

# Modeling Interstate Alliances with Constrained Random Dot Product Graphs

David J. Marchette<sup>1</sup> and Carey E. Priebe<sup>2</sup>

<sup>1</sup>Naval Surface Warfare Center, Code Q21, Dahlgren, VA 22448

<sup>2</sup>Applied and Mathematical Statistics, Johns Hopkins University, Baltimore, MD 21218

## Summary

A model of random graphs, the random dot product graph (RDPG) is described. This model is well suited to social networks, since it defines the edges in the graph in terms of a vector of “attributes”. The edge probabilities are modeled as the dot product of vectors associated with the vertices. A small set of distinct vectors is used, allowing the automatic grouping of vertices according to their attributes. We discuss various issues of model fitting and model selection for the reduced vector set version of the RDPG. We extend the basic model to model time series of graphs, and illustrate the model through application to a time series of graphs defined by the alliances between nation states.

**Keywords:** Random Graphs, Least Squares, Social Networks, Time Series of Graphs

## 1 Motivation

A graph is a pair  $G = (V, E)$ , where  $V = \{v_1, \dots, v_n\}$  is a set of vertices and  $E$ , the edge set, is a set of unordered pairs of distinct vertices (we do not allow loops in our graphs). The *order* of the graph is  $|V|$  and the *size* is  $|E|$ . As is the convention, we will write  $vw$  for  $\{v, w\} \in E$ . In this paper we will consider a time series of graphs, that is, a collection of graphs indexed by (discrete) time.

There are two viewpoints that one could take on time series of graphs. Dynamic graphs are graphs that change in time. Usually this means that the edge set changes in time: at each time step an edge may be added or removed. We will not concern ourselves with dynamic graphs in this sense. Instead, we assume that there is a (possibly time dependent) random model from which the graphs are drawn, and at each time  $t$  a new graph is drawn from this model (possibly depending on graphs drawn prior to  $t$ ). Throughout our discussion, the vertex set  $V$  will be fixed, and it is labeled, meaning that there is a known 1-1 correspondence between the vertices of the graph at time  $t$  and the one at time  $t + 1$ , allowing us to match up the vertices over time.

In this paper we focus primarily on model selection in the Random Dot Product Graph (RDPG) model of Kraetzl et al. [2007] and Scheinerman and Tucker [2007]. The RDPG is a random graph model in which a vector of attributes is assigned to each vertex, such that the probability of an edge is equal to (a function of) the dot product of the vectors. In this paper we consider the following model selection problem: can we choose a small number ( $K \ll |V|$ ) of distinct vectors to fit to the graph, so that each vertex is associated with one of these vectors? In this manner we obtain a natural grouping of vertices according to the selection of the vectors. There are several reasons to consider this problem:

1. Regularization: we reduce (dramatically) the number of parameters of the model. This reduces variance.
2. Interpretation: the much smaller number of vectors can make the interpretation and understanding of the graph much easier.
3. Clustering: the model provides a natural clustering of the vertices, which can be exploited to better understand the natural groupings of the data underlying the graph.

The selection of dimension ( $d$ ) for the vectors is also necessary. Our algorithm for fitting the vectors is a least squares algorithm, relying on the spectral decomposition of the adjacency matrix. Scree plots can be used to select the dimension  $d$ . We suggest that practitioners err on the side of smaller values for  $d$ , unless there is a compelling a priori or scientific reason to choose a

Table 1: Alliance codes in the alliance dataset.

0 or NA	No alliance	
1	Defense pact	intervene militarily if partner attacked
2	Neutrality	remain militarily neutral if partner attacked
3	Nonaggression pact	consultation and/or cooperation in a crisis

larger dimension. This aids in the interpretation (and visualization) of the model, as well as providing further regularization.

We consider a time series of graphs defined in terms of alliances between countries. We investigate data representing alliances between a total of 173 nations, collected from 1816 to 2000. The data are available at <http://www.correlatesofwar.org/datasets.htm> (see Gibler and Sarkees [2004]), and we provide a processed dataset (with more attributes than are considered in this paper) at [www.ams.jhu.edu/~marchette/igo.tgz](http://www.ams.jhu.edu/~marchette/igo.tgz). In this latter dataset, there are a total of 214 nations, but some of these do not have alliances (they do have other attributes such as trade). For each pair of nations, alliance is coded as in Table 1. While the edges are colored by alliance type (see Table 1), we will consider only the simplified graph with binary edges: existence or absence of an alliance.

We will construct an alliance graph for each year. Thus, the vertices of the graph will consist of the 173 nations for which an alliance is present in the data, and there will be an edge between two vertices if the corresponding nations had an alliance during the year. This produces a time series of graphs. We will describe a model of random graphs and apply this model to the time series.

## 2 Random Dot Product Graphs

We do not believe that most interesting random graphs have independent edges, and so we seek a model that relaxes this requirement. A simple model that has some interesting (and possibly relevant) properties is the random dot product model.

A random dot product graph (RDPG) is a random graph model containing Erdős-Renyí random graphs (Bollobás [2001]) as a sub-model, in which each vertex  $v$  is assigned a vector  $x_v \in \mathbb{R}^d$ . The probability of an edge from  $v$  to  $w$  is a function of the dot product of the vectors:

$$p_{vw} = f(x'_v x_w).$$

In this paper we will set  $f$  to be a simple threshold:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

The vectors  $x_v$  are fixed, and new graphs are drawn from the collection of all graphs on  $n$  vertices according to the edge probabilities defined above.

It should be noted that in addition to being a generalization of the Erdős-Renyí random graph, the RDPG is a generalization of random intersection graphs (Karonski et al. [1999]) and a sub-model of latent position models (Hoff et al. [2002], Hoff [2005]). This model is also very similar to the stochastic blockmodels of Nowicki and Snijders [2001]. In that work, the vectors are one-dimensional (scalars) and are constrained to be from a finite collection of values. Thus, the attribute vectors can be considered as class (or cluster) labels, and the probability of an edge is dependent on the labels for each vertex.

