Dynamic network data exploration through semi-supervised functional embedding

Alexei Pozdnoukhov
National Centre for Geocomputation
National University of Ireland Maynooth
Maynooth, Co. Kildare, Ireland
Alexei.Pozdnoukhov@nuim.ie

ABSTRACT

The paper presents a framework for semi-supervised non-linear embedding methods useful for exploratory analysis and visualization of spatio-temporal network data. The method provides a functional embedding based on a neural network optimizing the graph-based cost function. It exploits an online stochastic gradient descent which, avoiding the costly matrix computations and the out-of-sample problem, makes it naturally applicable for large-scale dynamic spatio-temporal problems. The semi-supervised schemes are introduced to guide the method with precisely defined locations, pairwise distances or norms of the selected data samples in the embedded space. The method is useful for exploring the complex fully dynamic networks with a variable number of geo-referenced nodes and evolving edges. The approach is illustrated with a case study devoted to the real-time embedding of the geo-referenced data on instant messaging on the internet.

Categories and Subject Descriptors

I.5 [**Computing Methodologies**]: Pattern Recognition—*Neural nets*; H.2.8 [**Database Management**]: Database Applications—*data mining, mining methods and algorithms, interactive data exploration and discovery*

General Terms

ALGORITHMS

Keywords

functional embedding, manifold learning, social networks, complex systems

1. INTRODUCTION

Understanding the evolution of the interaction between spatial and social phenomena is a challenging research direction providing interesting findings. For example, it was early noticed that the density of the flow of telephone calls is a good measure of the relationships binding together the economic interests of the region [13]. The way individuals interact can be described by the analysis of complex communication networks [25] as the data at human level is becoming available through telecommunication companies. The emergence of large datasets of emails and phone calls have recently make it possible to apply computational methods at a fine scale never accessible before for social science research [20].

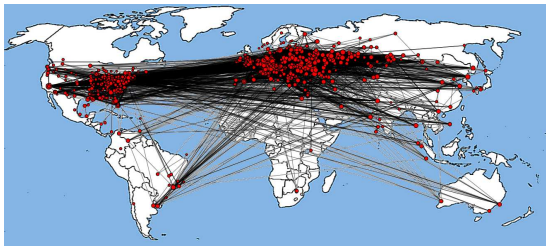

Figure 1: Instant messaging on the internet: a snapshot of new connections established during *one minute*. How does it change the space we live in? How does it evolve in time?

While the structure and dynamics of complex networks has already been widely studied [31, 2, 23], the use of geographical information brings some new perspectives to the field. It makes it possible to study the spatially-related properties of social interactions both at individual and spatially aggregated levels such as inter-city communications. Up to now, the research efforts were mainly concentrated on describing the useful statistics of these systems such as the spatial length of the communication links. For example, using a mobile phone communication dataset it was observed that the probability for a call between two people decreases by the square of their distance [19]. Recently studied larger-scale features of social networks where people are aggregated based on their geographical proximity also revealed gravitational law: the communication intensity between two cities appeared to be proportional to the product of their sizes divided by the square of distance [18]. Concerning instant messaging, an extensive descriptive study of the world-wide geography of users locations and communication between the countries is given in [21]. These works deal with static data aggregated over a significant period of time.

The temporal dynamics of communication and social net-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ACM GIS '09 Seattle, WA, USA
Copyright 2009 ACM 978-1-60558-649-6/09/11 ...$10.00.

works and their geospatial properties is far from being understood. Figure 1 presents an example of the communication network of short text instant messaging on the internet [1] shown by the geographic location of the parties involved in communication aggregated at the level of cities. The nodes of the network in Figure 1 are the locations of the initiators and recipients of the new communication sessions established during *one minute* of observation. Every minute tens of thousands of connections are being initiated and received from ever changing geographical locations. This system can be described as a fully dynamic network over a geographical space with a variable and evolving number of nodes and weighted edges. The analysis of such systems demands novel analytical and computational methods.

1.1 Embedding for dimensionality reduction

Non-linear embedding methods are becoming increasingly popular approaches for low-dimensional representation of data within many scientific domains. A number of powerful algorithms were proposed recently [29, 26, 3] and together with many extensions [9, 16, 4] the emerged field is referred to as manifold learning. The methods are of particular interest in machine learning, where they underpin both unsupervised and semi-supervised learning methods [7]. While the embedding of the metric spaces into a vector space is a common problem to consider in this framework, machine learning problems are mainly concerned with non-linear dimensionality reduction, that is a problem of reducing the dimensionality of a high-dimensional vector space of data to a low-dimensional space useful for visualization and exploratory analysis. A related problem of clustering have been intensely studied as graph-based approaches proven their usefulness on many real-life case studies [27, 24].

