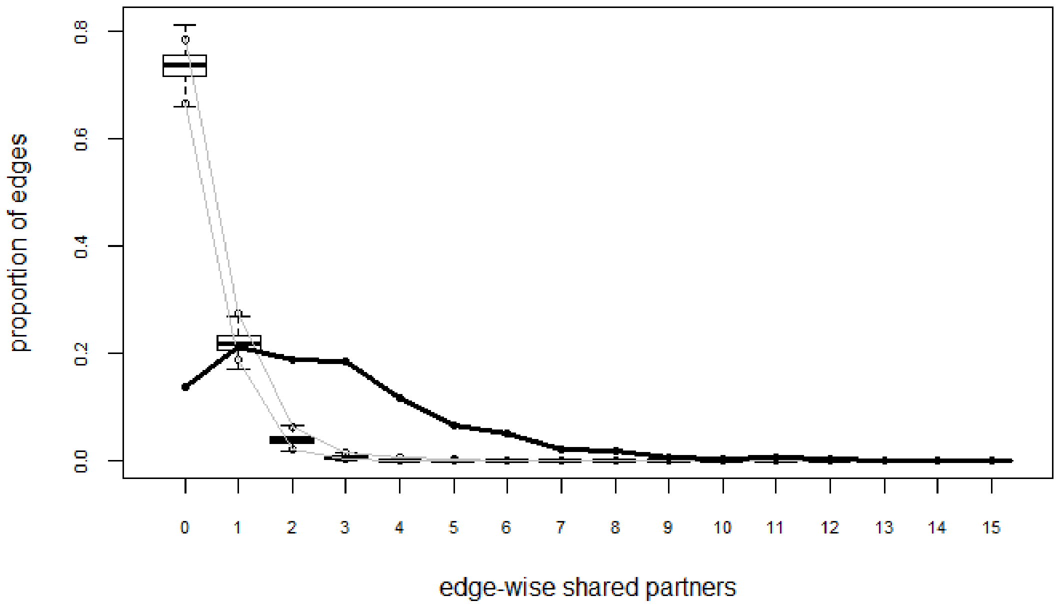
```
    Null Deviance: 26209  on 18906  degrees of freedom
 Residual Deviance:  5790  on 18897  degrees of freedom

AIC: 5808    BIC: 5878    (Smaller is better.)
```


Sample statistics
edges
edges
mutual
mutual
nodematch.gender.female
nodematch.gender.female

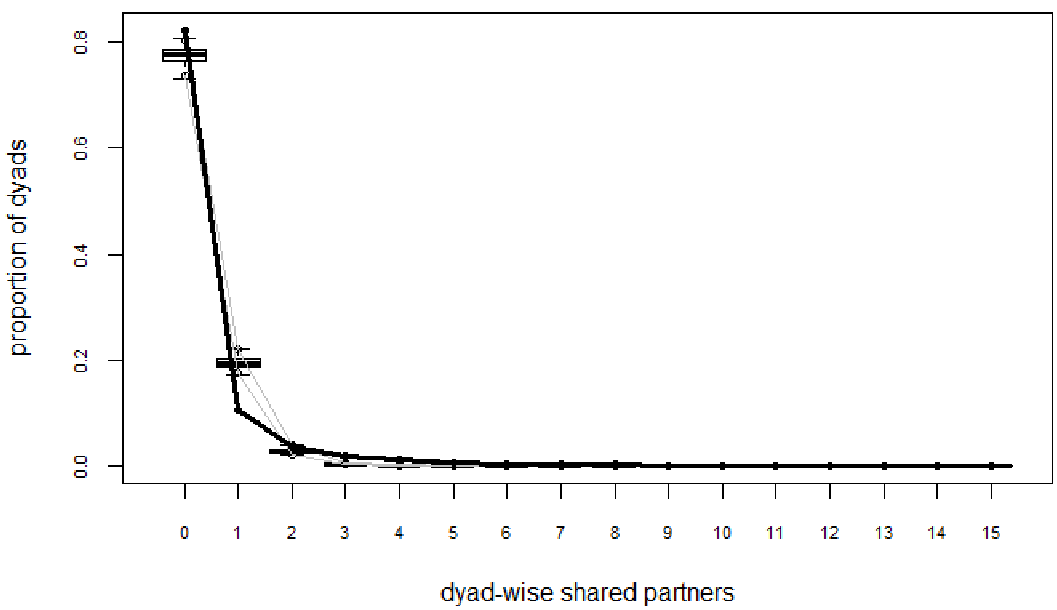

Sample statistics
nodematch.gender.male
nodematch.gender.male
nodematch.department.legal
nodematch.department.legal
nodematch.department.other
nodematch.department.other


Sample statistics
nodematch.department.trading
nodematch.department.trading
nodematch.seniority.junior
nodematch.seniority.junior
nodematch.seniority.senior
nodematch.seniority.senior
0e+00
1e+06
2e+06
3e+06
4e+06

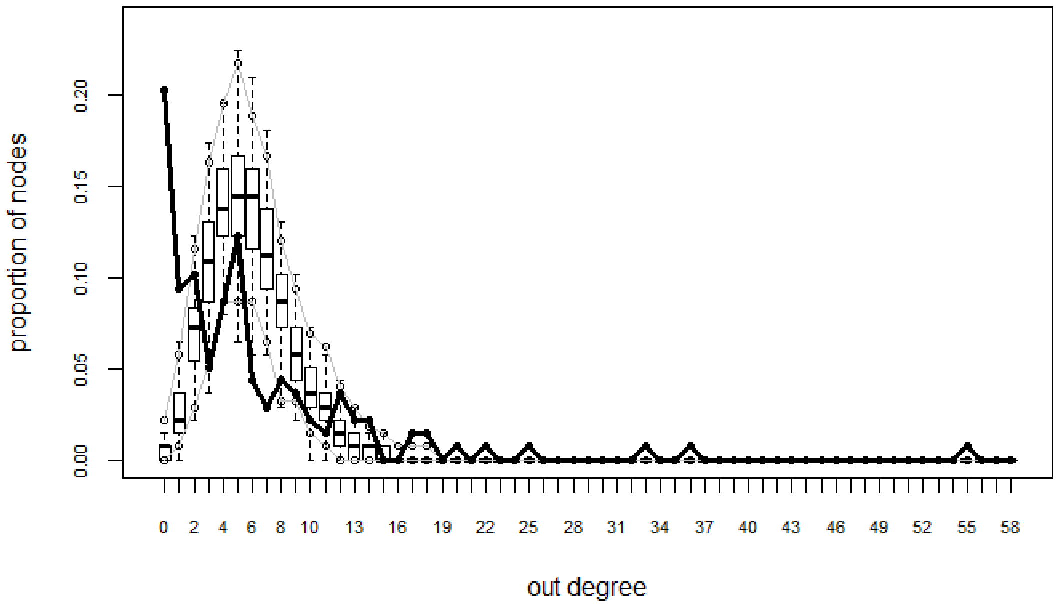

proportion of dyads
minimum geodesic distance
NR

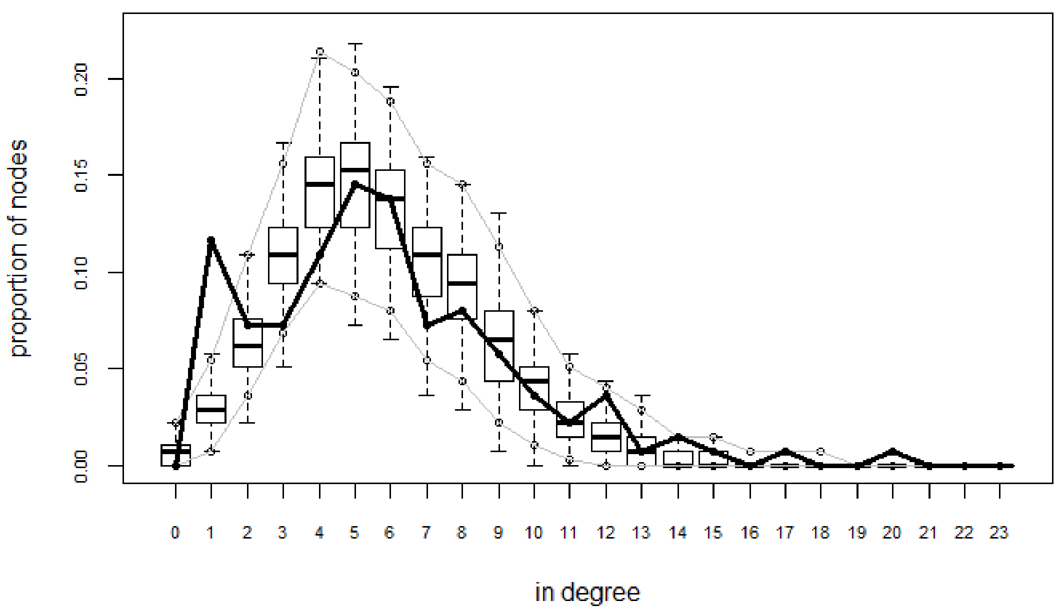


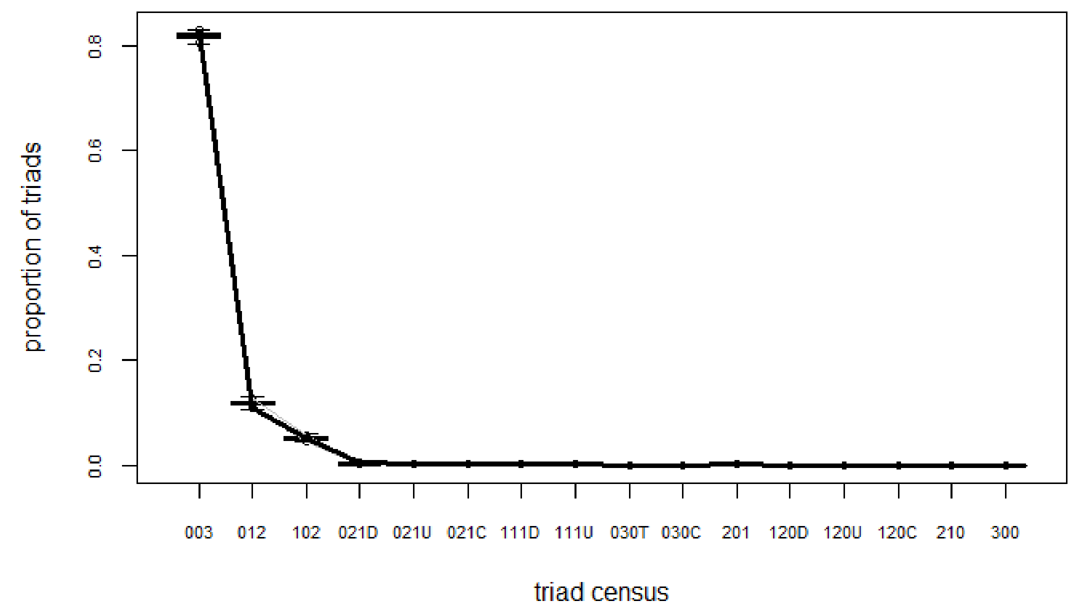
proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
edge-wise shared partners

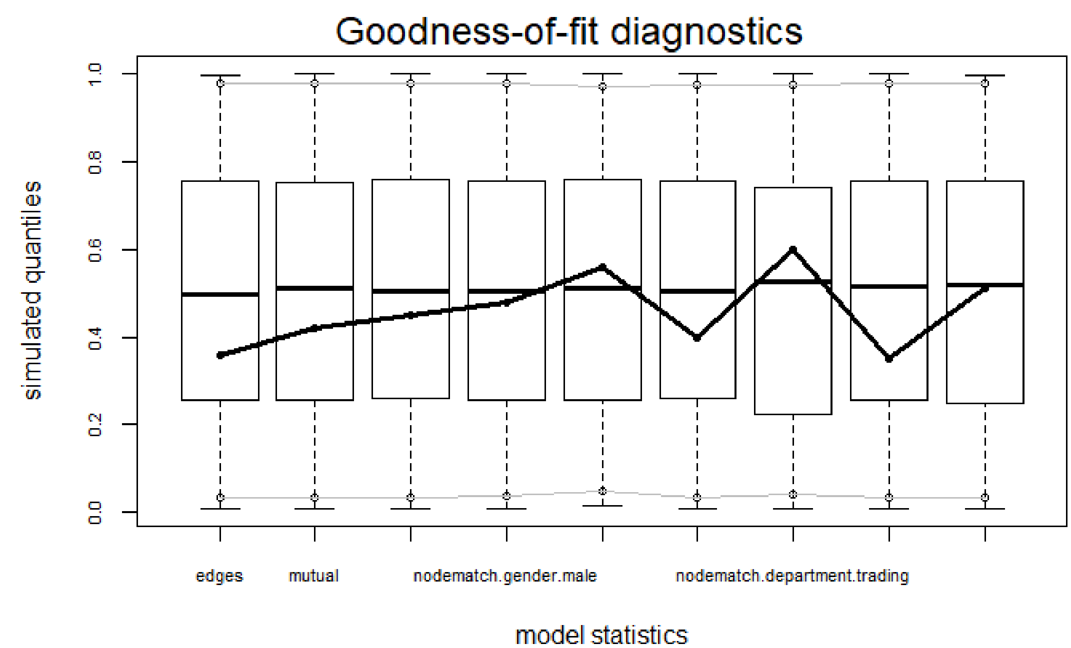


proportion of dyads
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
dyad-wise shared partners


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 2 4 6 8 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58
out degree