One motivation for this model comes from social networks applications. It seems reasonable to assume that individuals have a collection of attributes which (to a large extent) define the probabilities of connection with others. For example, in a friendship graph the attributes might correspond to interest areas and personality, and the probability of an edge (friendship) between two individuals could be largely driven by the overlap of their interests and the compatibility of their personalities. This is admittedly a simplification. It is not clear that in real problems such attribute vectors can be reliably identified. However, the model is a reasonable approximation, and one can fit the model without assuming any particular interpretation for the vectors. This will be our philosophy in this work.

One reason for choosing the RDPG over the distance model described in Hoff et al. [2002] is parsimony for a certain class of graphs. We are interested in finding groupings within the vertices, and the RDPG allows us to define these groups as being those which share the same vector. While (with suitable assumptions on the probabilities) the two models can be used to define the same random graphs, it turns out that the dimension  $d$  of the vectors needed in the distance model can be much larger for certain cases than is necessary for the dot product model.

Although computationally challenging, fitting the vectors to a given graph or set of graphs is relatively straightforward. Scheinerman (Scheinerman and Tucker [2007], see also Scheinerman [2005]) gives a linear algebra method that tries to minimize the Frobenius norm for the edge probabilities. The edge probabilities in the model are conditionally independent (conditioning on the vectors) and so for a given set of vectors the likelihood function  $L$  is

as follows:

$$L(x_1, \dots, x_n; A) = \prod_{i \neq j} (x'_i x_j)^{a_{ij}} (1 - x'_i x_j)^{1 - a_{ij}}. \quad (1)$$

Here,  $a_{ij}$  is the  $ij$ th element of the adjacency matrix  $A$ .

We will use a variant of the linear algebra method for fitting the vectors to a graph. The iterative algorithm, for fitting an order  $n$  graph with adjacency matrix  $A$  is:

1. Set  $D$  to be a vector of  $n$  zero.
2. Do
  - (a) Compute the spectrum of  $A + \text{diag}(D)$ .
  - (b) Set  $X = U\sqrt{e}$ , where  $U$  is the  $n \times d$  matrix of the top  $d$  eigenvectors and  $e$  is the vector of the top  $d$  eigenvalues, with negative eigenvalues set to 0.
  - (c) Set  $D = \text{diag}(XX')$ .
3. While  $D$  is changing, or up to a user supplied maximum number of iterations.
4. Return  $X$ .

In this algorithm, the function `diag` is like the R command of the same name: when applied to a matrix it returns the diagonal, and when applied to a vector it returns a diagonal matrix with the vector on the diagonal. This algorithm implements a least-squares fit, minimizing the Frobenius norm (sum of squares of the differences of the estimated probability matrix to the observed matrix). As such, it is much faster to produce a fit than a likelihood based method, and can be applied to weighted graphs and to estimated probability matrices as well as to single graphs, as will be seen below.

We consider a modification of this algorithm that fits the model in one step. The idea is to “guess” the values of the diagonal. The guess chosen is the scaled degree. We replace the zero in the  $i$ th diagonal entry with

$$\frac{\text{degree}(v_i)}{n - 1}.$$

The matrix of vectors is now the scaled eigenvectors of this augmented matrix. See Figure 1 for evidence of the performance of this new algorithm. As can be seen from the figure, the 1-step algorithm is no worse, in this simulation, than the iterative algorithm (and one could make the case that it actually performs better).

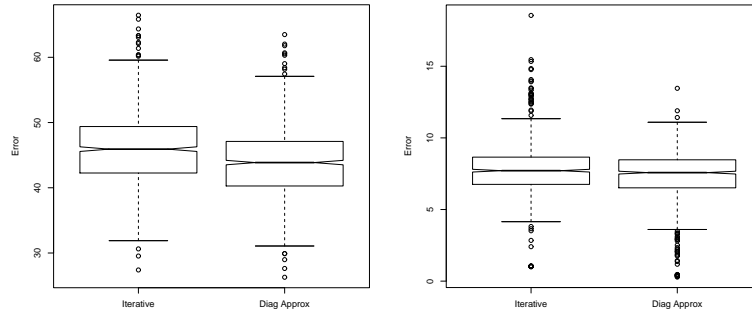


Figure 1: Comparison of the iterative algorithm with the augmented 1-pass algorithm. 1000 Monte Carlo replicates of 100 2-dimensional vectors were drawn from two distributions. Left plot: the vectors were sampled uniformly from the portion of the unit disk in the first quadrant. Right plot: the vectors were sampled uniformly from the portion of the unit disk in the first quadrant, within 5 degrees of the axes. A corresponding instantiation of the random dot product graph was formed, and the two algorithms were run. The error is the sum of the squared difference of the (lower triangle of the) estimated probability matrix and that of the true probability matrix. In this simulation, no thresholding was performed on the dot products to insure that they remained within the range  $[0, 1]$ .

The argument for this is the following. While we would like to put  $x'_i x_i$  in the  $i$ -th diagonal entry, we approximate this as  $x'_i \bar{x}_{(i)}$ , where  $\bar{x}_{(i)}$  is the mean of the other vectors. This comes from the observation that our estimate of the degree of the  $i$ -th node divided by  $n - 1$ , given the vectors, is:

$$\frac{1}{n-1} \sum_{j \neq i} x'_i x_j = \frac{x'_i}{n-1} \sum_{j \neq i} x_j = x'_i \bar{x}_{(i)}.$$

If the vectors are identically distributed, this estimates the diagonal element  $x'_i x_i$  by  $x'_i E\mathbf{x}$ . To illustrate the performance of this method for an extreme case, see Figure 1. An error of 45 corresponds to an average element-wise error of 0.0091, while an error of 7 corresponds to an average element-wise error of 0.0014.

For a graph  $G = (V, E)$ , a partition is a collection of subsets of vertices  $P = \{P_1, \dots, P_K\}$  such that  $P_i \cap P_j = \emptyset$  for  $i \neq j$  and  $\cup P_i = V$ . We will assume the number of partition cells  $K$  is known a priori. In a social network context, these groups might be defined via club membership, interest groups, religious affiliation, or some unobserved grouping that one would like

to discover. Thus, we are provided with a set of partition labels  $\mathcal{L}$ , and seek a map  $h : V \rightarrow \mathcal{L}$  to maximize the likelihood.