Given a dataset of samples which we consider as a set of vectors $\{x_i\}_{i=1,\dots\ell}$, in R^N, $N \gg 1$ the task of dimensionality reduction is to infer a mapping $x_i \to f_i, i = 1, \dots \ell$, $f \in R^M$, $M \ll N$, preserving some desirable properties of the data such as local pair-wise proximity and neighbourhood relations or distances. The general view on this problem used in manifold learning is based on graph-based representation of data. The samples are considered as a set of nodes of a graph G, with the vertices of G assigned according to the proximity of nodes in R^N. The general assumption here is that the data samples lie on a lower-dimensional (Riemannian) manifold in R^N, and the Euclidean distance in R^N is a proper similarity measure only locally. There are two basic approaches to form an adjacency matrix $W = \{W_{ij}\}$. In the first approach, one links the k-nearest neighbors, that is, $W_{ij} = 1$ if x_i and x_j are amongst the (symmetrised) k-nearest neighbors of each other and $W_{ij} = 0$ otherwise. Another approach is to consider samples linked if they are within an ϵ sphere, $W_{ij} = 1$ if $||x_i - x_j|| < \epsilon$ and $W_{ij} = 0$ otherwise. By defining these *local* similarities one can formulate the *global* optimization problem involving either affinity matrix or graph Laplacian and some constraints eliminating the translational and rotational degrees of freedom to solve for a stable embedding.

The adopted graph-based representation of data brings a common drawback to these methods. First, these methods necessarily involve matrix computations and, second, they are prone to the out-of-sample problem [4]. It restricts their usage for large scale problems and dynamic data analysis.

In this paper, we follow [14] and [32] to approach the embedding problem by introducing *functional embedding*. The paper first introduces the necessary theoretical setting and cost functions of graph-based embeddings in Section 2. It is then followed by extending this framework to account for semi-supervised embeddings useful when prior knowledge on the underlying data structure is available in the form of implicit information on the samples $\{x_k|\ x_k \to f_k^*\}$ or pair-wise relations of several kinds. Section 4 presents the formulation of the neural network based functional embedding solved by training the network with stochastic gradient descent. The approach is illustrated by an application in Section 5 where we embed the world geographic space enriched with socio-economic parameters into a 2-dimensional map enforced by the worldwide information exchange observed through real-time data on communications through instant messaging on the internet. The system processes up to 20000 newly established communications every 5 minutes resulting in 5.67 million aggregated samples of roughly 500 million individual communications daily. In Section 5.6 we show that the observed communication network is scale-free, discuss the implications of this in Section 6, and summarize the paper with the conclusions.

2. GRAPH-BASED MANIFOLD LEARNING

A central object to many algorithms which embed the dataset $\{x_i\},\ i = 1, \dots \ell$ with known pair-wise similarities W is the (generalized) graph Laplacian L, a matrix defined as:

$$L = D - W,$$

where D is a diaganal matrix of node degrees, $D_i = \sum_j W_{ij}$, and W is as defined above. For a connected graph, L has a single zero eigenvalue associated with the uniform vector $e = [1, 1, \dots 1]^{\mathrm{T}}$.

A low dimensional embedding can be given by the functions $f : V \to R$ defined on the vertices of the graph. These can be obtained by considering the minimum of the following cost function:

$$f^T L f = \frac{1}{2} \sum_{i,j} (f_i - f_j)^2 W_{ij}, \tag{1}$$

under the constraints $f^T f = 1$ and $f^T e = 0$ accounting for a freedom in scaling and translation correspondingly. The eigenvectors of L with smallest non-zero eigenvalues are the solutions to this problem [10]. These solutions can then be used as the embedding or followed by a post-processing step to obtain data clustering [24]. Different weights W_{ij} would lead to the different algorithms. A popular choice of $W_{ij} = e^{-\gamma|x_i - x_j|^2}$ is related to a diffusion process on the graph and results in the smooth functions f providing the embedding representation known as Laplacian eigenmaps [3].

2.1 Functional embedding and loss functions

By finding a functional embedding we mean the following rather general problem. For a given dataset $x_i, \dots x_\ell$ find a parametric function $f : R^N \to R^M$ defined by its parameters $\alpha \in \Lambda$ such that the functional

$$R(\alpha) = \sum_{i,j=1}^{\ell} L(f(x_i, \alpha), f(x_j, \alpha), W_{ij}), \tag{2}$$

is minimized[1] with respect to α. L is a loss function defined for a *pair* of examples, and a matrix of weights W is defined by either approaches described above or by prior knowledge on the problem at hand. The set of admissible functions is defined by the set of parameters Λ and has to be rich enough in order to learn complex non-linear transformations providing embeddings. We will use a multi-layer neural network of a Siamese architecture for this task. To simplify the notations, we denote the $f(x_i, \alpha)$ simply as f_i.