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
in degree

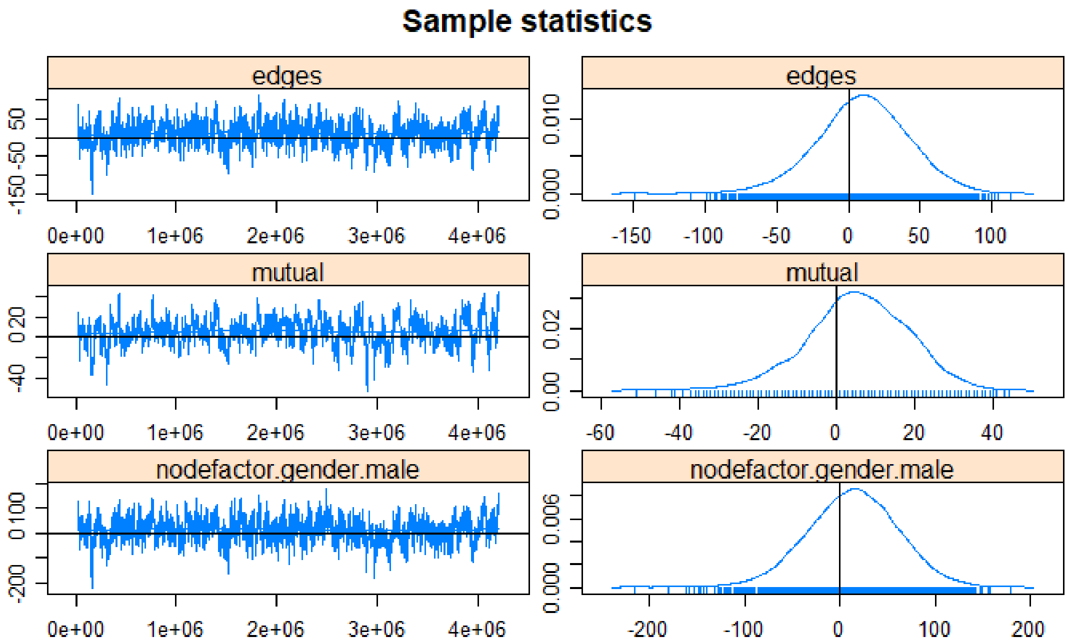


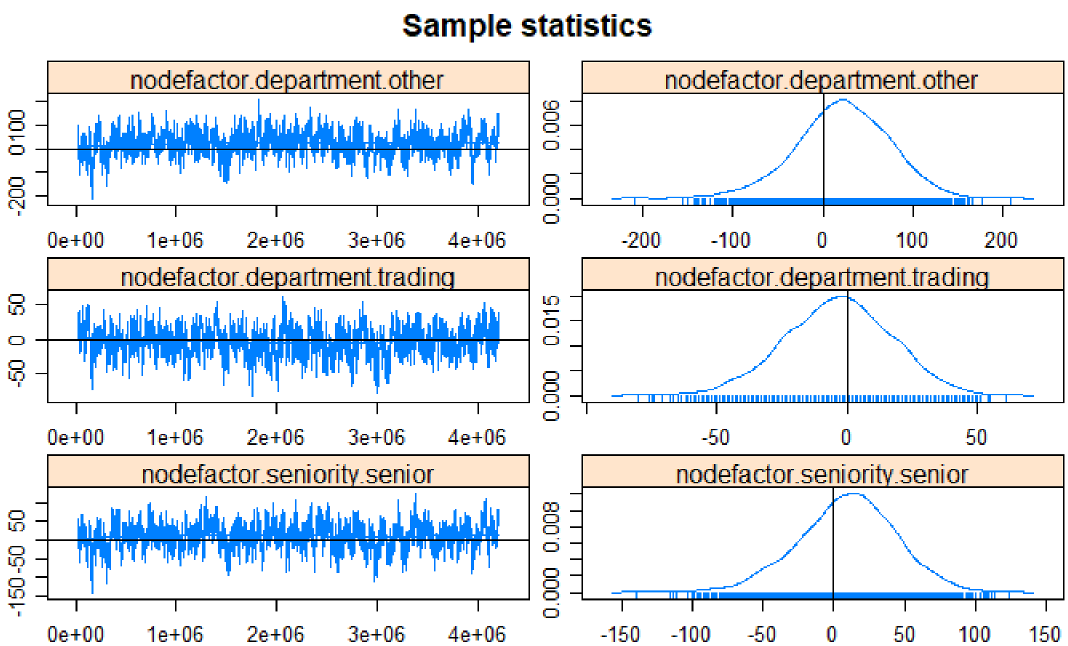


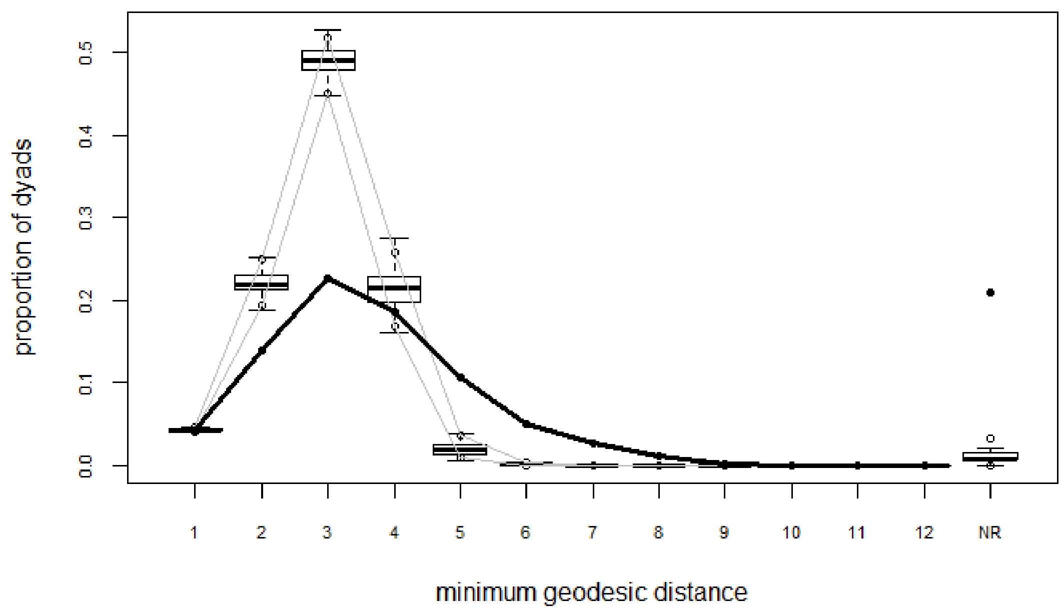
2001Q3 Nodefactor

==========================
Summary of model fit
==========================

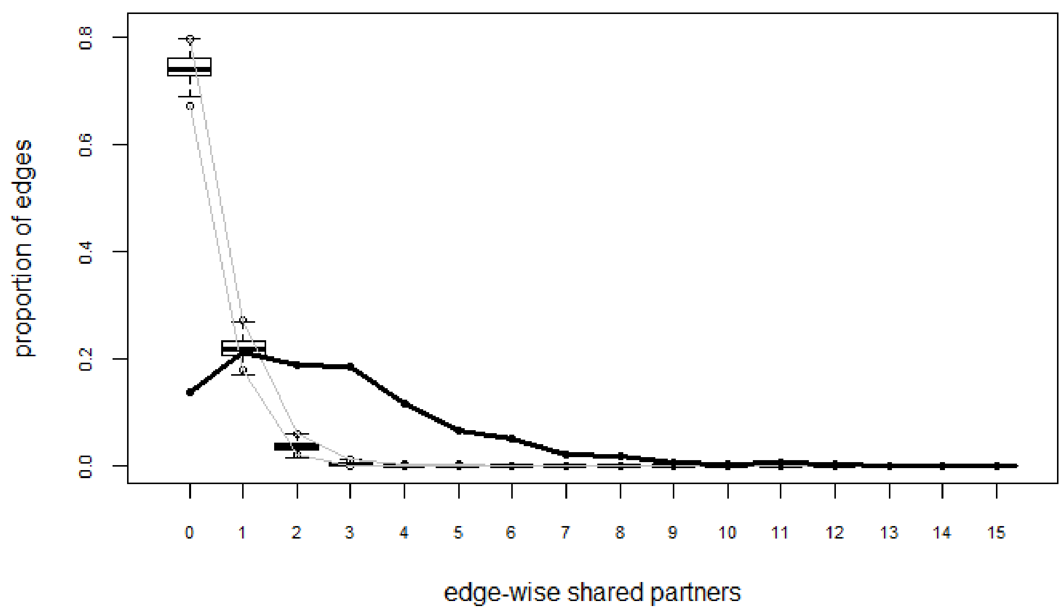
```
Formula:   net ~ edges + mutual + nodefactor("gender") + nodefactor("department") +
    nodefactor("seniority")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                         -3.32642    0.17386      0 -19.133  < 1e-04 ***
mutual                         3.53357    0.12419      0  28.452  < 1e-04 ***
nodefactor.gender.male        -0.14218    0.04873      0  -2.917  0.00353 **
nodefactor.department.other   -0.18787    0.08776      0  -2.141  0.03230 *
nodefactor.department.trading -0.39429    0.10006      0  -3.941  < 1e-04 ***
nodefactor.seniority.senior    0.34485    0.04631      0   7.447  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 26209  on 18906  degrees of freedom
 Residual Deviance:  5840  on 18900  degrees of freedom

AIC: 5852    BIC: 5899    (Smaller is better.)
```


Sample statistics
edges
edges
mutual
mutual
nodefactor.gender.male
nodefactor.gender.male
0e+00 1e+06 2e+06 3e+06 4e+06

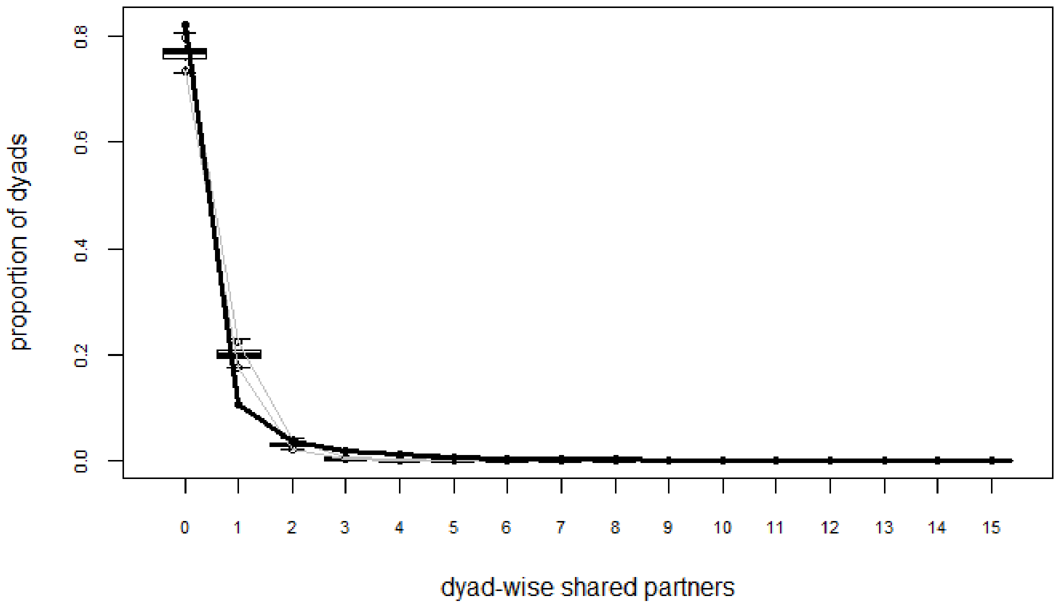

Sample statistics
nodefactor.department.other
nodefactor.department.other
nodefactor.department.trading
nodefactor.department.trading
nodefactor.seniority.senior
nodefactor.seniority.senior
0e+00 1e+06 2e+06 3e+06 4e+06

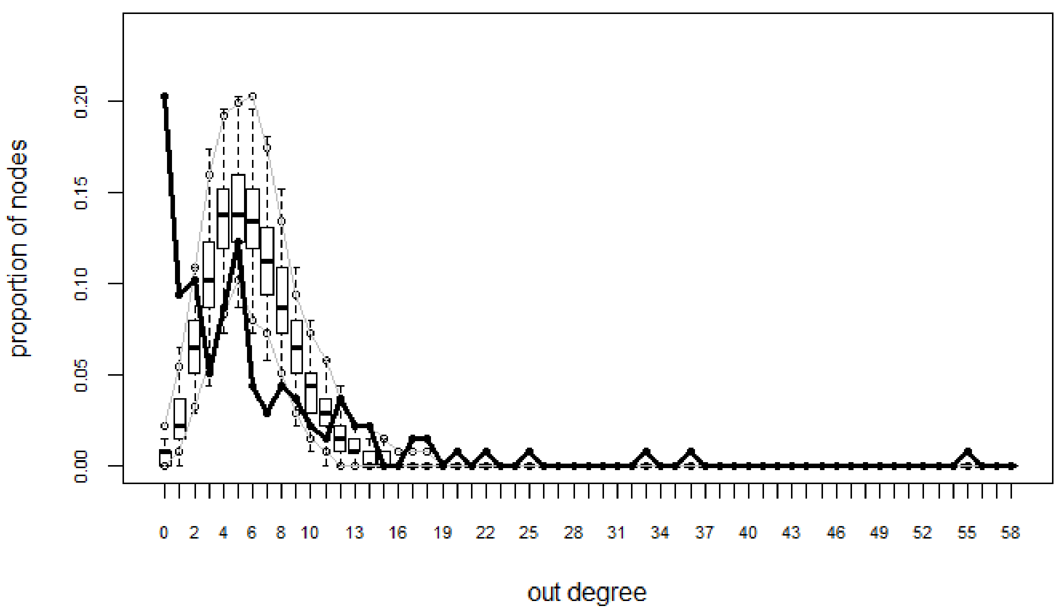


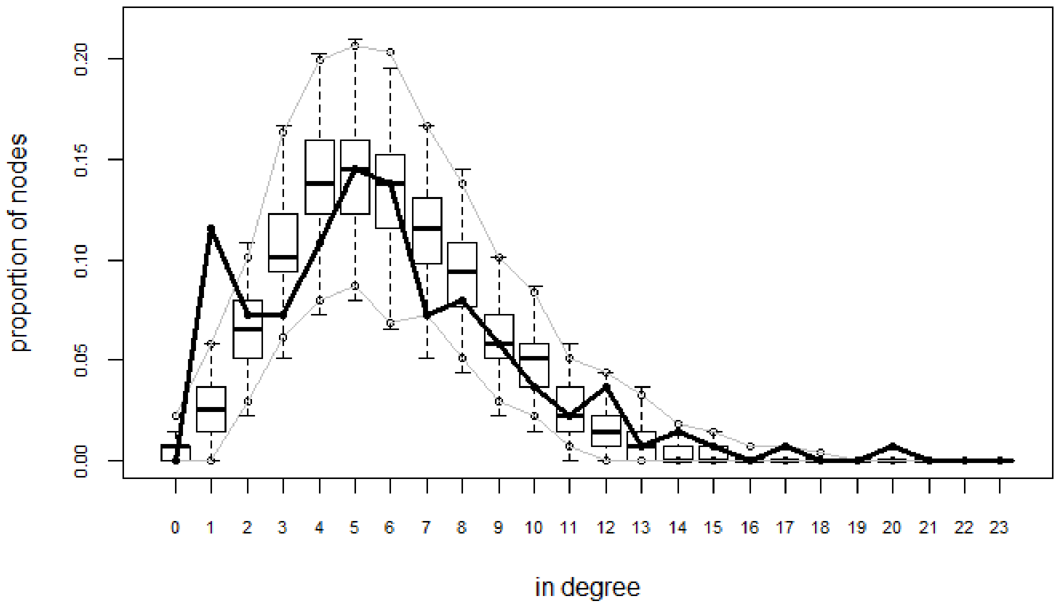
proportion of dyads
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
6
7
8
9
10
11
12
NR
minimum geodesic distance