Partitioning the vertices into groups, such that the edge probabilities depend only on the groups, is related to stochastic equivalence in the social network literature: all vertices within any group are stochastically equivalent. This is also called stochastic blockmodeling, or stochastic blockstructures. See Nowicki and Snijders [2001] and Wasserman and Faust [1994].

The approach we take in this paper is the well-known  $k$ -means algorithm. We use the similarities defined by the dot product, rather than Euclidean distance in the algorithm, but otherwise this is the standard  $k$ -means algorithm. Other approaches may be used for clustering, and this is an area of future research.

We have several model selection problems in this work. First, one must decide on the dimensionality  $d$  of the attribute vectors  $x_i$ . In a particular application, this should be in part driven by scientific considerations. One may hypothesize that there are two main factors defining the relationships, and thus choose  $d = 2$  and observe the model fit. It should be noted that we want to pick  $d$  as small as appropriate, due to the additional variance in the estimators as the dimension increases, but that we also want to ensure that  $d$  is large enough to properly account for the variation in the graph.

The basic statistics of the graphs are depicted in Figure 2. Some structure is immediately obvious, such as the changes in density at the years 1887 and 1937. The number of alliances, as well as the number of countries engaging in alliances, is increasing in time, while the density of edges in the graph tends to decrease after a peak in the early 1900s.

For the purposes of illustration, we consider the estimated probability matrix defined by averaging all the adjacency matrices. As will be obvious later on, one would not really do this, as the graphs are definitely not stationary over the full range of time. However, this will provide a single picture of the data, which will in turn provide some interesting observations. Consider Figure 3. The left plot shows two nearly orthogonal groups of countries: one consisting of most of western Europe, the other the Caribbean and some of South and Central America. These groups tend to have alliances among themselves, but not between the groups. The United States and Canada do bridge these groups, as is indicated by their positions in the plot. The large group of points near zero are mostly countries that either have very few alliances or which are not present for most of the time covered by the data.

The scree plot on the right indicates that a dimension of around  $d = 5$ , or perhaps even smaller, might be appropriate for these data. As mentioned above, this crude look at the data has several problems, and we will not discuss it further.

Another problem is the choice of  $K$ . Many papers have been written about

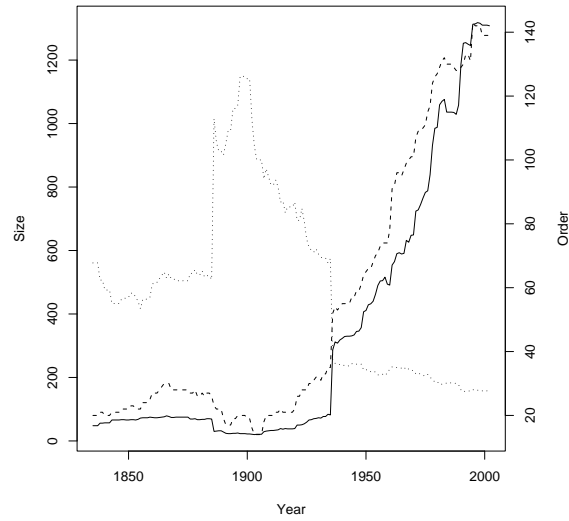


Figure 2: Sizes (solid), orders (dashed) and density (dotted, on a scale from 0.0 to 1) of the graphs defined by the alliances. The large jump in density occurs at 1887 and the drop is at 1937.

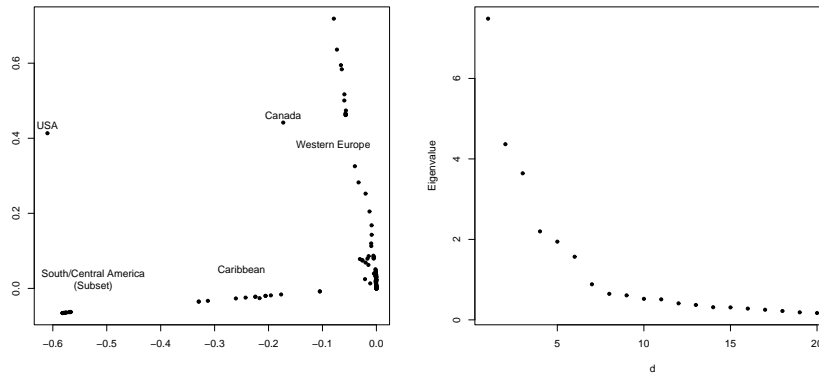


Figure 3: 2-d vectors (left) computed on the probability matrix defined as the average of the adjacency matrices of the alliance graphs. Scree plot (right) for the first 20 eigenvalues.

model selection similar to this one in various guises, but the basic idea of most model selection methodologies comes down to choosing an appropriate penalty for model complexity. We will try several values of  $K$  and plot the error, scaled by  $\binom{n}{2}$ , as a function of  $K$ . An appropriate penalty is an area of future research; in this paper we will use the plots and investigation of the graphs to make a subjective assessment of the appropriate value of  $K$ . Note that for  $K = 1$  the graph is an Erdős-Renyí random graph.

We will be using a window of width  $W$  on the time axis, and will average the graphs within the window, moving the window in time by a step of 1 year. The vectors will be computed on the averaged adjacency matrix, as was done above, but in this case the matrices are only averaged over the  $W$  most recent graphs. So, in addition to  $d$  and  $K$ , we need to select  $W$ .

To investigate the appropriate dimension for the alliance data, we plot scree plots (with all eigenvalues divided by the maximum eigenvalue to scale them equivalently) for each graph, and for windows of length 5, 10 and 20. See Figure 4. We would like to use a single  $d$  for all graphs, and these plots indicate that, while not appropriate for all graphs, a value of  $d = 5$  seems to be a good choice, particularly for the graphs averaged over a window.