The formulation in terms of a loss function minimization includes many popular methods into a convenient framework.

- Multidimensional scaling (MDS) and ISOMAP. The loss function for MDS is

$$L(f_i, f_j, W_{ij}) = (||f_i - f_j|| - W_{ij})^2. \quad (3)$$

It is a popular algorithm which tries to preserve the distances between points in the best possible way. Euclidean pair-wise distances between samples are used in its classical linear form [8]. It gave rise to well-developed approaches for graph drawing [11]. If one assumes the data lie on a low-dimensional manifold and indicates a way to compute geodesic distances therein, one arrives at the ISOMAP method for dimensionality reduction [29].

- Laplacian eigenmaps [3]. In this method, one preserves local distances by introducing the following cost function

$$L(f_i, f_j, W_{ij}) = W_{ij}(||f_i - f_j||)^2, \quad (4)$$

which leads to the optimization problem (1) which, being solved without balancing constraints is prone to a trivial solution $f(x) = const$. Indeed, any $f(x) = const$ puts (4) to zero and minimizes (2). The need to resolve this problem provokes the following modification of the loss function.

- DrLIM and Siamese networks. The loss function for DrLIM as proposed in [14] introduces the margin m which defines a radius around $f(x_i)$ such that the dissimilar points contribute to the loss function only if their distance is within this radius:

$$L(f_i, f_j, W_{ij}) = \begin{cases} ||f_i - f_j||^2, \text{if} \quad W_{ij} = 1 \\ \max(0, m - ||f_i - f_j||^2), \text{if } W_{ij} = 0 \end{cases} \quad (5)$$

The first term of the loss pulls the similar samples closer. The contrastive term pushing non-neighbour samples apart is very important as it prevents the method from a trivial collapsed solution $f(x) = const$. As this is achieved without any balancing constraints, it is much easier to optimize by gradient descent. Notice that no global knowledge (such as the pair-wise geodesic distances) is required here as only the local neighborhood relations are used. There is an interpretation of this method in terms of the energy-based learning approach [14] closely related to spring models. One can show that minimizing the global loss (5) one drives the spring system into an equilibrium state thus justifying the convergence of the optimization of the total loss.

[1] A balancing constraint may be required to avoid trivial solutions to this problem.

3. SEMI-SUPERVISED EMBEDDING

Sometimes (for better visual perception, for instance), a low-dimensional embedding might need to follow some desired properties to enhance the visualization and at the same time to satisfy the baseline criteria encoded with a loss function. In this case, it can be approached in a semi-supervised manner. We present here three extensions to the baseline algorithm by adding a semi-supervised term to an unsupervised loss function L^U.

First, the exact location for some samples might need to be preserved in the embedded space. That is, for a subset of samples $\{x_k\}, k = 1, ..., K$ one may require the embedding to satisfy $f(x_k, \alpha) = f_k^*$ for some given values of f_k^*. This can be achieved (in a mean square sense) by optimizing the following cost function with some trade-off constant η:

$$L^{SS} = L^U + \eta||f(x_k, \alpha) - f_k^*||^2. \quad (6)$$

Second, the pair-wise distances between pairs of samples may need to be preserved in the embedded space. That is, for a subset of samples $\{x_k, x_m\}_{m,k=1,...,M,K}$ one may require the embedding to satisfy $||f(x_k, \alpha) - f(x_m, \alpha)||^2 = {d^2}_{mk}$. This can be achieved (in a mean square sense) by optimizing the following cost function with some trade-off constant η:

$$L^{SS} = L^U + \eta(||f(x_k, \alpha) - f(x_m, \alpha)|| - d_{mk})^2. \quad (7)$$

Third, one may require the exact distances between a "central" node x_k and some or all of its "neighbours". This can be done by controlling the norm of the difference vectors by introducing the cost function

$$L^{SS} = L^U + \eta \sum_{m \in N(x_k)} (||f_k - f_m|| - d)^2. \quad (8)$$

A variety of combinations of the provided cost functions is possible. However, attention has to be paid when using this as enforcing the "hard" constraints may lead to an overfitted embedding that is not representative and misleading in respect to the inherent relations within the data.