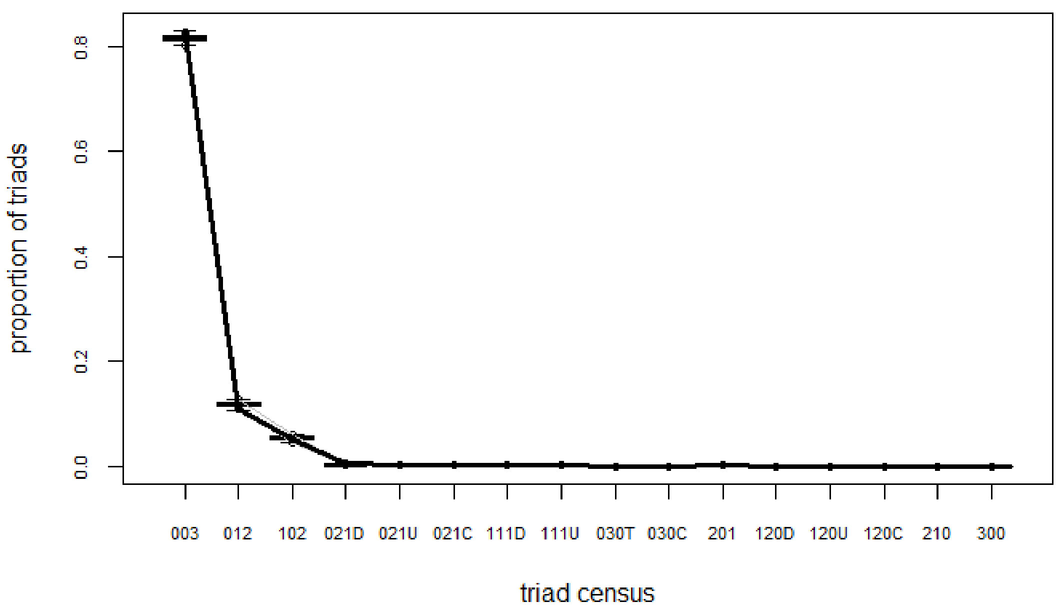


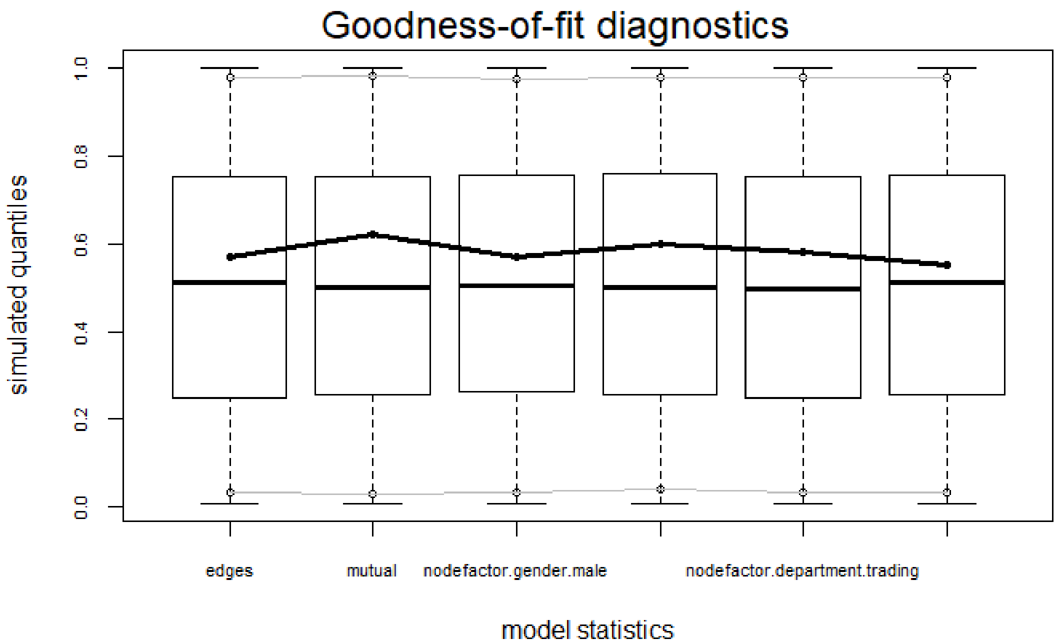
proportion of edges
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
edge-wise shared partners


proportion of dyads
0.0
0.2
0.4
0.6
0.8
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
dyad-wise shared partners

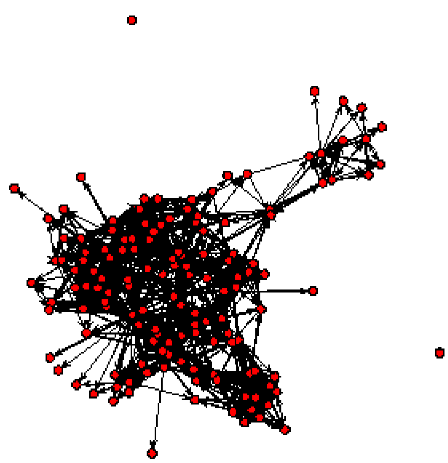

proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 2 4 6 8 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58
out degree


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
in degree


proportion of triads
0.0
0.2
0.4
0.6
0.8
003 012 102 021D 021U 021C 111D 111U 030T 030C 201 120D 120U 120C 210 300
triad census



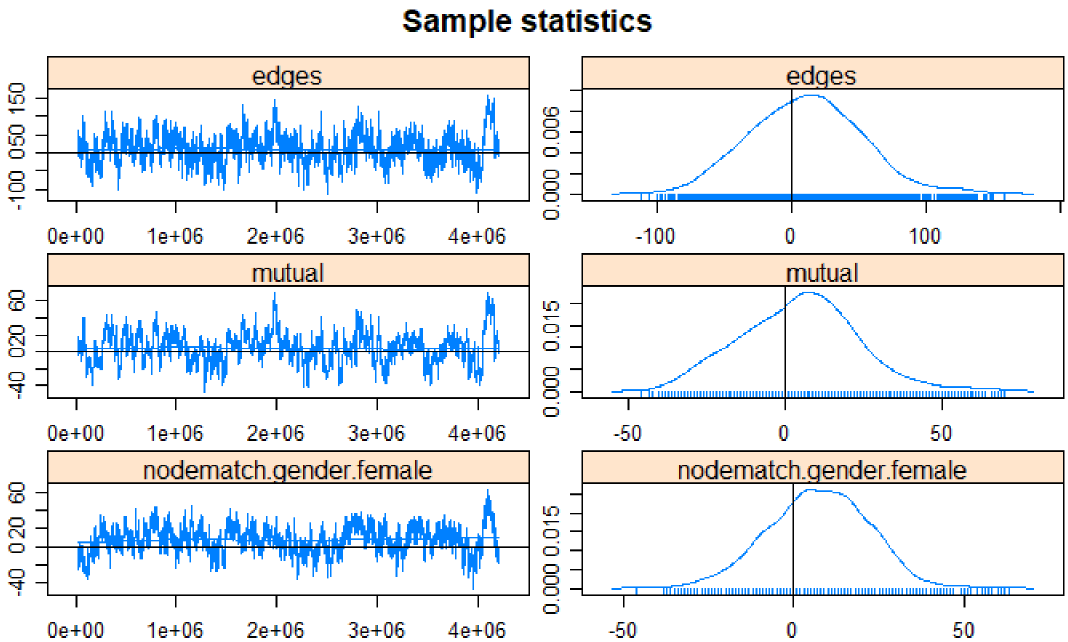
2001Q4

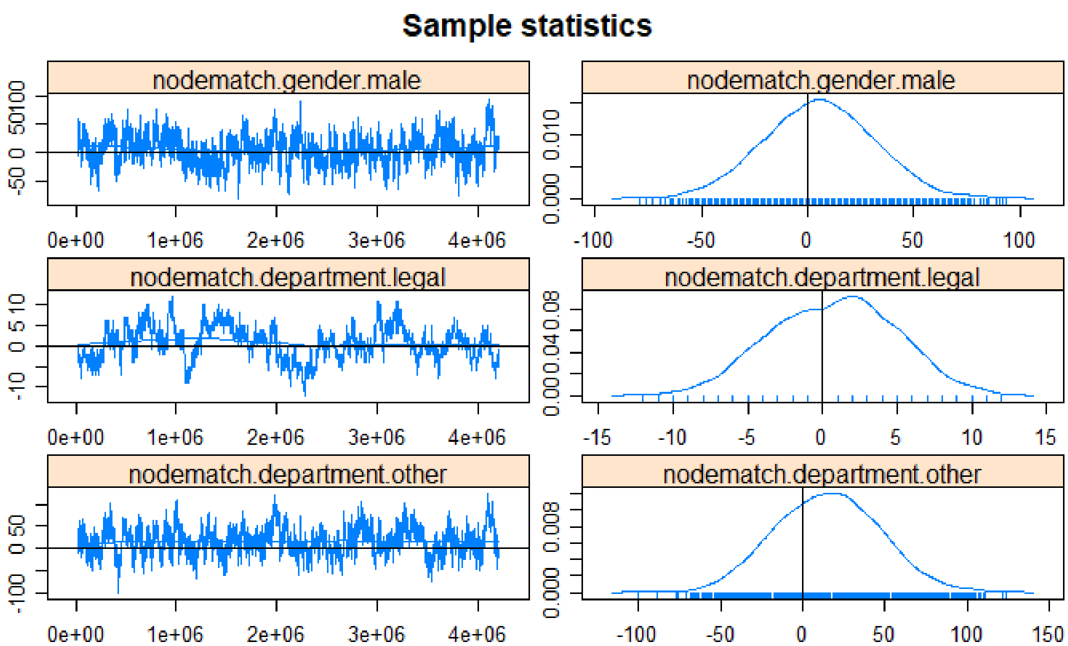


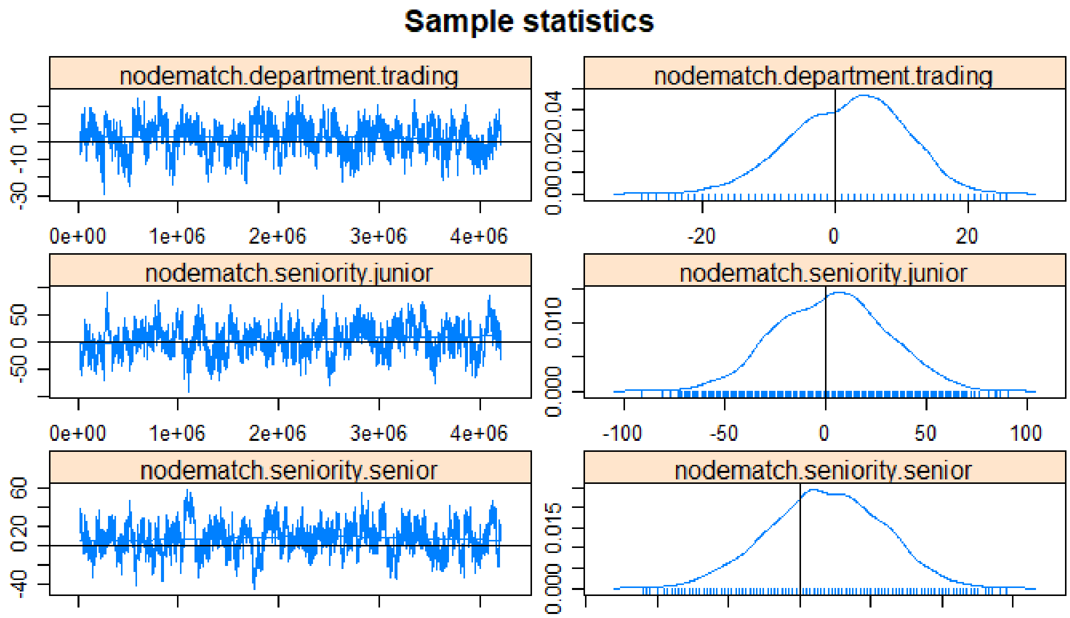
2001Q4 nodematch

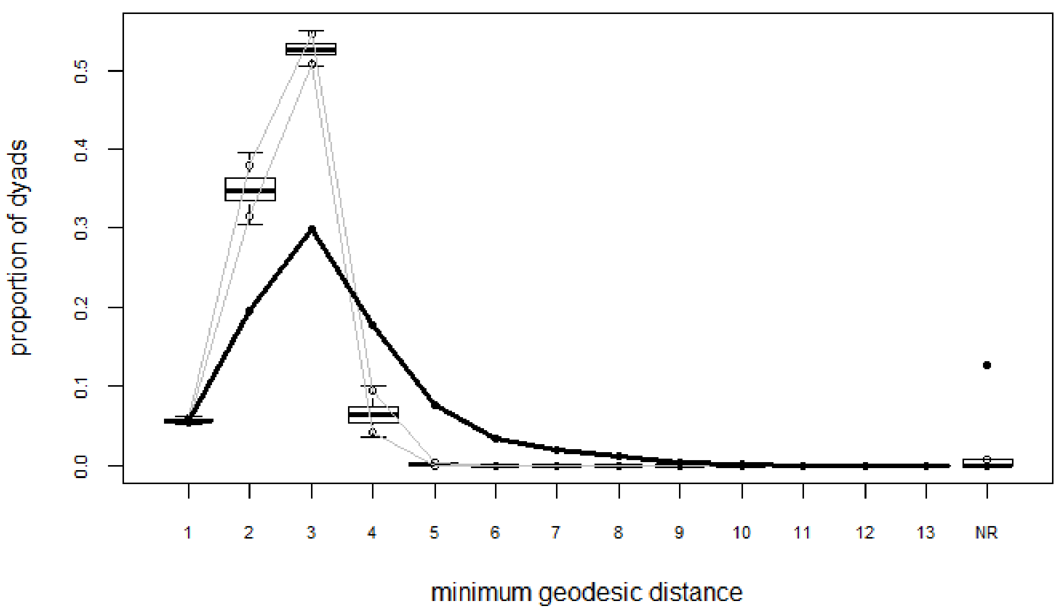
==========================
Summary of model fit
==========================

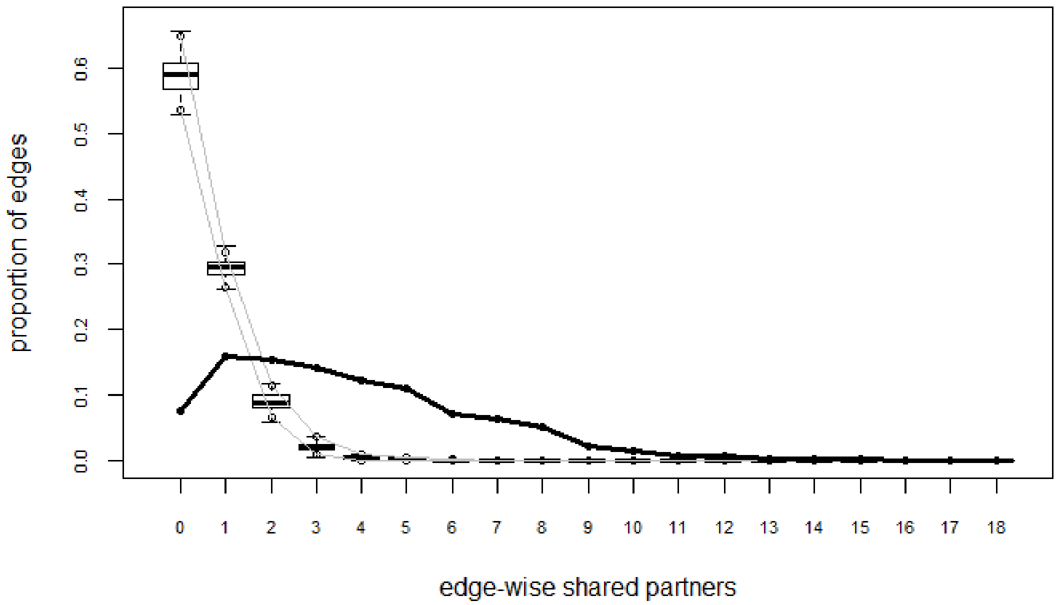
```
Formula:   net ~ edges + mutual + nodematch("gender", diff = T) +
    nodematch("department", diff = T) + nodematch("seniority",
    diff = T)

Iterations:  2 out of 20

Monte Carlo MLE Results:
                          Estimate Std. Error MCMC % z value Pr(>|z|)
edges                     -3.60288    0.07236      0 -49.789  < 1e-04 ***
mutual                     3.78842    0.10400      0  36.426  < 1e-04 ***
nodematch.gender.female    0.32240    0.08227      0   3.919  < 1e-04 ***
nodematch.gender.male     -0.18829    0.05584      0  -3.372 0.000746 ***
nodematch.department.legal    1.70337    0.25329      1   6.725  < 1e-04 ***
nodematch.department.other    0.14560    0.05613      0   2.594 0.009492 **
nodematch.department.trading  0.09740    0.13028      0   0.748 0.454685
nodematch.seniority.junior   -0.19071    0.05533      0  -3.447 0.000567 ***
nodematch.seniority.senior    0.44534    0.07700      0   5.784  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 27365  on 19740  degrees of freedom
 Residual Deviance:  7235  on 19731  degrees of freedom

AIC: 7253    BIC: 7324    (Smaller is better.)
```


Sample statistics
edges
edges
mutual
mutual
nodematch.gender.female
nodematch.gender.female