Rather than looking at the scree plots and “eye-balling” the dimension, one can use an automatic method such as the profile likelihood approach described in Zhu and Ghodsi [2006]. The results for a window of size  $W = 5$  are shown in Figure 5. If we want to choose a value for  $d$  that is adequate for all the graphs, we would choose the maximum in this plot ( $d = 9$ ). The value of  $d = 5$  works well for most of the graphs. Interestingly, there are many for which  $d = 2, 3$  are appropriate, which is convenient for plotting the vectors.

Figure 6 shows the scaled error for the model under various choices of dimension. The error is monotonic in dimension, and so one can easily see that, for much of the graph there is little reason to choose a dimension higher than 5. There is some reason to consider higher dimensional models in the latter half of the 20th century, but our choice of  $d = 5$  seems a reasonable compromise for these data.

The scree plots also seem to indicate little value in a window size of larger than  $W = 10$ ; however, we will investigate this question further while looking at the choice of  $K$ . Figures 7 and 8 show the errors (Frobenius norm divided by  $\binom{n}{2}$ ) of the estimated probability matrix under various choices of  $K$  and  $W$ .

A much more extensive study would look at each window separately, and investigate the trade-off between window size and dimension. By investigating the changes in the model, one might obtain insight into how these changes are reflected in the graphs, which in turn may give insight into the history. In effect, the dimension is an estimate of the number of factors needed to explain

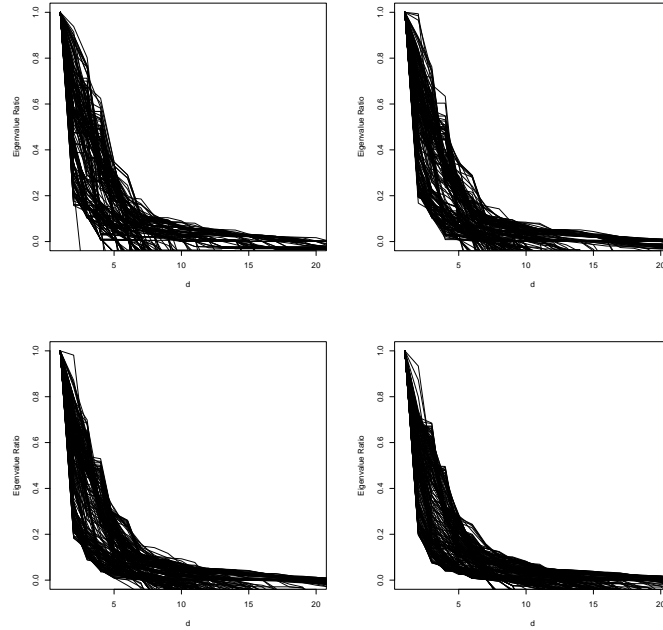


Figure 4: Scree plots for all 185 graphs (top left) and for windows of size  $W = 5$  (top right),  $W = 10$  (bottom left) and  $W = 20$  (bottom right).

the graph, and this could have important historical implications, provided one could then identify these factors. We believe that further research is warranted.

We will not choose a single value of  $K$  from these plots. Instead, we note that  $K = 1$  (Erdős-Renyí) is clearly not appropriate for any window size, and that  $K = 3, 4$  seem to be reasonable choices for most of the time. These values of  $K$  also produce nearly the same errors, and it would be interesting to investigate those years in which the errors are substantially different, as well as when they are the same.

A moving window analysis such as this assumes short-time stationarity of the data. The size of the window controls the bias-variance trade-off. Larger windows provide better parameter fits at the expense of quick detection of a change. One should certainly consider using different windows in regions in which different amounts of stability (stationarity) appear to be present. Again for illustration purposes, we show results using a window size of  $W = 10$ , which seems a reasonable compromise, in the next section.

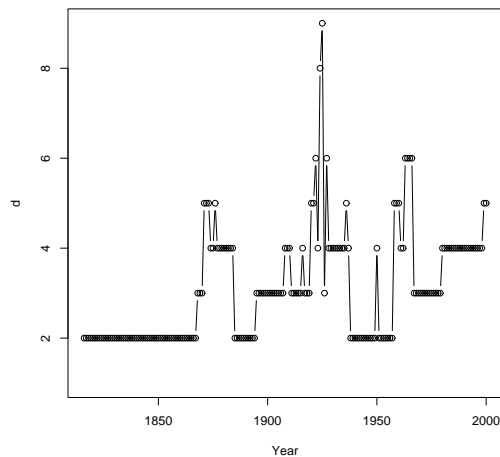


Figure 5: Dimension chosen by the profile likelihood method for a window of  $W = 5$ . The values above  $d = 5$  occur for years 1922, 1924, 1925 and 1963–1966.

### 3 Results

Prediction error seems to be a reasonable metric to use in this type of analysis. This measures how well the current model predicts the next graph. In Figures 6–8 we considered how well the model explains the data used to fit the model. Prediction error indicates how well the model (the probability matrix defined by the model) predicts the adjacency matrix of the next graph in the time series. By considering the prediction error for each vertex clustering, we can select the appropriate number of clusters. Outliers and change points in the prediction error can provide cues for investigation of changes in the underlying distributions of the graphs.

The prediction error (the scaled Frobenius error between the model’s probability matrix at time  $t$  and the adjacency matrix of the graph at time  $t + 1$ ) is shown in Figure 9. This shows that early on (prior to around 1880) the  $K = 3, 4$  models are as good as the full model, indicating that we may construct a more parsimonious model with as much information as the full model. After this point the full model performs better, but the  $K = 4, 5$  models seem to do quite well in this latter period.