4. TRAINING

The set of functions $f(x, \alpha)$ providing the embedding is implemented with a multi-layered neural network. The architecture consists of two copies of the network $f(x)$ sharing the same weights $\{\alpha\}$, known as a Siamese network [6]. Each network in this pair is a classical multi-layer perceptron with several layers of hidden neurons connected through a non-linear transfer function and concluded by a linear output neuron. The input of the network receives the pairs of the training samples $\{x_i, x_j\}$ and propagates them to the intermediate output $\{f_i, f_j\}$ which is in turn convolved with a desired measure depending on the selected loss function. The distance $||f_i - f_j||$ is a natural choice as it is used in most of the loss functions. This output is then used to compute the scalar loss depending on the value of W_{ij}. The weights of the network α are then updated with a stochastic gradient descent [5]. The gradient is computed by back-propagation through the loss, the distance convolution layer, and the hidden layers of the two instances of the Siamese part. The weights are updated at the forward pass following the sum of the contributions by each sample of the training pair respectively.

The algorithm minimizing the margin-based loss (5) is summarized below. It was implemented in the Torch5 machine learning toolbox [30].

Algorithm 1 Functional embedding.

Input Data:
Set of pairs $\{x_i, x_j\}$ and non-pairs $\{{x_i}^n, {x_j}^n\}$.
if Semi-supervised mode (SSL) **then**
  Set of labeled points $\{x_k \rightarrow f_k\}$
  and/or
  Pairs with desired distances d_{mk} such that for $\{x_k, x_m\}$, $||f(x_k, \alpha) - f(x_m, \alpha)||^2 = {d^2}_{mk}$.
end if
Input Parameters:
Learning rate $\lambda > 0$ and margin $m > 0$, see Eq. (5).
if Sem-supervised mode (SSL) **then**
  SSL loss coefficient $0 < \eta < 1$, see Eqs. (6)-(8).
end if

repeat
  if SSL **then**
    Select a random labeled sample $\{x_k \rightarrow f_k^*\}$ *and/or* pair $\{x_k, x_m\}$.
    Update network weights $\{\alpha\}$ w.r.t. $\eta\lambda\nabla L^{SS}$.
  end if
  Select a random pair of neighbours $\{x_i, x_j\}$,
    update network weights $\{\alpha\}$ w.r.t. $\lambda\nabla L(x_i, x_j, 1)$.
  Select a random pair of non-neighbours $\{{x_i}^n, {x_j}^n\}$,
    update network weights $\{\alpha\}$ w.r.t. $\lambda\nabla L(x_i, x_j, 0)$.
until convergence criteria met

Output: neural network from a Siamese pair implementing $f_i = f(x_i, \alpha^*)$.

5. INSTANT MESSAGING ON THE INTERNET

Instant messengers (IM) for on-line communication on the Internet have become extremely popular [21, 28]. In this study we explore the data on America Online instant messenger (AIM), one of the most widely used IM (with over 100 million users worldwide) [1]. The data on communications initiated during the last minute is available in real time as a web feed provided by AOL. The spatial locations of the users who initiated a communication and their respondents are available with a precision at the city level, as determined by geocoding the IP addresses. No private information is available as only the relatively large cities where personalization is impossible are included in the provided report. The data crawler implemented for this case study regularly requests the webservice for communications registered within every country. On average, we collect up to 20000 new connections (edges) from varying locations (nodes) registered during past five minutes. The raw data samples x_i are latitude and longitude of the initiation-recipient locations. The link is undirected and no information on the "intensity" or duration of communications is available. A sample snapshot of the data is illustrated in Figure 1.

5.1 GeoSpatial attributes

We extend the original vectors describing the locations of the nodes with socio-economic parameters of the region.

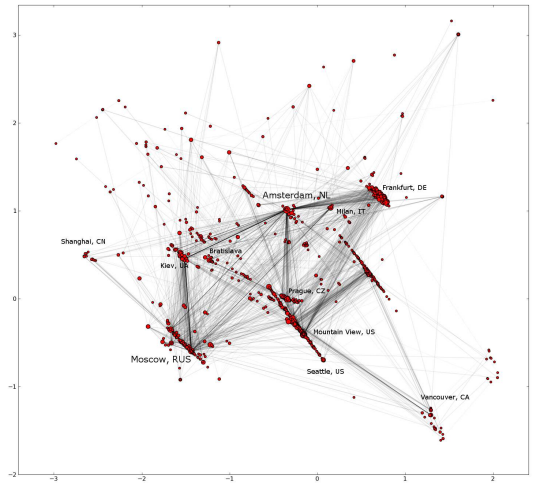

Figure 2: Unsupervised embedding of the inter-city instant messaging data at 12:00 CET in a week day of June 2009.

Table 1: The list of the geo-attributes (input features)

Name	Type	Range	Dim
Latitude	continuous	-90..90	1
Longtitude	continuous	-180..180	1
GeoID	nominal	ISO code	1
Country code	nominal	ISO code	1
Language code	indicator	[0, 1]	9
Time zone	categorical	1-24	1
Current time	continuous	hh:mm	1
Population	continuous	$[0 - 150 \times 10^6]$	1
Continent code	indicator	[0, 1]	7
Area	continuous	$[0 - 5 \times 10^6]$	1

The input features obtained from the GeoNames database [12] are listed in Table 1. The final feature vector is 25-dimensional description of the locality including the indicator variables for some major official languages spoken in the region.