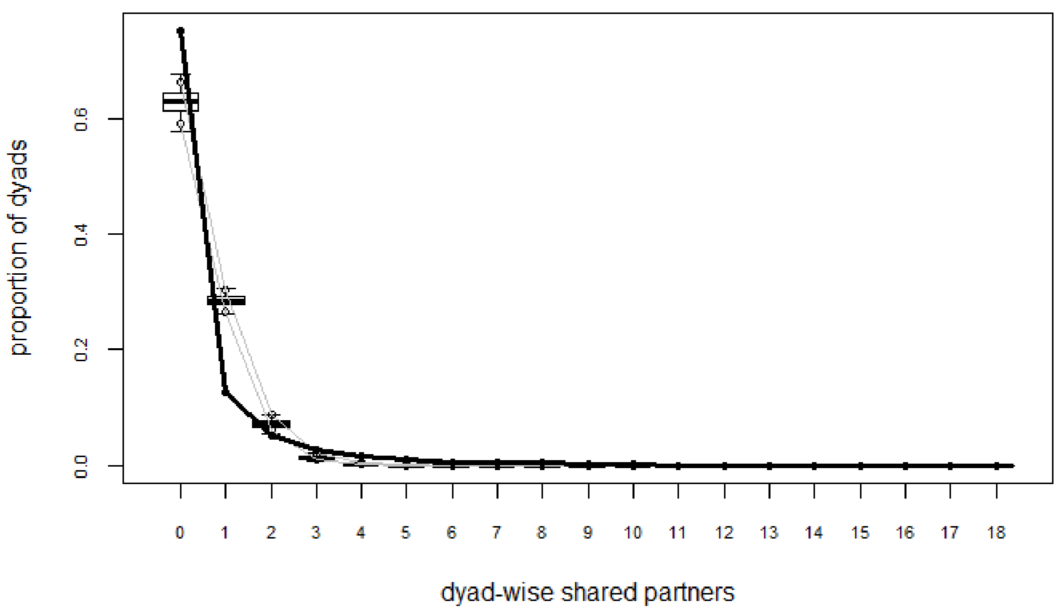

Sample statistics
nodematch.gender.male
nodematch.gender.male
nodematch.department.legal
nodematch.department.legal
nodematch.department.other
nodematch.department.other


Sample statistics
nodematch.department.trading
nodematch.seniority.junior
nodematch.seniority.senior

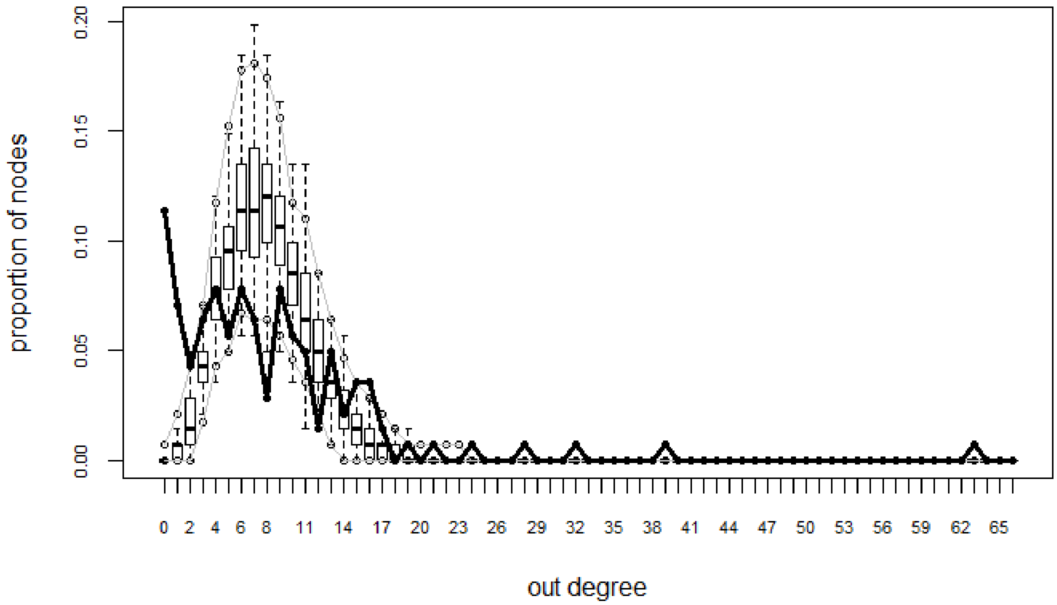

proportion of dyads
minimum geodesic distance
NR



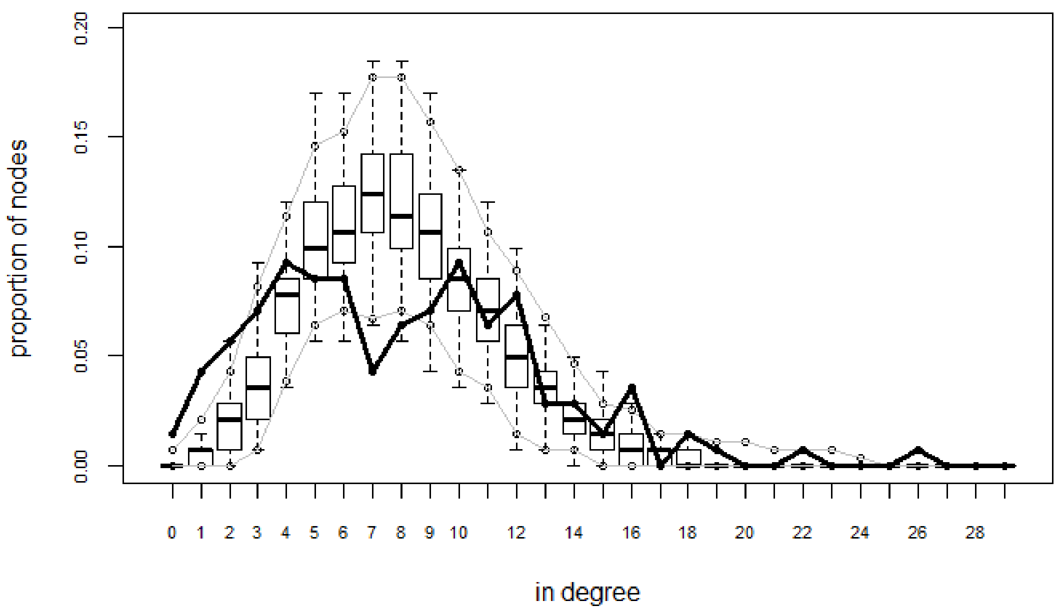
proportion of edges
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
edge-wise shared partners



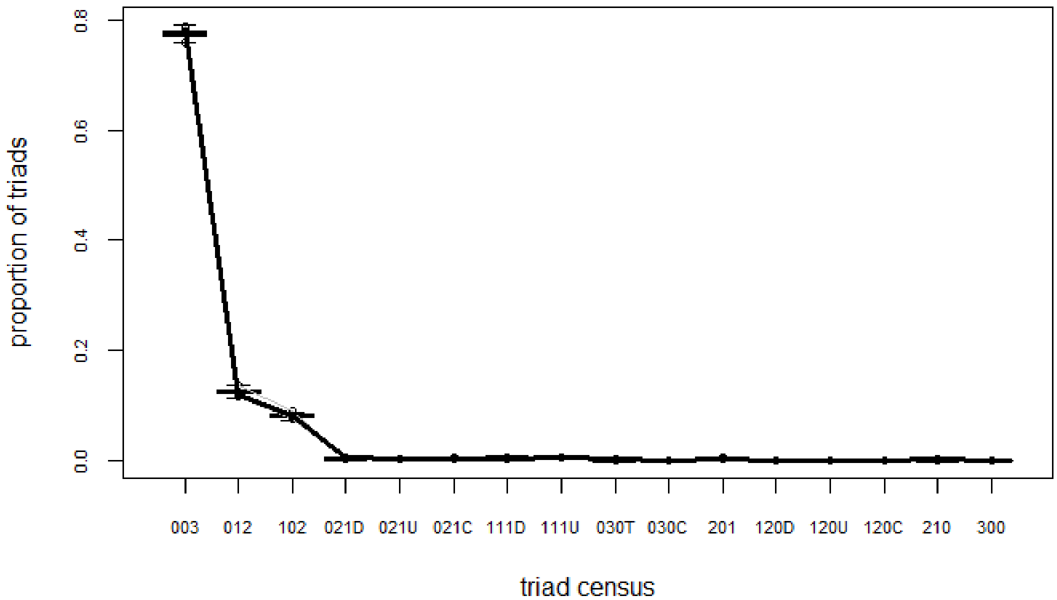
proportion of dyads
0.0
0.2
0.4
0.6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
dyad-wise shared partners

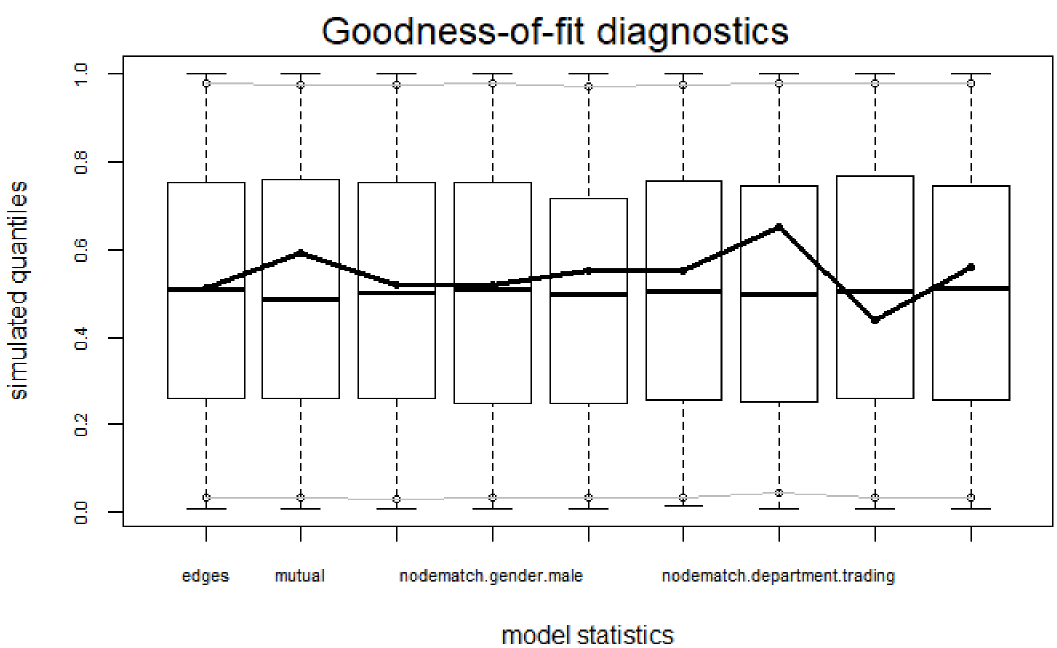


proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 2 4 6 8 11 14 17 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62 65
out degree



proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28
in degree





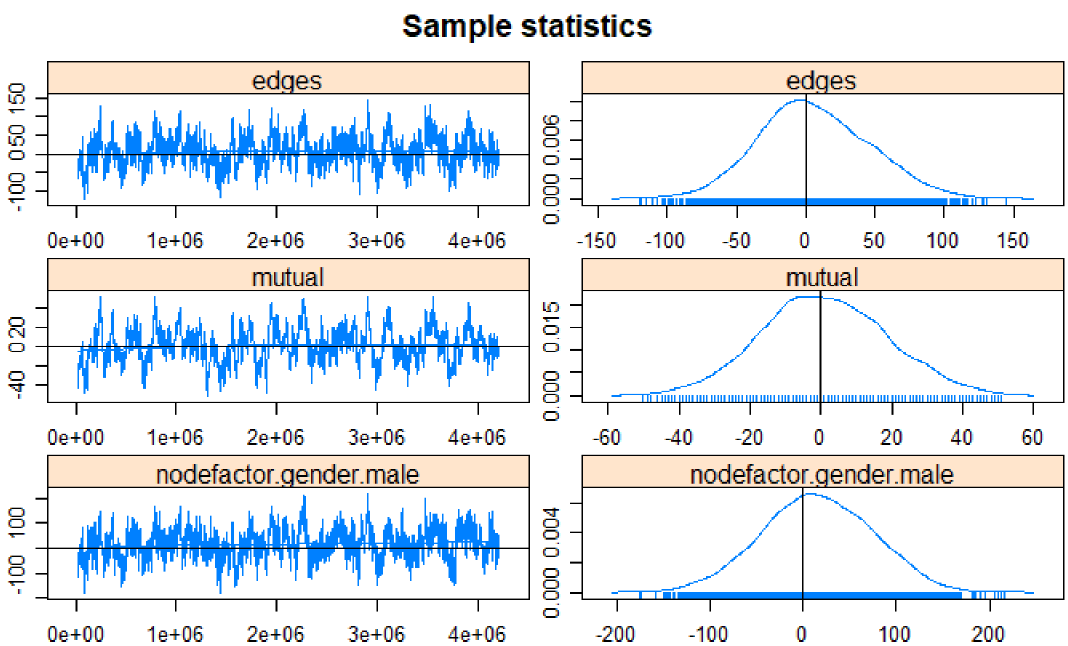
2001Q4 nodefactor

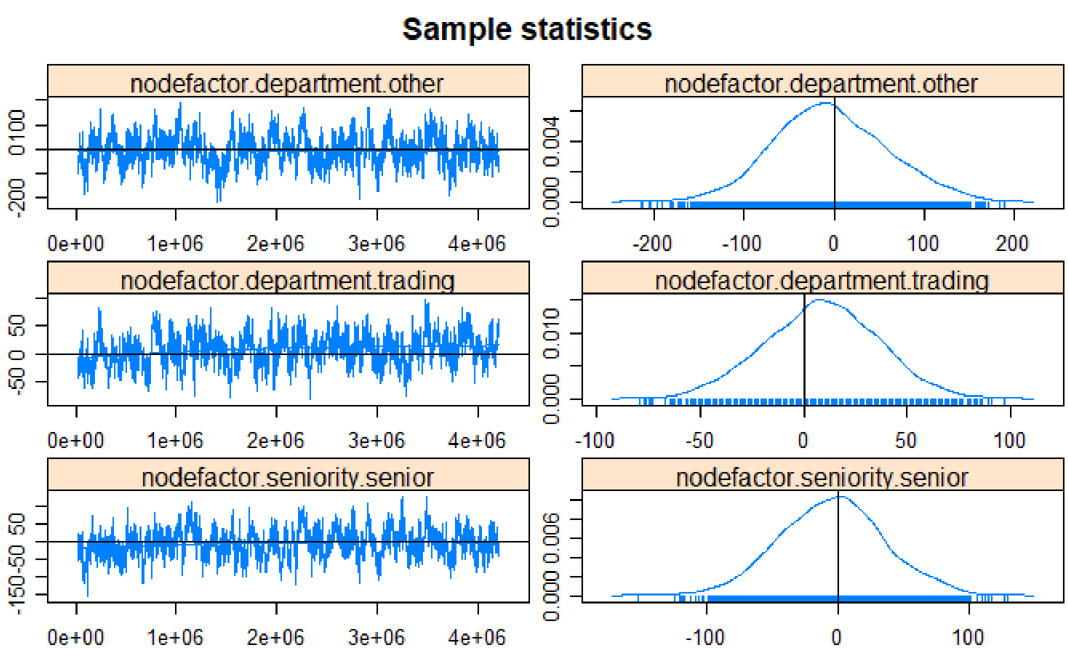
==========================

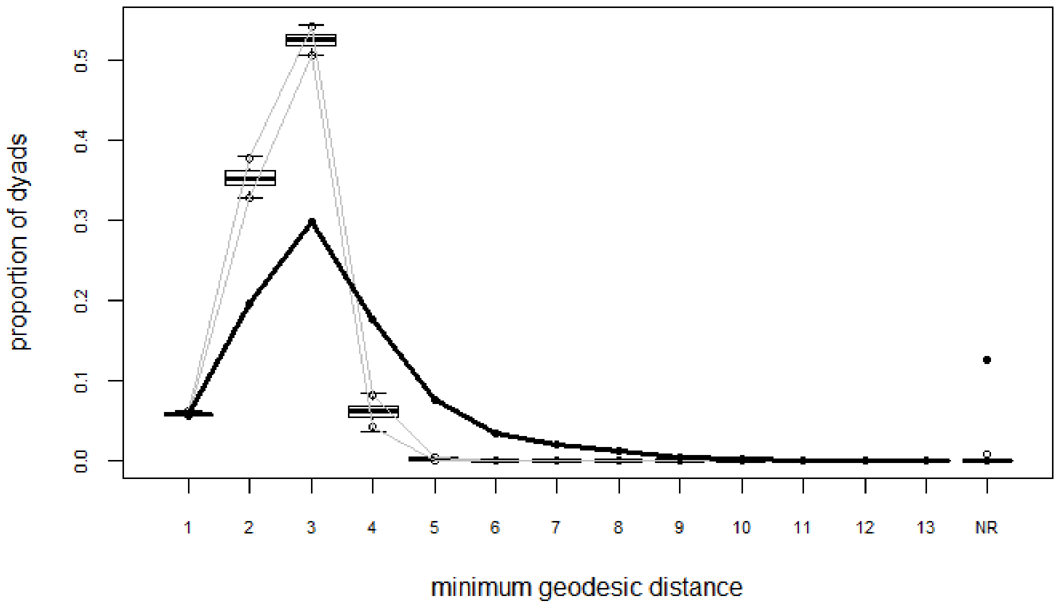
Summary of model fit

==========================

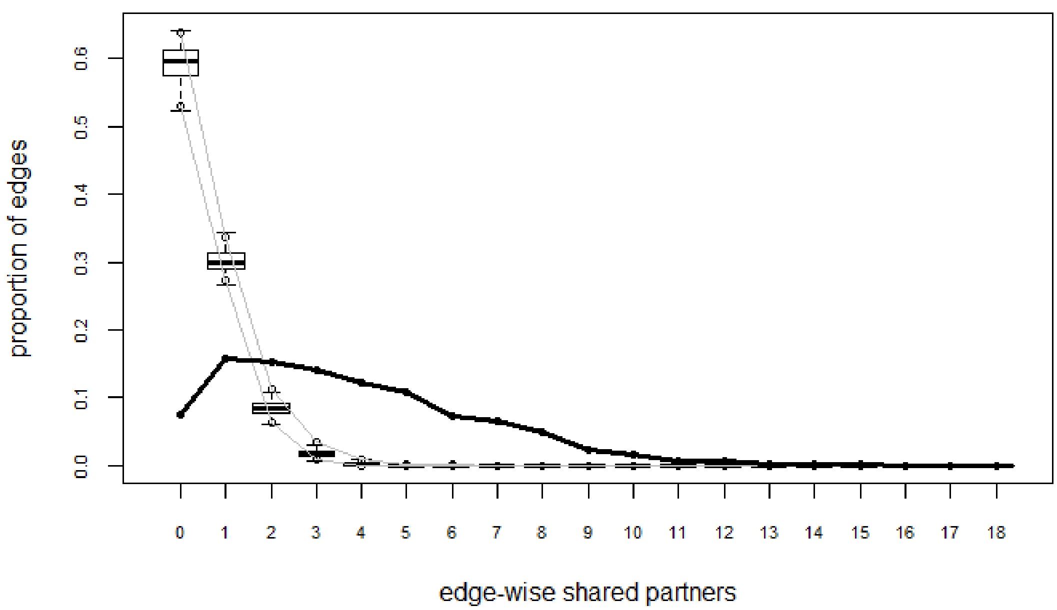
```
Formula:   net ~ edges + mutual + nodefactor("gender") + nodefactor("department") +
    nodefactor("seniority")

Iterations:  2 out of 20

Monte Carlo MLE Results:
                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                         -3.12769    0.15597      0 -20.053  < 1e-04 ***
mutual                         3.80451    0.10719      0  35.494  < 1e-04 ***
nodefactor.gender.male        -0.22771    0.03860      0  -5.900  < 1e-04 ***
nodefactor.department.other   -0.16181    0.07868      0  -2.057  0.03972 *
nodefactor.department.trading -0.27167    0.08534      0  -3.184  0.00146 **
nodefactor.seniority.senior    0.28595    0.03862      0   7.403  < 1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 27365  on 19740  degrees of freedom
 Residual Deviance:  7272  on 19734  degrees of freedom

AIC: 7284    BIC: 7331    (Smaller is better.)
```


Sample statistics
edges
mutual
nodefactor.gender.male

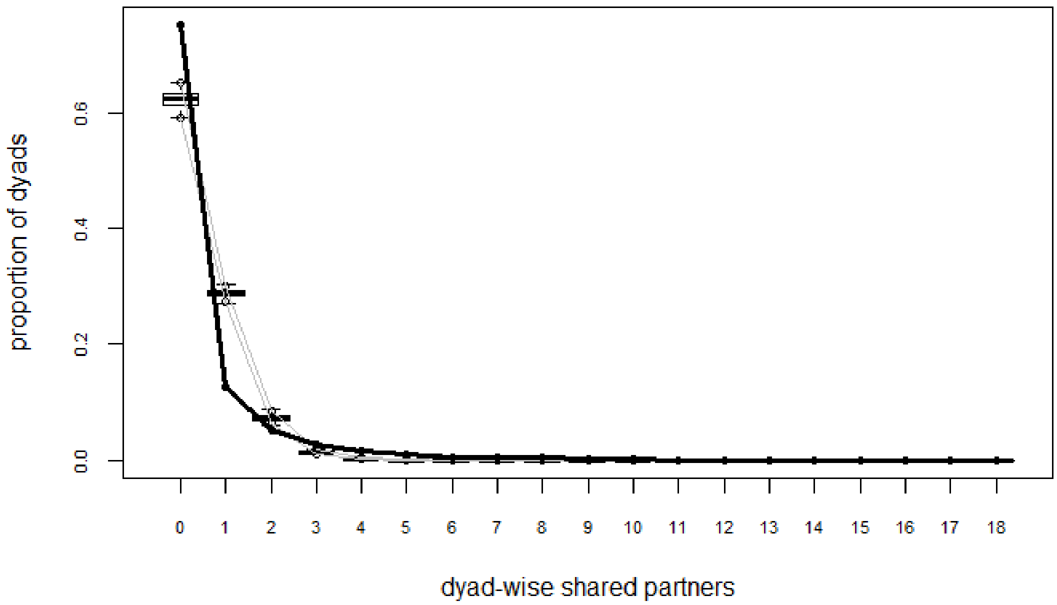

Sample statistics
nodefactor.department.other
nodefactor.department.trading
nodefactor.seniority.senior

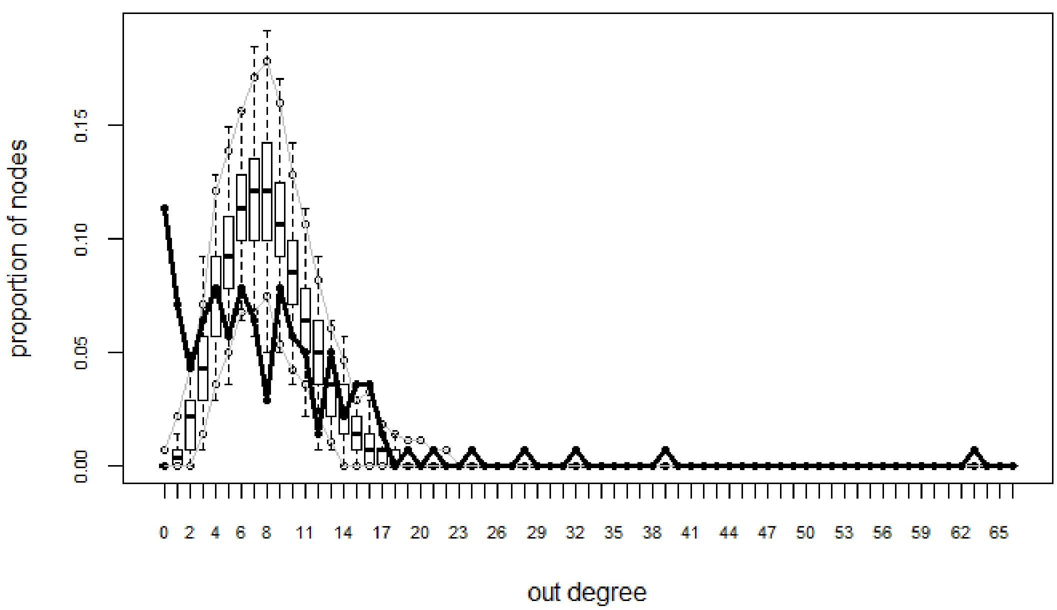


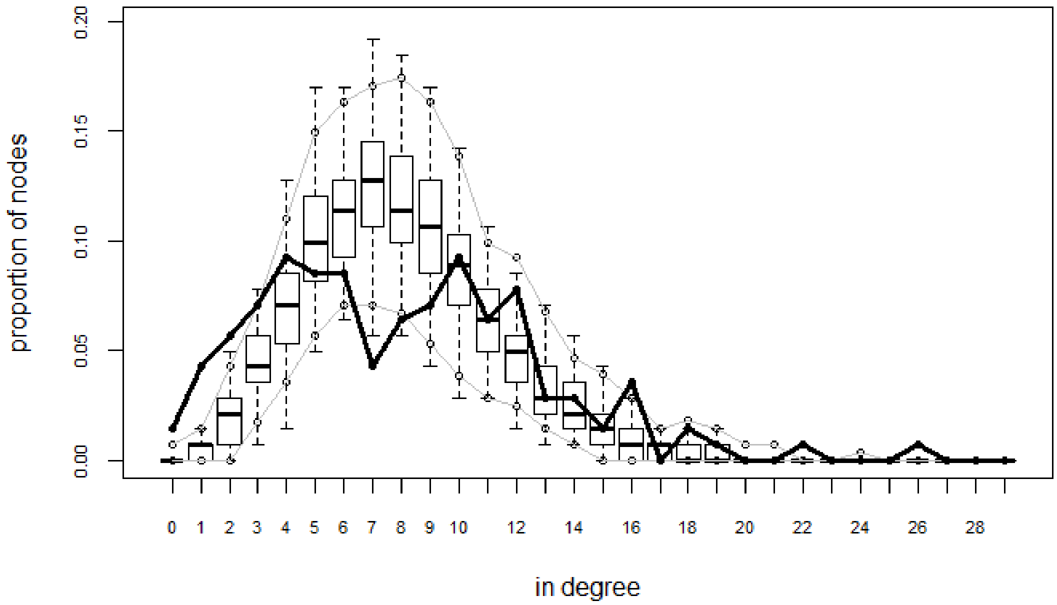
proportion of dyads
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
6
7
8
9
10
11
12
13
NR
minimum geodesic distance

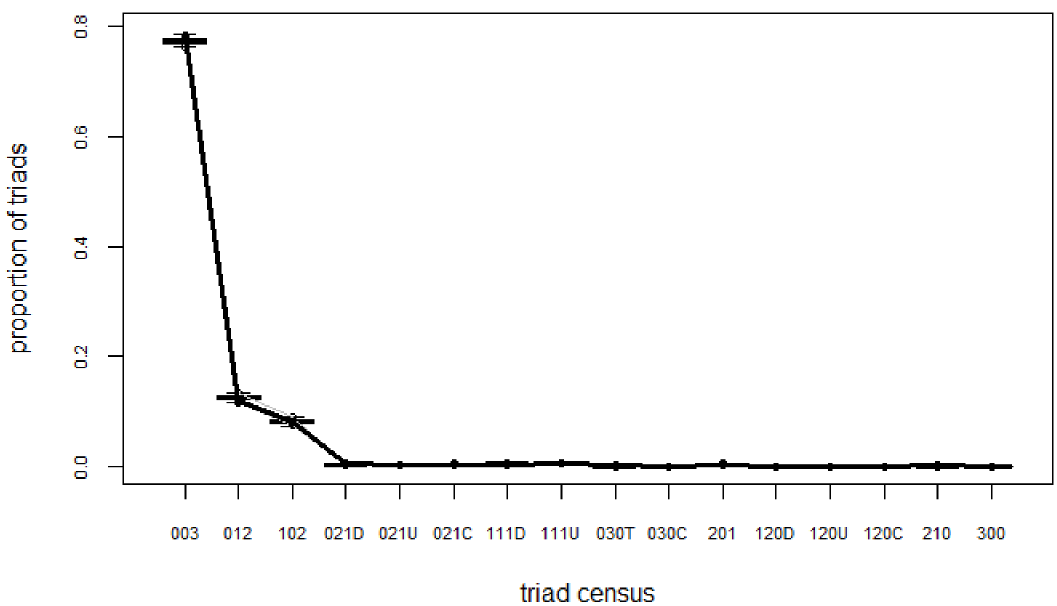


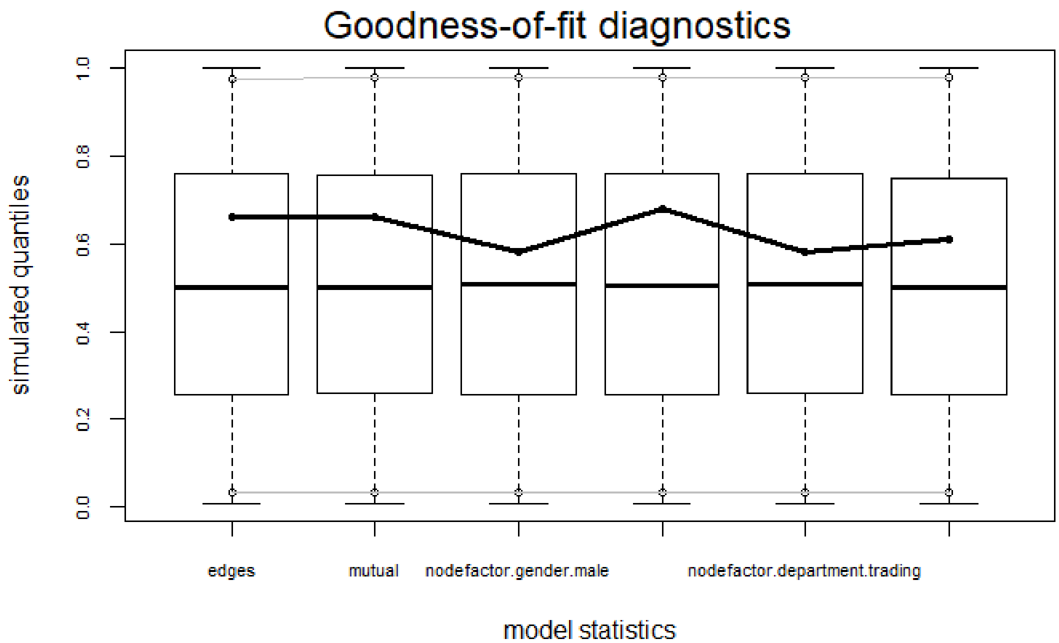
proportion of edges
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
edge-wise shared partners


proportion of dyads
0.0
0.2
0.4
0.6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
dyad-wise shared partners


proportion of nodes
0.00
0.05
0.10
0.15
0 2 4 6 8 11 14 17 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62 65
out degree