There has been much of work on dynamic graphs that utilize time series models of one type or another. See Snijders [2005] for a nice discussion of some of this recent work. There is clearly opportunity for more sophisticated

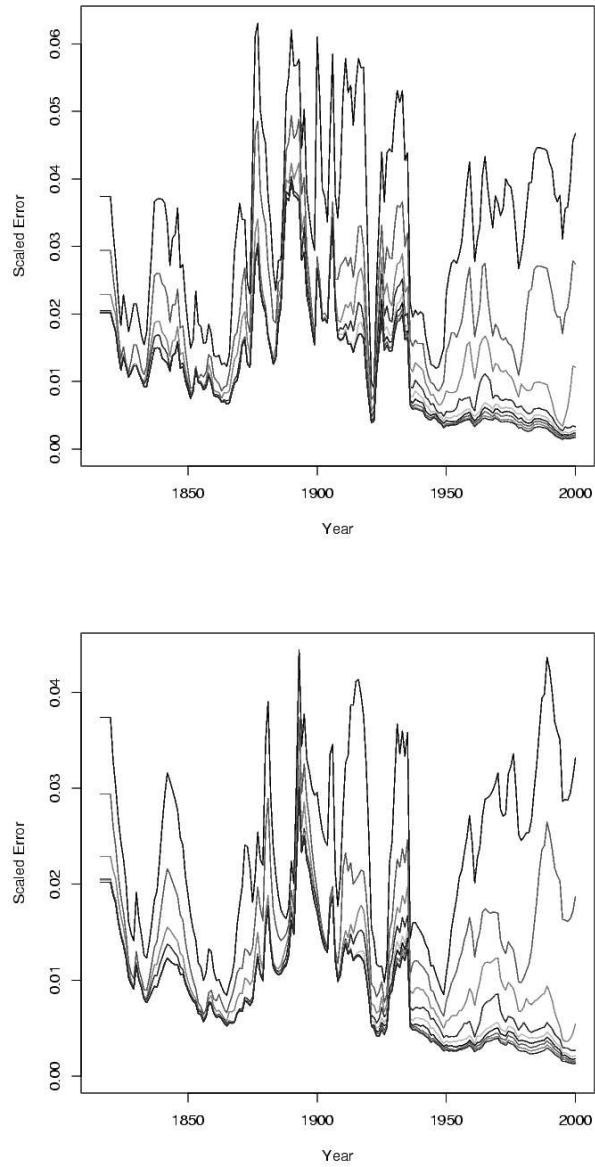


Figure 6: Scaled error for window size of  $W = 5$  (top) and  $W = 10$  (bottom). The curves are the error for a model of dimension  $d = 2, \dots, 10$ . The scaled error corresponds to the average error for an entry in the probability matrix. The curves show decreasing error in  $d$  (in the on-line version these are colored black, red, green, blue, cyan, black, red, green, blue as  $d$  increases from 2 to 10).

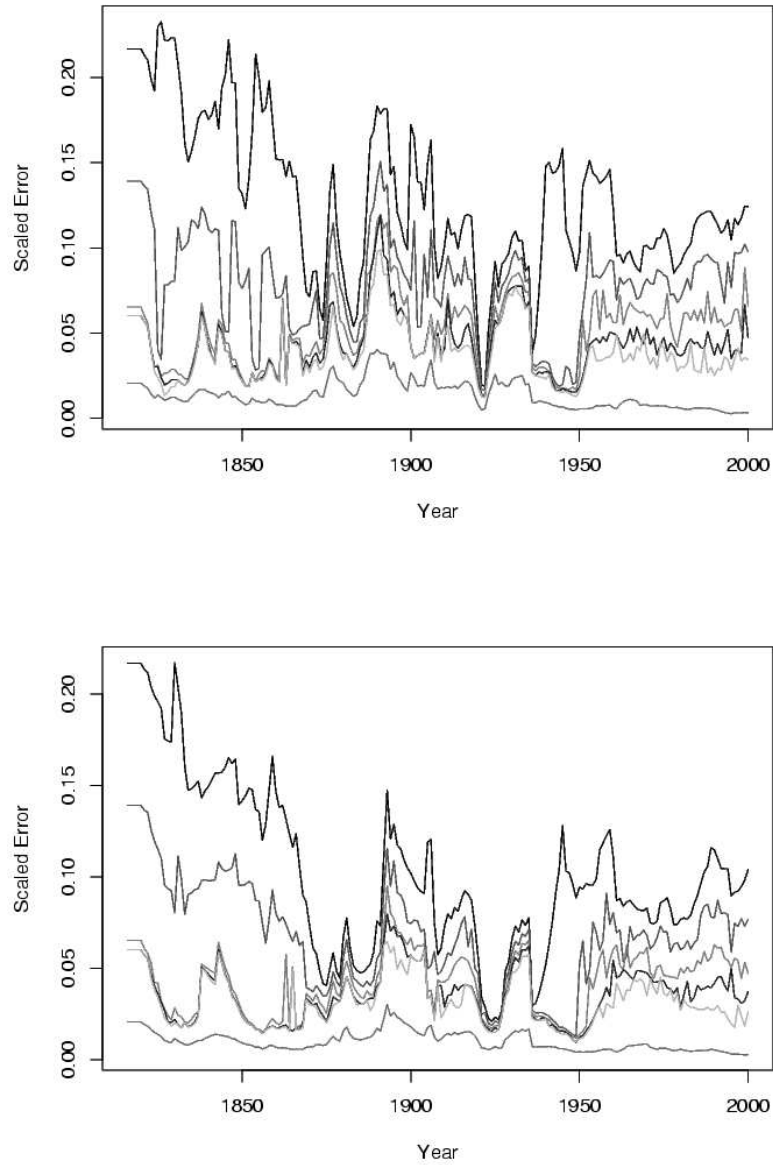


Figure 7: Errors for clusters of  $K = 1, 2, 3, 4, 5$  and no clustering ( $K = n$ ). The curves show decreasing error in  $K$  (in the on-line version these are colored black, red, green, blue, cyan and magenta). The top plot uses a value of  $W = 5$ , the bottom  $W = 10$ . In both plots  $d = 5$ .