5.2 Online embedding

It is quite straightforward to employ the algorithm presented in Section 4 for dynamic temporal setting. One needs to generate the training samples and feed them to the network on the fly, controlling the convergence and keeping the balance of making enough iterations and processing arriving data in real-time. In the presented case study, data preprocessing was as computationally intensive as training the neural network itself and these stages were synchronized.

The weights of the neural network are updated as the new data arrives. There is an important benefit from the selected loss function in terms of the computational time. As we employ the margin-based loss $\max(0, m - ||f_i - f_j||^2)$ for the non-linked samples which is zero for many samples not violating the margin m no modification is required to the network weights in this case. This trick can be employed also by introducing a small ϵ-insensitive region to the

Algorithm 2 On-line functional embedding.

loop
  Data: download new data block of initiated communications $\{x_i, x_j\}$.
  Preprocessing:
  Find non-pairs $\{{x_i}^n, {x_j}^n\}$.
  Generate input feature vectors (geo-attributes) for $\{x_i, x_j\}$ and $\{{x_i}^n, {x_j}^n\}$.
  (optional) Select labelled samples and/or pairs.
  (optional) Update loss function parameters m, η, and the learning rate λ.
  repeat
    Run Algorithm 1.
  until <new data uploaded>
end loop

MSE loss function of the semi-supervised part. It makes the online embedding algorithm more computationally effective and robust to noise and short-scale variations in input data.

5.3 Embedding with spatial attributes

We present here some illustrative embedding results and demonstrate the influence of semi-supervised constraints. An architecture with each Siamese network having 25 inputs, two fully connected hidden layers of 50 neurons with a hyperbolic tangent transfer function, and a linear 2-dimensional output was used. The learning rate was set as $\lambda = 0.001$, and the margin for the cost function at $m = 0.5$. Several initial iterations were made to pre-train the system before leaving it running in real time as described by the Algorithm 2. It was observed that it takes about 30 minutes for the system to enter a steady state when the output results evolve smoothly in time.

Figure 2 presents a snapshot[2] of the embedding obtained for the inter-city communication flows at 19:00 CET in a week day of June 2009. One can observe the cities arranged as implied by the instant messaging flows between them [3]. We could observe some typical configurations of information exchange for the weekdays, weekends, and a 24-hour periodicity due to the user activity variations depending on the time of the day.

5.3.1 Semi-supervised embedding

Here we show how one uses the semi-supervised embedding to manipulate the observed configuration. The fist example (Figure 3) is a configuration obtained using the loss function (6) with $\eta = 0.5$ to enforce the locations in the embedded space for Moscow, Amsterdam and Prague to [0,1], [-1,0] and [1,0] correspondingly.

In the second example shown in Figure 4 the loss function (7) was used with the pair-wise distances between Moscow, Amsterdam and Prague set as 5. The obtained configuration is a rotated version of the one observed in Figure 3 with similar relative positions of the communication nodes.

[2]The illustrations in this Section are made using the NetworkX package [15].

[3]All the interpretations provided in this paper are only illustrative examples of instant messaging flow embedding results for a particular time moment. They might be heavily influenced by unequal data sampling conditions and inhomogeneous use of the IM messaging service over the world and must not be considered as a basement for decision support of any kind.

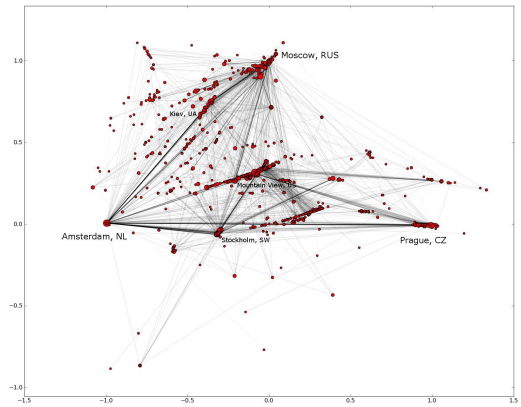

Figure 3: Semi-supervised embedding of instant messaging data at 12:00 CET. Locations for Moscow, Amsterdam and Prague are enforced with semi-supervised loss function at [0,1], [-1,0] and [1,0] correspondingly.