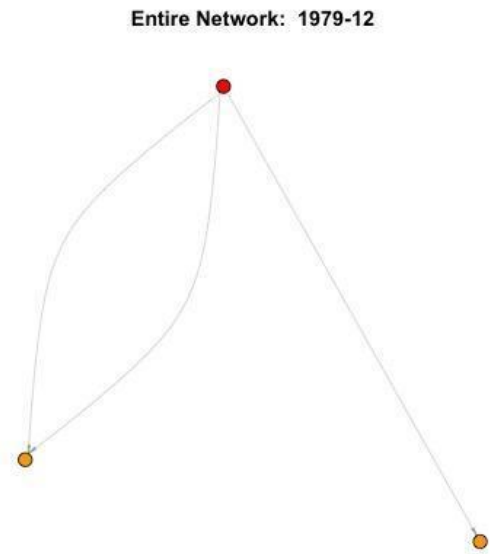

proportion of nodes
0.00
0.05
0.10
0.15
0.20
0 1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28
in degree

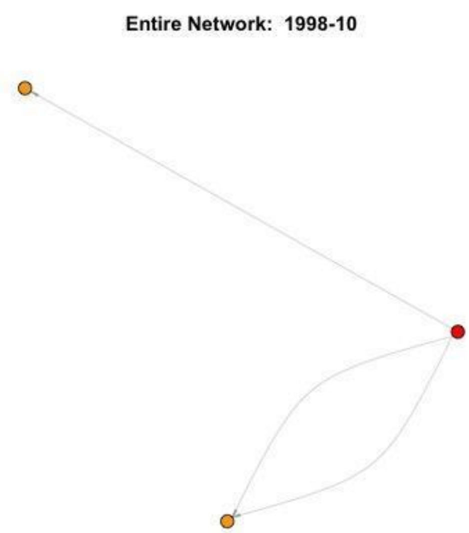

proportion of triads
0.0
0.2
0.4
0.6
0.8
003 012 102 021D 021U 021C 111D 111U 030T 030C 201 120D 120U 120C 210 300
triad census



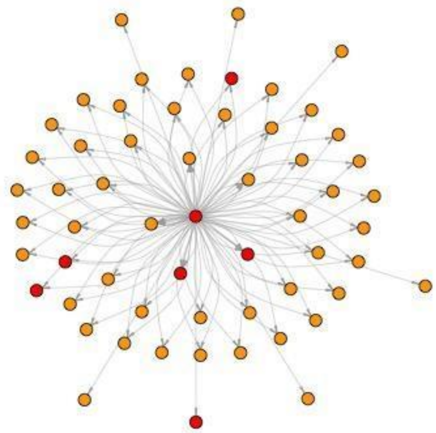
Goodness-of-fit diagnostics
simulated quantiles
1.0
0.8
0.6
0.4
0.2
0.0
edges
mutual
nodefactor.gender.male
nodefactor.department.trading
model statistics

Appendix II - Entire network visualization using igraph for each of the 47 months