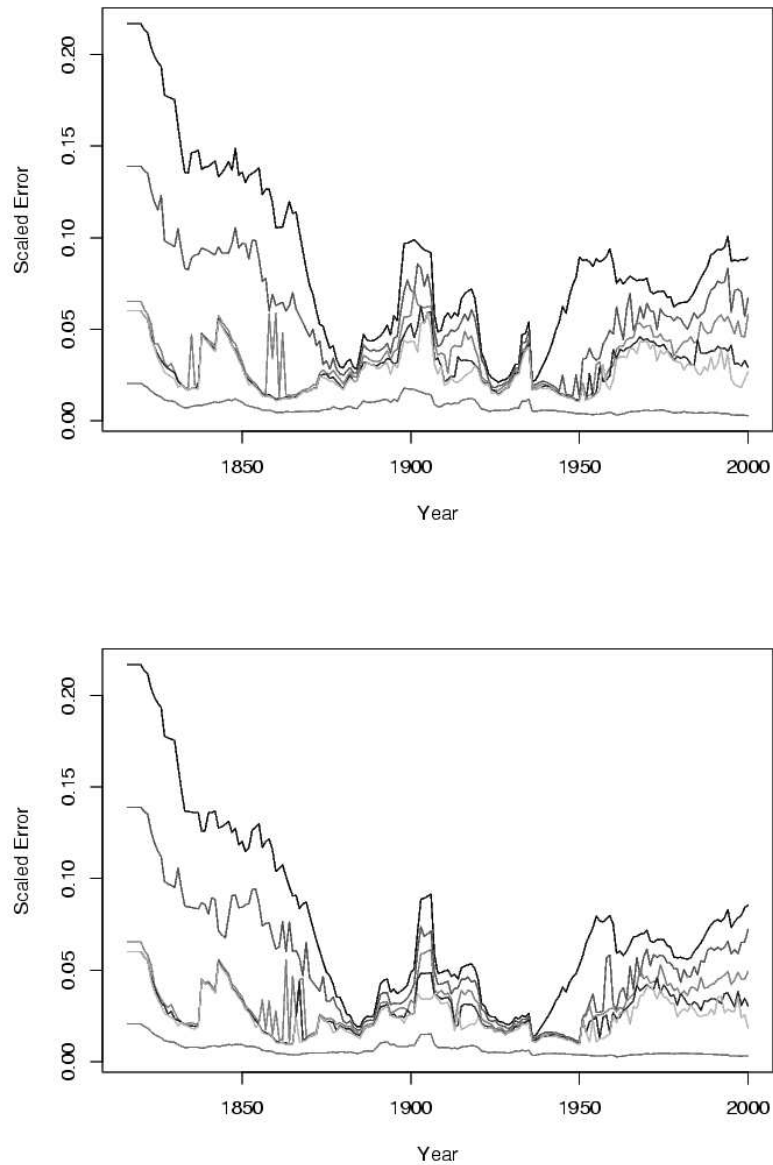


Figure 8: The same plots as in Figure 7 for two larger window widths. The top plot uses a value of  $W = 15$ , the bottom  $W = 20$ . Errors for clusters of  $K = 1, 2, 3, 4, 5$  and no clustering ( $K = n$ ). The curves show decreasing error in  $K$  (in the on-line version these are colored black, red, green, blue, cyan and magenta). In both plots  $d = 5$ .

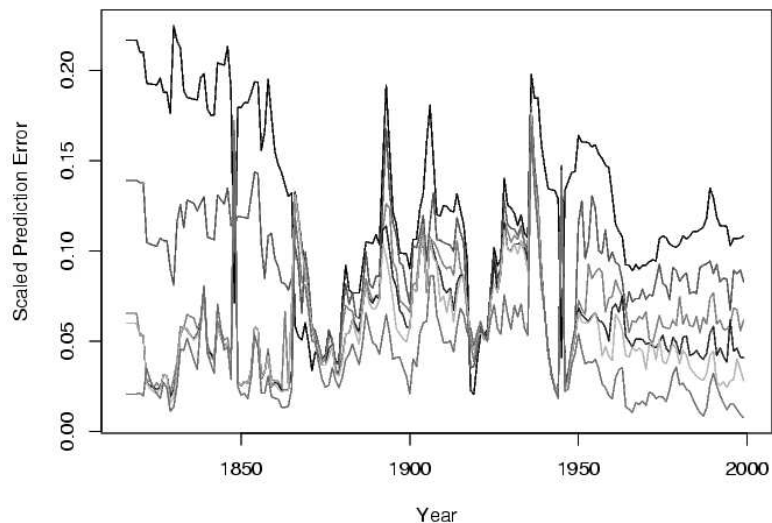


Figure 9: Prediction errors for clusters of  $K = 1, 2, 3, 4, 5$  and no clustering ( $K = n$ ), with a window width of  $W = 10$ . The dimension for the vectors is  $d = 5$ .

models, for instance by placing a time series model on the vectors. This is an area of future research.

The choice of least squares fitting over a maximum likelihood or Bayesian approach is in the interest of fast algorithms. With modern sparse matrix methods (social networks tend to be sparse), the algorithms discussed above can be performed on very large graphs. We have demonstrated that the algorithm can be applied to very large graphs (one simulated graph consisting of one million nodes with an average degree of 100 took less than a minute to produce a fit on a dual core processor). We do not claim that this timing is typical, but we have other anecdotal evidence that large graphs can be easily modeled with the least squares approach. While the alliance graphs are not large, we are interested in determining what this simple model can tell us about these graphs.

Note that throughout these plots (particularly Figure 9) there appear to be several distinct regimes: prior to 1880, the period up until the second world war, and the post-war period. In Figure 10 we have separated the data into four time periods: 1816–1880, 1881–1923, 1924–1948, and 1949–2000. As one can see, for the first period, a value of  $K = 4$  seems to be essentially as

good as possible under this model. For the second period, while there are differences between the  $K = 4$  and  $K = 5$  models, there is little support for choosing the latter over the former. The third period starts with a few years in which the two clustering models ( $K = 4$  and  $K = 5$ ) are slightly worse (in prediction error) than the full model, but all three models have essentially identical prediction errors from 1936 on. Finally, in the last period we see a trend that perhaps is not present in the other periods, and the  $K = 5$  model seems more appropriate. Note that in the first two regions the data appears to be reasonably stationary (as measured by prediction error).