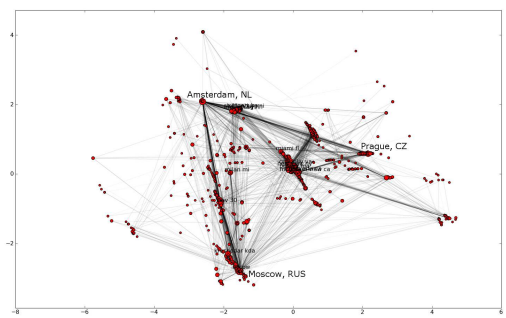

Figure 4: Semi-supervised embedding of instant messaging data at 12:00 CET. Pair-wise distances between Moscow, Amsterdam and Prague are enforced to be equal to 5 with semi-supervised loss function. Note that the obtained embedding follows the same structure as in Figure 3.

The semi-supervised constraints can be activated at the run-time to bias the current embedding to a desired configuration for better visual perception. We however observed that one should not enforce the system to follow the geometrically unachievable configuration as it slows down the convergence and increases the risk for a neural network to get stuck in a local minima. The use of visual analytics techniques for post-processing the visualization by node sizing and coloring, varying edge thickness, etc., facilitates perception.

5.4 Out-of-sample predictions

A particular feature of the proposed method is that it is possible to compute the embedding for any location, including those where instant messaging information is unavailable. One simply has to assign the attributes (Table 1) to the desired geographical location, form x^* and compute the $f(x^*)$. An interpretation for $f(x^*)$ can be given as a po-

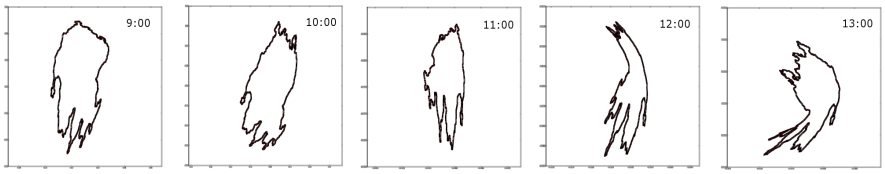

Figure 5: Temporal evolution of the border of Switzerland in the embedded space of instant messaging information flow from 9:00 CET to 13:00 CET.

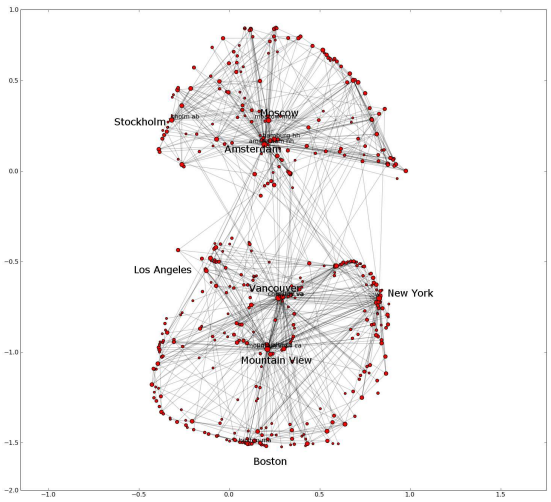

Figure 6: A sample embedding of instant messaging data at 01:00 CET (afternoon in the U.S). The information exchage is concentrated around the large cities, inter-continent communication is weak.

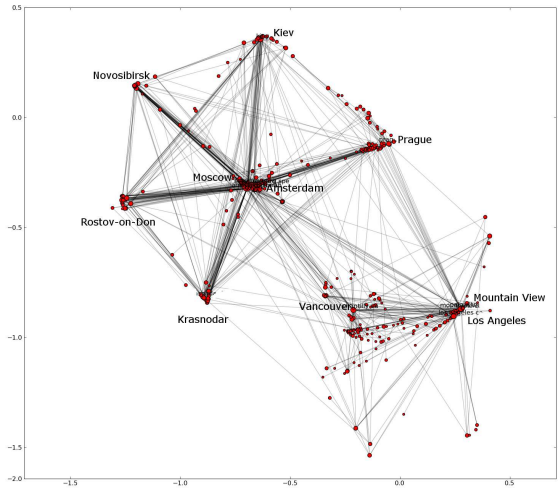

Figure 7: A sample embedding of instant messaging data at 9:00 CET (afternoon in the central Russia, morning in Europe, late night in the U.S.). Polycentric structure is observed with Amsterdam and Moscow in the middle.

sition of the geographical location x^* in the space implied by world-wide communication exchange. This can also be used for visualizing the joint location of regions or shapes of the countries in the embedded space. In Figure 5 we show the temporal variation of the polygon representing the border of Switzerland (central European location, 4 official languages) during 5 hours, from 9:00 to 13:00 in a weekday of June 2009. While it might look like an appealing property of the method, we have observed that it is often difficult to find a reasonable explanation for variations of shapes.