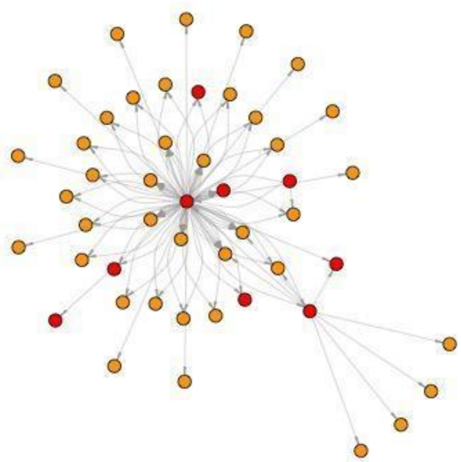




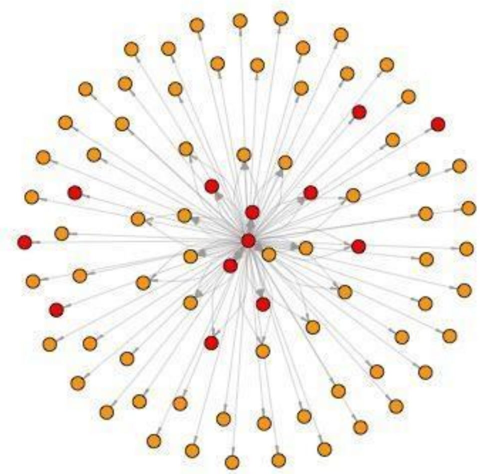
Entire Network: 1998-11



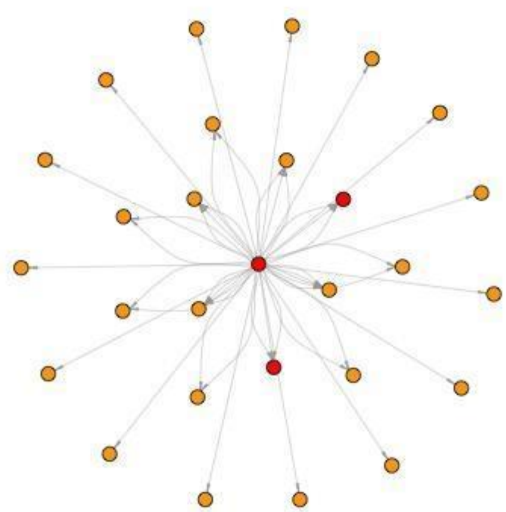
Entire Network: 1998-12

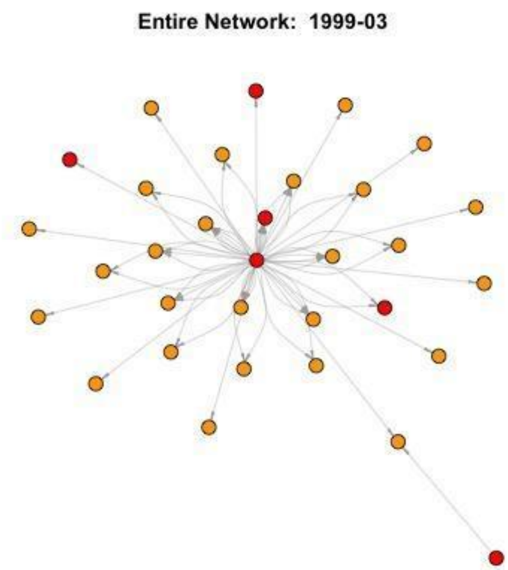


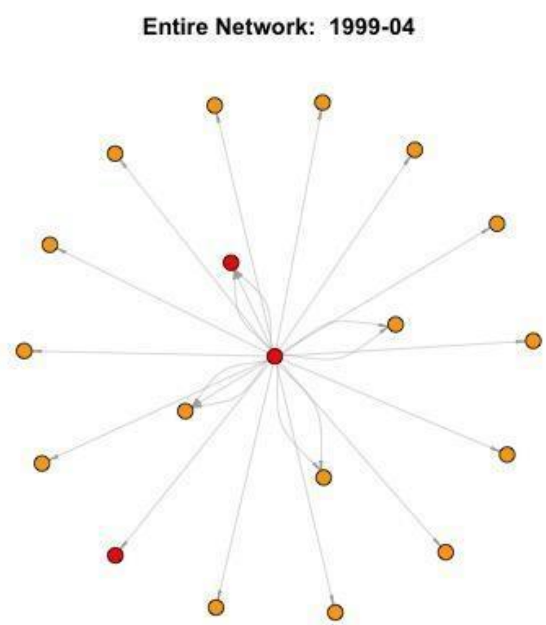
Entire Network: 1999-01



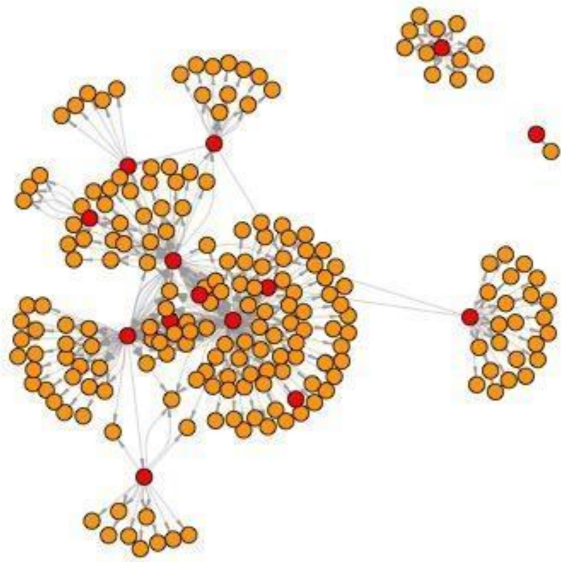
Entire Network: 1999-02



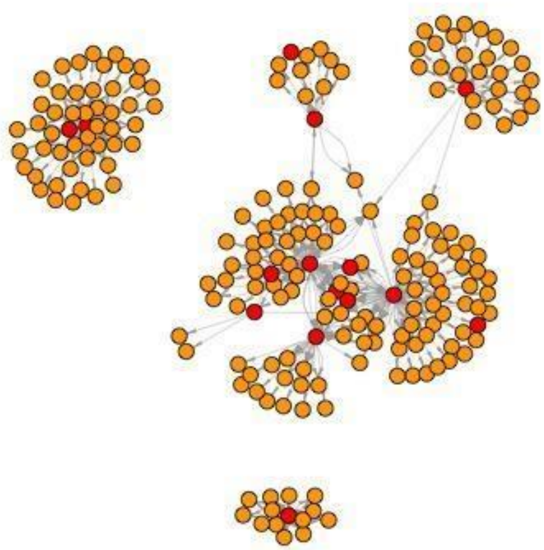

Entire Network: 1999-03


Entire Network: 1999-04

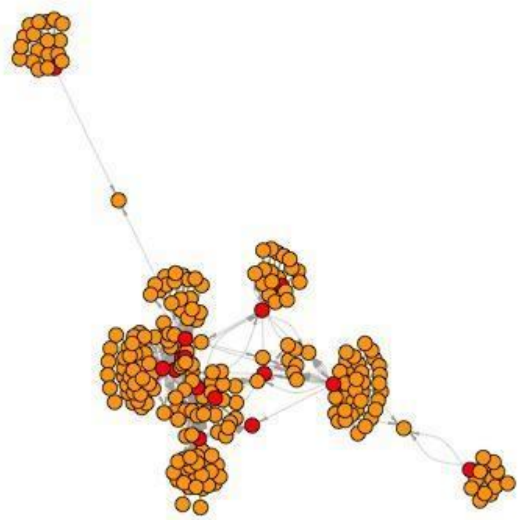
Entire Network: 1999-05



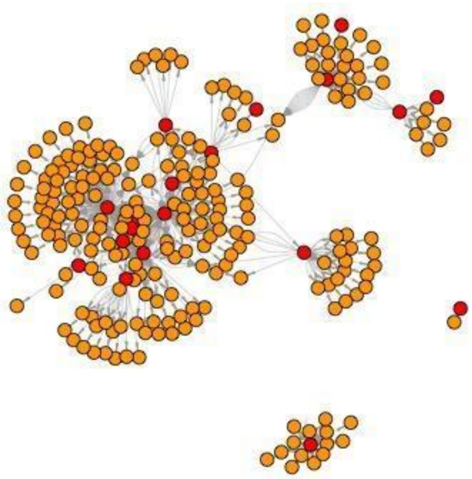
Entire Network: 1999-06



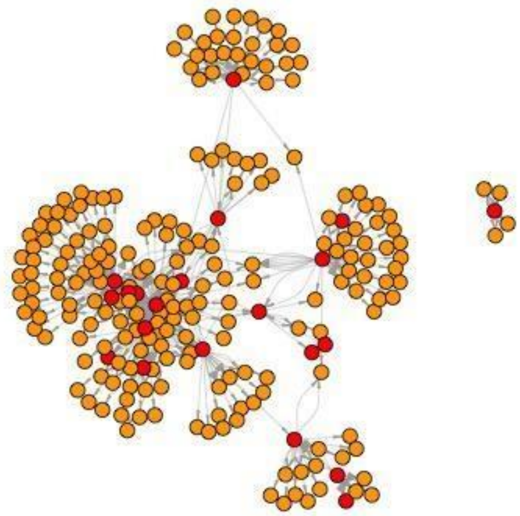
Entire Network: 1999-07



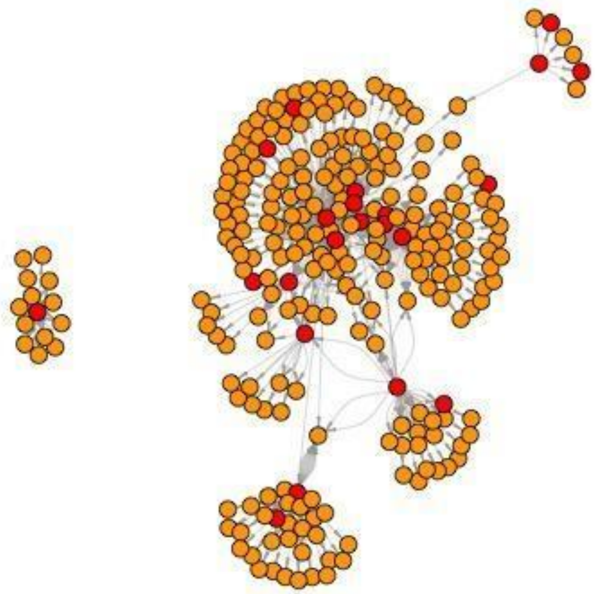
Entire Network: 1999-08



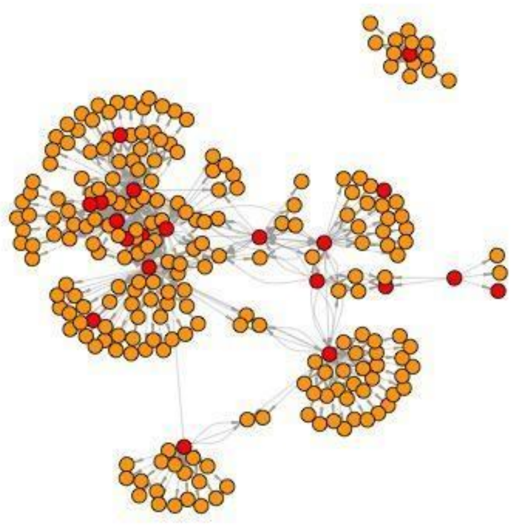
Entire Network: 1999-09



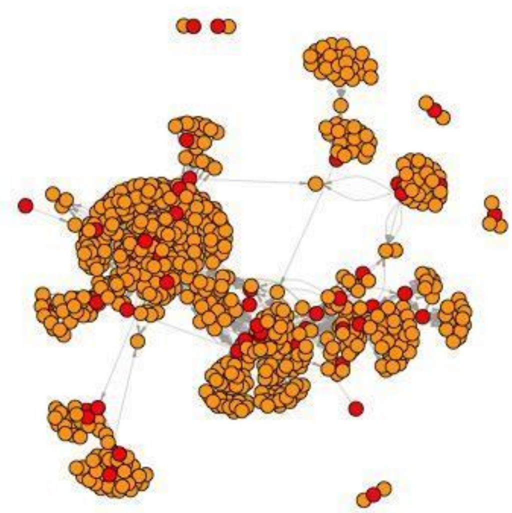
Entire Network: 1999-10



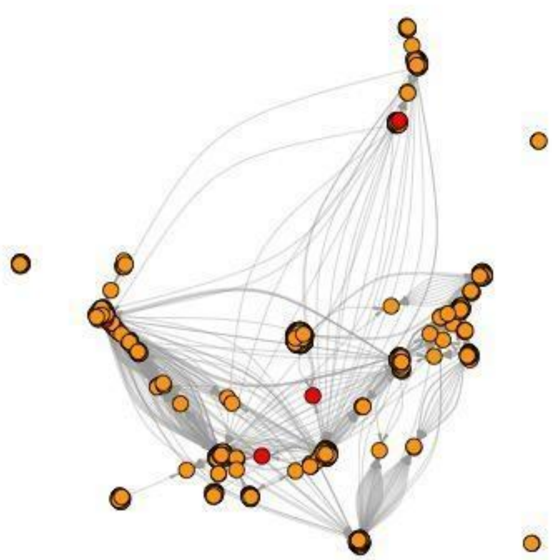
Entire Network: 1999-11



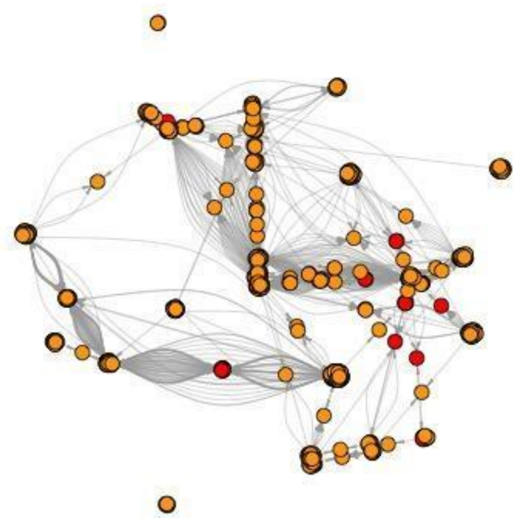
Entire Network: 1999-12



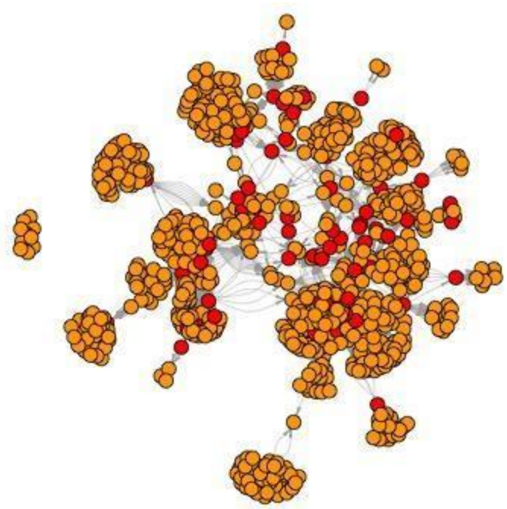
Entire Network: 2000-01



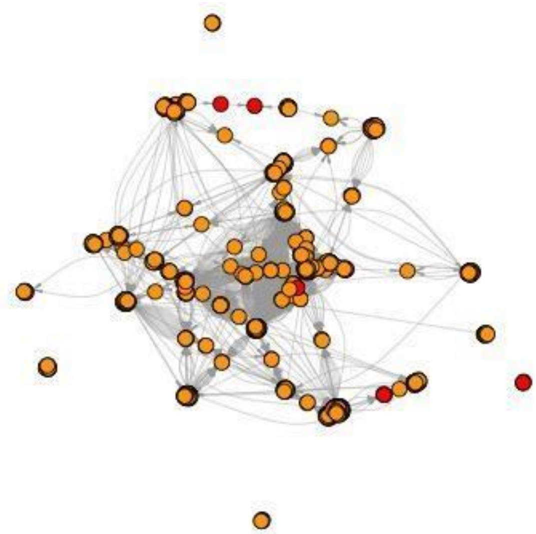
Entire Network: 2000-02



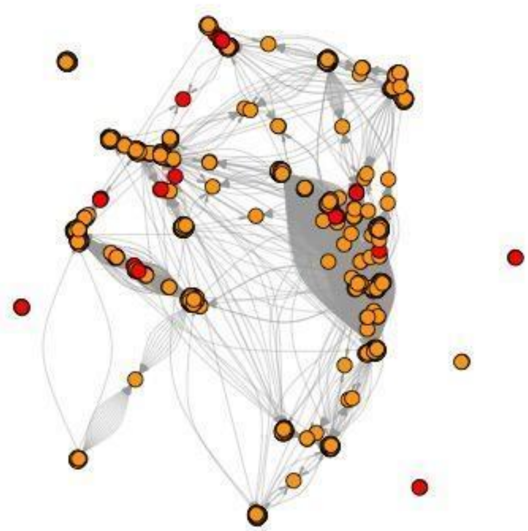
Entire Network: 2000-03



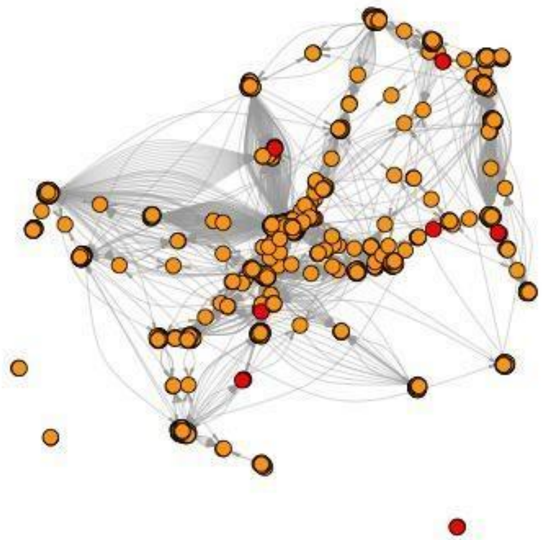
Entire Network: 2000-04



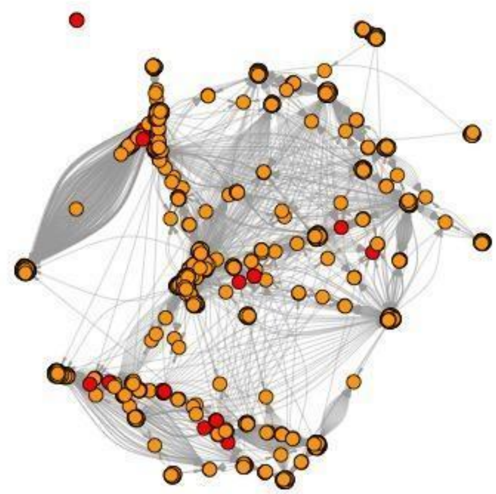
Entire Network: 2000-05



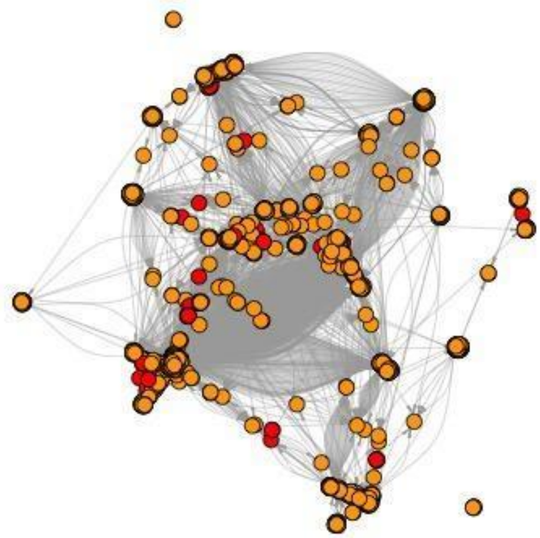
Entire Network: 2000-06



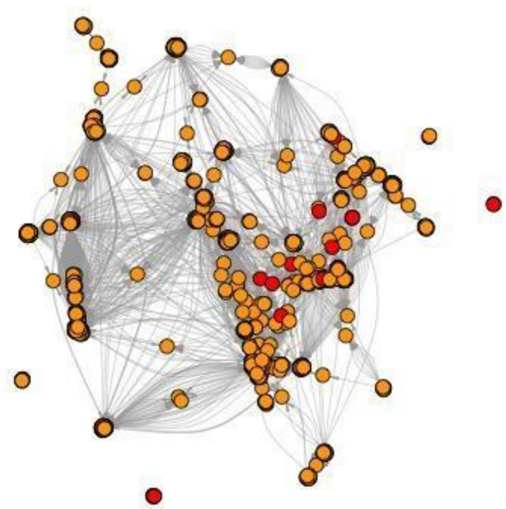
Entire Network: 2000-07



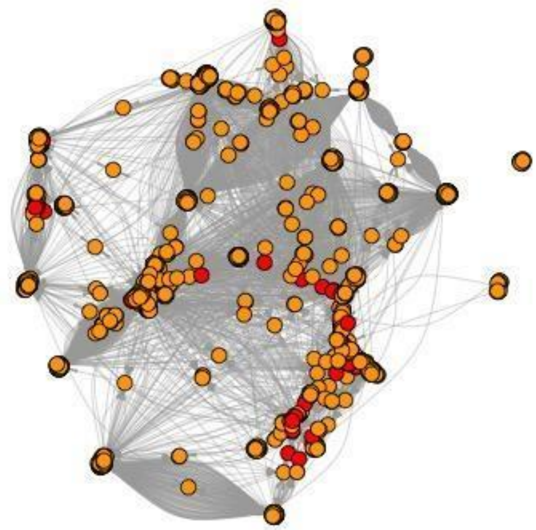
Entire Network: 2000-08



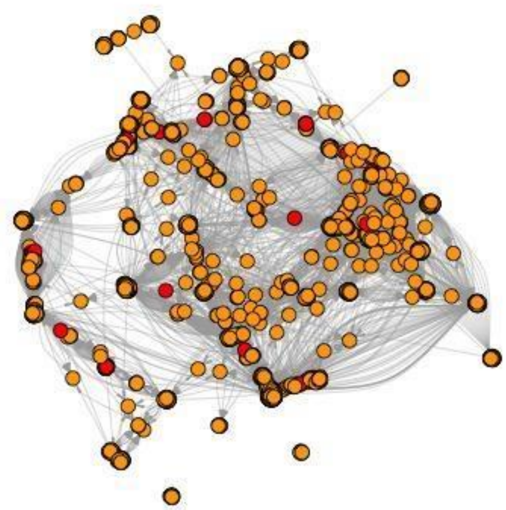
Entire Network: 2000-09



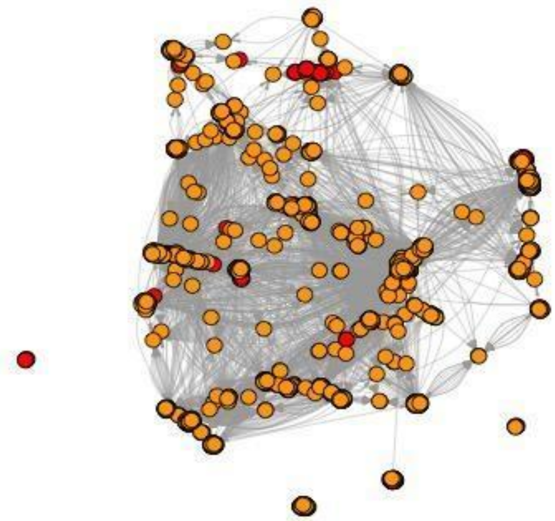
Entire Network: 2000-10



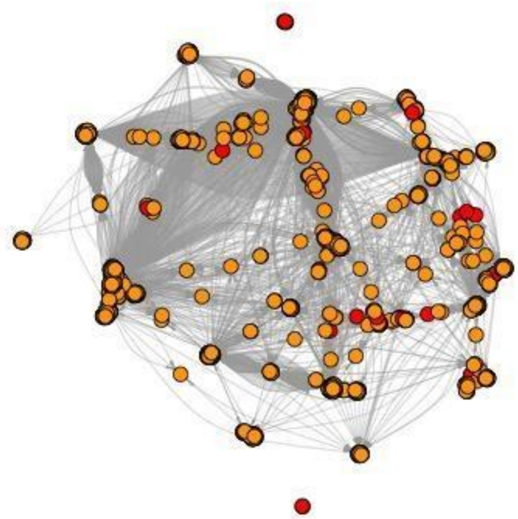
Entire Network: 2000-11



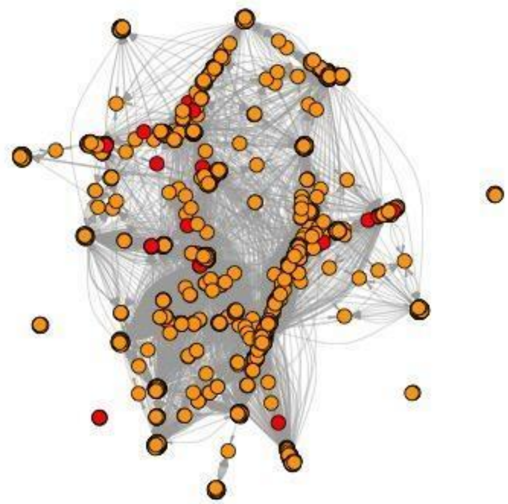
Entire Network: 2000-12



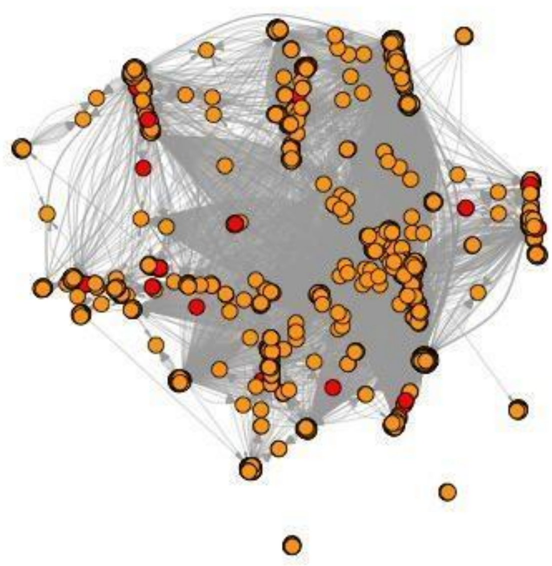
Entire Network: 2001-01



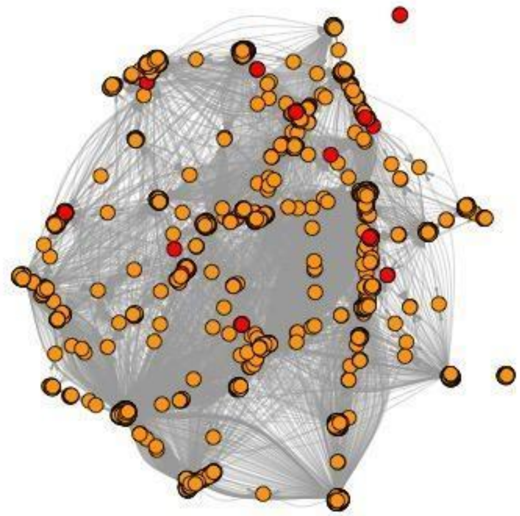
Entire Network: 2001-02



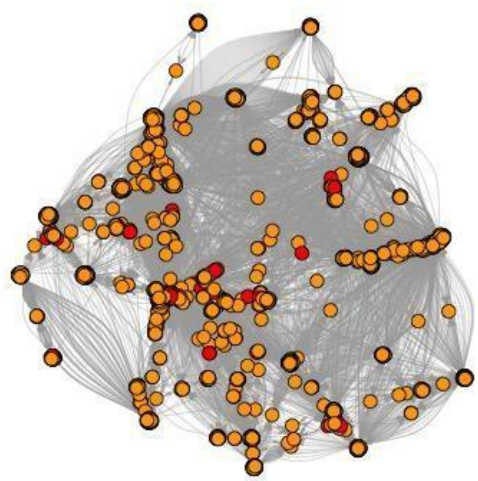
Entire Network: 2001-03



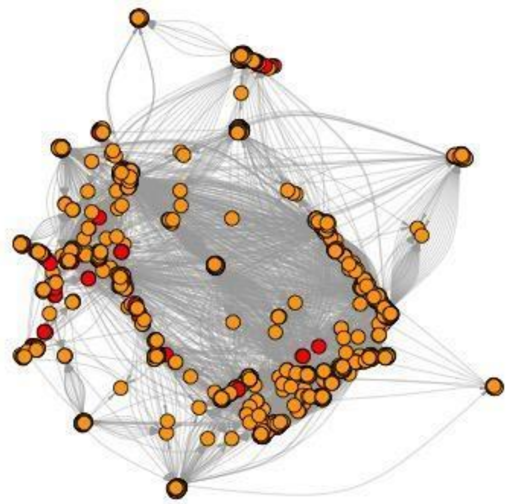
Entire Network: 2001-04



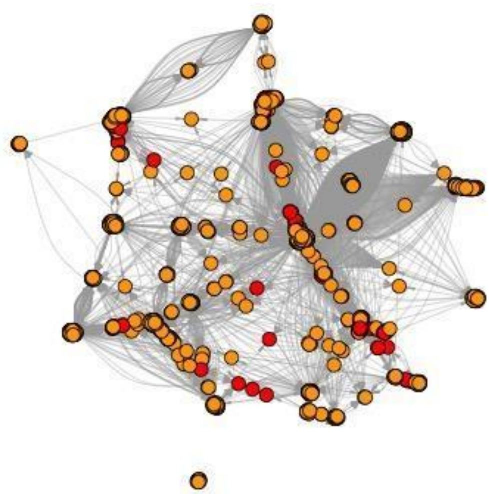
Entire Network: 2001-05



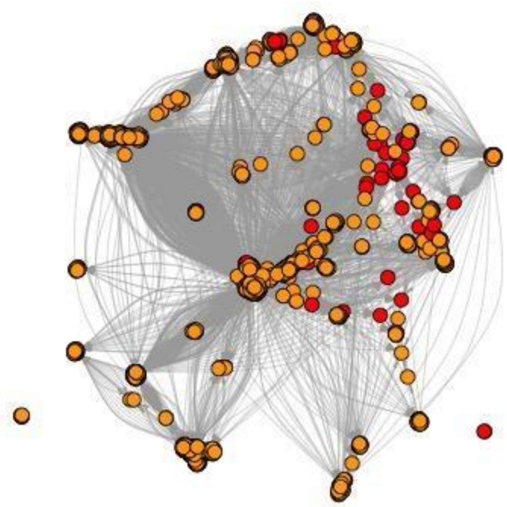
Entire Network: 2001-06



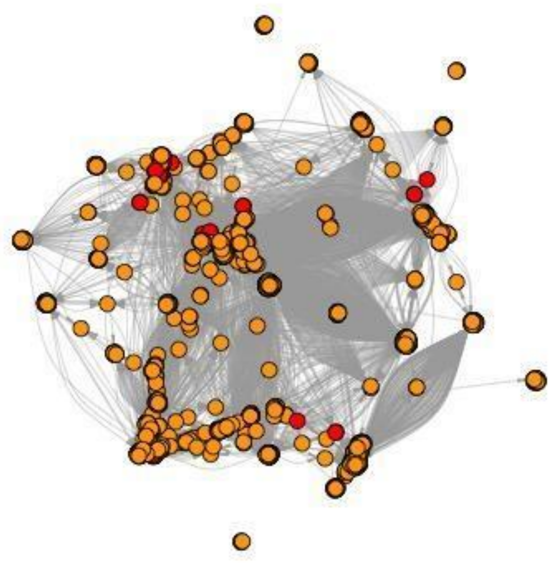
Entire Network: 2001-07



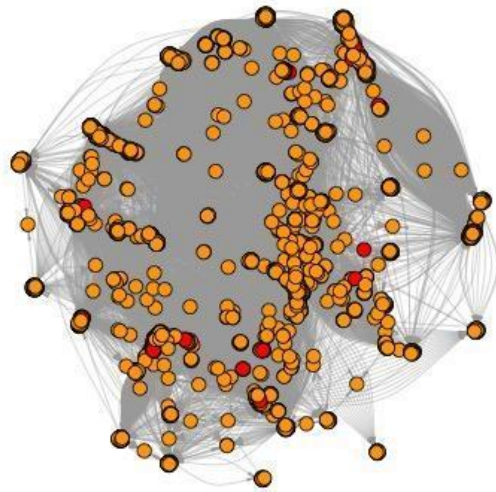
Entire Network: 2001-08



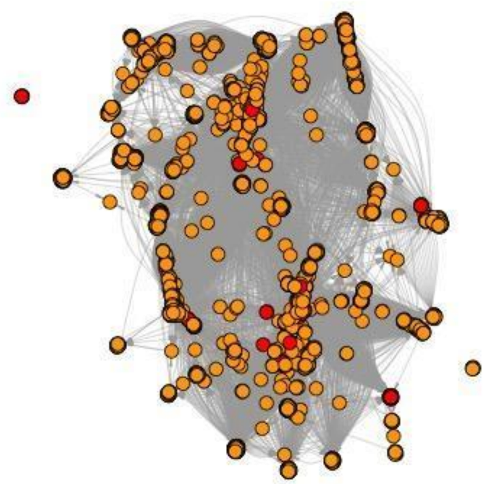
Entire Network: 2001-09



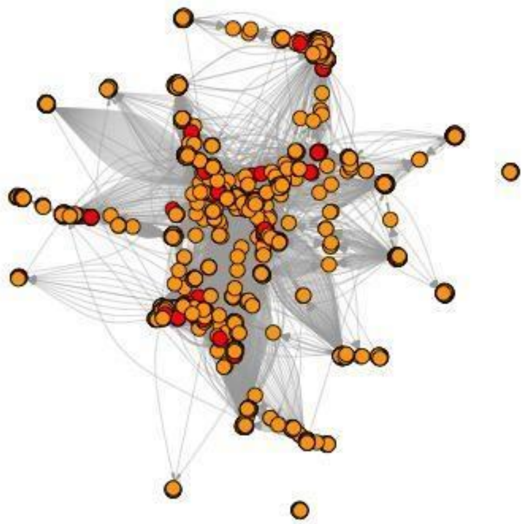
Entire Network: 2001-10



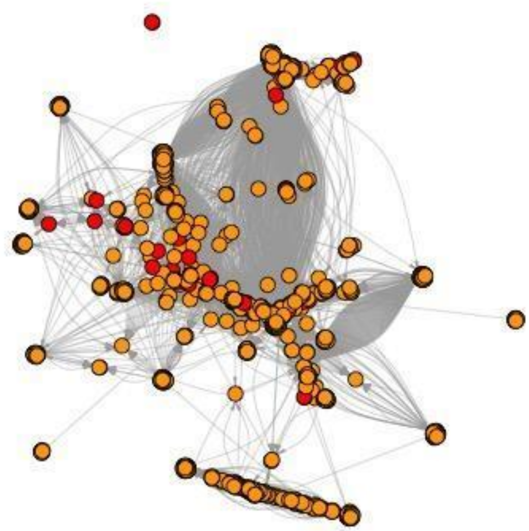
Entire Network: 2001-11



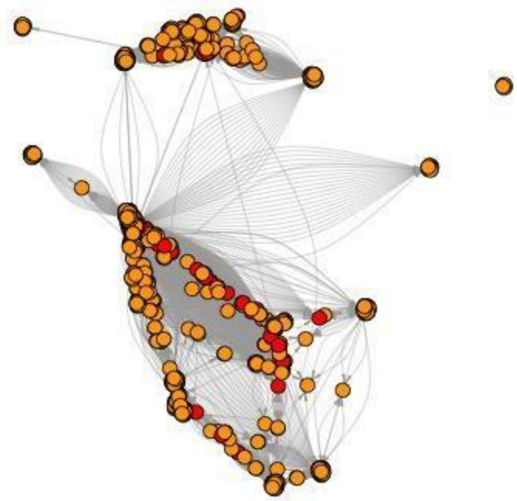
Entire Network: 2001-12



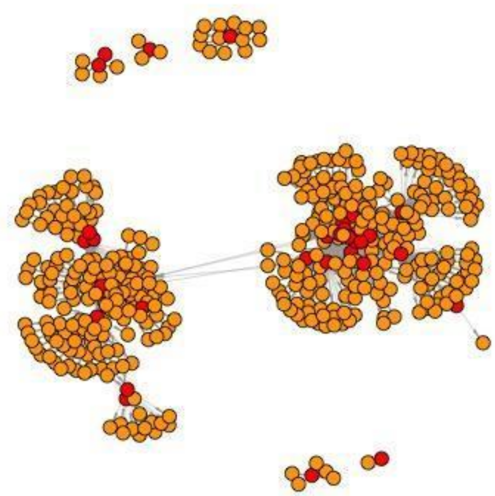
Entire Network: 2002-01



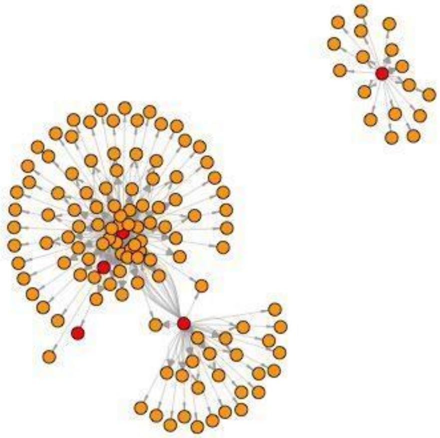
Entire Network: 2002-02



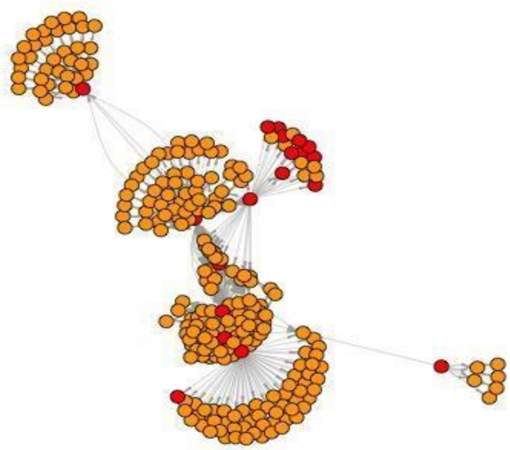
Entire Network: 2002-03