To get an idea of what is going on in these time periods, we again average the adjacency metrics, and fit the appropriate model. Figure 11 shows the scree plots for the four time periods. Note that the first and third time periods suggest a  $d = 3$ , while the second and fourth time periods suggest a larger dimension. We will use  $d = 5$  for these.

From the above analysis, we selected  $K = 4$  for the first three time periods, and  $K = 5$  for the final time period. The last two time periods show fairly strong evidence of nonstationarity, and so any results from the large-scale averaging should be considered to be only an approximation to the information in the time series. In particular, both the trend in the fourth period, the period from 1936–1944, and the year 1945 could bear further investigation. In each case, we show the graph consisting of all edges between country pairs that have an alliance within the period. The estimate of the probability matrix (average of the adjacency matrix) is used to construct a constrained random dot product graph, and the clusters are organized into circles around the center of the plot. Thus, in the Figures 12–15 we can see which countries are allied with which, and which group the algorithm has placed them in. In all graphs, we have removed any isolated vertices.

The first time period, from 1816–1880 is depicted in Figure 12. This graph shows a reasonable characterization of the data. The new world is a separate component of the graph; the German countries (with Russia) form a near clique; the other two clusters consist mostly of the southern/eastern and northern/western parts of Europe. The second graph, depicted in Figure 13, also splits the countries into groups that seem to make sense from a historical standpoint.

## 4 Conclusions

This analysis shows that the constrained RDPG model is a powerful tool for modeling social relationships. By comparing different models, we can learn about the dynamics of the different groups, and better understand the underlying structure of the relationships. The method provides a way to

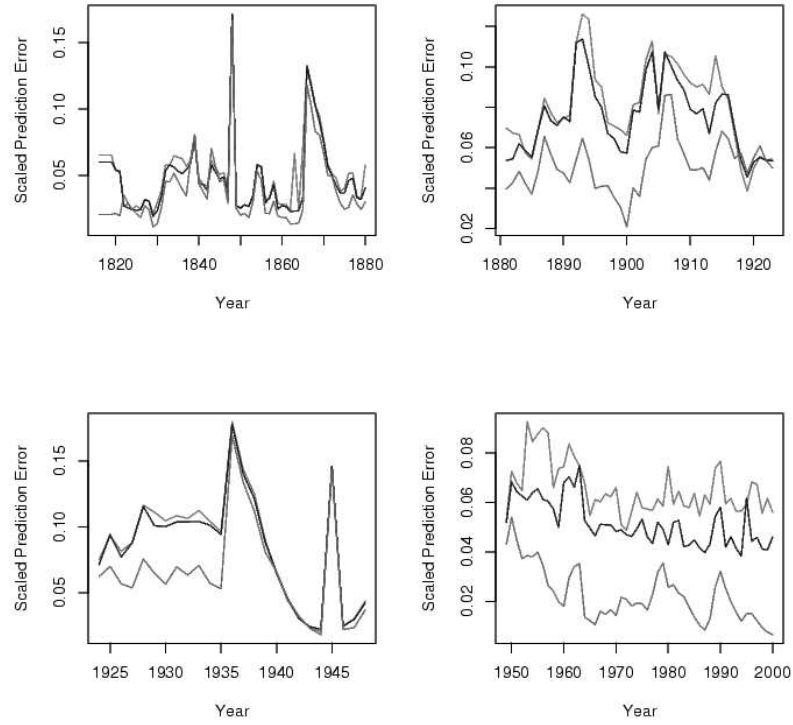


Figure 10: Prediction errors for clusters of  $K = 4, 5$  and no clustering ( $K = n$ ), with a window width of  $W = 10$  and  $d = 5$  for four distinct time periods.

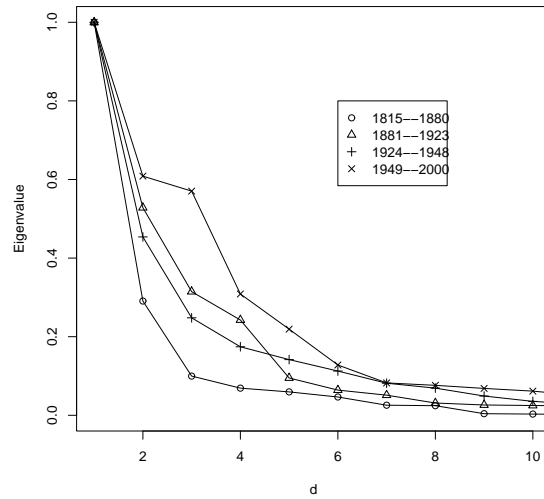


Figure 11: Scree plots for the different time periods, using the average of the graphs within each region.

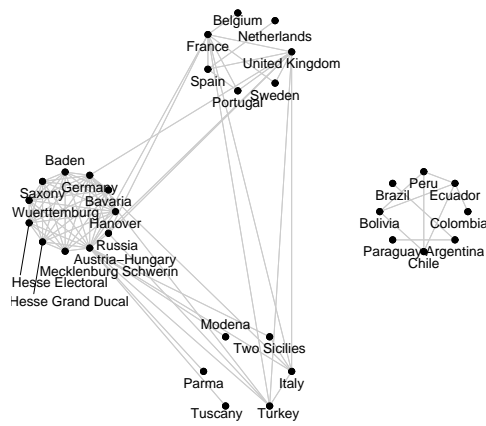


Figure 12: Graph defined by averaging all the graphs in the time period from 1816-1880.

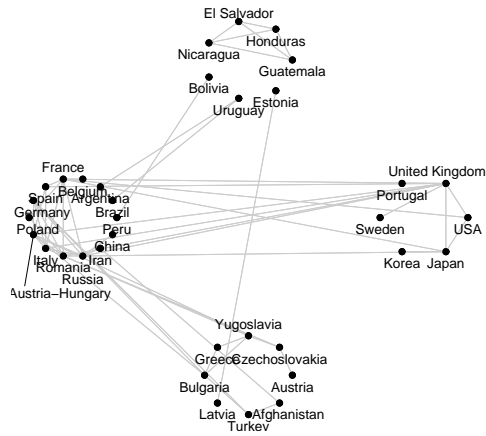


Figure 13: Graph defined by averaging all the graphs in the time period from 1881–1923.