5.5 Embedding with spatial coordinates only

As the neural network produces a continuous function of its inputs, including many spatial attributes enforces the model to follow the similarities encoded by the input features. Consider, for example, the coding for official languages used in the country. While it was observed [21] that language is an important factor in the inter-countries textual messaging, this information does not necessarily bring adequate value to presenting the communication flows. People may chat in a foreign language, or contact relatives and friends from abroad in their mother tongue, not mentioning that considerable volume of communication on the internet is expected to be in English.

To obtain a pure data-driven embedding which only relies on the prior smoothness related to the physical space (Tobler's law) we applied the method in an unsupervised way to the input space composed on geographical locations only. It leads to the non-linear transformation of the 2D geographical space into a 2D space implied by current inter-city communication flows. This highly non-linear transformation requires a more powerful network. A neural network with 2 inputs, 100 neurons in 2 fully connected hidden layers with a hyperbolic tangent transfer functions was used. The learning rate was set as $\lambda = 0.001$, and the margin parameter at $m = 0.2$. More iterations were required to obtain a reasonable minima of the cost function. Sample snapshots of the observed embedding are presented in Figures 6 and 7. We observed different communication modes to appear, with a distinct separation between American and European continents and several typical polycentric or monocentric structures within. While time is not included to the inputs explicitly, the 24-hour cycles related to the changing users activity depending on the time of the day were clearly observed.

5.6 Scale-free properties

It was observed that the instant messaging networks at the level of individual users are scale-free [28, 21]. In our

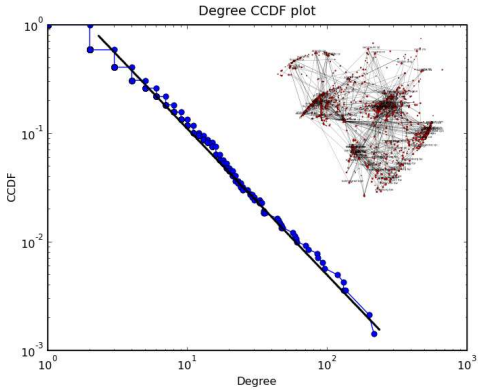

Figure 8: Cumulative node degree distribution for a particular time moment. Power law with an exponent of 2.3 is observed. A sample embedding of the network is presented in the top right corner.

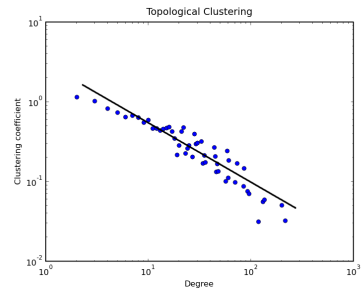

Figure 9: Average clustering coefficient follows a power law with an exponent of 0.77.

study, the nodes of the network are spatially aggregated at the city level, though a similar network structure is observed for every time moment, as illustrated in Figure 8 and 9. The exponent of the power law $p(k) \sim k^{-\gamma}$ of node degree distribution is $\gamma = 2.3$, and the clustering coefficient follows the power law with an exponent of $\gamma_c = 0.77$ which is rather typical for a landline communication network but higher then observed before [21] probably due to the aggregated highly clustered high degree intra-city communications. The power law can be explained by some kind of the preferential attachment mechanism [2]. It can be a direct consequence of the Zipf's law for population of cities, where the physical space may act as a critical structuring force for the popularity of a particular messaging service and its spread over the space. As the data source might be prone to preferential sampling when reporting the newly established connections[4], we can not make any rigor conclusions on this. What is important to note is that the proposed functional embedding method appeared to be capable of processing, visualizing, and modelling the dynamics (to the some data-driven extent, by providing the non-parametric mapping $x \rightarrow f(x)$) of a geo-referenced scale-free network.

6. DISCUSSION

Understanding the dynamical evolution of network processes and their relation to the physical space is crucial in the study of many complex systems. It is not straightforward to adopt the conventional intuitions for data visualization used in the graphical layout of simple networks [11] if the networks at hand are scale-free [17].

This paper provides a novel exploratory tool to investigate the evolution of large-scale networks in real time. Particularly, from the geo-referenced instant messaging communications one could observe the influence of geography on social networks, particularly how and when people communicate, and how the cities align in the space implied by these information flows, how different configurations emerge and evolve. However, some important factors were not available for this study, as, for example, the duration of the communication which is known to depend on distance [21, 18]. It was possible to observe some typical configurations and patterns related to 24h periodicity, but no quantitative description is available at the moment to analyse these changes. The graphical layout methods for small scale dynamic network visualization are well developed in the social network research [22], and can be combined with the data-driven large-scale methods as the one presented in this paper. The analysis of social networks overlaid on physical space at different spatial scales could help to improve our understanding of the structure and interaction of such systems.