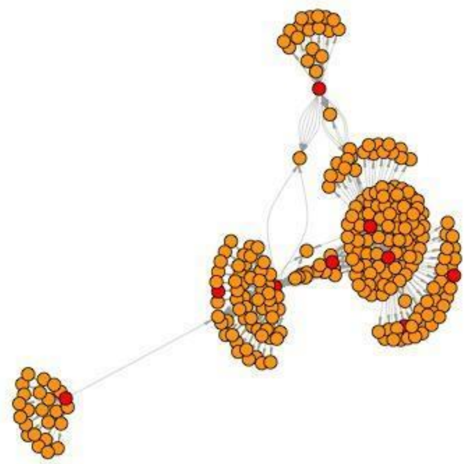
Entire Network: 2002-04



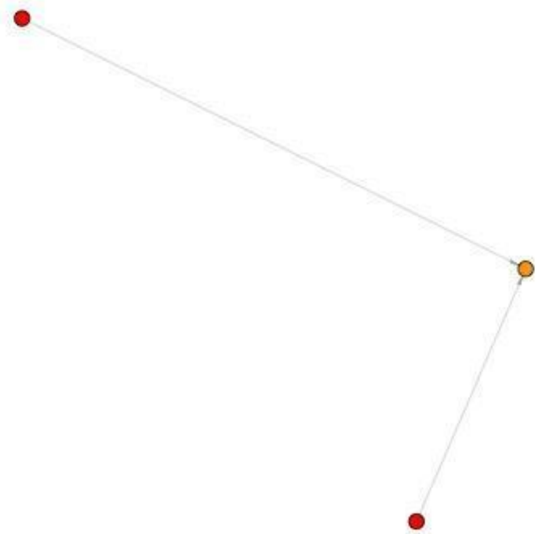
Entire Network: 2002-05



Entire Network: 2002-06



Entire Network: 2002-09



Appendix III - Kenneth Lay's network -Edge Lists at "2000-07", "2001-09", "2001-11" and "2001-12"

From	To	count
j.kaminski@enron.com	kenneth.lay@enron.com	2
j.kaminski@enron.com	steven.j.kean@enron.com	2
steven.j.kean@enron.com	j.kaminski@enron.com	2
steven.j.kean@enron.com	kenneth.lay@enron.com	2

July 2000 ("2000-07")

From	To	count
richard.shapiro@enron.com	steven.j.kean@enron.com	20
steven.j.kean@enron.com	greg.whalley@enron.com	16
steven.j.kean@enron.com	kenneth.lay@enron.com	8
steven.j.kean@enron.com	richard.shapiro@enron.com	4
steven.j.kean@enron.com	stanley.horton@enron.com	4
michelle.cash@enron.com	steven.j.kean@enron.com	2
richard.shapiro@enron.com	greg.whalley@enron.com	2
stanley.horton@enron.com	kenneth.lay@enron.com	2
greg.whalley@enron.com	kenneth.lay@enron.com	1
greg.whalley@enron.com	steven.j.kean@enron.com	1
kenneth.lay@enron.com	michelle.cash@enron.com	1
richard.shapiro@enron.com	kenneth.lay@enron.com	1
stanley.horton@enron.com	greg.whalley@enron.com	1
stanley.horton@enron.com	steven.j.kean@enron.com	1

September 2001 ("2001-09")

From	To	count
richard.shapiro@enron.com	steven.j.kean@enron.com	33
steven.j.kean@enron.com	kenneth.lay@enron.com	24
steven.j.kean@enron.com	greg.whalley@enron.com	23
richard.shapiro@enron.com	louise.kitchen@enron.com	22
steven.j.kean@enron.com	james.derrick@enron.com	16
steven.j.kean@enron.com	liz.taylor@enron.com	14
liz.taylor@enron.com	louise.kitchen@enron.com	13
liz.taylor@enron.com	steven.j.kean@enron.com	8
liz.taylor@enron.com	james.derrick@enron.com	7
liz.taylor@enron.com	kenneth.lay@enron.com	7
louise.kitchen@enron.com	greg.whalley@enron.com	6
steven.j.kean@enron.com	richard.shapiro@enron.com	6
richard.shapiro@enron.com	greg.whalley@enron.com	5
richard.shapiro@enron.com	kenneth.lay@enron.com	5
louise.kitchen@enron.com	richard.shapiro@enron.com	4
steven.j.kean@enron.com	louise.kitchen@enron.com	3
greg.whalley@enron.com	liz.taylor@enron.com	2
greg.whalley@enron.com	tom.donohoe@enron.com	2
kenneth.lay@enron.com	greg.whalley@enron.com	2
kenneth.lay@enron.com	james.derrick@enron.com	2
kenneth.lay@enron.com	steven.j.kean@enron.com	2
liz.taylor@enron.com	greg.whalley@enron.com	2
susan.w.pereira@enron.com	kenneth.lay@enron.com	2
tom.donohoe@enron.com	greg.whalley@enron.com	2
louise.kitchen@enron.com	kenneth.lay@enron.com	1
louise.kitchen@enron.com	steven.j.kean@enron.com	1
tom.donohoe@enron.com	kenneth.lay@enron.com	1

November 2001 ("2001-11")

From	To	count
danny.mccarty@enron.com	kenneth.lay@enron.com	2
kenneth.lay@enron.com	richard.shapiro@enron.com	2
kenneth.lay@enron.com	e..haedicke@enron.com	1

December 2001 ("2001-12")