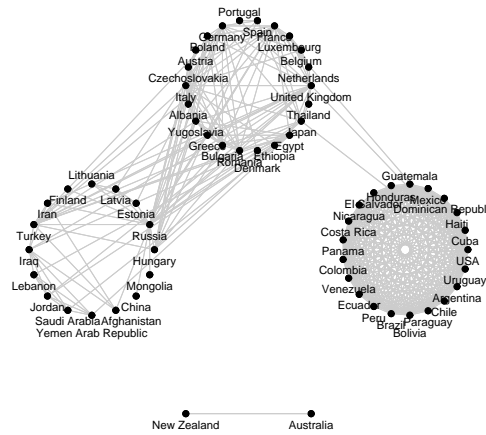


Figure 14: Graph defined by averaging all the graphs in the time period from 1923–1948.

Table 2: Countries in the clusters depicted in Figure 15. The clusters are counted anti-clockwise from the upper left corner of the plot.

GDR	Malta	Cape Verde	USA	Canada
Poland	Mauritania	Guinea-Bissau	Bahamas	UK
Hungary	Uganda	Gambia	Cuba	Netherlands
Czechoslovakia	Somalia	Mali	Haiti	Belgium
Albania	Djibouti	Senegal	Dominican Rep.	Luxembourg
Croatia	Morocco	Benin	Jamaica	France
Bosnia	Algeria	Niger	Trinidad	Spain
Slovenia	Tunisia	Ivory Coast	Barbados	Portugal
Bulgaria	Libya	Guinea	Dominica	Germany
Moldova	Sudan	Burkina Faso	Grenada	German FR
Romania	Iraq	Liberia	St. Lucia	Czech Rep.
Russia	Egypt	Sierra Leone	St. Vincent	Italy
Ukraine	Syria	Ghana	Antigua/Barbuda	Yugoslavia
Belarus	Lebanon	Togo	St. Kitts	Greece
Armenia	Jordan	Cameroon	Mexico	Cyprus
Georgia	Saudi Arabia	Nigeria	Belize	Norway
Azerbaijan	Yemen AR	Central Af. Rep.	Guatemala	Denmark
Finland	Yemen	Chad	Honduras	Iceland
Kenya	Yemen PR	Congo	El Salvador	Gabon
Tanzania	Kuwait	Dem. Rep. Congo	Nicaragua	Mauritius
Ethiopia	Bahrain	Burundi	Costa Rica	Iran
Angola	Qatar	Rwanda	Panama	Turkey
Mozambique	UAE	Madagascar	Colombia	Taiwan
Zambia	Oman		Venezuela	South Korea
South Africa			Guyana	Japan
Namibia			Suriname	Pakistan
Swaziland			Ecuador	Thailand
Afghanistan			Peru	Malaysia
Turkmenistan			Brazil	Philippines
Tajikistan			Bolivia	Indonesia
Kyrgyzstan			Paraguay	Australia
Uzbekistan			Chile	New Zealand
Kazakhstan			Argentina	
China			Uruguay	
Mongolia				
North Korea				
India				
Bangladesh				
Myanmar				
Cambodia				
Vietnam				

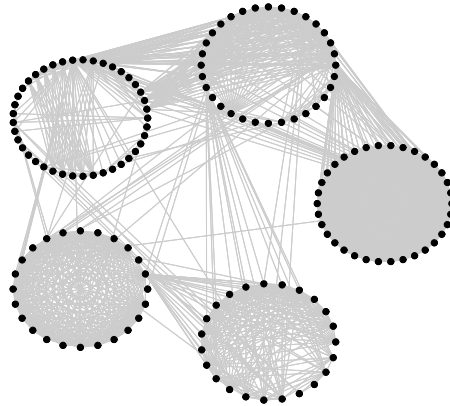


Figure 15: Graph defined by averaging all the graphs in the time period from 1949–2000. The labels for the nodes are provided in Table 2.

automatically group the vertices and provides a projection of the graph into a  $d$ -dimensional space. We have discussed methods for selecting the dimension and the number of groups.

Using multiple criteria to assess the model fit helps the user investigate different aspects of the models. This provides several different views of the graphs, and can provide insight into how changes in the graph structure can be reflected in changes in the appropriate model.

We considered only a simple time series model, based on windowing the data and assuming short-term stationarity. One could incorporate more sophisticated models, either by putting a Markov assumption on the probability matrix, and using this model prior to the RDPG, or by putting a time series model on the vectors. We believe that both of these approaches could bear interesting fruit.

## References

- Béla Bollobás. *Random Graphs*. Cambridge University Press, Cambridge, second edition, 2001.
- D. M. Gibler and M. Sarkees. Measuring alliances: the correlates of war

- formal interstate alliance data set, 1816-2000. *Journal of Peace Research*, 41:211–222, 2004.
- Peter D. Hoff. Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295, 2005.
- Peter D. Hoff, Adrien E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *JASA*, 97:1090–1098, 2002.
- M. Karonski, K. Singer, and E. Scheinerman. Random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159, 1999.
- Miro Kraetzl, Christine Nickel, and Edward Scheinerman. Random dot product graphs: a model for social networks, 2007. submitted.
- Krzysztof Nowicki and Tom A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 2001.
- Edward Scheinerman. Random dot product graphs, 2005. Talk given at IPAM, [www.ipam.ucla.edu/schedule.aspx?pc=gss2005](http://www.ipam.ucla.edu/schedule.aspx?pc=gss2005).
- Edward Scheinerman and Kimberly Tucker. Modeling graphs using dot product representations. *Computational Statistics*, 2007.
- Tom A. B. Snijders. *Models for Longitudinal Network Data*, pages 215–247. 2005.
- Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK, 1994.
- Mu Zhu and Ali Ghodsi. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics and Data Analysis*, 51:918–930, 2006.