The method described in this paper provides a neural network based functional embedding. Most of the common drawbacks of the neural network methods can be attributed to the presented approach including the problem of multiple local minima and the "black-box style" modelling. Moreover, some of the problems such as the choice of the network structure, the number of neurons, the optimal values for learning rates and cost function parameters can not be guided by a simple criteria such as cross-validation error since the problem is unsupervised. At the same time, it brings a flexibility to the method use which we illustrated within a semi-supervised setting.

7. CONCLUSIONS

This paper presented a functional embedding method useful for an exploratory analysis of the spatio-temporal evolution of complex networks of nodes with a set of vector attributes. The method is applicable for large-scale dynamic tasks. It provides a functional embedding of the attribute space of nodes implied by their inter-connections. It is achieved by training a neural network to minimize the cost functions originating from the modern graph-based manifold learning methods. The case study illustrated how the method can be used to explore the dynamics of the geo-referenced scale-free network of instant message communications.

8. ACKNOWLEDGMENTS

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) and Stokes Lectureship award by Science Foundation Ireland under the National Development Plan. The author gratefully acknowledge this support. The author would also like to thank Fergal Walsh for his help in correcting the text of the paper.

[4]For example, we did not observe Spain as a significant hub in South America to Europe communications as in [21].

9. REFERENCES

[1] AOL instant messenger. http://www.aim.com/.

[2] A.-L. Barabasi. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Plume Books, April 2003.

[3] M. Belkin and P.Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2003.

[4] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering. In *In Advances in Neural Information Processing Systems*, pages 177–184. MIT Press, 2003.

[5] L. Bottou and Y. LeCun. Large scale online learning. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[6] J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, and R. Shah. Signature verification using a siamese time delay neural network. In J. Cowan and G. Tesauro, editors, *Advances in Neural Information Processing Systems*, volume 6. Morgan Kaufmann, 1993.

[7] O. Chapelle, B. Schölkopf, and A. Zien, editors. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006.

[8] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman and Hall, London, 1994.

[9] V. de Silva and J. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. 2003.

[10] M. Fiedler. A property of eigenvectors of non-negative symmetric matrices and its applications to graph theory. *Czechoslovak Math Journal*, 100(25):619–633, 1975.

[11] E. R. Gansner, Y. Koren, and S. C. North. Graph drawing by stress majorization. In *Graph Drawing*, pages 239–250, 2004.

[12] GeoNames. The geonames geographical database. http://www.geonames.org/.

[13] J. Gottman. Megalopolis: the urbanization of the northeastern seaboard of the united states. *Economic Geography*, 7(33):31–40, 1957.

[14] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *Proceedings of Computer Vision and Pattern Recognition Conference, CVPR'06, IEEE press*, 2006.

[15] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, Aug. 2008.

[16] O. C. Jenkins and M. Mataric. A spatio-temporal extension to isomap nonlinear dimension reduction. In *Proceedings of the 21 st International Conference on Machine Learning, ICML'04, Banff, Canada*, 2004.

[17] Y. Jia, J. Hoberock, M. Garland, and J. Hart. On the visualization of social and other scale-free networks. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1285–1292, 2008.

[18] G. Krings, F. Calabrese, C. Ratti, and V. Blondel. Urban gravity: a model for intercity telecommunication flows. *arXiv:0905.0692v2 [physics.soc-ph]*, June 2009.

[19] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325, September 2008.

[20] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Social science: Computational social science. *Science*, 323(5915):721–723, February 2009.

[21] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 915–924, New York, NY, USA, 2008. ACM.

[22] J. Moody, D. Mcfarland, and S. Bender-Demoll. Dynamic network visualisation. *AJS*, 110(4):1206–1241, January 2005.

[23] M. Newman, A.-L. Barabasi, and D. J. Watts. *The Structure and Dynamics of Networks*. (Princeton Studies in Complexity). Princeton University Press, April 2006.

[24] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849–856. MIT Press, 2001.

[25] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, A. M. de Menezes, K. Kaski, A.-L. Barabasi, and J. Kertesz. Analysis of a large-scale weighted network of one-to-one human communication. Feb 2007.

[26] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[27] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 1997.

[28] R. D. Smith. Instant messaging as a scale-free network. arxiv:cond-mat/0206378v2 [cond-mat.stat-mech], Aug 2002.

[29] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[30] Torch5. A fast and efficient library for machine learning algorithms. http://torch5.sourceforge.net/.

[31] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[32] J. Weston, F. Ratle, and R. Collobert. Deep learning via semi-supervised embedding. In *Proceedings of the 25th International Conference on Machine Learning, ICML'08, Helsinki, Finland*, 2008.